Physically-Based Models for the Analysis of Raman Spectra

Physikalisch Basierte Modelle zur Auswertung von Raman Spektren

Von der Fakultät für Maschinenwesen der
Rheinisch-Westfälischen Technischen Hochschule Aachen
zur Erlangung des akademischen Grades eines Doktors
der Ingenieurwissenschaften genehmigte Dissertation

vorgelegt von

Peter Beumers

Berichter: Univ.-Prof. Dr.-Ing. André Bardow
Univ.-Prof. Dr.-Ing. habil. Andreas Siegfried Bräuer

Tag der mündlichen Prüfung: 23. Januar 2019

Aachener Beiträge zur Technischen Thermodynamik Band 23

Peter Beumers
Physically-Based Models for the Analysis of Raman Spectra

Physikalisch Basierte Modelle zur Auswertung von Raman Spektren

ISBN: 978-3-95886-319-4

Bibliografische Information der Deutschen Bibliothek
Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.ddb.de abrufbar.

Herstellung & Vertrieb:

1. Auflage 2019
© Wissenschaftsverlag Mainz GmbH - Aachen
Süsterfeldstr. 83, 52072 Aachen
Tel. 0241/87 34 34
Fax 0241/87 55 77
www.Verlag-Mainz.de

ISSN: 2198-4832

Satz: nach Druckvorlage des Autors
Umschlaggestaltung: Druckerei Mainz

printed in Germany
D82 (Diss. RWTH Aachen University, 2019)

Vorwort

Die vorliegende Arbeit entstand am Lehrstuhl für Technische Thermodynamik der RWTH Aachen im Rahmen meiner Tätigkeit als wissenschaftlicher Mitarbeiter.

An erster Stelle möchte ich meinem Doktorvater Prof. André Bardow für das entgegengebrachte Vertrauen, die Möglichkeit diese Arbeit anzufertigen und die fachliche Betreuung danken. Des Weiteren danke ich Prof. Andreas Bräuer für die Übernahme des Korreferats und Prof. Alexander Mitsos für die Übernahme des Prüfungsvorsitzes und die entspannte Leitung der Prüfung.

Von Herzen Danke ich allen Mitarbeitern des LTT: für die witzige Zeit und die großartige Unterstützung. Ohne Euch wäre ich nicht so gerne zur Arbeit am LTT gegangen!

Ein besonderer Dank gilt allen Mitgliedern meiner Arbeitsgruppe, der OMT. Es war großartig mit Euch zusammenzuarbeiten. Vielen Dank meinem Bürokollegen Ludger Wolff: für die entspannte Atmosphäre, den ein oder anderen fachlichen Tipp zum Thema numerische Optimierung und natürlich für das Ertragen von Kaffeetassen mit stattlicher Patina. Vielen Dank an Dominique Dechambre, Christine Peters und Julia Thien: für die Feierabendbierchen, für viele interessante Diskussion und dafür, dass es nie langweilig war. Ein weiterer Dank an die vierte Etage und im Besonderen Bastian Liebergesell und dem Mastermind der Weihnachtsfeiern: Carsten Flake. Vielen Dank meinen Gruppenleitern Christoph Pauls, Thorsten Brands und Hans-Jürgen Koss für die geduldige fachliche Unterstützung und die angenehme Arbeitsatmosphäre.

Nicht zuletzt Danke ich meine Eltern, die meinen Drang zu technischen Fragestellungen schon gefördert haben, als ich im Kinderzimmer meine Spielzeugautos auseinandergeschraubt habe und (angeblich!) nie wieder zusammengesetzt habe.

Leonie, vielen Dank für deine Geduld, dein Verständnis und deine Unterstützung in den Letzten Jahren. Ich freue mich auf ein Leben mit Dir.

Peter Beumers

Contents

List of Figures

List of Tables

Notation

Abbreviations

EtOH	Ethanol
FD-IHM	First-Derivative Indirect Hard Modeling
IHM	Indirect Hard Modeling
MAXPE	Maximum Absolute Prediction Error
MEA	Monoethanolamine
MeOH	Methanol
PLS	Partial Least Squares
PCR	Principle Component Regression
RMSE	Root-Mean-Square Error
VHM	Very Hard Modeling

Greek symbols

α	peak area
α_{CO2}	CO_2-loading
α_{sum}	peak sum
β	gaussian part
γ	peak position
δ	peak width (FWHM)
Ω	solid angle
λ	wavelength
ν	frequency
$\tilde{\nu}$	wavenumber
$\tilde{\nu}_m$	wavenumber of a discrete pixel m
$\tilde{\nu}_m^{\Delta}$	shifted wavenumber of a discrete pixel m
ω	relative weight
ζ	(absolute) weight
Θ_B	parameter set of a background model
Θ_{F}	stretching factor
$\boldsymbol{\Theta}_k^*$	parameter set of pure component model k
$\boldsymbol{\Theta}_k$	adjustable parameter set of component model k
$\boldsymbol{\Theta}_k'$	fixed parameter set of component model k
Θ_{S}	global shift parameter

Latin symbols

A	absorption in the medium
c	speed of light
c	concentration
E	energy
G	Gaussian profile
h	Planck constant
I	intensity
I_0	laser intensity
J	molar scattering parameter
k	calibration coefficient
K	overall spectrometer response
K_{ij}	binary calibration coefficient
L	Lorentzian profile
M	molar mass
P	polynomial coefficients for parametrized peak changes
$n_{\tilde{\nu}}$	number of discrete pixels in a spectrum
n_{calib}	number of calibration spectra
n_{C}	number of components
n_{m}	number of mixture spectra
n_{peaks}	number of peaks
n_{poly}	polynomial degree
$n_{\mathrm{refractive}}$	refractive index
R	residual
S	Raman spectrum (Raman signal)
t_{exc}	excitation time
V	pseudo-Voigt profile
w	mass fraction
x	mole fraction

General subscripts and superscripts

*	pure substance parameter
0	before the reaction
˜	uncorrected variable
calib	calibration
pred	predicted composition
sum	sum norm
true	true composition

Kurzfassung

Die Spektroskopie hat in den letzten Jahren sowohl für wissenschaftliche Fragestellungen als auch im industriellen Einsatz immer mehr an Bedeutung gewonnen, um Gemisch-Zusammensetzungen quantitativ zu bestimmen.

Mit der zunehmenden Verbreitung stellt sich die Frage, wie die gewonnenen Daten optimal genutzt werden können. Neben statistischen Verfahren wie Partial Least Squares (PLS) hat sich die Methode des Indirect Hard Modeling (IHM) zur Spektrenauswertung etabliert. IHM ist ein nichtlineares Verfahren und kann daher nichtlineare Effekte wie Peakverschiebungen besonders gut berücksichtigen. In dieser Arbeit wird IHM weiterentwickelt, um eine Anwendung auf neue Fragestellungen zu ermöglichen.

Da IHM nichtlineare Effekte schon in der Spektrenauswertung berücksichtigt, wird eine lineare Kalibrierung möglich, die die direkte Proportionalität des Raman-Signals zur Konzentration einer Komponente ausnutzt. Dieses lineare Kalibriermodell eignet sich sehr gut zur Extrapolation. Daher bietet sich IHM an, wenn es darum geht, reaktive Gemische zu untersuchen, in denen oft nur binäre Randsysteme zur Kalibration herangezogen werden können. Bisher lassen sich reaktive Systeme mit Zwischenkomponenten jedoch nur mit Hilfe von thermodynamischen Modellen kalibrieren. In der vorliegenden Arbeit wird ein Vorgehen vorgestellt, das eine Kalibration allein anhand von Reaktionsmechanismus sowie von Stöchiometrie und Elektroneutralitätsbedingung ermöglicht.

Spektrale Untergründe, z.B. Fluoreszenzen, lassen sich im Rahmen von IHM durch Spektrenvorbehandlung oder Untergrundmodelle behandeln. Allerdings sind Untergrundsignale trotzdem eine häufige Fehlerursache. Ableitungen von Spektren werden in statistischen Verfahren seit langem genutzt, um breitbandige Untergrundsignale von schmalbandigen Ramansignalen zu trennen. Daher wird IHM in dieser Arbeit so angepasst, dass sich auch die erste Ableitung eines Spektrums auswerten lässt.

In der Kalibration von IHM werden im Wesentlichen nur relative spektrale Intensitäten der beteiligten Stoffe berücksichtigt. Um die Information der Kalibrationsspektren vollständiger zu nutzen, wird in dieser Arbeit ein Verfahren gezeigt, das die Peakfreigabe von IHM durch eine feste konzentrationsabhängige Parametrisierung der nichtlinearen Effekte ersetzt.

Die üblichen Peakmodelle spiegeln die Physik in einem Detektor nur unzureichend wider. Das kann bei schmalen Peaks dazu führen, dass sich die Intensität des modellierten Peaks durch Peakverschiebung ändert. Um dieses unphysikalische Verhalten zu korrigieren, wird in dieser Arbeit ein neues Peakmodell vorgestellt, welches sich näher an der Physik im Detektor orientiert.

Abstract

In recent years, spectroscopy has developed into an increasingly valuable tool to determine the composition of mixtures; for scientific questions as well as for the industry.

The increasing use of spectroscopy raises the question how to best use the obtained data. For the analysis of spectral data, the method of Indirect Hard Modeling (IHM) has been established besides statistical methods like PLS. IHM is a nonlinear method that can therefore efficiently treat nonlinear effects such as peak-shifts. In the present work, the IHM method is expanded to increase its applicability.

IHM treats nonlinear effects in the spectral evaluation. Therefore, the direct proportionality between the concentration and the Raman signal of a component can be used for calibration. The resulting linear calibration model allows for reliable extrapolation. Thus, IHM can be used to study reactive systems, even if only binary subsystems can be used for calibration. However, thermodynamic systems with intermediates can so far only be calibrated by using thermodynamic models. In this work, a method is established that calibrates a reactive system with intermediates only based on the reaction mechanism as well as stoichiometry and electroneutrality.

Spectral backgrounds, *e.g.*, fluorescence, can be treated by a spectral pretreatment or via background models. However, spectral backgrounds are still a common source of error in IHM. Derivatives have long been used very effectively in statistical methods. Therefore, IHM is adapted so that it becomes possible to evaluate the first derivative of spectra.

The calibration of IHM is mostly limited to the relative spectral intensities of the involved components. In the present work, a method is presented that uses the information in the calibration spectra more thoroughly. For this purpose, nonlinear effects are parametrized as a function of concentration.

The commonly used peak profiles do not reflect the physical reality at a detector very well. As a result, narrow modelled peaks may change their apparent intensity if they are shifted. To correct these shortcomings, a new peak model is proposed in this work that is more closely aligned to the physical reality of a detector.

Chapter 1

Introduction

In the year 1930 C.V. Raman was awarded the Nobel Prize for the discovery of the effect that was later named after him. Raman used sunlight and complementary light filters to prove that "the diffuse radiation of the ordinary kind, having the same wavelength as the incident beam, is accompanied by a modified scattered radiation of degraded frequency" (Raman and Krishnan (1928)). Some years earlier the Raman effect had been theoretically predicted by Adolf Smekal (Smekal (1923)).

At the time, the experiments were motivated by fundamental questions in physics of the interaction of light with matter. In this thesis, we are concerned with a more mundane question: How can we determine the composition of fluid mixtures as precisely as possible using Raman spectroscopy? Nevertheless, this question has great significance for many areas of scientific research and industrial application. In these areas, spectroscopic techniques are a valuable tool to investigate material properties, chemical reactions and for process analysis and control.

Raman spectroscopy allows (as other spectroscopic techniques) for the simultaneous quantitative in-situ determination of all Raman active components of a mixture. Thanks to modern instruments, each acquired Raman spectrum usually consist of more than one thousand data points, *i.e.*, measured intensities. For the conversion of these data points to concentration measures, there is a huge variety of methods. Simple methods just measure the peak height or the peak area of one or more distinct peaks. Thus, these methods use only a fraction of the provided information.

A simple method that uses the whole spectrum is called Classical Least Squares (CLS, Martens (1992)). CLS fits a superposition of pure component spectra to a mixture spectrum. Thereby, CLS undoubtedly uses the whole spectrum. Still, CLS performs poorly for many mixtures as it assumes that a mixture spectrum is solely a linear superposition of pure component spectra, which is often not true. A more precise question would thus be: How can we best use the abundance of data points in

each spectrum to quantify the composition of mixtures?

A popular answer to this question are multivariate statistical methods – also called soft modeling approaches – such as Partial Least Squares (PLS) and Principal Component Regression (PCR), (Martens (1992)). These methods are based on linear algebra and thus a very fast and efficient means to extract the desired information from the data. It is even possible to determine properties of liquid mixtures that are only loosely correlated with the concentration of one specific component as, *e.g.*, the octane number in commercial gasoline (Flecher et al. (1997)). However, these methods come at the price of a high calibration effort and a very limited ability for extrapolation (Faber and Rajkó (2007); Martens (1992)).

One important obstacle in the spectral evaluation are so-called nonlinear effects, *i.e.*, peak shifts and peak deformations. Nonlinear effects are caused by interactions between molecules (*e.g.*, due to hydrogen bonds or Van-der-Vaals forces). These nonlinear effects cannot be modeled by CLS as stated above. The soft modeling approaches can handle nonlinear effects much better than CLS, but, due to their linear nature, still only to a certain extent.

The present work is based on Indirect Hard Modeling (IHM), a method that explicitly incorporates nonlinear effects (Alsmeyer et al. (2004)). IHM is a physically motivated method that uses sums of parametrized peak models. These peak models allow to model peak shifts and peak deformations. Therefore, IHM is particularly suited for highly interacting mixtures or even reactive systems. As IHM handles the nonlinearity in the spectral evaluation, it can use a purely linear calibration model. This calibration model allows for larger extrapolations and therefore greatly reduces the number of required calibration samples.

IHM has been successfully applied in multiple areas, *e.g.*, measurement of diffusion coefficients (Bardow et al. (2003, 2005, 2006); Blesinger et al. (2014); Peters et al. (2017)), mass transport in fuel cell membranes (Scharfer et al. (2007)), phase equilibria (Luther et al. (2013); Schuster et al. (2014); Glass et al. (2016); Luther et al. (2015); Thien et al. (2017); Liebergesell et al. (2017)), polymerization (Meyer-Kirschner et al. (2016)), determination of surfactant distribution (Baesch et al. (2016)), crystallization (Helmdach et al. (2013)), ionic liquids (Viell and Marquardt (2012)), hydrogels (Heinemann et al. (2005)), wood disintegration (Viell et al. (2016)) and the mixing behavior of switchable solvents (Hardy and Liauw (2013)). IHM is not restricted to optical spectroscopy but has also been used, *e.g.*, in nuclear magnetic resonance (NMR) spectroscopy (Michalik-Onichimowska et al. (2017)).

In the present work, we aim to increase the applicability and robustness of IHM. We focus on Raman spectroscopy here, but as stated above, IHM is applicable to other

spectroscopic techniques alike. In the following, we present the structure of this work and the topics addressed.

In Chapter 2, we briefly present some fundamentals of Raman spectroscopy. Furthermore, we discuss the state of the art of spectral evaluation with IHM.

In Chapter 3, we introduce a method to calibrate IHM models for reactive systems with intermediates, without relying on thermodynamic models. As intermediates are intrinsically unstable, gravimetric preparation of calibration samples is impossible.

In Chapter 4, we incorporate spectral derivatives into Indirect Hard Modeling. Using spectral derivatives has a long history in soft modeling approaches to reduce the negative effects of spectral background signals. The new method First-Derivative Indirect Hard Modeling (FD-IHM) combines the robustness of derivatives with respect to backgrounds with the ability of IHM to treat nonlinear effects.

The earlier Chapters discussed nonlinear effects only as a threat for the spectral evaluation. In Chapter 5, we present a method that exploits nonlinear effects in the spectral model to improve the spectral evaluation: Very Hard Modeling (VHM).

Despite having been used for a long time, the classical peak models in IHM are not entirely sound. The mathematical peak models are evaluated at discrete points that correspond to the wavenumbers the centers of each pixel on the detector. On the detector, however, the discrete pixels have a finite size and thus integrate the spectral signal over a finite range. For narrow peaks, this may lead to disadvantageous behavior of the peak models. In Chapter 6, we propose a new peak profile with increased physical soundness.

In the last Chapter (Chapter 7), we summarize the present work and present perspectives for future work.

Chapter 2

State of the Art

In this Chapter, we discuss the state of the art of the spectral evaluation of Raman spectra with Indirect Hard Modeling (IHM). In Section 2.1, we briefly introduce the aspects of Raman spectroscopy that are important for the spectral evaluation with IHM. In Section 2.2, the state of the art of Indirect Hard Modeling is presented.

2.1 Basics of Raman Spectroscopy

In this Section, we give a brief introduction on typical Raman setups and on the acquired Raman spectra. Much more detailed information can be found in textbooks as, *e.g.*, Ferraro et al. (2003).

Raman Setup

Figure 2.1 shows a typical Raman setup in a 90° arrangement, *i.e.*, the spectral signal is gathered at a 90° angle with respect to the laser beam. The monochromatic light of a laser is focused on a sample. In the sample, some photons of the light are scattered elastically (Rayleigh scattering) or inelastically (Raman scattering) by the molecules in the sample. The wavelength of the inelastically scattered photons is shifted compared to the wavelength of the photons of the laser light. The scattered light is focused on the entrance slit of a spectrometer. The spectrometer separates light of different wavelengths spatially by a diffraction grating. The spatially separated light falls onto a detector which measures the number of photons. A typical detector is a two-dimensional CCD chip. One dimension of the chip corresponds to the wavelength, the other dimension corresponds to the spatial dimension of the entrance slit which usually corresponds to the position on the laser line in the sample.

Another typical Raman setup is a confocal Raman setup (not shown) that collects the backscattered light. In a confocal Raman setup, spatial information is typically not available from a single measurement but can be attained by scanning the sample and acquiring multiple Raman spectra.

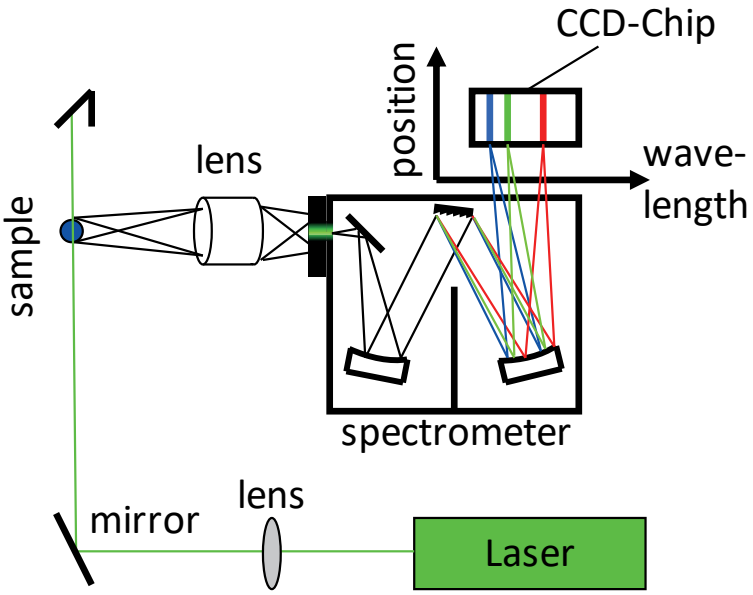

Figure 2.1: A typical Raman setup in a 90° arrangement

Raman Spectra

The commonly used quantity to display the energy of a photon in optical spectroscopy is wavenumbers $\tilde{\nu}$ (in cm^{-1}). Wavenumbers are proportional to the energy E of a photon:

$$E = h\,\nu = h\,c\frac{1}{\lambda} = h\,c\,\tilde{\nu}. \tag{2.1}$$

In Equation (2.1), h denotes the Planck constant, ν denotes the frequency, λ denotes the wavelength and $\tilde{\nu}$ denotes the wavenumbers. The conversion of frequencies ν or wavelengths λ to wavenumbers $\tilde{\nu}$ reads:

$$\tilde{\nu} = \frac{\nu}{c} = \frac{1}{\lambda}. \tag{2.2}$$

Raman scattering is an inelastic process. Thus, two different principal cases can be observed: In the case of Stokes-Raman scattering, a part of the energy of the incident photon is transferred to the intramolecular vibrations and/or rotations of the scattering molecule. Therefore, the scattered light has less energy than the incident light (red-shift). The energy difference between the incident light and the scattered light is called Raman shift. In the case of anti-Stokes-Raman scattering, energy of the intramolecular vibrations is transferred to the photon and the scattered light has more energy than the incident light (blue-shift). For anti-Stokes scattering, molecules must be in an excited state, which often requires elevated temperatures. In this thesis, we focus on processes near ambient temperature and thus on Stokes-Raman scattering solely.

Complex molecules often have many Raman-active intramolecular vibrations and rotations. These vibrations and rotations take place between single atoms (*e.g.*, C and H) or between groups of atoms (*e.g.*, two CH_2 groups). The energy of a vibration or a rotation is a function of the masses of the involved atoms and of the length and strength of the bonds between them. The energy difference between the excited and the ground state of the vibrations and rotations in a molecule lead to multiple distinct peaks which result in the specific Raman spectrum of this component.

We see from Equation (2.1) that wavenumbers are proportional to the energy of the photon. This proportionality is important as a Raman shift has a given energy difference, regardless of the excitation wavelength. Therefore, the positions of Raman peaks in the spectrum are invariant to the excitation wavelength.

So far, we discussed the *positions* of peaks in a Raman spectrum. The *intensities* of the peaks in a Raman spectrum depend on multiple variables. This thesis focuses on the quantitative spectral evaluation. Therefore, we limit the presentation of Raman spectroscopy to the fundamental aspects relevant for spectral evaluation.

The Raman intensity (in W/sr) is proportional to: (i) the laser intensity, (ii) the fourth power of the wavenumbers of the scattered radiation, (iii) the number of the molecules n in the measurement volume V, *i.e.*, the concentration c. Ferraro et al. (2003) give the following equation for the Raman intensity (Equation 2 on page 122 therein):

$$I_{\text{Raman}}\left(\tilde{\nu}\right) = K\left(\tilde{\nu}\right) \cdot A\left(\tilde{\nu}\right) \cdot \tilde{\nu}^4 \cdot I_0 \cdot J\left(\tilde{\nu}\right) \cdot c. \tag{2.3}$$

In Equation (2.3), $K(\tilde{\nu})$ denotes the overall spectrometer response. The spectrometer response accounts for the quantum efficiency of the grating of the spectrometer

and for the quantum efficiency of the camera. $A(\tilde{\nu})$ denotes the absorption of the scattered light in the medium. I_0 denotes the laser intensity. $J(\tilde{\nu})$ denotes a molar scattering parameter.

Equation (2.3) describes physical effects to the point that Raman scattered light is scattered from the sample only. For a measured spectrum S_i of component i, some more effects have to be taken into account. The Raman intensity is collected within a certain solid angle Ω that is defined by the optics and for a certain excitation time t_{exc}. Typical silicon-based detectors detect the number of photons rather than the intensity. Therefore, the exponent of the wavenumber $\tilde{\nu}$ in Equation (2.3) is reduced from 4 to 3. Due to the internal field effect, the refractive index $n_{\text{refractive}}$ has an influence on the Raman intensity (Nestor and Lippincott (1973)). We summarize this effect in a factor $f(n_{\text{refractive}})$. This leads to the following expression for the measured Raman spectrum S_i of component i:

$$S_i(\tilde{\nu}) = K(\tilde{\nu}) \cdot A(\tilde{\nu}) \cdot \tilde{\nu}^3 \cdot I_0 \cdot J_i(\tilde{\nu}) \cdot c_i \cdot t_{\text{exc}} \cdot \Omega \cdot f(n_{\text{refractive}}). \tag{2.4}$$

Equation (2.4) holds for every component i in a multi-component system. Note that the Raman spectrum S_i of component i is proportional to the concentration c_i of the same component. A spectrum of a multi-component system $S(\tilde{\nu})$ is the sum of the distinct Raman spectra $S_i(\tilde{\nu})$ of the molecules of all n_{C} components i:

$$S(\tilde{\nu}) = \sum_{i=1}^{n_{\text{C}}} S_i(\tilde{\nu}). \tag{2.5}$$

The Raman spectrum $S(\tilde{\nu})$ is still often referred to as the "Raman intensity". We adhere to this practice in this thesis. The spectrum is acquired by a sensor with discrete pixels. Therefore, the spectrum given in Equation (2.5) is represented by a vector of discrete intensities $S(\tilde{\nu}_m)$ and a vector of discrete wavenumbers $\tilde{\nu}_m$. The discrete wavenumber of each pixel is determined by calibration of the optical setup.

The distinct spectra of different molecules in a mixture and the proportionality of the intensities of the Raman spectra to concentration allow for a quantitative determination of the composition of the mixture in a spectral evaluation. The better the peaks of different components of a mixture are separated from one another, the easier it is to evaluate a mixture spectrum. Still, the peaks of different components often overlap, which aggravates the spectral evaluation. Furthermore, the intramolecular vibrations of a molecule are influenced by nearby molecules, *e.g.*, through H-bonds.

This influence leads to peak shifts and peak deformations such as peak broadening if the concentration is changed. Although the peak shapes and positions may change, the peak areas are usually assumed to remain constant (Pelletier (2003)). As the generic term for peak shifts and all kinds of peak deformation, we use the term nonlinear effects. Without nonlinear effects, a mixture spectrum would be a purely linear superposition of the pure component spectra.

2.2 Spectral Evaluation with Indirect Hard Modeling

The aim of spectral evaluation is the determination of the composition of multicomponent mixtures. We focus on the Indirect Hard Modeling method (IHM). The IHM method has been introduced by Alsmeyer et al. (2004). Further methodological papers focus on the automation of the method (Alsmeyer and Marquardt (2004); Kriesten et al. (2008a)) and on the identification of unknown pure component spectra in mixture spectra (Kriesten et al. (2008b)).

In this Section, we discuss the underlying theory of IHM as presented in Alsmeyer et al. (2004). The steps of the spectral fitting procedure with IHM are shown in Figure 2.2.

Mixture spectra are essentially a superposition of pure component spectra (see Equation 2.5). Thus, an obvious spectral evaluation method is fitting a weighted sum of pure component spectra to the mixture spectrum. This weighted sum is commonly referred to as Classical Least Squares (CLS, Martens (1992)). However, CLS is not able to model nonlinear effects due to its linear nature. If peak shifts occur, the respective peaks of the pure component spectra and of the mixture spectrum will not match, and the spectral evaluation will lead to large errors.

In principle, IHM follows an approach that is similar to CLS. But instead of using pure component spectra, IHM is based on the idea of using sums of mathematical peak shaped functions that represent the pure component spectra ("pure component models", *cf.* Figure 2.2a). The use of mathematical peak functions allows for easy compensation of nonlinear effects. We present the pure component models in Section 2.2.2. To account for nonlinear effects, the parameters of the component models are divided into adjustable parameters and fixed parameters (*cf.* Figure 2.2b). Then, the pure component models are fitted to mixture spectra in the component fitting step (*cf.* Figure 2.2c, Section 2.2.3). The calibration of the IHM model and the prediction of the composition of new spectra is presented in Section 2.2.4. To provide some context,

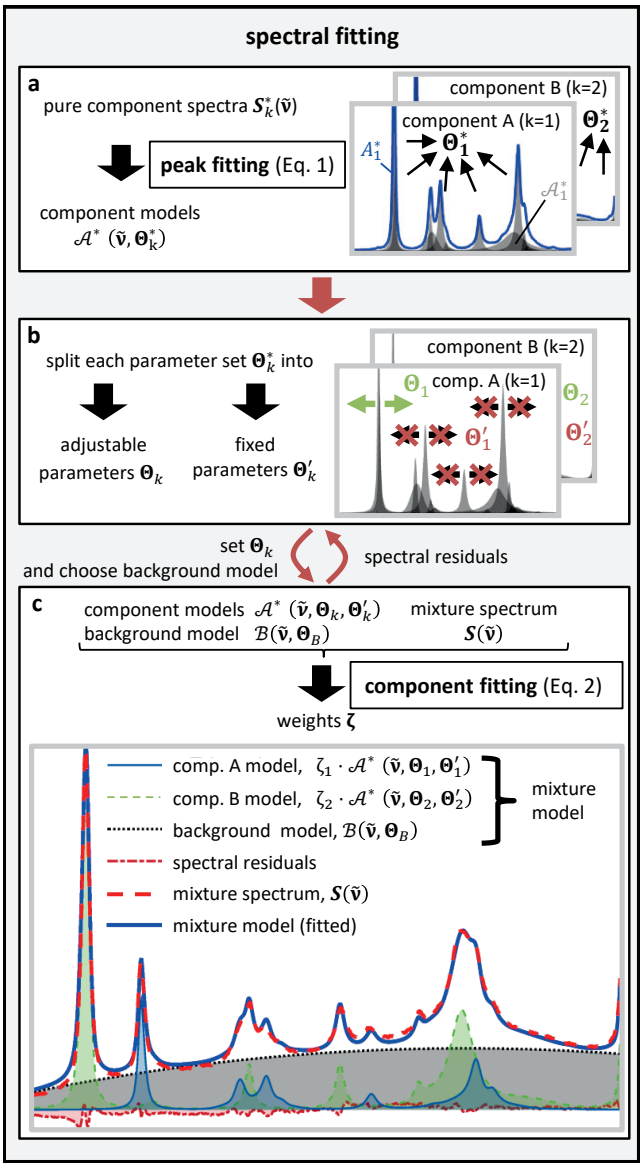

Figure 2.2: The spectral fitting procedure of Indirect Hard Modeling: (a) In the peak fitting step, a sum of pseudo-Voigt peaks (=component model) is fitted to each pure component spectrum. (b) The parameters of each component model are split into adjustable and fixed parameters. (c) A sum of weighted component models and a baseline model are fitted to mixture spectra. Based on the spectral residuals of the fitting (c), the choice of adjustable parameters (b) and the baseline model are refined. The baseline signal is modeled by a quadratic function for this illustration.

we discuss IHM in comparison to selected soft modeling approaches in Section 2.2.5. In Raman spectra, unwanted background signals are often present. In Section 2.2.1, we summarize common methods of mitigating the influence of these backgrounds by spectral pretreatment or by including a background model into IHM.

2.2.1 Baseline Models and Spectral Pretreatment

Background signals are a very common difficulty in Raman spectroscopy. Examples for common sources of background signals are fluorescence or background lighting. If background signals are not treated carefully, they can severely reduce the accuracy of the spectral evaluation. In the context of IHM, background signals can be removed in a pretreatment step or via background models.

Pretreatment aims to remove the background signal prior to the spectral evaluation with IHM. The downside of pretreatment is that the separation of the Raman signal from the background signal is generally not perfect. Background signals may be convex, concave or linear. In the region below an actual Raman peak, the pretreatment has to make some prediction about the shape of the background signal. Therefore, some parts of the actual Raman spectrum may be removed as background, and some parts of the background may still be present in the pretreated spectrum. For very narrow, distinct Raman peaks this prediction may be quite exact. For broad Raman peaks or several overlapping Raman peaks, the error introduced by this prediction may become the dominant source of error.

