

Multianalyte Quantifications by Means of
Integration of Artificial Neural Networks,
Genetic Algorithms and Chemometrics for
Time-Resolved Analytical Data

Multi-Analyt Quantifizierungen mit Hilfe der
Integration von künstlichen neuronalen Netzen,
genetischen Algorithmen und der Chemometrie
für zeitaufgelöste analytische Daten

DISSERTATION

der Fakultät für Chemie und Pharmazie
der Eberhard-Karls-Universität Tübingen

zur Erlangung des Grades eines Doktors
der Naturwissenschaften

2003

vorgelegt von

Frank Jochen Dieterle

Tag der mündlichen Prüfung: 25. Juli 2003

Dekan: Prof. Dr. H. Probst

1. Berichterstatter: Prof. Dr. G. Gauglitz

2. Berichterstatter: PD Dr. U. Weimar

3. Berichterstatter: Prof. Dr. J. Gasteiger

Table of Contents

1.	INTRODUCTION	6
2.	THEORY – FUNDAMENTALS OF THE MULTIVARIATE DATA ANALYSIS	10
2.1.	Overview of the Multivariate Quantitative Data Analysis	10
2.2.	Experimental Design	11
2.3.	Data Preprocessing	12
2.4.	Data Splitting and Validation	13
2.5.	Calibration of Linear Relationships	16
2.6.	Calibration of Nonlinear Relationships	19
2.7.	Neural Networks – Universal Calibration Tools	20
2.8.	Too Much Information Deteriorates Calibration	24
2.9.	Measures of Error and Validation	38
3.	THEORY – QUANTIFICATION OF THE REFRIGERANTS R22 AND R134A: PART I	39
3.1.	Experimental	39
3.2.	Single Analytes	40
3.3.	Sensitivities	44
3.4.	Calibrations of the Mixtures	45
3.5.	Variable Selection by Brute Force	48
3.6.	Conclusions	48
4.	EXPERIMENTS, SETUPS AND DATA SETS	50
4.1.	The Sensor Principle	50
4.2.	SPR Setup	51
4.3.	RfS Sensor Array	52
4.4.	4λ Miniaturized RfS Sensor	53
4.5.	Data Sets	54
5.	RESULTS – KINETIC MEASUREMENTS	60
5.1.	Static Sensor Measurements	60
5.2.	Time-resolved Sensor Measurements	61

5.3.	Makrolon – A Polymer for Time-resolved Measurements _____	63
5.4.	Conclusions _____	73
6.	RESULTS – MULTIVARIATE CALIBRATIONS _____	74
6.1.	PLS Calibration _____	74
6.2.	Box-Cox Transformation + PLS _____	80
6.3.	INLR _____	82
6.4.	QPLS _____	83
6.5.	CART _____	84
6.6.	Model Trees _____	86
6.7.	MARS _____	88
6.8.	Neural Networks _____	90
6.9.	PCA-NN _____	91
6.10.	Neural Networks and Pruning _____	92
6.11.	Conclusions _____	94
7.	RESULTS – GENETIC ALGORITHM FRAMEWORK _____	96
7.1.	Single Run Genetic Algorithm _____	96
7.2.	Genetic Algorithm Framework - Theory _____	98
7.3.	Genetic Algorithm Framework - Results _____	102
7.4.	Genetic Algorithm Framework – Conclusions _____	106
8.	RESULTS – GROWING NEURAL NETWORK FRAMEWORK _____	107
8.1.	Modifications of the Growing Neural Network Algorithm _____	108
8.2.	Application of the Growing Neural Networks _____	109
8.3.	Growing Neural Network Algorithm Frameworks _____	112
8.4.	Applications of the Growing Neural Network Frameworks _____	115
8.5.	Conclusions and Comparison of the Different Methods _____	121
9.	RESULTS – ALL DATA SETS _____	123
9.1.	Methanol and Ethanol by SPR _____	123
9.2.	Methanol, Ethanol and 1-Propanol by SPR _____	129
9.3.	Methanol, Ethanol and 1-Propanol by the RfS Array and the 4 λ Setup _____	137

9.4.	Quaternary Mixtures by the SPR Setup and the RfS Array _____	144
9.5.	Quantification of the Refrigerants R22 and R134a in Mixtures: Part II _____	148
10.	RESULTS – VARIOUS ASPECTS OF THE FRAMEWORKS AND MEASUREMENTS _____	149
10.1.	Single or Multiple Analyte Rankings _____	149
10.2.	Stopping Criteria for the Parallel Frameworks _____	150
10.3.	Optimization of the Measurements _____	152
10.4.	Robustness and Comparison with Martens' Uncertainty Test _____	155
11.	SUMMARY AND OUTLOOK _____	156
12.	REFERENCES _____	161
13.	PUBLICATIONS _____	179
14.	ACKNOWLEDGEMENTS _____	181
15.	APPENDIX _____	183

1. Introduction

During the last century, the instrumentation of analytical chemistry has dramatically changed. Advances in classical analytical setups, developments of new devices and applications of new measurement principles allow the acquisition of more information about an analytical problem in a shorter time. Faster working equipments and the parallelizing of devices enable measurements of more samples making in depth examinations of complex systems possible. State of the art devices allow the acquisition of more detailed information about samples by utilizing more wavelengths or additional sensors. Finally yet importantly, new measurement principles such as time-resolved measurements render the access to new sources of information possible.

This constantly increasing flood of information puts a new challenge to the field of data analysis, which can be considered as the link between the raw information provided by the instrumentation and the questions to be answered for the analyst. Being so universal the data analysis has many facets in the different areas of analytical chemistry such as qualitative analysis, quantitative analysis, optimization problems, identification of significant factors and many more. The diversity of data analysis for analytically relevant questions is also reflected in a number of different names for the same discipline like chemometrics, chem(o)-informatics, bioinformatics, biometrics, environmetrics, and data mining.

This work covers a wide variety of aspects of data analysis for chemical sensor systems ranging from the introduction and optimization of new measurement procedures to the preprocessing of the raw sensor signals and from the calibration of the sensors to the identification of important factors. Being interconnected and thus influencing each other, all these aspects have to be considered when setting up a sensor system for a certain analytical task. However, the main objectives of this work can be subsumed into two focuses.

The first focus is the introduction and optimization of kinetic measurements in chemical sensing. Thereby the effect is exploited that different analytes show different kinetics of sorption into the sensor coatings. This allows access to a completely new domain of information compared with commonly used measurement procedures of chemical sensing. The new approach of time-resolved measurements uses the kinetic information of the sensor responses not for the investigation of the interaction kinetics of the analytes with the sensor coatings but for the quantitative determination of several analytes in mixtures. In contrast to

some rare reports found in literature, which use the kinetic information as a given phenomenological effect to improve the multi-analyte quantification, a systematic investigation of the principles of time-resolved measurements is performed in this work. Thereby different aspects are investigated such as the interaction principles, the optimization of the measurement parameters, the relationships between the time-resolved sensor responses and the analytes, the transfer of the measurement principles to different setups and to different analytes and many more. This systematic investigation demonstrates that the principle of time-resolved measurements forms the basis for a simultaneous quantification of several analytes by single sensor systems. It is furthermore shown that sensor arrays also profit from this approach by the possibility of identifying and quantifying more analytes than before for a given sensor array setup. Consequently, this approach generally allows the reduction of the number of sensors resulting in smaller devices and less costs for the hardware. The systematic investigation also demonstrates that the principle is a very powerful and generic approach not limited to the setups, analytes, and interaction principles used in this study.

The large amount and the complexity of the data generated by time-resolved measurements necessitate the second focus of this work, which is the application and optimization of natural computing methods for the data analysis of sensors. The expression "natural computing" primarily refers to two concepts of computing copied from nature. The concept of neural networks has been inspired by the highly interconnected neural structures in the brain and the nervous system of mammals, whereas the concept of genetic algorithms has been inspired by the evolution in biology. For the data analysis in this study, the neural networks are used for the calibration of the data. It is demonstrated in this work that only the neural networks out of many multivariate calibration methods are capable of calibrating the nonlinear relationship between the sensor responses and the concentrations of the different analytes. Genetic algorithms are applied for the identification and selection of significant factors respectively variables and thus for the optimization of the calibration. Yet, it is shown that an often-reported combination of both concepts is faced with several problems with respect to the limited number of measurements. Thus, several frameworks are developed, implemented and optimized in this work, which use data sets limited in size in a very efficient way. These frameworks contain neural networks for the calibration, genetic algorithms respectively growing neural networks for the selection of significant variables and additional procedures and approaches from statistics and chemometrics for significance test. These new frameworks are designed to fulfill the needs of analytical chemistry such as a high performance of data analysis, an easy application of the algorithms, a portability to a wide range of data sets and

devices, an insight into the models built, an identification of important factors, a high robustness to noise in the data and the ability to cope with data sets limited in size. The frameworks are applied to several data sets, which were recorded by different devices in our laboratory. Two data sets have an environmental background based on the recycling of old refrigerants of air-conditioners and refrigerators. Additionally, the homologous series of the lower alcohols was measured several times allowing a systematic investigation of the time-resolved measurements. For all data sets under investigation, the frameworks show excellent results for calibration and variable selection. The frameworks also demonstrate that there are several possibilities to tweak the time-resolved measurements with respect to measurement time, properties of the sensitive layers, carrier gas and much more. The frameworks developed in this work are not limited to the calibration and optimization of sensor data, but can be used for virtually any multivariate calibration.

The outline of this study can be described as follows. The work starts with an overview of the multivariate data analysis. Several up-to-date concepts, methods and algorithms are presented and the advantages and problems are discussed. Thereby the focus is on two concepts, multivariate calibration and selection of variables. In the next chapter, a multivariate data analysis is performed using a data set recorded in our lab as an example for a data analysis, which is accepted as the current state of research in literature. Starting with this state of research, the studies and innovations of this work enhance several concepts presented in this and the previous chapter. Additionally, the different concepts of sorption of analytes into sensitive layers are presented and discussed in this chapter. The next chapter briefly presents the different sensor setups used for recording several data sets, which are presented afterwards.

In the following chapter, the principle of time-resolved measurements is introduced and explained. A systematic investigation of the time-resolved measurements is performed with respect to the theoretical background of this principle and with respect to the interaction principle between the sensitive layers and analytes used in this study. Thereby different properties of the sensitive layers, which are the basis for the time-resolved measurements, are investigated and modified allowing the optimization of the measurements.

Starting with chapter 6, all methods and concepts, which are developed, are demonstrated using one single data set. This allows an easy comparison of the methods. Thus, the improvements by the continually developed concepts can be monitored easily. First, common methods of multivariate calibration are applied resulting in rather poor calibrations. In the next chapter,

neural networks as the most promising method are further developed by the implementation of genetic algorithms, neural networks and statistical procedures into a framework, which is introduced in this work for the first time. The framework shows a superior calibration compared to the widespread methods for the multivariate calibration applied to the data in the previous chapter.

After that, two similar frameworks are introduced for the implementation of a new type of neural networks, which are called growing neural networks, resulting in the best calibration of the data set. These frameworks are unique with respect to finding automatically optimal neural network topologies with practically no input needed by the analyst. In chapter 9, an overview of the results is given for all data sets using commonly applied multivariate data analysis methods and the superior new frameworks for data analysis introduced in this work. Miscellaneous minor issues of the frameworks are discussed afterwards. The work ends with a summary of the results and some suggestions for further research.

2. Theory – Fundamentals of the Multivariate Data Analysis

2.1. Overview of the Multivariate Quantitative Data Analysis

Multivariate quantitative data analysis is part of the scientific field of chemometrics. In a recent review [1] chemometrics was defined as a process, in which measurements are made, data are collected and information is obtained. The multivariate quantitative data analysis, which tries to describe relationships between two groups of variables, also is subject to this process. A practical implementation of the process could look like this:

1. First, different factors like the analytes of interest and interfering substances have to be identified, which might influence the measurements.
2. Then, an experimental design has to be setup, which defines how many samples have to be measured and how to vary the different analyte concentrations and other factors for these samples.
3. Afterwards, these samples are measured, the responses of the device are recorded, and the raw data are optionally preprocessed.
4. After that, a calibration is performed, which tries to model a relationship between the factors such as the concentrations of the analytes, which are generally called independent variables, input variables or simply x -variables, and the responses of the device, which are called dependent variables, response variables or simply y -variables, ending up in a model. Usually, the quality of the calibration is judged by the prediction of additional validation data. Thereby the model does not know the true concentrations of the analytes but predicts these concentrations based on the input variables (device responses). These predictions are compared with the true concentrations in a mathematical manner by using a measure of error or in a graphical manner by using true-predicted plots.
5. Often, an optimization of the calibration or an interpretation of the established model follows. Finally, the model can be applied to new measurements in routine analysis (but has to be validated and updated from time to time).

In the next sections, several fundamental approaches and steps in multivariate calibration and their implementations in this work are explained in more detail.

2.2. *Experimental Design*

Having defined the type and concentration range of the analytes of interest and of the additional factors like temperature or humidity (or generally the independent variables), a plan has to be setup, which determines the number and compositions of the samples to be measured. This plan is known as experimental design in chemometrics. The experimental design tries to cover optimally the space spanned by the independent variables with as few samples as possible to understand the effects of these variables and to model the relationships between the dependent and independent variables. Among the many existing types of experimental designs, several designs are specialized for optimization strategies like the Central Composite Designs, Doehlert Design or Box-Behnken Design [2,3], several designs are mixture designs when all components add up to 100% and several designs such as the D-optimal designs [4-6] are specialized for a constrained variable space. In this study, the concentrations of the different analytes should be independently varied and the number of concentration levels and thus the number of samples should not be constrained rendering most of these designs useless. Thus, full factorial designs are used, which combine all levels of all independent variables (all defined concentration levels of all analytes). This results in a rapidly increasing number n of samples for an increasing number x of analytes and for an increasing number l of concentration levels per analyte:

$$n = l^x \quad (1)$$

In this work, full factorial designs with and without equidistant levels are used for the calibration data sets. For most validation data sets, also full factorial designs are used. Thereby, the meshes of the two designs are interleaved with a maximum distance of the meshes allowing the validation data to give a realistic estimation of the network performance in a real-world situation [7].

2.3. Data Preprocessing

Data preprocessing can be used for systematically modifying the raw signals of the device with the hope that the altered signals provide more useful input to the calibration method. Unfortunately, no general guidelines exist to determine the appropriate data preprocessing technique and thus the different preprocessing techniques are controversially discussed in literature [7,8].

In this work, the input variables are preprocessed by autoscaling according to:

$$x'_{ij} = \frac{x_{ij} - \bar{x}_j}{\sigma_{x_j}} \quad (2)$$

With x_{ij} as response of the i^{th} sample at the j^{th} variable, \bar{x}_j as the mean of the j^{th} variable and σ_{x_j} as the standard deviation of the j^{th} variable. Autoscaling involves a mean-centering of the data and a division by the standard deviation of all responses of a particular input variable resulting in a mean of zero and a unit standard deviation of each variable. For some calibration methods autoscaling can improve the calibration as autoscaling allows all variables to influence equally the calibration especially if different variables show different magnitudes of variation.

The dependent variables were range-scaled between -0.9 to 0.9, which is essential for calibration by neural networks with hyperbolic tangent activation functions, according to:

$$x'_{ij} = 1.8 \frac{x_{ij} - \min(x_j)}{\max(x_j) - \min(x_j)} - 0.9 \quad (3)$$

For the calculation of the prediction errors and the true-predicted plots, the range-scaling was reversed.

2.4. Data Splitting and Validation

A typical multivariate calibration procedure needs several separate data sets. The calibration or training data set is needed for setting up the model by estimating the parameters of an equation or for training a neural network. Often a second data set is needed to determine when to stop the training or to determine how many and which model components and variables to include. This second data set is usually called monitor data set. If several models are developed, a third data set called test set is required to select the most appropriate model. Finally, a validation data set is essential to estimate the quality of the final model. It has been shown that different data are needed for all these data sets, as otherwise the models and estimations are biased [9-12]. For example, if the same data set is used for the calibration and validation, the estimation of the prediction ability is overly optimistic. Additionally, each data set should be as large as possible. The larger the calibration data set the better the model and the larger the validation data set the better the estimation of the predictivity. If many data are available, representative large independent samples can be used for training, monitoring, testing and validating by simply partitioning the large pool of all samples. Typically in analytical chemistry, only data sets limited in size are available as measurements are expensive and work intensive. To solve the dilemma of partitioning a small pool of data into independent data subsets, which should be as large and as representative as possible, subsampling procedures, which are also known as resampling procedures, have become the quasi standard in chemometrics. There are many subsampling techniques, whereby the most important ones are described below.

Crossvalidation

The most popular subsampling technique is crossvalidation. For an n -fold crossvalidation, the data are partitioned into n equal parts. The first part is used as test data set; the rest is used as calibration data set. Then, the second part is used for the test data and the rest is used for a new calibration. This procedure is repeated n times and the predictions of the n test data are averaged. It is essential that no knowledge of the models is transferred from fold to fold. There exist no clear rules how many folds to use for the crossvalidation, whereby the simplest and clearest way of performing crossvalidation is to leave one sample out at a time. This special variant of crossvalidation is also called full crossvalidation, leave-one-out or jackknifing and gives a unique and therefore reproducible result. Yet, it has been shown that increasing the number of crossvalidation groups results in lower root mean square errors of

predictions giving overly optimistic estimations of predictivity [13-16]. This deficiency is known as asymptotically inconsistency in literature [17].

Bootstrapping

Bootstrap resampling was originally developed to help analysts determine how much their results might have changed if another random sample had been used instead and how different the results might be when a model is applied to new data. Bootstrapping has also gained an increasing popularity in the field of resampling small data sets [18]. Bootstrapping is based on sampling with replacement to form a calibration set. For the most popular variant, the 0.632 bootstrap, n times a sample is selected from n samples for the calibration set whereby the same sample can be selected several times. Then, the samples, which were not picked, are used for the test set. The chance that a particular sample is not picked for the calibration set is:

$$\left(1 - \frac{1}{n}\right)^n \approx e^{-1} \approx 0.368 \quad (4)$$

Consequently, the test set contains about 36.8% of the samples and the calibration set about 63.2% with some samples replicated in the calibration set. Bootstrapping is not affected by asymptotic inconsistency and might be the best way of estimating the error for very small data sets whereby the complete procedure can be repeated arbitrarily often [9].

Random Subsampling

Random subsampling, which is also known as Monte Carlo crossvalidation [19], as multiple holdout or as repeated evaluation set [20], is based on randomly splitting the data into subsets, whereby the size of the subsets is defined by the user [21]. The random partitioning of the data can be repeated arbitrarily often. In contrast to a full crossvalidation procedure, random subsampling has been shown to be asymptotically consistent [17] resulting in more pessimistic predictions of the test data compared with crossvalidation. The predictions of the test data give a realistic estimation of the predictions of external validation data [22].

Kennard Stones

The Kennard Stones algorithm [23-25] has gained an increasing popularity for splitting data sets into two subsets. The algorithm starts by finding 2 samples that are the farthest apart from each other on the basis of the input variables. These 2 samples are removed from the original data set and put into the calibration data set. This procedure is repeated until the desired number of samples has been reached in the calibration set. The advantages of this algorithm

are that the calibration samples map the measured region of the variable space completely and that the test samples all fall inside the measured region. Yet, this algorithm is only usable for a single subsampling run, as the partitioning of the data is unique rendering the algorithm for a resampling procedure unusable.

Kohonen Neural Networks

An interesting approach for splitting a data set into two subsets is the application of Kohonen neural networks [26,27,50]. These networks with two layers are unsupervised networks, which can be used as 2-dimensional mapping method. For the repartitioning, a Kohonen network is trained using the complete data set. Then, for each neuron a specific number of samples which excited this neuron during training, are selected for the first data set. The other samples are used for the second data set. This approach allows a very efficient distribution of the samples into subsets that cover the complete variable space.

Yet, using Kohonen networks for several subsampling runs is difficult, as the creation of different selection rules for samples exciting a neuron is rather subjective for an arbitrary number of runs and needs user input from data set to data set.

Conclusions

When comparing the advantages and disadvantages of the different subsampling algorithms bootstrapping and random subsampling are most suited for splitting the data into calibration, test and validation subsets. As the user definable ratio between the sizes of the different subset allows a high flexibility, the random subsampling procedure was used to split the data into calibration, test and monitor data sets in this work, whereas for most data sets a static external validation set was recorded and used. The monitor set for the early-stopping procedure of the neural networks (see section 2.7.3) was generated by a modified full crossvalidation procedure, which speeds up learning and which is described in detail in [28].

Besides of the averaging effect of the subsampling procedure, the comparison of the standard deviations between the predictions of the test data of the different subsets additionally allows an estimation of the robustness of the calibration method. A high standard deviation is an indication of the calibration being subject to the random partitioning of the data. If the quality of the calibration and prediction significantly depends on the perturbation of the data sub sets, the calibration method is not very robust.

2.5. Calibration of Linear Relationships

For multivariate calibration of the concentrations of one or more analytes (variables y) and the device responses (independent variables x) the simplest and the most common approach is the assumption of a linear relationship. A model for this linear relationship can be setup by a multiple linear regression (MLR), which is also known as inverse least-squares regression (ILS). This calibration can be seen as an extension of the univariate linear regression. For each response variable, a linear equation is formulated:

$$\begin{aligned} y_1 &= b_{11}x_1 + b_{12}x_2 + b_{13}x_3 + \dots + b_{1w}x_w + e_1 \\ y_2 &= b_{21}x_1 + b_{22}x_2 + b_{23}x_3 + \dots + b_{2w}x_w + e_2 \\ &\vdots \\ y_c &= b_{c1}x_1 + b_{c2}x_2 + b_{c3}x_3 + \dots + b_{cw}x_w + e_c \end{aligned} \quad (5)$$

with b_{ij} as regression coefficients and e_j as residuals. Using matrix notation equation (5) can be expressed as:

$$Y = BX + E \quad (6)$$

The regression coefficients can be estimated by

$$\hat{B} = (X^T X)^{-1} X^T Y \quad (7)$$

For the inversion of the variance-covariance matrix (7) at least as many samples as device responses have to be measured. This inversion (7) of the variance-covariance matrix of the independent variables, which is needed not only by MLR but also by most linear regression techniques to identify the model parameters, causes several problems. If the variables show a collinear behavior or if the variables are highly correlated, the resulting variance-covariance matrix will be ill-conditioned leading to unreliable model parameters and thus producing unstable calibration models.

Several methods have been developed to overcome these problems like principal component regression (PCR), partial least squares regression (PLS or PLSR), ridge regression (RR) and many more. Among these methods, "PLSR is the de facto standard for constructing a multivariate model" [29].

2.5.1. PLS

PLS, which was originally developed for IR- and UV-spectroscopy owes its wide application to its speed, robustness and user-friendliness. PLS performs a linear regression in a new coordinate system with a lower dimensionality than the original space of the independent variables. The new coordinates are called PLS factors or principal components (the latter is less correct but widely used analogous to the principal components for the PCA). The principal components are determined by the maximum variance of the independent variables and by a maximum correlation with the dependent variable(s). There are as many principal components as variables to predict, but for the actual model only the primary, most important principal components are used. This makes PLS robust to noise, as in theory the noise should be encapsulated in the less important principal components and the information of interest should be represented by the primary principal components. The actual regression is performed in the space spanned by the new reduced coordinate system of the orthogonal principal components. Different criteria exist for the number of principal components to be used: One of the easiest methods compares the corresponding eigenvalue of the principal component with the eigenvalues of the higher components using an F-test [30,107]. Eastman et al. [31] used a crossvalidation method for the determination of the optimal number of principal components, which corresponds to the minimal predicted residual sum of squares. This criterion is more conservative in terms of the number of principal components and is widespread in literature. Recently, Martens et al. introduced the Martens' Uncertainty Test [32,33], which uses a jackknifing procedure with many sub-models to determine the significant variables and the optimal number of principal components, which are found in an iterative procedure with an elimination of instable variables. Compared with the crossvalidation criterion, the number of principal components is biased towards a lower number rendering this criterion very conservative for the selection of the principal components (see also section 6.1). Other methods use knowledge of the size of the measurement error for the estimation of the optimal number of principal components [34] or add artificial noise to the data and determine the optimal number of components by comparing with the original data in a bootstrapping procedure [35]. In this study, the Martens' Uncertainty Test and the minimum crossvalidation error criterion are used.

In mathematical terms, PLS can be described as follows: The matrixes X and Y of the independent and dependent variables are decomposed according to

$$X = TP^T + E \quad (8)$$

$$Y = UQ^T + F \quad (9)$$

with E and F as residual matrices, T and U as score matrixes and P^T and Q^T as loading matrixes. For the decomposition, either a singular value decomposition (SVD) or the non-linear iterative partial least squares (NIPALS) algorithm can be used [36-41]. A linear model is assumed to relate the score matrixes T and U (with H as residual matrix and B as diagonal matrix):

$$U = TB + H \quad (10)$$

The PLS1 algorithm models one variable y at a time, whereas the PLS2 algorithm can model several variables in one run. In this study, the PLS1 algorithm is used, as multiple PLS1 models often perform better than a single PLS2 model [41,42]. Further details of the algorithms can be found in [39,41,43-45].

2.6. Calibration of Nonlinear Relationships

Linear modeling of relationships, which is the basis for the most common methods in chemometrics, is based on many laws of physics showing linear relationships in a first approach. For example, in the field of analytical chemistry, the relationship between the concentration of an analyte and the absorption of radiation by this analyte is linear if certain conditions are fulfilled, known as Beer's Law. In the same way, the relationship between the concentration of an analyte in the gaseous phase and the amount of analyte sorbed into an amorphous polymer can be linear, known as Henry's Law. Yet, both laws of physics are borderline cases and hardly fulfilled in many real world applications. For example, Beer's Law is not valid for high concentrations of analytes, interfering analytes and turbid solutions ending up in a nonlinear relationship. Similarly, Henry's Law is the borderline case for small concentrations of analyte in the gaseous phase in contrast to the nonlinear Langmuir sorption for a higher range of the concentration of analyte. If the relationships between the input variables and the response variables are nonlinear, widely used linear calibration methods show a systematical bias. Although it was shown in [39] that under certain circumstances a linear PLS model can be successfully used if some variables show a nonlinear relationship, linear models fail in most nonlinear real world applications. Especially if all independent variables show similar nonlinear relationships with the dependent variables [41] or if variables show interactions, which are often observed in mixtures of analytes, linear models often fail.

Several approaches can be found in literature dealing with calibrations when nonlinearities in the data are present. Besides of the application of methods, which are in principle nonlinear like neural networks, there exist a couple of methods trying to remove the nonlinearities in the data or extending linear models to cope with the nonlinearities. The quadratic PLS (QPLS) belongs to the latter whereby a quadratic term for the inner relation is used instead of a linear term. By adding squared terms and interaction terms to the input variables, the implicit nonlinear PLS (INLR) tries to account for the nonlinear relationship. On the other hand, the Box-Cox transformation for the response variables tries to linearize the relationship of the data directly. The locally weighted regression (LWR) and the modeling trees use a piecewise linear approximation whereas the classification and regression trees (CART) and the GFI-PLS use discrete variables to approximate a nonlinear behavior. The methods, which are used in this study, are explained in detail in chapter 6 when applied to a nonlinear data set. As neural networks play a major part of this study, a detailed description of neural networks follows directly in the next section.

2.7. Neural Networks – Universal Calibration Tools

During the last decade, artificial neural networks have gained an increasing popularity in several fields of chemistry [46-49] whereby the variety of applications in chemistry is best illustrated in a book written by Zupan and Gasteiger [50]. In the field of multivariate calibration, the class of the multilayer feedforward backpropagation networks is most popular as they allow calibrating relationships, which are linear and nonlinear, and as no assumption of a specific type of model is needed [51-55]. In this section, the basics of the multilayer feedforward backpropagation neural networks are briefly explained and then the issues, which are of interest for this study, are introduced. A very detailed discussion of neural networks in multivariate calibration can be found in an excellent tutorial by Despagne and Massart [8]. More information about the mathematical background and about other neural network topologies can be found in textbooks [56-58].

2.7.1. Principles of Neural Networks

Neural networks (NN) are parallel information processing systems consisting of a number of simple neurons (also called nodes or units), which are organized in layers and which are connected by links. The artificial neural networks imitate the highly interconnected structures of the brain and the nervous system of animals and humans whereby the neurons correspond to the cell bodies and the links are equivalent to the axons in biology. There are a number of different types of NN, whereby only multilayer feedforward neural networks are used and discussed in this study. An example of a multilayer feedforward neural network for three input variables x_1, x_2, x_3 and one response variable y is shown in figure 1.

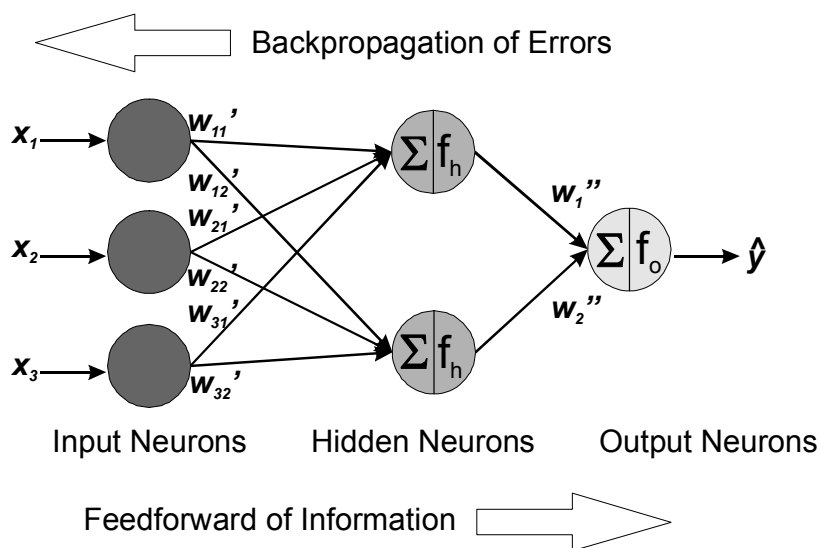


figure 1: Network elements of a multilayer feedforward backpropagation network.

The input variables are presented to the NN at the input units, which distribute the information by the connection links. Thereby the input variables are multiplied by the connection weights w'_{ij} between the input and hidden layer. The hidden neurons sum the weighted signals from the input neurons and then project this sum on an activation function f_h . The resulting activations of the hidden neurons are weighted by the connections w''_j between the hidden and output neurons and sent to the output neuron(s). The output neuron also performs a summation and projection on its activation function f_o . The output of this neuron is the estimated response \hat{y} . In the case of a single output neuron, the calculation of the estimated response can be summarized as:

$$\hat{y} = f_o \left[\Theta'' + \sum_{j=1}^{n_h} w''_j f_h \left(\sum_{i=1}^{n_v} w'_{ij} x_i + \Theta' \right) \right] \quad (11)$$

Thereby n_v and n_h are the number of input neurons and hidden neurons and Θ' and Θ'' are the biases of the hidden and output neurons, which shift the transfer functions horizontally. The weights w'_{ij} and w''_j and the biases Θ' and Θ'' are the adjustable parameters, which are determined by a learning algorithm during the calibration (often called training or learning) and which are assigned random values before the calibration. During the training, calibration samples with known response variables y (concentrations) are passed through the network. Then, the error between the predicted responses and the known experimental responses is calculated and used to adjust the parameters of the net in a backpropagation step to minimize the error. These two steps form an epoch (also called learn cycle or learn step) and are repeated until an acceptable low error is reached. The learning algorithm tries to find an acceptable minimum on the error surface, whereby in most cases the absolute minimum of the error surface is not found.

2.7.2. Topology of Neural Networks

All networks of this work are fully connected except of the non-uniform growing neural networks introduced in chapter 8. Fully connected means that a neuron is connected to all neurons of the proceeding layer. All networks except of the growing neural networks contain one layer of hidden neurons. If no special optimization technique is used, the number of hidden neurons is optimized by a gradient algorithm. Starting with 1 hidden neuron this algorithm adds fully connected neurons to the hidden layer until the error of prediction does not improve any more. For the hidden neurons, the hyperbolic tangent was used as activation function, which has some advantages referring to the convergence speed of learning in

contrast to other nonlinear functions [59]. The activation function of the output neurons is a linear function. The combination of linear and nonlinear activation functions allows an effective modeling of both, nonlinear and linear data sets.

In principle, neural networks can model several responses simultaneously. Therefore, it is possible either to use a neural network with as many output neurons as responses to model or to use a separate neural model with one single output neuron for each response. In congruence with Despagne et al. [8] and Moore et al. [60] several tests showed that for the calibration and prediction single networks with one output are superior in terms of lower errors of prediction. Thus, for all calibrations networks with single outputs are used if not stated differently. For the optimization of networks, like a variable selection, the choice of network type significantly influences the results as single output networks select variables, which are most predictive for one individual response whereas multi output networks select the variables, which model the ensemble of responses best. This issue is further discussed in section 10.1.

2.7.3. Training of Neural Networks

Many different learning algorithms exist for the training (calibration) of neural networks, which can be considered as an optimization problem. The first algorithm for the training of multilayer networks was proposed by Rumelhart and McClelland in 1986 [61]. This algorithm belongs to the gradient descent algorithms following the steepest descent of the error surface in the hyperspace of the adjustable parameters to find an acceptable minimum. To improve the excessively slow convergence of the algorithm for low gradients and to prevent trapping and oscillations in local minima a momentum term was introduced, which remembers the last change of the weights [62]. This algorithm, which smoothes the error surface, is also known as conjugate gradient descent. The introduction of individual learning rates and momentum terms for each weight significantly speeds up the algorithm and is known as delta bar algorithm [63] and SuperSAB algorithm [64]. One of the most modern algorithms of the gradient algorithms is the resilient propagation (Rprop) [65,66]. This algorithm combines the ideas of the algorithms described before. Yet, the weights are not adapted depending on the magnitude of the derivative but depending only on the sequence of the signs of the derivative. Consequently, the learning is spread equally over the complete network in contrast to all the other methods described before. Another advantage is the insensitivity of the algorithm to the different parameters [59], which can be all set to theses values proposed in [65,66] and implemented in [28,67]. The algorithm has been successfully used for several chemometric applications due to its speed [68,69].

Many modern learning algorithms with an amazing convergence speed belong to the second-order optimization methods, which utilize the Hessian matrix of partial second derivatives of the cost function, which describes how the error depends on the weights. The popular Levenberg-Marquardt algorithm [70] combines the gradient descent direction with the direction calculated from the estimated inverse Hessian matrix. The conjugate gradient algorithms also utilize the second-order information, but the estimation of the Hessian matrix is avoided. The scaled conjugate gradient algorithm (SCG) [71] makes use of the pseudo-second derivative, is insensitive to its parameters, which can be used as suggested in [71] and implemented in [65,66], and has proven to be a very efficient algorithm with respect to convergence speed and optimization quality [101]. Recently, genetic algorithms have been suggested for training NN [72,73]. As genetic algorithms are global optimization algorithms, they are more likely to find the global minimum of the error surface than gradient algorithms (see section 2.8.5 for more information about genetic algorithms). Yet, genetic learning algorithms are faced with several problems like long computing times, several parameters to be adjusted and troubles in fine-tuning the weights and biases. Combining genetic algorithms for the rough optimization and gradient algorithms for the fine-tuning has created a hybrid algorithm to overcome the last problem. Yet, the extremely high computing times render genetic algorithms for the training of neural networks unusable in practice, especially when several neural nets have to be trained for some kind of optimization (see chapters 7 and 8).

In this work, two learning algorithms are used. The Rprop algorithm is applied to the neural networks used in chapter 3. For all other networks SCG was used, as this algorithm shows a very fast initial convergence allowing the reduction in the number of training cycles during network optimization processes. All networks were trained with a maximum number of 2000 learning steps, whereby a method called early stopping was applied. This technique helps to anticipate the so-called overtraining effect [74]. An overtrained neural network learns a small calibration data set by heart. Thereby the noise in the data is learnt instead of generalizing the functional relationship of the data. For more details of overtraining, see section 2.8. Early stopping was implemented by monitoring the calibration data by a crossvalidation procedure (see section 2.4). The training is stopped when the error of crossvalidation of the calibration data starts going up, as the net may start losing its generalization ability at this moment. Early stopping is not an ultimate solution for preventing overtraining as a premature stopping of the training also stops the calibration of the functional relationship behind the data. Early stopping is only a tool, which should be used in combination with the network optimization procedures (see section 2.8.2) and becomes less important with more optimized networks.

2.8. Too Much Information Deteriorates Calibration

In chapter 5, a feature extraction is presented which generates many variables. The increase of information puts new challenges to multivariate calibration methods and especially to the data analysis by neural networks. Section 2.8.1 describes the fundamental relationships between the model complexity and the generalization and prediction abilities of the models, which are directly influenced by the number of independent variables utilized by the models. In section 2.8.2, the aspects, which influence the complexity of neural networks, are looked at. Then the different approaches for the optimization of the model complexity found in literature are discussed in the following sections.

2.8.1. Overfitting, Underfitting and Model Complexity

Neural networks are often referred to as universal function approximators since theoretically any continuous function can be approximated to a prescribed degree of accuracy by increasing the number of neurons in the hidden layer of a feedforward backpropagation network [75]. This can be proven by Kolmogorov's theorem stating that a neural network with linear combinations of $n(2n+1)$ monotonically increasing nonlinear functions of only one variable is able to fit any continuous function of n variables [76]. Yet in reality, the objective of a multivariate calibration is not to approximate a calibration data set with an ultimate accuracy, but to find a calibration with the best possible generalizing ability [77]. The gap between the approximation of a calibration data set and the generalization ability of a calibration becomes the more problematic the higher the number of variables and the smaller the data set, which will be further explained in the following sections.

The best measure for the generalizing ability is the error of prediction of as many independent separate validation data as possible. According to figure 2 the error of prediction is composed of two main contributions, the remaining interference error and the estimation error [39]. The interference error is the systematic error (bias) due to unmodeled interference in the data, as the calibration model is not complex enough to capture all the interferences of the relationship between sensor responses and analytes. The estimation error is caused by modeling measured random noise of various kinds. The optimal prediction is obtained, when the remaining interference error and the estimation error balance each other (arrow in figure 2). The effect of the prediction error increasing due to a too simple model is called underfitting whereas the effect of the increased prediction error due to a too complex model is called overfitting or overtraining. In figure 3 it is shown that the optimal complexity of the model highly depends

on the size and quality of the calibration data set. For data sets, which are noisy and limited in size, a simple calibration model is needed to prevent the overfitting. Neural networks, which are too complex (too big), are in danger of learning these data by heart and consequently model noise of the data. For big data sets, which contain only little noise, the best model is more complex resulting in an overall smaller prediction error for the same functional relationship. Consequently, for each data set an optimal model complexity has to be found [78] whereby the complexity of the models is directly related with the number of variables utilized by the model. The search of the optimal models is a very difficult task in the field of the multivariate calibration and is further discussed in section 2.8.2.

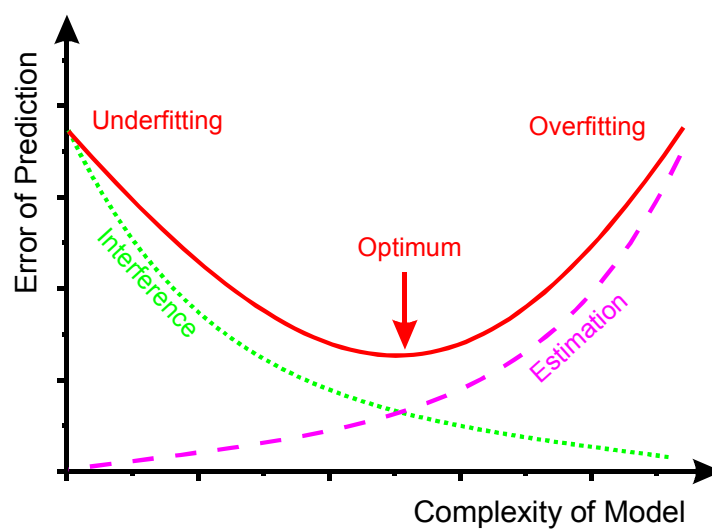


figure 2: Scheme for the error of prediction as a function of the complexity of the calibration model.

An overfitting can be detected, if the error of prediction of the independent validation data is significantly higher than the error of prediction of the calibration data whereby both data sets have to be within the same range of the response variables (for example within the same concentration range) to prevent additional biases due to extrapolation [79]. An underfitting manifests in high prediction errors for both data sets. Not only neural networks are affected by the effects of underfitting and overfitting, but also most modern multivariate calibration algorithms are subject to these effects [39]. In the following section, the discussion of the construction of optimal model complexities mainly refers to neural networks but can also be generalized for various multivariate calibration methods in many topics.

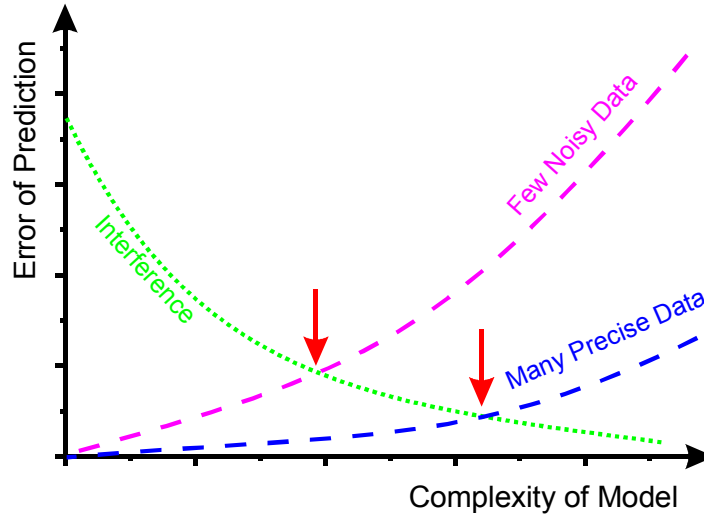


figure 3: Scheme for the error of prediction depending on the size and quality of the calibration data set, which influence the estimation error.

2.8.2. Neural Networks and the Complexity Problem

The complexity of neural networks can be mainly reduced to the number of adjustable parameters, namely the number of weights and the number of biases. Although the number of hidden layers also seems to influence the complexity of the neural networks on the first sight, the dimensionality of the optimization hyperspace and with it the complexity of the model is only defined by the number of adjustable parameters while the number of hidden layers influences only the direction of the optimization walk in the hyperspace. Consequently, the number of parameters can be ascribed to the sum of links, hidden neurons and output neurons. The number of adjustable parameters n can be calculated as:

$$n = n_h + n_o + n_l \quad (12)$$

Thereby n_l is the number of all links in the network, n_h is the number of hidden neurons and n_o is the number of output neurons. For uniform fully connected networks, the number of links can be replaced:

$$n = n_h (n_v + 1) + n_o (n_h + 1) \quad (13)$$

with n_v as the number of input variables (input neurons). Consequently, several strategies are possible for optimizing the model complexity and thus the neural networks [80]:

1. Optimizing the number of input variables n_v by a variable selection strategy.

2. Optimizing the number of hidden layers and the number of neurons in the hidden layers n_h .
3. Building networks with an arbitrary non-uniform structure optimized for the specific problem and thus directly optimizing the number of links n_l and hidden neurons n_h .