To avoid pretreatment, a background model $\mathcal{B}(\boldsymbol{\Theta}_B)$ with the parameter set $\boldsymbol{\Theta}_B$ can be fitted in the same optimization as the component models of IHM that will be discussed in the next Section. The component models provide information on the shape of the Raman signal. Therefore, the separation of Raman signal and background signal is facilitated. Still, the structure of the background model must be flexible enough to match the actual background signal without being too flexible and thereby matching the Raman signal. Commonly used background models are low order polynomials, exponential functions and custom background models that explicitly model a certain shape of a background signal as, e.g., the glass signal in a microchip (Peters et al. (2017)).

Still, background signals are always a source of error and must be treated very carefully. In general, one pretreatment method or one background model will not be able to handle every background.

2.2.2 Component Models: Representation of Pure Component Spectra

In IHM, pure component spectra are represented by pure component models. Each pure component model is a sum of peak functions. Commonly used peak functions are Gaussian and Lorentzian profiles. Gaussian and Lorentzian profiles can be parametrized in the height norm (peak height equals unity) and in the area norm (peak area equals unity). In Raman spectroscopy, the peak area is usually assumed to be constant (Pelletier (2003)). Therefore, the parametrization in the area norm is more useful and used in this thesis. Alsmeyer et al. (2004) use the height norm and a constraint to keep the peak areas constant. Both methods lead to the same results. A comparison of the height norm and the area is given in Appendix A.

To increase the flexibility of the peaks, combinations of the Gaussian and the Lorentzian profiles are common. The convolution of these profiles is called Voigt profile; a sum of these profiles is called pseudo-Voigt profile. As the convolution is numerically expensive, the pseudo-Voigt profile is used here. A peak in pseudo-Voigt parametrization has the following parameters:

- the position of the peak maximum δ,
- the full width at half maximum (FWHM) γ,
- the area of the peak α,
- and the Gaussian part β.

The Gaussian profile $G(\tilde{\nu})$ in area is given by:

$$G\left(\tilde{\nu}\right) = \left(\frac{2}{\gamma}\right) \cdot \sqrt{\frac{\ln 2}{\pi}} \cdot \exp\left\{\left[-\ln\left(2\right) \cdot \left(\frac{\tilde{\nu}-\delta}{\gamma/2}\right)^2\right]\right\}, \tag{2.6}$$

and the Lorentzian profile $L(\tilde{\nu})$ is given by:

$$L\left(\tilde{\nu}\right) = \left(\frac{2}{\gamma}\right) \cdot \frac{1}{\pi} \cdot \frac{1}{\left(\frac{\tilde{\nu}-\delta}{\gamma/2}\right)^2 + 1}. \tag{2.7}$$

With these definitions, the pseudo-Voigt profile $V(\tilde{\nu})$ reads:

$$V\left(\tilde{\nu}\right) = \alpha\left[\beta \cdot G\left(\tilde{\nu}\right) + \left(1-\beta\right) \cdot L\left(\tilde{\nu}\right)\right]. \tag{2.8}$$

For each Voigt profile, an *almost* identical pseudo-Voigt profile exists, *i.e.*, a pseudo-Voigt profile can reproduce a Voigt profile with an accuracy of 0.01% (Olivero and Longbothum (1977)). Both profiles differ in parametrization and an analytical parameter conversion is not available. However, empirical parameter conversions from Voigt to pseudo-Voigt are available (Olivero and Longbothum (1977); Thompson et al. (1987)). For very accurate parameter conversion, a peak in one parametrization can be fitted to a peak in the other parametrization.

For brevity, the parameters of all peak profiles of the pure component model $\mathcal{A}_k$ of component k are concatenated in a parameter matrix $\mathbf{\Theta}_k^* = [\boldsymbol{\delta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{\beta}]$. Each of the parameter vectors $\boldsymbol{\delta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\alpha}$, and $\boldsymbol{\beta}$ is a column vector so that each line of $\mathbf{\Theta}_k^*$ parametrizes one peak of the pure component model k. The equation of the pure component model $\mathcal{A}_k$ reads:

$$\mathcal{A}_k\left(\tilde{\nu}, \mathbf{\Theta}_k^*\right) = \sum_{p=1}^{n_{\text{peaks}}} V\left(\tilde{\nu}, \Theta_{k,p,1}^* \ldots \Theta_{k,p,4}^*\right). \tag{2.9}$$

The star (*) superscript denotes the parameter set of a pure substance, n_{peaks} denotes the number of peaks.

Setup of Component Models: IHM-Peak-Fitting

In the peak-fitting procedure, peaks are successively added to a pure component model $\mathcal{A}_k^*$ so that the pure component model matches the pure component spectrum S_k^* more and more accurately. The peak-fitting equation reads:

$$\mathbf{\Theta}_k^* = \underset{\mathbf{\Theta}_k^*, \mathbf{\Theta}_B}{\arg\min} \sum_{m=1}^{n_{\tilde{\nu}}} \left[S_k^*(\tilde{\nu}_m) - \mathcal{A}_k^*(\tilde{\nu}_m, \mathbf{\Theta}_k^*) - \mathcal{B}(\tilde{\nu}_m, \mathbf{\Theta}_B)\right]^2, \tag{2.10}$$

where $n_{\tilde{\nu}}$ denotes the number of discrete wavenumbers $\tilde{\nu}_m$ of the pure component spectrum S_k^*. A pure component spectrum may still include a background signal. Therefore, a background model $\mathcal{B}(\tilde{\nu}_m, \mathbf{\Theta}_B)$ with the parameter set $\mathbf{\Theta}_B$ is included in Equation (2.10), see previous Section 2.2.1. The parameters $\mathbf{\Theta}_B$ of the background model are usually of no further interest and are thus discarded after the fit. In manual peak-fitting, initial values of some peaks are commonly assigned to the model in a graphical user interface. These peaks are then fitted with Equation (2.10). Iteratively, more and more peaks may be added, until the pure component spectrum is adequately

represented by the pure component model. An automatic peak fitting algorithm has been published by Alsmeyer and Marquardt (2004). In Appendix B, we provide information on the numerical solution of Equation (2.10).

2.2.3 Component Fitting: Evaluation of Mixture Spectra

The spectral evaluation of mixture spectra with IHM is called component fitting. In component fitting, essentially a weighted sum of component models ("mixture model") is fitted to the mixture spectrum. The spectral evaluation results in weights ζ_k for each component k which correspond to the intensity of this component in the spectrum. Additionally, a background model $\mathcal{B}$ is usually included to mitigate the influence of background signals.

The equation for fitting a mixture model to a mixture spectrum S reads:

$$\boldsymbol{\zeta} = \underset{\boldsymbol{\zeta},\boldsymbol{\Theta}_1 \ldots \boldsymbol{\Theta}_{n_C},\boldsymbol{\Theta}_\mathrm{B},\Theta_\mathrm{S}}{\arg\min} \sum\nolimits_{m=1}^{n_{\tilde{\nu}}} \left\{ S\left(\tilde{\nu}_m\right) - \underbrace{\left(\sum_{k=1}^{n_\mathrm{C}} \left[\zeta_k \cdot \mathcal{A}_k \left(\tilde{\nu}_m^\Delta, \boldsymbol{\Theta}_k, \boldsymbol{\Theta}_k^{'} \right) \right] + \mathcal{B}\left(\tilde{\nu}_m^\Delta, \boldsymbol{\Theta}_\mathrm{B} \right) \right)}_{\text{mixture model}} \right\}^2 , \tag{2.11}$$

where $\tilde{\nu}_m^\Delta = \tilde{\nu}_m - \Theta_\mathrm{S}$.

Here, ζ_k denotes the weight of the component model $\mathcal{A}_k$ of component k in the mixture spectrum, $n_{\tilde{\nu}}$ denotes the number of discrete wavenumbers of the spectrum, S denotes the mixture spectrum, n_C denotes the number of components, and $\mathcal{B}$ denotes a background model with the parameter set $\boldsymbol{\Theta}_B$.

The distinguishing feature of IHM compared to CLS is that IHM incorporates nonlinear effects into the model. If nonlinear effects occur, the respective peak parameters are adjusted as part of the component fitting. For this purpose, the parameter set $\boldsymbol{\Theta}_k^*$ of each pure component model $\mathcal{A}_k$ is divided into adjustable parameters $\boldsymbol{\Theta}_k$ and fixed parameters $\boldsymbol{\Theta}_k^{'}$. Note that only the adjustable parameter sets $\boldsymbol{\Theta}_k$ are arguments to the arg min function in Equation (2.11).

If the optical setup is misaligned (*e.g.*, a badly repositioned grating), the whole measured spectrum apparently shifts. To account for such shifts, a global shift parameter Θ_S is introduced in Equation 2.11. $\tilde{\nu}_m$ denotes a discrete wavenumber of the spectrum and $\tilde{\nu}_m^\Delta$ denotes the shifted wavenumber used in the component model.

The choice of adjustable parameters $\boldsymbol{\Theta}_k$ is usually based on spectral residuals. An automated approach has been published by Kriesten et al. (2008a); the basic idea is to use a sensitivity analysis to choose the parameters with the greatest influence on the spectral residuals and to make these parameters adjustable. A separate sensitivity analysis is done for each calibration spectrum and for each evaluated spectrum.

Alternatively, a manual approach can be used. In this case, the aim is to find a set of adjustable parameters that works well for all calibration spectra. This choice is also primarily based on the spectral residuals, but care must be taken. On the one hand, more and more adjustable parameters always improve the spectral fit. On the other hand, too many adjustable parameters lead to ambiguous solutions: The same spectral feature can be fitted by various peaks of different component models. This ambiguity leads to poor calibration results. For this reason, the choice of free parameters should also be made on the calibration results. The choice of adjustable parameters is an iterative process (*cf.* Figure 2.2b-c) that is performed once during the calibration of a model.

The automatic approach as presented in Kriesten et al. (2008a) has the advantage that no expert knowledge is required, and that the adjustable parameters are chosen on an objective basis. However, two nearly identical spectra may be evaluated very differently, if different parameters are made adjustable by the automatic. The manual approach has the advantage that the same set of adjustable parameters is used for all evaluated spectra. By limiting the choice of adjustable parameters, more stable models can be acquired if the choice of free parameters is performed carefully. In this thesis, we always use the manual approach.

As a heuristic rule, we choose adjustable parameters in the following order:

1. assign adjustable position parameters δ to shifting peaks,
2. assign adjustable peak-width parameters γ to peaks that deform,
3. if the peak deformation cannot be handled adequately with γ, assign an adjustable Gaussian part β.

The area parameter α is never adjustable, as the peak areas of a Raman spectrum are assumed to be constant (Pelletier (2003)).

The calibration spectra and the pure component spectra should be acquired in a short time period without changing any settings of the optical setup. In this case, the shift between the calibration spectra becomes negligible and the shift parameter can be fixed to zero, *i.e.*, $\Theta_{\mathrm{S}} = 0$. This simplifies the selection of the set of adjustable parameters since the global shift is otherwise strongly correlated with peak shifts.

In an unconstrained optimization, some modeled peaks with adjustable parameters

tend to fit random features of the spectrum that are possibly part of the background signal. This behavior can often be observed if the modeled peaks are not present in the measured mixture spectrum, *e.g.*, because the composition is very close to a pure substance. To avoid this behavior, parameter constraints are used for the adjustable parameters. Typical IHM parameter constraints for Raman spectroscopy are given in Table 2.1.

Table 2.1: Typical IHM parameter constraints used for Raman spectroscopy in this thesis. The star superscript (*) denotes the pure parameter value of the pure component model.

	lower bound	upper bound
weight ζ	0	∞
position δ	$\delta^* - 20\,\mathrm{cm}^{-1}$	$\delta^* + 20\,\mathrm{cm}^{-1}$
FWHM γ	$\gamma^*/5$	$\gamma^* \cdot 5$
Gaussian part β	0	1

2.2.4 Calibration and Prediction

Calibration is used to translate the weights $\boldsymbol{\zeta}$ that are determined in Section 2.2.3 into concentration measures. For calibration, Raman spectra with known composition are needed. Typically, the calibration samples are prepared gravimetrically. However, this method is limited to nonreactive mixtures.

There are two common types of calibration for IHM: (i) an absolute calibration uses a concentration c in (mol/L) and (ii) a ratiometric calibration uses mass fractions or mole fractions.

Absolute calibration

From Equation (2.4), we see that the intensity $S_i(\tilde{\nu}_m)$ of a Raman spectrum of component i is generally proportional to the concentration c_i. However, if nonlinear effects are present(*i.e.*, peak shifts and deformations), not every pixel of the spectrum is proportional to c_i. To still obtain the composition, we use the assumption that all peak areas remain constant (Pelletier (2003)). With this assumption, the sum of intensities of the spectrum is still proportional to the concentration c_i:

$$\sum_{m=1}^{n_{\tilde{\nu}}} S_i\left(\tilde{\nu}_m\right) \propto c_i. \tag{2.12}$$

The other factors given in Equation (2.4) that influences S_i are discussed later.

In the component fitting step (Section 2.2.3), the weight ζ_i is determined. The component models are parametrized in a way that each component model has a constant sum of intensities (recall that the area parameter α is always constant for all peaks, Section 2.2.3). Therefore, the weight ζ_i is proportional to the sum of intensities of component i in the spectrum:

$$\sum_{m=1}^{n_{\tilde{\nu}}} S_i\left(\tilde{\nu}_m\right) \propto \zeta_i. \tag{2.13}$$

From Equations (2.12) and (2.13) follows:

$$c_i = k_i \cdot \zeta_i. \tag{2.14}$$

where k_i denotes a calibration coefficient. To determine the calibration coefficient k_i, at least one spectrum with a known concentration is required. After calibration, the concentration of component i in a mixture can be determined using Equation (2.14).

Nevertheless, there are factors that may prohibit the use of absolute calibration (*cf.* Equation (2.14)):

1. the absorption of the Raman signal $A(\tilde{\nu})$ in the medium may change for different compositions as the constituting components may have different absorption characteristics,
2. the refractive index $n_{\text{refractive}}$ of the mixture may change, which has an influence on the internal field effect $f(n_{\text{refractive}})$ (Nestor and Lippincott (1973)) and on the solid angle Ω (Woodward and George (1951)),
3. the laser intensity I_0 may fluctuate.

These factors change the intensity of spectrum S_i and thus the weight ζ_i. Therefore, the factors must be the same for calibration mixtures and for unknown mixtures.

The influence of the first two factors can be reduced, if calibration mixtures and evaluated mixtures have a similar composition. Still, absolute calibration is challenging for Raman spectroscopy and will not work for every system.

Ratiometric calibration

Ratiometric calibration can be used to determine mole fractions x_i or mass fractions w_i of all components in a mixture from the weights ζ_i.

The idea of ratiometric calibration is that the effects that may prohibit an absolute calibration, affect the Raman spectra S_i of all components i *in the same way*. Therefore, *ratios* of weights (ζ_i/ζ_j) are not affected. With Equation (2.14), this idea leads to the following equation for mole fraction x_i:

$$x_i = \frac{c_i}{\sum_{j=1}^{n_C} c_j} = \frac{\zeta_i \cdot k_i}{\sum_{j=1}^{n_C} \zeta_j \cdot k_j} \tag{2.15}$$

$$= \frac{1}{\sum_{j=1}^{n_C} \frac{\zeta_j \cdot k_j}{\zeta_i \cdot k_i}} = \frac{1}{\sum_{j=1}^{n_C} \frac{\zeta_j}{\zeta_i} \cdot \frac{1}{K_{ij}}}. \tag{2.16}$$

n_C denotes the number of components. As only ratios of calibration coefficients k_i/k_j are used, one calibration coefficient may be set to an arbitrary value. Alsmeyer et al. (2004) introduce binary calibration coefficients $K_{ij} = k_i/k_j$ instead of the calibration coefficients k_i and k_j. For multi-component systems, this notation introduces $(n_C)^2 - n_C$ binary calibration coefficients from which only $n_C - 1$ are independent. This excess of parameters introduces an unnecessary confusion. In this thesis, we therefore use the normal calibration coefficients k_i and set the first calibration coefficient to unity ($k_1 = 1$), instead of using K_{ij}. Both notations are equivalent. Note that $K_{i1} = k_i/k_1 = k_i$.

For mass fractions w_i, Equation (2.15) reads:

$$w_i = \frac{c_i \cdot M_i}{\sum_{j=1}^{n_C} c_j \cdot M_i} = \frac{\zeta_i \cdot k_i \cdot M_i}{\sum_{j=1}^{n_C} \zeta_j \cdot k_j \cdot M_j} = \frac{\zeta_i \cdot \hat{k}_i}{\sum_{j=1}^{n_C} \zeta_j \cdot \hat{k}_j}. \tag{2.17}$$

Equations (2.15) and (2.17) have the same structure. In Equation (2.17), we also set the first calibration coefficient to unity ($\hat{k}_i = 1$). As calibration thus works identically for mole fractions x and for mass fractions w, we only discuss the calibration for mole fractions below.

To determine the calibration coefficients $\mathbf{k}$, at least one calibration spectrum S with known composition $\mathbf{x}_{\text{true}}$ is required. If more calibration spectra are available, the calibration coefficients are determined by fitting the true mole fractions $x_{\text{true},s,i}$ of component i in calibration spectrum s to the mole fractions $x_{\text{pred},s,i}$ that are predicted with Equation (2.15) in a least squares sense:

$$\mathbf{k} = \arg\min_{\mathbf{k}} \sum_{s=1}^{n_{\text{calib}}} \sum_{i=1}^{n_{\text{C}}} \left(x_{\text{pred},s,i} \left(\boldsymbol{\zeta}_s, \mathbf{k} \right) - x_{\text{true},s,i} \right)^2 , \quad \text{where } k_1 = 1. \tag{2.18}$$

n_{calib} denotes the number of calibration spectra.

Measures for the predictive abilities of calibrations

As a measure for the predictive abilities of a model, the root-mean-square error RMSE can be applied. For mole fractions x, the RMSE reads:

$$RMSE = \sqrt{\frac{1}{n_m} \sum_{s=1}^{n_m} \frac{1}{n_{\text{C}}} \sum_{i=1}^{n_{\text{C}}} \left(x_{\text{pred},s,i} - x_{\text{true},s,i} \right)^2}. \tag{2.19}$$

Here, n_m denotes the number of mixture spectra and $x_{\text{pred},m,i}$ and $x_{\text{true},m,i}$ denote the predicted and true mole fraction of component i respectively.

If a calibrated model is used to predict the calibration spectra themselves, the RMSE is denoted root-mean-square error of calibration *RMSEC*. The *RMSEC* cannot detect overfitting, as the same spectra are used for calibration and prediction. Therefore, the root-mean-square error of cross validation *RMSECV* gives a more reliable assessment of the predictive abilities of a model. For the *RMSECV*, the composition of each spectrum is predicted by a calibration based on all other spectra. The best option is to use different spectra sets for calibration and validation. In this case, the root-mean-square error of prediction *RMSEP* is calculated for the validation spectra (Faber and Rajkó (2007)).

Another measure for the predictive abilities of a model is the maximum absolute prediction error *MAXPE*. This error is simply the worst case, *i.e.*, the largest error encountered in a set of predictions:

$$MAXPE = \max_{s=1\ldots n_m,\, i=1\ldots n_{\text{C}}} \left| x_{\text{pred},s,i} - x_{\text{true},s,i} \right| . \tag{2.20}$$

2.2.5 Differentiation of IHM and Statistical Methods

In this Section, we want to discuss the capabilities of IHM by comparing it briefly to Partial Least Squares (PLS, Martens (1992)), which is a widely popular soft modeling approach.

Soft modeling approaches such as PLS are data-driven approaches based on linear algebra. The analysis is based on data vectors and matrices. With linear algebra, nonlinear effects such as peak shifts can only be modeled to a certain extent by adding further latent vectors to the model (Zientek et al. (2016)). PLS requires large calibration sets, as the extrapolation ability is limited; in particular for multi-component systems (Faber and Rajkó (2007)). Spectral shifts due to misalignment cannot be modeled. In this case, recalibration or calibration transfer is required (Wehlburg et al. (2002)).

Still, PLS is one of the most frequently applied methods for spectral evaluation. One reason might be that it has long been implemented in commercially available software. E.g. the Unscrambler has been developed in 1986. In addition, there are more conceptual advantages of PLS over IHM. Unlike IHM, PLS does not require pure component spectra. PLS models can simply be trained with mixture spectra and results from a reference analytical technique. Furthermore, PLS can be calibrated to properties that are only remotely correlated to the actual composition as, *e.g.*, the octane number (Flecher et al. (1997)). IHM becomes imprecise if all spectral features of the components overlap. As more components in a system lead to more overlapping peaks, the number of components that can be evaluated by IHM is restricted. Given a large enough calibration set, PLS may perform better in these cases. Furthermore, IHM will sometimes struggle to predict small concentrations of one component, as the respective peaks become so small that they are not properly fitted by the peak models. In these cases, PLS might still be able to use other features of the spectrum to predict these concentrations accurately.

In a certain sense, the methods best complement each other: IHM has large extrapolation abilities and thus requires little calibration effort. IHM can therefore often predict the whole concentration range with a single model. PLS requires a larger calibration effort but can often still be applied to systems where IHM fails.

2.3 Contribution of this Thesis

IHM needs calibration mixtures of known composition for calibration. For chemical reactions, the concentrations of intermediates are typically unknown, as intermediates

are usually intrinsically unstable. Therefore, dedicated calibration samples cannot be prepared gravimetrically. If available, thermodynamic equilibrium models can be used to calculate the amount of intermediates in a mixture in equilibrium. However, using these models for calibration bears the risk of overfitting the spectral data to the thermodynamic equilibrium models. To enable the application of IHM to reactive systems with intermediates, we present in this thesis the **model-free calibration of Raman measurements of reactive systems**. In Chapter 3, instead of a thermodynamic equilibrium model, only a reaction mechanism and well-known quantities such as material balances and electroneutrality are used. The method is demonstrated for the system monoethanolamine–water–CO_2.

Background signals are challenging for the spectral evaluation with IHM. Model setup and calibrations are often performed with very pure laboratory grade chemicals. In later application, however, new unforeseen sources of background signals might arise. Possible sources are fluorescent signals from, *e.g.*, a lubricant that entered the system or some source of background lighting. In statistical methods, spectral derivatives are a very common method to suppress background signals. In Chapter 4, we introduce this concept to IHM and test its performance to allow for a **robust analysis of spectra with strong background signals by First-Derivative Indirect Hard Modeling (FD-IHM)**.

One of the major features of IHM is the small number of required Raman spectra needed for calibration. Only very few calibration coefficients need to be determined. Therefore, one pure component spectrum per component plus one mixture spectrum is theoretically sufficient. However, this also means that more than a few calibration spectra are of limited value to IHM. An IHM calibration cannot be arbitrarily refined by adding more and more calibration spectra. The reason is that IHM only uses a part of the information provided by the calibration spectra, *i.e.*, the relative intensities of the components. However, calibration spectra also provide information on nonlinear effects that are dependent on the composition of the spectrum, *i.e.*, peak shifts and peak deformations. Due to its mathematical structure, IHM very successfully *mitigates* the negative impact of these nonlinear effects on the performance of the spectral evaluation. To include the information of these nonlinear effects into the IHM analysis, we parametrize all nonlinear effects as a function of the composition of the mixture. Thereby, we *use* the provided information to make the model more robust. The new method is called **Very Hard Modeling** and presented in Chapter 5.

The commonly used peak model evaluates pseudo-Voigt functions once for the central wavenumber of each pixel. This evaluation at a single point is not entirely correct since each pixel integrates the intensity function over space and time. In Chapter 6,

we present a **physical sound peak profile** that also implements this integration to the model. The new model is compared to the conventional peak model.

Chapter 3

Model-Free Calibration of Reactive Systems with Intermediates

Parts of this Chapter have been published in:

P. Beumers, T. Brands, H.-J. Koss, A. Bardow, "Model-Free Calibration of Raman Measurements of Reactive Systems: Application to Monoethanolamine/Water/CO_2", *Fluid Phase Equilibria*, vol. 424, pp. 52-57, 2016

Contribution report: Principal author, author of the first draft, development and implementation of the method, data evaluation.

To use Raman spectroscopy quantitatively, calibration is required. For calibration, the concentrations of the species need to be known. However, if Raman spectroscopy is used to monitor complex chemical reactions, the concentration of intermediates is typically unknown. Since intermediates are usually intrinsically unstable, dedicated calibration samples cannot be prepared. However, if available, thermodynamic models can be used to calculate the concentrations of the intermediates. Using thermodynamic models bears the risk of overfitting spectral data to reproduce the models. Methods for calibration of reactive systems without the use of a thermodynamic model have therefore been developed:

For an aqueous MEA solution loaded with CO_2, calibration without using a thermodynamic model has been successfully performed by Wong et al. (2015). Based on deep understanding of the reacting mixture, these authors were able to calibrate the aqueous reactive mixture using non-aqueous reactive mixtures which are assumed to achieve full conversion. In this case, no equilibrium model is required. The calibration of the non-aqueous mixture was then used for the aqueous system. The speciation results look very promising. However, they are not compared to any thermodynamic model or to other measurements. As full conversion of all species is not always guar-

anteed, the applicability of this method depends on the specific chemical system.

A very elegant approach is to directly parametrize a thermodynamic equilibrium model during the spectral evaluation as demonstrated for FT-IR measurements of aqueous solutions of the amines monoethanolamine (MEA), diethanolamine (DEA) and 2-amino-2-methyl-1-propanol (AMP) (Richner and Puxty (2012)). A similar approach was used to identify thermodynamic model parameters from FT-IR spectra of the synthesis of diphenylurethane from phenylisocyanate and phenol (Spegazzini et al. (2012)). This direct method should be also applicable to Raman spectra but still requires the functional form of the thermodynamic model as input.

In this Chapter, we present a calibration-free method based on stoichiometry and electroneutrality only. We illustrate the method with an aqueous MEA solution loaded with CO_2. This system has already been thoroughly studied by Raman spectroscopy (Vogt et al. (2011); Souchon et al. (2011); Samarakoon et al. (2013); Wong et al. (2015); Richner and Puxty (2012)) and NMR spectroscopy (Böttinger et al. (2007); Hilliard (2008); Jakobsen et al. (2005)). Thermodynamic models are available in the literature (Aronu et al. (2011); Wagner et al. (2013)). We use a model based on UNIQUAC (Aronu et al. (2011)) as reference for the presented method.

In Section 3.1, the experimental data employed in this work is briefly described. Section 3.2 summarizes the spectral evaluation to determine the weights of each species in the mixture spectrum. In Section 3.3, the model-free calibration method is explained. The concentrations of the species are determined from the weights calculated in the spectral evaluation. A case study is presented in Section 3.4 to validate our method by comparison to a thermodynamic model (Aronu et al. (2011)).

3.1 Raman Measurements of CO_2-loaded Aqueous MEA Solutions

Only the most important aspects of the experiments are described, as this work focuses on the calibration method of reactive systems. The experiments evaluated here were carried out by the group of Maiwald and a detailed presentation can be found elsewhere (Esche et al. (2014); Meyer et al. (2015)).

The core part of the setup consists of a temperature-controlled equilibrium vessel coupled to the optical setup. The vessel is filled with a solution of MEA and water. The solution is stirred to reduce the time to reach equilibrium. The vessel can be accessed to add CO_2 to the mixture and to measure the pressure in the vessel. The optical setup consists of a RamanRxn1TM Analyzer equipped with a 785 nm Invictus^{TM}

laser and a fiber-coupled immersion probe. The excitation wavelength of 785 nm has been chosen to avoid fluorescence within the spectrum (Esche et al. (2014); Meyer et al. (2015)).

The measurements evaluated in this work were performed at $40°C \pm 0.68\,K$ (Meyer et al. (2015)). The aim of the experiments was to analyze the MEA/water solution with increasing CO_2 loadings in chemical equilibrium. The CO_2 loading α_{CO_2} is defined as the total amount of absorbed CO_2 ($n^0_{CO_2}$) per initial amount of MEA (n^0_{MEA}):

$$\alpha_{CO_2} = \frac{(n^0_{CO_2})}{(n^0_{MEA})}. \tag{3.1}$$

The superscript 0 denotes the initial amount of substance before any reaction has taken place. One Raman spectrum is recorded with the unloaded aqueous MEA solution. Then, gaseous CO_2 is introduced to increase the CO_2 loading. The CO_2-loadings were measured gravimetrically. For this purpose, a reservoir filled with pressurized CO_2 is weighed. The CO_2 reservoir is then connected to equilibrium vessel to start loading CO_2. During CO_2-loading, the pressure in the vessel first rises sharply and then slowly decays as CO_2 is absorbed in the solution. Once the pressure decay in the system almost stops, the reservoir is decoupled and weighed again. The change in weight is corrected for the amount of CO_2 in the gas phase of the equilibrium vessel. As no gas is leaving the equilibrium vessel during CO_2-loading, the water loss is assumed to be negligible. After the reservoir is decoupled, the system equilibrates. To ensure chemical equilibrium, pressure, temperature, and Raman spectra are recorded over time. Once pressure and temperature are constant for a certain time, chemical equilibrium is assumed. Raman spectra were collected at different times to ensure chemical equilibrium (Meyer et al. (2015)).

3.2 Spectral Evaluation of the MEA - H_2O - $MEAH^+$ - $MEACOO^-$ System

In the spectral evaluation step, the Raman spectra of the mixtures are used to determine weight of every component in the spectrum. These weights are proportional to the concentration and are used in the model-free calibration.

The set of Raman spectra used in this work is shown in Figure 3.1. In the works of Wong et al. (2015) and Richner and Puxty (2012), the assignment of the peaks to

the respective species is thoroughly discussed. Due to the long excitation wavelength of 785 nm, the large peaks of water are located in the range of wavenumbers around 3500 cm^{-1}. The detector efficiency is very low in this area. Thus, the maxima of the water peaks were not detected. Therefore, we chose not to evaluate the water content in this work.

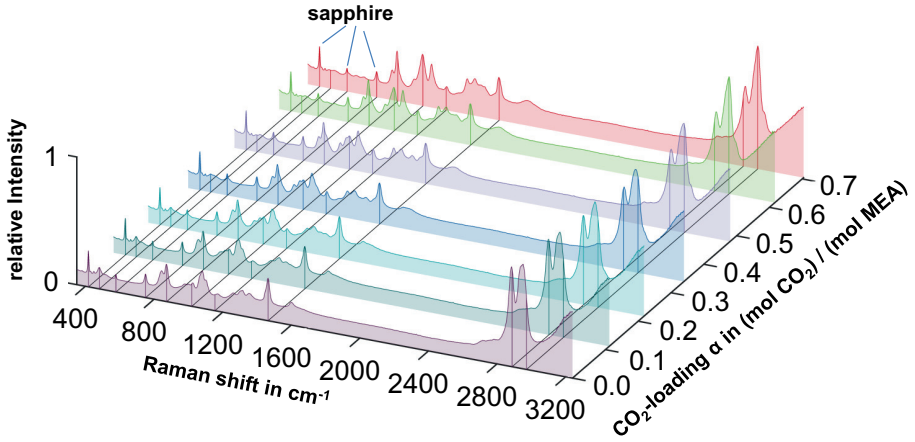

Figure 3.1: Typical set of Raman spectra of an aqueous MEA solution with varying CO_2-loading α_{CO_2}. The sapphire peaks are indicated which were used to normalize the measured weights of all components. Some peaks are marked to guide the eye.

In this work, the pure component model of MEA was determined from a mixture of MEA and water with a concentration of 30 wt-% MEA. First, a mixture of both components, MEA and water, was modeled. Then, the peaks were left out which were known to belong to water from a pure water spectrum. Using an aqueous MEA solution with a MEA concentration close to the concentration of the reactive system instead of pure MEA is advantageous: As most peaks have already shifted in the same manner as they will in the reactive mixture, less adjustable parameters are needed resulting in a more robust fit.

To determine the pure component spectra of $MEAH^+$ and HCO_3^-, experiments comparable to the ones described by Souchon et al. (Souchon et al. (2011)) have been performed. The pure component spectrum of $MEAH^+$ was determined from an aqueous solution of ethanolamine hydrochloride (MEA•HCl) which dissociates resulting in $MEAH^+$ and Cl^-. As the monoatomic Ion Cl^- has no Raman spectrum, the pure component model of $MEAH^+$ was determined by leaving out the water peaks as for MEA/water. The pure component spectrum of HCO_3^- was determined from an

aqueous solution of $KHCO_3$ which dissociates to HCO_3^- and K^+. Depending on the amount of dissolved HCO_3^-, part of HCO_3^- may react further to CO_3^{2-}. The easily distinguishable major peak of CO_3^{2-} at $1065\,cm^{-1}$ (Wong et al. (2015)) was therefore not modeled.

The pure component model of $MEACOO^-$ has been determined as follows from a loaded MEA/water solution (30 wt-% MEA) with a CO_2 loading of $\alpha_{CO_2} = 0.3$ mol CO_2 per mol MEA. At this loading, no HCO_3^- is found in the spectrum. Therefore, we assumed that all CO_2 reacted to $MEACOO^-$. As the pKa-value of MEA/MEAH+ equals 9.5 (Hall Jr (1957)), the concentration of OH^--Ions does not exceed 0.02 mol/L for typical initial concentrations of MEA of roughly 5 mol/L. Therefore, the concentrations of OH^- and H_3O^+ are also neglected. To maintain electroneutrality, equal amounts of $MEAH^+$ and $MEACOO^-$ are formed. The spectra of all species different from $MEACOO^-$ were subtracted from the mixture spectrum according to their concentration in the mixture. The remaining spectrum was used to prepare the spectral pure component model of $MEACOO^-$.

The pure component models were fitted to mixture spectra following the IHM approach outlined in Section 2.2. From the fit, the weights under the spectrum of each species can be determined.

3.3 Model-free Calibration

The calibration method detailed in this Section calculates the concentration of all components from the weights of all components. We start from Equation (2.14):

$$c_i = \tilde{k}_i \cdot \tilde{\zeta}_i. \tag{3.2}$$

The setup allows for a correction of fluctuating laser intensity or different exposure times that is described below. In Equation (3.2), the tilde superscript denotes that these are the uncorrected variables.

It is important to note that Equation 3.2 might not hold, if the refractive index of the solution changes significantly during the experiment. On the one hand, for an increasing refractive index, the laser focus moves away from the probe which decreases the Raman signal (Woodward and George (1951)). On the other hand, the Raman signal increases by the internal field effect as the refractive index increases (Nestor and Lippincott (1973)). Although both effects do not cancel each other completely,

we refrained from any corrections related to the refractive index for the sake of mathematical simplicity and assumed a constant refractive index. In general, the refractive index could be measured alongside with the Raman spectra to correct for the effect of the refractive index.

To determine the weights $\tilde{\zeta}_i$ required in Equation 3.2, we used Indirect Hard Modeling (IHM, Alsmeyer et al. (2004)) as outlined in Section 3.2. Still, model-free calibration is not limited to IHM. Any spectral evaluation method providing the weights $\tilde{\zeta}_i$ of the species i in the spectrum can be used.

The calibration task is to deduce the calibration coefficients $\tilde{k}_i$ for all components. The dependency of $\tilde{k}_i$ on the optical setup in Equation 3.2 is a challenge for the model-free calibration. If, e.g., the laser intensity fluctuates, the weights of all components $\tilde{\zeta}_i$ in the measured spectrum are affected linearly with the laser power.

To remove the influence of the optical setup in Equation 3.2, we used the Raman peaks of the sapphire window of the optical setup which have been visible in all spectra. The weight of the sapphire $\tilde{\zeta}_{\text{sapphire}}$ signal also changes according to changes in the optical setup. Assuming that the sapphire signal is affected by changes in the optical setup in the same way as the signal for all species in the mixture, we use the sapphire signal to normalize the weights $\tilde{\zeta}_i$. In Equation 3.2, $\tilde{\zeta}_i$ is replaced by $\zeta_i = \tilde{\zeta}_i / \tilde{\zeta}_{\text{sapphire}}$ resulting in

$$c_i = k_i \cdot \zeta_i, \tag{3.3}$$

with k_i denoting the calibration coefficients for the normalized weights ζ_i which do not depend on the optical setup anymore. This normalization is comparable to the normalization performed by Esche et al. (2014). The resulting weights ζ_i are shown in Figure 3.2. For brevity, we use Equation 3.3 in the subsequent discussion.

Conventional calibration requires the concentrations c_i of all components i as input. For a reactive system, this requirement cannot be met if the concentrations of the intermediates are unknown.

The proposed model-free calibration assumes only knowledge of well-known characteristics of the chemical reaction mechanism, namely the stoichiometry and electroneutrality. As input, we assume that the initial amounts of material are known for all reactants. In this work, we considered the following chemical reactions for the MEA-H2O-CO_2 system (reactions II, IV and V in Wagner et al. (2013)):

$$\mathrm{MEA} + \mathrm{H^+} \rightleftharpoons \mathrm{MEAH^+} \tag{3.4}$$

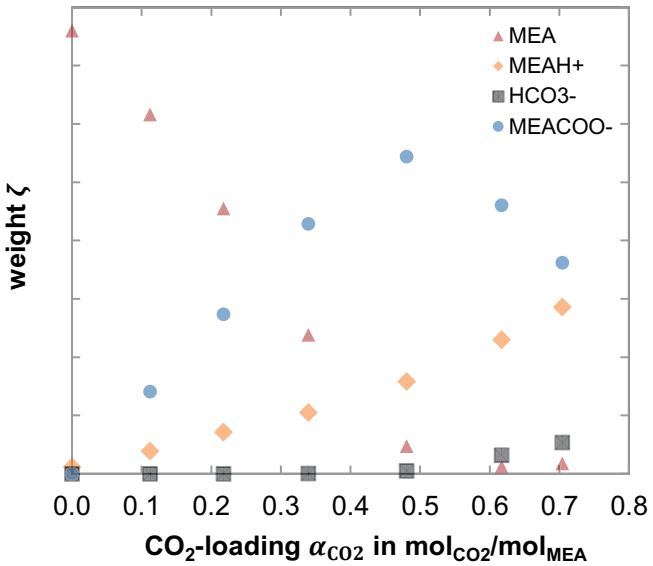

Figure 3.2: Weights ζ_i of the species of the reactive system MEA/water/CO_2 in the Raman spectrum determined with the IHM-approach. The shown weights have been normalized by the weight of the sapphire signal.

$$\mathrm{MEA} + \mathrm{CO_2} \rightleftharpoons \mathrm{MEACOO^-} + \mathrm{H^+}, \tag{3.5}$$

$$\mathrm{H_2O} + \mathrm{CO_2} \rightleftharpoons \mathrm{HCO_3^-} + \mathrm{H^+}. \tag{3.6}$$

The carbonate formation and the autoprotolysis of water (reactions I and III in Wagner et al. (2013)) are not taken into account. As shown in Section 3.1, the concentrations of H_3O^+ and OH^- are negligible. CO_3^{2-} has not been found in the spectrum and was therefore assumed to have a negligible concentration as well. Furthermore, water is not taken into account as explained in Section 3.1. Thus, the considered species are MEA, $MEAH^+$, $MEACOO^-$, and HCO_3^-.

For reactions 3.4-3.6, the material balance equations can be established. The material balance of MEA reads

$$n^0_{\mathrm{MEA}} - (n_{\mathrm{MEA},s} + n_{\mathrm{MEAH^+},s} + n_{\mathrm{MEACOO^-},s}) = 0. \tag{3.7}$$

The superscript 0 denotes the initial amount of substance before the reaction. The index s denotes the number of the measurement at different CO_2-loadings.

The material balance for CO_2 reads

$$n^0_{\mathrm{CO_2}} - (n_{\mathrm{HCO_3^-},s} + n_{\mathrm{MEACOO^-},s}) = 0. \tag{3.8}$$

As a third balance equation, we use electroneutrality:

$$n_{\mathrm{MEAH^+},s} + n_{\mathrm{H_3O^+},s} - n_{\mathrm{OH^-},s} - n_{\mathrm{HCO_3^-},s} - n_{\mathrm{MEACOO^-},s} = 0. \tag{3.9}$$

As discussed in Section 3.2, the concentrations of OH^- and $H3O^+$ are negligible compared to the concentration of the other ions. Equation (4.8) thus simplifies to:

$$n_{\mathrm{MEAH^+},s} - n_{\mathrm{HCO_3^-},s} - n_{\mathrm{MEACOO^-},s} = 0. \tag{3.10}$$

When using Raman spectroscopy, the concentration c_i is determined rather than the amount of material n_i, as stated in Section 2.2. To introduce Equation 3.3 into the balance equations, the molar volumes of the $\mathrm{MEA} - \mathrm{H_2O} - \mathrm{CO_2}$ system were taken from the literature (Amundsen et al. (2009)) and slightly extrapolated to higher loadings using a cubic polynomial fit. Alternatively, it would also be possible to measure the liquid volume together with the Raman spectra during the loading of the solution. We introduce $n_i = c_i \cdot V$ and $c_i = k_i \cdot \zeta_i$ into the balance Equations 3.7, 3.8 and 3.10. This introduces 4 unknown k_i, which are to be determined by 18 balance equations (3 independent balance equations per measurement, 6 independent measurements with a CO_2-loading $\neq 0$). As the system is overdetermined, the balance equations will not hold exactly. We replace the zeros on the right-hand side by residuals. This yields the following equations:

$$\begin{aligned}
n^0_{\mathrm{MEA}} - V \cdot (\zeta_{\mathrm{MEA},s} \cdot k_{\mathrm{MEA}} + \zeta_{\mathrm{MEAH^+},s} \cdot k_{\mathrm{MEAH^+}} + \zeta_{\mathrm{MEACOO^-},s} \cdot k_{\mathrm{MEACOO^-}}) \\
= R_{\mathrm{MEA},s}(\bar{\bar{\zeta}}, \mathbf{k}), \quad (3.11) \\
n^0_{\mathrm{CO_2}} - V \cdot (\zeta_{\mathrm{HCO_3^-},s} \cdot k_{\mathrm{HCO_3^-}} + \zeta_{\mathrm{MEACOO^-},s} \cdot k_{\mathrm{MEACOO^-}}) \\
= R_{\mathrm{CO_2},s}(\bar{\bar{\zeta}}, \mathbf{k}), \quad (3.12) \\
V \cdot \left[\zeta_{\mathrm{MEAH^+},s} \cdot k_{\mathrm{MEAH^+}} - (\zeta_{\mathrm{HCO_3^-},s} \cdot k_{ceHCO3-} + \zeta_{\mathrm{MEACOO^-},s} \cdot k_{\mathrm{MEACOO^-}})\right] \\
= R_{\mathrm{elec},s}(\bar{\bar{\zeta}}, \mathbf{k}). \quad (3.13)
\end{aligned}$$

The residuals of these equations are now minimized in a least-squares sense to determine the calibration coefficients $\mathbf{k}$:

$$\mathbf{k} = \arg\min_{\mathbf{k}} \sum_{s=1}^{n_\mathrm{m}} R^2_{\mathrm{MEA},s}(\bar{\bar{\zeta}}, \mathbf{k}) + R^2_{\mathrm{CO_2},s}(\bar{\bar{\zeta}}, \mathbf{k}) + R^2_{\mathrm{elec},s}(\bar{\bar{\zeta}}, \mathbf{k}). \tag{3.14}$$

If a nonreactive subsystem can be measured, the calibration coefficients for this subsystem can be determined directly via Equation 3.3. For the MEA-H2O-CO_2 system, MEA and water (without CO_2) could be calibrated independently by using a binary mixture or even the pure substances. To demonstrate the power of the model-free calibration, we refrained from using any subsystems for calibration in this work. The impact on the accuracy was minor for the studied system (not shown). In general, if the method is applied to systems with many different reaction paths or if only few experiments are available, the calibration of subsystems would probably increase stability.

3.4 Results and Discussion

The model-free calibration explained in Section 3.3 is applied to the Raman data for the 30.4 wt-% MEA/water solution described in Section 3.1. As a first step, the weights of all components have been determined as described in Section 3.2 (*cf.* Figure 3.2). The model-free calibration is validated by comparing model-free calibration results with a thermodynamic model. In this work, only Raman measurements of the system in chemical equilibrium have been used. Therefore, the results can be compared to an equilibrium model from the literature (Aronu et al. (2011)).

The speciation obtained by the model-free calibration is compared to the thermodynamic model in Figure 3.3. As quantitative measure, the root-mean-square error (RMSE) is used (*cf.* Equation (2.19)). The speciations by the model-free calibration and the thermodynamic model are in good agreement with a RMSE of 0.141 mol/L.

It is important to keep in mind that the thermodynamic model has not been used for the calibration. Equation 3.3 which was used for the fit of the model-free calibration is substantially different from Equation (2.19).

We observe the highest deviations for the speciation of $MEAH^+$. The spectral pure component model of $MEAH^+$ has been obtained at CO_2-free conditions, and thus the increasing nonlinear effects at higher loadings might not be captured completely. This can be shown by calibrating the same data using a model-based calibration with the thermodynamic model (Figure 3.4). The model-based fit yielded only a slightly

improved RMSE of 0.115 mol/L. While the calibration results of $MEACOO^-$ and HCO_3^- are slightly improved, the calibration results for $MEAH^+$ remain practically unchanged. Therefore, we believe that the increased deviation of $MEAH^+$ is mainly due to the spectral pure component model of $MEAH^+$ rather than the model-free calibration. Furthermore, we assumed a constant refractive index as stated in Section 3.3. Correcting for the effects of the refractive index might even slightly improve the RMSE. Overall, the good agreement of the model-based and model-free calibration validates the model-free calibration method.

The model-free calibration has been performed with a freshly prepared MEA/water solution. Therefore, we assumed that neither oxidative or thermal degradation products nor other impurities were present. As we did not include any side reactions in the model-free calibration, the absence of impurities and degradation is a prerequisite for the model-free calibration. Nevertheless, the calibrated model could then in principle be used to analyze process streams containing impurities. We expect that the analysis of process streams containing impurities should work if the spectral evaluation is not severely affected by the degradation products (*e.g.*, by fluorescence). No further calibration would be required as the calibration coefficients $\mathbf{k}$ are not affected by impurities. The impurities and degradation products, however, could not be quantified.

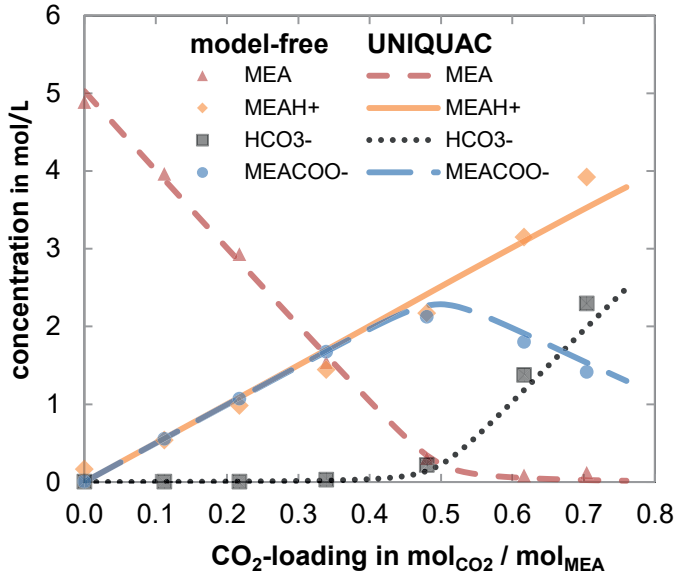

Figure 3.3: Results of the model-free calibration (symbols) compared to data calculated with a UNIQUAC model from Aronu et al. (2011) (lines). The composition of the unloaded solution was 30.4 wt-% in 69.6 wt-% water. The thermodynamic model (lines) has not been used within the calibration procedure and serves as reference only.

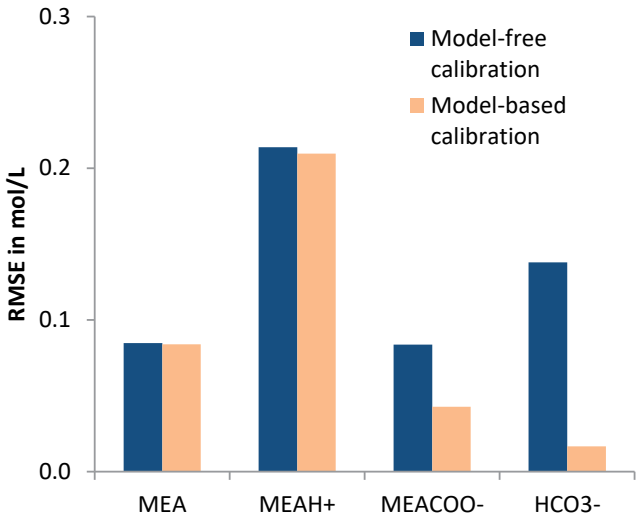

Figure 3.4: Root-mean-square errors (RMSE) of a model-based calibration and the model-free calibration for each species. The RMSE indicates the deviation between concentrations determined by the respective calibration and concentrations determined by the thermodynamic model from Ref. Aronu et al. (2011).

3.5 Conclusions

A method for model-free calibration of spectra is presented. To demonstrate the method, Raman spectroscopy was chosen as it is a quantitative method which is able to quantify multiple components in-situ. The Raman spectra of aqueous solutions of monoethanolamine (MEA) loaded with CO_2 were evaluated with Indirect Hard Modeling (IHM) to extract the weights of the pure components in the spectra. The present study focused on Raman spectroscopy; the model-free calibration is generic and should thus be applicable to most spectroscopic techniques.

For model-free calibration, only well-known relations of the chemical reaction mechanism such as stoichiometry and electroneutrality were applied. Thereby, the method is applicable to novel amines or other chemical reaction mechanisms. The method has been validated by comparison to results of a thermodynamic model. It has been demonstrated that very accurate speciation can be obtained.

Model-free calibration can even be advantageous if a thermodynamic model is available. Measurement errors become obvious, as the method calibrates independently of the thermodynamic model while calibration to a thermodynamic model bears the risk of overfitting the spectral data.

Chapter 4

First-Derivative Indirect Hard Modeling (FD-IHM)

Parts of this Chapter have been published in:

P. Beumers, D. Engel, T. Brands, H.-J. Koss, A. Bardow, "Robust Analysis of Spectra with Strong Background Signals by First-Derivative Indirect Hard Modeling (FD-IHM)", *Chemometrics and Intelligent Laboratory Systems*, vol. 172, pp. 1-9, 2017

Contribution record: Principal author, author of the first draft, development and implementation of the method, planning and conduction of the experiments, data evaluation.

IHM is sensitive to background signals. Background signals may lead to poor results if they are not taken into account by data pretreatment or within the mixture model. (see Section 2.2.1).

In statistical methods (e.g., Partial Least Squares, PLS), backgrounds are often removed by using the first or second derivative of a spectrum (Lieber and Mahadevan-Jansen (2003); Mariani et al. (2015); Ozaki et al. (2001); Rahman et al. (2016); Zhou et al. (2012)). The underlying rationale is that the narrow peaks of the spectral signal are more pronounced in the derivatives of a spectrum than the broad peaks of the background signal (Brown et al. (2000)). Using derivatives is a well established and successful method to pretreat data. In statistical methods, derivatives are a pure pretreatment method since the methods themselves work the same for the spectra and their derivatives.

Due to the well established effectiveness of derivatives to pretreat spectra, it seems natural to apply derivatives to IHM. Unlike statistical methods, which can often be applied to derivatives of spectra without any change of the method itself, IHM can-

not be applied to derivatives of spectra directly, as IHM is unable to model negative parts of a spectrum. To overcome these limitations, we introduce First-Derivative indirect Hard Modeling (FD-IHM). The basic idea is to replace all mathematical peak functions by their analytical derivative. We show that FD-IHM leads to excellent calibration results while reducing manual effort, since the need for background treatment is removed. FD-IHM is validated with multiple sets of Raman spectra. Nevertheless, FD-IHM should be applicable to other spectroscopic techniques alike (e.g., NMR or Mid-Infrared).

We introduce FD-IHM in Section 4.1. In the experimental Section 4.2, we apply IHM and FD-IHM to the system ethanol-acetone with intentionally introduced large backgrounds using (i) external light sources, (ii) yeast, and (iii) rhodamine B. We discuss our results in Section 4.3. Finally, we give our conclusion in Section 4.4.

4.1 First-Derivative Indirect Hard Modeling (FD-IHM)

FD-IHM facilitates the evaluation of spectra with pronounced background signals. In classical IHM, background signals require dedicated background models, which have to be selected very carefully (see Section 2.2.1). In FD-IHM, the main idea is to apply the well-established concept to differentiate the spectra to filter the background signals. The first derivative of a broad background signal is generally much smaller than the first derivative of the peaks (Baek et al. (2009)). To clearly distinguish IHM (no derivative) and FD-IHM (first derivative), we use "FD" as prefix for all models and spectra that are differentiated, *e.g.*, "FD-spectral fitting".

The FD-spectral fitting follows the same steps as outlined in Figure 2.2 for the spectral fitting of IHM: (a) FD-peak-fitting: fit FD-component models to pure component spectra; (b): choose free parameters; (c) FD-component fitting: fit the FD-mixture model to mixture spectra.

An FD-component model is a sum of analytical derivatives of pseudo-Voigt functions. Each FD-component model is fitted to the numerical derivative of a pure component spectrum in the FD-peak-fitting. As numerical derivative, we use a finite difference quotient of two neighboring pixels: $\left(\frac{\Delta S_k^*}{\Delta \tilde{\nu}}\right)_m = \frac{S_K^*(\tilde{\nu}_{m+1}) - S_k^*(\tilde{\nu}_m)}{\tilde{\nu}_{m+1} - \tilde{\nu}_m}$. The analytical derivative is calculated in the center of these two pixels: $\left(\frac{\mathrm{d}}{\mathrm{d}\nu}\right)_m \mathcal{A}^*\left(\tilde{\nu}, \Theta_k^*\right) = \frac{\mathrm{d}}{\mathrm{d}\nu}\mathcal{A}^*\left(\frac{\tilde{\nu}_m + \tilde{\nu}_{m+1}}{2}, \Theta_k^*\right)$. With these definitions, the equation for FD-peak fitting reads:

$$\boldsymbol{\Theta}_k^* = \arg\min_{\boldsymbol{\Theta}_k^*} \sum_{m=1}^{n_{\tilde{\nu}}-1} \left[\underbrace{\left(\frac{\Delta S_k^*}{\Delta \tilde{\nu}}\right)_m}_{\substack{\text{numerical derivative} \\ \text{of pure component spectrum } k}} - \underbrace{\left(\frac{\mathrm{d}}{\mathrm{d}\tilde{\nu}}\right)_m \mathcal{A}^*\left(\tilde{\nu}, \boldsymbol{\Theta}_k^*\right)}_{\text{FD-component model}} \right]^2 . \tag{4.1}$$

In contrast to the peak fitting equation of IHM (Equation (2.10)), FD-IHM peak-fitting Equation (4.1) uses derivatives. Most importantly, FD-IHM employs no background model. For the numerical derivative of the measured spectra, a simple finite difference quotient has been found to be sufficient for the studied spectra in this work, as the spectra have a very favorable signal-to-noise ratio. In other cases, more advanced methods may be considered (Hanke and Scherzer (2001)), *e.g.*, Savitzky-Golay filters (Savitzky and Golay (1964)) or Smooth Noise Robust Differentiators (Holoborodko (2008)). There is already a large experience on the differentiation of noisy spectra from the pre-treatment for statistical methods (for a review on spectral pre-processing methods see Rinnan et al. (2009)). Nevertheless, within the scope of this work, there was no need to apply any such noise filtering.

As the FD-component models of FD-IHM are the analytical derivatives of the component models of IHM, the parameters used, $\boldsymbol{\Theta}_k^* = (\boldsymbol{\delta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}, \boldsymbol{\beta})$, are the same for both methods. Therefore, the parameters of an FD-component model fitted with FD-peak fitting can be used as parameters for an IHM component model and vice-versa. Nevertheless, the objective functions for peak-fitting, Equations (2.10) and (4.1), will lead to different parameter values. Based on our experience, we recommend to use peak fitting (Equation (2.10)) to establish component models for IHM and to use FD-peak fitting (Equation (4.1)) to establish FD-component models for FD-IHM.

The choice of adjustable parameters $\boldsymbol{\Theta}_k$ is based on the residuals of the FD-component fitting step, similar to IHM (Figure 2.2). In the FD-component fitting step of FD-IHM (Figure 2.2c), the analytical derivatives of the FD-component models are fitted to the numerical derivative of the mixture spectra. The FD-component fitting is thus obtained from:

$$\boldsymbol{\zeta} = \underset{\boldsymbol{\zeta},\boldsymbol{\Theta}_1 \ldots \boldsymbol{\Theta}_{n_C}}{\arg\min} \sum_{m=1}^{\tilde{\nu}-1} \left\{ \underbrace{\left(\frac{\Delta S^*_{\text{mixture}}}{\Delta \tilde{\nu}} \right)_m}_{\substack{\text{numerical derivative} \\ \text{of the spectrum}}} - \underbrace{\sum_{k=1}^{n_C} \underbrace{\left(\frac{\mathrm{d}}{\mathrm{d}\tilde{\nu}} \right)_m \mathcal{A}^* \left(\tilde{\nu}, \boldsymbol{\Theta}_k, \boldsymbol{\Theta}_k^{'} \right)}_{\text{FD-component models}}}_{\text{FD-mixture model}} \right\}^2 . \tag{4.2}$$

Note that compared to Equation (2.11) no background model is needed, *i.e.*, ($\mathcal{B}(\tilde{\nu}_m, \Theta_B) = 0$).

As discussed earlier, the areas of the (non-differentiated) pseudo-Voigt functions are commonly assumed to be constant. To meet this condition in FD-IHM, the area parameter (α) is kept constant during the FD-component fitting. The FD-component fitting results in the weights ζ, as in IHM. The weights ζ are processed with the same calibration and data analysis procedure as described for IHM in Section 2.2.4.

4.1.1 Illustrating the FD-IHM Idea

To illustrate the benefits of using the first derivative in IHM, we created artificial spectra by adding three types of background signals to the same signal (Figure 4.1 a-c). The artificial spectra were modeled by fitting a parametrized pseudo-Voigt function. As background signal, we added a constant, a linear, and a quadratic polynomial. It can be seen that IHM is disturbed by any type of non-vanishing background signal (Figure 4.1a-c). To account for these background signals, a background model has to be incorporated within the mixture model (see Equation (2.11)) or pretreatment of the spectrum has to be applied to remove the background signal prior to component fitting. To illustrate the advantage of FD-IHM, we fitted the derivative of the pseudo-Voigt function to the derivatives of the artificial spectra (Figure 4.1d-f). Constant background signals vanish completely in the first derivative (Figure 4.1d). In the least squares fit, even a linear background (constant gradient) does not deteriorate the fit results (Figure 4.1e). Such a constant gradient leads to a constant offset in the first derivative that cannot be fitted by the derivative of a peak function. As any broad background is almost linear in the range of a single peak (see also Figure 4.2) neither a background model nor the second derivative of a spectrum is needed for FD-IHM to give very accurate results. Only strongly curved background signals disturb the spectral fitting of FD-IHM (Figure 4.1f).

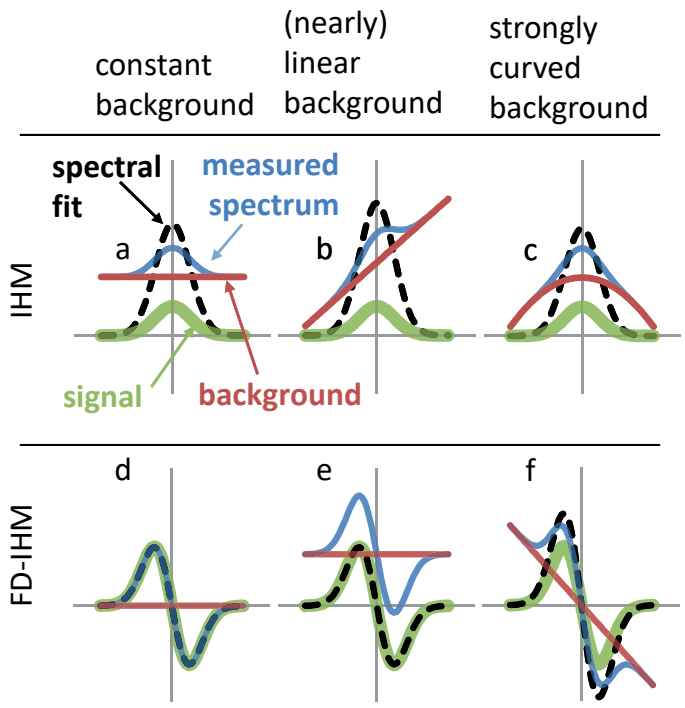

Figure 4.1: Robustness of IHM and FD-IHM for uncorrected background signals (red) which superimpose the signals (green, bold) leading to the measured spectra (blue). For FD-IHM, the derivative of the measured spectrum, the signal and the background are shown. Ideally, the spectral fit results (black, dashed) have the same shape and amplitude as the signals. The spectral fits of IHM are disturbed by any type of non-vanishing background (a-c), which makes a thorough background treatment necessary. The spectral fit of FD-IHM is not disturbed by constant or linear background signals (d, e). Only strongly curved background signals disturb the spectral fit of FD-IHM (f).

The FD-IHM method is further illustrated in Figure 4.2 with a real Raman spectrum of a mixture of ethanol, acetone and, to create a large background, small amounts of rhodamine (for details see Section 4.2). In Figure 4.2a, the component fitting step of FD-IHM is illustrated: A sum of weighted component models is fitted to the numerical derivative of an experimental mixture spectrum in a least squares fit (Equation (4.2)).