The first strategy is known as feature selection or variable selection [81] and has been proposed and applied in numerous publications. Thereby an optimal subset of variables is selected, avoiding an overfitting and the inclusion of too noisy and redundant variables, which could hide meaningful variables [68,82]. Additionally, the danger of chance correlation, which is further discussed in chapter 7, decreases with fewer variables [83]. Many variable selection strategies have been proposed and compared [84] such as the two stepwise algorithms forward addition and backward elimination [85-89], such as orthogonal descriptors and successive projection algorithms for linear calibration models [90], and such as simulated annealing [91-96] and genetic algorithms [10,93-103], which contain random search steps. Especially the last two approaches have been demonstrated to be superior in the case of many variables [93-96,104]. A compression of the input variables by the use of a principal component analysis can also help to reduce the number of parameters [105-109]. The most important approaches are discussed in detail in the sections 2.8.3 to 2.8.10.

The second strategy optimizes the inner topology of the neural networks by varying the number of hidden layers and number of neurons in the hidden layers, whereby the networks are fully connected. In most publications dealing with neural networks, this task is performed by a trial and error procedure and relies on the intuition of the user. Yet, there are two non-manual approaches for finding an optimal inner topology of neural networks. The first approach is based on the use of genetic algorithms for finding the optimal number of hidden layers and hidden neurons whereby the number of neurons of each layer is directly coded into the genetic string [110]. This approach is usually combined with a variable selection based on genetic algorithms [111,112]. The second approach is based on neural networks with growing hidden layers. Thereby full-connected hidden neurons are added to a single hidden layer [96,113-116] or to a prespecified location (known as cascade correlation architecture [117]) until the error of prediction does not improve any more. All approaches based on the second strategy are faced by 2 major drawbacks: Only fully connected neurons are added to the network resulting in a simultaneous addition of many parameters per addition step and the place where to add the neurons is not optimized by any of the algorithms.

The third strategy, which is also known as topology optimization, should be superior to the first two strategies since both, the number of input variables and the number and location of the hidden neurons are simultaneously optimized resulting in a synergetic effect. Additionally, the presence of each single possible link and with it each single parameter is decided on when building non-uniform networks. Thus, the number of adjustable parameters is effectively reduced, allowing the iterative calibration procedure of neural networks to find a better solution [12] than fully connected networks. Three general approaches can be found in literature following this strategy. The first approach is based on a structure generation and evaluation by the use of genetic algorithms [118-121]. In the second approach unimportant links of a big network are removed by so-called pruning techniques [59,109,122-124]. The third approach is based on growing non-uniform neural networks proposed by Vinod et al. [125]. In the next sections, the different methods for selecting and compressing variables are discussed and then the three approaches for the optimization of the network topology are looked at in more detail.

2.8.3. Brute Force Variable Selection

The most obvious method of selecting a subset of variables is the examination of all combinations of variables. Thereby a subset of variables is selected, a neural network utilizing only these variables is calibrated, and the error of prediction of an independent test data set is calculated. Finally, the combination with the smallest error of prediction is chosen. Besides of some problems due to the random weight initialization of the networks and the limitation of the size of the data set, this so-called brute force variable selection is the most accurate approach. However, this approach is only feasible for a very limited number of variables, as the number of variable subsets increases dramatically with the number of variables.

For a fixed number n_v of variables to be selected from n_{tot} variables in total, the number n of different variable subsets can be calculated as [12,126,127]:

$$n = \frac{n_{tot}!}{n_v!(n_{tot} - n_v)!} \quad (14)$$

In the common case, when an optimal solution is searched, the number of variables to select is not fixed resulting in even more possible combinations n of variable subsets:

$$n = \sum_{n_v=1}^{n_{tot}} \frac{n_{tot}!}{n_v!(n_{tot}-n_v)!} = 2^{n_{tot}} - 1 \quad (15)$$

For example 40 variables (refrigerant data introduced in section 4.5.1.1) result in 1 099 511 627 775 different combinations to be examined. If a fast up-to-date computer needs 1 minute for the training of a neural net (the time needed for the prediction can be neglected) the examination of all possible combinations needs 2 090 540 years computing time rendering the brute force variable selection useless for this work.

2.8.4. Variable Selection by Stepwise Algorithms

The two stepwise algorithms forward addition (forward selection) and backward elimination (backward selection) are also sometimes called gradient methods as the next addition or elimination step is performed on the basis of the steepest gradient of the error surface. The forward selection begins by selecting the variable, which results in the lowest error of prediction. In the next step, the variable out of the remaining variables is added, which minimizes the error in combination with the first variable. The stepwise addition of further variables is repeated until an optimal subset is found with a maximum of n_{tot} steps. The backward elimination works in the opposite direction by starting with all variables and eliminating single variables. In addition, combinations of both methods are known as stepwise multiple regressions [12]. Yet, the stepwise algorithms fail to take the information into account that involves the combined effect of several variables. Thus, these algorithms hardly find an optimal solution, which requires several independent variables to be selected [25,128]. The stepwise algorithms walk during the minimum search in the valleys of the error surface and cannot find minima surrounded by high mountains. In figure 4, the error surface of the selection of 2 variables out of 40 is shown for the refrigerant data introduced in section 4.5.1.1. Even in this figure, which represents only a highly constrained 2-dimensional lateral surface of the 40-dimensional error surface, it is visible that the error surface is too rough for the stepwise algorithms finding an optimal solution and not usable for high dimensional data sets with many correlated variables.

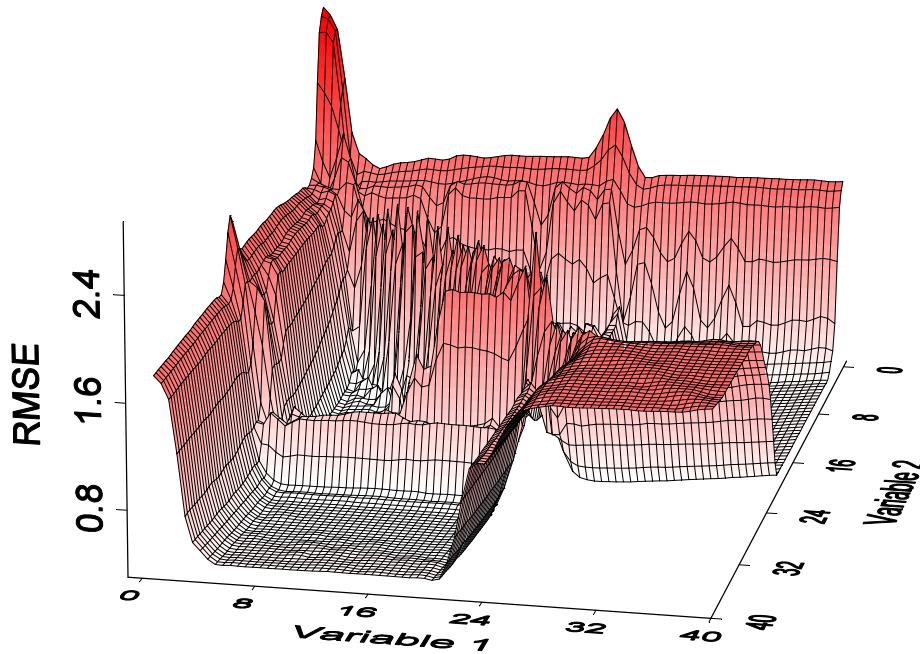


figure 4: Root mean square of prediction versus the index of the 2 variables selected.

2.8.5. Variable Selection by Genetic Algorithms

Genetic algorithms (GA) have become a popular optimization method as they often succeed in finding the best optimum in contrast to most common optimization algorithms. Genetic algorithms imitate the natural selection process in biological evolution with selection, mating reproduction and mutation. On the left-hand side of figure 5, the sequence of the different operations of a genetic algorithm is shown.

The parameters to be optimized are represented by a chromosome whereby each parameter is encoded in a binary string called gene. Thus, a chromosome consists of as many genes as parameters to be optimized. A population, which consists of a given number of chromosomes, is initially created by randomly assigning "1" and "0" to all genes. On the top right part of figure 5, the different terms are graphically shown for a population of 4 chromosomes with 4 genes (in the case of variable selection a gene contains only a single bit string for the presence and absence of a variable). The best chromosomes have the highest probability to survive evaluated by a so-called fitness function. The next generation is reproduced by selecting the best chromosomes, mating the chromosomes to produce an offspring population and by an occasional mutation. The evaluation and reproduction steps are repeated until a certain number of generations, until a defined fitness or until a convergence criterion of the population are reached. In the ideal case, all chromosomes of the last generation have the same genes representing the optimal solution. The theory and benefits of GA in variable

selection have been described several times in literature [98-103] and will not be repeated here, as there are uncountable variants of the different genetic operators. Instead, a description of the GA implementation, which has been used in this work and its special features like the implementation of the fitness function and the other genetic operators will be further discussed and are illustrated on the right side of figure 5.

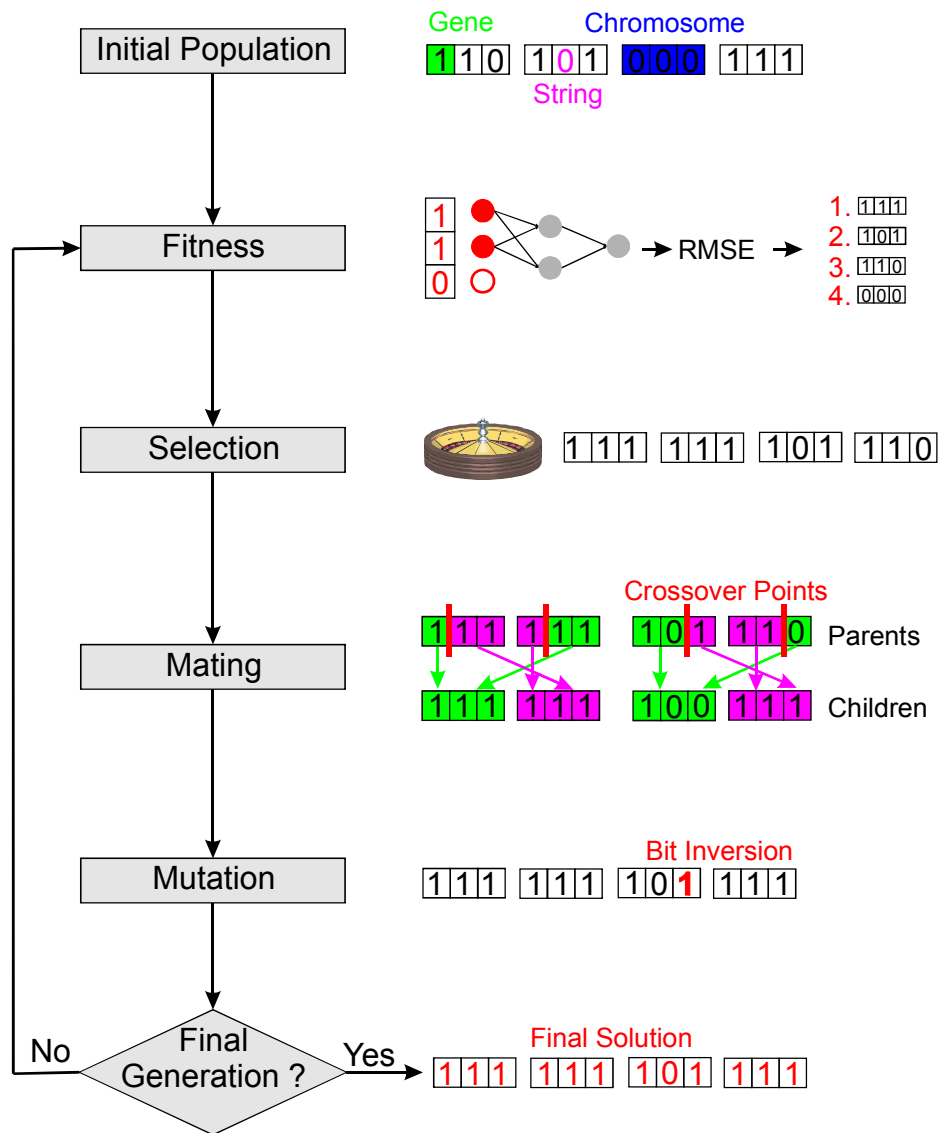


figure 5: Flow chart of a genetic algorithm (left side) and explanations of the different operations by the application of the genetic algorithm to a variable selection problem with 3 variables and neural networks for the evaluation of the fitness.

In this work, the initial population of the GA is randomly generated except of one chromosome, which was set to use all variables. The binary string of the chromosomes has the same size as variables to select from whereby the presence of a variable is coded as "1" and the absence of a variable as "0". Consequently, the binary string of a gene consists of only one single bit. After evolving the fitness of the population, the individuals are selected by

means of the roulette wheel. Thereby, the chromosomes are allocated space on a roulette wheel proportional to their fitness and thus the fittest individuals are more likely selected. In the following mating step, offspring chromosomes are created by a crossover technique. A so-called one-point crossover technique is employed, which randomly selects a crossover point within the chromosome. Then two parent chromosomes are interchanged at this point to produce two new offspring. After that, the chromosomes are mutated with a probability of 0.005 per gene by randomly changing genes from "0" to "1" and vice versa. The mutation prevents the GA from converging too quickly in a small area of the search space.

A crucial point in using GA is the design of the fitness function, which determines what a GA should optimize. In the case of a variable selection for calibration, the goal is to find a small subset of variables, which are most significant for a regression. In this work, the calibration is based on neural networks for modeling the relationship between the input variables (many time-dependent sensor signals) and the responses (concentrations of the different analytes). Thus, the evaluation of the fitness starts with the encoding of the chromosomes into neural networks whereby "1" indicates that a specific variable is used and "0" that a variable is not used by the network. Then the networks are trained with a calibration data set and after that, a test data set is predicted. Finally, the fitness is calculated by a so-called fitness function f . In contrast to many GA for variable selection found in literature [99,101,129-131,243], the fitness function used for the GA variable selections in this work takes into account not only the prediction error of test data but also partially the calibration error and primarily the number of variables used to build the corresponding neural nets:

$$f = 0.2 \cdot \text{MRMSE}_{\text{cal}} + 0.8 \cdot \text{MRMSE}_{\text{test}} - \alpha \left(1 - \frac{n_v}{n_{\text{tot}}} \right) \quad (16)$$

Thereby n_v is the number of variables used by the neural networks, n_{tot} is the total number of variables and MRMSE is the mean root mean square error of the calibration respectively test data. The MRMSE is defined in equation (17) with N as total number of samples predicted, M as total number of analytes, $\hat{y}_{i,j}$ as predicted concentration of analyte j in sample i and $y_{i,j}$ as the corresponding known true concentration:

$$\text{MRMSE} = \sum_{k=1}^{20} \frac{\sum_{j=1}^M \sqrt{\frac{\sum_{i=1}^N (\hat{y}_{i,j} - y_{i,j})^2}{N}}}{M} \quad (17)$$

The fitness function (16) can be broken up into three parts. The first two parts correspond to the accuracy of the neural networks. Thereby $\text{MRMSE}_{\text{cal}}$ is based on the prediction of the calibration data used to build the neural nets, whereas $\text{MRMSE}_{\text{test}}$ is based on the prediction of separate test data not used for training the neural networks. It was demonstrated in [11] that using the same data for the variable selection and for the model calibration introduces a bias. Thus, variables are selected based on data poorly representing the true relationship. On the other hand, it was also shown [11,132] that a variable selection based on a small data set is unlikely to find an optimal subset of variables. Therefore, a ratio of 1:4 between the influence of calibration and test data was chosen. Although being partly arbitrary this ratio should give as little influence to the calibration data as to bias the feature selection yet taking the samples of the larger calibration set partly into account. The third part of the fitness function rewards small networks using only few variables by an amount proportional to the parameter α . The choice of α influences the number of variables used by the evolved neural nets. A high value of α results in only few variables selected for each GA whereas a small value of α results in more variables being selected.

As the initial weights of a neural network are randomly set, the network finds another local minimum of the error surface for each calibration run with a slightly different performance of prediction. In order to reduce the variance of the error of prediction due to random weight initialization the fitness is averaged in expression (17) over 20 training and prediction sessions per network topology (evaluation of 20 parallel neural networks with different initial weights).

2.8.6. Variable Selection by Simulated Annealing

Simulated Annealing (SA) is a method that simulates the thermodynamic process in which metal is heated to its melting temperature and then slowly cooled to its crystal configuration of lowest energy. The system is in thermal equilibrium when the probability of a certain state is governed by a Boltzmann distribution:

$$p(\Delta E) = e^{\frac{-\Delta E}{kT}} \quad (18)$$

with E as energy, T as temperature and k as Boltzmann's constant. Kirkpatrick et al. [133] applied SA to an optimization problem. During the minimization of a multivariate function, a candidate solution is generated by randomly perturbing the current configuration and the energy (similar to the fitness of GA) is calculated. If the new energy is lower than the current,

the displacement is accepted. If the energy is higher, the displacement is accepted with a probability given by the Boltzmann distribution (18). These uphill steps allow the algorithm to escape from local minima. The probability of accepting an uphill step is a function of the change of the energy and of the temperature, which is gradually lowered during the search process. Due to the similar approach with random steps in the search process, SA has several times been compared with GA for the selection of variables whereby SA achieved comparable or slightly worse results and consequently will not be used in this study [91-96].

2.8.7. Variable Compression by Principal Component Analysis

The Principal Component Analysis (PCA), which originates from psychometrics, can be used as preprocessing tool for neural networks. Thereby the PCA compresses the independent variables into fewer principal components, which are then used as new input variables for the neural networks. The PCA finds the direction in space along which the variance of the data is the largest. This direction is called the first principal component. The second principal component is the direction in space orthogonal to the first principal component, which describes maximum variance not covered by the first principal component, and so on. The data matrix is decomposed by the PCA into a product of a loading matrix P^T and of a score matrix T and a matrix containing the residuals E :

$$X = TP^T + E \quad (19)$$

Similar to the PLS only the first few principal components are used with similar methods to determine the optimal number (see section 2.5).

Yet, the variable compression by principal components is affected by some (at least theoretical) drawbacks. Using only few principal components does not ensure that the information preserved in these components is useful for the calibration of the relationship of interest. For example, if noise dominates the variations of the input variables, the variations caused by the sensor responses due to the analytes might not be included as the corresponding principal components with small singular values are discarded [107]. Additionally, nonlinear relationships are often spread over many principal components, which might not be included in the model (see also discussions in the sections 6.1 and 9.2.5). As in contrast to the PLS the principal components are determined only on the basis of the variances of the independent variables and not on the basis of an optimal regression, no synergetic effects of the combination of the PCA and the neural networks can be expected.

2.8.8. Topology Optimization by Pruning Algorithms

The oldest and very popular approach to topology optimization of neural networks is based on destructive pruning algorithms [59,109,122-124]. These algorithms start with a trained fully connected neural network, remove certain network elements and then retrain the network. The steps of removing and retraining are repeated until a certain stopping criterion is fulfilled. Besides of the algorithm "Skeletonization" [134], only links are considered as candidates to be removed. Neurons are automatically removed, if all corresponding input or output links have been removed. In this work the two pruning algorithms "Optimal Brain Surgeon" and "Magnitude Based Pruning" are used, which were identified as most promising pruning methods for sensor data in a comparison [28]. A combination of both algorithms can be found in [153].

Magnitude Based Pruning

The Magnitude Based Pruning (MP) is the simplest approach among the pruning algorithms, which removes the links with the smallest weights. This approach originates in the linear regression, as the variables with the smallest regression coefficients of a linear regression are least important. Yet, Despagne et al. demonstrated that for data sets with both, linear and nonlinear relationships, this assumption is not valid as primarily links of linear variables are removed [135]. On the other hand, it was shown in [28,136] that this simple and very fast algorithm performs as well as the other algorithms as long as not too many links are removed.

Optimal Brain Surgeon

The Optimal Brain Surgeon (OBS) algorithm [137] estimates the increase of the training error when a link is deleted. Then the links with the smallest increase, which is also called saliency, are deleted. To determine the saliencies, the changes of the cost function ∂E associated with a perturbation of the weight matrix W has to be evaluated. The term ∂E is approximated by a second-order Taylor series expansion:

$$\partial E = \sum_i \frac{\partial E}{\partial w_i} \Delta w_i + \frac{1}{2} \sum_i \frac{\partial^2 E}{\partial w_i^2} \Delta w_i^2 + \frac{1}{2} \sum_{i \neq j} \frac{\partial^2 E}{\partial w_i \partial w_j} \Delta w_i \Delta w_j \quad (20)$$

As the training is stopped when the neural network has converged to a minimum, the first term can be considered as being zero. The second and third term are the so-called Hessian Matrix H , whereby the Optimal Brain Damage algorithm [138] neglects the third term resulting in faster yet less reliable pruning results [28]. The saliency of a weight k is then calculated as

$$S_k = \frac{w_k^2}{[H^{-1}]_{kk}} \quad (21)$$

Discussion

Although the pruning algorithms have been applied in several fields of chemistry and analytics [55,109,139-141, 153] the general application of these algorithms is controversially discussed in literature. In congruence with the theory that pruned sparse neural networks should generalize better, improved predictions were found in references [28,142,143] whereas a deterioration of the prediction ability was found in references [136,144]. Another problem of the pruning algorithms is the need of a reference network for the removal of the network elements. This reference network has to be designed by the user and the final optimal network found by the algorithm cannot be more complex than this reference network. This means that the "solution" is biased by the choice of the user for the reference network as the pruning algorithms are highly sensitive to local minima of the error surface ending up in completely different network topologies and different variables selected for different reference networks.

2.8.9. Topology Optimization by Genetic Algorithms

The application of genetic algorithms for evolving neural networks is not only limited to the selection of input variables but can also be used for optimizing the complete topology of the neural networks [118-121]. The different approaches found in literature can be classified according to their encoding mechanism as direct and indirect methods. In the direct methods, all the information about the structure is directly represented in the chromosome. The most common way is the representation of the connections in a matrix (connectivity matrix) and then linking this matrix row by row into the chromosome. The indirect encoding methods are also called grammatically encoding methods as the chromosome contains development rules, which have to be interpreted to build the corresponding net. This allows a compression of the topology resulting in a smaller length of the chromosome and thus a better scalability.

A promising approach of genetic algorithms for the optimization of neural network topologies was proposed by Boozarjohery et al. [145] using a grammatically encoding procedure. This algorithm is also applied to real world data of a neutralization process besides of simple benchmark problems. Yet, a problem of this approach is that the solutions found by the algorithm are randomly depending on the initial weight initialization and on the parameters. Another very complex approach for evolving neural networks by genetic algorithms using a direct encoding was proposed by Braun et al. [146,147]. The corresponding software ENZO is

available for free [148] and has been applied to several real world problems [149-151]. The problem of this approach is its complexity with more than 100 parameters, which can be adjusted by the user. Although the default settings work well in many cases, the excellent results demonstrated in the references mentioned before need adjustments of these parameters rendering a general application of this approach with only little input by the analyst virtually impossible.

In general all approaches of optimizing the topology of neural networks by genetic algorithms are faced by a poor scalability [152] and by complex genetic operations [147]. An example is the structural mapping causing problems to the crossover operator. For two networks with an identical topology, the contributions of the hidden neurons to the overall solution may be internally permuted (only visible by a permutation of the weights). If a crossover operator is applied to these networks, one offspring is created with partly doubled internal contributions and one offspring is created with partly missing internal contributions. In most cases, the optimization of the neural network topology has been used only for simple benchmarks like the XOR problem [152]. Another general problem is that similar to the pruning algorithms the networks cannot be bigger and more complex than a largest possible reference network predefined by the user (see also section 2.8.8). Due to these quite complex problems, the genetic algorithms are used only for a variable selection and not for the optimization of the topology in this work.

2.8.10. Topology Optimization by Growing Neural Network Algorithms

An interesting algorithm for building non-uniform optimized neural network topologies was initially proposed by Vinod et al. [125]. The algorithm starts with a feedforward backpropagation neural network, which has no hidden neurons and no links ("empty network"). The algorithm grows the network by adding one neuron at a time. The neuron is connected to one output neuron and to two other neurons whereby these links are selected on the basis of the maximum estimated error decrease for the calibration data. The insertion of the neurons is stopped when a prescribed error has been reached. It was demonstrated that each growing step does improve the calibration error. It was also shown that the algorithm is able to approximate complex continuous functions (like a sinus wave) using very small networks.

2.9. Measures of Error and Validation

Besides of the true-predicted plots, which will be introduced in section 3.4, the root mean square errors (RMSE) are used for the validation of the models in this work as it is one of the most common measures for the quality of calibrations and predictions in chemometrics:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \quad (22)$$

Thereby \hat{y}_i is the predicted concentration of the sample, y_i is the true concentration of the sample and N is the total number of samples. The RMSE, which has the dimension and units of the concentrations predicted, is a strict measure of the error as it penalizes poor predictions by a quadratic term. For a relative measure of the error, the relative RMSE, which is also sometimes misnamed standard error of prediction (SEP) [41,153,154], is used in this work:

$$\text{RMSE}_{\text{rel}} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}} \bigg/ \frac{\sum_{i=1}^n y_i}{n} \quad (23)$$

3. Theory – Quantification of the Refrigerants R22 and R134a: Part I

In this chapter, an example of a multivariate calibration in chemical sensing is shown whereby the focus of this example is the demonstration of the methods, which are widely accepted and which can be found in literature over and over again. This allows an easy comparison with the new approaches proposed in this study, which are going far beyond the widespread techniques in the areas of multivariate calibration and measurement principles. Furthermore, this data set will be examined later again (section 9.5) using the new approaches proposed in this work. Although the recording of the data set was not optimized for these approaches, the new methods of data analysis show better results. Additionally some concepts and theories of chemical sensing of vapors by polymer-based sensors are introduced in this chapter.

The objective of this example is the quantitative detection of the ozone depleting R22 (chlorodifluoromethan) in the vapor of its harmless substitute R134a (1,1,1,2-tetrafluoroethan) and in air for preliminary studies of on-line measurements in recycling stations. More details of the environmental background of these refrigerants can be found in section 4.5.1. First, the sorption characteristics of 6 different polymers, which are exposed to different concentrations of the refrigerants R22 and R134a, are investigated with a sensor array setup in respect to sensitivities, sensitivity patterns, and calibration curves. Based on these investigations two polymers are selected for the application in a miniaturized low-cost 4λ sensor setup, which complies best with the conditions for on-site measurements at recycling stations. Finally, different binary mixtures of R22 and R134a are measured by both setups and a multivariate calibration is performed by the use of neural networks.

3.1. *Experimental*

Both setups, which are based on the reflectometric interference spectroscopy, are described in detail in section 4.3 (Array set-up) and in section 4.4 (4λ setup). For the sensor array setup, 6 sensitive polymer layers were prepared using the polymers Polyetherurethane (PUT), Polydimethylsiloxane (PDMS), a hyperbranched polyester (HBP), Ultrason (UE 2010) and Makrolon (M 2400). Besides of measurements of single analyte vapors for a sensitivity analysis, two data sets of binary mixtures were measured based on an equidistant 6-level full factorial design [155]. Thereby the relative saturation pressures and thus the concentrations of

the analytes R22 and R134a were varied between 0 and 0.1 with synthetic air as ambient gas. The first data set was generated by measuring the experimental design 4 times with the sensor array RfS setup and the second data set was produced by measuring the experimental design twice with the miniaturized 4λ RfS setup. The sensor signals were recorded after 10 minutes of exposure to analyte and a recovery time of 2 hours was chosen.

A 20-fold random subsampling procedure described in section 2.4 was used for splitting the data into a calibration data set (75%) and a test data set (25%) with the confinement that all repeated measurements of a concentration combination went into one subset to prevent overoptimistic predictions [156]. The neural networks implemented for this example had a topology of 1 output neuron, 4 neurons in 1 hidden layer and 6 respectively 2 input neurons with all features and parameters described in section 2.7.3 in detail.

3.2. Single Analytes

The first step was the characterization of the polymers in respect to the sensor responses for the analytes. The responses of the 6 sensors coated with the different polymers were simultaneously measured by the array setup. The sensor responses of the polymers PUT, PDMS and HBP are shown on the left side of figure 6. These 3 polymers contain polar groups and should therefore show different response characteristics for analytes with different polarities or polarizabilities [157]. The sensor responses of the polymers UE 2010 and M 2400 are shown on the left side of figure 7. Both polymers are amorphous glassy polymers with a microporous structure whereby the mean size of the pores of M 2400 is 0.1 nm^3 [158] and the mean size of the pores of UE 2010 is 0.08 nm^3 [159]. These polymers can discriminate analytes due to different sizes of the analytes as only analytes with a smaller volume of the molecules than the volume of the pores sorb into the pores of the polymers. Further discussions regarding the pores can be found in chapter 5.

According to both figures, all polymers show a fast and reversible swelling when exposed to R22 whereby the polar polymers PUT and PDMS are reaching an equilibrium state instantly. These two polymers were measured above their static glass transition temperature. The glass transition temperature is the temperature, above which the molecules in the polymer backbone can move relatively to one another resulting in a quasi-liquid state [160,161]. Thus, the interactions between vapor and coating can be described as dissolution of a solute vapor in a solvent coating resulting in a very fast sorption and desorption of the analytes, which can be

modeled by linear solvation energy relationships (LSERs) [162]-[164]. Consequently, these two polymers show also an immediate sorption and desorption of R134a. On the other hand, exposed to R134a the microporous polymers UE 2010 and M 2400 do not reach an equilibrium state within 30 minutes and the signals need 2 hours to return to the baseline. Due to the bigger volume of the molecules of R134a, the sorption process is kinetically inhibited and the molecules are less (and more slowly) sorbed into the polymers.

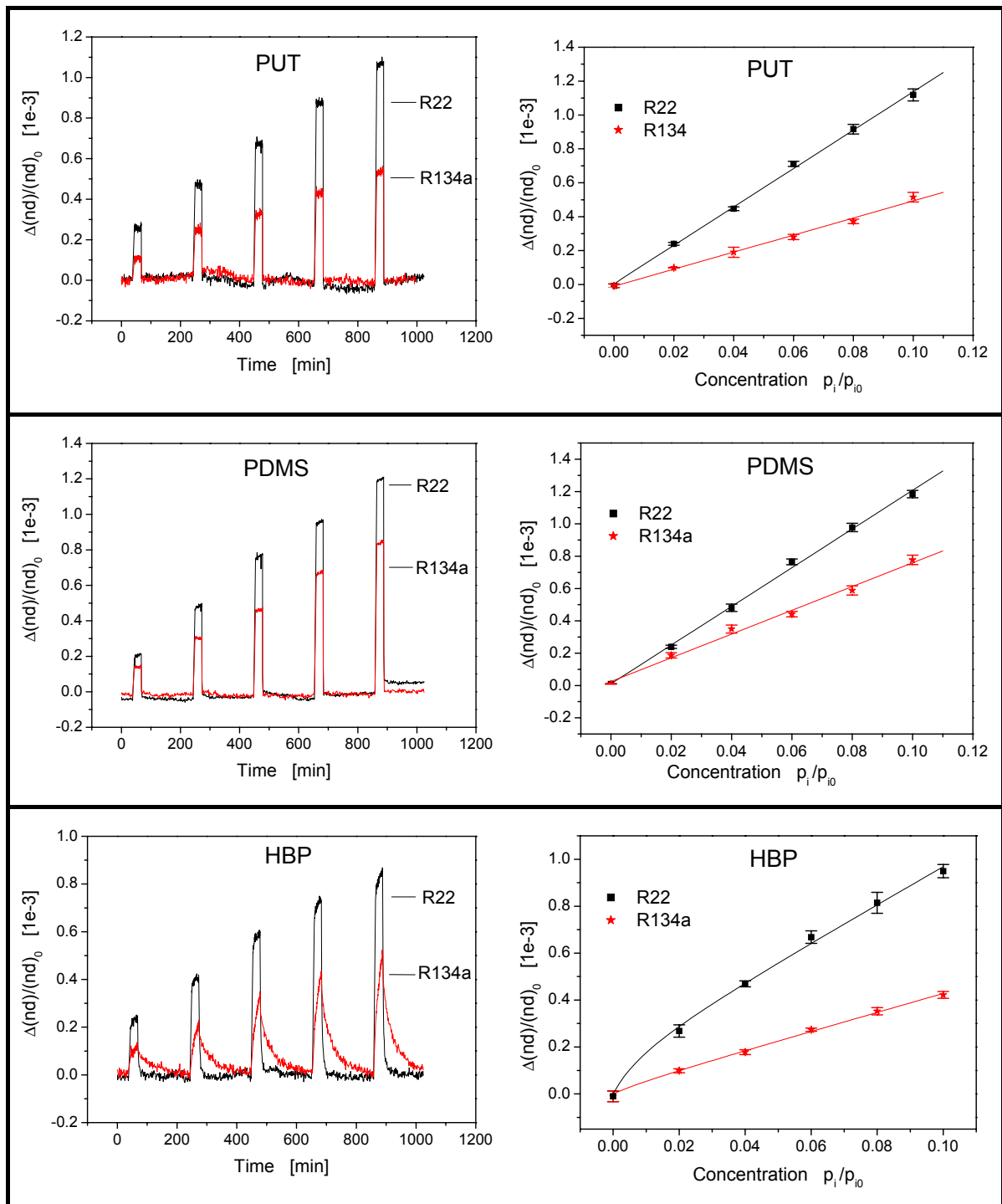


figure 6: Sensor responses, calibration curves and standard deviations of 3 measurements of the polar polymers recorded with the array setup.

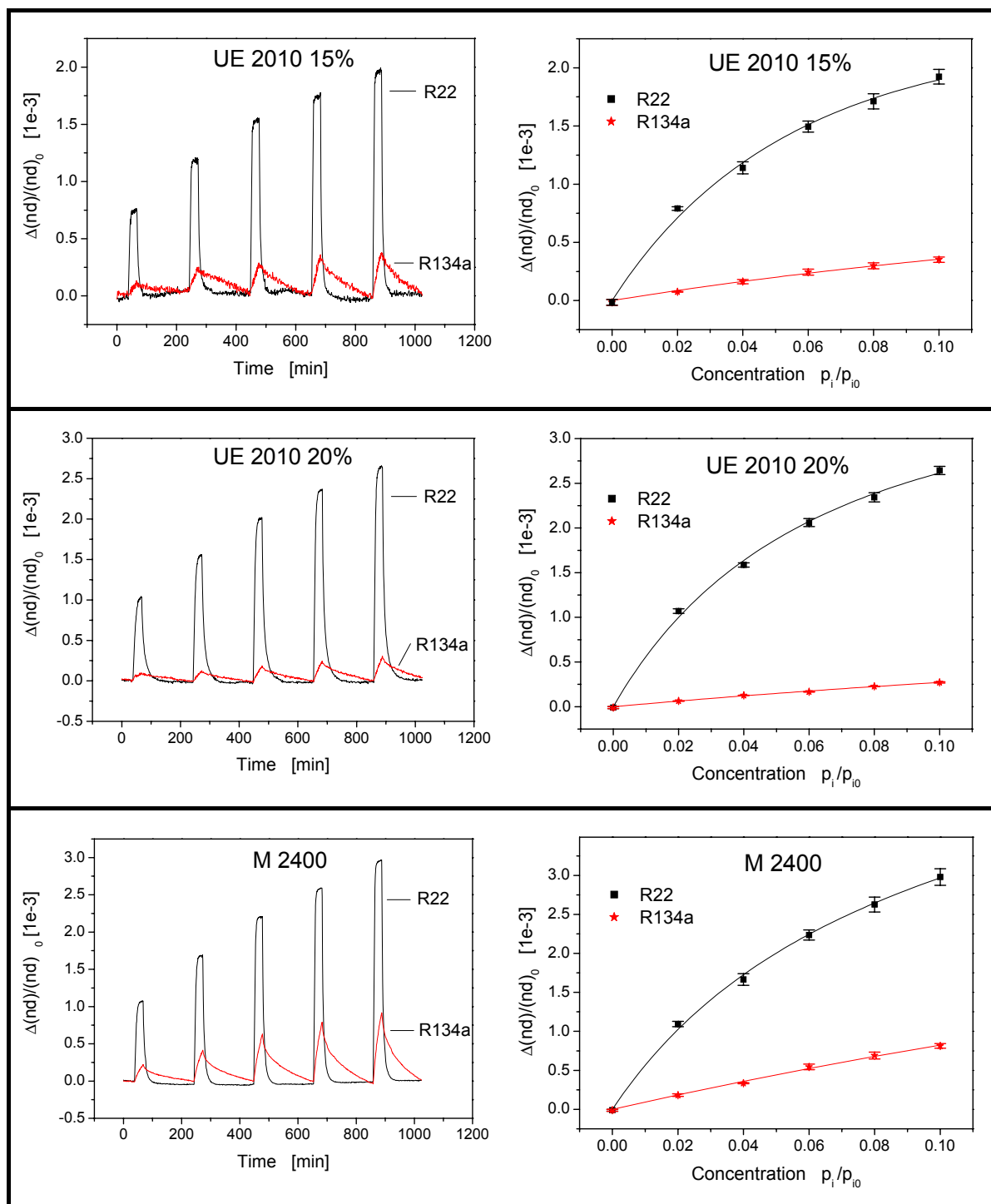


figure 7: Sensor responses, calibration curves and standard deviations of 3 measurements of the microporous polymers recorded with the array setup.

The right sides of figure 6 and figure 7 show the signals of the 6 sensors versus the concentrations of the 2 analytes. This type of plots is often referred to as calibration curve. The correlation of the polymer swellings (the sensor signals) with the analyte concentrations is often described as a Henry sorption [165], a Langmuir sorption [166] or a combination of both types [167]-[169]. For all polymers under investigation, the sorption of R134a is best

described as a linear Henry type sorption. This type of sorption process is an indication of an unspecific sorption process [170]. The molecules of R134a are too big for the micropores and therefore only an unspecific sorption process into the polymer matrix of the microporous polymer can be observed. The sorption causes a swelling of the polymer matrix, which can be observed as an increase of the thickness of the sensitive layer. The sorption of R134a and of R22 into the two polar polymers is also of the Henry type, since both analytes do not have distinctive polar groups and thus do not specifically interact with the polymers.

On the other hand, the sorption of R22 into the microporous polymers UE 2010 and M 2400 shows a calibration curve, which can be best described by the combination of the Henry and Langmuir sorption. The Langmuir type sorption can be found, if there is a specific interaction and if the amount of sorption and interaction sites is limited [170]. The combination of both sorption types can be best detected when examining the curve for high analyte concentrations, as the Henry sorption and the Langmuir sorption are identical for small concentrations. If the sorption is a pure Langmuir sorption, the calibration curve should pass into saturation for high concentrations whereas the combination of both types of sorption results in a curve with a positive slope for high concentrations. In figure 7, both polymers show for R22 this latter case of a Langmuir sorption with a small portion of a Henry sorption. Furthermore, the figures demonstrate that the microporous polymers show a much higher slope for R22 than for R134a. Both findings can be explained by a sorption of the small molecules of R22 into the micropores. This results in higher signals of the R22 sorption, as the unspecific Henry sorption into the polymer matrix (which is also present for R134a) is overlaid by a specific Langmuir sorption into the pores (which is not present for the bigger R134a molecules). The number of pores is limited and consequently the sorption of the molecules into the pores and with it the Langmuir part of the sorption reaches saturation for higher concentrations.

3.3. Sensitivities

For this study, primarily the differences of the sensitivities of the polymers for the two analytes are of interest to select the two most suitable polymers for the low-cost 4λ setup. The sensitivities can be calculated as the slopes of the calibration curves. For the Henry sorption, a constant sensitivity can be easily specified for the examined concentration range, whereas the sensitivity for the Langmuir sorption is a function of the concentrations. Thus, the sensitivities for the sorption of R22 in the polymers UE 2010, Makrolon, and HBP were calculated using the derivative of the calibration function at zero concentration.

In figure 8 the sensitivities for both analytes and all 6 sensors are shown. As already mentioned the microporous polymers show a very different swelling for the 2 analytes whereas the sorption process of both analytes into the polar polymers is very similar. Consequently, the sensitivities of M 2400 and UE 2010 for both analytes are very different whereas the sensitivities of PUT and PDMS for both analytes are similar. For a quantification of binary mixtures with only 2 sensors, 2 polymers should be chosen, which show the most possible different sensitivity patterns for the 2 analytes (with the extreme of 2 selective sensors). Consequently, for the quantification of the 2 analytes, 1 microporous polymer and 1 polar polymer should be chosen. The polymers UE 2010 and PDMS, which show the most different sensitivity patterns, would be the best choice. Due to the technical limitation of the thickness and consistence of the layers [178], the combination UE 2010 / PUT was chosen for the 4λ setup.

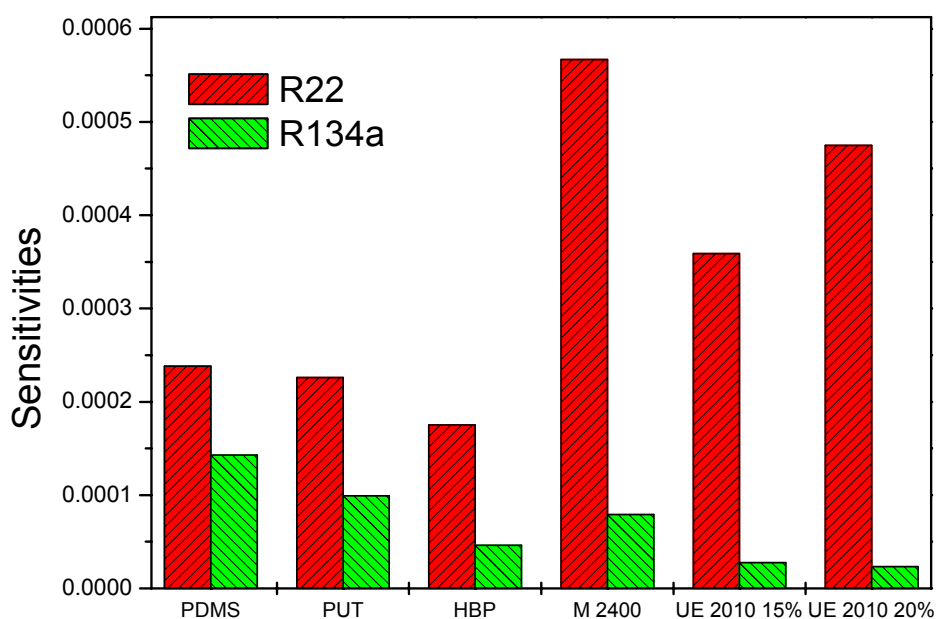


figure 8: Sensitivities of the different polymers.

3.4. Calibrations of the Mixtures

For the quantification of binary mixtures, neural networks were trained by the calibration data and subsequently predicted the independent test data. The root mean square errors (RMSE) of the test data and the crossvalidated calibration data measured with the 6-sensor array setup are listed in the first row of table 1. The RMSE of the prediction of the R134a concentrations is about three times higher than the RMSE of R22, whereby the relative RMSE of the test data are 3.5% for R22 and 13.0% for R134a. The RMSE of the test data of both analytes are only slightly higher than the RMSE of the crossvalidation of the calibration data indicating good calibration models [79]. The predictions of the test data are graphically shown in the true-predicted plots in figure 9. Thereby the predicted concentrations of the test data are plotted versus the true known concentrations. Ideal predictions are consequently located on the diagonal. As the single predictions cannot be graphically resolved, the predictions of each concentration level are represented by the mean and the standard deviation. It is visible that the predictions of both analytes are not biased as the means of all concentration levels do not significantly deviate from the diagonal. The smaller standard deviations of the plot of R22 show that the predictions of the concentrations of R22 are more precise than the predictions of R134a. Both, the RMSE and the true predicted plots allow the conclusion that both, R22 and R134a, can be quantitatively determined in binary mixtures using the RIFS array setup. Especially the more important detection of the harmful R22 has proven to be very accurate over the whole concentration range.