To facilitate the visual assessment of the fit results, the fitted mixture model and the measured spectrum are plotted in the non-differentiated form (Figure 4.2b). Since derivative spectra are usually more difficult to interpret for users, one of the major

disadvantages of FD-IHM encountered in our experience can be overcome by the possibility to directly visualize the spectra rather than the derivatives. A residual background is calculated by subtraction of the mixture model from the measured spectrum. A residual background without any peaks of the pure components (see yellow plot in Figure 4.2b) indicates a good fit. The comparison of the spectrum and the FD-spectrum shows that noise is amplified in the derivative. Still, FD-IHM distinguishes well between signal and noise leading to stable and robust identification of the spectral information without requiring a custom-made background model.

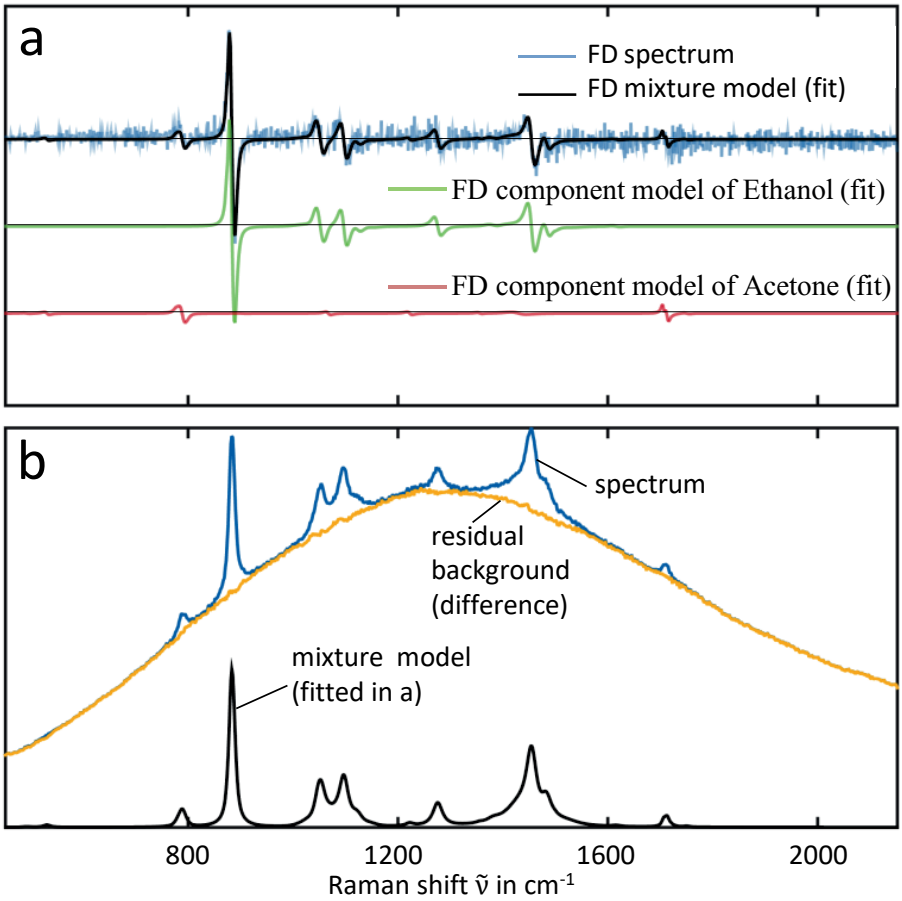

Figure 4.2: FD-IHM for Raman spectra of ethanol, acetone and small amounts of rhodamine B. (a) FD-Component fitting: An FD-mixture model is fitted to an FD-spectrum; (b) integrated FD-model and spectrum. The residual background obtained by subtraction of the integrated FD-model from the spectrum does not show any peaks.

4.2 Experimental

To compare the robustness of IHM and FD-IHM with respect to background signals, we apply both classical IHM and the proposed FD-IHM method to Raman spectra of mixtures of ethanol and acetone. To introduce realistic background signals that differ in shape and intensity, we used (i) background lighting, (ii) fluorescence dye, and (iii) scattering bodies.

4.2.1 Sample Preparation and Spectral Acquisition

Experiments were carried out with acetone (Uvasol) and ethanol (Uvasol) purchased from Merck Millipore (Germany). All chemicals were used without further purification. To create a fluorescent background, we added small random amounts of rhodamine B to some mixtures of acetone and ethanol. As scattering bodies, we used Yeast (saccharomyces cerevisiae) from a local convenience store.

We prepared three physical sample sets of ethanol-acetone solutions. From each sample set, we recorded one or more spectra data sets. To distinguish physical sample sets from spectra data sets, we use Roman numerals for physical sample sets I, II, and III and lowercase letters for spectra data sets a-h.

Sample set I: 11 samples in steps of 10 wt-% ethanol without anything else added. From these samples, we recorded a data set of clean Raman spectra (no added background signal) for calibration (Figure 4.3). From the same sample set, we recorded Raman spectra containing various types of background signals induced by background lighting: (1) green LED, (2) white LED, (3) photographic flash with a red filter attached, (4) photographic flash, (5) 12V halogen spotlight (Figure 4.4c-g). The approximate spectra of the used light sources can be deduced from Figure 4.4. Any light source that does not have distinct peaks in the spectral range can serve as exemplary background in this context.

Sample set II: 11 samples in steps of 10 wt-% ethanol. For this sample set, we recorded a clean spectra data set to validate the models with a background free spectra data set (Figure 4.4a). Then, we added small amounts of yeast to the samples and recorded another spectra data set (Figure 4.4h). The yeast is not considered for the determination of mass fractions, *i.e.*, these mixtures are still evaluated as a binary mixture.

Sample set III: Mixtures with rhodamine were prepared with pure ethanol, pure acetone, and a solution of small amounts of rhodamine in ethanol. To prepare the ethanol/rhodamine solution, we dissolved a small amount of rhodamine in 10 ml of

ethanol (less than 0.1 mg rhodamine). This solution was further diluted (roughly 1:100), so that a few drops of the ethanol/rhodamine solution were sufficient to create the desired background signal in a total mixture of roughly 2 mL. The mass fraction of rhodamine in the sample is in the order of $w = 10^{-7}$ or lower and has therefore been neglected in the analysis of the composition. We treated the ethanol/rhodamine solution as pure ethanol in the gravimetrical preparation of the mixtures. In the following analysis, we assume a binary mixture of ethanol and acetone in terms of mass fractions. Note that we used some drops of rhodamine/ethanol solution and pure ethanol for each sample. Therefore, the amount of rhodamine is not proportional to the ethanol amount. The rhodamine-ethanol-acetone spectra data set is shown in Figure 4.4b.

From the accuracy of the balance used for gravimetrical preparation, we calculate an absolute uncertainty in mass fractions of $\Delta w < 10^{-4}$ for all samples.

Table 4.1 summarizes all recorded spectra data sets and the underlying sample sets. For sample sets I and II, we recorded 10 and 20 repeated Raman spectra respectively for each combination of sample and background. For samples without artificially induced background, we recorded 5 repeated Raman spectra. For sample set III, we recorded 1 spectrum per sample.

To build the mixture models, we use the averaged calibration set. Here, the term "averaged set" indicates that we used the average of the repeated Raman spectra of the same sample. For validation, we used the individual Raman spectra of each spectra data set.

All Raman spectra were recorded with an inverse confocal Raman microscope at 532 nm excitation wavelength (Renishaw inVia: Leica microscope, 1800 l/mm grating, spectral resolution 1-2 cm^{-1}) at room temperature. Exposure time was adjusted so that the maximal intensity was in the range of approximately 40k to 60k counts for each individual spectrum (see Figures 4.3 and 4.4). 60k counts correspond to approximately half of the maximum intensity of the detector for the selected settings. In Figures 4.3 and 4.4, averaged Raman spectra are shown. Note that some Raman spectra of spectra data set h (Figure 4.4h yeast) have a very low signal due to absorption of the laser signal at the yeast.

4.2.2 Model Setup

Following the IHM approach and the FD-IHM approach, we built mixture models for the mixture of ethanol and acetone. The IHM mixture model is used as basis for

Table 4.1: Spectra data sets used for calibration and test-set validation.

spectra data set	induced background	sample set	number of repeated Raman spectra per sample
calibration	none	I	5
test set a	none	II	5
test set b	rhodamine	III	1
test set c	green LED	I	10
test set d	white LED	I	10
test set e	red flash	I	10
test set f	white flash	I	10
test set g	halogen bulb	I	10
test set h	yeast	II	20

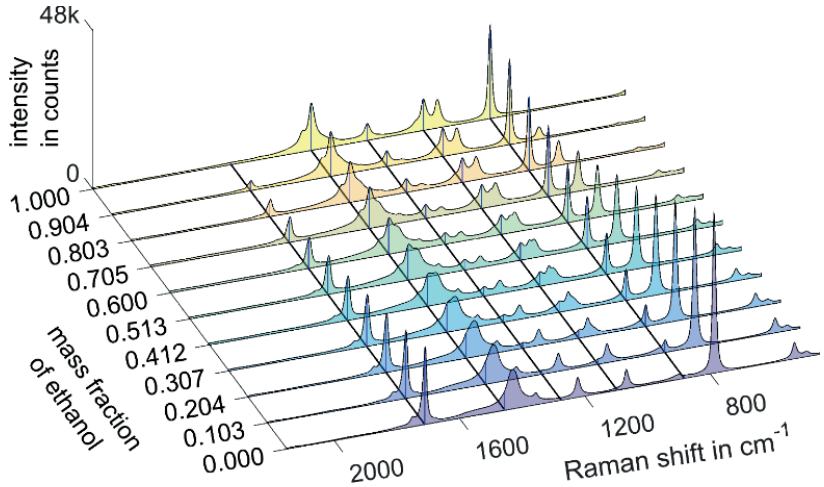

Figure 4.3: Averaged data set for the acetone-ethanol system used for calibration. Some peaks are marked to guide the eye.

different approaches of background treatment detailed in the following. The FD-IHM mixture model is used without any treatment of the background.

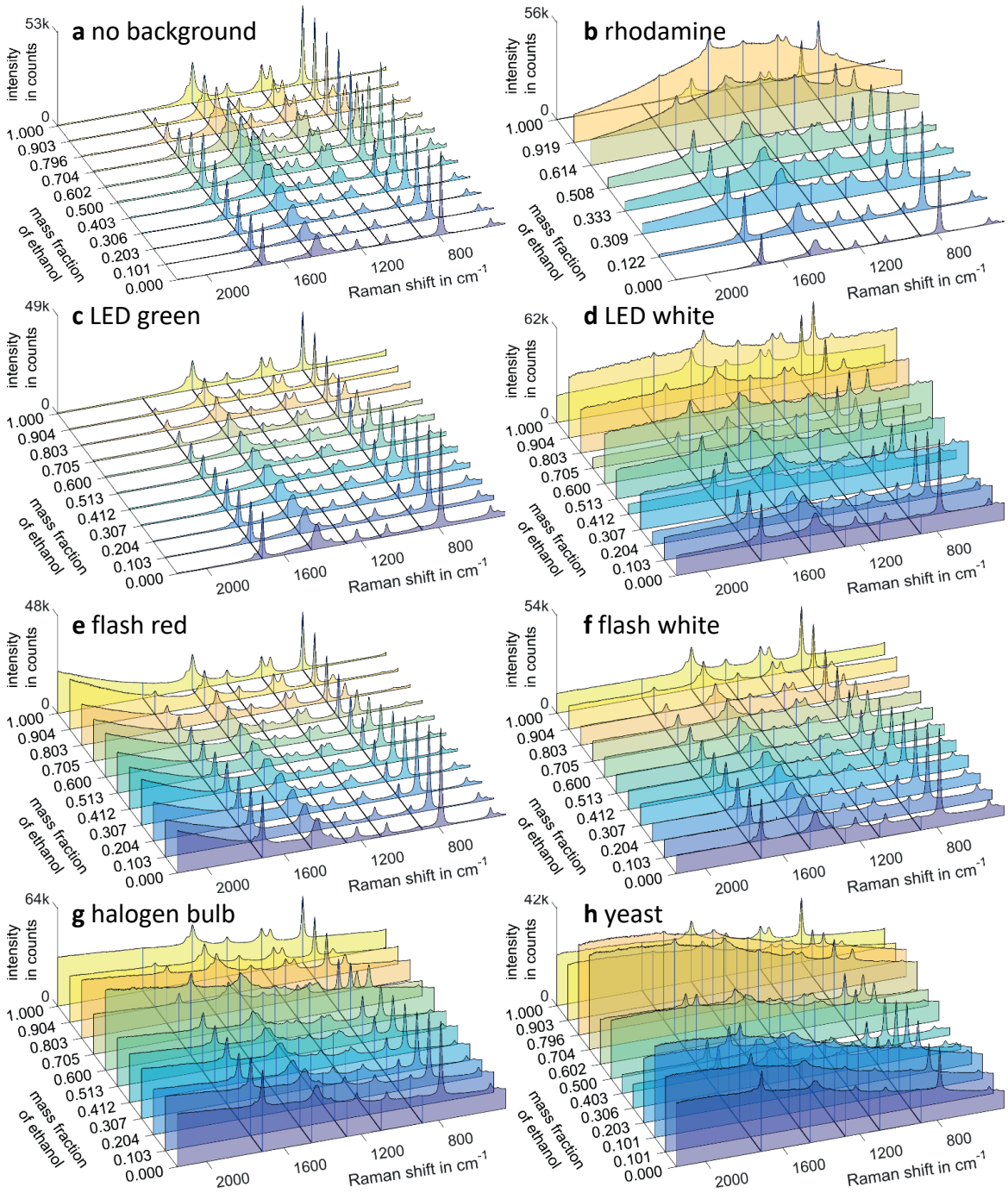

Figure 4.4: Averaged test sets for the acetone-ethanol system with varying backgrounds. Backgrounds of b and h are induced by substances added to the samples. Other backgrounds are induced by external light sources. Some peaks are marked to guide the eye. (For details, see Table 4.1)

4.2.3 Spectral Evaluation (Component Fitting)

We evaluate the spectra data sets using the FD-IHM mixture model, which contains no background model. To compare FD-IHM to alternative strategies, we also created mixture models following the IHM approach and incorporated a variety of background treatment methods:

- NONE: no background treatment,
- CONSTANT: a constant background model (one degree of freedom),
- LINEAR: a linear background model (two degrees of freedom),
- PARABOLIC: a polynomial background model (parabola, three degrees of freedom),
- PEAK2: background modeled by six broad peaks, fitted to the shape of the largest background signal of the specific rhodamine background ($w_{\mathrm{EtOH}} = 0.919$ in Figure 4.4 b) via complemental hard modeling (Kriesten et al. (2008b)). The sum of these peaks is scaled during the spectral fitting and a variable offset is fitted (1+1=2 degrees of freedom),
- PEAK14: same as (Peak2) but with two degrees of freedom per peak (position δ and FWHM γ, 6*2+1+1=14 degrees of freedom),
- RBB: background subtraction as a pretreatment step, using the rubber band method as implemented in the commercial software PEAXACT (version 4.0, S-PACT GmbH, Germany). The rubber band method computes the convex hull of the spectrum and subtracts the convex hull from the original spectrum to get the baseline corrected spectrum. Optionally, piecewise removal of the convex hull is possible to better account for concave baselines. The latter option has been applied in this work.

The IHM mixture models NONE, CONSTANT, LINEAR, PARABOLIC, PEAK2, and PEAK14 are fitted to unpretreated Raman spectra. The IHM mixture model RBB is fitted to Raman spectra pretreated by the rubber band method. The FD-IHM mixture model is fitted to the numerical derivatives of the Raman spectra.

4.2.4 Calibration and Validation

The calibration and validation steps are generic and thus applied for all mixture models as detailed in Section 4.2.3. Each mixture model is used for component fitting (see Equation 2.11 for IHM and Equation 4.2 for FD-IHM).

Based on the component fitting, we calibrate each mixture model separately with the calibration set shown in Figure 4.3 using Equation (2.18). For validation of the calibrated mixture models, we predict the mass fractions for each individual spectrum in each test set (Equation 2.17) and compare the results to the true mass fractions.

To assess the predictive abilities of the calibrated mixture models, we employ the root mean square error of prediction (*RMSEP*, *cf.* Equation (2.19)) and the maximum absolute prediction error (*MAXPE*, *cf.* Equation (2.20)). Both values reflect the robustness against changing spectral background signals.

4.3 Results and Discussion

Raman Spectra of ethanol-acetone mixtures with strong background signals have been used for validation employing IHM and FD-IHM. Multiple strategies to handle backgrounds with IHM (NONE, CONSTANT, LINEAR, PARABOLIC, PEAK2, PEAK14, and RBB) as well as FD-IHM are tested (for details, see Section 4.2.3). All mixture models were calibrated using the calibration set shown in Figure 4.3 and validated with the clean test set a (not shown). The test sets are shown in Figure 4.4. Details on all spectra data sets are given in Table 4.1.

In Figure 4.5, we show the validation results for all mixture models using the clean Raman spectra from test set a. FD-IHM outperforms all IHM mixture models with an *RMSEP* of 0.19 wt-%. Most IHM mixture models perform similarly (*RMSEP* close to 0.5 wt-%). Only the IHM model without any background model (NONE) performs significantly worse. It is surprising that FD-IHM performs better for spectra data set a. We anticipated IHM to be suited better for spectra data sets without backgrounds, as no background treatment is needed. The maximum error *MAXPE* is approximately twice as large as the *RMSEP* for all mixture models. From the *MAXPE* we see that no significant outliers are present.

In Figure 4.6, we show the validation results for test set b (rhodamine). FD-IHM outperforms all IHM methods by a factor of at least 2.6 for the RMSEP (Equation (2.19)) and a factor of at least 2.2 for the MAXPE (Equation (2.20)). Interestingly, the rubber band method still delivers slightly better results than the method PEAK14 that uses a background model adapted specifically to the rhodamine background (see Section 4.2.3). This emphasizes how challenging it is to model a large fluorescent background accurately. In terms of practical applicability, a background model with 14 parameters has the major disadvantage that it greatly extents evaluation time. The rigid modelling of the background (PEAK2) and the generic polynomial background

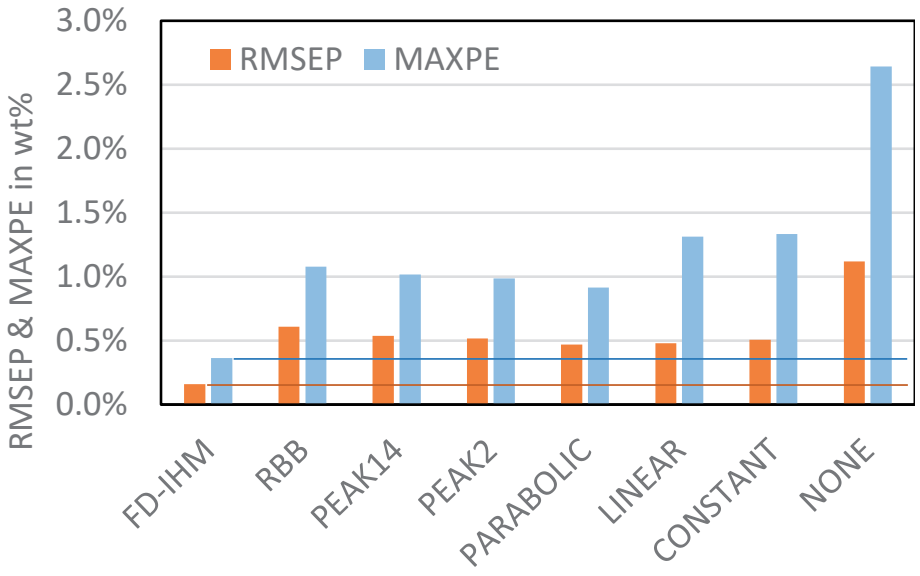

Figure 4.5: Validation results (root-mean-square error of prediction RMSEP (Equation (2.19)) and maximum absolute prediction error MAXPE (Equation (2.20)) using averaged test set a (no background).

models (NONE, CONSTANT, LINEAR, and PARABOLIC) perform worst. For these poor models, the RMSEP errors are in the range from 3 wt-% to more than 10 wt-%.

To provide a broader basis of validation results, we compare FD-IHM to the most promising mixture models of IHM for all test sets (a-h). For IHM, we limit the choice of mixture models to one generic pretreatment method (RBB) and to one reasonably flexible generic background model (PARABOLIC). The mixture models with rhodamine specific background models (PEAK2, PEAK14) and the mixture models with inflexible background models (NONE, CONSTANT, LINEAR) are discarded. The resulting RMSEP and MAXPE are given in Figures 4.7 and 4.8 respectively.

For all studied test sets, FD-IHM delivers the best results concerning both root-mean-square error RMSE (*cf.* Figures 4.7) and maximum-absolute-prediction error MAXPE (*cf.* Figure 4.8). The good performance of FD-IHM concerning MAXPE for averaged and individual Raman spectra shows that FD-IHM rarely produces large outliers, even though we calculated the derivatives with a finite difference quotient without any smoothing. The largest absolute prediction error of FD-IHM for all sets (over 800 predictions) is less than 5 wt-%.

Compared to FD-IHM, IHM with rubberband pretreatment (RBB) comes closest to FD-IHM. RBB leads to larger RMSEPs (plus 135% on average) and MAXPEs (plus

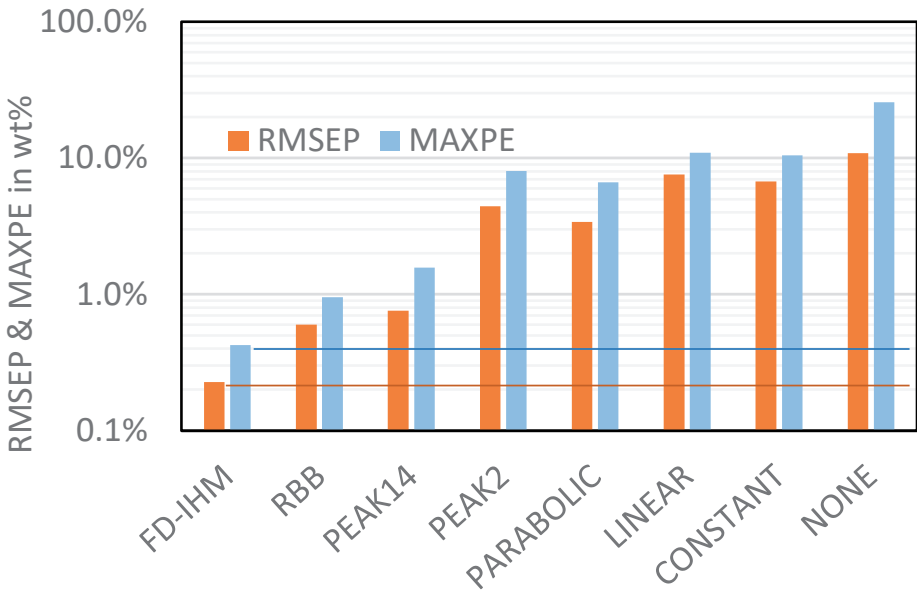

Figure 4.6: Validation results (root-mean-square error of prediction RMSEP, Equation (2.19)) and maximum absolute prediction error MAXPE (Equation (2.20)) for test set b (rhodamine).

60% on average) alike.

For the spectra data sets d and h, FD-IHM shows slightly elevated RMSE and MAXPE values (*cf.* Figures 4.7 and 4.8). This is probably due to an elevated noise level in these spectra data sets. Note that Figures 4.7 and 4.8 show the results for the evaluation of individual spectra while Figure 4.4 shows the average for multiple repeated spectra of these spectra data sets (*cf.* Section 4.2.1). While FD-IHM still outperforms the non-derivative methods for these examples, we would recommend FD-IHM rather for applications where backgrounds are the primary concern than for spectra with very poor signal-to-noise ratios. The performance of FD-IHM is also expected to deteriorate if actual peaks show similar behavior as the background (*e.g.*, a very broad peak).

The parabolic background treatment (PARABOLIC) is clearly not sufficiently flexible to handle the various backgrounds. Only if the shape does not deviate too much from a parabola (*cf.* Figures 4.7 and 4.8 “LED green” and “flash white”) the performance is still acceptable.

Besides FD-IHM showing the best performance in these controlled experiments with their relatively clean conditions during both calibration and validation, even greater benefit lies in the expected long-term robustness for industrial applications. In our

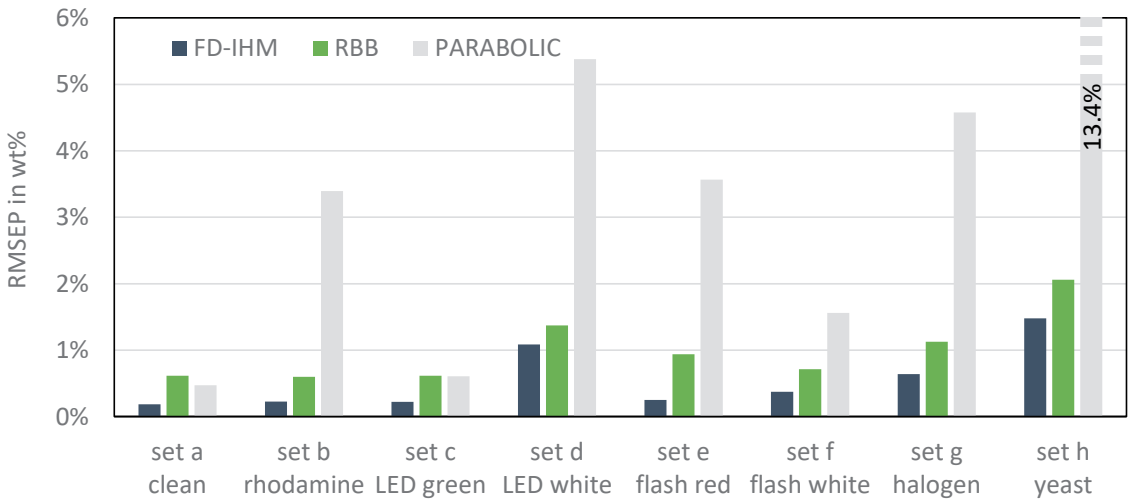

Figure 4.7: Root-mean-square error of prediction (RMSEP, Equation (2.19)) for all test sets (see Table 4.1).

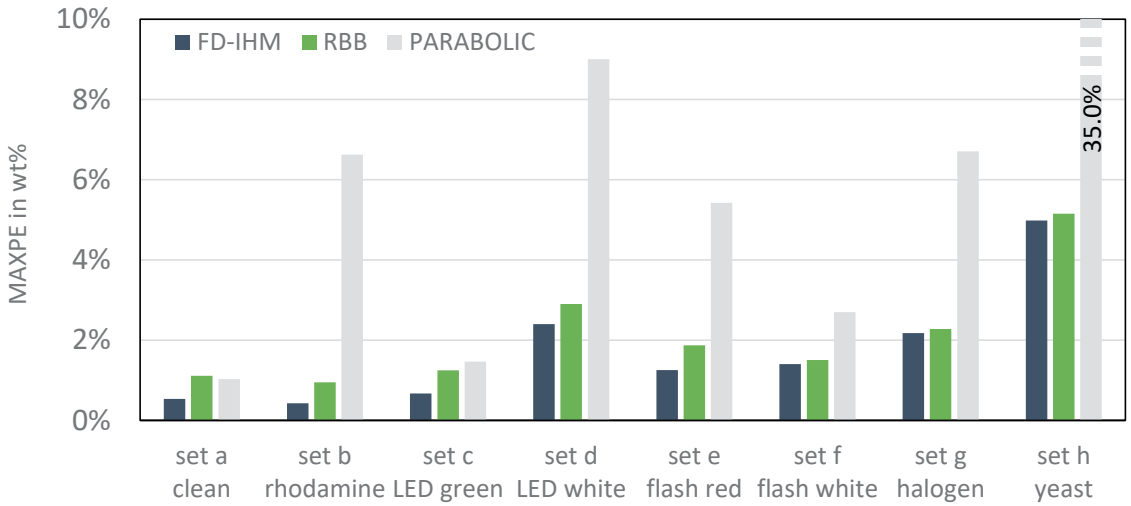

Figure 4.8: Maximum absolute prediction error (MAXPE, Equation (2.20)) for all test sets (see Table 4.1).

experience, production conditions slightly change over time and lead to baseline drifts, fluorescence, etc. Having a robust FD-IHM model that inherently corrects for those changes keeps the prediction error low and reduces model maintenance costs in the long run.

4.4 Conclusions

First-Derivative Indirect Hard Modeling (FD-IHM) is presented as an extension to Indirect Hard Modeling (IHM). FD-IHM combines the strength of IHM regarding treatment of nonlinear effects with the benefit of using the first derivative for handling background signals, as often used in statistical methods such as PLS. IHM is a physically motivated method for spectra analysis. A major advantage is the incorporation of nonlinear effects such as peak shifts into the spectral fitting procedure. Background signals are usually treated by polynomials, via pretreatment, or with peaks fitted to the background. In contrast to IHM, FD-IHM does not need any models of background signals.

For the spectral analysis of a binary system (ethanol and acetone) with significant background signals, the benefits of FD-IHM over IHM were demonstrated. We investigated fluorescent backgrounds induced by rhodamine, backgrounds induced by external light sources, and backgrounds due to stray light induced by yeast. Our results show that a very good calibration can be achieved using classical IHM. However, background signals must be treated very carefully with IHM. Validation provides root-mean-square error of prediction (RMSEP) in the range of 0.60 wt-% to 2.06 wt-% for Raman spectra pretreated with the rubberband method. IHM with rubberband pretreatment was the only robust IHM method that provided stable results for all investigated spectra sets. Attempts to model the background by generic polynomial models failed for large backgrounds.

Using FD-IHM for the studied system reduces the RMSEP errors even further. The difference between IHM with rubberband pretreatment and FD-IHM is in the range of 0.29 wt-% and 0.69 wt-% which corresponds to 26% to 277%. From the maximum error MAXPE, we see that FD-IHM does not produce large outliers even though derivatives of the Raman spectra have not been smoothed. For each of the investigated systems, FD-IHM led to the best results concerning RMSEP and MAXPE.

FD-IHM does not require any background treatment, it should thus work for any background as long as the background does not have any sharp peaks interfering with peaks of the pure component Raman spectra. The same FD-IHM model is shown to work for all investigated backgrounds without any further adaption.

In summary, FD-IHM outperforms all standard IHM methods for the studied samples. We see the promising results as validation of the FD-IHM method. While no general claims can be made from this case study, we believe that FD-IHM is a powerful and efficient method to treat strong varying background signals in spectral analysis. Consequently, we recommend FD-IHM when robustness against unpredictable back-

ground signals is important.

Chapter 5

Robust Spectra Evaluation: From Indirect Hard Modeling to Very Hard Modeling

In this Chapter, we present Very Hard Modeling (VHM) which is based on the IHM framework but in contrast to IHM, VHM does not treat the concentration-dependent peak changes (*i.e.*, nonlinear effects) as a threat to the spectral evaluation. Instead, VHM exploits the additional information contained in the nonlinear effects to improve the spectral evaluation. VHM parametrizes the peak changes as function of composition. Thereby, VHM uses more information provided by the calibration spectra than IHM. As a result, a VHM model reduces the number free parameters in the spectral evaluation and shows increased robustness. We use a methanol+water system to demonstrate that VHM models are more reliable and have an increased robustness in the face of noisy spectra and spectra with backgrounds. In this Chapter, we first present the VHM-method in Section 5.1. We present a comparison of VHM and IHM at the end of Section 5.1. In Section 5.2 we present the experimental procedure which comprises the spectral acquisition as well as the calibration and application of VHM and IHM models. IHM serves as a benchmark for the VHM method. In Section 5.3 we present the results of IHM and VHM together with a discussion. In Section 5.4 we give our conclusion.

5.1 The Very Hard Modeling Method

VHM uses the same pure component models to represent pure component spectra as IHM (*cf.* Section 2.2.2). For IHM, the parametrized form of the pure component models (in contrast to pure component spectra) enables the modeling of peak changes

in an easy manner. VHM uses the pure component models as a basis and models the spectral changes occurring in the mixture as function of composition. The VHM algorithm is presented in Section 5.1.1. In Sections 5.1.2 and 5.1.3, we present the calibration and application of a VHM model respectively. In Section 5.1.4, we compare VHM to IHM.

5.1.1 The VHM Mixture Model

A VHM mixture model is basically a weighted superposition of the pure component models. In the VHM mixture model, all peak shifts and deformations depend on the composition of the mixture.

In the following, we describe the algorithm to calculate a spectrum with a VHM mixture model step by step as shown in Figure 5.1. The procedure assumes known mass fractions w and calibration factors k from calibration (Section 5.1.2).

Step 1: Determine peak parameters for the actual composition

In this step, the actual peak parameters $\tilde{\mathbf{\Theta}}$ are determined. The parameters are based on the pure component parameters $\mathbf{\Theta}_i^*$ determined by peak fitting (see Section 2.2.2) and an empirical function $f(\mathbf{w})$ that describes composition-dependent variations. $\mathbf{w}$ denotes the composition in mass fractions. The equation for the actual peak parameters $\tilde{\mathbf{\Theta}}$ reads:

$$\tilde{\Theta}_{i,p,q}\left(\mathbf{w}\right) = \Theta^*_{i,p,q} + f\left(\mathbf{w}\right). \tag{5.1}$$

Here, p denotes the peak of component i, and q denotes the parameters of the peak ($q = [\alpha, \beta, \gamma, \delta]$), *cf.* Section 2.2.2. In a spectroscopically ideal system, the mixture spectrum would be a linear superposition of the pure component spectra. VHM addresses non-ideal systems where peak changes occur in the mixture. These deviations from ideal behavior are modeled by the empirical function $f(w)$, which can thus be regarded as spectral excess function. In this work, we use polynomial functions for $f(\mathbf{w})$. Mass fractions are employed as basis as they have proved to work best. Alternatives are discussed in Appendix C. In this Chapter, we limit the presentation to binary mixtures where Equation (5.1) for the concentration-dependent peak parameters $\tilde{\mathbf{\Theta}}$ reads:

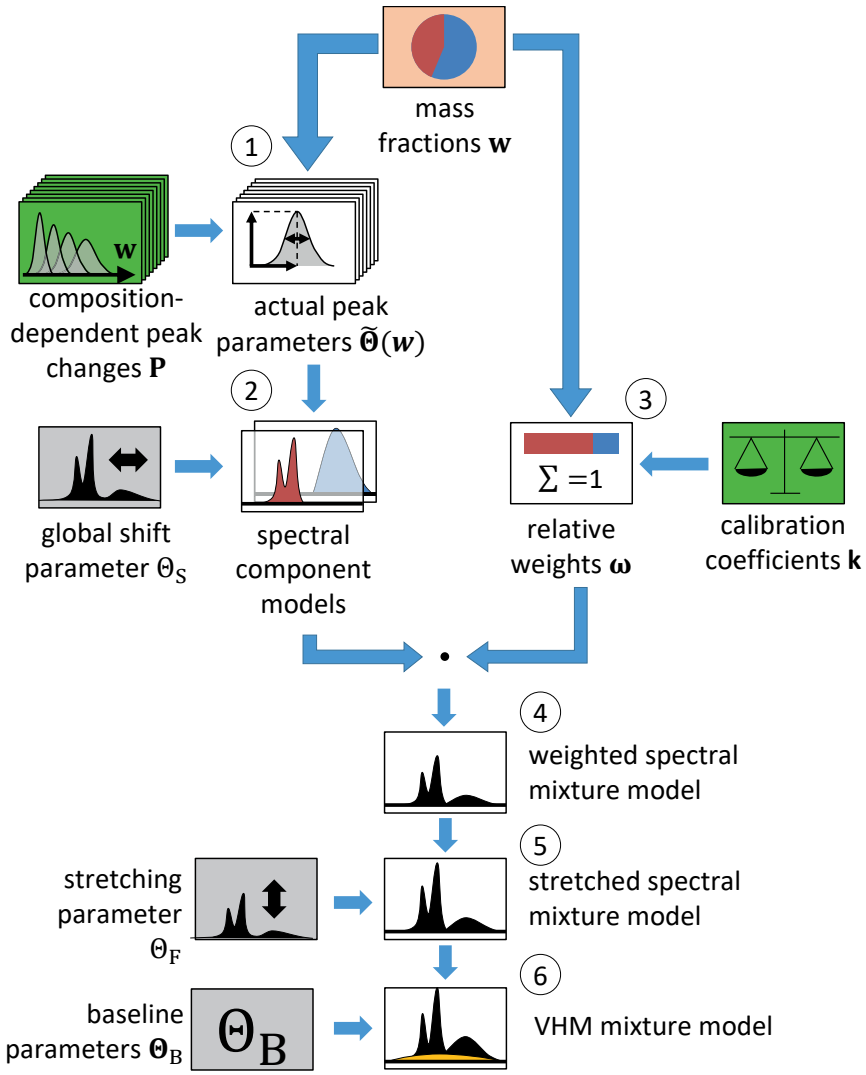

Figure 5.1: Scheme of the algorithm calculating a very hard model (VHM). Boxes with colored backgrounds are input parameters. Green background: Parameter fitted in calibration (Section 5.1.2). Grey background: Parameter fitted to each spectrum to account for random changes that are attributed to the optical setup. Red background: Mass fractions. The calculation steps are labeled in the circles.

$$\tilde{\Theta}_{i,p,q}\left(\mathbf{P},\mathbf{w}\right) = \Theta^{*}_{i,p,q} + \sum_{r=0}^{n_{\mathrm{poly}}} P_{i,p,q,r} \cdot w^{r}_{j\neq i}. \tag{5.2}$$

Here, polynomials for any peak of component i are expressed as function of the mass

Table 5.1: Sets of VHM peak parameters investigated in this work. We add the following subscripts to VHM: "P" if the all position parameters δ are parametrized, "W" if all peak width γ are parametrized, "A" if all peak areas α are parametrized. The index value (*e.g.*, P1) of the VHM indicates the order of the polynomial.

parametrized peak parameters	model name
position δ	VHM_{P1}
position δ & width γ	VHM_{PW1}
position δ & area α	VHM_{PA1}
position δ & area α & width γ	VHM_{PWA1}

fraction of the other component $j \neq i$. The r^{th} polynomial coefficient for changes of the q^{th} peak parameter of the p^{th} peak of the i^{th} component, $\tilde{\Theta}_{i,p,q}(\mathbf{w})$, is denoted by $P_{i,p,q,r}$. These changes are based on the mass fraction of component $j \neq i$. The degree of the polynomial is denoted by n_{poly}. Equation (5.2) can be extended to multi-component mixtures by adding an additional polynomial for each additional component. The flexibility of the VHM model can be tuned by the degree of the polynomial. This flexibility can be different for different types of peak parameters. *E.g.*, we do not always model all types of peak-parameters (*e.g.*, peak width γ) as concentration-dependent but leave them fixed at their pure component values $\tilde{\Theta}_{i,p,q}(w) = \Theta^*_{i,p,q}$, *i.e.*, $f(w) = 0$. Thereby, the number of parameters is reduced. Still, for simplicity, we parametrize one parameter type (e.g. the peak width γ) either for all peaks of all components or not at all. For the same reason, we use the same degree for all polynomials. Probably, individual parametrizations for all peaks would be beneficial. However, identifying these individual parametrizations would currently take a lot of manual effort, which would be restrictive in practice. As shown in Section 5.2.3, a heuristic rule is that the parametrization of peak changes may be finer (*i.e.*, with more parametrized peak parameters) if more calibration spectra are available. To describe the resulting parametrization of a VHM model, we introduce the nomenclature given in Table 5.1.

Step 2: Calculate spectral component models

The spectral component models in the mixture are determined similar to the pure component models (Equation (2.9)). However, the concentration-dependent parameters $\tilde{\Theta}_{i,p,q}(\mathbf{w})$ are used instead of the pure component parameters $\Theta^*_{i,p,q}$. The resulting

equation for the spectral component model i reads:

$$\mathcal{A}_i\left(\tilde{\Theta}_{i,p}\left(\mathbf{w}\right),\tilde{\nu}_m,\Theta_S\right)=\sum_{p=1}^{n_{\text{peaks},i}}V_{i,p}\left(\mathbf{w},\tilde{\nu}_m^{\Delta}\right), \tag{5.3}$$
$$\text{where } \tilde{\nu}_m^{\Delta}=\tilde{\nu}_m-\Theta_{\text{S}}.$$

Here, the total number of peaks of the pure component model for component i is again denoted by $n_{\text{peaks},i}$. Θ_S denotes the global shift parameter as introduced in Section 2.2.3.

Step 3: Calculate relative weights ω

In step 3, the relative weights $\boldsymbol{\omega}$ of the weighted superposition of the component models are determined as function of the composition in mass fractions $\mathbf{w}$. Absolute weights ζ cannot be determined from a composition in mass fractions, as the absolute intensity of the mixture depends also on factors such as laser intensity and excitation time (*cf.* Section 2.1). As closure condition, the sum of the relative weights is fixed to unity ($\sum\omega_i=1$). A quotient of the relative weights ω_i/ω_j equals a quotient of absolute weights ζ_i/ζ_j. Starting from Equation (2.17) and introducing the boundary condition for the relative weights ($\sum\omega_i=1$), the equation for the relative weights $\boldsymbol{\omega}$ reads:

$$\omega_i=\frac{w_i/\hat{k}_i}{\sum_{j=1}^{n_C}w_j/\hat{k}_j}. \tag{5.4}$$

So, n_C-1 independent relative weights are determined from n_C-1 independent mass fractions, where n_C denotes the number of components. Note that in Equation (5.4) the relative weights $\boldsymbol{\omega}$ are determined from the mass fractions $\mathbf{w}$, while in Equation (2.17) the mass fractions $\mathbf{w}$ are determined from the absolute weights $\boldsymbol{\zeta}$. $\hat{k}_i$ denotes the calibration coefficient for a composition given in mass fractions.

Step 4: Calculate the weighted spectral mixture model

The mixture model is obtained by adding the pure component models $\mathcal{A}_i\left(\tilde{\nu}_m^{\Delta},\boldsymbol{\Theta}_i\right)$ weighted by the relative weights as follows:

$$\sum_{i=1}^{n_\mathrm{C}} \left[\omega_i \cdot \mathcal{A}_i\left(\tilde{\nu}_m^{\Delta}, \boldsymbol{\Theta}_i\right)\right]. \tag{5.5}$$

Step 5: Stretch the spectral mixture model with stretching factor Θ_F

The mixture model is stretched by a factor Θ_F which accounts for the absolute intensity of the mixture spectrum due to, *e.g.*, variations of laser intensity or excitation time (see Section 2.1):

$$\Theta_\mathrm{F} \cdot \sum_{i=1}^{n_C} \left[\omega_i \cdot \mathcal{A}_i\left(\tilde{\nu}_m^{\Delta}, \boldsymbol{\Theta}_i\right)\right]. \tag{5.6}$$

Step 6: Add a baseline model

In the final step, a baseline model $\mathcal{B}(\tilde{\nu}_m, \boldsymbol{\Theta}_\mathrm{B})$ is added with a parameter set $\boldsymbol{\Theta}_\mathrm{B}$ (see Section 2.2.1):

$$\Theta_\mathrm{F} \cdot \sum_{i=1}^{n_C} \left[\omega_i \cdot \mathcal{A}_i\left(\tilde{\nu}_m^{\Delta}, \boldsymbol{\Theta}_i\right)\right] + \mathcal{B}\left(\tilde{\nu}_m^{\Delta}, \boldsymbol{\Theta}_\mathrm{B}\right). \tag{5.7}$$

The complete VHM model

Steps 1 to 6 lead to the following equation that describes the VHM model:

$$VHM\left(\tilde{\nu}_m, \mathbf{P}, \hat{\mathbf{k}}, \mathbf{w}, \boldsymbol{\Theta}_\mathrm{B}, \Theta_\mathrm{F}, \Theta_\mathrm{S}\right) = \Theta_\mathrm{F} \cdot \sum_{i=1}^{n_\mathrm{C}} \left[\omega_i \cdot \mathcal{A}_i\left(\tilde{\nu}_m^{\Delta}, \tilde{\boldsymbol{\Theta}}_i\right)\right] + \mathcal{B}\left(\tilde{\nu}_m^{\Delta}, \boldsymbol{\Theta}_\mathrm{B}\right), \tag{5.8}$$

$$\text{where } \omega_i = \frac{w_i/\hat{k}_i}{\sum_{j=1}^{n_\mathrm{C}} w_j/\hat{k}_j},$$

$$\tilde{\Theta}_{i,p,q} = \Theta^*_{i,p,q} + \sum_{r=0}^{n_\mathrm{poly}} P_{i,p,q,r} \cdot w^r_{j\neq i},$$

$$\tilde{\nu}_m^{\Delta} = \tilde{\nu}_m - \Theta_\mathrm{S}.$$

A calibrated VHM mixture model can predict the shape of a mixture spectrum for given mass fractions $\mathbf{w}$. Assumed inputs are the calibration parameters (see Figure 5.1). Additionally, random effects of the optical setup can be compensated by additional parameters (global shift parameter Θ_S, stretching parameter Θ_F, background parameters Θ_B). The effects of the optical setup are assumed to be unpredictable in this work. Thus, the corresponding parameters have to be fitted to each individual spectrum.

Calibration (Section 5.1.2) and prediction (Section 5.1.3) with a Very Hard Model rely on the same equations and differ by the choice of free and fixed parameters as explained in the following.

5.1.2 Calibration

In the calibration of a very hard model, both the polynomial coefficients $\mathbf{P}$ and the calibration coefficients $\mathbf{k}$ are determined. The mass fractions of the calibration spectra are known from gravimetrical preparation and are assumed to be the true mass fractions w_true. All calibration spectra are used at once, *i.e.*, one optimization is solved for calibration (Figure 5.2a).

Therefore, more calibration spectra usually lead to better results for $\mathbf{P}$ and $\mathbf{k}$, as more data is available to fit the same number of parameters. In the calibration spectra, backgrounds may be present and the global intensity may vary. Therefore, the background model $\mathcal{B}(\tilde{\nu}_m, \mathbf{\Theta}_\mathrm{B})$ and the stretching parameter Θ_F are fitted to each individual spectrum. The calibration spectra can usually be recorded in a short time period. In this case, changes of the optical alignment may be neglected, and the shift parameter may be fixed to zero ($\Theta_S = 0$), which reduces the number of fitted parameters within the calibration procedure. The calibration equation for a very hard model reads:

$$\begin{aligned}\mathbf{P}, \mathbf{k} = \underset{\mathbf{P},\mathbf{k},\mathbf{\Theta}_\mathrm{F},\mathbf{\Theta}_{\mathrm{B},1}\ldots\mathbf{\Theta}_{n_\mathrm{calib}}}{\arg\min} & \sum_{s=1}^{n_\mathrm{calib}} \sum_{m=1}^{n_{\tilde{\nu}}} \\ & \left\{S_s\left(\tilde{\nu}_m\right) - \mathit{VHM}\left(\tilde{\nu}_m, \mathbf{P}, \mathbf{k}, \mathbf{w}_{\mathrm{true},s}, \mathbf{\Theta}_{\mathrm{B},s}, \Theta_{\mathrm{F},s}, \Theta_{\mathrm{S},s} = 0\right)\right\}^2, \\ & \text{s.t. } k_1 = 1.\end{aligned} \tag{5.9}$$

Here, n_calib denotes the number of calibration spectra, $n_{\tilde{\nu}}$ the number of discrete wavenumbers $\tilde{\nu}_m$, $\mathbf{\Theta}_{\mathrm{B},s}$ the parameter set of the background model for calibration

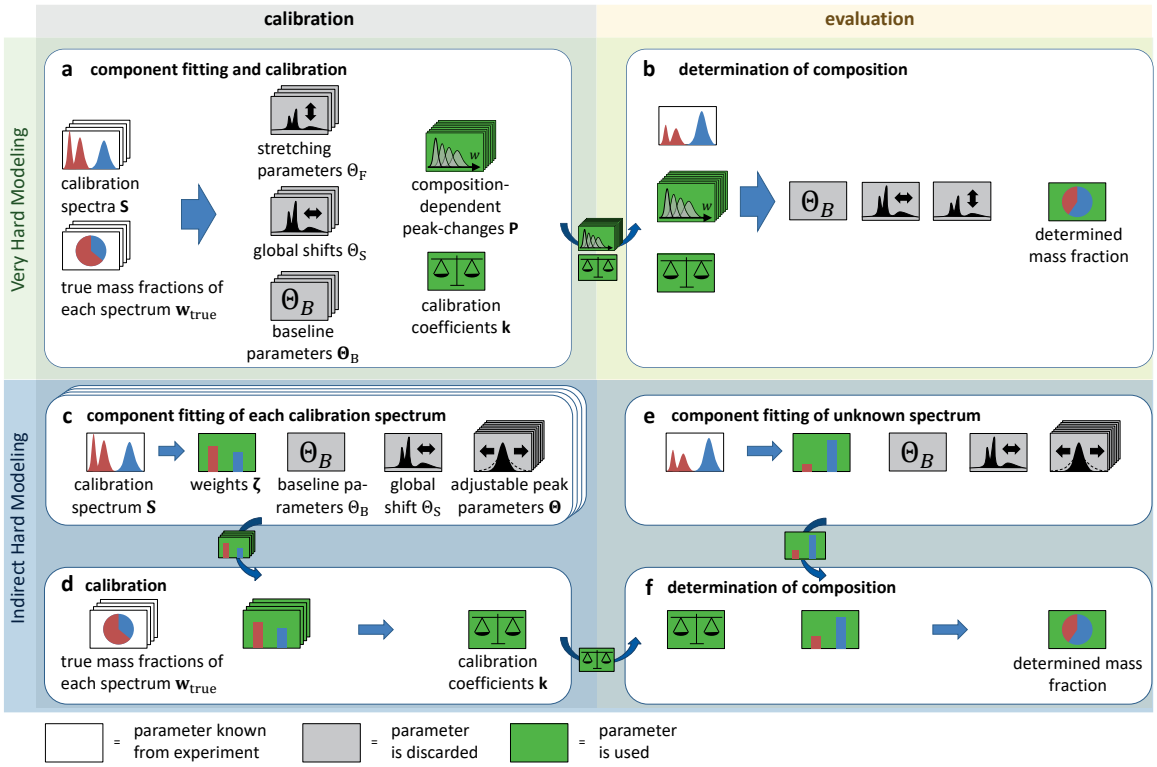

Figure 5.2: Calibration and spectral evaluation with VHM (a-b) and IHM (c-f). VHM: (a) calibration and component fitting are integrated and all calibration spectra are fitted at once; (b) determination of composition. IHM: (c) component fitting of each calibration spectrum to determine the weights of each component in each spectrum, *i.e.*, calibration spectra are fitted consecutively (shown as stacked boxes); (d) calibration; (e) component fitting of an unknown spectrum; (f) determination of composition.

spectrum s, $\Theta_{\mathrm{F},s}$ the stretching parameter for calibration spectrum s, and $\Theta_{\mathrm{S},s} = 0$ indicates that all global shift parameters are fixed to zero during calibration.

A fitted VHM model can predict the shape of a mixture spectrum based on the composition. In Figure 5.3, a calibrated VHM model is shown for the MeOH–H_2O system which is discussed in the experimental Section of this Chapter. The calibration spectra and the VHM models match very well for the composition of the calibration spectra. It is shown that the VHM model can then determine mixture spectra for any composition. The pronounced peak shifts of the methanol peaks at 2800-3000 cm^{-1} are represented very well by the VHM models.

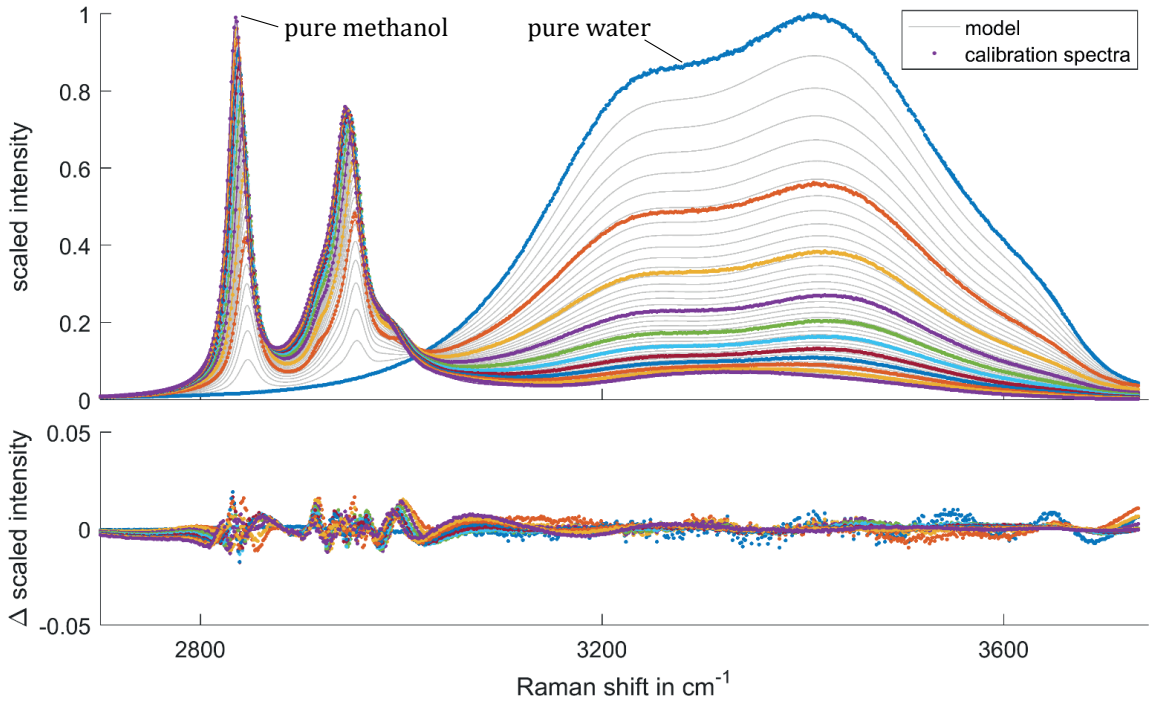

Figure 5.3: Calibration results of Very Hard Modeling (VHM_{PW1}) for MeOH–H_2O. Colored Dots: Calibration spectra with scaling and background correction based on the VHM fit; colored lines: Corresponding VHM fit; grey lines: Spectra predicted by VHM in steps of 0.25 wt-% methanol. Bottom: Difference between calibration spectra and VHM.

5.1.3 Evaluation of Mixture Spectra

In the spectral evaluation, the mass fractions $\mathbf{w}$ of unknown mixture spectra are predicted based on the polynomial coefficients $\mathbf{P}$ and the calibration coefficients $\mathbf{k}$ known from calibration. Additionally, parameters for random influences of the optical setup are applied as before: One stretching factor Θ_F, one global shift group Θ_S, and a set of background parameters Θ_B (see Figure 5.2b). In contrast to the calibration procedure, the spectra are evaluated successively. The equation for the spectral evaluation reads:

$$\mathbf{w} = \underset{\mathbf{w},\Theta_F,\boldsymbol{\Theta}_B,\Theta_S}{\arg\min} \sum_{m=1}^{n_{\tilde{\nu}}} \left\{ S\left(\tilde{\nu}_m\right) - VHM\left(\tilde{\nu}_m, \mathbf{P}, \mathbf{k}, \mathbf{w}, \boldsymbol{\Theta}_B, \Theta_F, \Theta_S\right) \right\}^2, \tag{5.10}$$

$$\text{s.t.} \sum w_j = 1.$$

$S(\tilde{\nu}_m)$ denotes the measured spectrum at the discrete wavenumber $\tilde{\nu}_m$. Note that in Equation (5.10) only the desired mass fractions $\mathbf{w}$ and parameters that account for random effects of the optical setup are fitted.

5.1.4 Comparison to Indirect Hard Modeling

In this Section, we compare Very Hard Modeling (VHM) to Indirect Hard Modeling (IHM).

IHM and VHM use the same framework. Both methods are based on component models, which model pure component spectra by a sum of pseudo-Voigt profiles. However, VHM parametrizes peak changes as function of composition to account for nonlinear effects while IHM uses selected adjustable parameters to give the model the flexibility to follow nonlinear effects. In IHM, the adjustable parameters are fitted to each evaluated spectrum. The component fitting procedure of IHM is the same for calibration (Figure 5.2c) and evaluation (Figure 5.2e). In contrast to VHM, IHM fits all calibration spectra consecutively (Figure 5.2c). VHM uses all calibration spectra in one optimization (stacked calibration spectra in Figure 5.2a). The component fitting of IHM results in absolute weights ζ that are either used for calibration (Figure 5.2d) or for determination of the composition of an unknown mixture spectrum (Figure 5.2f).

The major difference of VHM compared to IHM is that more information from the calibration is passed to the evaluation (*cf.* boxes a&b and d&f in Figure 5.2). VHM models the composition-dependent peak changes, stores them in the matrix $\mathbf{P}$, and uses this matrix as input to determine the composition.

As IHM fits the calibration spectra consecutively, less parameters have to be fitted simultaneously compared to VHM (Figure 5.2a&c). Therefore, IHM calibration is roughly 10 to 100 times faster than VHM calibration. Still, the total number of parameters fitted to the calibration spectra by IHM is larger, as adjustable parameters are fitted to each calibration spectrum. The VHM calibrations performed in this work took about 5 to 30 minutes on a standard consumer PC. For spectral evaluation, the calculation effort is reversed, VHM is faster by a factor of approximately 10 since the number of fitted parameters is reduced compared to IHM (Figure 5.2b&e). As a consequence, once calibrated, VHM models save time, which is beneficial for, *e.g.*, online applications.

5.2 Experimental

5.2.1 Sample Preparation

To validate and evaluate the VHM method, we prepared three independent physical sample sets (labeled A, B, and C) of mixtures of methanol and water. The sample sets were prepared independently on different days. Set A and B range from pure methanol to pure water in steps of 0.1 mol/mol (11 spectra). Set C comprises both pure components and samples from 0.05 to 0.95 mol/mol in steps of 0.1 mol/mol (12 spectra in total). We prepared all samples by volume but determined the exact compositions gravimetrically. The uncertainty of the composition of the samples is below 0.015 wt-%.

5.2.2 Spectral Acquisition

Several spectra were recorded from each sample of the three physical sample sets A, B, and C to study reproducibility and the influence of noise. We defined 50,000 counts as target intensity of the largest peak of each spectrum. This is approximately half of the maximum intensity that can be measured with the confocal Raman spectrometer used (Renishaw inVia confocal Raman microscope). The exposure time was adjusted for each sample so that the maximum intensity of the spectra is in the range of 40,000 to 60,000 counts.

With these settings, we recorded 10 spectra of each sample. From these spectra, we calculated the averages to produce a set of very "clean" spectra with little noise for calibration purposes. These averaged spectra data sets are denoted $\bar{\alpha}$, $\bar{\beta}$, and $\bar{\gamma}$ for the sample sets A, B, and C, respectively. To study the influence of noise, we prepared spectra with lower excitation times in 3 steps: (i) We used the spectra of sample set A individually. This spectra data set is denoted α. (ii) We recorded 10 spectra for each sample of set A with 1/10th of the exposure time mentioned above (spectra data set $\alpha_{1/10}$). (iii) We recorded 10 spectra for each sample of set A with 1/100th of the exposure time or a minimal exposure time of 0.01 s (spectra data set $\alpha_{1/100}$). As VHM uses extensive parametrization, overfitting might be an issue. Therefore, we defined subsets of the calibration data set $\bar{\alpha}$ to study the effect of the number of calibration spectra: Spectra data sets $\bar{\alpha}_3$, $\bar{\alpha}_4$ $\bar{\alpha}_5$, and $\bar{\alpha}_6$. The indices indicate the number of spectra in each set. Compositions of all spectra data sets are given in Table 5.2 in terms of mole fractions.

Figure 5.4 provides a detailed view of the spectra data set $\bar{\alpha}$. In the range of 3000

Table 5.2: Compositions of the spectra data sets in mole fractions. The spectra data sets α, $\alpha_{1/10}$, and $\alpha_{1/100}$ have the same compositions as set $\bar{\alpha}$ and are thus not shown.

	0.00	0.05	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60	0.65	0.70	0.75	0.80	0.85	0.90	0.95	1.00
$\bar{\alpha}$	x		x		x		x		x		x		x		x		x		x		x
$\bar{\beta}$	x		x		x		x		x		x		x		x		x		x		x
$\bar{\gamma}$	x	x		x		x		x		x		x		x		x		x		x	x
$\bar{\alpha}_3$	x										x										x
$\bar{\alpha}_4$	x						x								x						x
$\bar{\alpha}_5$			x				x				x				x				x		
$\bar{\alpha}_6$	x				x				x				x				x				x

to 3600 cm^{-1}, broad OH peaks of both components are present. In the range of 2800 to 3000 cm^{-1}, sharp CH peaks of methanol are visible. A peak shift of both CH peaks is clearly visible (compare spectra $w_{H_2O} = 0.0$ and $w_{H_2O} = 0.943$). The spectra are quite challenging for the spectral evaluation, as water has only broad peaks which are additionally completely overlapped by peaks of methanol. For low water contents, it becomes challenging to distinguish the water signal from the background and from the methanol signal (see also Figure 5.5). In Figure 5.5 (left), the spectral sets $\bar{\alpha}$, $\bar{\beta}$, and $\bar{\gamma}$ are shown (normalized). Some of the spectra of set $\bar{\gamma}$ have slightly elevated backgrounds that are visible to the eye in the range of 2400 to 2800 cm^{-1}. These spectra are used to challenge the spectral evaluation methods. The spectra data sets to study noise (sets α, $\alpha_{1/10}$, and $\alpha_{1/100}$) are shown in Figure 5.5 (right). The increasing noise level with decreasing exposure time is clearly visible.