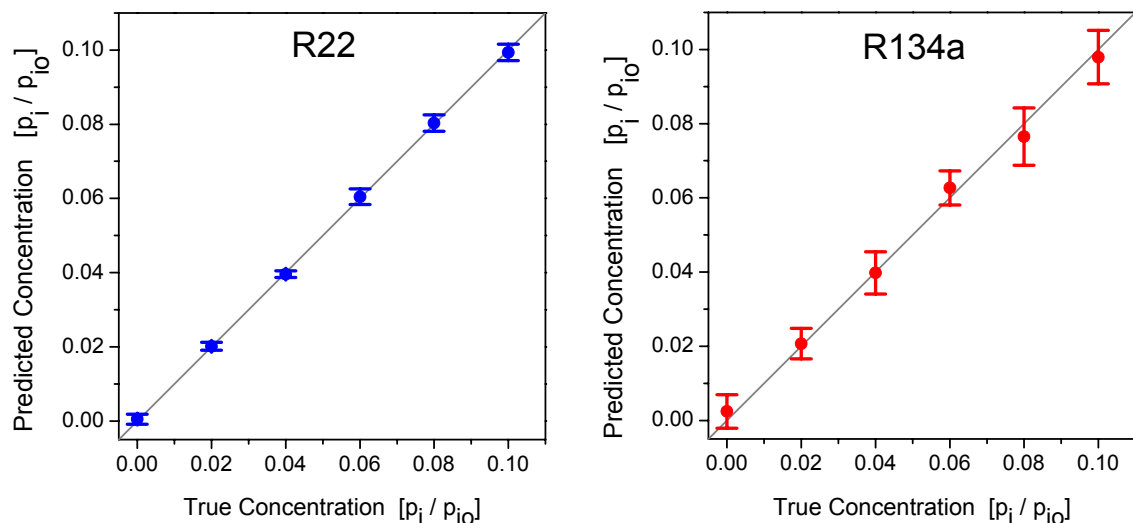


figure 9: Predicted concentrations versus true concentrations of the test data measured by the array setup.

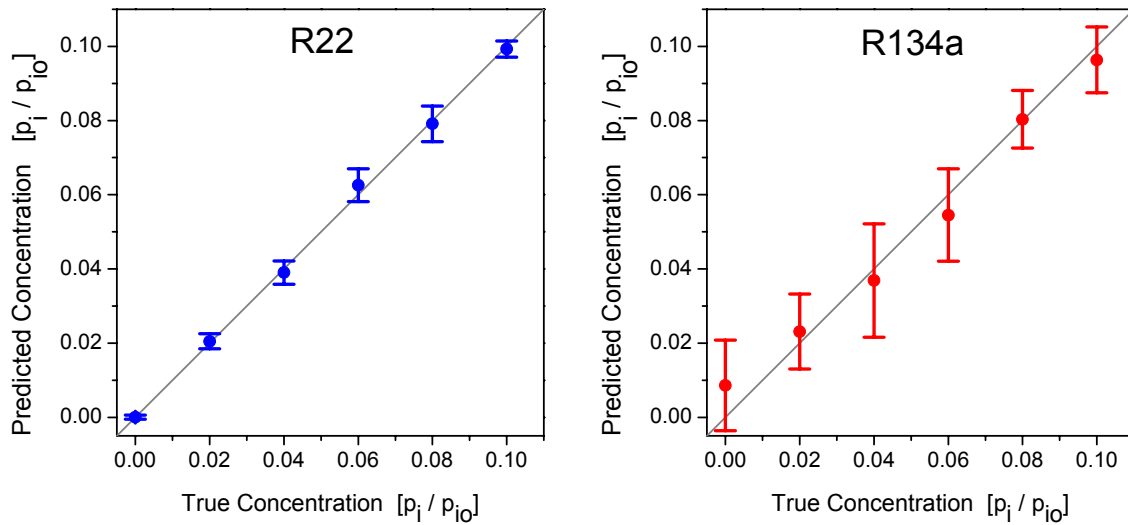


figure 10: Predicted concentrations versus true concentrations of the test data measured by the 4λ setup.

The RMSE for the measurements performed by the 2-sensor 4λ setup are listed in the second row of table 1, whereby the relative RMSE of the test data are 7.4% for R22 and 21.4% for R134a. The RMSE of the calibration data and the test data are also not significantly different indicating again the absence of typical problems of calibrations by neural networks like overtraining effects. Yet, the RMSE of both analytes are nearly twice as high as the RMSE of the measurements performed by the array setup with 6 sensors. The true-predicted plots in figure 10 correspond with this higher RMSE. The high standard deviations of the true-predicted plot of R134a show that the predictions are rather scattered. Additionally, the predictions of the low concentrations are a little biased with the absence of R134a being predicted as a relative saturation pressure of 0.0086 on average. On the other hand, the true-predicted plot of R22 shows that the predictions of R22 are not biased and more precise than the predictions of R134a demonstrating that a quantification of R22 in the presence of R134a is possible by using the low-cost 4λ setup.

The deterioration of the quantification of both analytes using the 4λ setup instead of the array setup raises the question, if this deterioration is caused by the reduction of the number of sensors or by the type of setup. Therefore, the data set recorded with the array setup was calibrated with neural networks again, but this time only the sensors PUT and UE2010 15% were used as input variables. For the comparison the UE 2010 15% layer is used instead of the 20% layer, as the UE 2010 layer used in the 4λ setup is more similar to the UE 2010 15% layer in respect to the thickness of the layer. The RMSE of this experiment are listed in row 3

of table 1. The RMSE of R22 is nearly equal for the 4λ setup and for the array setup with 2 sensors. Thus, the deterioration of the prediction of R22 can be primarily ascribed to the limitation of the number of sensors. On the other side, the RMSE of R134a of the 2-sensor array lies in between the 4λ setup and the 6-sensor array. Hence, not only the limitation of the number of sensors, but also the type of setup plays a role for the quantification of R134a.

Setup	Crossv.	Crossv.	Test	Test	Test
	R22	R134a	R22	R134a	Mean
Array: 6 Sensors	0.00162	0.00572	0.00183	0.00630	0.00406
4λ : 2 Sensors	0.00325	0.01169	0.00317	0.01146	0.00732
Array: PUT + UE 2010 15%	0.00283	0.00750	0.00334	0.00891	0.00612
Array: PUT + UE 2010 20%	0.00182	0.00626	0.00203	0.00678	0.00440
Array: PDMS + UE 2010 20%	0.00184	0.00596	0.00206	0.00667	0.00436
Array: PDMS + M 2400	0.00238	0.00673	0.00268	0.00772	0.00520
Array: PDMS + UE 2010 15%	0.00276	0.00656	0.00329	0.00810	0.00569
Array: PUT + M 2400	0.00241	0.00776	0.00273	0.00881	0.00577
Array: PUT + PDMS	0.00611	0.01156	0.00685	0.01259	0.00972
Array: HBP + PDMS	0.0613	0.01122	0.00714	0.01257	0.00986
Array: HBP + UE 2010 20%	0.00177	0.01902	0.00199	0.02139	0.01169
Array: HBP + PUT	0.00657	0.01531	0.00773	0.01789	0.01281
Array: M 2400 + UE 2010 20%	0.00182	0.02371	0.00211	0.02371	0.01291
Array: HBP + UE 2010 15%	0.00316	0.02368	0.00384	0.02828	0.01606
Array: M 2400 + HBP	0.00320	0.02447	0.00393	0.02930	0.01661
Array: M 2400 + UE 2010 15%	0.00296	0.02774	0.00368	0.03436	0.01902
Array: UE 2010 20% + UE 2010 15%	0.00187	0.02962	0.00210	0.03431	0.01821

table 1: Root mean square errors for the calibration and prediction of the data measured by the array setup using 6 polymer sensors (1st row) and using all combinations for 2 sensors out of 6 sensors (3rd row to 17th row). The root mean square errors for the miniaturized 4λ setup, which uses 2 sensors, are listed in the 2nd row.

3.5. Variable Selection by Brute Force

The problem of finding the best 2-sensor combination for the discrimination of the two analytes is a typical variable selection problem. The rather low number of 2 variables out of 6 allows a brute force variable selection, as according to equation (15) only 15 combinations are possible. The RMSE of these 15 combinations, which are listed in row 3 to row 17 of table 1, confirm the conclusion drawn when analyzing the sensitivities of the sensors for the pure analytes. The best combination with the lowest mean RMSE of the test data is the PDMS + UE 2010 20% layer and the second best combination is the PUT + UE 2010 20% layer. The combinations with the 6 lowest mean RMSE of the test data in between 0.00440 and 0.00612 consist all of one polar polymer and one microporous polymer, which showed the biggest differences of the sensitivity patterns in figure 8. The combination of two different interaction principles seems to be optimal for the discrimination of the two refrigerants as the combinations of 2 polar or 2 microporous polymers all showed significantly higher errors.

3.6. Conclusions

In this example, 6 polymers were investigated for application in a sensor system for the quantification of the refrigerants R22 and R134a in mixtures. By the use of all polymers an accurate quantification of both analytes could be performed. Additionally, 2 polymers were selected on the basis of the sensitivity patterns for the application in a small low-cost 4 λ RIFS setup. By the use of these 2 polymers, the 4 λ RIFS setup could quantify R22 in the presence of R134a quite well but not vice versa.

It was shown that the best selection of 2 sensors is the combination of 1 microporous polymer with 1 polar polymer enabling a discrimination on the basis of two interaction principles. The two interaction principles show the most different sensitivity patterns for the two analytes, which is the common selection criterion of sensor coatings for an analytical problem.

The polymers used in this example show two different types of sorption, the specific Langmuir sorption and the unspecific Henry sorption. For the classical feature extraction, which uses the height of the raw signal after a definite time of exposure to analyte, the unspecific Henry sorption is advantageous as the immediate sensor responses allow very fast measurements. Yet, a drawback of this common feature extraction is the extraction of only one single variable per sensor. This limits the number of analytes to be quantified to the

number of sensors in the ideal case. This means that a 2-sensor setup can be calibrated for 2 analytes or contaminants and the 6 sensor array setup can be calibrated for up to 6 analytes and contaminants. Contaminants in the samples, which do not sorb into the sensitive layers, do not interfere the determination of the concentration of both analytes and can be ignored during the calibration. Yet, contaminants, which sorb into the polymer layers, bias the predictions of the analyte concentrations unless they can be considered during the calibration process as additional analytes in combination with additional sensors.

In chapter 5, the principle of a time-resolved feature extraction will be introduced. Thereby the kinetics of sorption and desorption of analytes is exploited allowing the extraction of a virtually unlimited number of variables per sensor. Thus, the number of analytes, which can be quantified per sensor, is limited only by similarities of kinetics and is not fixed by the device. The time-resolved approach removes the limitation of the common sensor array approach of being able to quantify simultaneously only a maximal theoretical number of analytes. The time-resolved approach also changes dramatically the search and the rating of polymers, which might be suitable for a specific analytical problem. The static measurements with a single feature extraction need sensitivity patterns as different as possible, whereas the kinetic feature extraction needs different shapes of the sensor responses during sorption or desorption.

The selection of the best combination of polymers based on the sensitivity pattern in this example could be verified by a brute force variable selection approach. Yet, the time-resolved measurements introduced in this work generate many variables putting new challenges to calibration methods and variable selection techniques, as a brute force variable selection is rendered impossible. Therefore, the introduction of new calibration techniques combined with variable selection methods are one of the focuses of this work.

4. Experiments, Setups and Data Sets

4.1. *The Sensor Principle*

For this work data sets were used, which were measured by several different setups to demonstrate that the principles of data analysis introduced in this study can be generally applied to a broad spectrum of devices. All setups belong to the category of the sensor devices, which transform chemical information into physically measurable quantities. Sensors allow establishing durable and economical devices for fast measurements without the need of sample pretreatments rendering these devices ideal for monitoring environmental pollutions, for process monitoring and for all kinds of continuous on-line and in-line measurements. Depending on the sensitive layer, which is responsible for the recognition of the chemical properties, the sensors are often divided into two groups: The sensors belonging to the first group have polymers, metals or metal oxides as sensitive layers. These sensors are called chemosensors. The sensitive layers of the second group of sensors use biochemical interactions like antigen – antibody or DNA – DNA interactions resulting in the name biosensor. In this work, data sets are analyzed for the detection of gases and vapors of volatile organic compounds. These data sets were measured by colleagues using 3 different devices, which are all based on polymer chemosensors. The polymer-based chemosensors recognize the presence of analyte by changes of the thickness and changes of the refractive index of the sensitive layer when analyte sorbs into the sensitive polymer layer. Two devices are based on the **Reflectometric Interference Spectroscopy (RIfS)** as detection principle and one device is based on the **Surface Plasmon Resonance Spectroscopy (SPR)**.

4.2. SPR Setup

The SPR device used for the measurements of this work is a single sensor device, which is based on the resonance of surface plasmons as detection principle. The setup, which is shown in figure 11 and which was assembled by Dr. Birgit Kieser [171], was initially suggested by Kretschmann in 1968 [172]:

At constant angle, parallel-polarized white light is focused through a glass prism onto a silver film on condition of total reflection. The surface plasmons are excited at the back surface of the silver layer, which is evaporated onto the glass prism. The reflected light is coupled into a multimode fiber and detected by a diode array spectrometer. A sensitive polymer layer is coated onto the top of the silver film. The resonance wavelength of the surface plasmons is influenced mainly by the refractive index n and partly by the thickness d of the sensitive layer (for all sensitive layers used in this work practically only changes of the refractive index n play a role [173]). Due to a high temperature dependency of the refractive index, the temperature was kept constant by a thermo-regulator. More details of the setup can be found in literature [174].

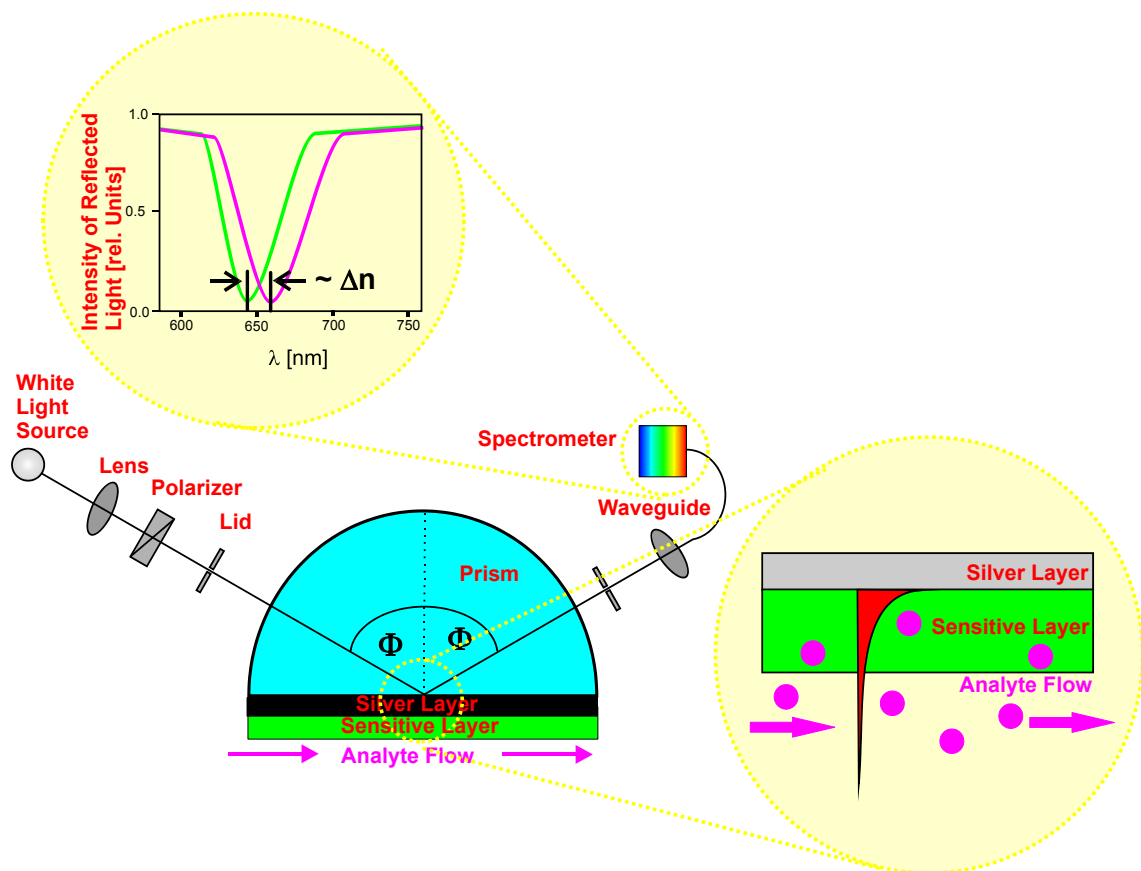


figure 11: Schematic of the RI/S array setup. The yellow bubbles show the effects when analyte sorbs into the polymer layer (magenta).

4.3. RIfS Sensor Array

The Reflectometric Interference Spectroscopy (RIfS), which was initially proposed by Gauglitz and Nahm [175], evaluates the interference pattern of light, which is reflected at thin transparent films. The interference is caused by different partial beams, which are reflected in a different way at the film interfaces depending on the wavelength. Changes of the optical thickness nd of the sensitive layer can be detected as shifts of the maxima and minima of the interference pattern [176].

The RIfS sensor array shown in figure 12 contains 7 sensors allowing the application of 7 sensitive layers. The configuration of the sensor array is based on a white light source. The light is transmitted through a lens and filter system via polymer fibers to the glass substrates with 7 different polymer layers. The reflected light is transmitted to an optical multiplexer, which is directly connected to a diode array spectrometer. More details can be found elsewhere [177].

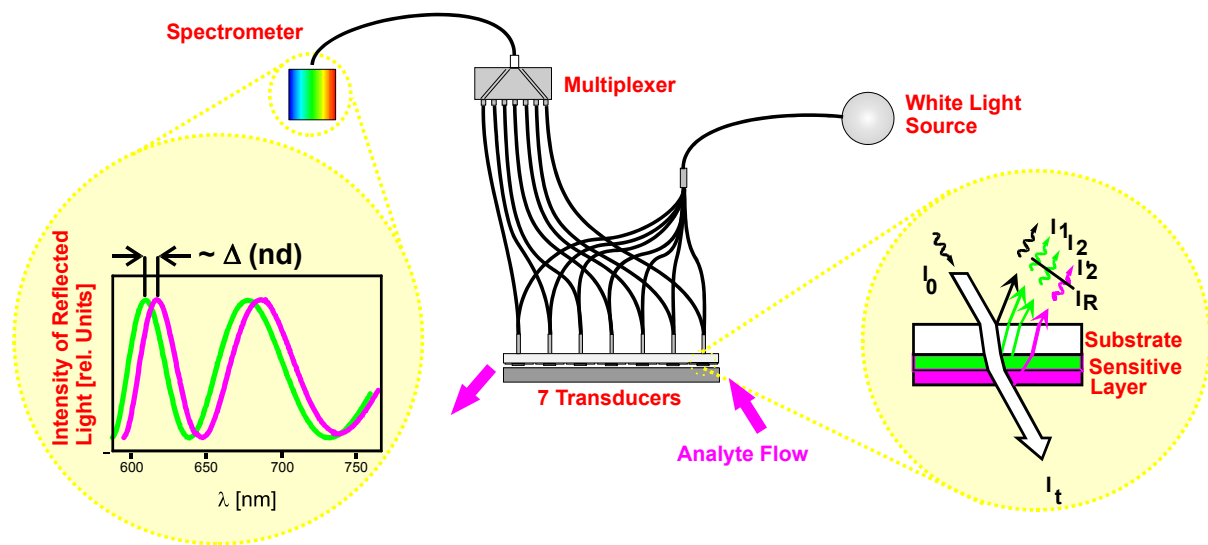


figure 12: Schematic of the RIfS array setup. The yellow bubbles show the effects when analyte sorbs into the polymer layer (green before sorption of analyte, magenta after sorption of analyte).

4.4. 4λ Miniaturized RIfS Sensor

The second setup, which is based on RIfS as detection principle, is shown in figure 13. This so-called 4λ system bases on a wavelength separation in the light source. Individual wavelengths are sequentially emitted by 4 light emitting diodes rendering a spectral recording of the reflected light unnecessary. Thus, the reflected light is detected by a simple photodiode enabling a miniaturization of the system and a reduction of the costs. The setup, which has been used for the detection of single analytes and sum parameters, is described elsewhere in detail [178,179]. As the parallelization of this setup is still in progress, the measurements using two different sensitive layers were performed in series.

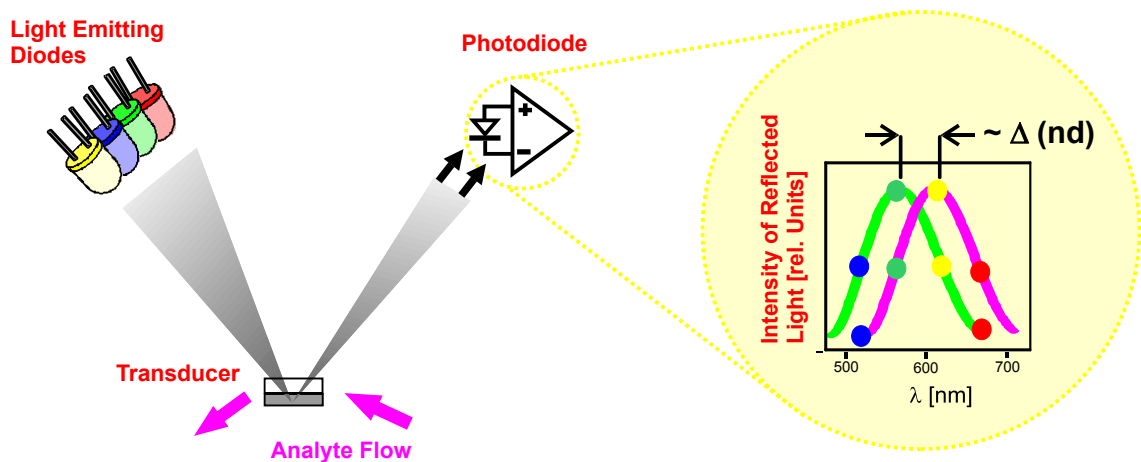


figure 13: Schematic of the 4λ setup. The yellow bubble shows the reconstruction of the shifted interference pattern after the sorption of analyte into the polymer layer (green before sorption of analyte, magenta after sorption of analyte).

4.5. Data Sets

4.5.1. Refrigerants R22 and R134a

For more than four decades a chlorofluorocarbon known as R22 (chlorodifluoromethan) has been the refrigerant of choice for application in freezers, refrigerators and air-conditioners. Yet, in the early 70s, Rowland and Molina showed that the chlorine of chlorofluorocarbons and hydrochlorofluorocarbons is released when the molecules are exposed to ultraviolet radiation in the stratosphere. A free chlorine atom in the stratosphere can act as catalyst for breaking down many stratospheric ozone molecules [180-182]. Consequently, in 1987 the Montreal Protocol was signed by almost all industrial countries. The Montreal Protocol is an international agreement on the phase-out of refrigerants based on chlorofluorocarbons. The Montreal Protocol is carried out in national laws of all participating countries like the "Clean Air Act, Title VI" in the United States [183] or the "FCKW-Halon-Verbotsverordnung" in Germany [184].

In consequence, national laws [185,186] regulated the substitution of the chlorofluorocarbons and hydrochlorofluorocarbons by hydrofluorocarbons, which do not deplete the ozone layer. Among these substitutes, the hydrofluorocarbon R134a (1,1,1,2-tetrafluoroethane) plays the most important role. According to the "Alternative Fluorocarbons Acceptability Study" [187] the portion of R134a among the worldwide sold refrigerants for air-conditioners and refrigerators increased from 0% to 31% in the period from 1990 to 2000. In the same time, the portion of the hydrochlorofluorocarbons decreased from 95% to 64% whereby in 2000 practically no hydrochlorofluorocarbons except of R22 were sold for the use in refrigerants and air-conditioners.

This gradual substitution process of the refrigerants and the resulting diversity of the refrigerants found in old freezers, refrigerators and in air-conditioners prevent a reuse of the refrigerants. At the moment, the recycling process is a downcycling process whereby the refrigerants are converted to hydrogen chloride and hydrogen fluoride. A high quality recycling is not possible since the refrigerants are sometimes returned with no labels or wrong labels and since the refrigerants are often mixtures originating from different refrigerators or even were used as blends like the so-called R400 and R500 series. On the other hand, an analysis of the refrigerants of each refrigerator by a laboratory is too expensive. This results in the need of a detection method, which enables a fast, durable and economical on-line quantification of different mixtures of refrigerants. As sensor setups can come up to these

requirements, data sets were recorded to examine how feasible an application of sensor setups for the quantification of R22 and R134a in mixtures in air is. Additionally several polymers were investigated to find an optimal polymer respectively an optimal combination of polymers for the discrimination of the refrigerants R22 and R134a (see chapter 3).

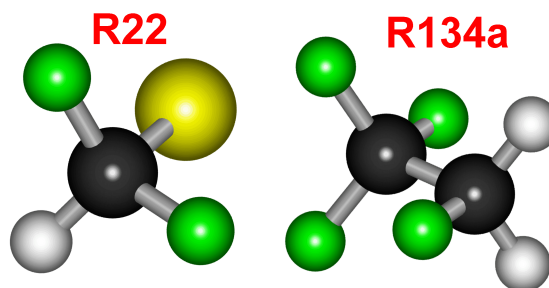


figure 14: 3D ball and sticks molecular models of the refrigerants R22 and R134a.

The gas mixtures of R22, R134a and air were generated using a gas mixing station with computer-driven mass-flow controllers (MKS, Munich, Germany). A 4-way valve before the cell ensured that the path length was the same for all analytes. Dry synthetic air was used as carrier gas. All measurements were performed at a constant flow rate.

4.5.1.1. R22 and R134a by the SPR Setup

The SPR setup, which is described in section 4.2 in detail, was used for measuring the two refrigerants in air. The polycarbonate Makrolon (Makrolon M2400, Bayer AG, Leverkusen Germany) was used as sensitive layer. Details of the preparation of the 60 nm thick sensitive layer can be found in [188].

In addition to single analyte measurements with the sensor responses shown in figure 20 and in figure 22, two multicomponent data sets were recorded [189], which are used for the systematic investigation and development of different multivariate data analysis methods. These data sets will be further referred to as "refrigerant data". The first data set is a 21-level full factorial design whereby the concentrations (relative saturation pressures) of the two analytes were varied between 0 and 0.1 p_i/p_{i0} relative saturation pressure. This data set is used for the calibration and optimization of the different multivariate models and will be further referred to as calibration data set. The second data set is a 20-level full factorial design and will be referred to as validation data set. The concentrations of this independent external validation data set were varied between 0.0025 and 0.0975 p_i/p_{i0} in 0.005 steps. Thus, all concentration levels of the two data sets are different and consequently the validation data set should give a realistic estimate of the network performance in a real world situation [7]. In

total, 841 different mixtures of the refrigerants R22 and R134a were measured by SPR. All measurements were performed in random order. The sensitive layer was exposed to the analyte-air mixtures for 60 seconds and afterwards to synthetic air for 300 seconds for recovery. During the sorption and desorption process the signal was recorded with a resolution of 45 data points (see figure 20). The signals at each time point were used as independent variables whereby only the 40 time points until 125 seconds of measurement time were used for the multivariate data evaluation (see chapter 5 for details of the time-resolved measurement proceedings).

4.5.1.2. R22 and R134a by the RlfS Array and the 4 λ -Setup

This data set was already introduced and described in chapter 3 and will be further investigated in section 9.5 as the sensor response of the Makrolon layer was also recorded in a time-resolved mode. The signal of the Makrolon sensor could be extracted from the raw measurements at 18 time points during sorption and desorption. Thus, 5 independent variables of the static measurements of 5 polymers and additionally 18 independent variables of the time-resolved measurements were available.

4.5.2. Homologous Series of the Low Alcohols

In a homologous series, the analytes differ from each other only by additional CH₂ groups. This allows a systematic variation of the size of the analytes with other molecular properties like polarizability or polarity being nearly constant. In this study the alcohols methanol, ethanol, 1-propanol and 1-butanol are measured with the microporous polymer Makrolon as sensitive layer of different setups. The differences of the analyte sizes (0.068 nm³ to 0.180 nm³) allow a systematic investigation of the quantification based on different kinetics of sorption into and desorption out off the micropores of Makrolon. The mixtures of the analytes were generated by thermo-controlled bubblers filled with pure liquid analytes sorbed in chromosorb and a subsequent gas mixing station with the same path length for all analytes.

4.5.2.1. Methanol and Ethanol by the SPR Setup

An about 300 nm thick Makrolon layer was coated on the glass prism of the SPR setup. Further details of the experimental setup can be found in [190]. In addition to several single analyte measurements, 2 datasets were recorded for a multicomponent analysis [191]. The calibration data set is a 9-level full factorial design whereby the relative saturation pressures (p_i/p_{i0}) of methanol and ethanol were varied between 0 and 0.045. The validation data set is an

8-level full factorial design with relative saturation pressures between 0.0025 and 0.0425. 24 measurements were identified as outliers according to [192], whereby 20 measurements of the calibration set and 4 measurements of the test set were removed. These measurements showed outlying high signals and were recorded in series. During all measurements, the polycarbonate was exposed to the analyte mixtures for 120 s and after that to synthetic air during 300 s. During the exposure to analyte the data were recorded every 5 s and during the exposure to synthetic air with a time resolution of 10 s resulting in 53 time points (see also figure 15).

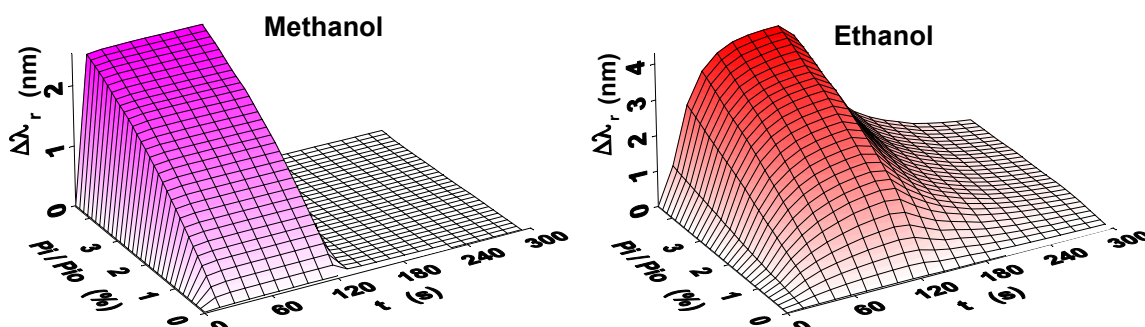


figure 15: Sensor responses of different concentrations of single analytes recorded during 120 seconds of exposure to analyte and afterwards during exposure to synthetic air. The x-axis represents the exposure time, the y-axis represents different relative saturation pressures (concentrations) of the analytes and the z-axis represents the signal shifts during the SPR measurement.

4.5.2.2. Methanol, Ethanol and Propanol by the SPR Setup

For the multicomponent analysis of ternary alcohol mixtures, 2 data sets were recorded [193] using the SPR setup with a 300 nm Makrolon layer (more details in [194]). The calibration data set is a 6-level equidistant full factorial design whereby the relative saturation pressures of methanol, ethanol and 1-propanol were varied between 0 and 0.035. 18 concentration combinations were measured two more times for an estimation of the experimental signal inaccuracies. The validation data set is a 5-level equidistant full factorial design with relative saturation pressures between 0.0035 and 0.0315. Additional 15 measurements were performed with the pure analytes at the five concentration levels. 9 measurements were identified as outliers according to [192] whereby 7 measurements of the calibration data set and 2 measurements of the validation data set were removed. During all measurements the polycarbonate was exposed to the analyte mixtures for 600 seconds and afterwards to dry synthetic air during 4760 seconds with 28 time points recorded during exposure to analyte and 22 time points recorded during exposure to air resulting in 50 input variables in total. The sensor responses are shown in figure 16 for different concentrations of the single analytes. All measurements were performed in random order within the two data sets. The validation data

set was recorded one month after the calibration data set and the signals of this data set were averaged using two measurements.

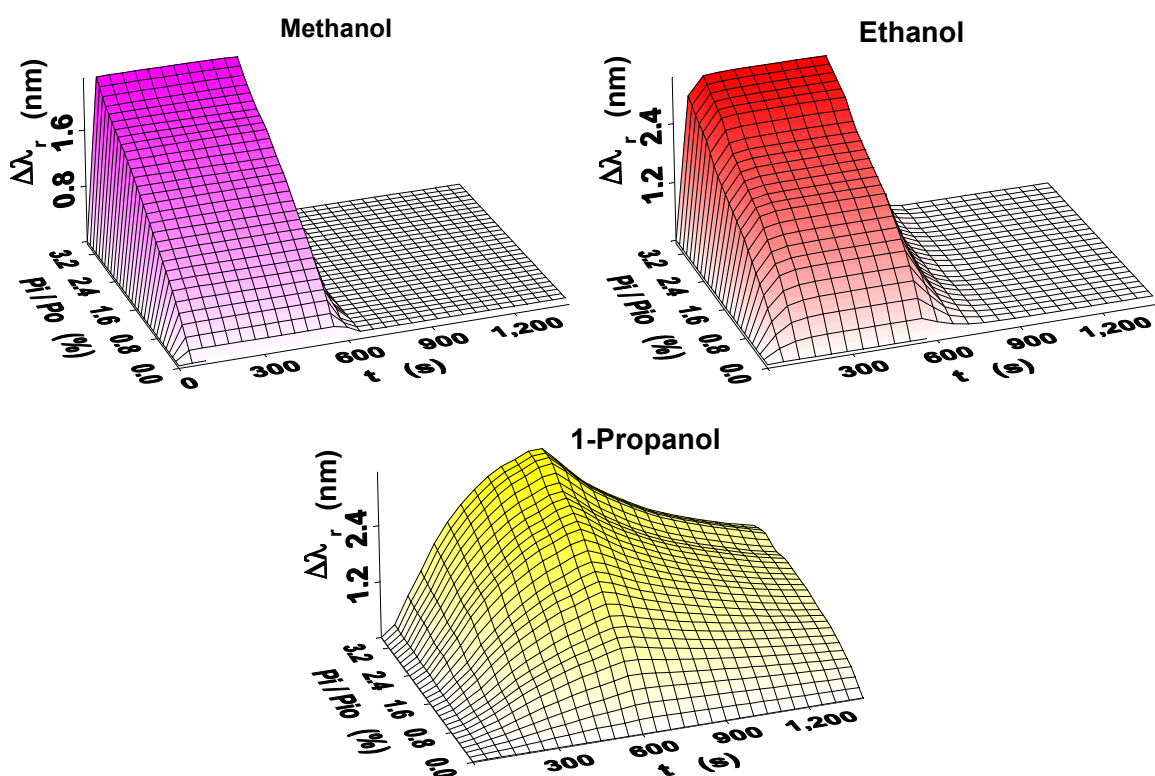


figure 16: Signals of the SPR device during exposure to analyte and then to synthetic air for different concentrations of methanol ethanol and 1-propanol.

4.5.2.3. Methanol, Ethanol and Propanol by the RIfS Array and the 4λ Setup

Ternary mixtures of methanol, ethanol and 1-propanol were measured by the RIfS Array setup and the 4λ setup to compare the performance of these setups [195]. Additionally the effect of smoothing the sensor responses is investigated, as the sensor signals of the RIfS setups are noisier than the sensor signals of the SPR setup using Makrolon as sensitive layer. For both setups, a 6-level equidistant full factorial design with relative saturation pressures between 0 and 0.1 was measured twice for the calibration data set. The validation data set is a 5-level equidistant full factorial design with relative saturation pressures between 0.01 and 0.09 also measured twice. The data of the 4λ setup, which was equipped with a 310 nm Makrolon layer, were recorded during 180 seconds of exposure to analyte (44 time points recorded) and afterwards 575 seconds during exposure to synthetic air (149 time points recorded). The data set of the array setup, which was equipped with a 80 nm, a 120 nm, and a 160 nm Makrolon layer and additionally with a PUT layer, were recorded during 240 seconds exposure to analyte and afterwards during 210 seconds of exposure to synthetic air with a

time resolution of 50 data points per sensor. Due to a bug in the program for the gas mixing station, 15 measurements of the calibration set and 18 measurements of the validation sets had to be removed.

4.5.2.4. Quaternary Mixtures of Alcohols by the RIfS Array and the SPR Setup

Finally, quaternary mixtures of methanol, ethanol 1-propanol and 1-butanol were measured by the RIfS array setup and by the SPR setup to compare the performance of both setups [196]. A 4-level equidistant full factorial design with relative saturation pressures between 0 and 0.09 was measured for the calibration data set. The validation data set is a 3-level equidistant full factorial design with relative saturation pressures between 0.015 and 0.075. The data of the SPR setup, which was equipped with a 100 nm Makrolon layer, were recorded during 124 seconds of exposure analyte (40 time points) and afterwards for 806 seconds during exposure to synthetic air (116 time points). The sensor responses for different concentrations of the single analytes are shown in figure 24. The data of the array setup, which was equipped with a 95 nm and a 165 nm Makrolon layer and additionally with a PUT layer, were recorded during 236 seconds exposure to analyte (16 time points recorded per Makrolon layer) and afterwards during 3240 seconds of exposure to synthetic air (34 time points recorded per Makrolon layer). The sensor signals of the PUT layer were recorded only once at 236 seconds as the swelling kinetics of this polymer cannot be differentiated at the chosen time resolution of the measurements.

5. Results – Kinetic Measurements

In this chapter, the principle of the time-resolved measurements is introduced. In contrast to the common principles of static measurements and static feature extractions, this approach allows the simultaneous quantification of a virtually unlimited number of analytes. Although some sporadic reports of time-resolved measurements in chemical sensing can be found in literature, no systematic investigations of time-resolved measurements, no systematic exploitations of the time domain and no transfers to different setups and measurement principles have been reported for chemical sensing up to now. Although the approach of the time-resolved measurements can be used for many different transduction and interaction principles, this study focuses on the application of one specific polymer for the time-resolved measurements. Thus, after an explanation of the static and the time-resolved sensor measurements this polymer is introduced. Then different properties of the polymer and different interactions of the polymer with analyte molecules are investigated as the understanding of these properties and interactions allows tweaking the sensors for different analytical tasks. This tuning is demonstrated in the last sections of this chapter.

5.1. *Static Sensor Measurements*

The quantification of multiple analytes in mixtures by sensors is faced by several problems. For the determination of the regression parameters of the MLR (expression (7)) and of all other multivariate calibration methods as many sensor signals as analytes to be quantified are needed since otherwise the set of equations would be statistically underdetermined [197]. In real world measurements, even more signals are needed as noise, drifts, interfering unknown analytes and other non-ideal influences "consume" additional degrees of freedom. Typically, the sensors used are not selective for a single analyte also known as crossreactivity. The resulting covariance of the sensor signals causes additional problems for the calibration resulting in the need of as many sensors as possible, which show different sensitivity patterns for the analytes of interest. Thus, the common approach to quantify several analytes simultaneously is the combination of many sensors on a sensor array with the sensors showing different sensitivity patterns for the different analytes and a subsequent multivariate data analysis [198-202]. An advantage of this unspecific multisensor approach is the possibility of using the same array setup for many different analytes without the need of finding specific sensor materials [203]. Usually only one single feature of the sensor response

(value of the sensor signal) is extracted, such as the height of the sensor response at equilibrium or the slope of the sensor response resulting in the need of more sensors in the sensor array than the number of analytes to be quantified.

5.2. Time-resolved Sensor Measurements

Recently a new approach has been proposed by several groups for reducing the number of required sensors [204-214]. This approach is based on the exploitation of the time specific information of sensor responses. If various analytes show different kinetics for the sorption into the sensitive layer, the resulting sensor response recorded versus time features a different shape for these analytes. This additional temporal information can render the parallel information of different sensors in an array redundant allowing reducing the number of the sensors. For that purpose the shape of the sensor responses has to be digitized by recording the sensor responses over time (further referred to as time points) and by performing a multivariate analysis of these time points. These time-resolved measurements were performed in various areas of sensor research. For example, Yan et al. [204] quantified binary mixtures of solvents in water by a single reflectometric interference spectroscopic sensor whereby the time shift of the highest signal after analyte exposure depended on the composition of the mixture similar to gas or liquid chromatography. The components of binary and ternary mixtures of organic analytes in water could also be determined by the use of a single amperometric sensor [207,214]. Thereby, the consumption of oxygen by the metabolism of microorganisms with different time constants for the analytes was detected. In the gaseous phase, time-resolved measurements were used in combination with sensor arrays to obtain additional variables. Using time-resolved measurements with an array of quartz microbalances coated with three different polymer films the classification of six solvent vapors was improved compared with the classification using only the saturation mass [215]. Johnson et al. [210] classified 20 different analytes with only 4 fiber optic sensors and also classified these analytes semi-quantitatively into low, medium and high analyte concentrations using ten fiber optic sensors with 90 % of the test data being assigned to the correct concentration class. Podgorsek et al. [216] used a glassy polyimide for the detection of methanol and ethanol and showed the difference of the response times, which could be used for the discrimination of both analytes. However, a detection of both analytes in mixtures was not performed.

Yet, all these publications used the time-resolved measurements as phenomenological tool and no systematic research was performed concerning the optimization of the different mechanisms and components like the time delaying effects, the time-resolution of the recorded sensor response, the feature extraction, the data preparation and the multivariate data analysis, which was rather basic in most of the publications cited. For example in most of the approaches the sensor responses were recorded using a rather coarse time resolution [205, 206,212] or even a coarse resolution combined with other more or less arbitrary features [210,211,217]. Although a coarse time-resolution grants an easily manageable quantity of information, the risk of losing important information can be high, especially if several analytes show similar or very fast sensor responses. Thus, a fine time grid should be generally preferred, which nevertheless needs a more sophisticated data analysis. Additionally, all approaches cited above were isolated applications and no transfer to other systems was performed.

In this work, an extensive and systematic research on time-resolved measurements is performed from the investigation of the effects causing the time delays to an optimization of the data analysis. The principle of time-resolved measurements is applied to several analytical problems and to several different sensor devices. In this work, the time-delaying effects are based on a microporous polymer. For the data analysis several methods are developed, which allow a highly efficient evaluation of a fine time grid of the sensor responses. These methods are embedded in frameworks, which allow a simultaneous variable selection and calibration of nonlinear data and which can be applied to any linear and nonlinear multivariate relationship.

5.3. Makrolon – A Polymer for Time-resolved Measurements

5.3.1. General Properties of Makrolon

The time-resolved measurements of the SPR and RIfS setups are based on sensitive layers, which were prepared using a polycarbonate polymer. This polymer is commercially available as Makrolon M2400 from Bayer AG, Leverkusen, and will be further referred to as Makrolon. The structure of the monomer unit is shown in figure 17. The interesting property of Makrolon is the microporous structure of the glassy polymer. According to [218], the size of these pores follows a distribution with a mean pore size of 0.1 nm^3 determined by the use of the positron annihilation lifetime spectroscopy (PALS). The PALS detects the decay of positrons and exploits the effect that the positrons are more probably present in the pores than in the polymer matrix due to the lower electron density in the pores [218,219].

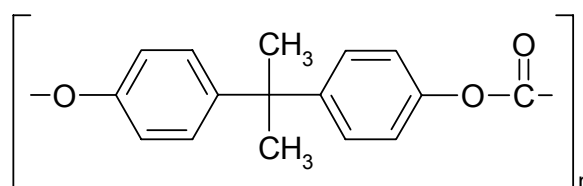


figure 17: Structure of the monomer unit of the polymer Makrolon M2400.

In figure 18, the shift of the SPR wavelength is shown while the device was exposed to an alternating sequence of synthetic air and relative saturation pressures of methanol of 0.31, 0.62 and 0.80. At these high concentration levels, the sensor signal increases rapidly at the beginning of the analyte exposure meaning that the refractive index of the sensitive layer increases. This increase of the refractive index can only be explained by methanol ($n_{D20}=1.329$) sorbing into the micropores of the polymer and replacing air ($n_{D20}=1$) in these polymers as the sorption of methanol into the polycarbonate matrix ($n_{D20}=1.58$) would decrease the refractive index of the sensitive layer. The molecules of methanol have less volume (0.068 nm^3) than the mean size of the pores of Makrolon and therefore can easily sorb into the micropores.

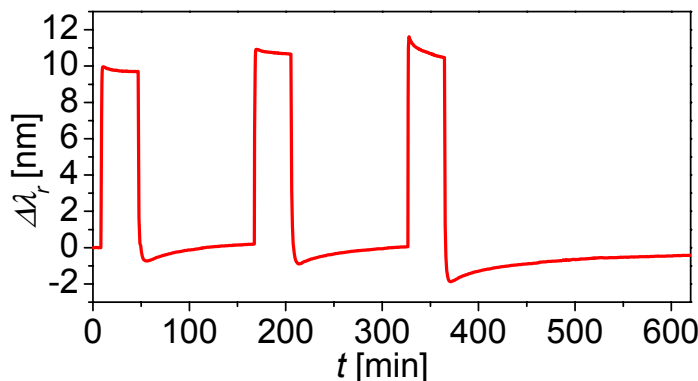


figure 18: SPR sensor response of the SPR device when exposed to different relative saturation pressures of methanol of 0.31, 0.62 and 0.80 and alternating to synthetic air.

During a long-term exposure of Makrolon to methanol (see figure 18), the refractive index decreases. This decrease reinforces with higher concentrations of methanol and can be explained by two effects. The first effect is based on an expansion of the micropores of the polycarbonate matrix when these micropores are occupied by analyte molecules resulting in a decrease of the refractive index of the sensitive layer. When the exposure to methanol stops the refractive index decreases rapidly as the methanol molecules quickly desorb. After the exposure to high concentrations of methanol, the refractive index decreases below the initial value, as the pores and with it the Makrolon matrix are still expanded. The initial refractive index is gradually reached while the Makrolon layer is shrinking to the initial thickness. The expansion of the Makrolon matrix can only be detected when exposed to analyte for a long time. This effect allows the modification of the kinetics of sorption and desorption for bigger molecules and is exploited in section 5.3.4. The second effect is an unspecific sorption of the analyte into the polymer matrix besides of the specific sorption into the pores resulting in a swelling of the matrix. As the analyte has a lower refractive index than the polymer matrix, the shift of the SPR wavelength decreases. This unspecific sorption is a Henry type sorption (see section 3.2) whereas the sorption into the micropores can be considered as a Langmuir sorption (see section 3.2). In figure 19 the isothermal calibration curve for the sorption of methanol into Makrolon is shown. The curve shows the typical shape of the Langmuir sorption and hardly any portion of Henry sorption can be detected whereas the isothermal calibration curves of R22 and R134a show a significant Henry sorption (see figure 7 in section 3.2 and additionally the discussion in section 3.2). At the moment, further research is done by combining measurements of an ellipsometric device, an SPR device and a RIFS device for distinguishing the sorption into the pores from the sorption into the matrix and for

distinguishing the long-term expansion of the matrix from the swelling due to analyte sorption. First experiments allow following preliminary conclusions:

- The Henry sorption plays only a role if the analyte molecules are so big that a specific Langmuir sorption into the pores is not possible or very slow, whereby the total amount of the Henry sorption is always very low due to the high glass temperature of Makrolon.
- For exposure times less than 30 minutes, the expansion of the polymer matrix can be neglected.
- The sorption of analyte is reversible and the layer is highly stable. For example, the measurements, which are partly shown in figure 18, had a mean drift of $5.3 \cdot 10^{-4}$ nm/h, which corresponds to only 0.01 % of the maximum sensor response.

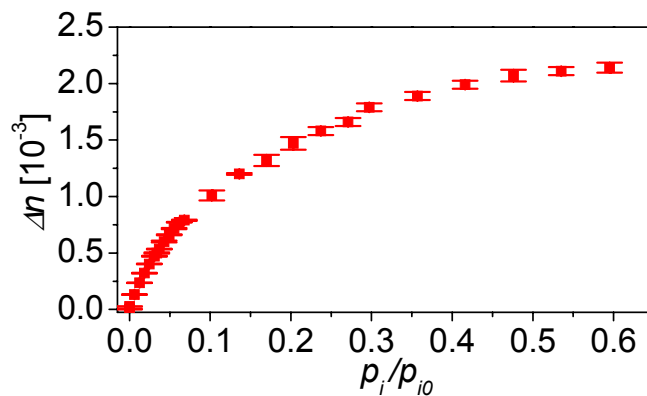


figure 19: Isothermal calibration curve after 30 minutes of exposure to different concentrations of methanol. The standard deviations of 3 measurements are represented by the error bars.

5.3.2. Time-resolved Measurements

The interesting property of Makrolon is the microporous structure, which allows a discrimination of analytes due to different sizes of the analyte molecules [220-222]. Small analytes show a very fast sorption (of the Langmuir type) into the pores, medium sized analytes slower kinetics and bigger molecules hardly sorb into the pores of the polymer. For the two analytes R22 and R134a, the sensor responses of the SPR device are plotted in figure 20 during 60 seconds of exposure to analyte with a relative saturation pressure p_i/p_{i0} of 0.1 and afterwards during exposure to synthetic air. The smaller molecules of R22 immediately sorb into the

pores of the polymer and desorb very fast resulting in a rectangular shape of the sensor response. The bigger molecules of R134a show a slower kinetics resulting in a curvature of the sensor response. A quantitative correlation of the different interaction kinetics with the absolute volume of the analyte molecules is difficult, as different mechanisms play a role. For the diffusion of the molecules within a microporous polymer, often the kinetic volumes of the molecules are used. These kinetic volumes can be calculated with the help of the Van-der-Waals parameters, the critical pressure, the critical temperature, the Lennard-Jones potential and the Stockmayer potential [223-225]. The kinetic volume is 0.120 nm^3 for R22 and 0.140 nm^3 for R134a. On the other hand, for the inclusion of the molecules into the pores, the easiest description is the Van-der-Waals volume, which is 0.048 nm^3 for R22 and 0.056 nm^3 for R134a. For the sorption process into the pores, which can be regarded as a combination of both volumes, the effective volume of the molecules consequently depends on the thickness of the layer as the path length of diffusion increases with the thickness of the layer. This effect is exploited for tailoring the kinetics of a specified analyte by changing the thickness of the sensitive layer (see section 5.3.3).

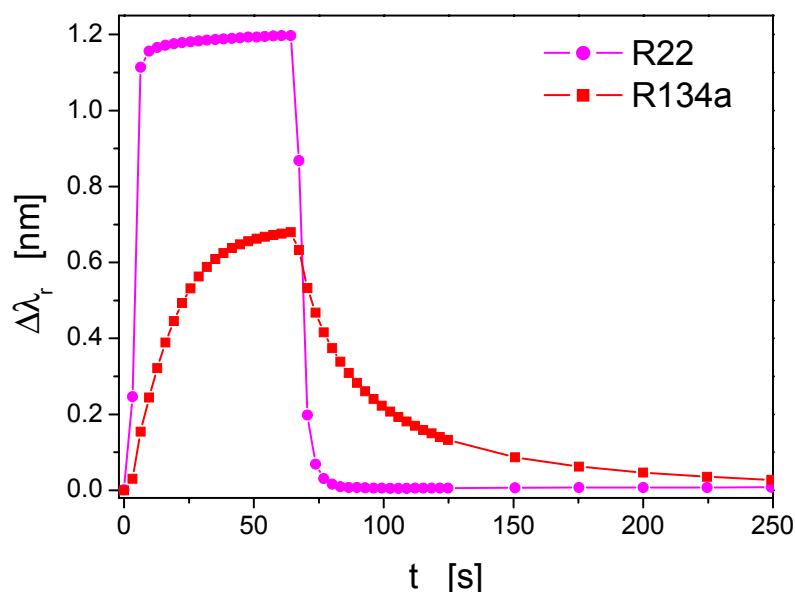


figure 20: SPR sensor responses for 60 seconds of exposure to analyte with a relative saturation pressure of 0.1 and after that for exposure to air.

For a quantitative determination of analytes in mixtures, which is based only on the different kinetics of the analytes, the kinetics has to differ for the complete concentration range of interest. In figure 21, the sensor responses of the sorption of two analytes are plotted versus the time of sorption (x -axis) and versus the concentration of the analytes (y -axis). Both analytes show a linear relationship between the sensor response and the concentration. In

addition, the relationship between the sensor responses and the time of sorption are also linear for both analytes. The similar shape of the sensor responses along the time axis means that both analytes cannot be discriminated by the kinetics although analyte 2 has twice the sensitivity of analyte 1. For example, the sensor response recorded over time is similar if analyte 1 is present with a concentration of 0.8 or analyte 2 is present with a concentration of 0.4 rendering a simultaneous quantification of both analytes impossible. For a successful discrimination, the sensor responses of the analytes have to show a different shape along the time axis and not a different slope. In mathematical terms, the analytes have to show different partial derivatives of the second order (twice derived by time) over the complete concentration range for at least as many time points as analytes to be quantified:

$$\frac{\partial^2 z_1}{\partial t_j^2} \neq \frac{\partial^2 z_2}{\partial t_j^2} \neq \frac{\partial^2 z_3}{\partial t_j^2} \neq \dots \frac{\partial^2 z_n}{\partial t_j^2} \quad (24)$$

With z_i as sensor response of the analyte i and t_j as time point j .

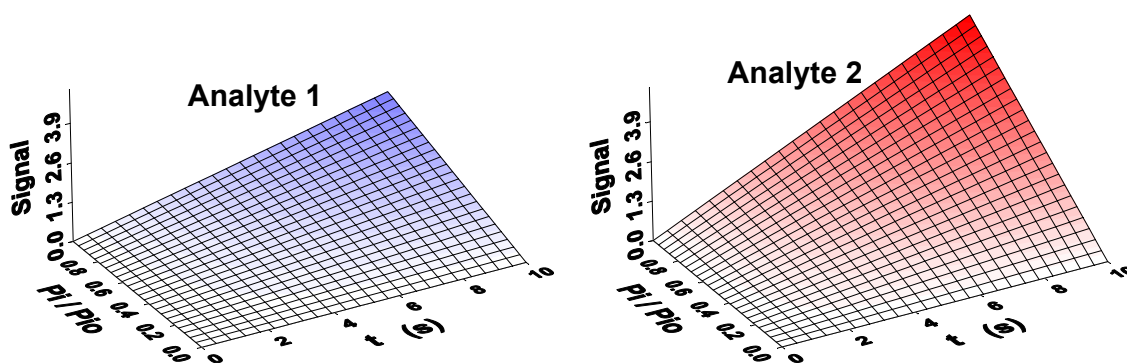


figure 21: Linear sensor response of two analytes for a sorption process versus the concentration and the time.

To test the feasibility of the quantification of R22 and R134a in mixtures by the use of only 1 sensor, the sensor responses of the SPR device were time-resolved recorded for 20 single analyte measurements with relative saturation pressures between 0 and 0.1. The sensor responses of these measurements are shown in figure 22. It is obvious that the kinetics of sorption and desorption differs for both analytes over the complete concentration range enabling the quantitative determination of both analytes in binary mixtures on the basis of the kinetics recorded with only 1 single SPR sensor. Thereby the sensor responses at definite time points are used as input variables for a multivariate calibration. These time points, which are for example displayed as dots and boxes in figure 20, have also been referred to as "virtual sensor array" in literature. In contrast to typical setups for chemical sensing, many variables

can be generated by one single sensor and not only 1 variable per sensor. The number of variables generated depends on the scanning speed (digitization speed) for the recorded sensor response and on how long the sensor responses are recorded. In contrast to some approaches, which try to fit a model to the sensor response and which use the model parameters as input variables, the scanning of the sensor response allows an accuracy only limited by the sensor recording system, whereas the model approach often has to deal with lacks of fit or instable parameters.

The time-resolved approach does not need to wait until an equilibrium state between the analyte and the analyte sorbed in the polymer has been reached (R22 in contrast to R134a in figure 22) proposed in static approaches [226]. This allows shorter measurements and the application of a broader variety of polymers and analytes.

In figure 22 it is also visible that the sensor responses of many time points are very similar. This means that the variables are highly correlated rendering the application of simple calibration methods like the MLR useless (see section 2.5).

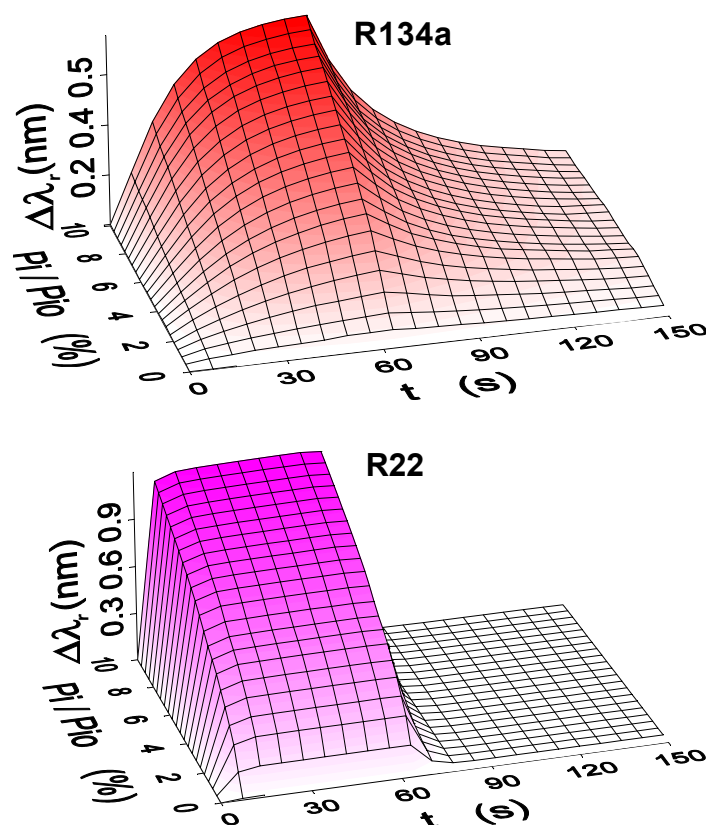


figure 22: SPR sensor response during 60 seconds of exposure to analyte and afterwards during exposure to synthetic air. The x-axis represents the time and the y-axis represents different relative saturation pressures of the analyte.

The experimental design of multivariate calibrations highly depends on the interactions of the analytes in mixtures with the sensitive layer compared to the interactions of the single analytes with the sensitive layer. A first approach is the assumption that the sensor responses of the analytes in mixtures are additive of the sensor responses of the single analytes. In this case, the experimental design only needs few levels of concentrations for a successful calibration. In figure 23, the sensor responses of the measured binary mixtures of the refrigerants (both of equal concentrations) minus the sensor responses of the single analytes of the corresponding concentrations are plotted versus the time and versus the total concentration of both analytes. It is visible that the plot strongly deviates from the ideal case of additive signals, which would result in a plane. The signal during exposure to analyte is lower than expected. This effect reinforces with higher concentrations of the analytes and can be ascribed to a saturation of the sensitive layer with analyte. Additionally a faster kinetics directly after the beginning and after the end of exposure to analyte is visible for higher concentrations of the analytes. The deviations from the signals being additive make an extensive calibration like a several-level full factorial design of the mixtures necessary.

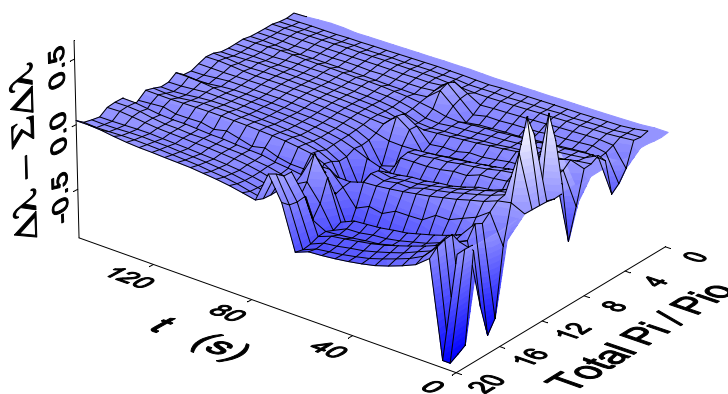


figure 23: Deviations from additive signals of both analytes when measured in mixtures of equal concentrations of both analytes.

A systematic variation of the size of the molecules with other molecular properties like polarity being similar can be found when a homologous series of analytes is investigated. In figure 24, the sensor responses of the SPR device are shown for 120 seconds of exposure to analyte of the homologous series of low alcohols and after that for exposure to air. Similar to the refrigerants, the shape of the sensor response changes from a rectangular to a wave with an increasing size of the alcohol molecules. The differences of the sorption and desorption kinetics should allow the quantification of quaternary mixtures using only the time resolved measurements of this single SPR sensor.

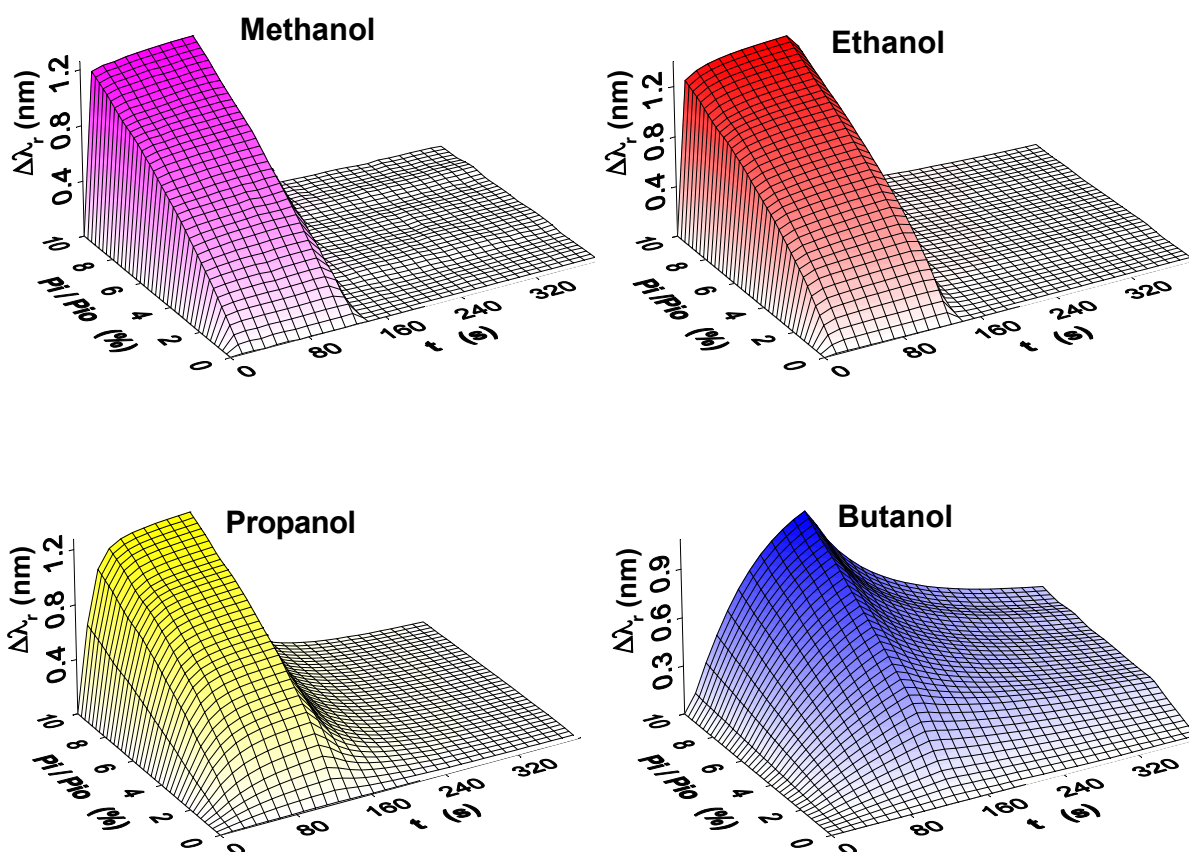


figure 24: SPR sensor response during 120 seconds of exposure to analyte and afterwards during exposure to synthetic air. The x-axis represents the time and the y-axis represents different relative saturation pressures of the analyte.

5.3.3. Thickness of the Sensitive Layer

In figure 25, the sorption and desorption of ethanol is shown for two Makrolon layers with a thickness of 100 nm and 300 nm. It is obvious that the thinner layer shows a significantly faster kinetics than the thicker layer. This effect can be explained by the different mechanisms of sorption proposed in section 5.3.2. A thicker layer means a longer diffusion path resulting in a higher proportion of the kinetic volume in the effective volume, which is consequently bigger. This effect allows tailoring the sensitive layer to a specific analytical problem. For example, the thicker layer allows an efficient discrimination of ethanol from the smaller and consequently "faster" methanol (see figure 15) whereas the thinner layer allows discriminating ethanol from the bigger and consequently slower 1-propanol and 1-butanol (see figure 24). The ability of tailoring the layers is only limited for very thick layers by the consumption of time needed for a measurement and for thin layers by a poor signal to noise ratio as only

few molecules can sorb into very thin layers. Further investigations of the influences of the thickness of the sensitive Makrolon layer can be found in [227].

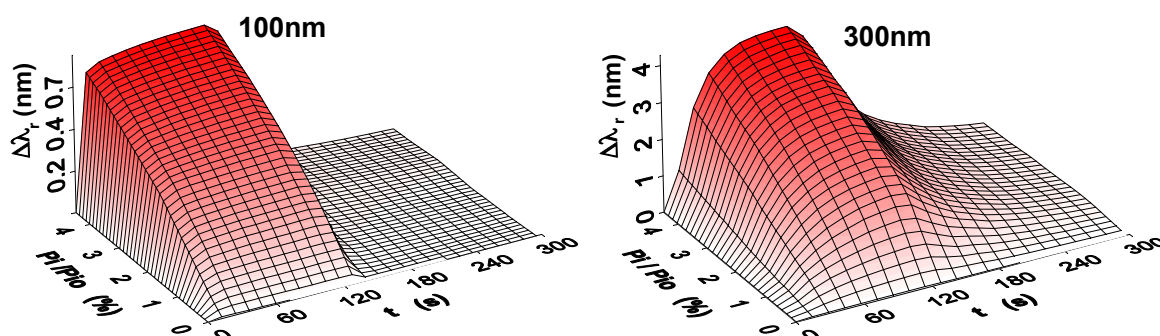


figure 25: Kinetics of sorption and desorption of ethanol for a thin and a thick sensitive layer.

5.3.4. Influence of the Carrier Gas

In section 5.3.1, it was shown that the Makrolon matrix expands when exposed to analyte for a long time. This effect can also be used for adapting the sensitive layer to a specific analytical problem by modifying the composition of the carrier gas. In figure 26, the kinetics of R134a is shown if pure air and air mixed with R22 is used as carrier gas. It is visible that even the quite low amount of adding 1% R22 to air as carrier gas significantly reduces the time needed for the desorption of R134a. In figure 27 the autoscaled signal of the R134a kinetics with pure air as carrier gas minus the autoscaled signal of the R134a kinetics in an air-R22 mixture is plotted. The positive difference in the plot demonstrates that the faster kinetics caused by the modified carrier gas plays a role for the desorption of highly concentrated R134a. Consequently, the addition of bigger analytes to air as carrier gas can be used to modify the desorption kinetics and to accelerate measurements. At the moment the accelerated kinetics of sorption and desorption can be best explained by the expansion of the micropores during the occupation of the pores by rather big molecules of the carrier gas and its additives (see also section 5.3.1). During the following exposure to analyte, the analyte molecules can sorb faster into the expanded micropores replacing the molecules of the carrier gas. Yet, further research on the exact mechanisms of the expansion of the polymer matrix and on the influence of different carrier gas mixtures has to be done. Up to now, it is only

certain that the concentrations and the sizes of the molecules of the additives to the carrier gas play a role.

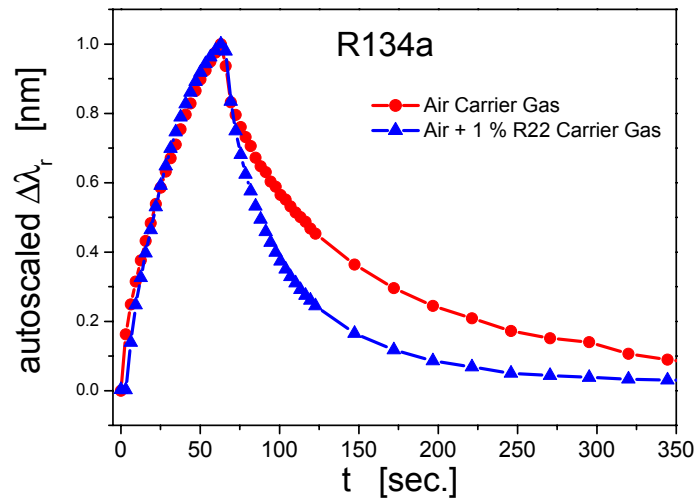


figure 26: Autoscaled signal of the sorption and desorption of R134a in pure air as carrier gas and in an air – R22 mixture as carrier gas.

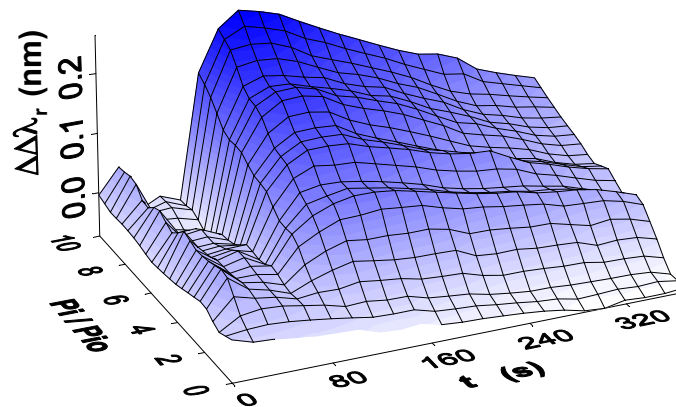


figure 27: The difference of the autoscaled signals of the sorption and desorption of R134a in pure air and in an air - R22 mixture is plotted versus the concentrations of R134a and the time (60 seconds of sorption and 290 seconds of desorption).

5.4. Conclusions

In this chapter, the principle of time-resolved measurements has been introduced as a new approach in chemical sensing. Thereby the time-resolved measurements performed in this work are based on Makrolon as sensitive polymer layer. The microporous structure of this polymer allows the discrimination of different analytes based on the size of the molecules. The combination of the polymer and a time-resolved recording of the sensor responses during sorption and desorption of the analytes allows a simultaneous quantification of a virtually unlimited number of analytes. The analytes to be quantified are only limited by too similar kinetics and by too slow kinetics whereby the kinetics can be modified by the type of carrier gas and by the thickness of the sensitive layer. The new time-resolved approach allows reducing the number of sensors to be used for an analytical problem. In this work, several analytical tasks are solved by the use of a single sensor setup rendering sensor arrays unnecessary, which would have been used for these analytical problems in the common approaches.

Additional research on other microporous polymers should allow the application of this approach to a broader spectrum of analytes. Yet, the time-resolved measurements are not limited to a size sensitive detection but can also be applied to other interaction principles. The time-resolved approach with a discrimination of the analytes based on exploiting the different shapes of the time-resolved sensor responses also dramatically changes the search for suitable sensitive layers for an analytical problem. Optimal sensitive layers of the common static sensor evaluation show most different sensitivity patterns for the analytes (like in section 3.3) whereas an optimal sensitive layer for the time-resolved sensor evaluation shows different shapes of the sensor responses for the different analytes.

In contrast to the sparse reports of time-resolved measurements in chemical sensing found in literature, the properties and interactions of the sensitive layer with the analytes is systematically investigated in this study allowing the tweaking of the kinetics described above. In the next chapters additional systematic investigations follow, which concern the exposure time, the recording speed of the of the sensor responses, the parallelization of the sensors and last but not least the multivariate data analysis rendering this work unique in respect to time-resolved measurements in chemical sensing.

6. Results – Multivariate Calibrations

In this chapter, the data sets of the refrigerants R22 and R134a, which were introduced in section 4.5.1.1, are investigated by the use of the most common methods of multivariate calibration starting with the PLS. Thereby models for the relationship between the 40 time-resolved sensor responses and the concentrations of both analytes are established. As the linear PLS calibration cannot deal with the nonlinearities present in the data sets, several methods, which are known to be capable of dealing with nonlinearities, are applied to this data set afterwards. These methods originate from different fields of scientific research such as multivariate spectroscopic calibration, quantitative structure activity relationship, machine learning, medical decision support systems, psychometrics, economic research and artificial intelligence whereby all these methods are able to calibrate multivariate relationships. An overview of the prediction errors for the calibration data and the validation data is shown in table 2 in section 6.11 for all methods used in this chapter. It is obvious that the calibration quality of the different methods shows a very broad variety ranging from unacceptable results for the widely used PLS calibration to promising results for neural network based calibrations.

6.1. PLS Calibration

PLS1 models were built for the calibration data set of the refrigerants (see section 4.5.1.1) with Martens' Uncertainty Test [32,33] for the determination of the insignificant variables and with the recommended determination of the number of principal components (factors) according to:

$$\text{Min}[V_{ytot_{PC0}} \cdot 0.01 \cdot a + V_{ytot_{PCa}}] \quad (25)$$

with a as the current dimensionality (principal component number), V_{ytot} as the total residual y -variance validated with a principal components respectively 0 principal components (total y -variance). This means that for each principal component 1% of the total y -variance is added to the variance at principal component number a . This ensures that only principal components are added, which significantly improve the model.

For R22 a model was calibrated with 2 principal components and for R134a a model with 3 principal components. Martens' Uncertainty Test removed no variable for R22 and only the time points 15 s and 67 s for R134a. In figure 28, the x and y loadings are shown for R134a. It

is visible that nearly all variables are highly correlated whereby the time points during desorption form a cluster (top right) and the time points during sorption form a second cluster (bottom left). The three dots near the origin of coordinates are the first 3 time points during sorption and the remaining 2 dots are the first 2 time-points during desorption. The corresponding loadings plot for R22 is analogous with only the location of the two clusters exchanged.

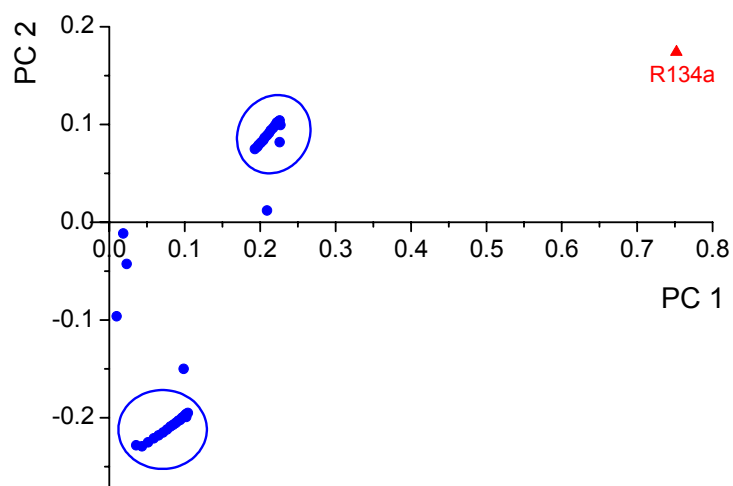


figure 28: Loadings plot of the x and y variables for the calibration of R134a.

The relative root mean square errors of the predictions of the calibration data are very high with 11.89% for R22 and 11.40% for R134a. The prediction of the independent validation set also showed disappointing high relative errors of 10.27% for R22 and 9.94% for R134a. In figure 29 the predictions of the validation data are shown as true-predicted plots. It is visible that the plots show a curvature and consequently the predictions are highly biased. The curved true-predicted plots are a typical sign of nonlinearities in the data, which are not successfully calibrated by a linear model. In figure 30, the T scores are plotted versus the U scores of the model of R134a for the first principal component, which explains 92.5% of the y -variance and 54.2% of the x -variance. The different concentration levels of R134a are visible along the axis of the U scores and the different concentrations of the interfering analyte R22 are visible as scattering along the axis of the T scores. As this type of plot shows the inner relationship of the PLS model (see expression (10)), the nonlinearities, which are visible in figure 30 and which are also visible in the less scattered T scores versus U scores plots for the higher principal components, demonstrate that the linear PLS model cannot deal with a nonlinear relationship present in the data. The corresponding plots for R22 are analogous and will not be discussed here.

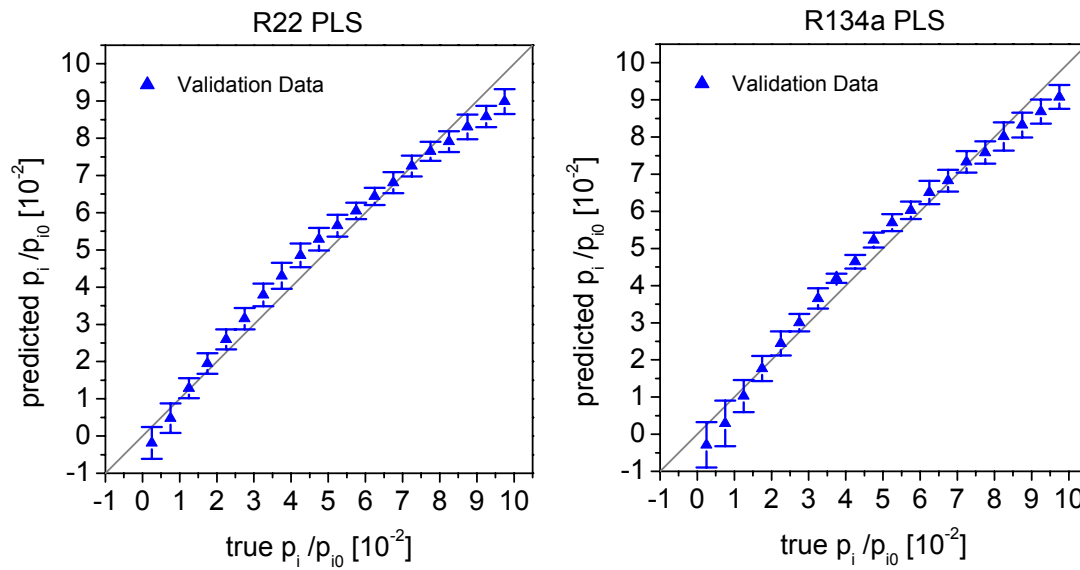


figure 29: True-predicted plots of the PLS calibration for the validation data. The optimal number of principal components was determined by Martens' Uncertainty. The predictions are poor due to systematic biases.

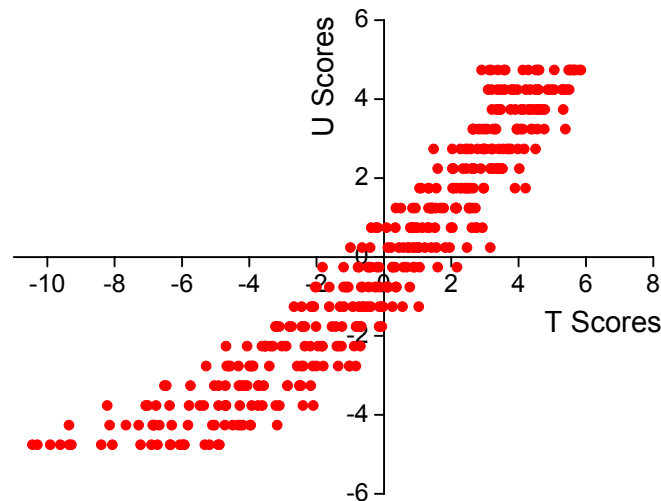


figure 30: T scores versus U scores for the first principal component of the model for the calibration of R134a.

It is known [39, 228-230] that the linear PLS sometimes can be successfully applied to nonlinear problems when minor important higher principal components are included into the model. These principal components may encapsulate not only noise but also significant information about the nonlinear relationship. Thus, the minimum crossvalidation error was used as selection criterion for the optimal number of principal components, as this criterion is less conservative in terms of the number of principal components compared to expression (25). The optimal models contain 5 components for R22 and 6 components for R134a, which were selected using the minima in figure 31. The calibration data were predicted with relative RMSE of 10.47% for R22 and 8.51% for R134a. The predictions of the validation data are

visualized in figure 32 with relative RMSE of 8.69% for R22 and 7.63% for R134a. Although the increasing number of principal components improved the errors of prediction, the plots show that the improvement can be put down to less scattered predictions but not to an improved calibration of the nonlinearities. The true-predicted plots can also explain why the predictions of the calibration data are worse than the predictions of the validation data. At both ends of the concentration range, the predictions are most biased. As the concentration range of the calibration data is wider ($p_i/p_{i0}=0-0.1$) than the concentration range of the validation data ($p_i/p_{i0}=0.05-0.095$) the high bias at the lower and upper concentration limit significantly increases the RMSE of the prediction of the complete calibration data set.

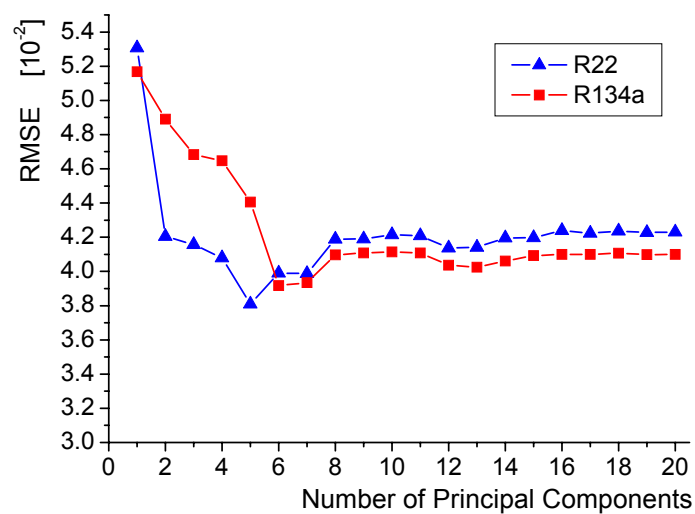


figure 31: Crossvalidated root mean square errors for the first 20 principal components for the calibration of R22 and R134a.

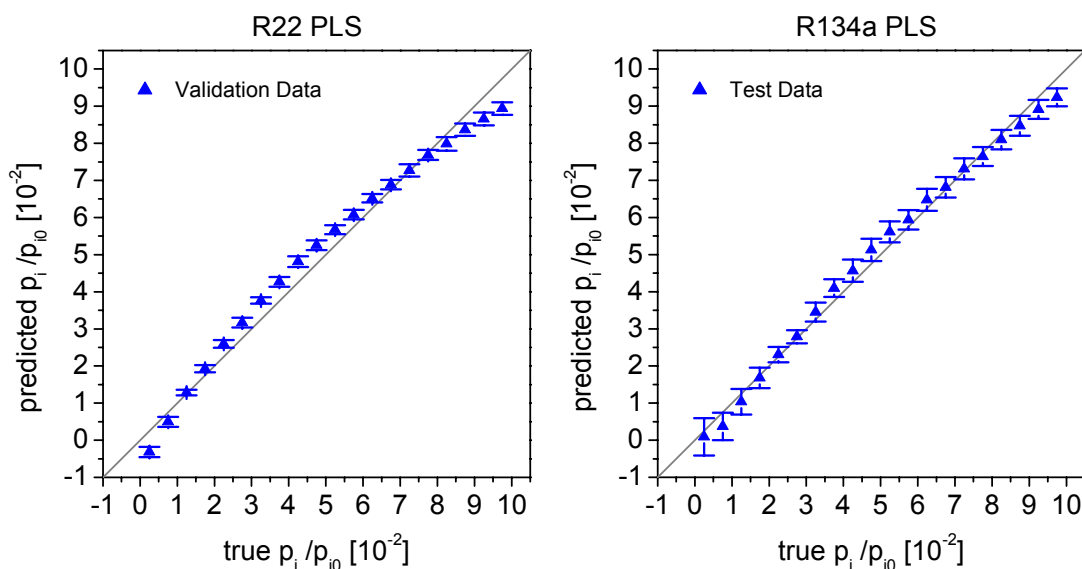


figure 32: True-predicted plots of the PLS for the validation data. The optimal number of principal components was determined by a full crossvalidation of the calibration data. The predictions are still rather poor visible as systematic biases.

In order to find a statistical basis to evaluate the ability of a method to calibrate nonlinearities, statistical methods can be used, which investigate the randomness of residuals of sequential observations. These tests like the Durbin-Watson statistics were originally developed for sequential observations equidistant in space or time. Centner et al. [231] showed that the Durbin-Watson Statistics can be used to test the residuals of purity data without a space or time component while implementing the test into a SIMPLISIMA algorithm for a mixture analysis. In this study, the mean residuals in figure 29 and figure 32 are treated with these statistics by using the increasing concentration levels as "virtual equidistant space component". The Durbin-Watson statistics is used to investigate the correlation of the residuals and additionally the Wald-Wolfowitz Runs test is used to test if the signs of the residuals occur in a random sequence.

Wald-Wolfowitz Runs Test

A good calibration of a relationship results in a random sequence of the positive and negative residuals of prediction. Calibrations with systematic errors show longer sequences of positive and negative residuals. For example, in figure 29 the mean residuals of R22 show only 3 sequences, first with negative signs, then with positive signs and finally with negative signs again. These sequences are also known as runs. The Wald-Wolfowitz method tests whether the number of runs is small enough or big enough for the null hypothesis of a random distribution of the signs to be rejected [232]. Thereby the number of positive and negative runs is compared with the tabulated value for the number of observations and for a given error level [233].

Durbin-Watson Statistics

The Durbin-Watson statistics can be used to test for a correlation (non-randomness) of the residuals [231,234,235]. The null hypothesis that there is no correlation between the successive residuals is evaluated by:

$$d = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2} \quad (26)$$

Thereby e_i is the residual corresponding to the observation i and e_{i-1} is the residual of the preceding observation. The d value is compared to the lower and upper critical values d_L and d_U proposed by Durbin and Watson [236]. If the calculated d is smaller than the tabulated lower critical value d_L , the null hypothesis is rejected meaning that the residuals are

correlated. If the calculated d is bigger than the tabulated upper critical value d_U , the correlation of the residuals is negligible and if the calculated d is between both tabulated critical values, the test is inconclusive.

For the residuals of the predictions of both analytes by all PLS models (Martens' Uncertainty Test and minimum crossvalidation criterion), the Wald-Wolfowitz Runs test ($p=0.000$) shows that the sequence of the residuals is highly non-random. The null hypotheses of the Durbin-Watson Statistics have to be rejected at the 5% error level for both analytes and both models indicating a correlation of the residuals. The combination of both statistics and the true-predicted plots indicate that the PLS cannot deal with nonlinearities, which are present in the relationship between the concentrations of the analytes and the signals of the device.