5.2.3 Calibration and Spectral Evaluation

For calibration, we used the spectra data sets $\bar{\alpha}$, $\bar{\beta}$, and $\bar{\gamma}$ as well as the spectra data sets $\bar{\alpha}_3$, $\bar{\alpha}_4$, $\bar{\alpha}_5$, and $\bar{\alpha}_6$. With each spectra data set, we calibrated a VHM_{P1}, a VHM_{PW1}, a VHM_{PA1}, and a VHM_{PWA1} model (*cf.* Section 5.1.1). Furthermore, we calibrated IHM models for comparison. To improve comparability, all VHM and IHM models are based on the same component models. For the spectral evaluation, we used the spectra data sets $\bar{\alpha}$, $\bar{\beta}$, and $\bar{\gamma}$ as well as the spectra data sets α, $\alpha_{1/10}$, and $\alpha_{1/100}$. Each of these sets is evaluated with all calibrations.

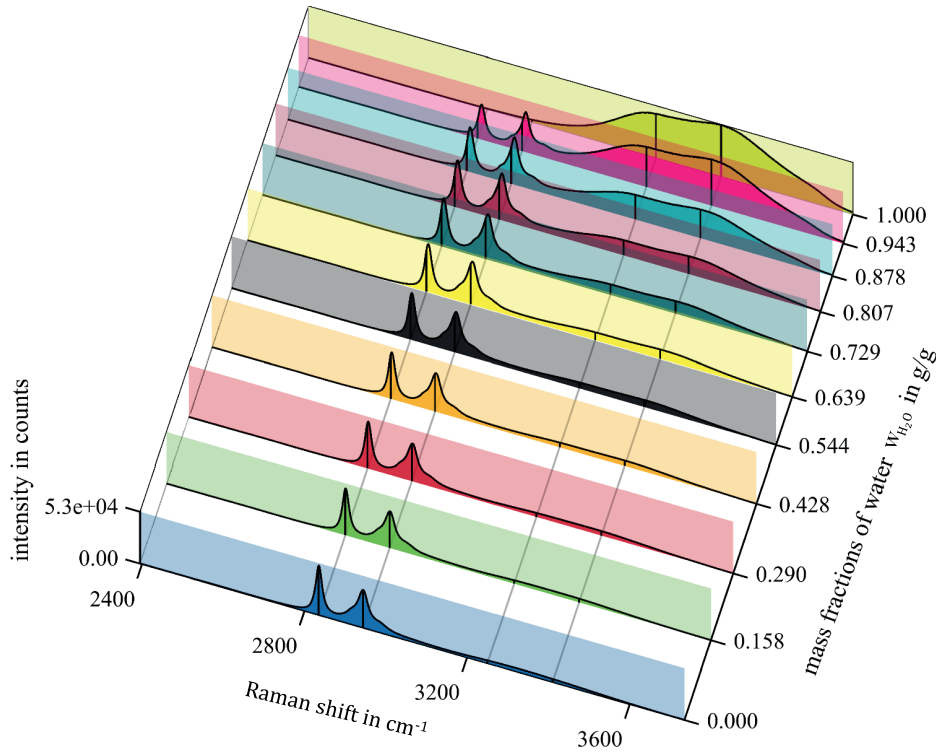

Figure 5.4: Raman spectra of mixtures of methanol and water (set $\bar{\alpha}$). Some peaks of the pure components are marked to guide the eye. The peak shift of the sharp MeOH CH-peaks in the range of 2800-3000 cm^{-1} is clearly visible.

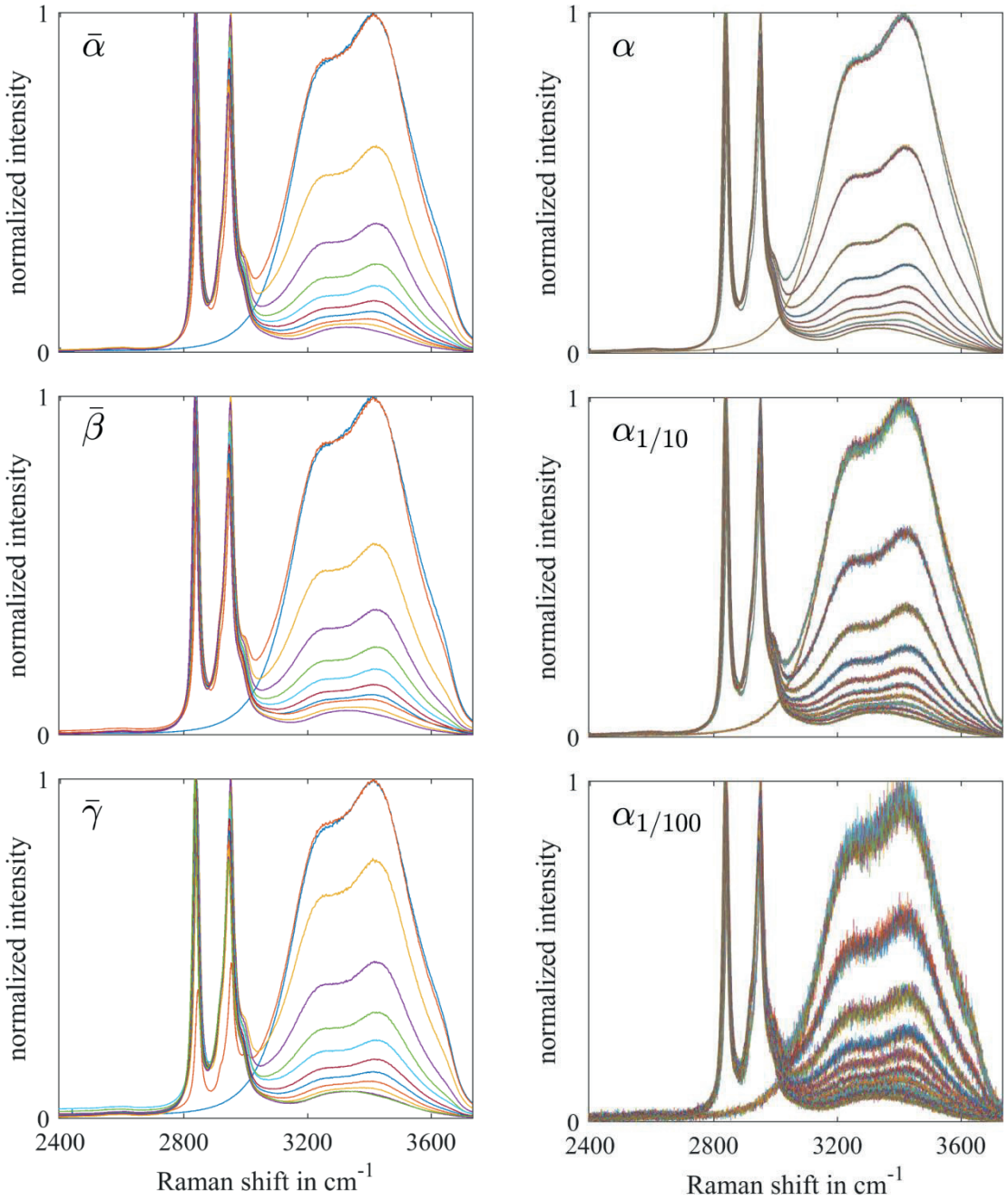

Figure 5.5: Left: Raman spectra data sets $\bar{\alpha}$, $\bar{\beta}$, and $\bar{\gamma}$ (mixtures of methanol and water). Each spectrum is scaled individually to a maximum intensity of one. For spectra data set $\bar{\gamma}$, a background is visible to the eye in the range of 2400-2800 cm^{-1}. Right: Raman spectra data sets α, $\alpha_{1/10}$, and $\alpha_{1/100}$ (mixtures of methanol and water). Each spectrum is scaled individually to a maximum intensity of one.

5.3 Results and Discussion

The results of calibration and evaluation are assessed by the root-mean-square error *RMSE* (*cf.* Equation (2.19)).

The results are presented in the following steps: (i) We perform a test-set validation with the "clean" sets (low noise) $\bar{\alpha}$, $\bar{\beta}$, and $\bar{\gamma}$ (Section 5.3.1). Each of these sets is used for calibration and each of these sets is evaluated with all calibrations. (ii) We study the influence of noise. The clean spectra data sets are used for calibration. The increasingly noisy spectra data sets α, $\alpha_{1/10}$, and $\alpha_{1/100}$ are evaluated (Section 5.3.2). (iii) We study the influence of the number of calibration spectra. Calibration is based on the spectra data sets $\bar{\alpha}_3$, $\bar{\alpha}_4$, $\bar{\alpha}_5$, $\bar{\alpha}_6$, and $\bar{\alpha}$ and evaluation is assessed for the spectra data sets $\bar{\alpha}$, $\bar{\beta}$, $\bar{\gamma}$, α, $\alpha_{1/10}$, and $\alpha_{1/100}$ (Section 5.3.3).

In Section 5.3.4, we use the VHM framework to predict the composition of the binary mixture using only peaks of methanol. Thereby, we demonstrate the additional information content of the parametrized nonlinear effect.

5.3.1 Test-set Validation (Spectra Sets $\bar{\alpha}$, $\bar{\beta}$, and $\bar{\gamma}$)

Figure 5.6 shows a test-set validation of the low noise spectra data sets $\bar{\alpha}$, $\bar{\beta}$, and $\bar{\gamma}$. Each set has been used for calibration and the resulting model has been used to analyze all spectra data sets. VHM is compared to IHM. The deviations range from RMSE 0.21 wt-% to 1.09 wt-%. As expected, we see for IHM that the RMSE is always lowest if the same data set is used for prediction and calibration. Still, the three sets perform differently with IHM: Set $\bar{\beta}$ is always predicted best, set $\bar{\alpha}$ is always second, and set $\bar{\gamma}$ always has the largest RMSEs. The larger errors RMSE of set $\bar{\gamma}$ are most probably caused by the elevated backgrounds that are highlighted in Section 5.2.2. For the same reason, calibrations with set $\bar{\gamma}$ lead to the worst errors RMSE if the calibration is used to predict one of the other sets.

For the VHM model, we see that set $\bar{\beta}$ is predicted best as for IHM, but the difference between the errors RMSE of the predicted sets differ much less than for IHM. Sets $\bar{\alpha}$ and $\bar{\gamma}$ are predicted almost equally well with VHM. We see the largest difference between IHM and VHM for the prediction of set $\bar{\gamma}$. Compared to IHM, VHM reduces the RMSE by 50% to 68%. Our interpretation is that the backgrounds in set $\bar{\gamma}$ are a real challenge for IHM. IHM fails to separate the broad OH peaks of water accurately from the background and the methanol peaks in the same range. The separation is challenging for low water contents in particular, as the water peaks are very small in this concentration range (*cf.* Figure 5.5). VHM probably cannot separate the OH

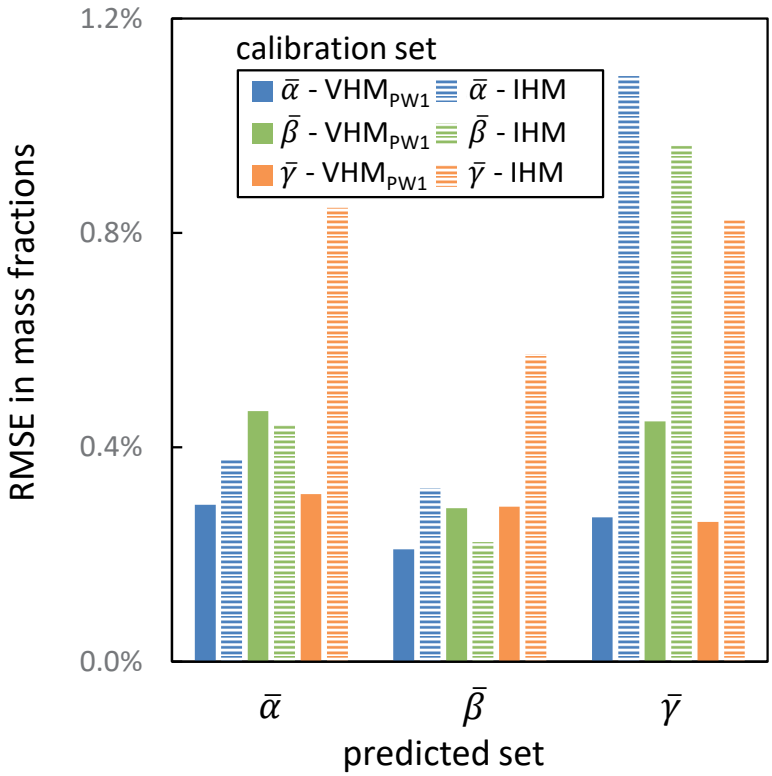

Figure 5.6: *RMSE* (Equation (2.19)) in mass fractions for test-set validation of sets $\bar{\alpha}$, $\bar{\beta}$, and $\bar{\gamma}$ with IHM and VHM$_{PW1}$.

peaks better, but additional information is available: The peak shifts of the CH peaks of methanol are parametrized in the VHM model and therefore improve the accuracy. This feature is particularly helpful for low water contents where CH peaks are large.

For the prediction of set $\bar{\beta}$ with VHM, the calibration with set $\bar{\alpha}$ performs better than calibration with set $\bar{\beta}$ itself, which is rather counter-intuitive. Still, as the results are quite close, we believe that this might be an artefact from the numerical optimization.

5.3.2 Influence of Noise (Spectra Sets α, $\alpha_{1/10}$, and $\alpha_{1/100}$)

To assess the robustness of IHM and VHM with regards to noise, we studied the spectra data sets α, $\alpha_{1/10}$, $\alpha_{1/100}$ after calibration with sets $\bar{\alpha}$, $\bar{\beta}$, $\bar{\gamma}$.

For IHM, we see that the calibrations with sets $\bar{\alpha}$ and $\bar{\beta}$ perform similarly. For these sets, the spectra data set with the largest noise (set $\alpha_{1/100}$) shows an increased

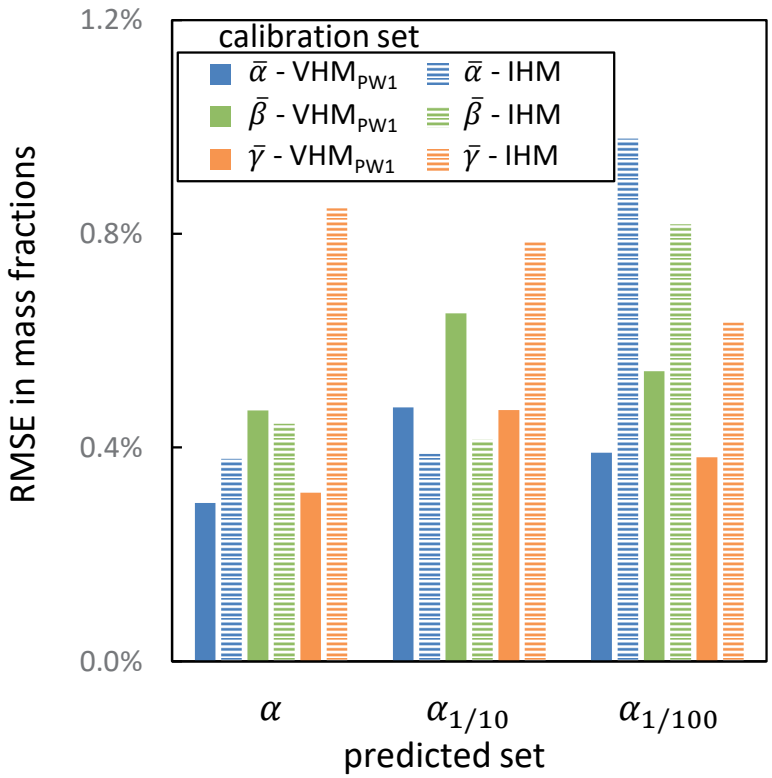

Figure 5.7: RMSE (Equation (2.19)) in mass fractions for varied noise in the spectral evaluation with VHM and IHM. From set α to set $\alpha_{1/10}$ and from $\alpha_{1/10}$ to $\alpha_{1/100}$, the signal intensity is roughly decreased to $1/10^{\text{th}}$.

RMSE. For calibration with set $\bar{\gamma}$, we see the opposite behavior: The larger the noise, the lower the *RMSE*. Probably, the effect of noise and the effect of the elevated backgrounds accidentally cancel each other out to a certain extent.

For VHM the results are more constant overall. For the noisiest spectra, VHM leads to significantly lower RMSE than IHM: The RMSE of VHM are lower by 34% to 60%.

5.3.3 Number of Calibration Sets vs. Flexibility of the Very Hard Model

The high number of parameters in the VHM model bares the risk of overfitting in the calibration step. To study when overfitting affects the calibration results, we calibrated VHM models with different parametrizations that have varying parameter

counts: VHM_{P1}, VHM_{PW1}, VHM_{PA1}, and VHM_{PWA1} (see Table 5.1). Furthermore, we built separate calibrations with spectra data sets that have different numbers of calibration spectra: Set $\bar{\alpha}$ provides 11 calibration spectra and the sets $\bar{\alpha}_3$ to $\bar{\alpha}_6$ provide 3 to 6 calibration spectra (index: Number of calibration spectra; detailed composition: *cf.* Table 5.2). With these calibrations, we predicted the sets $\bar{\alpha}$, $\bar{\beta}$, $\bar{\gamma}$, α, $\alpha_{1/10}$, and $\alpha_{1/100}$. As benchmark, we performed the same calibrations and validations with IHM.

The results are summarized in Figure 5.8. For each method (IHM and all VHM methods), the mean error RMSE (Equation (2.19)) is given for each calibration set. Furthermore, the best and the worst RMSE is indicated with an error bar. For IHM, we see that the number of calibration spectra has almost no influence on the RMSE. Set $\bar{\alpha}_3$ contains only the two pure component spectra and one mixture spectrum. Still, this set performs practically identical as the sets with more calibration spectra. This constant performance demonstrates one of the key features of IHM: The need for very few calibration spectra and high extrapolation ability.

For VHM, the calibration results are less uniform. Still, we see a clear trend for VHM in line with intuitive expectations: (i) more calibration spectra generally improve the RMSE; (ii) models with more parameters need more calibration spectra (see VHM_{PWA1} in Figure 5.8); (iii) the model with the highest parametrization (VHM_{PWA1}) gives the best results if sufficient calibration spectra are available.

Compared to IHM, the VHM_{P1} model (only position is parametrized) leads to worse mean errors RMSE values, but the worst RMSE of VHM is always lower than the worst RMSE of IHM. The methods VHM_{PW1} and VHM_{PA1} lead to very similar results: High RMSE values for only 3 calibration spectra (RMSEs of 2 wt-% and over 4 wt-%). For more than 3 calibration spectra the results of VHM are better than IHM by factors of 1.7 to 2.9 for the same combinations of calibration set and predicted set. The method VHM_{PWA1} needs a minimum of 6 calibration spectra to produce good results for the studied system. If less calibration spectra are used, the RMSE values are in the range of 1.4 to 4.3 wt-%. For more calibration spectra, VHM_{PWA1} produces the best RMSE we found in this Chapter with 0.2 wt-%.

Another feature of VHM is that the distribution of RMSE is always very narrow. Although the RMSE for set $\bar{\alpha}_3$ is 4.3 wt-% on average for the model VHM_{PWA1}, the largest and the lowest RMSE differ by only 0.15 wt-%. So, all spectra data sets are predicted equally poorly. The narrow distribution of RMSEs can even prove as an advantage in practice: If some test spectra with compositions different from the compositions of the calibration spectra are available, an overfitted model will almost certainly predict these calibration spectra very poorly. Therefore, a bad spectral model can easily be recognized and discarded. If the test spectra are predicted well

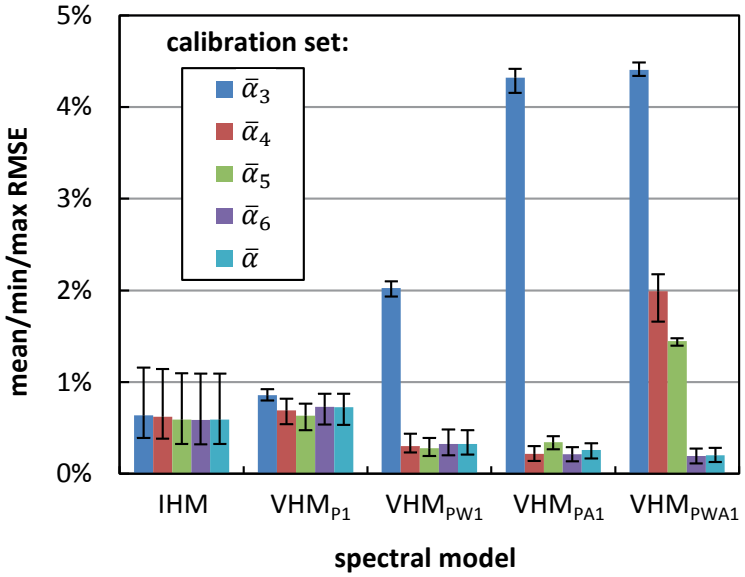

Figure 5.8: Results for IHM as well as VHM with different sets of parametrized peak parameters for the calibration sets $\bar{\alpha}_3$-$\bar{\alpha}$. For each method, the mean value of the RMSEs of the prediction of sets $\bar{\alpha}$, $\bar{\beta}$, $\bar{\gamma}$, α, $\alpha_{1/10}$, and $\alpha_{1/100}$ is plotted as filled bars. The error bars indicate the minimum and maximum value of these RMSE values.

on the other hand, the VHM model is very likely to perform similarly well for new mixture spectra.

5.3.4 Spectral Evaluation Based on Peak Shape Only

In this Section, we show that it is possible in principle to predict the composition of the binary mixture of methanol and water from the peak shapes and positions of the methanol peaks alone to emphasize that the nonlinear effects contain valuable information for the spectral evaluation.

For calibration and validation, we narrow the investigated spectral range to the range of the CH-Peaks (2700-3000 cm^{-1}, see Figure 5.4). In this spectral range, there are no water peaks. Therefore, we only fit methanol peaks in the calibration of this model. As no water peaks are fitted, no information is available on the relative weights of MeOH and water. For this reason, we cannot fit any calibration coefficients k (*cf.* Equation (5.8)). Instead, the relative weight of MeOH is fixed to unity and the relative weight of water is fixed to zero. As a result, the relative position of the two methanol peaks and the shape of these peaks are the only usable parameters for the VHM model

Table 5.3: RMSE (Equation (2.19)) of the prediction of sets $\bar{\alpha}$, $\bar{\beta}$, and $\bar{\gamma}$ each without the spectrum of pure water after calibration of a $\mathrm{VHM_{PW3}}$ model with the same sets (vertical) in wt-%.

		validation sets		
		$\bar{\alpha}\backslash H2O$	$\bar{\beta}\backslash H2O$	$\bar{\gamma}\backslash H2O$
clb. set	$\bar{\alpha}\backslash H2O$	0.27%	0.81%	0.87%
	$\bar{\beta}\backslash H2O$	0.89%	0.13%	1.28%
	$\bar{\gamma}\backslash H2O$	0.89%	0.94%	0.49%

to predict the composition of the mixture.

For calibration, we use sets $\bar{\alpha}$, $\bar{\beta}$, and $\bar{\gamma}$ without the pure water spectrum containing no methanol peaks. As the peak shape and position is the only usable information, we use higher polynomials to account for the peak shapes more accurately (VHM also works with the linear polynomials used before with errors RMSE from 1.2%-2.6%, details not shown). For our system, a polynomial degree of 3 provided the best results. The test-set validation results for the method $\mathrm{VHM_{PW3}}$ are presented in Table 5.3. The results indicate overfitting as the results on the main diagonal are always by far the best results. Still, all other results range from 0.87 to 1.28 wt-%, which might be sufficient for many applications. The results could be further improved by very accurate calibration of the instrument with respect to wavelength such that the global shift parameter Θ_S can be set to zero in Equation (5.9). Thereby, the absolute position of both peaks could be parametrized. Currently, only the relative position is used effectively in Equation (5.9) due to the free global shift parameter Θ_S. Still, within the scope of this work, we did not optimize our measurements to predict the composition based on methanol peaks alone. The results given in Table 5.3 already demonstrate the value of parametrized nonlinear effects for the spectral evaluation.

5.4 Conclusion

Very Hard Modeling (VHM) is presented as a novel spectral evaluation method. VHM is based on the framework of Indirect Hard Modeling (IHM). Indirect Hard Modeling uses sums of peak-shaped functions (component models) instead of pure component spectra. During the fitting with IHM, a subset of the peak parameters is made adjustable to account for nonlinear effects such as peak shifts.

VHM uses the same component models but parametrizes the peak shapes based on

the composition of the mixture. This parametrization leads to models that are more robust as they have less free parameters in the spectral evaluation. The flexibility of VHM models can be tuned by different parametrizations. We show that larger parametrizations may be used if more calibration spectra are available. By adapting the parametrization to the amount of calibration spectra VHM achieves more accurate prediction results. In contrast, IHM always determines the same number of calibration factors from the calibration spectra – regardless of the number of calibration spectra. Therefore, more calibration spectra than a few are of limited value for IHM.

As all peak parameters are parametrized in the calibration step, VHM has a lower count of free parameters in the spectral evaluation. Therefore, the time for the prediction of a spectrum is reduced by a factor of approximately 10 compared to IHM. On the other hand, the more extensive parametrization of VHM leads to longer calibration times by a factor of 10 to 100, compared to IHM. Still, calibrations are usually only performed once and a calibration takes no more than 30 minutes on a standard PC. VHM even allows to predict the composition of a binary mixture solely based on the peak shape of one component. Such a prediction is not possible with IHM.

The larger robustness of VHM models leads to improved robustness in the face of noise and backgrounds. For the noisiest spectra of the studied system of methanol and water, VHM reduces the root-mean square error RMSE by 34% to 60% compared to IHM. For a spectra set with slightly elevated backgrounds, VHM reduces the RMSE by 50% to 68% compared to IHM. Although no general claims may be drawn from a single case study, we see the promising results as a validation of the Very Hard Modeling method.

Chapter 6

Physically Sound Peak Profiles

In this Chapter we introduce a new peak profile called sum norm. Compared to the common area norm peaks (see Section 2.2), the sum norm reflects the physics of a CCD detector more accurately. The name *sum* norm is chosen, as the modeled intensities of the spectrum (*i.e.*, a vector of discrete intensities) have a constant sum even if the peak is deformed or shifted. For a peak in the area norm, the integral of the underlying continuous function – the area – is constant. We show in this Chapter that the sum of the modeled intensities of an area norm peak is not constant under certain circumstances. A constant intensity of each peak is one of the main assumptions of IHM.

In Section 2.1, the Raman intensity was introduced as power per solid angle with the unit (W/sr). The optics collect the scattered light in a certain solid angle and the spectrometer separates the light by wavelength so that each pixel of the detector receives a certain light intensity (in W/m^2), which is then effectively integrated over time and space by counting the photons. This piecewise integration leads to the measured spectrum S.

The commonly used peak profile in the area norm presented in Section 2.2.2 (Equations (2.6)-(2.8)) models the continuous light intensity that is received by the CCD detector. This continuous function is evaluated for the wavenumber of each pixel center to model a spectrum. This peak model is physically inaccurate: If the intensity profile on the detector is a continuous function which is integrated piecewise over space, a realistic peak model should also include this piecewise integration.

This inaccuracy can lead to errors in the spectral analysis as illustrated in Figure 6.1. For broad peaks, the modeled spectrum (red dots) in the area norm reflects the shape of the continuous light intensity (blue) quite accurately. For very narrow peaks, however, the area norm leads to erratic results (Figure 6.1 top right). Depending on the peak position, the sum of modeled intensities changes from zero to a large value.

For the sum norm, the sum of intensities is always constant, regardless of the peak position Figure 6.1 bottom right). The varying sum of modeled intensities of the area norm is a problem, as IHM is based on the assumption that each peak retains the same intensity, even if the peak is deformed or shifted (Section 2.2.2).

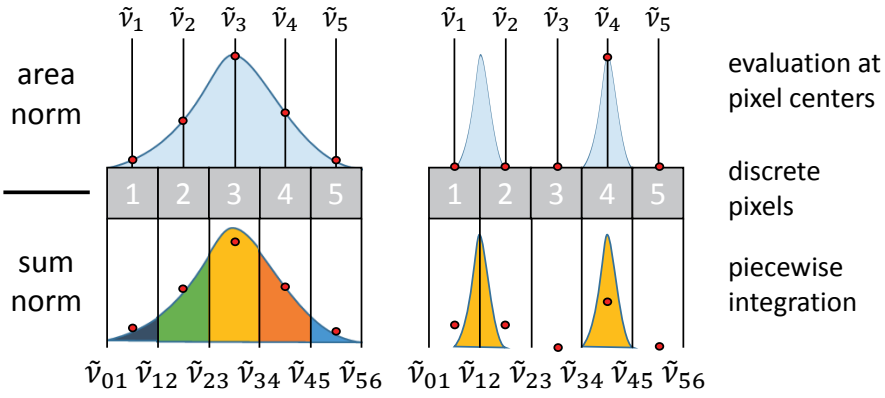

Figure 6.1: Comparison of area norm profiles and sum norm profiles for a large peak and a narrow peak (sketch). The red dots denote the intensities, modeled with the respective norm. The blue curve denotes the continuous light-intensity that is modeled by the area norm. The area norm profiles are evaluated at the pixel centers ($\tilde{\nu}_1 \dots \tilde{\nu}_5$). In the area norm, the integral of this continuous function (light blue) is constant. The sum norm peaks are integrated piecewise between the edges of the pixels ($\tilde{\nu}_{01} \dots \tilde{\nu}_{56}$). The piecewise integrals are shown as multiple colors. In the sum norm, the sum of the discrete intensities (red dots) is constant.

6.1 The Sum Norm Peak Profile

To account for the spatial integration of the signal on the chip, we introduce the sum norm profile that integrates the continuous light-intensity over a certain range of wavenumbers. The piecewise integration over wavenumbers corresponds to an integration over the spatial dimension of the pixels. The sum norm is based on pseudo-Voigt functions, just as the area norm presented in Section 2.2.2. The exact dimensions of a pixel (expressed in wavenumbers) is unknown. For the piecewise integration the edges of each pixel are needed. From the calibration of the optical instrumentation, the wavenumber of each pixel center is known. As wavenumbers of a pixel edge, we

simply assume the average of the wavenumbers of the two adjacent pixel centers. For the wavenumbers of the edges of the first and the last pixel, the wavenumbers are slightly extrapolated.