The nonlinearities, which are somehow visible in figure 22, can be made more prominent by plotting the first and second partial derivatives of the signals with respect to the relative saturation pressure versus the relative saturation pressure of the analyte R22 and versus time in figure 33. It is obvious that during exposure to R22 the first partial derivative is not constant along the concentration axis (y -axis) confirming that the relationship between the sensor signals and the relative saturation pressures of R22 is not linear. The partial derivative of the second order is neither zero nor constant. This means that the relationship also cannot be exactly described by polynomials with quadratic terms. The same nonlinear relationship between the signal and the relative saturation pressure can also be observed for R134a, but is not shown here. The nonlinearity can be ascribed to a saturation effect, as the number of micropores is limited (see discussion in section 3.2). For the high-concentrated mixtures, the nonlinearity is even worse since the total concentrations add up to $0.2 p_i/p_{i0}$. As for both analytes this nonlinearity can be observed and as a model for the nonlinear relationship between the concentrations and the signals is not known, several extensions to the linear models and several variable transformations are used for the calibration of this data set in the next sections. Additionally, "true nonlinear" methods, the neural networks, are applied to the data set in the last sections of this chapter.

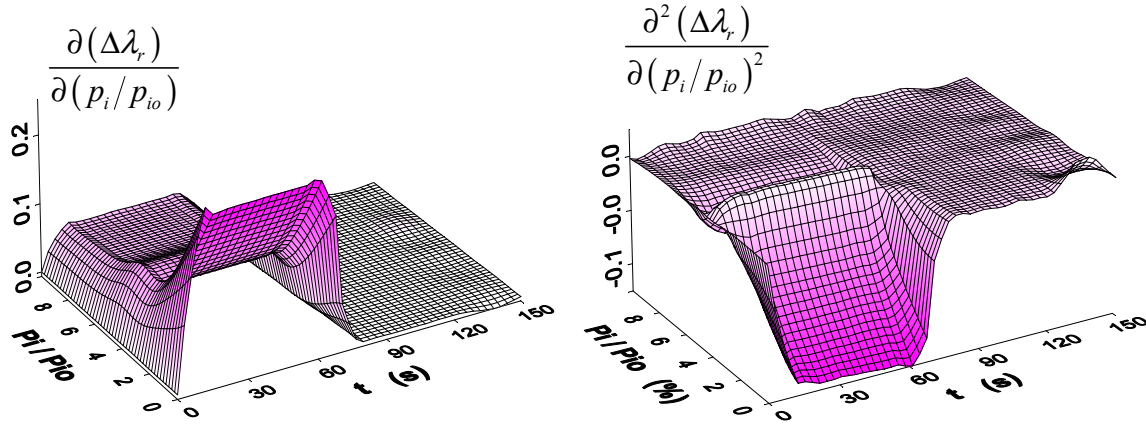


figure 33: First and second partial derivatives of the LOESS plot of figure 22 for R22 with respect to the relative saturation pressure.

6.2. Box-Cox Transformation + PLS

The Box-Cox transformation or power transformation is a general and widely used linearization procedure when no theory exists, which indicates that a certain transformation of the input and/or response variables will result in a more linear model [39]. The idea is to model a power of the response variable y as a linear function of x :

$$\frac{y^\lambda - 1}{\lambda} = b_0 + xb \quad (27)$$

The value of λ , which fits the linear function of x best, is estimated using the available data of the pure analytes. After the estimation of l and of the regression coefficients b and b_0 , the response variable can be transformed according to:

$$y = ((b_0 + xb)\lambda + 1)^{\frac{1}{\lambda}} \quad (28)$$

If $\lambda=0$, it is common to transform y according to:

$$y = \ln(y) \quad (29)$$

The Box-Cox transformation (27) was determined for the measurements of the single refrigerants of the calibration data set. For R22 $\lambda=0.68$ and for R134a $\lambda=0.74$ were estimated. Then the relative saturation pressures of the refrigerants of the calibration and the validation data were transformed according to expression (28). Similar to section 6.1 PLS models were

built for the transformed calibration data and then the validation data were predicted whereby the optimal number of principal components was determined by the minimum error of crossvalidation of the calibration data. The optimal model for R22 contained 11 principal components and the model for R134a used 10 principal components. The calibration data were predicted with a relative RMSE of 2.97% for R22 and 4.50% for R134a. The prediction of the validation data, which is also shown in figure 34 was performed with rel. RMSE of 3.09% for R22 and 5.04% for R134a. Both, the Durbin-Watson Statistics and the Wald-Wolfowitz Runs test are significant at the 5% error level. In figure 34 it is visible that the prediction of both analytes shows slightly a wave. Compared with the standard PLS the Box-Cox Transformation allows a highly improved calibration while a few nonlinearities remain uncalibrated. Although a rather high number of principal components are needed, only a slight overfitting can be observed, as the errors of the validation data are only moderately higher than the errors of the calibration data.

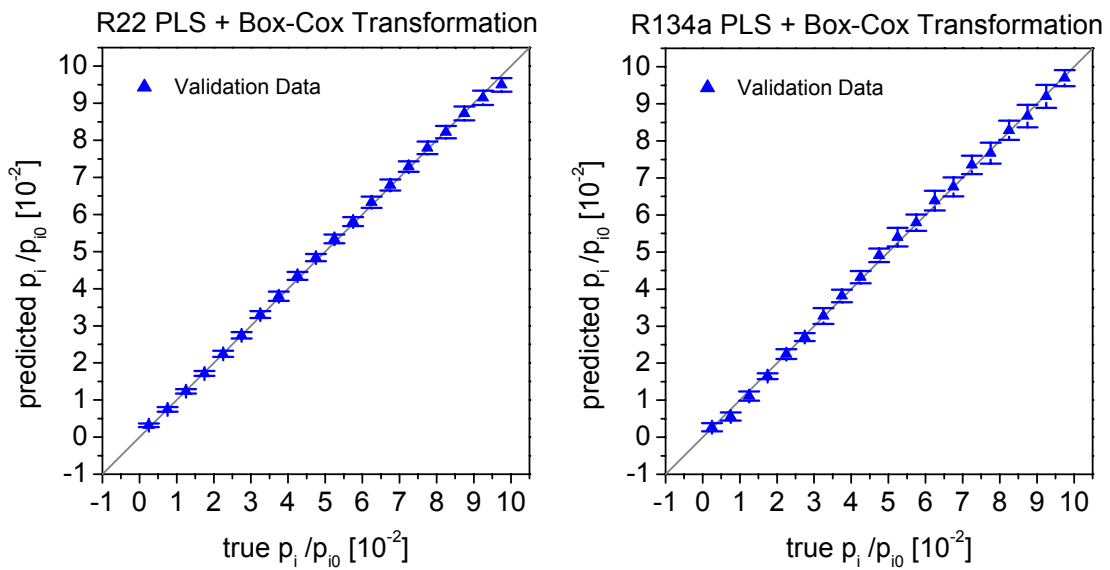


figure 34: True-predicted plots of the PLS for the validation data. The data were linearized by a Box-Cox transformation.

6.3. INLR

The Implicit Nonlinear PLS Regression (INLR) [237,238] is also called Nonlinear PLS in many publications. The INLR introduces nonlinearities into the regression model by adding squared terms ($x_1^2, x_2^2, x_3^2 \dots$) and optionally the cross-product terms ($x_1x_2, x_1x_3, x_2x_3 \dots$) to the set of "original" independent variables ($x_1, x_2, x_3 \dots$) [239]. For this study, only the squared terms were added as the addition of the cross-product terms would increase the number of independent variables to an unmanageable number of about $2 \cdot 10^{46}$. PLS models were built for the increased number of 100 independent variables with the optimal number of principal components selected by the minimum crossvalidation criterion.

The prediction of R22 by the optimal model with 16 principal components showed a relative error of 2.25% for the calibration data and 2.81% for the validation data. For R134a the optimal model with 17 principal components predicted the calibration data with a rel. RMSE of 3.47% and the validation data with a rel. RMSE of 4.02%. The addition of the squared variables can be also seen as a polynomial approach, which might explain why a rather many principal components are needed. This high number of principal components increases the relative gap between the error for the calibration data and the validation data compared with the Box-Cox Transformation and PLS due to the increased number of parameters (see also section 2.8.1). Yet, the INLR compensates the nonlinearities better than these two methods, as only for R22 the Wald-Wolfowitz Runs test and the Durbin-Watson Statistics are significant.

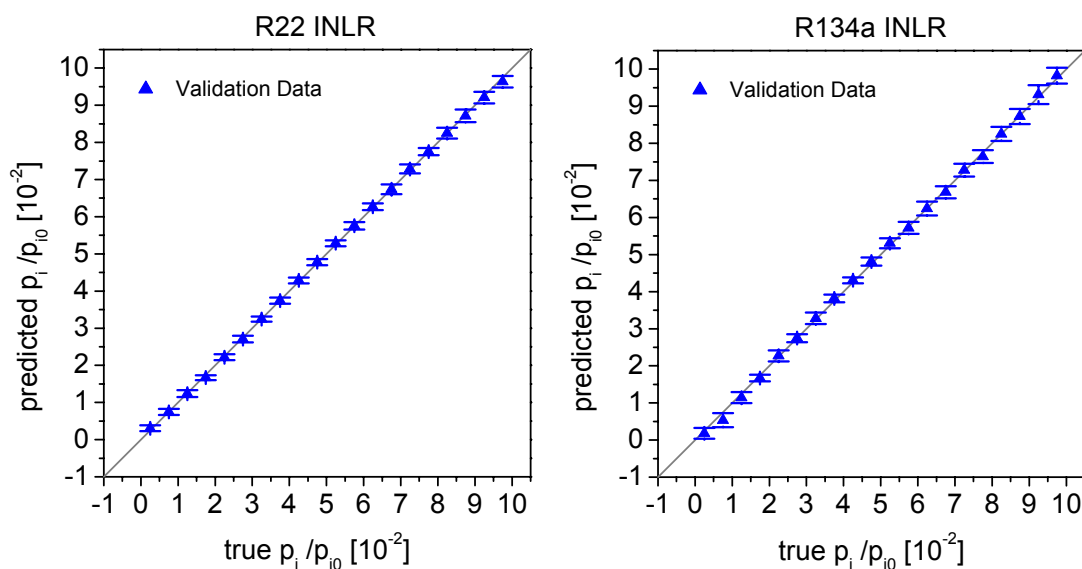


figure 35: True-predicted plots of the INLR for the validation data.

6.4. QPLS

In the field of QSAR research, a polynomial expansion of the inner relationship, which is linear in the original PLS, has become popular to model nonlinear relationships [38,228,240-247]. If the polynomial terms are of the second order, this approach is also known as QPLS (quadratic PLS). Instead of the linear relationship (10) between the score matrixes U and T , following polynomial expression is used:

$$U = C_0 + C_1T + C_2T^2 + H \quad (30)$$

The coefficients C_0 , C_1 and C_2 are determined by the least squares method in an iterative procedure similar to the PLS. For the calibration data, QPLS models were built with an optimal number of principal components determined by a minimum crossvalidation error of the calibration data. For R22 the optimal model with 2 principal components predicted the calibration data with a rel. RMSE of 2.31% and the validation data with a rel. RMSE of 2.41%. For R134a the optimal model with 4 principal components predicted the calibration data with a rel. RMSE of 3.87% and the validation data with a rel. RMSE of 3.92%. The sensibly low number of principal components is "rewarded" by practically identical validation errors and calibration errors. Both, the Wald-Wolfowitz Runs test and the Durbin-Watson Statistic cannot find a significant non-randomness in the prediction of the validation data. In combination with the true-predicted plots (figure 36) and the low errors of prediction, it is obvious that among the different PLS approaches the QPLS can deal best with the nonlinear data set. In the next sections, several non-PLS methods, which are also know to be able to account for nonlinearities, are applied to the calibration and validation data set.

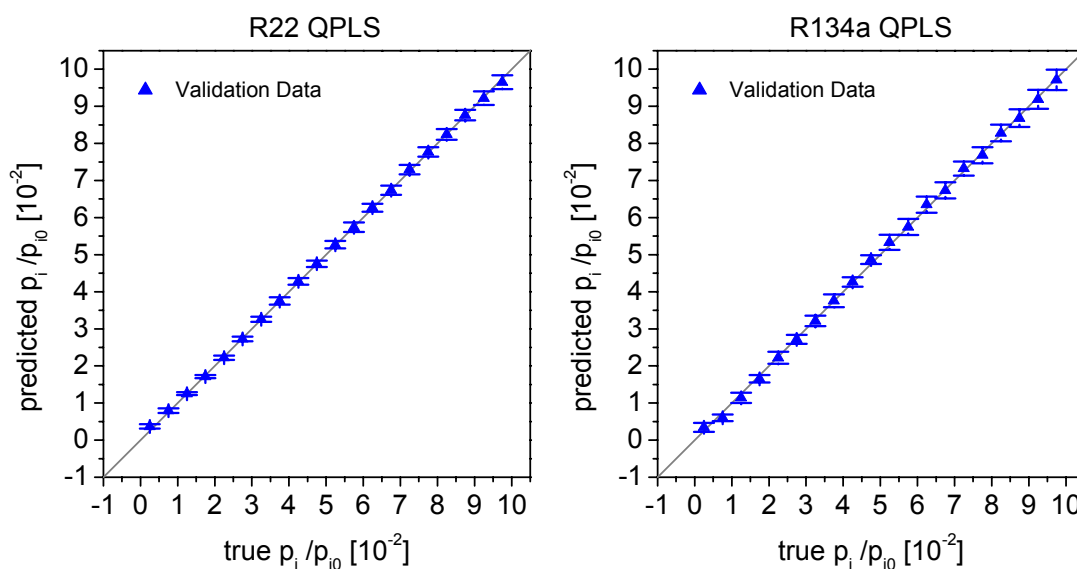


figure 36: True-predicted plots of the QPLS for the validation data.

6.5. CART

The classification and regression trees (CART) belong to the family of the decision trees, which are very common in the field of machine learning and medical decision support systems [248,249]. Thereby a classification or regression problem is split into sub-problems by a binary recursive partitioning procedure. In doing so, a tree is built, which consists of nodes that assign a subpart of the tree being responsible for a specific sample and of leaves that finally assign classes or regression values to specific samples. The CART principle is an automatic machine learning method, which can deal with nonlinear classification and regression problems. The tree building process is a two-step algorithm. In the first step, the tree is built by recursively partitioning the nodes into two child nodes starting with the root node. When trying to find the maximum average "purity" of the two child nodes during a partitioning process, the CART algorithm looks for the best input variable and the corresponding decision criterion by a brute force approach. For the regression, the average purity is calculated by the least squares of the response variables. In easier words, the CART algorithm tries to put similar samples into the same sub-trees. This first step stops when the tree cannot grow any more resulting in an overfitting. In the second step, child nodes are pruned away, which increase the error of the calibration data less than a "size corrected" value determined in a crossvalidation procedure. Finally, for the regression the mean of the calibration samples, which are passed through the tree to a specific terminal node (leaf), is assigned to the corresponding terminal node. The prediction of a new sample is performed by passing the sample through the tree and assigning the value of the corresponding leaf.

For the calibration data set of the refrigerants, trees were built, which are shown in figure 37. For both trees, there are more leaves than concentration levels (21) of each analyte in the calibration data. This is caused by the interference of the second analyte in the mixtures. Although the CART principle is quite simple, the predictions of the calibration data are quite good with relative RMSE of 3.81% for R22 and 4.85% for R134a (see table 2). Yet, the predictions of the external validation data show the major drawback of the CART principle for the application in regressions instead of classifications with relative RMSE of 8.79% for R22 and 11.20% for R134a. The gap between the prediction of the calibration data and the validation data is enormous and dwells in the transformation of the continuous response variables into discrete variables. The experimental design used for the calibration data contains 21 different concentrations for each analyte, which are learnt as discrete values by the regression trees. The concentrations of the validation data are exactly in the middle of two neighboring concentrations of the calibration data. Consequently, the trees assign one of these

two adjacent concentrations to the validation sample, which corresponds to a systematic quantization error of 5%. Thus, the gap between the validation error and the calibration error should be at least 5% for a 21-level experimental design. The rather high prediction errors of the validation data are visible in the true-predicted plots (figure 38) as rather high standard deviations. The absence of a bias in the true-predicted plots and the statistical tests for the residuals being not significant demonstrate that in principle CART is capable of accounting for the nonlinearities in the data although the quantization error renders CART impractical for this type of data.

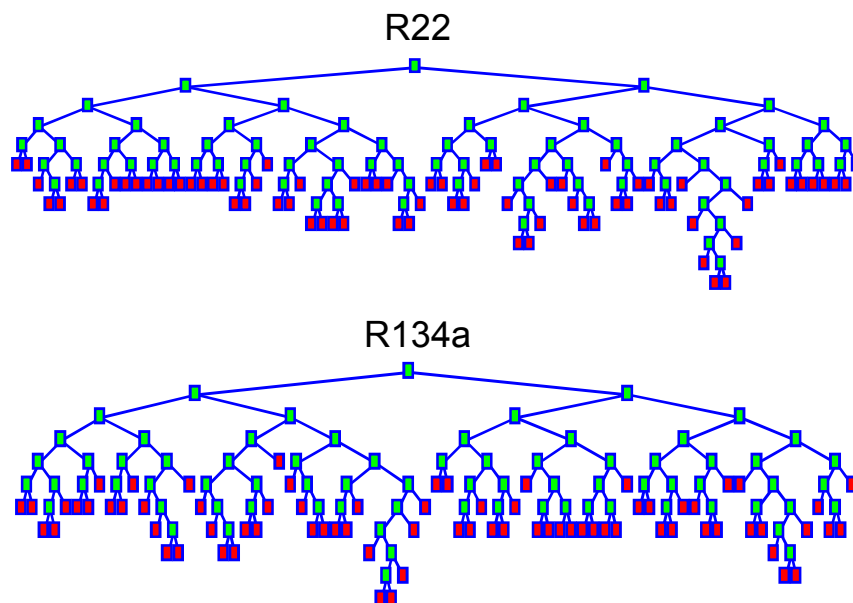


figure 37: Decision trees built by the CART for the calibration data with green nodes and read leaves.

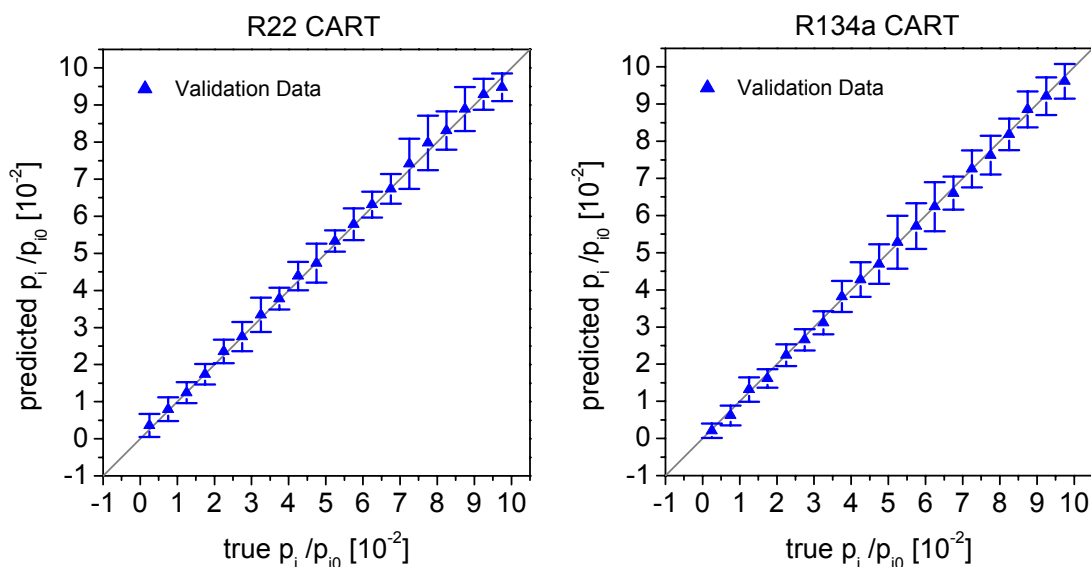


figure 38: True-predicted plots of the CART for the validation data.

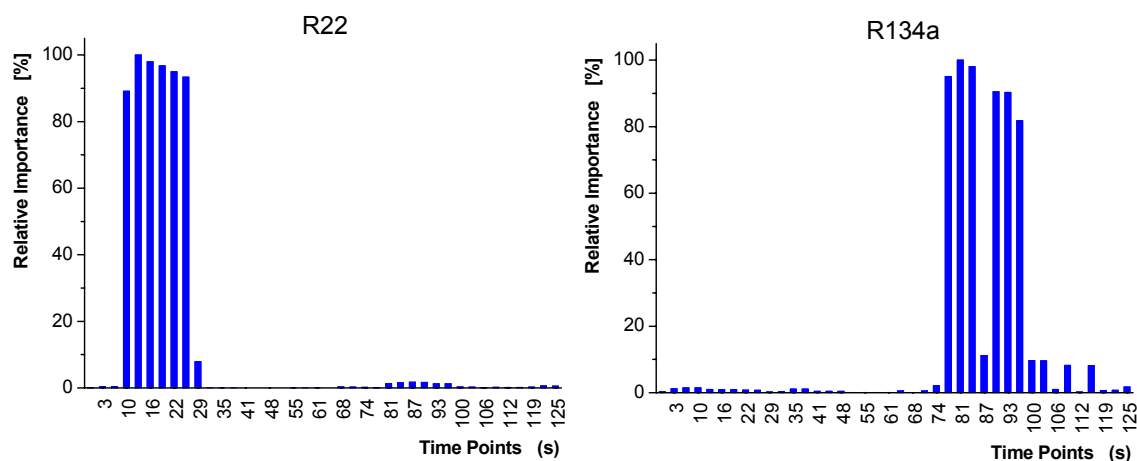


figure 39: Relative importance of the variables for the CART predictions measured by the frequency of being used for the prediction.

In figure 39 the relative importance of the variables for the decision tree is shown. The importance reflects how often the variables are used when all samples are passed through the decision tree. The two plots show that the 2 decision trees for the 2 analytes use complementary variables for the regression. The decision tree for R22 only uses variables during the sorption of analyte whereas the decision tree for R134a only uses variables during the desorption of analyte. When looking at figure 22, it is clear that the variables used for the prediction of R22 represent time points when mainly R22 has sorbed into the polymer whereas the variables used for the prediction of R134a reflect time points when all R22 has already desorbed and when practically pure R134a is left in the polymer.

The GIF-PLS, which was proposed by Berglund et al. [250] for the calibration of nonlinear data, also uses a transformation of continuous response variables into discrete variables. Consequently, the GIF-PLS approach is subject to the same quantization error and will not be investigated here any further.

6.6. Model Trees

The model trees are very similar to the CART principle and are often applied in the field of economic research [9,251]. Yet, each leaf contains a local linear regression model instead of a single discrete value for the samples passed to this leaf. Similar to CART an oversized tree is built in a first step. Thereby the optimal criterion for the splitting of a node is the minimization of the 2 standard deviations of the response variables of the samples assigned to the 2 child nodes. In the second step, a pruning of the subtrees is performed. Similar to the CART

procedure, the nodes and leaves are pruned, which increase the error of the calibration data less than a specified "size corrected" value. For the calibration data of the refrigerant data set, a tree with 33 nodes and 35 leaves was built for R22 and a tree with 29 nodes and 32 leaves was built for R134a. Both, the predictions of the validation data with relative RMSE of 7.19% for R22 and 7.59 % for R134a and the predictions of the validation data with relative RMSE for R22 of 10.29% and 11.20% for R134a were disappointing. In principle, the model trees should be superior to the regression trees as many local regression models are used instead of single discrete values. The true-predicted plots in figure 40 show that the predictions of the different concentration levels are rather inconsistent indicating differences of the quality of the various local linear regression models. This means that among the more than 30 local regression models per analyte not all models are calibrated well. The data set might be too limited in size to calibrate 30 linear regression models successfully with single local models spoiled by noise and outliers. Therefore, some local models are overfitted resulting in the significant increase of the prediction error of the validation data. In figure 40, no significant bias of the residuals can be detected in agreement with the statistical tests. The locally weighted regression (LWR) also uses the principle of many local linear regression models. In contrast to the model trees, which separate the sample space by a tree into local regions, the LWR generates local models at prediction time by weighting samples in the neighborhood more. As the principle of local regression models seems not to work for this highly correlated nonlinear refrigerant data set, LWR and other methods based on local model are not investigated any further.

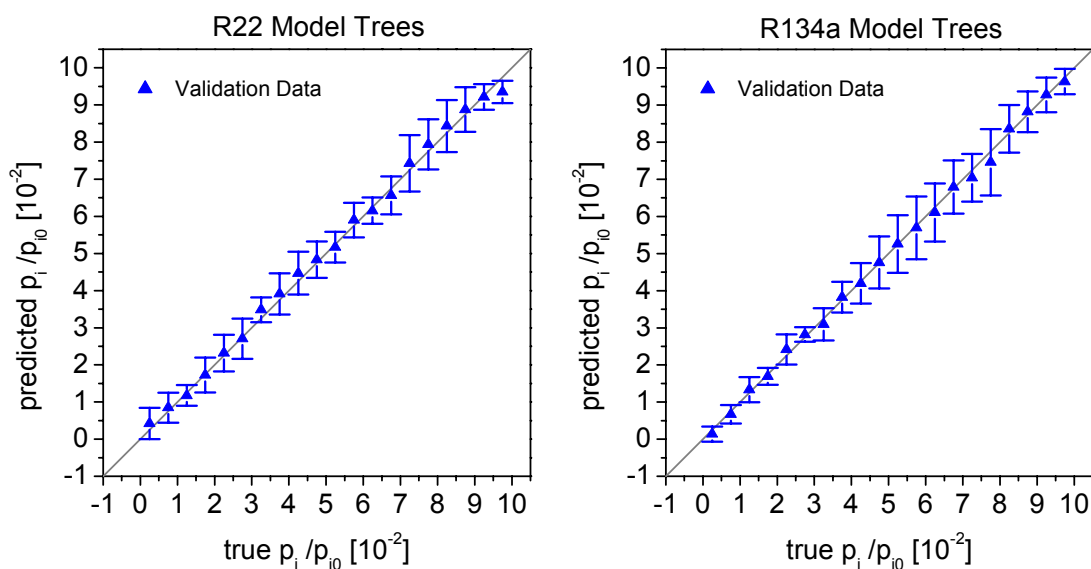


figure 40: True-predicted plots of the model trees for the validation data.

6.7. MARS

The multivariate adaptive regression splines (MARS) were introduced by Friedman [252,253] as a multivariate nonparametric regression procedure. The MARS procedure fits separate splines, which are also called basis functions, to distinct intervals of the input variables. The basis functions have the general form:

$$\text{BF1} = \max(0, x - a) \quad (31)$$

with BF1 as basis function, x as input variable and a as so-called knot. The transformation of the input variable is nonlinear, although the basis functions are piecewise linear. A regression using two basis functions can be described by:

$$y = b_0 + b_1 \cdot \text{BF1} + b_2 \cdot \text{BF2} \quad (32)$$

with y as response variable and b_0 , b_1 and b_2 as regression coefficients. Additionally, interactions up to a prescribed degree are also possible by the multiplication of two basis functions. The variables, the interactions and the locations of the knots are all found by a brute force approach and the regression coefficients are determined by a least squares procedure. The optimal model is found by a two-step algorithm similar to the CART principle. First, a model is grown by adding basis functions until an overfitting occurs. In the second phase, basis functions are deleted (pruned) until an optimal balance between overfitting and underfitting measured by the generalized crossvalidation error (GCV) has been reached for N samples and M basis functions:

$$GCV(M) = \frac{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2}{\left(1 - \frac{DOF(M)}{N}\right)^2} \quad (33)$$

$DOF(M)$ represents the degrees of freedom used by the basis functions. For linear regressions, $DOF(M)$ is usually set to M . Increasing DOF prefers smaller models.

The MARS principle was applied to the data set of the refrigerants. The models for R22 and R134a were built by the use of the calibration data. Thereby the optimal DOF was determined by a 10-fold crossvalidation implemented in the MARS package [253]. The degree of allowed interaction was systematically varied whereby the optimum for the crossvalidated calibration data was found allowing second order interactions.

For R22, the optimal MARS model contained 43 basis functions forming 3 additive and 27 interaction effects. In total 20 variables were used whereby the importance of the variables is shown in figure 41 measured by the relative amount of the reduction of the GCV by the corresponding variable. For R134a, the optimal model contained 43 basis functions forming 7 additive and 24 interaction effects. The relative importance of the 21 variables used by the model is also shown in figure 41. It is obvious that for both models the relative importance of the variables is very similar with the important variables forming two blocks after the beginning of exposure to analyte and after the end of exposure to analyte (>60 s). These blocks are similar to the blocks built by the CART, but in contrast to the CART both blocks are used for both analytes.

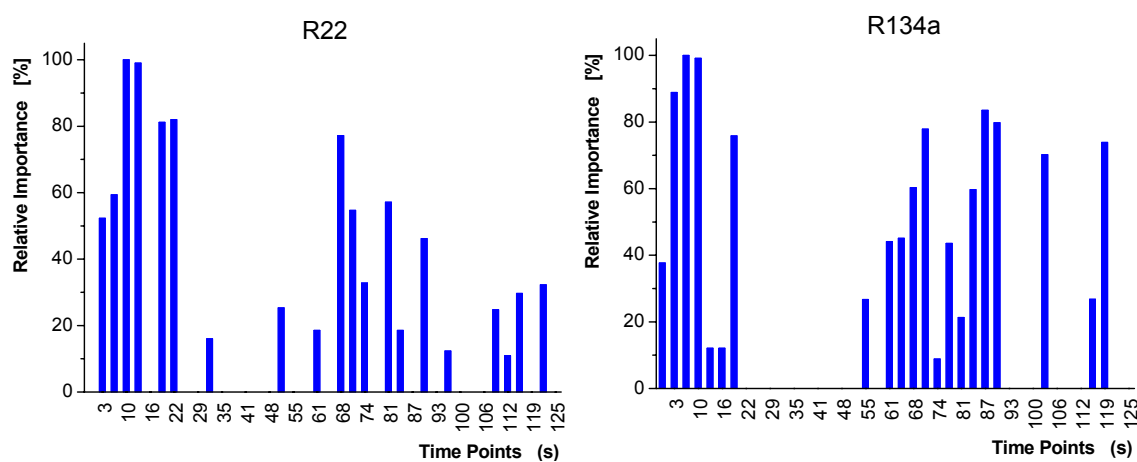


figure 41: Relative importance of the variables for the 2 MARS models measured by the reduction of the GCV.

According to table 2 the predictions of the calibration data are very promising with relative RMSE of 1.46% for R22 and 2.27% for R134a. The prediction errors of the validation data are significantly worse with 2.96% for R22 and 3.71% for R134a. The rather high numbers of basis functions used for models seem to overfit the calibration data. The true-predicted plots of the validation data in figure 42 demonstrate that the MARS deal well with the nonlinearities in the data and no significant bias of the predictions can be observed in agreement with the Wald-Wolfowitz Runs test and the Durbin-Watson statistics.

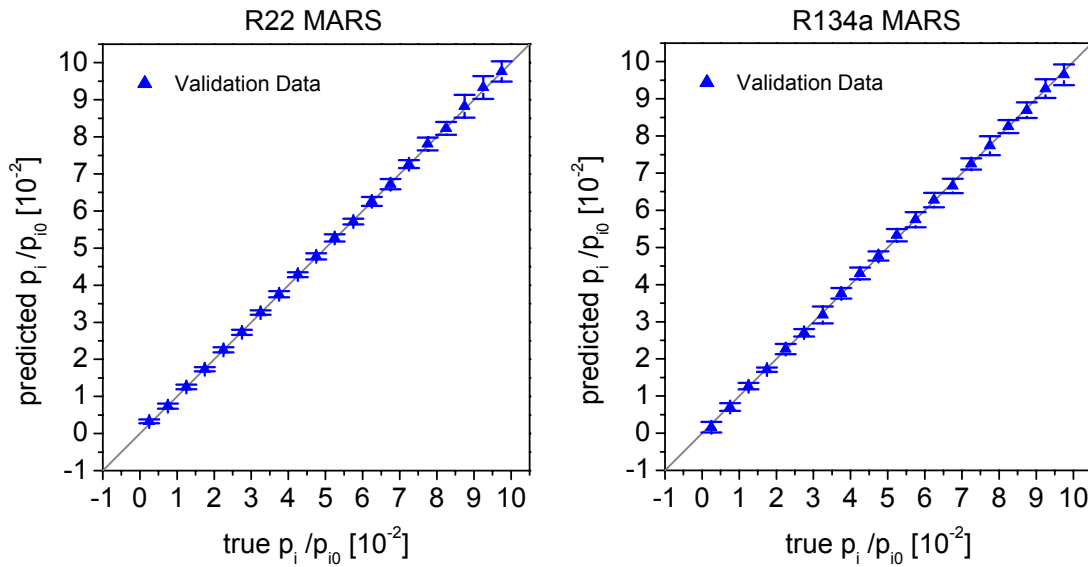


figure 42: True-predicted plots of the MARS for the validation data.

6.8. Neural Networks

In this section, uniform fully connected feedforward neural networks are used for the calibration of the refrigerant data set with the calibration data set used for the training of the neural networks and the validation data set predicted by the trained networks. Separate networks were used for both analytes. For the training, the SCG algorithm with early stopping was used (see section 2.7.3). More details of the implementation of the networks and the learning algorithms are described in the sections 2.7.2 and 2.7.3. Different network topologies were systematically investigated by varying the number of hidden layers and the number of neurons in the hidden layer. The best topology for both analytes evaluated by the lowest crossvalidation error of the calibration data set are fully connected networks with 6 hidden neurons organized in 1 hidden layer. Twenty networks of this topology were trained using different random initial weights and the network with the lowest crossvalidation error was used for the prediction of the validation data. According to table 2, the validation data were predicted with relative RMSE of 2.18% (R22) respectively 3.26% (R134a). The predictions of the validation data (see figure 43) do not show any type of nonlinearity in conformity with the Wald-Wolfowitz Runs test and the Durbin-Watson statistics. Among the different calibration methods used for the refrigerant data set so far, the neural networks can model the nonlinear relationships between the concentrations of the analytes and the time-resolved sensor signals best resulting in the lowest errors of prediction of the validation data. Yet, the calibration data were predicted with significantly lower relative RMSE of 1.47% (R22) respectively 2.62%

(R134a) indicating an overfitting. According to equation (13), the amount of 247 adjustable parameters of this network topology seems to be quite high for the calibration problem with many redundant or unused links resulting in overoptimistic low calibration errors but a decreased generalization ability. Thus, different methods like a variable selection and an optimization of the network topology are used in the next sections expecting that the reduction of the number of parameters further improves the generalization ability.

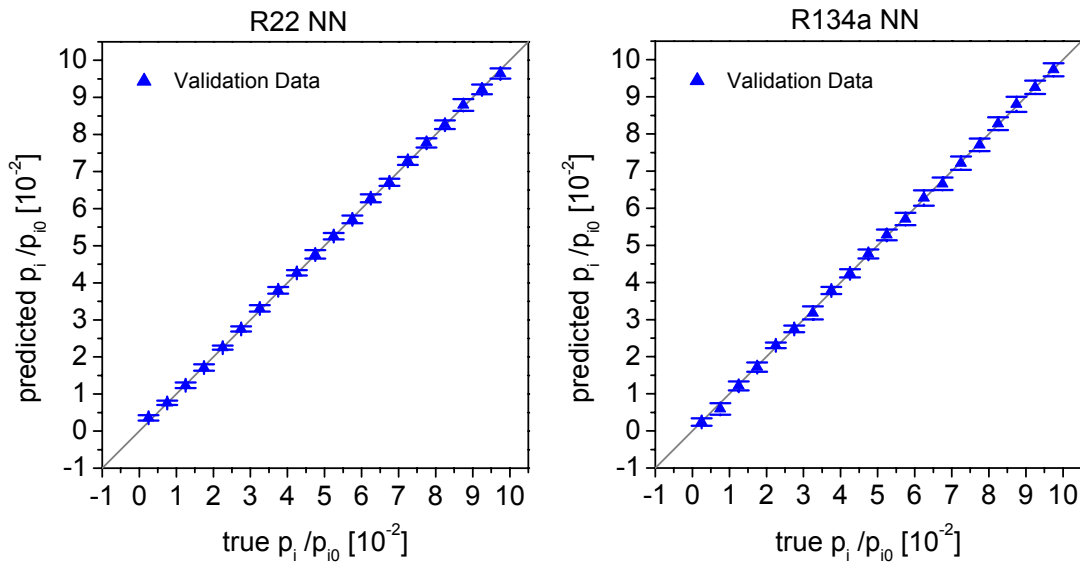


figure 43: Predictions of the validation data by fully connected neural networks.

6.9. PCA-NN

The combination of a principal component analysis with neural networks is a fast and efficient way of compressing the information fed to the neural networks. Yet, the decision how many principal components to use for the neural networks remains a problem as this determines the extent of compression and the extent of information loss similar to the PLS (see section 2.5). Thus, neural networks with 6 hidden neurons in 1 hidden layer were trained with a systematically increasing number of principal components from 1 to 40 and the cross-validation error of the calibration data was determined. The optimal models in terms of lowest crossvalidation errors were obtained by networks using 25 principal components for R22 and 16 principal components for R134a. According to table 2, the prediction errors of the external validation data (2.16% for R22 and 3.24% for R134a) are practical identical with the fully connected neural networks. In addition, the true-predicted plots and the statistical tests are practically identical and will not be discussed any further here. The negligible gap between

the calibration errors (1.98% for R22 and 3.08%) and the validation errors indicates that the reduction of the number of parameters (157 for R22 and 103 for R134a) successfully prevents an overfitting of the calibration data. Yet, the predictions of the validation data and with it the generalization ability are not significantly improved, which might be ascribed to some general drawbacks of the variable compression by the PCA already discussed in section 2.8.7.

6.10. Neural Networks and Pruning

For the pruning of neural networks, which is described in section 2.8.8 in detail, separate neural networks for both analytes were trained using the calibration data set. The networks were fully connected with 8 hidden neurons serving as reference networks for the pruning algorithms. Then, the two pruning algorithms Magnitude Based Pruning and Optimal Brain Surgeon were used to remove network links until the estimated increase of the error for the calibration data reached 2%. After that, the networks were retrained. This procedure was repeated 3 times in total. Finally, the calibration data and the external validation data were predicted. For both pruning algorithms, 50 networks were trained and optimized by this procedure using different initial random weights.

Magnitude Based Pruning

For R22, the network with the smallest crossvalidated calibration error consisted of 20 input neurons, 2 hidden neurons and 25 links. This network predicted the calibration data with a rel. RMSE of 2.34% and the validation data with a rel. RMSE of 2.48% (see table 2). For R134a the network with the smallest crossvalidated calibration error consisted of 33 input neurons, 3 hidden neurons and 64 links. The predictions by this network showed relative errors of 3.16% for the calibration data and 3.34% for the validation data. Compared with the fully connected neural networks, the number of adjustable parameters (27 respectively 67) were dramatically reduced resulting in a smaller gap between the prediction errors of the calibration data and the prediction errors of the validation error. Yet, the predictions of the validation data are worse than the predictions of the fully connected neural networks rendering this approach to improve the generalization ability of neural networks useless.

Optimal Brain Surgeon

For R22, the network with the smallest crossvalidated calibration error consisted of 25 input neurons, 3 hidden units and 37 links. This network predicted the calibration data with a rel. RMSE 2.10% and the validation data with a rel. RMSE of 2.12% (see table 2). For R134a the

network with the smallest crossvalidated calibration error consisted of 17 input neurons, 4 hidden neurons and 24 links. The predictions by this network showed relative errors of 3.22% for the calibration data and 3.32% for the validation data. The low number of adjustable parameters (40 respectively 26) successfully helped to prevent an overfitting with practically no gap between the predictions of the calibration and validation data visible. Compared with the fully connected neural networks the predictions of the validation data are slightly better for R22 and slightly worse for R134a. This demonstrates the possibility of modeling the relationship between the concentrations of the analytes and the time-resolved sensor responses using by far less adjustable parameters. It is also visible that the sophisticated OBS algorithm performs better than the simple MP approach.

Summary

The predictions of both pruning algorithms did not show unmodeled nonlinearities and the true-predicted plots were similar to the true-predicted plots of the fully connected networks (see figure 43). The most severe drawback of both pruning algorithms is the instability of the algorithms resulting in a totally different network topology for each run with different initial weights. For example, the 50 networks created by OBS for R22 used 7 to 27 input neurons, 1 to 4 hidden neurons and 8 to 40 links and showed prediction errors of the external validation data between 2.12% and 3.38%. The 50 networks for R134a used 8 to 36 input neurons, 2 to 6 hidden neurons and 12 to 49 links with no repeated topology. The predictions of the validation data varied between 3.32% and 5.48%. The variation of the networks created by the MP algorithm was even worse. Although the pruning algorithms demonstrated that significantly sparser network topologies are enough for modeling the relationships between the time-resolved sensor responses and the concentrations of the analytes, the high variations of the network topologies and of the qualities of prediction render the pruning approach useless for an easy reproducible application.

6.11. Conclusions

In this chapter, several multivariate calibration methods were used for the quantification of the refrigerants R22 and R134a in mixtures measured by a single sensor SPR setup. It was demonstrated that the combination of the time-resolved measurement approach with a multivariate data analysis allows the quantification of 2 analytes using only 1 single sensor.

Yet, the most common multivariate calibration method, PLS, was not capable of dealing with the nonlinear relationship between the sensor signals and the concentrations of the analytes resulting in unacceptably high prediction errors and systematical biases. Thus, different well-known methods for linearizing the data or for introducing nonlinearities into linear models were used. In principle, these methods successfully compensated the nonlinearities in the data structure. Yet, compared with the standard deviations of the sensor signals with 0.2% for R22 and 1.3% for R134a, which were calculated for 60 seconds exposure to analyte of different concentrations using reproduced measurements, the overall quality of predictions seems to be improvable.

The application of uniform fully connected neural networks showed the best results with respect to generalization ability. In contrast to the methods mentioned before, the neural networks make no assumption of the type of relationship between the input and the response variables (linear, quadratic...) and thus can approximate the relationship between the time-resolved sensor responses and the concentrations of the analytes quite well, for which no model exists at the moment. A significant drawback, which can also be seen in table 2, is the overfitting of the neural networks, which is observable as gap between the prediction errors of the calibration data and of the validation data. In contrast to all other methods mentioned before, the neural networks perform no variable selection or compression of the input information resulting in a high number of adjustable parameters and consequently running into the danger of overfitting. Therefore, neural networks combined with a principal component analysis were used for the compression of the input variables and two pruning algorithms were applied for thinning out the network structure. Whereas the pruning algorithms showed unstable and worse results, the combination of the PCA with neural networks demonstrated that in principle smaller neural network could calibrate the relationships at least as well as the non-optimized networks. As the PCA-NN is a very simple method to reduce the neural network size, a more sophisticated variable selection method is expected to show even better results. Thus, in the next chapter genetic algorithms are

combined with neural networks to perform a variable selection and thus to reduce the network size.

Method	Calibration Data Set		Validation Data Set		Non-linearity	Over-fitting
	R22	R134a	R22	R134a		
PLS (Martens' Uncertainty)	11.89	11.40	10.27	9.94	High	No
PLS (Min. Crossvalidation)	10.47	8.51	8.69	7.63	High	No
Box-Cox Transformation	2.97	4.50	3.09	5.04	Low	Medium
INLR	2.25	2.81	3.47	4.02	No	Medium
QPLS	2.31	3.87	2.41	3.92	No	No
CART	3.81	4.85	8.79	11.20	No	High
Model Trees	7.19	7.59	10.29	11.20	No	Medium
MARS	1.46	2.27	2.96	3.71	No	Medium
Neural Networks	1.47	2.62	2.18	3.26	No	Medium
PCA - NN	1.98	3.08	2.16	3.24	No	No
MAG-Pruning - NN	2.34	3.16	2.48	3.34	No	No
OBS-Pruning - NN	2.10	3.22	2.12	3.32	No	No

table 2: Comparison of the rel. RMSE of the calibration and validation data in %. Additionally the degrees of nonlinearity and overfitting of the predictions are listed.