The Gaussian profile in the sum norm reads:

$$G_{\mathrm{sum}}\left(\tilde{\nu}_m\right) = \left[\frac{1}{2}\,\mathrm{erf}\left\{\sqrt{\ln(2)}\cdot\left(\frac{\tilde{\nu}-\delta}{\gamma/2}\right)\right\}\right]_{\frac{\tilde{\nu}_{m-1}+\tilde{\nu}_m}{2}}^{\frac{\tilde{\nu}_m+\tilde{\nu}_{m+1}}{2}}. \tag{6.1}$$

The Lorentzian profile in the sum norm reads:

$$L_{\mathrm{sum}}\left(\tilde{\nu}_m\right) = \left[\frac{1}{\pi}\arctan\left\{2\cdot\left(\frac{\tilde{\nu}-\delta}{\gamma/2}\right)\right\}\right]_{\frac{\tilde{\nu}_{m-1}+\tilde{\nu}_m}{2}}^{\frac{\tilde{\nu}_m+\tilde{\nu}_{m+1}}{2}}. \tag{6.2}$$

The modeled intensities of Gaussian and Lorentzian profiles in the sum norm both have sum of unity, as long as there are no parts of the peak outside the evaluated range of wavenumbers; hence the name *sum norm*. The constant sum is easily verifiable as the piecewise integrals of all pixels sum up to an integral over the entire range and as $1/2\cdot\int_{-\infty}^{+\infty}\mathrm{erf}\left(x\right)\mathrm{d}x = 1$ and $1/\pi\cdot\int_{-\infty}^{+\infty}\arctan\left(x\right)\mathrm{d}x = 1$. Similar to Equation (2.8), the pseudo-Voigt profile in sum norm reads:

$$V_{\mathrm{sum}}\left(\tilde{\nu}_m\right) = \alpha_{\mathrm{sum}}\left[\beta\cdot G_{\mathrm{sum}}\left(\tilde{\nu}_m\right) + (1-\beta)\cdot L_{\mathrm{sum}}\left(\tilde{\nu}_m\right)\right]. \tag{6.3}$$

Here, α_{sum} denotes the sum of the sum norm profile $\sum_{\tilde{\nu}_m} V(\tilde{\nu}_m)$. The constant sum of the sum norm profile is shown in Figure 6.1 (bottom). In contrast to the are norm peak, the sum norm peak retains a constant sum of intensities even if the peak is shifted. An approximate parameter conversion between the area norm and the sum norm is given in Appendix A.4.

For the first derivative of a sum norm profile (to be used in FD-IHM) we use the same numerical derivative as we use for the spectral data (*cf.* Section 4.1). The sum norm aims to model the intensity of each pixel on a physically sound basis. Therefore, it is logical to treat the modeled spectrum in the same way as we treat the measured data to calculate the derivative. In fact, the analytical derivatives of the area norm used in FD-IHM (Section 4.1) may have unfavorable properties for very narrow peaks as we demonstrate in Figure 6.2. The narrow peaks may vanish completely or the sum of its first derivative may not equal zero. Note that the sum of the first derivative of a peak in a real spectrum will always equal zero if a difference quotient is used for the differentiation.

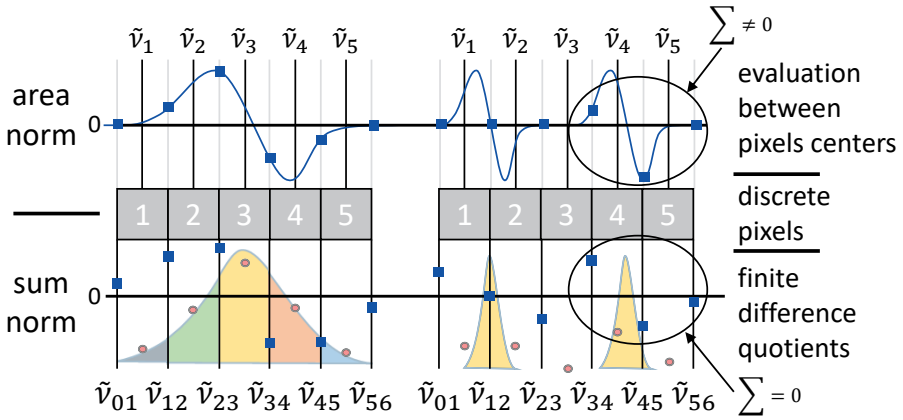

Figure 6.2: Comparison of area norm profiles and sum norm profiles for a large peak and a narrow peak (sketch) for FD-IHM. The blue squares mark the calculated intensities of the first derivative for both norms. The area norm evaluates the analytical derivative of a pseudo-Voigt profile (blue line) between pixels. The derivative of the sum norm is the finite difference quotient of the non-derivative intensities (red dots) of neighboring pixels.

6.2 Comparison of the Sum Norm and the Area Norm

In Figure 6.1, a comparison of the sum norm and the area norm is shown. The area norm parameter α (Equation (2.8)) parametrizes the area of the peaks in the upper part of the figure. The sum norm parameter α_{sum} parametrizes the sum of the intensities which are marked as red dots in the lower part of the figure.

The value of the increased physical soundness of the sum norm becomes obvious in the following example: Many instruments allow for hardware-binning of various pixels to reduce the readout noise of the detector for low intensity spectra. If hardware binning is applied, the charges (*i.e.*, the intensities) of several pixels are combined on the CCD detector and the signal equals the sum of the combined signals. If the five pixels in Figure 6.1 are binned to one big pixel, the area norm peak is only evaluated once at central position of the binned peak (here: The position $\tilde{\nu}_3$) and thus assigns a *lower* summed intensity to the binned peak than the sum of the five intensities. The sum norm assigns the same intensity to the peak as before, as the binned peak is integrated from $\tilde{\nu}_{01}$ to $\tilde{\nu}_{56}$, which is correct.

Note that the computational costs of a peak in sum norm are comparable to a peak

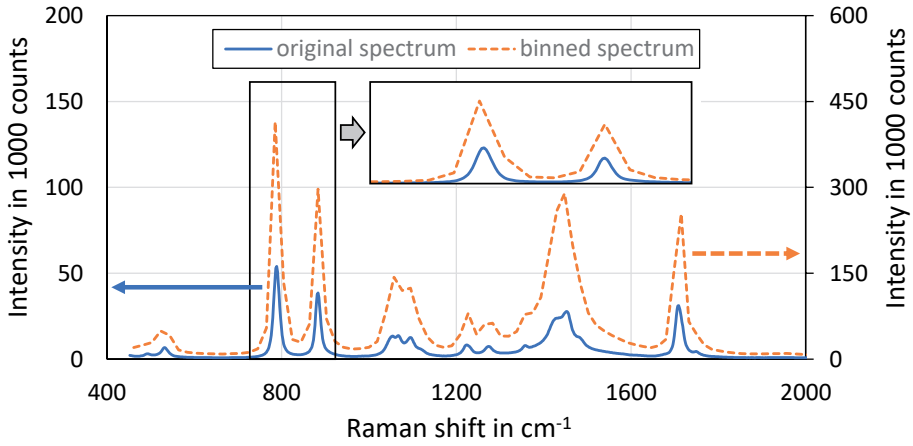

Figure 6.3: Comparison of a binned spectrum and the original spectrum for one spectrum of spectra data set *a*.

in area norm. Equations (6.1) and (6.3) have to be evaluated $n_{\tilde{\nu}} + 1$ times for $n_{\tilde{\nu}}$ pixels, which is only one more function evaluation compared to a peak in area norm (*e.g.*, 1601 vs. 1600 function evaluations). The computational overhead of the sum norm is thus limited to $n_{\tilde{\nu}}$ differences.

6.3 Experimental

To test the applicability of the sum norm, we apply the sum norm to a large data set. For this purpose, we use all EtOH–acetone spectra data sets that we used in the FD-IHM Chapter (Figure 4.4 on page 46). We apply the exact same procedure as in Section 4.2 with the sole difference that we use the sum norm instead of the area norm. As these spectra sets have a high resolution, the area norm and the sum norm should lead to similar results.

To investigate the effect of binning, we use spectra data set *a* which is shown in Figure 4.4a (page 46). Set *a* is one of the sets without added backgrounds. From this spectra set, we derive a spectra set of binned spectra numerically. Always 11 pixels are binned. The original spectra have 1015 discrete pixels each; the binned spectra have 92 pixels each. One sample spectrum is shown in Figure 6.3 together with the binned spectrum. The binned spectrum lacks some details but the main peaks are present. We use the original spectra set to calibrate mixture models in the area norm and in the sum norm for IHM and FD-IHM. These models are then used to predict the binned spectra sets.

6.4 Results and Discussion

The comparison of the area norm and the sum norm for FD-IHM is shown in Figure 6.4 for the original (not binned) spectra. The results of both norms are almost identical. The average values of the prediction errors *RMSEP* are 0.55% and 0.56%, for the sum norm and the area norm respectively. The results show that the sum norm is just as applicable as the conventional area norm for highly resolved spectra. We did not encounter any new challenges.

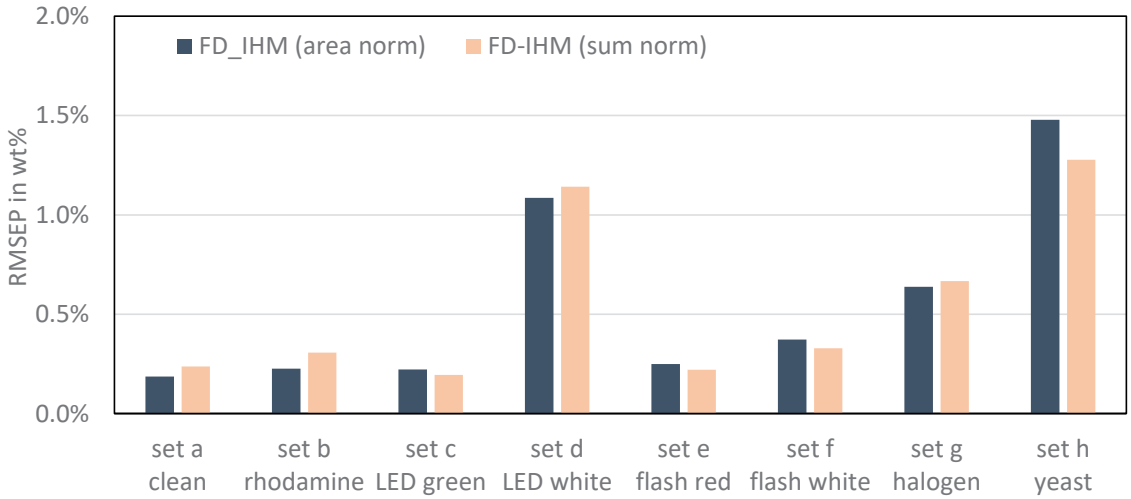

Figure 6.4: Comparison of the area norm and the sum norm for the spectral evaluation with FD-IHM. The evaluated spectra sets are discussed in Section 4.2.

The comparison of the performance of the area norm and the sum norm for the binned data sets is given in Tables 6.1 and 6.2. We see that sum norm and area norm calibrate the non-derivative spectra set equally well, for IHM and FD-IHM alike. The *RMSEC* of the calibrations is almost exactly the same for the area norm and the sum norm. If these calibration models are applied to the binned test set, however, the sum norm leads to *RMSE* values that are more than three times lower than the same values for the area norm. We interpret these results as a consequence of the better physical soundness of the sum norm.

Table 6.1: Comparison of area and sum norm using IHM. RMSEC for the calibration with a highly resolved spectra set and RMSEP for the prediction of a binned low resolution spectra set.

	RMSEC calibration set	RMSEP test set (binned)
area norm	0.48%	2.67%
sum norm	0.47%	0.75%

Table 6.2: Comparison of area and sum norm using FD-IHM. RMSEC for the calibration with a highly resolved spectra set and RMSEP for the prediction of a binned low resolution spectra set.

	RMSEC calibration set	RMSEP test set (binned)
area norm	0.13%	5.39%
sum norm	0.14%	1.72%

6.5 Conclusions

We present a new peak model that reflects the physics of the CCD detector of a common Raman setup better than the previously available peak model. The previously available peak model (area norm) simply evaluates pseudo-Voigt functions at the center of each pixel. The new peak model (sum norm) reflects that the intensity of a spectrum is a continuous function of the wavelength that is integrated within a certain range of wavenumbers by each pixel. Thus, the sum norm peak profile increases physical soundness.

We show on a theoretical level that the sum norm peaks are better applicable to very narrow peaks than the area norm peaks. For peaks in the area norm, a constant sum of the modeled intensities is not guaranteed if the peaks deform or shift, which violates one of the basic assumptions of IHM. However, the more pixels a peak spans, the more the advantages of the sum norm vanish compared to the area norm and they lead to very similar results.

Our results for the system EtOH–acetone confirm that the sum norm peak profiles perform very similar compared to the area norm profiles for highly resolved spectra. The prediction of the composition of eight spectra sets using FD-IHM with both peak norms lead to almost identical results.

To test the new peak model with low resolution spectra, we artificially decreased the resolution of the spectra of an EtOH–acetone spectra set by binning. After calibration with the high resolution spectra sets, the low resolution spectra sets are predicted. For IHM as well as for FD-IHM, the models that apply the sum norm lead, in the case of low resolution spectra, to predictions with three times lower RMSE errors. We regard these results as a proof of the increased physical soundness of the sum norm, compared to the area norm.

In our view, the acquired results suggest the use of the sum norm peak profile for all spectral evaluations with IHM or FD-IHM. The increased physical soundness may help if binning is applied, if the detector of the setup is replaced by a detector with a different resolution, or if there are very narrow peaks in the spectrum. In other cases, the results of the area norm and the sum norm are almost equivalent. More research concerning possible benefits of the sum norm for spectra with very narrow peaks seems to be worthwhile.

Chapter 7

Summary and Outlook

7.1 Summary

In many scientific and industrial applications, a precise measurement of the composition of liquid and gaseous mixtures is important. Precise compositions can be used to study, *e.g.*, diffusion, phase equilibria, and chemical reactions. For the chemical industry, the composition is related to, *e.g.*, quality control, process monitoring, and process control.

Optical spectroscopic techniques allow for the fast simultaneous in-situ quantification of the concentration of all components in multi-component systems. Optical spectroscopic measurements can be performed typically in a few seconds, while alternative techniques such as chromatographic techniques usually need several minutes for the same task. Therefore, optical spectroscopy is interesting for online applications in particular.

Raman spectroscopy is a very versatile technique. In contrast to IR-spectroscopy, visible laser light can be used and the optics can therefore be made from glass instead of salt crystals. Furthermore, the Raman signal of water is relatively weak. Therefore, Raman spectroscopy is suitable also for aqueous systems as the signals of the other components are not dominated by the water signal.

In recent years, Indirect Hard Modeling (IHM) has become a popular method to evaluate Raman spectra. Unlike statistical methods, IHM explicitly accounts for nonlinear effects such as peak shifts and peak deformations. Nonlinear effects are common in Raman spectra and caused by molecular interactions in mixtures that differ from the interactions in a pure substance.

In the present work, we present new methods that extent the applicability of Indirect Hard Modeling (IHM).

Model-free calibration of mixture spectra: As IHM can handle large nonlinear effects, IHM is particularly suited for the evaluation of systems with strong interactions and even for reactive systems. However, IHM needs calibration samples with known composition. For chemical systems with intermediates, such samples cannot be prepared gravimetrically, as intermediates typically are intrinsically unstable. Therefore, IHM calibration relies on thermodynamic models for such systems, which are not always available.

In the present work, a method to calibrate reactive mixtures with intermediates is presented that does not rely on such a thermodynamic model. Instead, only the chemical reaction mechanism, the material balances, and electroneutrality are used. The calibration coefficients are determined by minimizing the combined residuals of the material balances and of the electroneutrality. We test the method for the reactive system MEA–H_2O–CO_2 that forms the intermediates $MEAH^+$ and $MEAHCOO^-$. For comparison, we use a thermodynamic model from the literature to predict the same concentrations. The results from the model-free calibration and the thermodynamic model are in very good agreement. Our results can be seen as a validation for both, the model-free calibration as well as the thermodynamic model.

First-Derivative Indirect Hard Modeling: As Raman signals are very weak, background signals as, *e.g.*, fluorescent signals are a common challenge. Therefore, IHM needs careful background treatment. Background models or methods for spectral pretreatment are usually set up manually, which requires expert knowledge and time, and is still potentially erroneous. If backgrounds in an application differ substantially from the backgrounds in the calibration spectra, large errors may occur. Statistical methods have used the first derivative of the spectra for a long time to filter backgrounds.

We introduce First-Derivative Indirect Hard Modeling (FD-IHM) to remove the need for any further background treatment. FD-IHM combines the straightforward but effective background treatment of the first derivative with the capacity of IHM to handle nonlinear effects. We validate the method with 8 sets of spectra that contain multiple backgrounds that are very different with regards to shape and intensity. As reference, we use calibrations with IHM combined with several different ways of background treatment. Our results show that FD-IHM performs significantly better than IHM for the studied spectra sets. Furthermore, FD-IHM requires less manual effort.

Very Hard Modeling: One of the key features of IHM is the calibration effort. In an IHM calibration, only $n_C - 1$ calibration factors are determined (n_C denotes the number of components). Therefore, a multi-component system can principally be

calibrated with a single calibration spectrum. Nonetheless, more calibration spectra lead to a higher confidence in the model performance. However, once a few calibration spectra are available, an IHM calibration cannot be significantly improved any further by adding more spectra.

We propose the Very Hard Modeling method (VHM) to use the calibration spectra more thoroughly. VHM is based on the IHM framework. In contrast to IHM, VHM exploits the nonlinear effects to improve the spectral evaluation. VHM parametrizes the nonlinear effects as function of the composition of the calibration spectra. Consequently, the information of the calibration spectra is used more thoroughly. The VHM models show a better robustness than IHM when confronted with noisy spectra. Furthermore, VHM performs significantly better than IHM if sufficient calibration spectra are provided.

Physically sound peak profiles: A pixel on a CCD detector has a finite size that translates to a finite range of wavenumbers. The continuous intensity of a Raman spectrum is integrated in this range. However, this integration is not represented correctly by the current peak models. Instead of an integration, the current peak models are evaluated once at the pixel center. For peaks with a width in the order of the width of a pixel, the current peak models lead to unfavorable effects: If a peak is shifted, the apparent sum of intensities of the peak model may change. The reason is that the maximum of the peak model can be more or less aligned with a pixel center. Changing peak intensities violate the basic assumption of IHM that peak intensities are constant.

Therefore, we introduce a new peak profile that integrates a continuous mathematical intensity profile for the range of each pixel. The new peak profile is called "sum norm" as it guarantees a constant sum of modeled intensities in a spectrum. The sum norm can replace the commonly used "area norm" peak profile for IHM, FD-IHM, and VHM. The sum norm is particularly useful for low resolution spectra. We validate the sum norm peak profile with the same spectra set as used for FD-IHM. The area norm and the sum norm lead to very similar results for these spectra. Therefore, we assume that for highly resolved spectra, both peak norms can be used equally well. To show the advantage of the sum norm, we calibrated IHM and FD-IHM models with both norms to highly resolved spectra but evaluated binned spectra with a 11-fold lower resolution. For the binned spectra, the sum norm peak profile leads root-mean-square errors that are more than three times lower; for IHM and FD-IHM alike. We see the reduction in errors as a benefit of the increased physical soundness of the sum norm peak profile. The sum norm might prove useful for, *e.g.*, low-cost setups with limited resolution as well as for setups with excellent optics and thus very sharp peaks.

In summary, we introduced methods that extent the applicability of IHM to reactive systems with intermediates, to systems with large, unpredictable backgrounds, to systems with significant noise levels, and to low resolution spectra.

7.2 Perspectives for Future Work

In this Section, we propose ideas for further studies that might broaden the applicability of IHM even further.

Improvement of IHM

IHM uses a global shift parameter to account for shifts of the whole spectrum due to misalignment of the optical setup. So far, this global shift parameter shifts the whole modeled spectrum for the same amount of wavenumbers. An optical misalignment usually shifts the spectrum spatially on the detector. The spatial coordinate on the detector is approximately proportional to wavelengths. As the relationship between wavenumbers and wavelengths is not linear, the global shift parameter should shift the spectrum in wavenumbers rather than wavelengths. This corrected shift could prove worthwhile for setups where large alignment errors cannot be avoided.

IHM is based on symmetrical peak profiles. Asymmetrical peaks of the spectrum are commonly modeled with a sum of two or more peaks. It could prove useful to introduce asymmetrical peak profiles as, *e.g.*, proposed by Stancik and Brauns (2008), into IHM. Thereby, the total number of peaks could be reduced, which results in reduced calculation times. Furthermore, more versatile peaks would probably reduce the effort in the setup of a pure component model.

In this work, we parametrized the VHM models based on the mass fraction. However, peak shifts are caused by changes of the intramolecular vibrations. These changes can probably be related to macroscopic thermodynamic variables as, *e.g.*, the excess enthalpy, to a certain extent. Therefore, it would be worthwhile to investigate if VHM models can be parametrized even better when such variables are used.

New Applications

The introduced VHM method parametrizes peak changes based on the composition of the mixture. Nevertheless, the peak parameters could also be parametrized as a function of other influences as, *e.g.*, temperature. Temperature often has an influence

on peaks, *e.g.*, when OH-groups are present. If the peak shifts and deformations can be parametrized correctly, it is principally possible to predict composition and temperature from the same spectrum.

The proposed FD-IHM method treats backgrounds very effectively. FD-IHM uses the difference quotient to calculate the derivative of a normal spectrum. A very promising application would be the shifted excitation Raman difference spectroscopy (SERDS, Wei et al. (2015)). SERDS acquires Raman spectra with two slightly different excitation wavelengths. The Raman peaks have the same Raman shift for both excitation wavelengths (*cf.* Section 2.1). Therefore, the absolute wavelength of the peaks in the Raman spectrum differ just as the excitation wavelength does. The wavelengths of fluorescent background signals or any source of background lighting, however, do not change. If both spectra are subtracted, these background signals are suppressed very effectively. With this method, even sharp peaks that overlap with the Raman spectra can be suppressed in principle. As two slightly shifted spectra are subtracted, the result is basically the first derivative of the Raman spectrum. Therefore, a slightly adapted FD-IHM method should be able to evaluate such spectra very well.

We are keen to see where future developments of spectral evaluation techniques for Raman spectroscopy will lead.

Bibliography

Alsmeyer, F., Koß, H.-J., and Marquardt, W. (2004). Indirect Spectral Hard Modeling for the Analysis of Reactive and Interacting Mixtures. *Appl. Spectrosc.*, 58(8):975–985.

Alsmeyer, F. and Marquardt, W. (2004). Automatic Generation of Peak-Shaped Models. *Appl. Spectrosc.*, 58:986–94.

Amundsen, T. G., Øi, L. E., and Eimer, D. A. (2009). Density and Viscosity of Monoethanolamine + Water + Carbon Dioxide from (25 to 80)°C. *J. Chem. Eng. Data*, 54(11):3096–3100.

Aronu, U., Gondal, S., Hessen, E., Haug-Warberg, T., Hartono, A., Hoff, K., and Svendsen, H. (2011). Solubility of CO2 in 15, 30, 45 and 60 mass% MEA from 40 to 120 °C and model Representation using the Extended UNIQUAC Framework. *Chem. Eng. Sci.*, 66:6393–6406.

Baek, S.-J., Park, A., Kim, J., Shen, A., and Hu, J. (2009). A Simple Background Elimination Method for Raman Spectra. *Chemom. Intell. Lab. Syst.*, 98(1):24 – 30.

Baesch, S., Siebel, D., Schmidt-Hansberg, B., Eichholz, C., Gerst, M., Scharfer, P., and Schabel, W. (2016). Comparison of Surfactant Distributions in Pressure-Sensitive Adhesive Films Dried from Dispersion under Lab-Scale and Industrial Drying Conditions. *ACS Appl. Mater. Interfaces*, 8(12):8118–8128.

Bardow, A., Göke, V., Koß, H.-J., Lucas, K., and Marquardt, W. (2005). Concentration-Dependent Diffusion Coefficients from a Single Experiment using Model-Based Raman Spectroscopy. *Fluid Phase Equilib.*, 228:357–366.

Bardow, A., Göke, V., Koß, H.-J., and Marquardt, W. (2006). Ternary Diffusivities by Model-Based Analysis of Raman Spectroscopy Measurements. *AIChE J.*, 52:4004–4015.

Bardow, A., Marquardt, W., Göke, V., Koß, H.-J., and Lucas, K. (2003). Model-Based Measurement of Diffusion using Raman Spectroscopy. *AIChE J.*, 49(2):323–334.

Blesinger, C., Beumers, P., Buttler, F., Pauls, C., and Bardow, A. (2014). Temperature-Dependent Diffusion Coefficients from 1D Raman Spectroscopy. *J. Solution Chem.*, 43(1):144–157.

Brown, C., Vega-Montoto, L., and Wentzell, P. (2000). Derivative Preprocessing and Optimal Corrections for Baseline Drift in Multivariate Calibration. *Appl. Spectrosc.*, 54(7):1055–1068.

Böttinger, W., Maiwald, M., and Hasse, H. (2007). Online NMR Spectroscopic Study of Species Distribution in MEA-H2O-CO2 and DEA-H2O-CO2. *Fluid Phase Equilib.*, 263:131–143.

Esche, E., Kraemer, B., Müller, D., Meyer, K., Zientek, N., Maiwald, M., and Wozny, G. (2014). Improved Desorption Control via Raman Spectroscopy. In *20th International Conference Process Engineering and Chemical Plant Design*.

Faber, N. and Rajkó, R. (2007). How to avoid Over-Fitting in Multivariate Calibration–The Conventional Validation Approach and an Alternative. *Anal. Chim. Acta*, 595(1):98 – 106.

Ferraro, J. R., Nakamoto, K., Brown, C. W., and ISBN: (2003). *Introductory Raman Spectroscopy*. Elesevier, 2nd edition.

Flecher, P. E., Welch, W. T., Albin, S., and Cooper, J. B. (1997). Determination of Octane Numbers and Reid Vapor Pressure in Commercial Gasoline using Dispersive Fiber-Optic Raman Spectroscopy. *Spectrochim. Acta, Part A*, 53(2):199–206.

Glass, M., Aigner, M., Viell, J., Jupke, A., and Mitsos, A. (2016). Liquid-Liquid Equilibrium of 2-methyltetrahydrofuran/Water over Wide Temperature Range: Measurements and Rigorous Regression. *Fluid Phase Equilib.*, 433:212–225.

Hall Jr, H. (1957). Correlation of the Base Strengths of Amines. *J. Am. Chem. Soc.*, 79(20):5441–5444.

Hanke, M. and Scherzer, O. (2001). Inverse Problems Light: Numerical Differentiation. *Am. Math. Monthly*, 108(6):512–521.

Hardy, S. and Liauw, M. A. (2013). Mixing Behaviour Investigation of a Switchable Solvent Synthesis using ATR-IR Spectroscopy. *Chem. Eng. J.*, 233:292–296.

Heinemann, M., Meinberg, H., Büchs, J., Koß, H.-J., and Ansorge-Schumacher, M. (2005). Method for Quantitative Determination of Spatial Polymer Distribution in Alginate Beads using Raman Spectroscopy. *Appl. Spectrosc.*, 59(3):280–285.

Helmdach, L., Feth, M. P., Minnich, C., and Ulrich, J. (2013). Application of ATR-MIR Spectroscopy in the Pilot Plant-Scope and Limitations using the Example of Paracetamol Crystallizations. *Chem. Eng. Process.*, 70:184 – 197.

Hilliard, M. D. (2008). *A Predictive Thermodynamic Model for an Aqueous Blend of Potassium Carbonate, Piperazine, and Monoethanolamine for Carbon Dioxide Capture from Flue Gas*. PhD thesis, Graduate School of the University of Texas at Austin.

Holoborodko, P. (2008). Smooth Noise Robust Differentiators. online. http://www.holoborodko.com/pavel/numerical-methods/numerical-derivative/smooth-low-noise-differentiators/.

Jakobsen, J. P., Krane, J., and Svendsen, H. F. (2005). Liquid-Phase Composition Determination in CO2-H2O-Alkanolamine Systems: An NMR Study. *Ind. Eng. Chem. Res.*, 44(26):9894–9903.

Kriesten, E., Alsmeyer, F., Bardow, A., and Marquardt, W. (2008a). Fully Automated Indirect Hard Modeling of Mixture Spectra. *Chemom. Intell. Lab. Syst.*, 91(2):181 – 193.

Kriesten, E., Mayer, D., Alsmeyer, F., Minnich, C., Greiner, L., and Marquardt, W. (2008b). Identification of Unknown Pure Component Spectra by Indirect Hard Modeling. *Chemom. Intell. Lab. Syst.*, 93:108–119.

Lieber, C., A. and Mahadevan-Jansen, A. (2003). Automated Method for Subtraction of Fluorescence from Biological Raman Spectra. *Appl. Spectrosc.*, 57:1363–1367.

Liebergesell, B., Flake, C., Brands, T., Koß, H.-J., and Bardow, A. (2017). A milliliter-scale setup for the efficient characterization of isothermal vapor-liquid equilibria using Raman spectroscopy. *Fluid Phase Equilib.*, 446:36–45.

Luther, S. K., Schuster, J. J., Leipertz, A., and Braeuer, A. (2013). Non-Invasive Quantification of Phase Equilibria of Ternary Mixtures Composed of Carbon Dioxide, Organic Solvent and Water. *J. Supercrit. Fluids*, 84:146–154.

Luther, S. K., Stehle, S., Weihs, K., Will, S., and Braeuer, A. (2015). Determination of Vapor-Liquid Equilibrium Data in Microfluidic Segmented Flows at Elevated Pressures using Raman Spectroscopy. *Anal. Chem.*, 87(16):8165–8172.

Mariani, N., de Almeida Teixeira, G., de Lima, K., Morgenstern, T. B., Nardini, V., and Júnior, L. (2015). NIRS and iSPA-PLS for Predicting Total Anthocyanin Content in Jaboticaba Fruit. *Food Chem.*, 174:643 – 648.

Martens, H. (1992). *Multivariate calibration*. John Wiley & Sons.

Meyer, K., Kraemer, B., Esche, E., Müller, D., Zientek, N., Paul, A., Wozny, G., and Maiwald, M. (2015). Combined On-line NMR and Raman Spectroscopic Study and Chemometric Model of Species Distribution in MEA/H2O/CO2. *in preparation*.

Meyer-Kirschner, J., Kather, M., Pich, A., Engel, D., Marquardt, W., Viell, J., and Mitsos, A. (2016). In-Line Monitoring of Monomer and Polymer Content During Microgel Synthesis using Precipitation Polymerization via Raman Spectroscopy and Indirect Hard Modeling. *Appl. Spectrosc.*, 70(3):416–426.

Michalik-Onichimowska, A., Kern, S., Riedel, J., Panne, U., King, R., and Maiwald, M. (2017). "Click" Analytics for "click" Chemistry - A Simple Method for Calibration-Free Evaluation of Online NMR Spectra. *J. Magn. Reson.*, 277:154 – 161.

Nestor, J. R. and Lippincott, E. R. (1973). The Effect of the Internal Field on Raman Scattering Cross Sections. *J. Raman Spectrosc.*, 1(3):305–318.

Olivero, J. J. and Longbothum, R. (1977). Empirical Fits to the Voigt Line Width: A Brief Review. *J. Quant. Spectrosc. Radiat. Transfer*, 17(2):233–236.

Ozaki, Y., Sasic, S., and Jiang, J. (2001). How Can we Unravel Complicated Near Infrared Spectra? - Recent Progress in Spectral Analysis Methods for Resolution Enhancement and Band Assignments in the Near Infrared Region. *J. Near Infrared Spectrosc.*, 9(2):63–95.

Pelletier, M. (2003). Quantitative Analysis Using Raman Spectrometry. *Appl. Spectrosc.*, 57(1):20A–42A.

Peters, C., Wolff, L., Haase, S., Thien, J., Brands, T., Koß, H.-J., and Bardow, A. (2017). Multicomponent Diffusion Coefficients from Microfluidics using Raman Microspectroscopy. *Lab. Chip*, 17(16):2768–2776.

Rahman, A., Kondo, N., Ogawa, Y., Suzuki, T., and Kanamori, K. (2016). Determination of K Value for Fish Flesh with Ultraviolet-Visible Spectroscopy and Interval Partial Least Squares (iPLS) Regression Method. *Biosyst. Eng.*, 141:12 – 18.

Raman, C. V. and Krishnan, K. S. (1928). A New Type of Secondary Radiation. *Nature*, 1928(121):501–502.

Richner, G. and Puxty, G. (2012). Assessing the Chemical Speciation during CO2 Absorption by Aqueous Amines Using in Situ FTIR. *Ind. Eng. Chem. Res.*, 51(44):14317–14324.

Rinnan, Å., van den Berg, F., and Engelsen, S. B. (2009). Review of the Most Common Pre-Processing Techniques for Near-Infrared Spectra. *Trends Anal. Chem.*, 28(10):1201–1222.

Samarakoon, P. G. L., Andersen, N. H., Perinu, C., and Jens, K.-J. (2013). Equilibria of MEA, DEA and AMP with Bicarbonate and Carbamate: A Raman Study. *Energy Procedia*, 13:2002–2010.

Savitzky, A. and Golay, M. (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.*, 36(8):1627–1639.

Scharfer, P., Schabel, W., and Kind, M. (2007). Mass Transport Measurements in Membranes by Means of in situ Raman Spectroscopy–First Results of Methanol and Water Profiles in Fuel Cell Membranes. *J. Membr. Sci.*, 303(1-2):37–42.

Schuster, J., Will, S., Leipertz, A., and Braeuer, A. (2014). Deconvolution of Raman Spectra for the Quantification of Ternary High-Pressure Phase Equilibria Composed of Carbon Dioxide, Water and Organic Solvent. *J. Raman Spectrosc.*, 45(3):246–252.

Smekal, A. (1923). Zur Quantentheorie der Dispersion. *Naturwissenschaften*, 11(43):873–875.

Souchon, V., de Oliveira Aleixo, M., Delpoux, O., Sagnard, C., Mougin, P., Wender, A., and Raynal, L. (2011). In situ Determination of Species Distribution in Alkanolamine-H2O-CO2 Systems by Raman Spectroscopy. *Energy Procedia*, 4:554–561.

Spegazzini, N., Siesler, H. W., and Ozaki, Y. (2012). Sequential Identification of Model Parameters by Derivative Double Two-Dimensional Correlation Spectroscopy and Calibration-Free Approach for Chemical Reaction Systems. *Anal. Chem.*, 84(19):8330–8339.

Stancik, A. L. and Brauns, E. B. (2008). A simple Asymmetric Lineshape for Fitting Infrared Absorption Spectra. *Vib. Spectrosc.*, 47(1):66–69.

Thien, J., Peters, C., Brands, T., Koß, H.-J., and Bardow, A. (2017). Efficient Determination of Liquid–Liquid Equilibria Using Microfluidics and Raman Microspectroscopy. *Ind. Eng. Chem. Res.* In press.

Thompson, P., Cox, D., and Hastings, J. (1987). Rietveld Refinement of Debye–Scherrer Synchrotron X-Ray Data from Al2O3. *J. Appl. Crystallogr.*, 20(2):79–83.

Viell, J., Inouye, H., Szekely, N. K., Frielinghaus, H., Marks, C., Wang, Y., Anders, N., Spiess, A. C., and Makowski, L. (2016). Multi-Scale Processes of Beech Wood Disintegration and Pretreatment with 1-ethyl-3-methylimidazolium Acetate/Water Mixtures. *Biotechnol. Biofuels.*, 9:7.

Viell, J. and Marquardt, W. (2012). Concentration Measurements in Ionic Liquid-Water Mixtures by Mid-Infrared Spectroscopy and Indirect Hard Modeling. *Appl. Spectrosc.*, 66(2):208–217.

Vogt, M., Pasel, C., and Bathen, D. (2011). Characterisation of CO2 Absorption in Various Solvents for PCC Applications by Raman Spectroscopy. *Energy Procedia*, 4:1520–1525.

Wagner, M., von Harbou, I., Kim, J., Ermatchkova, I., Maurer, G., and Hasse, H. (2013). Solubility of Carbon Dioxide in Aqueous Solutions of Monoethanolamine in the low and high Gas Loading Regions. *J. Chem. Eng. Data*, 58(4):883–895.

Wehlburg, C. M., Haaland, D. M., Melgaard, D. K., and Martin, L. E. (2002). New Hybrid Algorithm for Maintaining Multivariate Quantitative Calibrations of a Near-Infrared Spectrometer. *Appl. Spectrosc.*, 56(5):605–614.

Wei, D., Chen, S., and Liu, Q. (2015). Review of Fluorescence Suppression Techniques in Raman Spectroscopy. *Appl. Spectrosc. Rev.*, 50(5):387–406.

Wong, M., Bustam, M., and Shariff, A. (2015). Chemical Speciation of CO2 Absorption in Aqueous Monoethanolamine Investigated by in situ Raman Spectroscopy. *Int. J. Greenh. Gas Control*, 39:139–147.

Woodward, L. and George, J. (1951). Light Flux received by a Spectrograph from a Spatially Extended Refracting Source. *Proc. Phys. Soc. B*, 64(9):780.

Zhou, Y., Zheng, C., Cao, H., Li, X., and Sha, J. (2012). Measurement of Hemoglobin in Whole Blood using a Partial Least Squares Regression Model with selected Second Derivative Near Infrared Transmission Spectral Signals. *Biochem Bioph. Res. Co.*, 420(1):205 – 209.

Zientek, N., Meyer, K., Kern, S., and Maiwald, M. (2016). Quantitative Online NMR Spectroscopy in a Nutshell. *Chem. Ing. Tech.*, 88(6):698–709.

Appendix A

Height Norm, Area Norm, and Sum Norm

In this Chapter, we give the definition of peak profiles in height norm, area norm, and sum norm. Furthermore, we show how to change the parametrization from one norm to another.

A.1 The Height Norm

For a peak in the height norm, the modeled Gaussian and Lorentzian function have a maximum height of unity. If used in IHM, a constant area must be enforced by a constraint in the optimization (Alsmeyer et al. (2004)).

The Gaussian profile G_{height}, its indefinite integral $\int G_{\text{height}}$, and its area $\alpha_{\text{G,height}}$ in the height norm read:

$$G_{\text{height}}(\tilde{\nu}) = \exp\left\{[-\ln(2)] \cdot \left(\frac{\tilde{\nu}-\delta}{\gamma/2}\right)^2\right\}, \tag{A.1}$$

$$\int G_{\text{height}}(\tilde{\nu})\,\mathrm{d}\tilde{\nu} = \left(\frac{\gamma}{2}\right) \cdot \sqrt{\frac{\pi}{4\cdot\ln 2}}\,\mathrm{erf}\left[\ln(2) \cdot \left(\frac{\tilde{\nu}-\delta}{\gamma/2}\right)\right], \tag{A.2}$$

$$\alpha_{\text{G,height}} = \left(\frac{\gamma}{2}\right) \cdot \sqrt{\frac{\pi}{\ln(2)}}. \tag{A.3}$$

The Lorentzian profile L_{height}, its indefinite integral $\int L_{\text{height}}$, and its area $\alpha_{\text{L,height}}$ in the height norm read:

$$L_{\text{height}}(\tilde{\nu}) = \left[\left(\frac{\tilde{\nu}-\delta}{\gamma/2}\right)^2 + 1\right]^{-1}, \tag{A.4}$$

$$\int L_{\text{height}}(\tilde{\nu})\,\mathrm{d}\tilde{\nu} = \left(\frac{\gamma}{2}\right) \cdot \arctan\left(\frac{\tilde{\nu}-\delta}{\gamma/2}\right), \tag{A.5}$$

$$\alpha_{\text{L,height}} = \left(\frac{\gamma}{2}\right) \cdot \pi. \tag{A.6}$$

The pseudo-Voigt profile V_{height}, its indefinite integral $\int V_{\text{height}}\mathrm{d}\tilde{\nu}$, and its area α in the height norm read:

$$V_{\text{height}}(\tilde{\nu}) = h \cdot \left[\beta_{\text{height}} \cdot G_{\text{height}}(\tilde{\nu}) + (1-\beta_{\text{height}}) \cdot L_{\text{height}}(\tilde{\nu})\right], \tag{A.7}$$

$$\int V_{\text{height}}(\tilde{\nu})\,\mathrm{d}\tilde{\nu} = h \cdot \left[\beta_{\text{height}} \cdot \int G_{\text{height}}(\tilde{\nu})\,\mathrm{d}\tilde{\nu} + (1-\beta_{\text{height}}) \cdot \int L_{\text{height}}(\tilde{\nu})\,\mathrm{d}\tilde{\nu}\right], \tag{A.8}$$

$$\alpha = h \cdot \left[\beta_{\text{height}} \cdot \alpha_{\text{G,height}} + (1-\beta) \cdot \alpha_{\text{L,height}}\right]. \tag{A.9}$$

A.2 The Area Norm

For a peak in the area norm, the modeled Gaussian and Lorentzian profiles have an area of unity. If used in IHM, the area parameter should not be changed to ensure a constant peak area.

The Gaussian profile G_{area}, its indefinite integral $\int G_{\mathrm{area}}\mathrm{d}\tilde{\nu}$, and its height $h_{\mathrm{G,area}}$ in the area norm read:

$$G_{\mathrm{area}}\left(\tilde{\nu}\right) = \frac{1}{\alpha_{\mathrm{G,height}}} \cdot G_{\mathrm{height}}\left(\tilde{\nu}\right), \tag{A.10}$$

$$\int G_{\mathrm{area}}\left(\tilde{\nu}\right)\mathrm{d}\tilde{\nu} = \frac{1}{2} \cdot \mathrm{erf}\left[\ln\left(2\right) \cdot \left(\frac{\tilde{\nu}-\delta}{\gamma/2}\right)\right], \tag{A.11}$$

$$h_{\mathrm{G,area}} = \frac{1}{\alpha_{\mathrm{G,height}}}. \tag{A.12}$$

The Lorentzian profile L_{area}, its indefinite integral $\int L_{\mathrm{area}}$, and its height $h_{\mathrm{L,area}}$ in the area norm read:

$$L_{\mathrm{area}}\left(\tilde{\nu}\right) = \frac{1}{\alpha_{\mathrm{L,height}}} \cdot L_{\mathrm{height}}\left(\tilde{\nu}\right), \tag{A.13}$$

$$\int L_{\mathrm{area}}\left(\tilde{\nu}\right)\mathrm{d}\tilde{\nu} = \frac{1}{\pi} \cdot \arctan\left(\frac{\tilde{\nu}-\delta}{\gamma/2}\right), \tag{A.14}$$

$$h_{\mathrm{L,area}} = \frac{1}{\alpha_{\mathrm{L,height}}}. \tag{A.15}$$

The pseudo-Voigt profile V_{area}, its indefinite integral $\int V_{\mathrm{area}}$, and its height h in the area norm read:

$$V_{\mathrm{area}}\left(\tilde{\nu}\right) = \alpha \cdot \left[\beta \cdot G_{\mathrm{area}}\left(\tilde{\nu}\right) + \left(1-\beta\right) \cdot L_{\mathrm{area}}\left(\tilde{\nu}\right)\right], \tag{A.16}$$

$$\int V_{\mathrm{area}}\left(\tilde{\nu}\right) = \alpha \cdot \left[\beta \cdot \int G_{\mathrm{area}}\left(\tilde{\nu}\right) + \left(1-\beta\right) \cdot \int L_{\mathrm{area}}\left(\tilde{\nu}\right)\right], \tag{A.17}$$

$$h = \alpha \cdot \left[\beta \cdot h_{\mathrm{G,area}} + \left(1-\beta\right) \cdot h_{\mathrm{L,area}}\right]. \tag{A.18}$$

A.3 The Sum Norm

In contrast to the area norm, the sum norm does not parametrize the area of the modeled function, but the sum that a peak has in the modeled spectrum. The height of a sum norm peak cannot be calculated without knowing the actual wavenumbers of the spectrum, as larger pixels lead to larger integration ranges and therefore larger heights.

The Gaussian, Lorentzian and pseudo-Voigt profiles in the sum norm for the discrete pixel m read:

$$G_{\mathrm{sum}}\left(\tilde{\nu}_m\right) = \int_{(\tilde{\nu}_{m-1}+\tilde{\nu}_m)/2}^{(\tilde{\nu}_m+\tilde{\nu}_{m+1})/2} G_{\mathrm{area}}\left(\tilde{\nu}\right) \mathrm{d}\tilde{\nu}, \tag{A.19}$$

$$L_{\mathrm{sum}}\left(\tilde{\nu}_m\right) = \int_{(\tilde{\nu}_{m-1}+\tilde{\nu}_m)/2}^{(\tilde{\nu}_m+\tilde{\nu}_{m+1})/2} L_{\mathrm{area}}\left(\tilde{\nu}\right) \mathrm{d}\tilde{\nu}, \tag{A.20}$$

$$V_{\mathrm{sum}}\left(\tilde{\nu}_m\right) = \alpha_{\mathrm{sum}}\left[\beta \cdot G_{\mathrm{sum}}\left(\tilde{\nu}_m\right) + (1-\beta) \cdot L_{\mathrm{sum}}\left(\tilde{\nu}_m\right)\right]. \tag{A.21}$$

A.4 Parameter Conversion

Parameter conversion between Height Norm and Area Norm

For height norm and area norm, an exact parameter conversion is possible.

Equation (A.9) can be used to calculate the area parameter α of the area norm from a parametrization in the height norm.

Equation (A.18) can be used to calculate the height parameter h of the height norm from a parametrization in the area norm.

The Gaussian part of the height norm β_{height} denotes the relative heights of a Gaussian and a Lorentzian profile. The Gaussian part of the area norm β denotes the relative areas of a Gaussian and a Lorentzian profile.

The Gaussian part of the height norm β_{height} as function of the Gaussian part of the area norm β reads:

$$\beta_{\text{height}} = \frac{\beta \cdot h_{\text{G,area}}}{\beta \cdot h_{\text{G,area}} + (1-\beta) \cdot h_{\text{L,area}}}. \tag{A.22}$$

The Gaussian part of the area norm β as function of the gaussian part of the height norm β_{height} reads:

$$\beta = \frac{\beta_{\text{height}} \cdot \alpha_{\text{G,height}}}{\beta \cdot \alpha_{\text{G,height}} + (1-\beta) \cdot \alpha_{\text{L,height}}}. \tag{A.23}$$

The peak width δ and the peak position γ are identical for the height norm and the area norm.

Parameter conversion between Area Norm and Sum Norm

There is no exact parameter conversion between the sum norm and the area norm. Nevertheless, the parameters can be translated approximately for wavenumbers of an actual spectrum. The ratio of a Gaussian profile in the sum norm and in the area norm reads:

$$\frac{G_{\text{sum}}\left(\tilde{\nu}_m\right)}{G_{\text{area}}\left(\tilde{\nu}_m\right)} = \frac{\int_{(\tilde{\nu}_{m-1}+\tilde{\nu}_m)/2}^{(\tilde{\nu}_m+\tilde{\nu}_{m+1})/2} G_{\text{area}}\left(\tilde{\nu}\right) \mathrm{d}\tilde{\nu}}{G_{\text{area}}\left(\tilde{\nu}_m\right)}, \tag{A.24}$$

$$\approx \frac{G_{\text{area}}\left(\tilde{\nu}_m\right) \cdot \left(\frac{\tilde{\nu}_{m+1}-\tilde{\nu}_{m-1}}{2}\right)}{G_{\text{area}}\left(\tilde{\nu}_m\right)}, \tag{A.25}$$

$$= \left(\frac{\tilde{\nu}_{m+1}-\tilde{\nu}_{m-1}}{2}\right). \tag{A.26}$$

Equation (A.25) holds exactly if G_{area} is constant in the range of wavenumbers of pixel m ($[(\tilde{\nu}_{m-1}+\tilde{\nu}_m)/2 \ldots (\tilde{\nu}_m+\tilde{\nu}_{m+1})/2]$). The broader a peak is, relative to a pixel, the better Equation (A.25) holds.

We can do the same derivation and get the same relative intensity for a Lorentzian function (not shown). As there is only a constant factor between the Lorentzian and the Gaussian profile in both norms, the Gaussian part β is the same. The quotient of a pseudo-Voigt peak in the area norm and a pseudo-Voigt peak in the sum reads:

$$\begin{aligned}\frac{V_{\text{sum}}\left(\tilde{\nu}_m\right)}{V_{\text{area}}\left(\tilde{\nu}_m\right)} &= \frac{\alpha_{\text{sum}}\left[\beta \cdot G_{\text{sum}}\left(\tilde{\nu}_m\right) + (1-\beta) \cdot L_{\text{sum}}\left(\tilde{\nu}_m\right)\right]}{\alpha \cdot \left[\beta \cdot G_{\text{area}}\left(\tilde{\nu}_m\right) + (1-\beta) \cdot L_{\text{area}}\left(\tilde{\nu}_m\right)\right]} \\ &\approx \frac{\alpha_{\text{sum}}\left[\beta \cdot G_{\text{area}}\left(\tilde{\nu}_m\right) + (1-\beta) \cdot L_{\text{area}}\left(\tilde{\nu}_m\right)\right] \cdot \left(\frac{\tilde{\nu}_{m+1}-\tilde{\nu}_{m-1}}{2}\right)}{\alpha \cdot \left[\beta \cdot G_{\text{area}}\left(\tilde{\nu}_m\right) + (1-\beta) \cdot L_{\text{area}}\left(\tilde{\nu}_m\right)\right]} \\ &= \frac{\alpha_{\text{sum}}}{\alpha}\left(\frac{\tilde{\nu}_{m+1}-\tilde{\nu}_{m-1}}{2}\right). \end{aligned} \tag{A.27}$$

As the pseudo-Voigt profiles in both norms should have the same intensity ($\frac{V_{\text{sum}}(\tilde{\nu}_m)}{V_{\text{area}}} = 1$), we get the following equation for the parameter conversion:

$$\alpha \approx \alpha_{\text{sum}} \cdot \left(\frac{\tilde{\nu}_{m+1}-\tilde{\nu}_{m-1}}{2}\right). \tag{A.28}$$

For the position parameter δ and the peak width γ, we use the same parameter values for both parametrizations.

Appendix B

Remarks on the Numerical Optimization with Matlab

In this thesis, a Matlab program has been implemented to perform IHM, FD-IHM, and VHM. We fit the spectra with 'lsqcurvefit' as implemented in Matlab. lsqcurvefit is a nonlinear least-squares solver that offers a Levenberg-Marquardt algorithm and a Trust-Region-reflective algorithm. Both algorithms are gradient based.

Lower and Upper Bound Constraints for the Levenberg-Marquardt Algorithm

The Levenberg Marquardt algorithm is usually faster but does not support bound constraints. The Trust-Region-Reflective solver supports bound constraints but is slower.

We need certain bound constraints for peak fitting and component fitting. The peak area α must be positive and the gaussian part β must be in the interval of zero to one. To be able to use the Levenberg-Marquardt solver, we use auxiliary functions that project the solver parameter par_{solver} to the peak parameter par_{peak} in the way that $\mathbb{R} \rightarrow [lb..ub]$. lb denotes the lower bound constraint, ub denotes the upper bound constraint, par_{solver} denotes the parameter that is visible to the solver, and par_{peak} denotes the peak parameter.

For lower and upper bound constraints, the absolute function can be sometimes used (*e.g.*, $par_{\text{peak}} = |par_{\text{solver}}| + lb$). However, when using the absolute function, the gradient of the solver parameter has a discontinuity at $par_{\text{solver}} = 0$. For a gradient based solver this discontinuity greatly prolongs the calculation time for values close to zero. Therefore, we use the following auxiliary functions that do not have any discontinuity in the gradient:

if only a lower bound constraint is needed, we use:

$$par_{\text{peak}} = lb - 3 + \sqrt{9 + par_{\text{solver}}^2}, \tag{B.1}$$

if only an upper bound constraint is needed, we use:

$$par_{\text{peak}} = ub + 3 - \sqrt{9 + par_{\text{solver}}^2}, \tag{B.2}$$

if lower and upper bound constraints are needed (as for the Gaussian part β), we use:

$$par_{\text{peak}} = lb + \left[\frac{\arctan\left(par_{\text{solver}}\right)}{\pi} + 0.5\right] \cdot (ub - lb). \tag{B.3}$$

Equations (B.1) and (B.3) are plotted in Figure B.1.

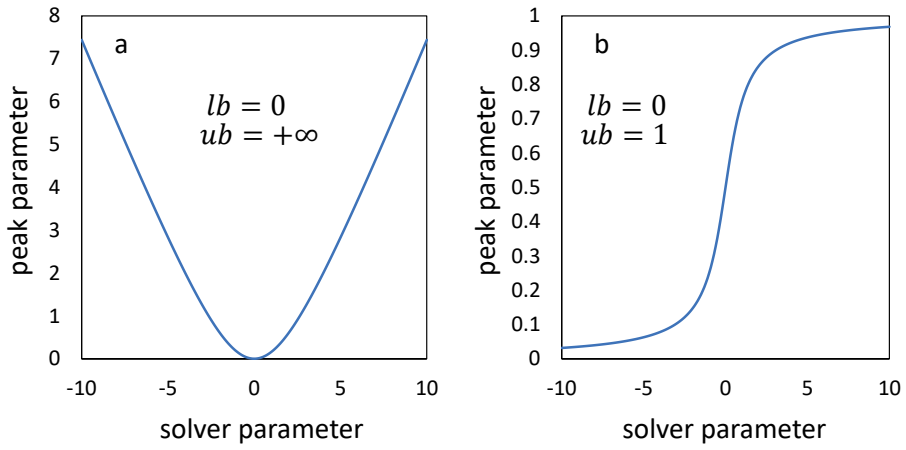

Figure B.1: Auxiliary functions to add lower and upper bound constraints to the Levenberg-Marquardt solver. (a) for only a lower bound constraint of 0 (*cf.* Equation (B.1)), (b) for a lower bound constraint 0 and an upper bound constraint of 1 (*cf.* Equation (B.3)).

Providing a Jacobian for the Solver

Providing a Jacobian J to the lsqcurvefit solver, greatly cuts the number of needed function evaluations. For each fitted parameter, lsqcurvefit expects the partial derivative of the modeled spectrum $\mathbf{M}$ with respect to each parameter Θ_i:

$$\mathbf{J}(\Theta_i) = \frac{\partial \mathbf{M}}{\partial \Theta_i}. \tag{B.4}$$

In principle, we have two options to calculate the partial derivative $\frac{\partial \mathbf{M}}{\partial \Theta_i}$. We can provide the partial derivative for each parameter analytically; if an analytical form exists. Or we can simply calculate the partial derivative numerically:

$$\mathbf{J}(\Theta_i) \approx \frac{\mathbf{M}(\Theta_i + \epsilon) - \mathbf{M}(\Theta_i)}{\epsilon}. \tag{B.5}$$

While an analytical partial derivative is more sound, the numerical partial derivative facilitates changes to the code as, *e.g.*, introduction of a new peak profile. We chose to implement the numerical derivative as we are more concerned with changeability than with maximum performance.

As model spectrum $\mathbf{M}$, we consider an IHM model pure component model, which is essentially a sum of pseudo-Voigt profiles V. We recall Equation (2.9):

$$\mathcal{A}_k\left(\tilde{\nu}, \mathbf{\Theta}_k^*\right) = \sum_{p=1}^{n_{\text{peaks}}} V\left(\tilde{\nu}, \Theta_{k,p,1}^* \ldots \Theta_{k,p,4}^*\right). \tag{B.6}$$

The additive structure of the pure component model allows for an efficient calculation of the Jacobian. Each peak must be calculated once anyway to calculate the model spectrum of the current parameter set. For the partial derivative of each parameter, we only calculate the respective peak with the parameter changed by epsilon. The partial derivative for each peak parameter reads:

$$\mathbf{J}(\Theta_{k,p,i}^*) \approx \frac{\mathbf{V}\left(\Theta_{k,p,i}^* + \epsilon\right) - \mathbf{V}\left(\Theta_{k,p,i}^*\right)}{\epsilon}. \tag{B.7}$$

In Equation (B.7), $\mathbf{V}$ denotes the vector of the pseudo-Voigt function for all discrete wavenumbers $\tilde{\nu}_m$.

Appendix C

Different Concentration Measures for Very Hard Modeling

We use mass fractions w as the argument for the polynomial polynomial $f(w)$ (*cf.* Equations (5.1) and (5.2)), as mass fractions proved to work best. However, the argument of the polynomial has a large influence on the quality of the model, in particular for low order polynomials. Beyond the mass fraction we also investigated a mole fraction and a "generalized fraction" in a preliminary study. The generalized fraction $(\cdot)$ reads:

$$(\cdot)_i = \frac{a_i w_i}{\sum_{j=1}^{n_C} a_j w_j} \quad \text{where } a_1 = 1 \tag{C.1}$$

Equation (C.1) is a generalized fraction to transform a mass fraction into another $(\cdot)$ -fraction. For example, if a_j is chosen as a quotient of molar masses ($a_j = \frac{M_i}{M_j}$), $(\cdot)_i$ equals the mole fraction. Similarly, a mass fraction can be converted into a volume fraction by a quotient of densities. Equation (C.1) introduces one extra parameter to the very Hard Model for a binary system (parameter a_2) or $n_C - 1$ extra parameters for a multi component system. In the preliminary study, mass fractions and the generalized fraction led to very similar results. For the sake of simplicity we thus use mass fractions in Section 5.1.1.

Aachener Beiträge zur Technischen Thermodynamik

ABTT 1
Philip Voll
Automated Optimization-Based Synthesis of Distributed Energy Supply Systems
1. Auflage 2014
ISBN 978-3-86130-474-6

ABTT 2
Johannes Jung
Comparative Life Cycle Assessment of Industrial Multi-Product Processes
1. Auflage 2014
ISBN 978-3-86130-471-5

ABTT 3
Franz Lanzerath
Modellgestützte Entwicklung von Adsorptionswärmepumpen
1. Auflage 2014
ISBN 978-3-86130-472-2

ABTT 4
Thorsten Brands
Einfluss der Gemischzusammensetzung auf die Verbrennung im Diesel- und GCAI-Motor
1. Auflage 2014
ISBN 978-3-95886-006-3

ABTT 5
Dominique Dechambre
Efficient Measurement of Liquid-Liquid Equilibria using Automation and Optimal Experimental Design
1. Auflage 2016
ISBN 978-395886-077-3

ABTT 6
Niklas von der Aßen
From Life-Cycle Assesement towards life-Cycle Design of Carbon Dioxide Capture and Utilization
1. Auflage 2016
ISBN 978-3-95886-080-3

ABTT 7
Matthias Lampe
Integrated Process and Organic Rankine Cycle Working Fluid Design in the Continuous-Molecular Targeting Framework
1. Auflage 2016
ISBN 978-3-95886-086-5

ABTT 8
Thomas Hülser
Optische Untersuchung der Zündvorgänge und deren Auswirkung auf die Verbrennung in PKW-Motoren
1. Auflage 2016
ISBN 978-3-95886-090-2

Aachener Beiträge zur Technischen Thermodynamik

ABTT 9
Malte Döntgen
Reaction Models from Reactive Molecular Dynamics and High-Level Kinetics Predictions
1. Auflage 2016
ISBN 978-3-95886-156-5

ABTT 10
Heike Schreiber
Experiments and Validated Models for Adsorption Thermal Energy Storage in Industrial and Residential Application
1. Auflage 2017
ISBN 978-3-95886-178-7

ABTT 11
André Dirk Sternberg
System-Wide Perspective for Life Cycle Assesment of CO_2-based C1-Chemicals
1. Auflage 2017
ISBN 978-3-95886-193-0

ABTT 12
Uwe Bau
From Dynamic Simulation to Optimal Design and Control of Adsorption Energy Systems
1. Auflage 2018
ISBN 978-3-95886-216-6

ABTT 13
Christian Jens
Modellbasiertes Design von Produkt, Lösungsmittel und Prozess für die Ameisensäure-synthese aus CO_2 und H_2
1. Auflage 2018
ISBN 978-3-95886-231-9

ABTT 14
Jan David Scheffczyk
Integrated Computer-Aided Design of Molecules and Processes using COSMO-RS
1. Auflage 2018
ISBN 978-3-95886-236-4

ABTT 15
Björn Bahl
Optimization-Based Synthesis of Large-Scale Energy Systems by Time-Series Aggregation
1. Auflage 2018
ISBN 978-3-95886-240-1

Aachener Beiträge zur Technischen Thermodynamik

ABTT 16
Bastian Liebergesell
A Milliliter-Scale Setup for the Efficient Characterization of Multicomponent Vapor-Liquid Equilibria Using Raman Spectroscopy
1. Auflage 2018
ISBN 978-3-95886-247-0

ABTT 17
Stefan Wilhelm Graf
A Design Approach for Adsorption Energy Systems Integrating Dynamic Modeling with Small-Scale Experiments
1. Auflage 2018
ISBN 978-3-95886-258-6

ABTT 18
Sebastian Kaminski
Quantum-Mechanics-Based Prediction of SAFT Parameters for Non-Associating and Associating Molecules Containing Carbon, Hydrogen, Oxygen and Nitrogen
1. Auflage 2019
ISBN 978-3-95886-270-8

ABTT 19
Maike Renate Hennen
Decision Support for the Synthesis of Energy Systems by Analysis of the Near-Optimal Solution Space
1. Auflage 2019
ISBN 978-3-95886-277-7

ABTT 20
Peyman Yamin
COSMO-RS-Based Methods for Improved Modelling of Complex Chemical Systems
1. Auflage 2019
ISBN 978-3-95886-288-3

ABTT 21
Meltem Erdogan
Assessement of Adsorbents for Drying by Experiments and Dynamic Simulations
1. Auflage 2019
ISBN 978-3-95886-303-3

ABTT 22
Christian Schulz
SRS/LIF-Messungen zur Charakterisierung rußarmer dieselähnlicher Flammen von alternativen Kraftstoffen und n-Heptan
1. Auflage 2019
ISBN 978-3-91886-310-1

Aachener Beiträge zur Technischen Thermodynamik

ABTT 23
Peter Beumers
Physically-Based Models for the Analysis of Raman Spectra
1. Auflage 2019
ISBN 978-3-95886-319-4