7. Results – Genetic Algorithm Framework

In the previous chapter, it was demonstrated that neural networks show the best calibration for the refrigerant data set due to the nonlinearities present in the data. It was also shown that a compression of the input variables by a simple combination of a PCA and neural networks shows results comparable with the neural networks using all variables. Therefore, it is expected that a more sophisticated selection of the input variables can improve the generalization ability and thus the calibration quality. Hence, a genetic algorithm for the variable selection is combined with neural networks for the calibration in this chapter. As single applications of this combination neither show superior calibrations nor reproducible variable selections, a framework is setup, which uses many parallel runs of the genetic algorithm for different data subsets resulting in improved calibrations and a high reproducibility.

7.1. Single Run Genetic Algorithm

For the variable selection, a combination of a genetic algorithm and neural networks described in section 2.8.5 was used. For the evaluation of the fitness function (16), the calibration data set of the refrigerant measurements (see section 4.5.1.1) was randomly split into a calibration (75%) and a test data subset (25%). The neural networks were fully connected with 4 hidden neurons and 2 output neurons (1 network for both analytes together). The genetic algorithm evaluated 50 populations during 76 generations whereas the stopping criterion was set to a convergence of the standard deviation of the genes below 0.04. The parameter α of the fitness function was set to 0.9, which resulted in the selection of 8 time points (0, 12, 15, 51, 67, 93, 122 and 125 seconds) as most dominant solution in the last generation. The corresponding neural network (8 hidden neurons, fully connected and 1 output neuron) predicted the test data subset with excellent low rel. RMSE of 1.87% for R22 and 2.50% for R134a. Yet, the prediction of the external validation data by this network, which had been trained using the complete calibration data set, shows RMSE of 2.32% for R22 and 2.93% for R134a comparable with the non-optimized neural networks using all time points (see table 3 in section 7.4). A second run of the genetic algorithm using a different partitioning of the calibration data into calibration and test data subsets showed even worse results. After 86 generations 8 time points (0, 3, 6, 51, 74, 90, 115 and 125 seconds) were selected with rel. RMSE of 1.84% (R22) and 2.62% (R134a) for the prediction of the test subsets. The prediction of the external validation data showed disappointing high errors of

2.63% for R22 and 3.35% R134a (see table 3). For both runs, the predictions of the external validation data are significantly worse compared with the test data subset used for the evaluation of the fitness for the genetic optimization. Additionally, the selection of the time points is not reproducible. This instability of the variable selection can also be seen in figure 46, which shows the frequency of the time points being selected during 100 runs of the GA. Although some time points are more often selected than other time points, there is no time point, which was never selected. Both findings, the instability of the variable selection and the deterioration of the prediction ability for external validation data can be ascribed to a general problem of single run genetic algorithms. The variables are selected on the basis of a fitness function with a static test and calibration data set. Consequently, the optimal solution is only valid for one individual partitioning of the data into calibration and test data subsets and is not representative for the complete data set. Although the fitness function (16) tries to compensate for the overestimation of the test data by partly considering the calibration data (in contrast to most GA found in literature), the drawbacks of a static partitioning cannot be completely compensated. Apart from these problems known in literature (approximately 99% of all GA are based on static data sets), the single run algorithms are faced by additional problems:

1. Both, the chromosomes of the initial population and the weights of the neural networks are randomly generated. As there is no guarantee that the walk of the genetic algorithm in the search space, which also contains random steps, can always find the best subset of variables before converging, different runs (even with identical test and calibration data subsets) often find similar but not exactly identical subsets of variables [254].
2. Jouan-Rimbaud et al. [255] recently demonstrated that by chance correlation of variables often irrelevant variables are selected by GA or have at least a significant influence on the final model, even if validation procedures are used.

7.2. Genetic Algorithm Framework - Theory

In recent literature, several approaches are reported to solve the different problems of single runs of GA:

1. Massart and Leardi [98,256] use a very refined algorithm for the variable selection, which is based on parallel runs of many GA with different combinations of test and calibration data. Then a validation step is performed to find the best variable subset. The GA is a hybrid algorithm using a stepwise backward elimination of variables to find the smallest possible subset of variables. Although this approach is very promising, Jouan-Rimbau et al. [255] showed that this algorithm is still partly subject to chance correlation.
2. In [99] Leardi et al. use 100 runs of GA with the same calibration and test data sets. The final model is obtained by adding systematically the variables, which are ranked according to the frequency of selection of the GA runs and by using the combination with the smallest error of prediction. In [97] this algorithm is modified by the different GA runs learning from each other.
3. In [126] the predictions are averaged by several models found by different GA runs. Yet, the average prediction was not better than the prediction by a single model.
4. In [254] 10 runs of GA are performed by using different calibration and test data subsets. The final model uses all variables, which were selected at least 5 times, whereby this limit is rather arbitrary.

The genetic algorithm framework proposed in this work picks several elements of the studies mentioned above and is presented in the flow diagram in figure 44. The framework can be divided into three steps. The first step consists of multiple parallel runs of the GA presented in section 2.8.9 and in section 7.1 using different calibration and test data subsets (yellow boxes in the flow diagram). Variables, which are represented higher than average in the final population of each GA run, are collected over all GA runs and are ranked according to the frequency of appearance in the final populations. The second step of the framework finally selects the variables in an iterative procedure by adding the variables to the neural network model according to their rank in a stepwise procedure. The neural network is evaluated by the use of different calibration and test data subsets (green boxes in figure 44). The RMSE of prediction of all test data sets is compared with the RMSE of the previous model. If the

RMSE is lower (see section 10.2), the last variable is accepted and the procedure is repeated adding the next important variable until the predictions are not improved any more.

In the third step, the final neural network topology is determined. First, the number of hidden neurons of a single hidden layer is optimized in an iterative procedure, which is shown in figure 45. Starting with fully connected neural networks with 1 hidden neuron additional fully connected neurons are added until the error of prediction of the test data doesn't improve any more, whereby the l different test data subsets are generated by a data subsampling procedure. Finally, this neural network topology is trained with the complete data set several times, and the neural net with the smallest error of crossvalidation should be used as final optimized model and should be validated by an external data set not used during the complete variable selection algorithm.

In all three major steps of the framework, the complete data set is split several times into a calibration (75 %) and a test (25 %) subset, which was done by a random subsampling procedure (see section 2.4) resulting in rather pessimistic predictions of the test data. Consequently, according to expression (16) models are preferred, which are more predictive and which yield a better interpolation.

As already stated in section 2.8.5, the choice of α in the fitness function (16) influences the numbers of variables being selected during each run of a GA. A too high value of α ignores partly the accuracy of the neural nets and ends in only few variables being selected. Consequently, there might be too few variables selected in the first step to be added to the neural net in the second step. This problem can be recognized by all variables with a ranking higher than "0" being used for the neural net in the second step. On the other side, a too low value of α results in too many variables being selected. This can be detected by the absence of a differentiation of the variables in the ranking. An empirical way to select an optimal α is based on running a single GA with different values of α and on choosing that α , which results in the selection of the number of variables expected to be needed for the calibration. A good choice to start with is setting α to "1" for these single runs of the GA. Yet, preliminary studies showed that the parallel runs of the GA make the framework quite robust towards the choice of α and to the population size, which is suggested to be set to the number of variables to select from. Although the framework seems to be complex on the first sight, this robustness renders the algorithm quite user-friendly.

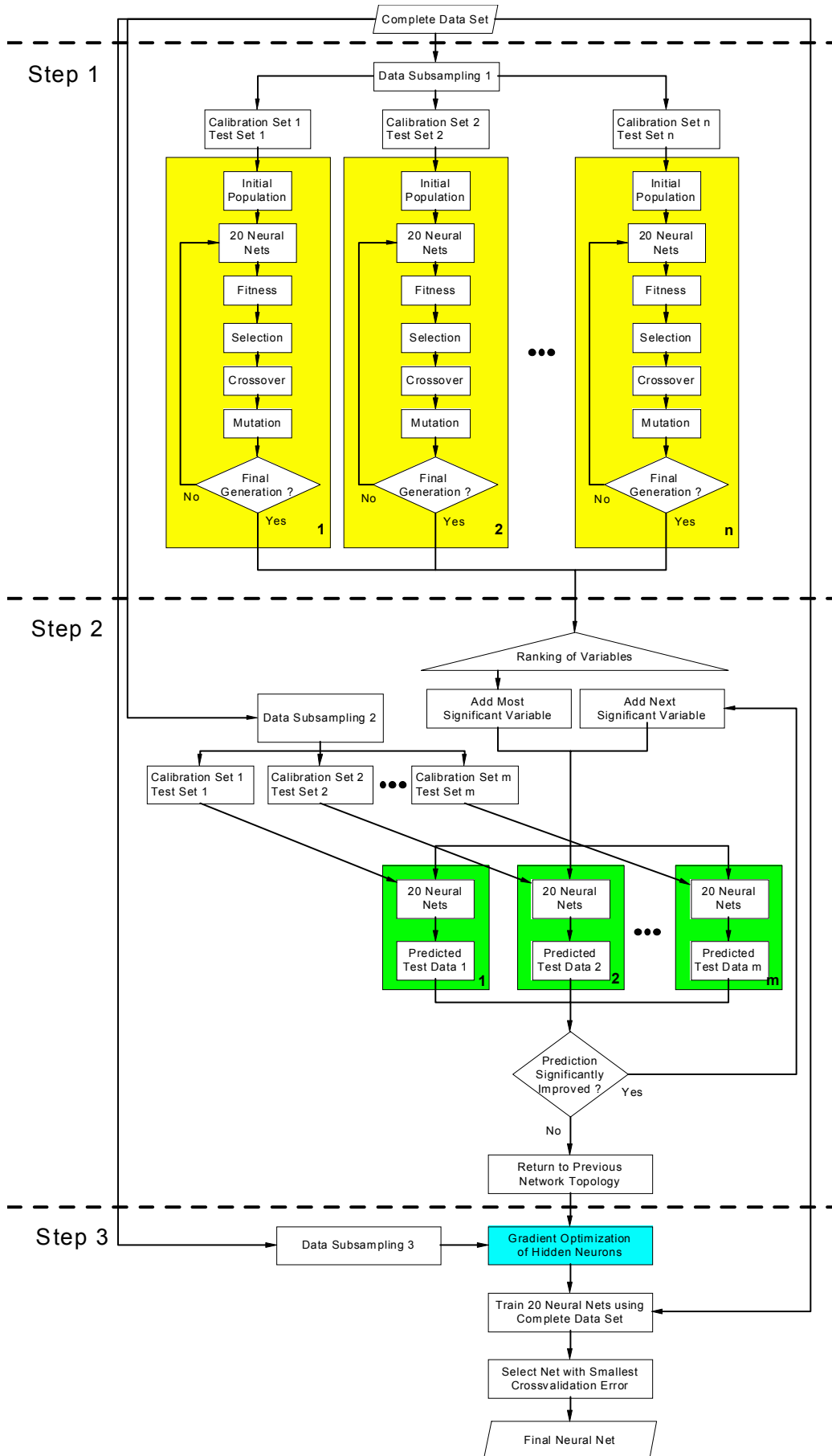


figure 44: Flow chart of the genetic algorithm framework.

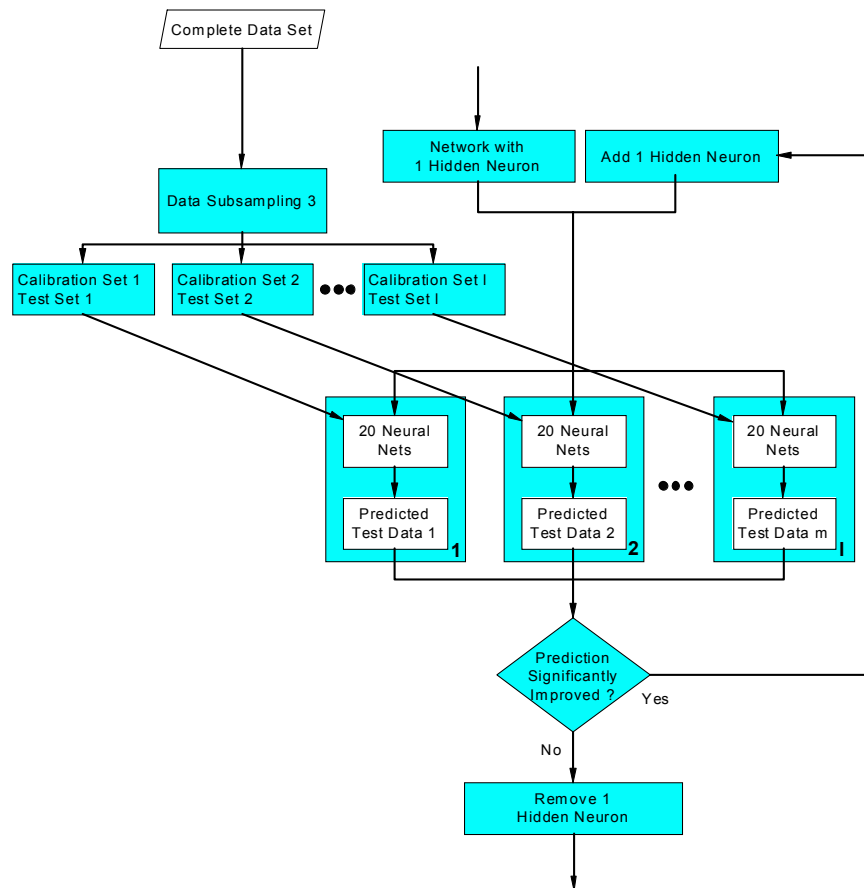


figure 45: Optimization of the number of hidden neurons. This figure is a detailed flow chart of the blue box of the genetic algorithm framework shown in figure 44.

7.3. Genetic Algorithm Framework - Results

The genetic algorithm framework was applied to the calibration data of the refrigerant data set. Thereby 100 parallel runs of the GA were used with the same settings of the parameters as described in section 7.1. The results of the first step of the global algorithm are shown in figure 46. Thereby the ranking of the variables is shown as frequency of the variables being present in the last population of the genetic algorithms. In the second step, these variables enter the model according to their rank until the prediction of the test data of a 20-fold random subsampling does not improve any more. The iterative procedure stopped after the addition of 10 time points, which are labeled in figure 46. As the labels are rounded seconds of the time points, the most important time point "0" does not represent the absolute beginning of the measurement, but 0.3 seconds after the beginning of exposure to analyte.

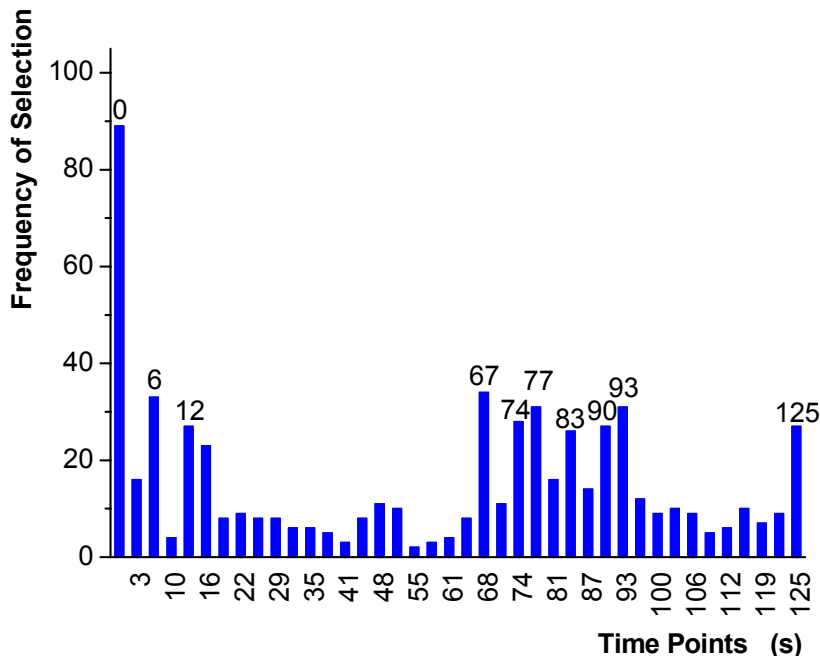


figure 46: Frequency of the selection of the time points for 100 parallel runs of the genetic algorithms. The 10 time points selected by the algorithm are labeled additionally.

The optimized networks (10 input neurons, 8 hidden neurons, 1 output neuron, fully connected) predicted the test data of the 20-fold random subsampling procedure (figure 45) with rather low rel. RMSE of 1.94% (R22) and 3.05% (R134a). The predictions of the external validation data by these networks, which had been trained by the complete calibration data set, were best of all methods used so far with 2.04% for R22 and 2.89% for R134a (see table 3 in section 7.4). Practically no gap between the prediction of the calibration and validation data is noticeable indicating much more stable models compared with the non-

optimized networks. The predictions of the validation data, which are shown in the true-predicted plots in figure 47, are not biased and hardly scattered.

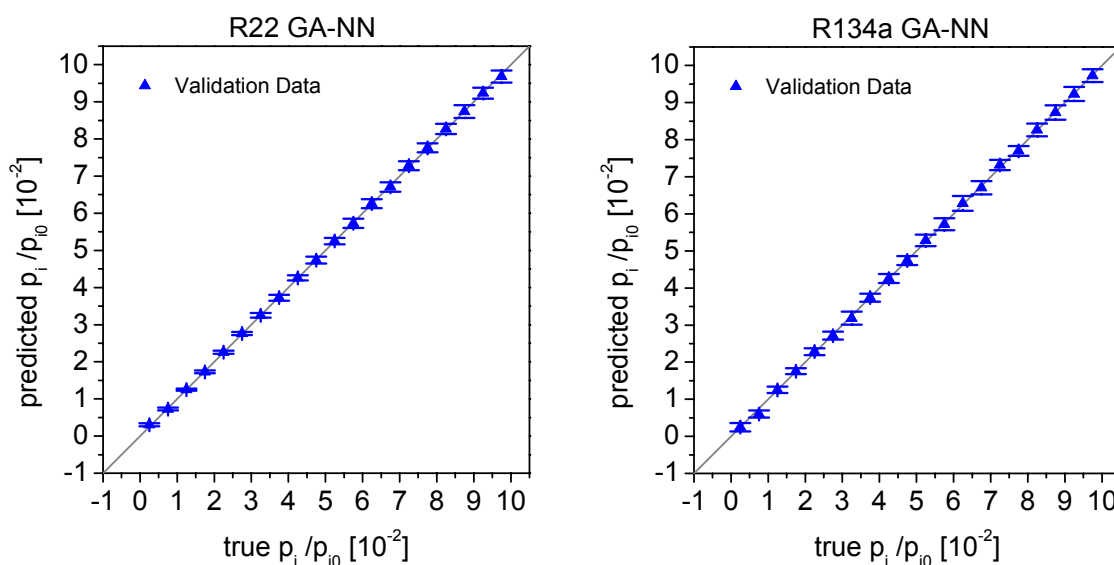


figure 47: Predictions of the validation data by neural networks optimized by the genetic algorithm framework.

The residuals of the predictions of the neural networks were further examined in respect to the compositions of the predicted analyte concentrations. In figure 48, the absolute residuals of the predictions of the analyte concentrations are plotted versus the corresponding analyte concentrations of the predicted sample. The plot for R22 demonstrates that the absolute residual for the prediction of R22 increases with an increasing concentration of R22, but is practically randomly distributed along the axis representing the concentration of R134a. For R134a the plot shows that the residuals of the predictions of R134 show a higher dependency on the concentration of R134a than on the concentration of R22. This means that the concentration of the interfering analyte does practically not influence the prediction quality of the analyte of interest. Thus, it should be possible that the system of the time-resolved measurements, the variable selection and the calibration by neural networks can be extended to parallel quantifications of even more analytes.

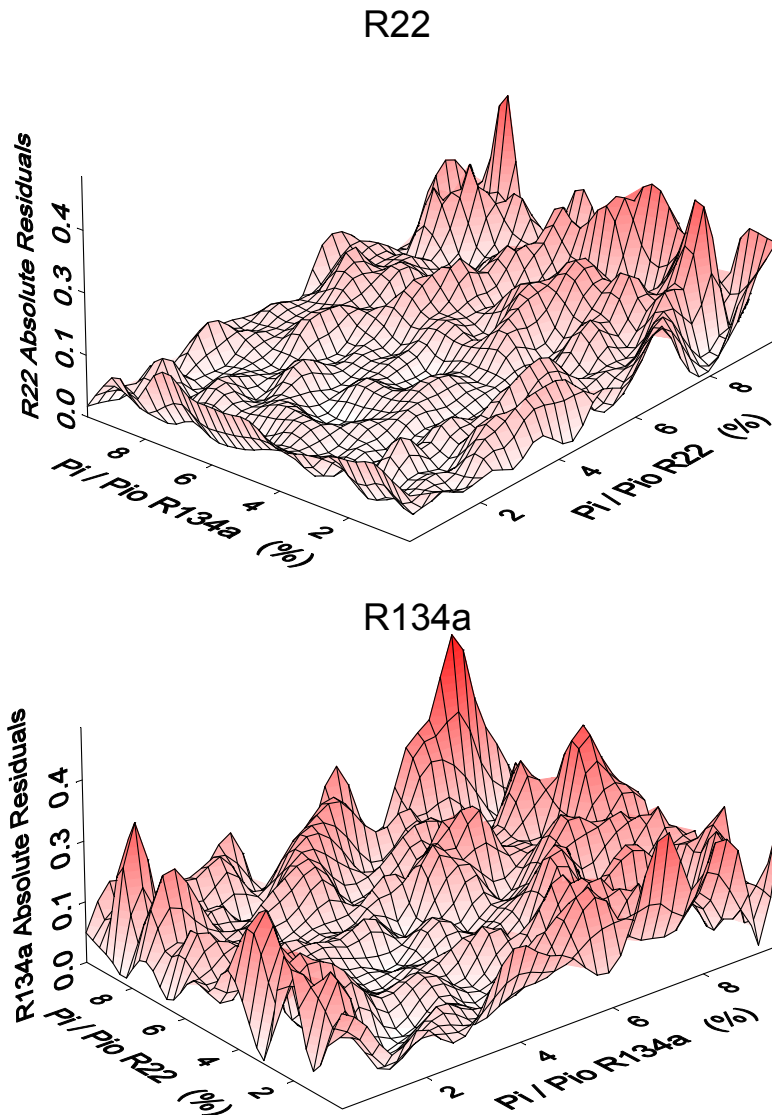


figure 48: Absolute residuals of the predictions of the concentrations of R22 (top) and R134a (bottom) versus the compositions of the corresponding samples.

An unattended use of many genetic algorithms is often limited by chance correlations of variables. This can happen if variables are noisy, if the number of samples is limited and if there are many variables to select. In that case, it can happen that the GA models noise instead of information and consequently selects randomly correlated variables. Therefore, a test similar to [126,255] was performed to investigate the robustness of the variable selection algorithm proposed in this study. In this test, the number of variables is increased by adding meaningless artificial variables, which contain only random numbers, to the meaningful original variables. Then, the algorithm for the variable selection is run using the increased amount of variables. A well performing algorithm should not select any of the artificial random variables, which contain no meaningful information. For this study, 40 random variables were added to the set of 40 original time points. The random variables were created

by uniformly distributed random numbers with the same variation as the original time points. The genetic algorithm framework was used for this extended data set same way as described before except of two parameters adapted for the increased data set: The population size was increased to 100 resulting in about 120 generations until the convergence criterion was reached and the parameter α was set to 1, which resulted in approximately 6 variables being selected in single runs of the GA.

The variable ranking after the first step of the algorithm is shown in figure 49. It is obvious that all random time points are ranked very low and no random variable can be found among the most important 18 time points. The parallel runs of multiple GA with different combinations of test and calibration data seem to prevent the selection of randomly correlated variables whereas single runs of the GA selected random variables evident by non-zero frequencies of random variables in figure 49. Additionally, the left side of figure 49 looks very similar to figure 46 demonstrating the reproducibility of the ranking of meaningful variables when running the global algorithm repeatedly. The top 11 time-points are ranked the same way as for the algorithm applied to the original data (figure 46). Consequently, the same 10 variables are selected in the second step of the algorithm demonstrating the reproducibility of the selection of the variables by the genetic algorithm framework.

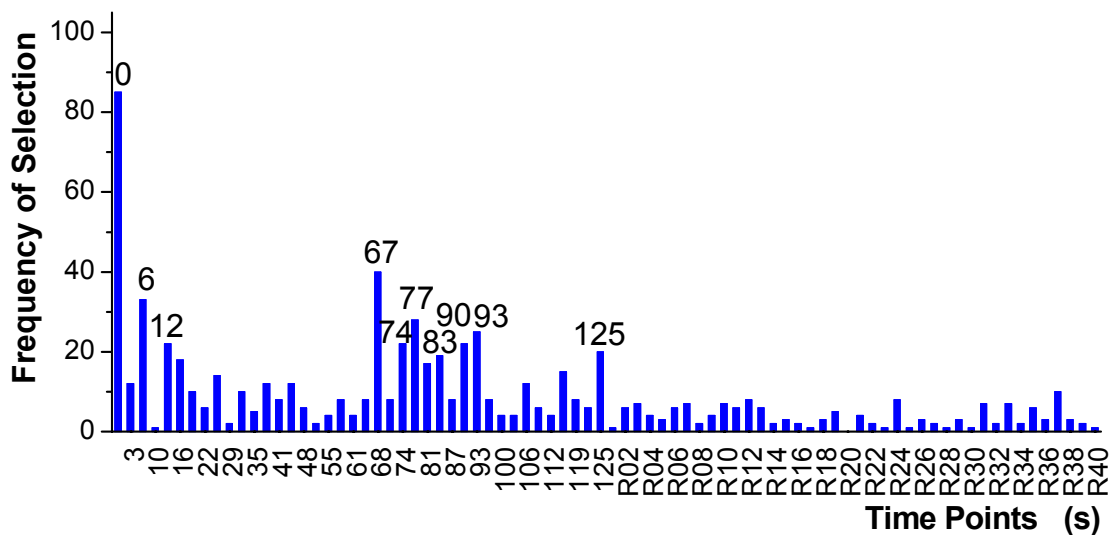


figure 49: Frequency of selection for 40 time points and 40 additional random variables (R1 – R45) after the first step of the genetic algorithm framework.

7.4. Genetic Algorithm Framework – Conclusions

In this chapter, a genetic algorithm was used for the optimization of neural networks by variable selection. It was shown that single runs of the GA are faced with irreproducible variable selections and with unstable predictions of the external validation data, which were even worse than the predictions by non-optimized neural networks. A genetic algorithm framework was suggested, which uses repeated runs of the GA. The predictions of the neural networks optimized by this framework were superior to the commonly used non-optimized neural networks and to all other calibration methods used before. It was also shown that the variable selection is reproducible and not subject to noise. Additionally, it was demonstrated that the predictions are hardly influenced by interfering analytes rendering the combination of time-resolved measurements, the genetic algorithm framework for the variable selection and for the calibration open to the parallel quantification of more analytes using only one single sensor. The unique framework introduced in this works is not restricted to the calibration and variable selection of sensor signals, but can be used for the variable selection and multivariate calibration of virtually any data set as long as a sensible number of data is available.

Method	Calibration Data Set		Validation Data Set	
	R22	R134a	R22	R134a
Non-optimized Neural Networks	1.47	2.62	2.18	3.26
GA-NN (1st run)	2.13	2.75	2.32	2.93
GA-NN (2nd run)	2.19	2.87	2.63	3.35
GA-NN Framework	1.89	2.69	2.04	2.89

table 3: Comparison of the rel. RMSE of the calibration and validation data in % for the genetic algorithm approaches and the non-optimized neural networks used in section 6.8.

8. Results – Growing Neural Network Framework

The genetic algorithms for a variable selection, which were proposed and applied in section 2.8.5 and in chapter 7, successfully improved the calibration of the refrigerant data set by selecting the most predictive variables and by optimizing the number of hidden neurons in a single hidden layer using a simple gradient method. As already stated in section 2.8.2, a total non-uniform optimization of the topology of the neural networks should be superior to a pure variable selection and to a simple optimization of the number of hidden neurons. The algorithms for a structure optimization decide on the need of each single network element resulting in sparse yet effective non-uniform networks. In addition, these algorithms can be used for networks with several hidden layers. The oldest and most popular methods of structure optimization are the pruning algorithms, which were introduced in section 2.8.8 and applied in section 6.10. Yet, as already seen and discussed in both sections the pruning algorithms are faced by several drawbacks rendering the application of these algorithms in practice doubtful. The sophisticated approach of optimizing the network topology by the use of genetic algorithms is also faced by several problems and limits discussed in section 2.8.9 rendering the application of these algorithms nearly impossible for analytical data sets.

The growing neural network algorithm, which was initially proposed by Vinod et al. [125] and which was introduced in section 2.8.10, has already been successfully applied to the calibration of sensor data sets [28]. In this chapter, several modifications of the algorithm are introduced. The application of this algorithm to the refrigerant data set shows an improved calibration with similar prediction errors like the genetic algorithm framework. In order to improve the reproducibility of the algorithm, two frameworks for the growing neural networks similar to the genetic algorithm framework are introduced. Both frameworks show an extraordinary calibration and generalization ability and a good reproducibility.

8.1. Modifications of the Growing Neural Network Algorithm

The original algorithm was modified in several points, which were partly introduced and described in [28] and which are partly introduced in this work, to fit better to the needs of the calibration of sensor data sets and to improve the generalization ability of the networks built:

1. Not only neurons with two input links and one output link but also neurons with one input link and one output link can be added to the network. In addition, links can be added between any neuron and a neuron of a proceeding layer ensuring that practically any feedforward network topology can be built. In contrast to the stepwise algorithms, the addition of neurons with two input links takes interactions of 2 variables during the addition step into account. Higher interactions can be modeled later by the addition of additional links.
2. The estimation of the reduction of the calibration error was replaced by temporarily inserting a network element then training the network and subsequently predicting a monitor data set. This procedure is repeated for all possible locations and all possible elements. The type and the location where to insert the new element are decided by the maximum reduction of the prediction error of the monitor data not used for training. This ensures that the neural network not only approximates the calibration data well, but also primarily generalizes well. Using different data subsets for the calibration of the model (training data) and for the building of the model (monitor data) prevents the introduction of a bias demonstrated by Kupinsky et al. [11]. The change of the network topology by adding a network element between two training steps helps to escape local training minima similar to a random mutation of genetic algorithms.
3. The stopping criterion of an absolute error limit for the algorithm was replaced by a stopping criterion of a relative minimal error decrease, which is independent from the scaling of the data sets. Thereby the insertion of the network elements is repeated until the insertion of a new network element improves the error of prediction less than this prescribed relative limit.
4. The algorithm can start with nearly any arbitrary network topology, not only with an empty network. As the current implementation of the algorithm only supports networks with 1 output neuron, a separate network has to be used for each analyte.

8.2. Application of the Growing Neural Networks

For the application of the growing neural net algorithm, the calibration data set was split into a training (80 %) and a monitor (20%) subset by a random subsampling procedure (see section 2.4). Using the stopping criterion of 0.1% minimal error decrease the growing network algorithm built the network for R22 shown in figure 50 with 11 input neurons, 22 links and 7 hidden neurons organized in 2 hidden layers. For R134a the network consisted of 13 input neurons, 23 links and 7 hidden neurons organized in 2 hidden layers shown in figure 51.

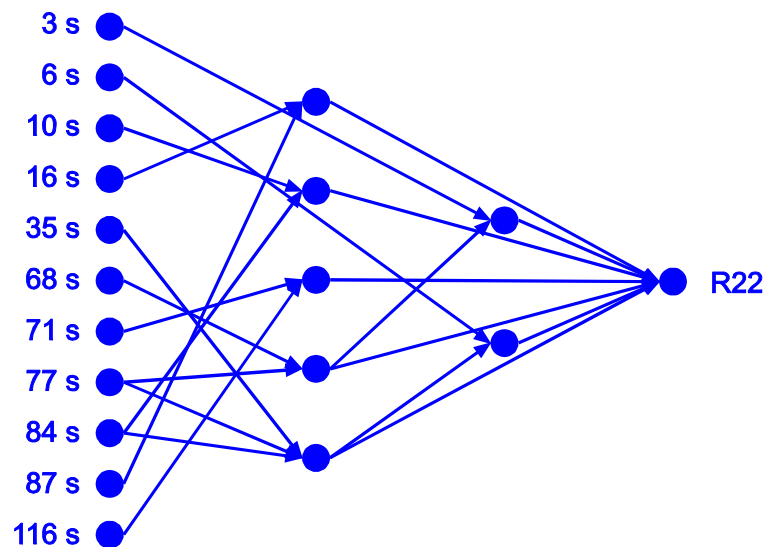


figure 50: Neural network built by the first run of the growing neural network algorithm for R22.

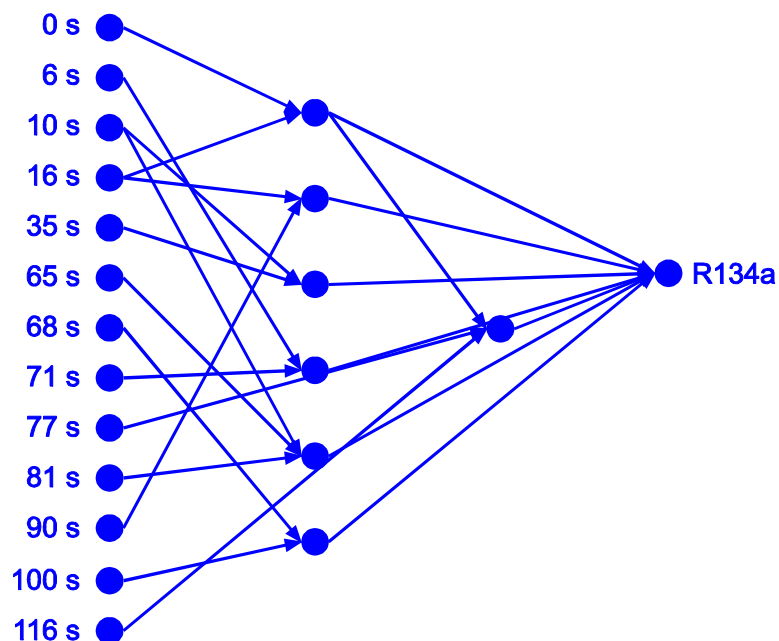


figure 51: Neural network built by the first run of the growing neural network algorithm for R134a.

These network topologies were trained using the complete calibration data set and then predicted the concentrations of the external validation data. According to table 4 in section 8.5, the grown neural networks predicted the external validation data not used for the network growing process significantly better than non-optimized static neural networks and no significant gap between the predictions of the calibration and validation data is visible.

Yet, similar to the application of single run genetic algorithms for the optimization of neural networks (see section 2.8.9), the topology of the grown networks depends highly on the partitioning of the data set. A second run of the algorithm with differently subsampled training and monitor data subsets resulted in other network topologies for both analytes. The network for R 22 of this second run is shown in figure 52. Although several substructures, which are printed in green, are equal to the network shown in figure 50, both networks also show significant differences, which are printed in blue in figure 52. In principle, these differences of the network topology are not necessarily bad as for a given set of input variables the approximation of a functional relationship between the input and the response variables can be performed by a neural network on nearly uncountable ways. Yet, the selection of different variables during different runs is by far more problematic. For example, the second network uses the time points 13 s, 22 s and 29 s instead of 16 s and 116 s as input variables, which are printed in red in figure 52. The selection of different variables irreversibly changes the possibilities of the functional mapping and significantly influences the quality of calibration. As can be seen in table 4, the predictions of the validation and calibration data differ for the nets built during the different runs whereby the growing neural nets performed generally better for the validation data than the static neural nets during several runs. Also, the network of the second run for R134a with 10 input neurons, 18 links and 5 hidden neurons organized in 1 hidden layer differs significantly from the network of the first run for R134a in respect to the topology and even worse in respect to the selected variables.

Similar to the single runs of genetic algorithms the topology and more important the selection of the variables are representative for only one partitioning of the data set into calibration and monitor data set and not for the complete data set. Analogous to the framework of the genetic algorithms (section 7.2), two frameworks are proposed in the next section to make the variable selection of the growing neural networks less sensitive to the partitioning of the data into different subsets and to different random initial weights. In section 8.4, these two frameworks are applied to the refrigerant data sets resulting in improved calibrations.

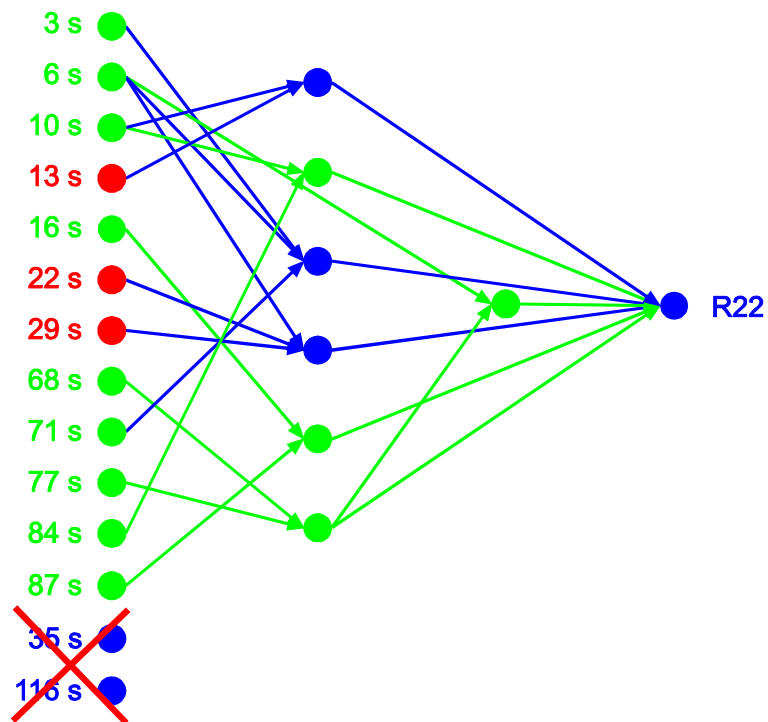


figure 52: Neural network built by the second run of the growing neural network algorithm for R22. Elements equal to the network of the first run are printed in green.

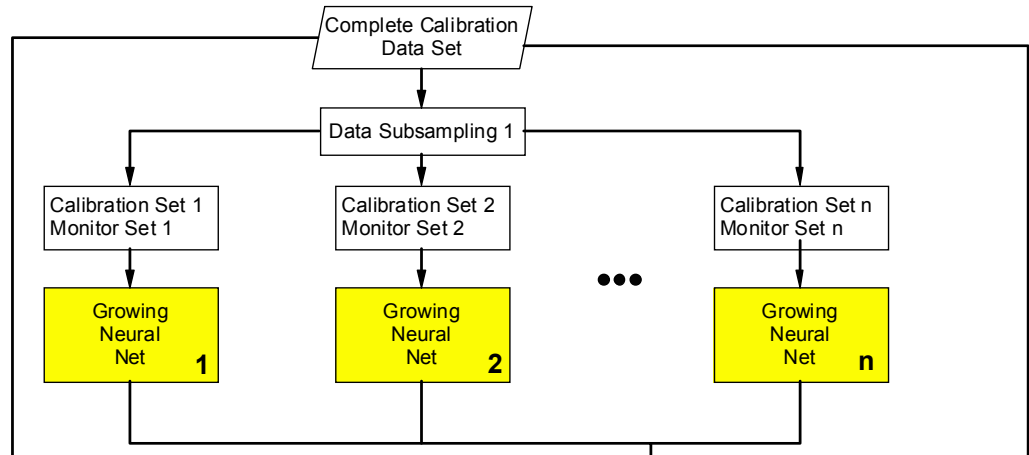
8.3. Growing Neural Network Algorithm Frameworks

The first framework for the growing neural networks is similar to the genetic algorithm framework, which was proposed in section 7.2 and will be further referred to as parallel growing neural network framework. The framework, which is presented in figure 53, can be divided into three steps. The first step consists of multiple parallel runs of the growing neural network algorithm with different training and monitor data subsets. When all networks have finished growing, the variables are ranked according to the frequency of being used in the different networks. In the second step, the algorithm builds the final neural network in an iterative procedure by adding the variables according to their rank step by step to a fully connected neural network. During each step, the performance of the neural network is evaluated by the use of different training and test data subsets. The iterative algorithm stops when the addition of the next variable does not improve the predictions of the test data subsets any more. As the number of hidden variables (organized in 1 hidden layer) is determined by the mean number of hidden variables of the networks built during the first step, the third step of the growing neural net framework, which only trains the final network with the complete calibration data set, is less complicated than the third step of the genetic algorithm framework.

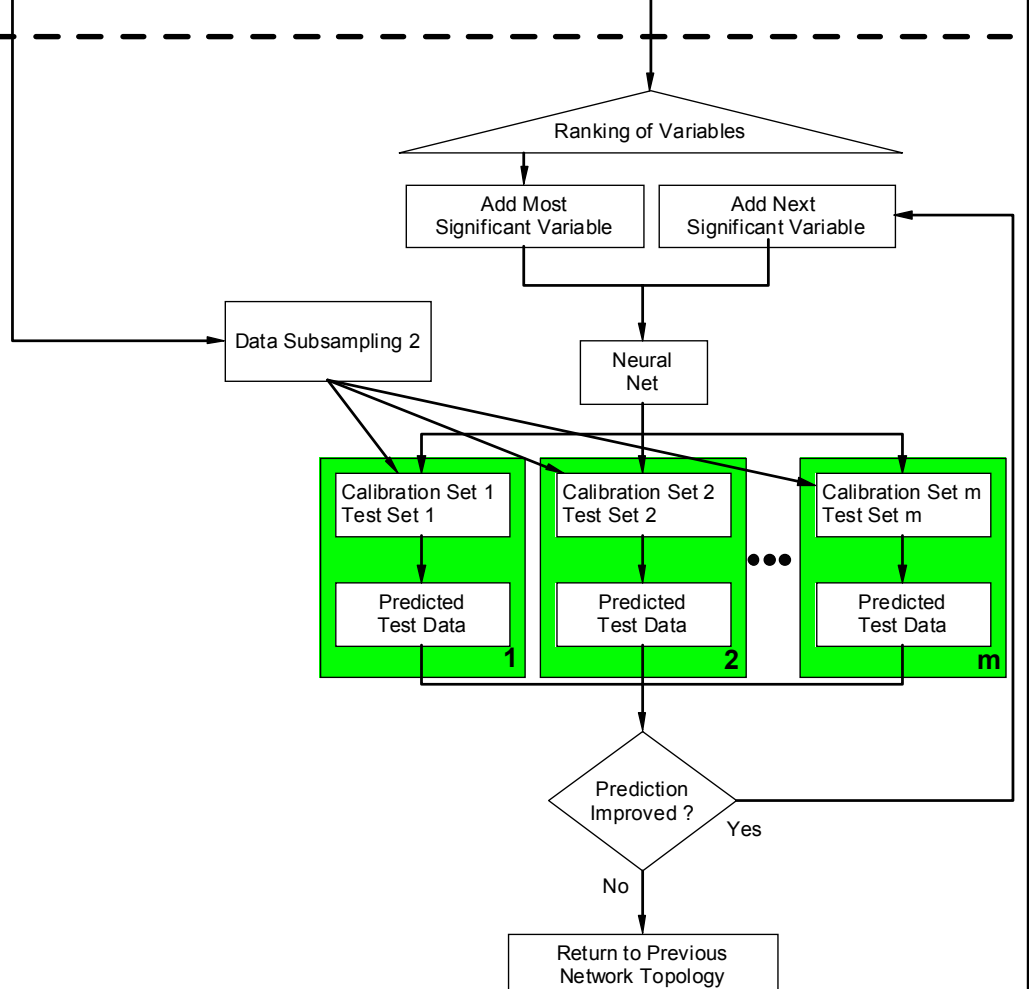
The second framework is based on loops using the knowledge of previous runs of the growing algorithm and will be further referred to as loop-based framework. The concept is illustrated in figure 54. Several parallel runs of the growing neural network algorithm are performed using different random initial weights but the same data subsets for training, monitoring and selection. The network showing the best prediction of the data subset for selection is used as initial network for the growing algorithm of the next loop cycle. Thereby the algorithm starts not with an empty neural network but with the selected built network and new subsets of the data for training, monitoring and selection. This procedure is repeated, until the selected network has the same topology than the network selected in the previous loop cycle.

In both frameworks, the complete calibration data set has to be split several times into subsets for training, monitoring, testing and selection. Similar to the genetic algorithm framework, this was performed by a random subsampling procedure.

Step 1



Step 2



Step 3

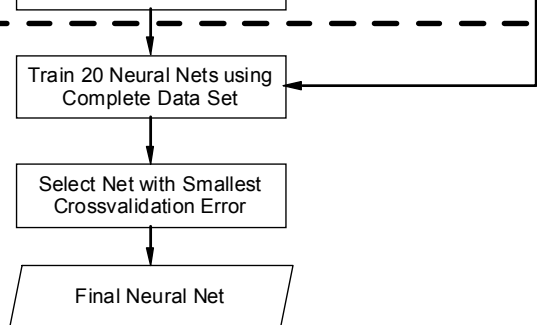


figure 53: Flow chart of the parallel growing neural network framework.

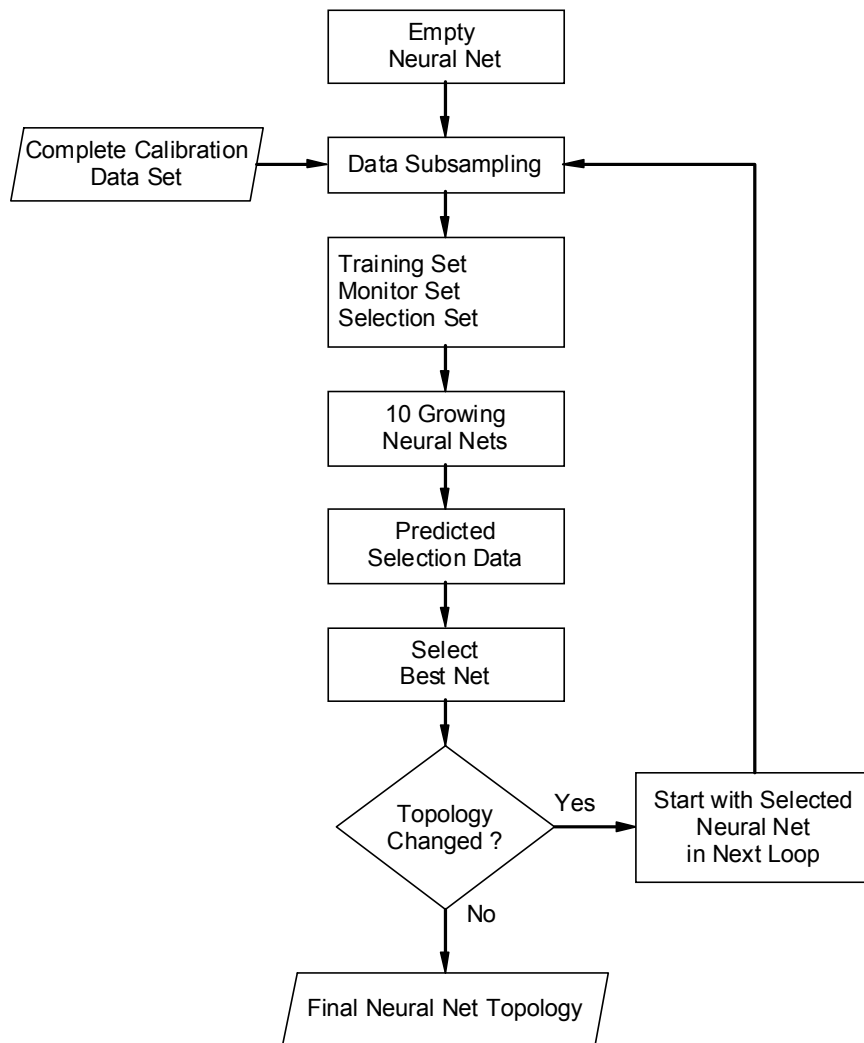


figure 54: Flow chart of the loop-based growing neural network framework.

According to the introduction to the network optimization in section 2.8.2, both frameworks follow different strategies. The parallel framework only uses the variable selection property of the growing neural networks ignoring the information of the inner topology of the grown networks. This framework is suitable for the implementation on massive parallel computer systems as there are only few nodes for information exchange. A high number of parallel runs of the algorithm allows a variable selection practically independent from the partitioning of the data set. The second framework uses the possibility of the algorithm to build non-uniform neural networks with an optimized internal topology adapted for a specific problem. In contrast to the first framework, this framework is less suitable for parallel computer systems and less partitions of the data set are used, since the algorithm typically stops after several loop cycles.

8.4. Applications of the Growing Neural Network Frameworks

8.4.1. Parallel Framework

For the parallel approach, 100 runs of the growing neural network algorithm were performed simultaneously for each analyte. The ranking of the variables after the first step of the frameworks is shown in figure 55. In contrast to the genetic algorithm framework, the variables are ranked separately for each analyte, as for each analyte separate networks were grown. The final topology resulted in fully connected networks for R22 with 5 input neurons and 6 hidden neurons and for R134a with 6 input neurons and 7 hidden neurons. The input variables, which were used by these networks, are labeled in figure 55. These networks were subsequently trained with the complete calibration data set and then used for the prediction of the validation data whereby the errors are shown in table 4. The predictions of the validation data not used during the network building process are the best out of all methods used up to now with 2.04% for R22 and 2.61% for R134a. The small size of the networks with only 43 respectively 57 adjustable parameters is rewarded by an excellent generalization ability. The true-predicted plots look pretty much like the true-predicted plots of the genetic algorithm framework (figure 47) with low standard deviations and no bias present and are not shown here. For both analytes, all variables used by the networks were recorded within the first 16 seconds of exposure to analyte and within the first 30 seconds after the end of exposure to analyte (60 seconds to 90 seconds). The selected time points are located within the same time intervals as the time points selected by the genetic algorithm framework (except of the time-point 125 s). Yet, the parallel growing neural network framework achieves better predictions using fewer variables and smaller networks. Additionally, the variable selection of the parallel framework suggests to reduce the exposure time to 20 seconds and to record the signal of the sensor for 90 seconds. The reduced time of exposure to analyte would also reduce the time needed for the recovery of the sensor signal resulting in a significantly shorter repetition times.

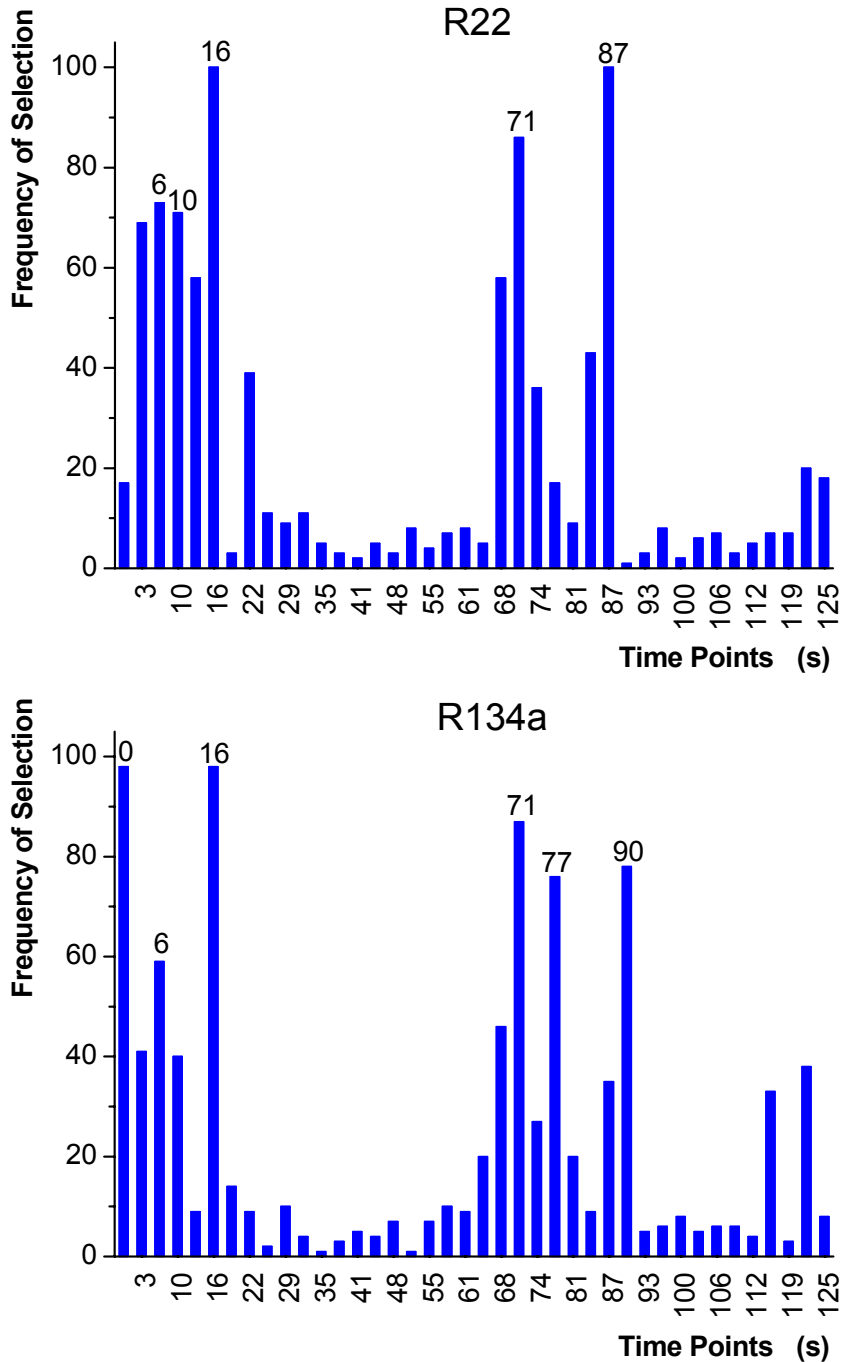


figure 55: Ranking of the time-points represented as frequency of being present in the grown neural networks of the first step of the parallel framework.

The same test for chance correlation and reproducibility was performed as for the genetic algorithm framework before. The parallel framework was used for the increased data set with 40 additional autoscaled random variables the same way as described for the original data set. In the ranking of the variables after the first step, no random variable was found in the top 34 variables for R22 and in the top 25 variables for R134a. This means that the growing neural network algorithm and especially the parallel framework are very robust to selecting

randomly correlated variables. In figure 56, it is shown that for R22 the first 8 variables and for R134a the first 7 variables in the ranking were exactly the same as in the previous run of the parallel algorithm resulting in exactly the same neural network topologies and predictions. Thus, the parallel framework shows a high reproducibility not sensitive to the partitioning of the calibration data set.

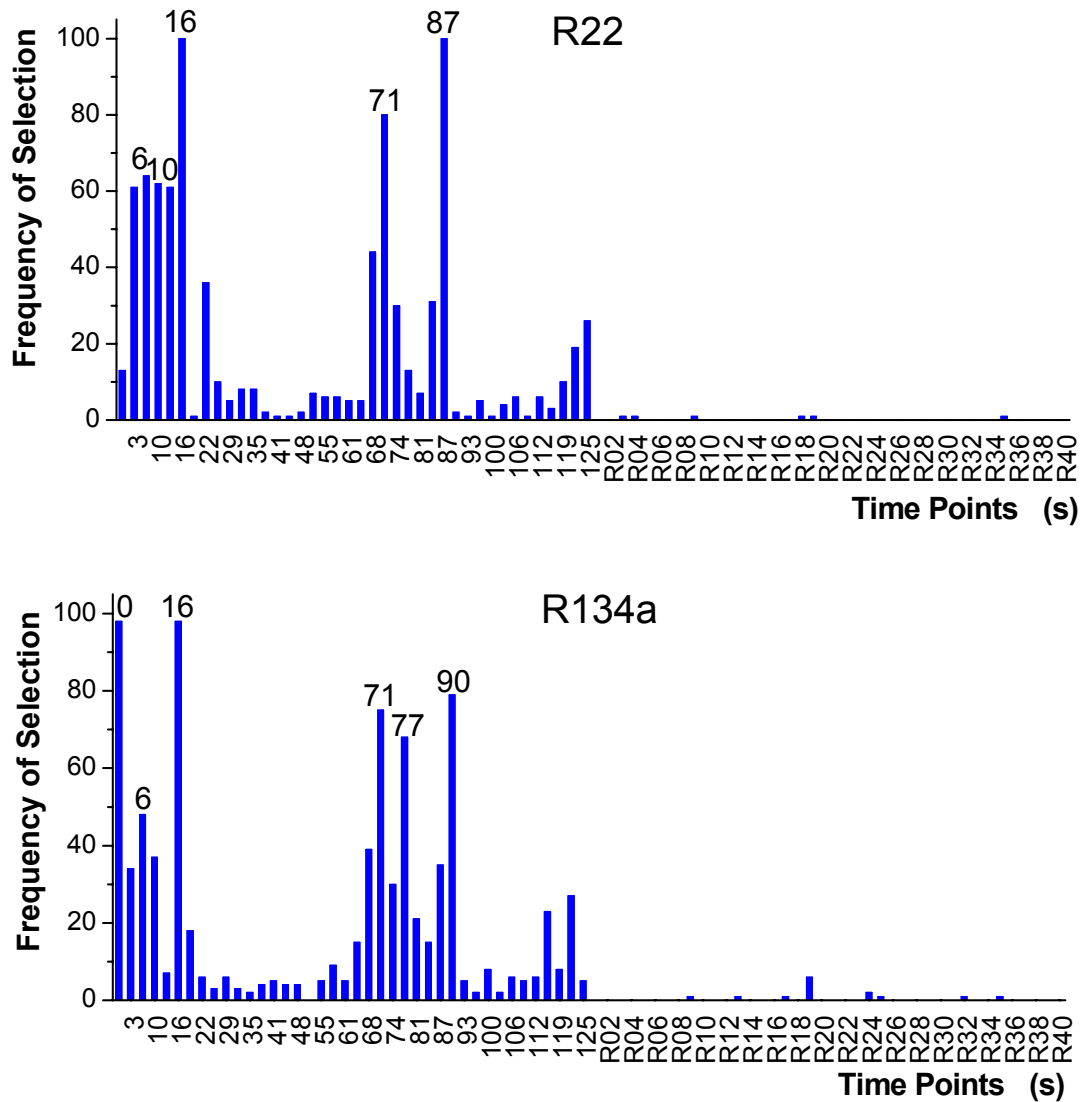


figure 56: Ranking of the time-points and of the random variables after the first step of the parallel framework for the increased data set.

8.4.2. Loop-based Framework

The loop-based approach was performed according to the flow chart shown in figure 54, whereby 10 neural networks were grown in each loop cycle. The complete calibration data set was split into training (60 %), monitor (20 %) and selection (20 %) subsets for each loop cycle. For R22 the framework stopped after 8 loop cycles with a network topology consisting of 16 input neurons, 56 links and 15 hidden neurons organized in 4 hidden layers shown in figure 57. For R134a the framework stopped after 7 loop cycles with a network topology consisting of 13 input neurons, 35 links and 8 hidden neurons organized in 3 hidden layers shown in figure 58. The predictions of the validation data by these network topologies show the best results of all multivariate calibration methods used for this data set with relative errors of 1.50% for R22 and 2.37% for R134a (see table 4). The true-predicted plots show no bias and very low standard deviations for all concentration levels (see figure 59). Compared with the parallel approach the loop-based network topologies use rather many input variables. It is also remarkable that the number of 3 respectively 4 layers of hidden neurons is unusually high. Yet, the non-uniform network design helps to keep the number of adjustable parameters low by building a sparse network topology with only few links. The topologies of the grown neural networks show that the common recommendation [8,257-259] to use only 1 or at the furthest 2 hidden layers for fully connected networks is only a vague rule, as the growing neural network algorithm automatically decides, how many hidden layers are optimal. The good generalization ability demonstrates that the non-uniform topology efficiently uses small networks and is superior to fully connected networks.

The same test for chance correlation and reproducibility was performed for the loop-based approach as already described for the genetic algorithm framework. Thereby the network topologies are by far more reproducible than using single runs of the growing neural network algorithm. The network for R22 of the second run uses the same variables than the network of the first run except of one variable being not used. The network for R134a uses the same variables of the first run except of one variable, which was exchanged by another one. Both networks did not use a random variable, whereby inside a loop cycle some networks used a random variable but these networks were not selected for the next loop cycle due to worse predictions of the selection data sets.

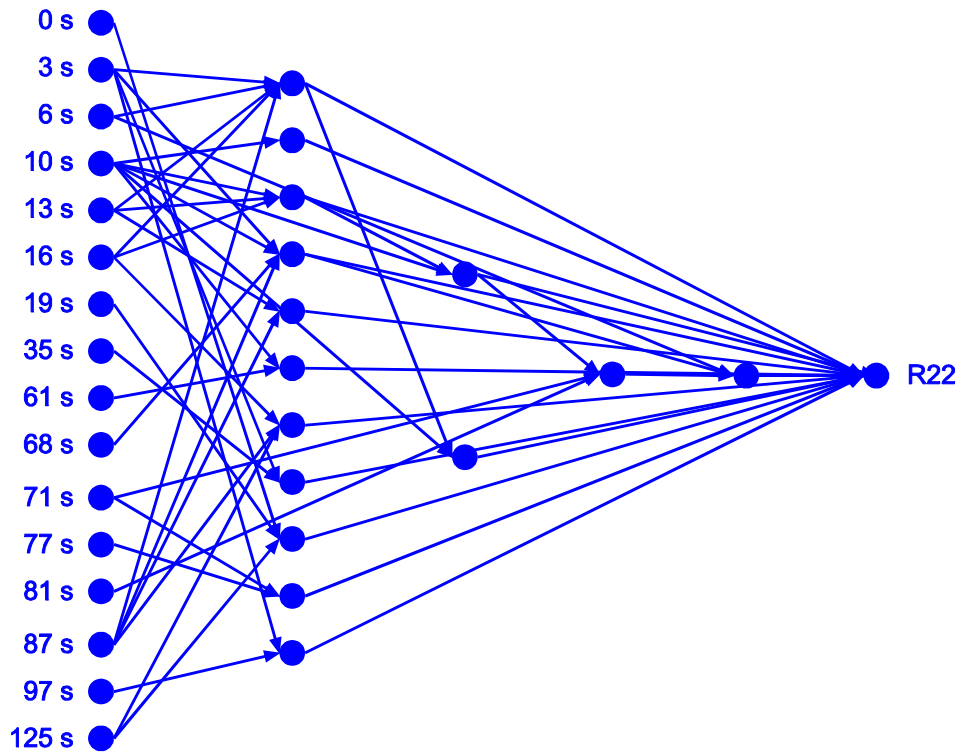


figure 57: Neural network with 4 hidden layers built by the loop-based framework for R22.

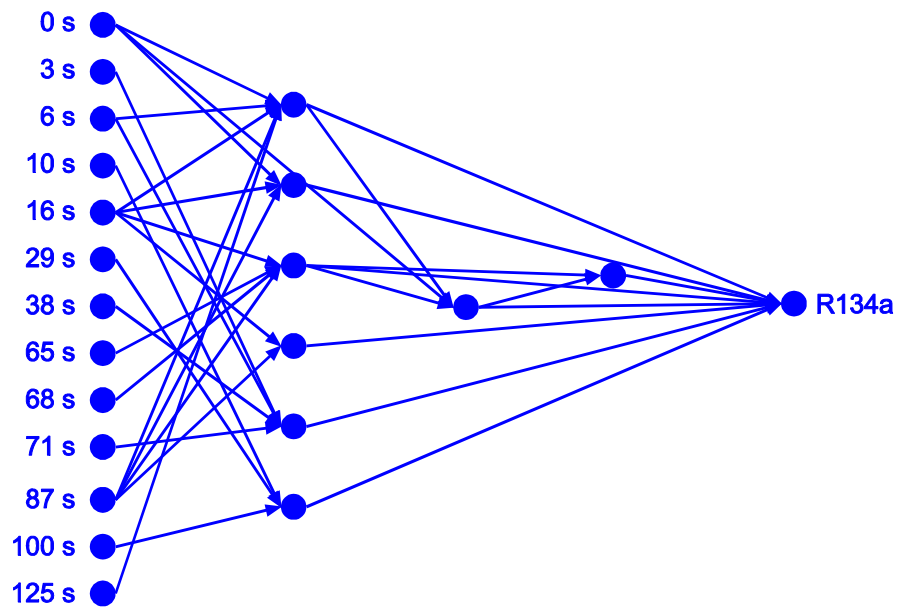


figure 58: Neural network with 3 hidden layers built by the loop-based framework for R134a.

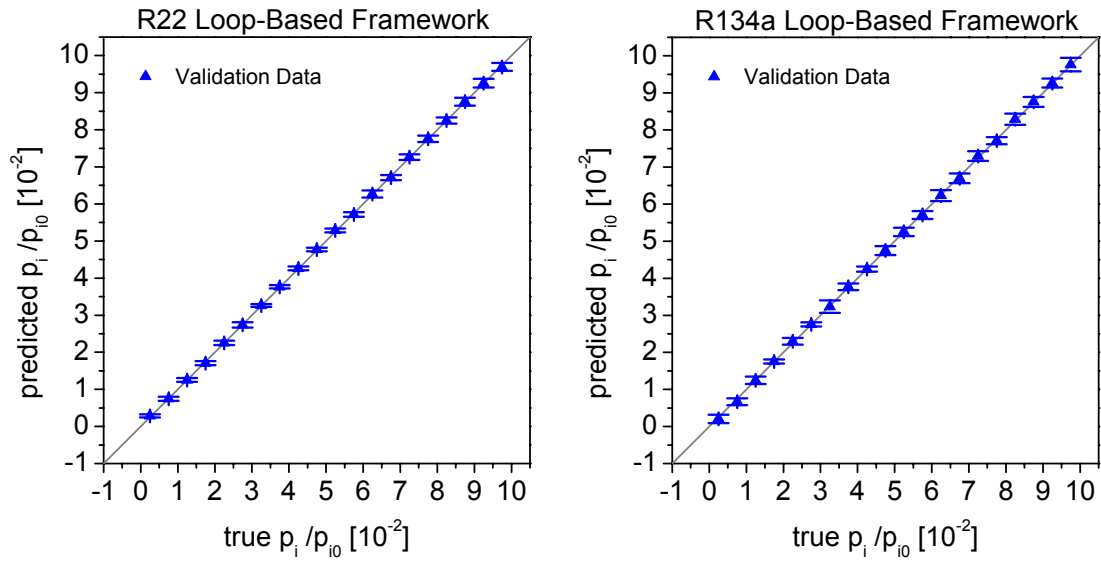


figure 59: Predictions of the validation data by neural networks optimized by the loop-based growing network framework.

8.5. Conclusions and Comparison of the Different Methods

In this chapter, a growing neural network algorithm for building non-uniform neural networks was applied to the refrigerant data set. The algorithms showed improved calibrations compared with the common non-optimized neural network. Yet, the variable selection and the topology of the networks were only partly reproducible. Thus, similar to the genetic algorithms a parallel framework and additionally a loop-based framework were introduced to improve the reproducibility and to improve the calibration quality further.

The loop-based framework showed the best generalization ability of all multivariate data analysis methods introduced and applied in this work as it allows building non-uniform neural networks of any arbitrary size and topology exploiting a data set limited in size to a maximum extent. The predictions of the external validation data showed impressively low rel. RMSE of 1.50% for R22 and 2.37% for R134a, which are only slightly higher than the standard deviations of the sensor signals of reproduced measurements. The loop-based framework needs a lot computing power (2 weeks for each analyte of the refrigerant data set using an up-to-date personal computer) and is hardly suited for parallel computing hardware as it is a loop based approach and not a parallel approach.

If a reproducible variable selection is important, both, the parallel growing network framework introduced in this chapter and the genetic algorithm framework introduced in the previous chapter are a good choice, whereby the latter scales better with an increasing number of variables, but shows a slightly worse generalization ability. Both parallel frameworks showed improved calibrations compared with the common neural networks. Both frameworks are well suited for parallel computer hardware rendering both methods ideally suited for computer pools.

The single run growing neural network algorithm is a good choice for data sets with not too many variables to find an optimized non-uniform network topology without the danger of overfitting (about 3 hours computing time for each analyte of the refrigerant data set). Different single runs of the growing neural network all showed better calibrations than the non-optimized neural networks.

Although single runs of genetic algorithms are frequently reported in literature, this method has been proven to be inferior compared with the different new algorithms and frameworks introduced in this work. Though being the fastest methods for a successful variable selection (about 1 hour for refrigerant data set), the instability of the variable selection and the resulting

diversity of the quality of calibration and of prediction render single runs genetic algorithms rather useless for most applications.

In summary it may be said, that the growing neural networks and all three frameworks introduced in this work performed better than the common non-optimized neural networks. Among these new methods introduced in this work, no general recommendation for a specific method can be given as the method of choice for the optimization of neural networks depends on the needs of the user and on the data set.

Method	Adjustable Parameters		Calibration Data Set		Validation Data Set	
	R22	R134a	R22	R134a	R22	R134a
Non-optimized Neural Networks	247	247	1.47	2.62	2.18	3.26
Growing Neural Networks (1st run)	30	31	1.84	2.73	1.99	2.63
Growing Neural Networks (2nd run)	31	24	2.14	2.73	2.12	2.87
Parallel Framework Growing Neural Networks	43	57	1.89	2.71	2.04	2.61
Loop-based Framework Growing Neural Networks	72	44	1.39	2.41	1.50	2.37

table 4: Comparison of the rel. RMSE of the calibration and validation data in % for the growing neural network approaches and the non-optimized neural networks. Additionally the number of adjustable parameters used by the networks are listed.

9. Results – All Data Sets

In this chapter, the principle of the time-resolved measurements introduced in chapter 5 and the different new data analysis methods introduced in chapter 7 and chapter 8 are applied to several data sets. Thereby the methods are compared with the commonly used methods of data analysis and with static measurements to demonstrate the superiority of the new approaches. Different data sets of mixtures of the homologous series of the low alcohols were recorded by different setups. The application of the new approaches to these data sets demonstrates that the methods proposed in this study can be easily transferred to different setups and to different analytical task, such as a varying number of analytes in mixtures and varying types of analytes. In the last part of this chapter, the principle of the time-resolved measurements is applied to the data set introduced in chapter 3 also resulting in an improved calibration.

9.1. *Methanol and Ethanol by SPR*

Measurements of the two homologues Methanol and Ethanol allow investigations of the difference of the kinetics of sorption and desorption, which is mainly based on the different sizes of the two analytes. Therefore several measurements of single analytes and measurements of mixtures were performed, which are described in section 4.5.2.1 and in [190] in detail. Besides of non-optimized neural networks, the parallel growing neural network framework is applied to the data resulting in a very small number of variables being selected. This small number of variables allows the insight into the "black box" of the calibration by neural networks by the use of a sensitivity analysis. Additionally, the selection of the variables is confirmed by a brute-force variable selection.

9.1.1. **Single Analytes**

The changes of the resonance wavelength when the device was exposed to alternating sequences of different concentrations of analyte and to synthetic air are shown in figure 15 in section 4.5.2.1 with the shift of the SPR resonance wavelength plotted versus the time and versus the relative saturation pressure of analyte (0 to 0.045). It is visible that 10 seconds of exposure to methanol vapor are enough to reach equilibrium between the methanol vapor and the methanol, which is sorbed into the polycarbonate layer, and thus the resonance wavelength does not change any more. In contrast, even after 120 s of exposure of the

polycarbonate layer to ethanol vapor equilibrium has not been reached. However, after 120 s the signals of ethanol are higher than the signals of methanol for the same relative saturation pressure as ethanol has a higher refractive index ($n_{D20} = 1.362$) than methanol ($n_{D20} = 1.329$). Since the diffusion of ethanol is slower, the signals at short exposure times are lower compared to methanol.

9.1.2. Parallel Growing Neural Network Framework

For the multivariate calibration, neural networks using all time points of the calibration data were trained set and subsequently predicted the validation data set. The root mean square errors of the crossvalidation and of the prediction of the independent validation data set are listed in the first row of table 5. The concentrations of the test samples were predicted with a relative RMSE of 3.32% for methanol and of 4.11% for ethanol. The fully connected networks used for the calibration consisted of 53 input neurons, 4 hidden neurons and 1 output neuron. In order to improve the calibration, the parallel growing neural network framework introduced in chapter 8 was applied to the calibration data. Thereby the growing network algorithm was repeated 200 times for each analyte. In contrast to chapter 8, the frequencies of the variables being selected after the first step of the algorithm are combined for both analytes by summing up the individual frequencies of each analyte. The corresponding plot (figure 60) shows that the variables in three time intervals are prominent: The beginning of exposure to analyte (10-45 s), when exposure to analyte has ended and exposure to synthetic air starts again (130-140 s), and additionally hundred seconds after the start of exposure to synthetic air (220-230 s). The first two intervals are easily interpretable as in accordance with figure 15 the sensor responses for both analytes differ most during these time intervals. The time interval around 220 s might be considered as reference signal with practically no analyte remaining to compensate possible drifts of the baseline.

Number of Time Points	Calibration Data		Validation Data	
	Methanol	Ethanol	Methanol	Ethanol
53	4.77	3.14	3.32	4.11
3	2.01	1.91	1.97	2.37
2	2.28	2.36	2.22	2.70

table 5: Relative RMSE in % for the prediction by the neural networks using different time points.

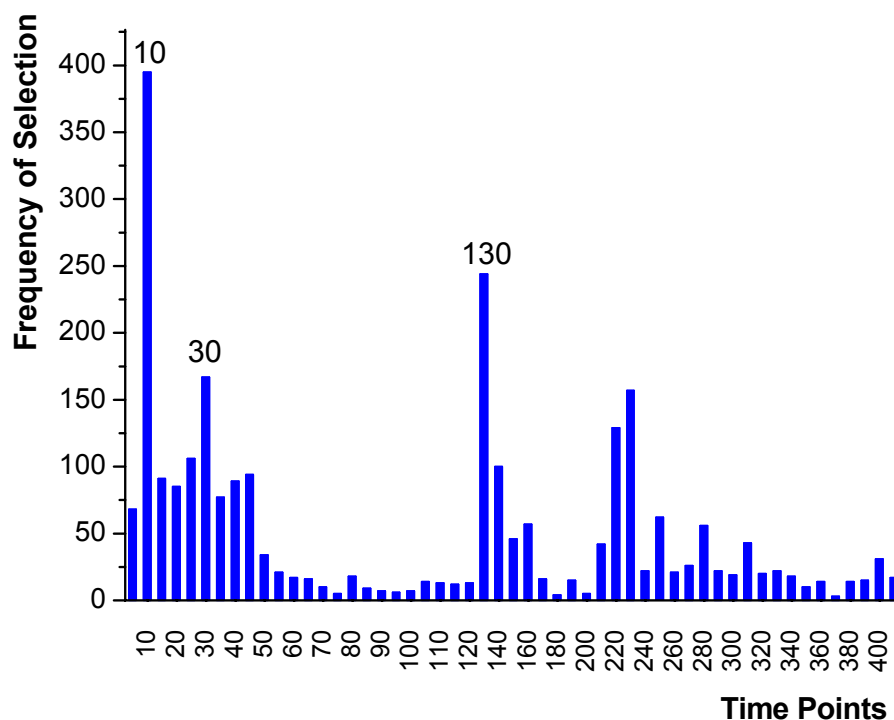


figure 60: Ranking of the time points after the first step of the parallel framework for both analytes together.

The final model obtained by a stepwise addition of time points (step 2 in figure 53) uses only the 3 time points 10 s, 30 s and 130 s (topology 3-4-1). According to table 5 the prediction errors are significantly lower compared with the fully connected networks for both, the crossvalidated calibration data and for the validation data with excellent low errors of 1.97% for methanol and 2.37 % for ethanol. The third time interval is not used by this network and consequently seems not to contain significant additional information not covered by the 3 other sensor signals. Using reproduced measurements of the single analytes in the same concentration range, a standard deviation of the signals was calculated with 0.62% for methanol and 0.98 % for ethanol. These errors are caused by the noise of the spectrometer, inaccuracies of the gas mixing station and fluctuations of the temperature and thus also exist for the measurements of the mixtures. The only moderate increase of the errors for the calibration and prediction shows the potential of the calibration and variable selection by the growing neural network framework. In figure 61 the true-predicted plots of the validation data and the calibration data are shown. The predictions of all concentrations are characterized by very small standard deviations and by the absence of systematic errors.

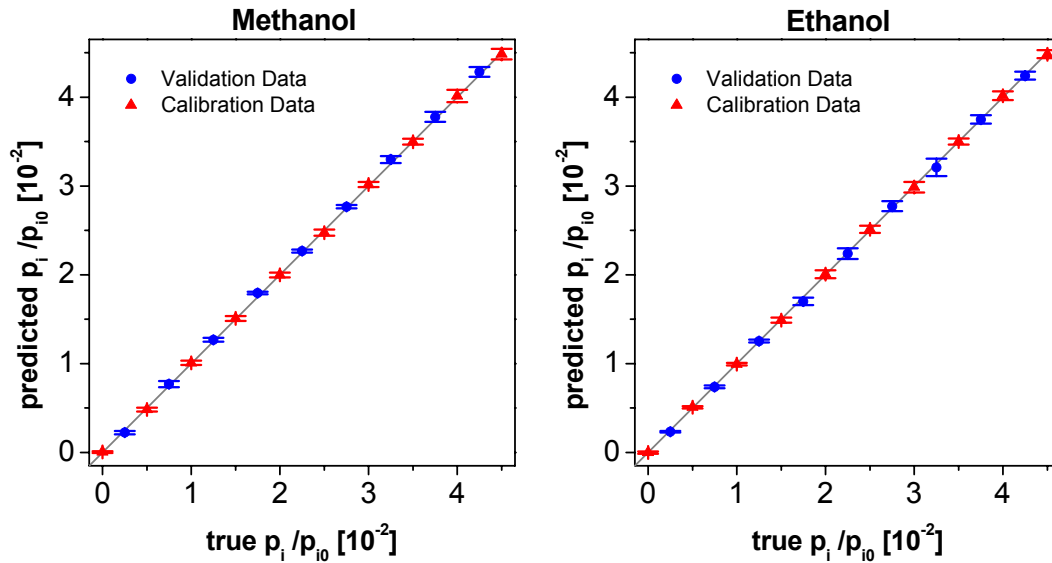


figure 61: Predictions of the calibration and validation data by the optimized neural networks using only three time points.

9.1.3. Sensitivity Analysis

The small number of 3 variables allows an efficient sensitivity analysis [95,101] to give a little insight into the "black box" of neural networks. Thereby the values of two time points were systematically varied between the measured minimum and maximum sensor responses of these time points with the third time point kept constantly in the middle of the corresponding range of the measured sensor responses. The trained neural nets described above were fed with the values of the 3 time points. The concentrations calculated by the neural nets are plotted versus the two varying time points in figure 62.

The top row of this figure shows the prediction of the concentrations of methanol (left) and ethanol (right) depending on the sensor responses of the time points 10 s (x -axis) and 30 s (y -axis). The prediction of methanol is determined by the ensemble of both time points. This is in accordance with figure 15, which demonstrates that the sorption process of methanol has come to a steady state after 10 s. Hence, a high sensor response after 10 s caused by a high concentration of methanol automatically induces a high sensor response after 30 s. The top right plot of figure 62 shows that the prediction of ethanol is practically not influenced by the time point 10 s explainable by the variance of this time point being mainly caused by the sorption of methanol (see figure 15). Nevertheless, the prediction of ethanol is nearly linearly correlated with the sensor response of the time point 30 s. The bottom row of figure 8 shows the prediction of the concentration depending on the signal of the time points 10 s (x -axis) and 130 s (y -axis). The plane parallel to the y -axis demonstrates that the prediction of methanol is practically independent from the time point 130 s. However, the prediction of ethanol highly

depends on the sensor response of the time point 130 s. The plots of figure 62 top right and bottom right are nearly identical except of the higher dynamics of time point 130 s.

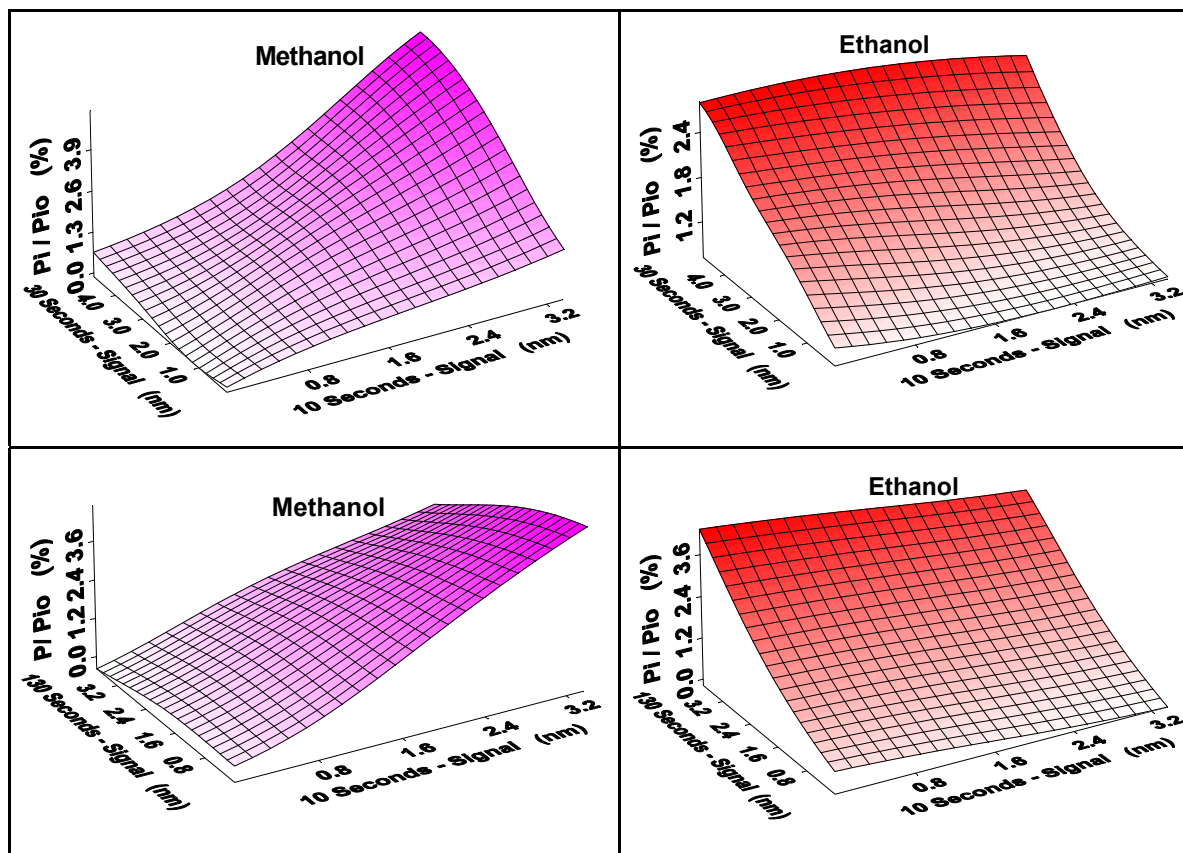


figure 62: Predicted concentrations of the sensitivity analysis versus the sensor responses of the two time points varied.

In summary, it may be said that the prediction of methanol depends on the combination of the time points 10 s and 30 s whereas the prediction of ethanol depends on the time points 30 s and on the time points 130 s. The similar dependencies of the predictions of ethanol on the time points 30 s and 130 s indicate that the time point 130 s could be rendered unnecessary by calculating the concentrations of methanol using the sensor response of the time point 10 s and by calculating the concentration of ethanol by the ratio of the sensor responses after 10 s and 30 s. Thus, neural networks of the topology 2-4-1 were trained using only the time points 10 s and 30 s. These small networks perform quite well predicting both analytes with relative errors of only 2.22% and 2.70% (see table 1). The only small deterioration of the predictive ability is overcompensated by the fact that the measurement time can be reduced from 130 s to 30 s. In addition, the shortening of the time with the polymer being exposed to the analyte results in the sorption of less ethanol into the polymers. Therefore, the desorption of ethanol needs less time to be completed additionally shortening the time needed between 2 measurements.

9.1.4. Brute Force Variable Selection

The low number of 3 variables selected by the growing neural network framework for an optimal model also allows a comparison with the variable selection by a brute force method. According to expression (14), there are 140556 different realizations for selecting 3 variables out of 53. For all these realizations neural networks (fully connected with 4 hidden and 1 output neuron) were trained using the calibration data set and then the mean error of the prediction of the validation data set for both analytes was calculated (similar to the prediction error shown in figure 4 for 2 variables of the refrigerant data set). This procedure was repeated 25 times using different initial weights for the neural networks (a higher number of runs is desirable but limited by the computing time). Among the 25 best networks in respect to the lowest mean prediction error of the validation set, only 1 combination of 3 variables was selected more than once. This selection was the same than the 3 time points selected by the growing networks, whereas the 23 other best selections all were different. Thus, the best combination of the variables highly depends on the initial weights of the training, which is an indication of a high correlation of the variables rendering many realizations of 3 variables very similar. Nevertheless, the 3 variables selected by the parallel growing network framework were the most frequently individually selected variables among the 25 selections by the brute force method confirming the variable selection quality of the parallel growing network framework.

9.1.5. Conclusions

It has been shown that mixtures of methanol and ethanol vapors can be quantified by the use of a single SPR sensor coated with Makrolon and a subsequent data analysis by neural networks. Compared with fully connected neural networks the variable selection and calibration by the parallel growing neural network framework again showed several benefits. First, the calibration was significantly better with impressive low prediction errors, which were only slightly higher than the standard deviations of the signals of reproduced measurements. Secondly, the small number of variables of the optimized networks allowed an insight into the calibration model using a visual interpretation, whereby each variable could be assigned a chemical sense. Finally, the comparison with a brute force variable selection demonstrated that the parallel growing neural network framework selected the probably most predictive combination of 3 variables.

9.2. Methanol, Ethanol and 1-Propanol by SPR

Although the measurements of methanol, ethanol and 1-propanol by a single sensor SPR device can be seen as an extension of the previous sections with 1 more analyte to be quantified, this data set was recorded as the first data set of the series. Thereby the measurement parameters like the thickness of the sensitive layers, the number of recorded time points and the measurement time had been less optimized. Details of the measurements and data sets are explained in section 4.5.2.2 and in [194]. An interesting point with consequences for some data analysis methods is the fact that the validation data set in contrast to the calibration data set was measured by averaging 2 repeated measurements. Several common data analysis method, like PLS, INLR, NN, PCA-NN and the new methods like the genetic algorithm framework and the parallel growing neural network framework are applied to the data set. This allows a comparison of the performance of the different methods on the basis of a second data set in addition to the refrigerant data.

9.2.1. Single Analytes

The sensor responses for the measurements of different concentrations of the 3 analytes are plotted in figure 16 (section 4.5.2.2) during 600 seconds of sorption and a subsequent desorption. The sensor reacts spontaneously to the presence and absence of methanol vapors resulting in a rectangular shape along the time axis. The sensor response for ethanol is slower resulting in shapes, which are more curved, whereas the sorption of 1-propanol into the sensitive polymer layer is very slow not reaching equilibrium between the vapors of 1-propanol and the 1-propanol sorbed into the polymer layer. Therefore, the shape of the sensor response for 1-propanol does not reach a plateau during sorption and the signals need a long time to come down to the baseline during desorption.

9.2.2. Multivariate Calibrations of the Mixtures

Using different data analysis methods introduced in chapter 6, models were built using the calibration data set. Then the validation data were predicted with the relative RMSE listed in table 6.

Method	Calibration Data			Validation Data		
	Meth.	Eth.	Prop.	Meth.	Eth.	Prop.
PLS	8.82	12.68	16.99	6.95	11.01	16.12
INLR	9.02	12.57	13.21	8.67	12.39	17.57
Neural Networks	2.83	6.25	7.41	4.33	8.65	12.58
GA Framework	3.24	6.91	7.84	3.58	6.20	7.60
Growing NN Framework	2.62	6.26	7.16	3.31	6.05	7.33
PCA-NN	2.42	6.39	7.31	3.63	8.68	13.36

table 6: Relative RMSE in % for different calibration methods.

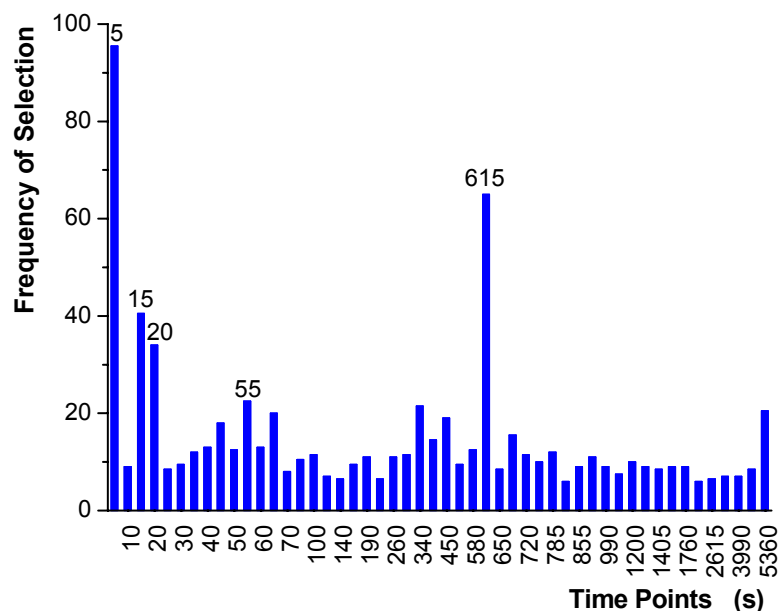
For the calibration by the PLS, the optimum number of principal components was determined by the minimum crossvalidation error with 4 principal components for methanol, 8 principal components for ethanol and 3 principal components for propanol. The predictions of both, the calibration and the validation data are unacceptably high. Similar to section 6.1, the calibration by the PLS cannot deal with the nonlinearities of the data with systematic deviations of the predictions. The INLR (see section 6.3) also showed disappointing predictions of both data sets. Compared with the PLS, the systematic deviations of the predictions were lower but the scattering was higher. In contrast, the calibration by fully connected neural networks (6 hidden neurons and 3 output neurons) was significantly better. Yet, the gap between the calibration and validation data shows that there is still room for an optimization of the neural networks to prevent an overfitting.

9.2.3. Genetic Algorithm Framework

The genetic algorithm framework introduced in chapter 7 was applied to the calibration data set with 100 parallel runs of the GA. Each GA run evaluated 50 populations using about 60 generations whereas the stopping criterion was set to a convergence of the standard deviation of the genes below 0.04. The parameter α of the fitness function was set to 0.9 resulting in the selection of approximately 6 variables per single GA.

The ranking of the variables after the first step is shown in figure 63. In the second step, these variables entered the model according to their rank until the prediction of the test data did not

improve significantly. In contrast to chapter 7, a Kruskal-Wallis non-parametric test ($p < 0.05$) was used to test the significance of improvement for the 20-fold random subsampling procedure (see section 10.2 for a detailed discussion). The iterative procedure stopped after the addition of 5 variables with the selection of the time points 5 s, 15 s, 25 s, 55 s and 615 s, which are labeled in figure 63.



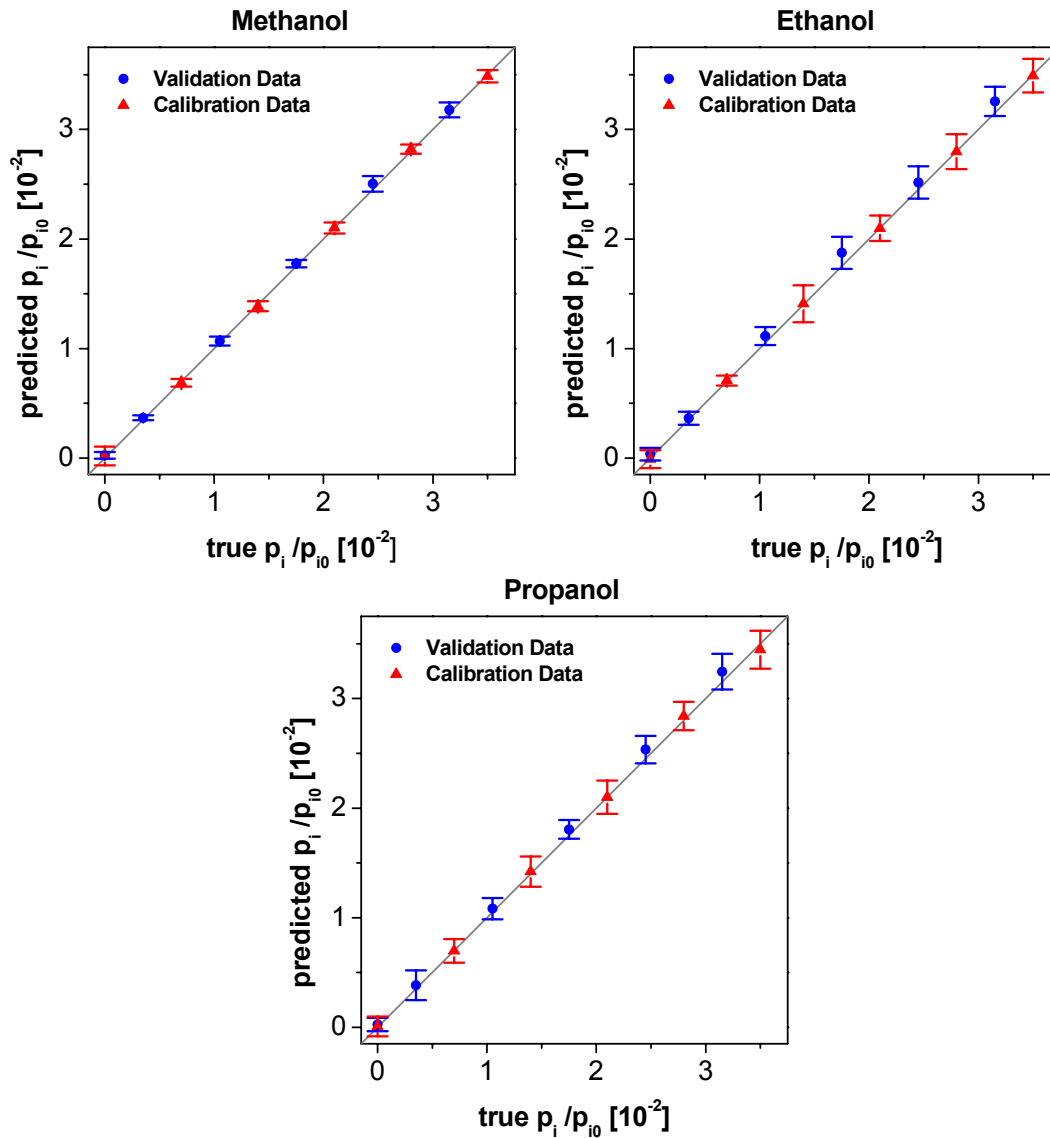


figure 64: Predictions of the calibration and validation data by the neural networks optimized by the genetic algorithm framework.

The 5 time points selected by the framework (5 s, 15 s, 25 s, 55 s and 615 s) can be analyzed in more detail when looking at the sensor response plots (figure 16). The response surface of methanol shows that after 5 seconds the response has practically reached the plateau of the highest sensor signals whereas ethanol and 1-propanol hardly show any sensor signal. The same applies to the 615 s signal, which is situated 15 seconds after the end of exposure to analyte: Methanol has already desorbed whereas the sensor response of ethanol is still very high and 1-propanol shows practically no decrease of the sensor signal. Thus, the 5 s signal represents the concentration of methanol, whereas the 615 s signal represents the sum of the concentrations of ethanol and propanol. Large parts of the variance of the sensor signals after 15 and 25 seconds can be identified with ethanol since the signal of methanol has already reached the plateau and 1-propanol attributes negligibly to the total signal. On the other hand,

the variance of the sensor signal at 55 seconds can be mainly ascribed to 1-propanol whereby the sensor response of methanol has completely and the sensor response of ethanol has nearly reached equilibrium. In summary, it may be said that all 5 time points selected by the algorithm can be associated with the characteristic sensor responses of the pure analytes and consequently make sense in a chemical respect. Another benefit from the variable selection results from the direct relation of the variables with the time needed for the analysis. Only information during 55 seconds of exposure to analyte and 15 seconds after the end of exposure is evaluated. Thus, it should be enough to reduce the time of exposure to analyte to 55 seconds and to record the sensor responses during 70 seconds. This would dramatically reduce the analysis time.

Similar to section 7.3, a randomization test was performed to test the reproducibility and robustness of the variable selection and of the calibration. For this test, 50 uniformly distributed autoscaled random variables were added to the set of 50 original time points. The genetic algorithm framework was used for this extended data set the same way as described before except of increasing the population size to 100 resulting in about 110 generations until the convergence criterion was reached and except of setting the parameter α to "1", which resulted in approximately 6 variables being selected in single runs of the GA. The ranking of the variables after the first step of the algorithm is shown in figure 65.

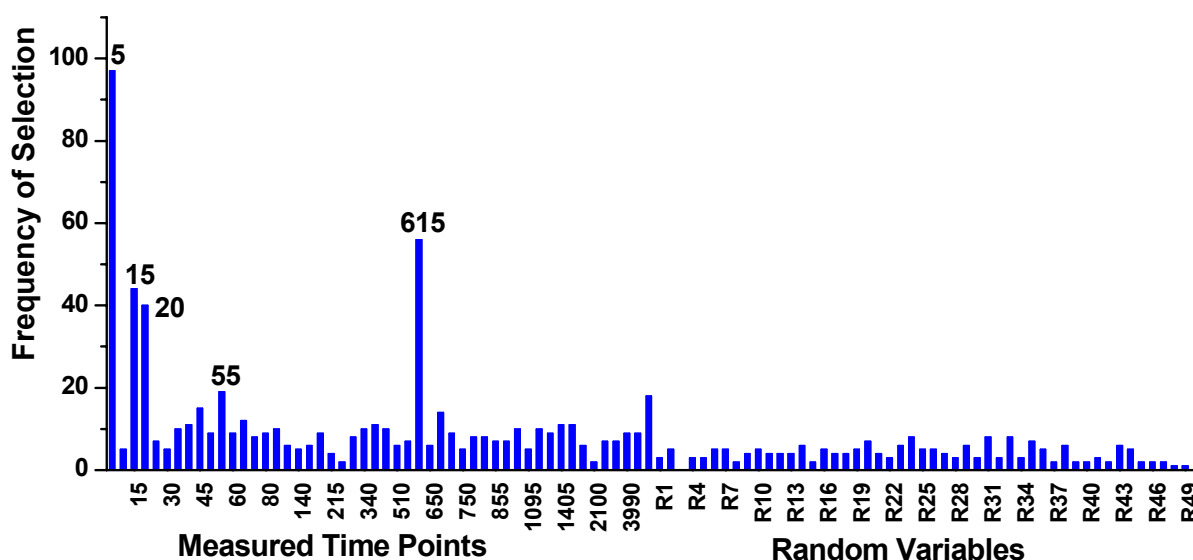


figure 65: Ranking of variables for 50 time points and for 50 additional random variables.

It is obvious that all random time points are ranked very low and no random variables can be found among the most important 27 time points. Similar to section 7.3 the parallel runs of multiple GA prevented the selection of randomly correlated variables whereas single runs of the GA selected random variables. The left side of figure 65 looks very similar to figure 63 demonstrating the reproducibility of the ranking of meaningful variables. The top 5 time

points are ranked similarly to the algorithm applied to the original data. Consequently, the same 5 variables are selected in the second step of the algorithm demonstrating the reproducibility of the selection of the variables by the genetic algorithm framework.

9.2.4. Parallel Growing Neural Network Framework

Also, the parallel growing neural network framework introduced in chapter 8 was applied to the data with 500 parallel runs of the growing neural networks for each analyte. The ranking of the variables, which is combined for all analytes similar to section 9.1.2, is shown in figure 66. The second step of the algorithm (20-fold random subsampling sets) stopped after the addition of 5 variables, which are labeled in figure 66. Compared with the variable selection by the GA framework, the selection by the parallel growing network framework looks similar, but not identical. Instead of the signal at 20 s the time point 35 s is used (variation of ethanol and 1-propanol) and instead of the time point 55 s the signal at 650 s is used (main variation of 1-propanol). According to table 6, the corresponding optimized neural networks (4 hidden neurons and 1 output neuron) showed the best predictions of all methods used for this data set.

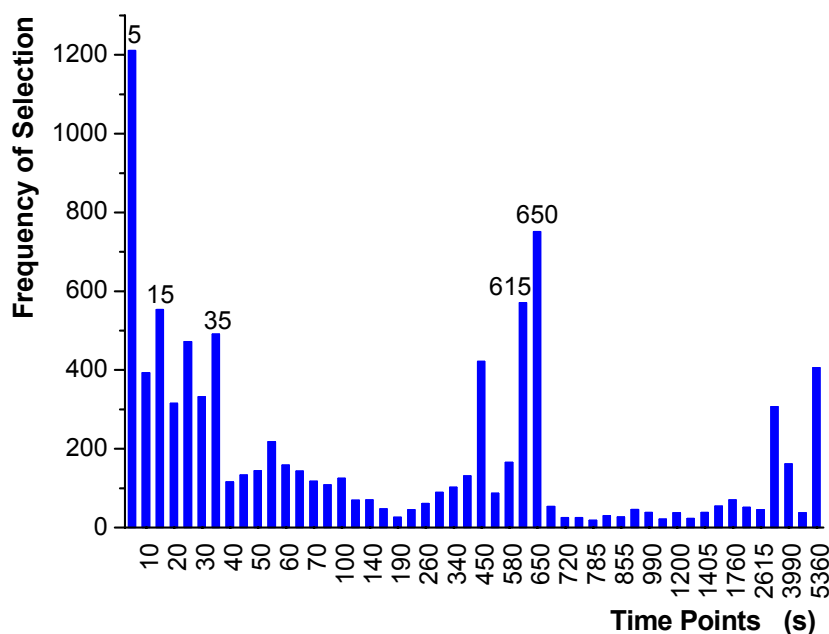


figure 66: Frequency of the variables selected in the first step of the parallel growing neural network framework

The randomization test (200 parallel runs of the growing nets) demonstrates that the parallel growing neural network framework is highly reproducible with the selection of the same 5 variables (see figure 67). When comparing figure 67 with figure 65, it is obvious that the

growing neural nets are less subject to selecting random variables due to chance correlation than the genetic algorithm.

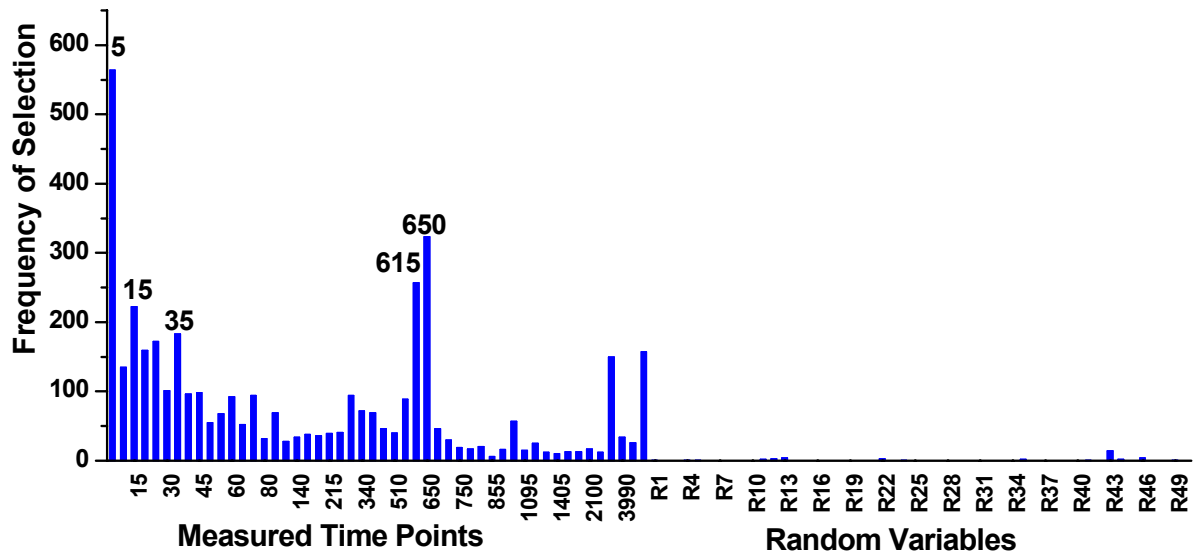


figure 67: Ranking of the variables for 50 time points and for 50 additional random variables.

9.2.5. PCA-NN

Similar to section 6.9, a compression of the input space for the data analysis by neural networks was performed by a principal component analysis. The optimal number of 19 principal components was determined by the minimum crossvalidation error of the calibration data. The predictions of the calibration data are promising whereas the predictions of the validation data are significantly worse (see table 6). The true-predicted plots (figure 68) demonstrate that the predictions of the validation data are biased towards too high predictions.

The bias can be explained by the different amount of noise of the validation and calibration data sets in combination with the nonlinearities in the data sets (see discussion in section 6.9): The linear PCA projection spreads the nonlinearities over many principal components resulting in the selection of the high number of 19 components by the minimum crossvalidation error criterion. On the other hand, the typical noise of the calibration data set is included in these components. Thus, most of these components contain a combination of important information about the model and information about noise. As the validation data set was recorded by averaging two measurements, the noise is significantly reduced resulting in a changed data structure and thus a changed projection by the PCA causing the significant bias of prediction.

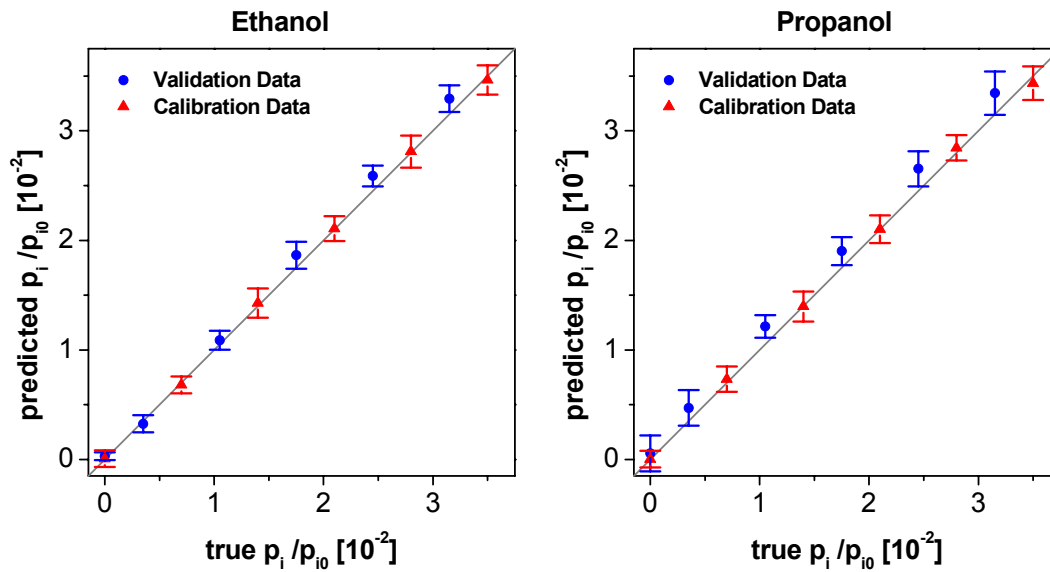


figure 68: Predictions of the calibration and validation data by PCA-NN.

9.2.6. Conclusions

Similar to the previous chapters, it has been demonstrated that both, the genetic algorithm framework and the parallel growing neural network framework are superior to the common methods of multivariate calibration in terms of calibration quality. Both frameworks allow a reproducible variable selection resulting in smaller optimized neural network models with a better generalization ability compared with common neural networks. Both frameworks show reproducible and robust results whereby the parallel growing neural network framework seems to be more robust and shows a slightly better generalization ability, which has also been observed for the binary mixtures of the refrigerants. The classical chemometric methods like the PLS and INLR once more have problems to model the nonlinear relationships of the data similar to chapter 6. It was also shown that the PCA-NN could not deal with changes of the noise in the data resulting in significant biases of the predictions.

9.3. *Methanol, Ethanol and 1-Propanol by the RIfS Array and the 4λ Setup*

The data sets of ternary mixtures of methanol, ethanol and 1-propanol in air, which are described in section 4.5.2.3 in more detail, were recorded for transferring the principle of time-resolved measurements to other setups and to other detection principles. The measurements were performed by the two RIfS devices, the sensor array setup and the low-cost 4λ setup. As the signal to noise ratio is rather poor for the measurement by the RIfS array setup (3 Makrolon layers and 1 PUT layer), it is investigated if a smoothing of the sensor responses can improve the calibration. Additionally single-sensor RIfS setups are simulated by evaluating only single sensors of the array setup. Finally, the low-cost 4λ setup is compared with the array setup.

9.3.1. **Signals and Data Preparation**

The RIfS principle detects changes of the optical thickness nd of the sensitive layer whereby in most cases the changes of the thickness d are dominant [260]. As Makrolon is a hard polycarbonate with a high glass transition temperature, mainly the refractive index n and not the thickness d changes during exposure to analyte resulting in bad signal to noise ratios for the measurements using the 2 RIfS setups (see figure 69). Although the signal to noise ratio can be improved by the use of thicker sensitive layers and longer times of exposure to analyte (for the bigger molecules), both approaches drastically increase the time needed for measurements and between measurements (see figure 25), which is not desired for sensor applications. Additionally, the thickness of the sensitive layer of the 4λ setup can be varied only within certain limits [261]. Therefore, it is investigated if the reduction of noise by the use of a smoothing technique is beneficial for the calibration. In figure 69, the sensor signals of the 80 nm Makrolon layer are shown for 1-propanol and for methanol before and after the application of an FFT filter for smoothing. It is visible that 1-propanol has the poorest signal to noise ratio, as not all micropores are occupied by the analyte in contrast to methanol. Thus, the sensor signals for propanol benefit most from smoothing. On the other hand, the sensor response of methanol shows a counterproductive effect of smoothing. The rectangular sensor signal of methanol before smoothing changes into a rounder sensor profile after smoothing whereas the shape of the wave-like sensor signal of 1-propanol is practically not affected by smoothing. This means that the shapes of the sensor responses of the different analytes are made more similar by smoothing. Thus, the quantification of the analytes is rendered more difficult as the quantification is based on the differences of the shapes. To investigate the

effects of smoothing the data are evaluated separately with and without smoothing. Additionally, the effect of smoothing for sensitive layers with a different thickness is investigated as the thickness influences the signal to noise ratio.

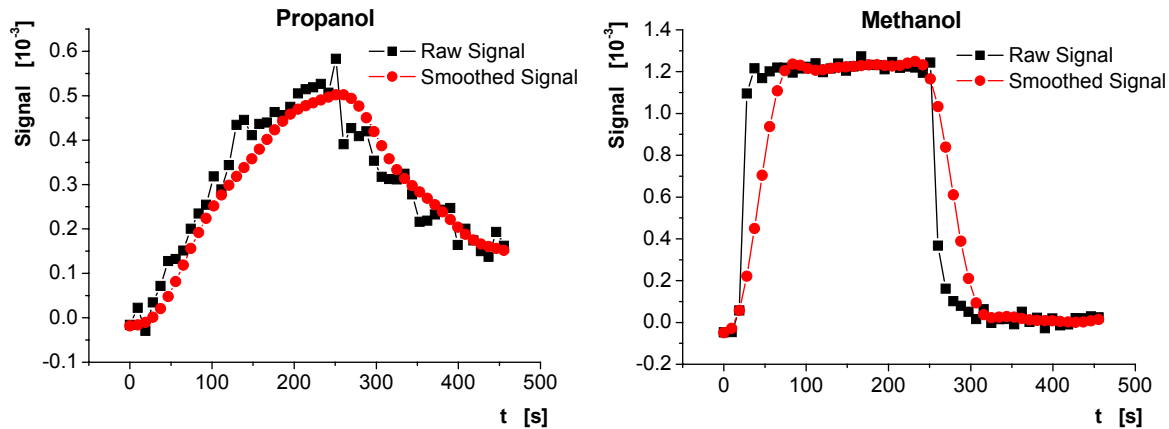


figure 69: Sensor responses before and after filtering with an FFT-Filter for the 80 nm layer.

9.3.2. Mixtures by the RIFS Array

First, a multivariate calibration was performed using the sensor responses of all time points of all 4 sensors resulting in 200 independent variables. For the unsmoothed data, prediction errors of the validation data between 22.18% and 23.96% were achieved (see row 1 of table 7). The prediction errors of the validation data for the smoothed sensor signals are between 14.61% and 34.85% (see row 2 of table 7). When using the sensor signals of all 4 sensors no clear decision can be made if smoothing is beneficial for the calibration. As the 200 input variables contain too much redundant information for an optimal calibration, the parallel growing network framework (50 networks per analyte) was applied to the calibration data of the smoothed and the raw sensor signals. The importance of the different variables is shown in figure 70 as frequency of selection. For the raw signals the 2 Makrolon sensors of 160 nm and 80 nm dominate, whereas the PUT sensor and the 120 nm Makrolon are by far less important. The variable ranking of the smoothed signals looks similar with two differences: Although the 160 nm and 80 nm sensors still dominate, the importance of the other two sensors increased and the important time points of the 80 nm sensor shifted from the end to the beginning of desorption. Compared with the 160 nm layer, the 80 nm layer has gained importance after smoothing.

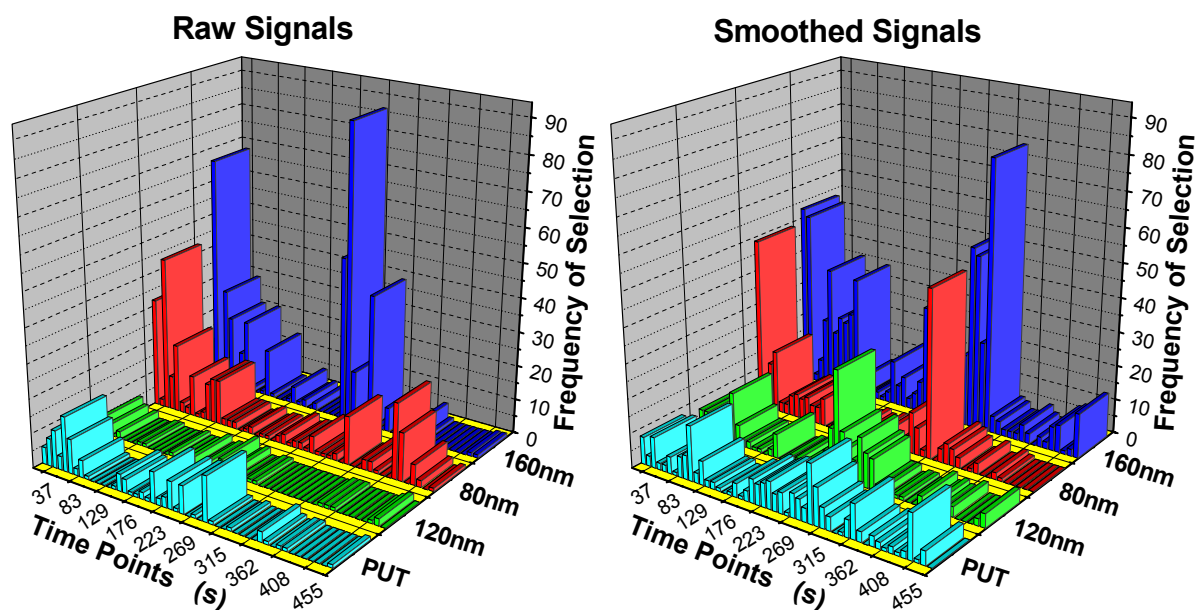


figure 70: Frequency of selection of the different variables after the first step of the parallel growing network framework.

The second step of the growing network framework stopped after the addition of 7 variables for the raw sensor signals and after the addition of 10 variables for the smoothed data. The variable selections for both data sets are similar and very astonishing. For both data sets, only time points of the 80 nm and of the 160 nm Makrolon layer are used. Additionally, only time points within the first 90 seconds of sorption and within the first 75 seconds of desorption are used (instead of 240 seconds of sorption and 210 seconds of desorption) suggesting that faster measurements are possible (see also discussion in section 10.3). Both, the predictions of the validation data and the predictions of the calibration data are significantly better for the raw and the smoothed data when compared with the calibrations using all time points of all sensors (see row 3 and row 4 of table 7). The quantification of methanol is better for the raw data, whereas the quantification of ethanol and 1-propanol is better for the smoothed data whereby for this combination of a thick and a thin layer no method can be generally preferred. The true-predicted plots for the raw data are shown in figure 71.

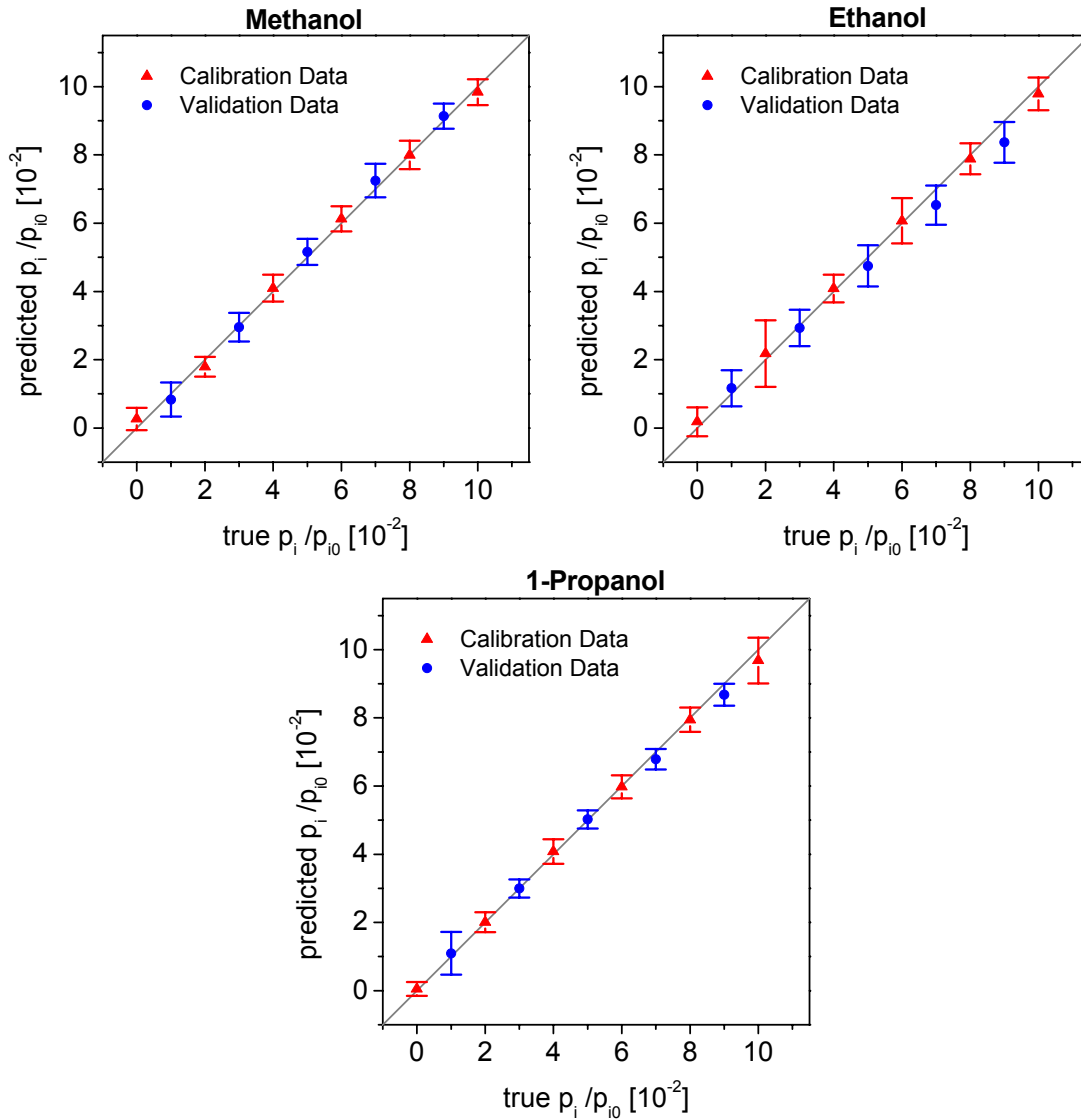


figure 71: True-predicted plots for the raw sensor signals of the array setup whereby only the sensor responses of 2 sensors are evaluated.

In order to see the interactions of the thickness of layers and of smoothing, the sensor responses of the single sensors are calibrated using unoptimized networks (50 input neurons, 5 hidden neurons and 1 output neuron). The predictions of these single sensor calibrations for the raw and for the smoothed data are listed in row 5 to row 12 of table 7. First of all, the single sensor calibrations confirm the variable selection of the framework. The 160 nm layer shows the best calibrations whereas the PUT sensor and the 120 nm Makrolon sensor show poor calibrations. From the chemical point of view, the poor single sensor performance of the PUT sensor can be ascribed to the immediate sensor response without any time-resolution possible whereas the poor performance of the 120 nm Makrolon sensor cannot be explained.

The effect of smoothing is quite interesting for the 3 Makrolon layers with a different thickness. The 80 nm layer clearly benefits from the smoothing while the 160 nm layer shows

worse calibration results if the smoothed sensor signals are used instead of the raw sensor signals. The 120 nm layer with the medium thickness shows no clear preference. The benefits of smoothing for thin layers can be explained by the improvement of the signal to noise ratio overcompensating the changes of the shapes of the sensor responses. On the other hand, the thick layers with a rather good signal to noise ratio are mainly affected by the disadvantageous changes of the shapes of the sensor signals without any real improvement of the signal to noise ratio.

Method	Calibration Data			Validation Data		
	Meth.	Eth.	Prop.	Meth.	Eth.	Prop.
4 Sensors Raw Data	16.21	22.38	21.98	22.77	23.96	22.18
4 Sensors Smoothed Data	14.65	15.11	16.21	34.85	14.61	19.23
Framework Raw	7.86	12.48	8.32	9.17	13.27	7.99
Framework Smoothed	8.81	9.22	6.94	10.32	11.56	7.23
Raw (80 nm M2400)	25.36	24.87	19.26	28.05	31.24	19.65
Smoothed (80 nm M2400)	22.68	20.69	10.78	25.86	22.09	10.58
Raw (120 nm M2400)	21.29	24.27	27.38	24.47	36.61	38.53
Smoothed (120 nm M2400)	23.67	25.27	24.35	26.46	40.99	36.22
Raw (160 nm M2400)	10.54	14.15	12.00	9.81	13.77	11.79
Smoothed (160 nm M2400)	12.57	15.07	13.56	9.91	14.44	14.45
Raw (PUT)	33.72	47.98	14.91	34.49	43.55	12.53
Smoothed (PUT)	35.06	43.36	16.07	45.67	42.39	23.89
4 Sensors Static Eval.	36.61	40.66	37.38	38.73	42.20	37.96
4λ Setup Framework	22.43	24.77	20.87	17.15	25.20	21.32

table 7: Relative RMSE for different data analysis methods and for different setups.

9.3.3. Mixtures by the 4λ Setup

For the quantification of the ternary mixtures measured by the 4λ setup, the data were evaluated with the growing neural network framework. The optimized networks (13 input variables, 5 hidden neurons and 1 output neuron) predicted the validation data with relative errors between 17.15% and 25.2%. The corresponding true-predicted plots are shown in figure 72. The results of the 4λ setup can be best compared with the 80 nm (smoothed) single sensor calibration of the array setup. Yet, the thick layer of the 4λ setup needs significantly more time for the desorption of the analytes (nearly 600 seconds), whereas the 80 nm layer recovers in less than 60 seconds. The 160 nm layer of the array, which needs about 200 seconds for recovery, shows significantly better predictions than the 4λ setup. This means that although the 4λ setup can be successfully used for the multicomponent analysis, the price of miniaturization and simplification has to be paid in terms of longer measurement times or worse calibrations.

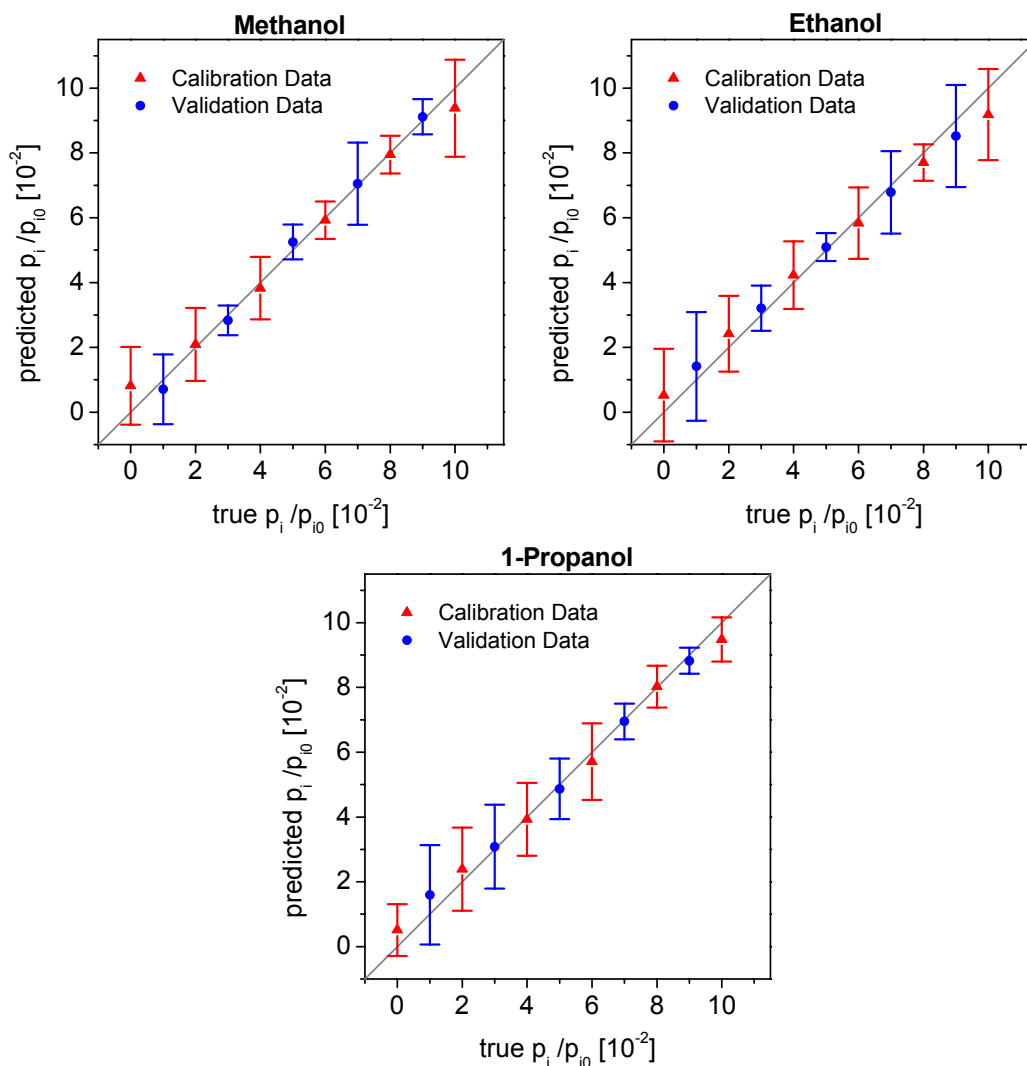


figure 72: True-predicted plots for the 4λ -setup.

9.3.4. Conclusions

Ternary mixtures of methanol, ethanol and 1-propanol measured by the RIFS array setup and by the 4λ setup could be successfully quantified. The application of the growing neural network framework instead of the non-optimized neural networks resulted in significantly improved calibrations again. The variable selection of the framework for the array setup is quite astonishing, since only the sensor signals of 2 sensors out of 4 sensors are used. As the framework selects the most predictive variables, it can be concluded that the time domain of 2 sensors contains more information than the parallel (static) information of all 4 sensors together. This is an impressive demonstration, how the time-resolved measurements of few sensors can render the application of many parallel sensors with different sensitivities redundant. The static evaluation of the 4 sensors, which corresponds to the sensor signals at the end of exposure to analytes, shows that the time-resolved evaluation of the sensor signals is highly superior even though the static sensor evaluation is not a mathematically underdetermined system. It was demonstrated that smoothing improves the calibration of measurements performed by thin sensitive layers whereas the calibration deteriorates for measurements performed by thick sensitive layers when smoothing the sensor signals. Furthermore, the 4λ setup can be used as single sensor device for a multicomponent quantification. Compared with the array setup, the price of miniaturization and cost reduction has to be paid in terms of extended measurement times respectively higher prediction errors.

9.4. Quaternary Mixtures by the SPR Setup and the RIfS Array

9.4.1. Introduction

In this chapter, the quaternary mixtures of methanol, ethanol, 1-propanol and 1-butanol measured by the RIfS array and the SPR setup are investigated. This allows a comparison of the two setups. The RIfS array shows a significantly worse signal to noise ratio than the SPR setup using Makrolon as sensitive layer. Thus, more time of exposure to analyte was necessary for the RIfS array setup than for the SPR setup. Additionally, it is investigated if the smoothing of the data of the RIfS array can improve the calibration (similar to the last section). More details of the data sets can be found in section 4.5.2.4.

9.4.2. Results

For all data sets, several calibrations using the corresponding calibration data set were performed, whereby the predictions of the external validation data are summarized in table 8. First of all, calibrations were performed using all time channels of the SPR setup and all smoothed time channels of the 3 sensors of the RIfS array setup (1st and 5th row of table 8). Then the same data sets were used for the parallel growing neural network framework (2nd and 6th row). It is obvious that for both setups the framework significantly improves the calibration of all analytes. The SPR setup performs better on the calibration of methanol and ethanol and the RIfS array with 3 sensors performs better on the calibration of 1-propanol and 1-butanol whereby the overall mean predictions of the RIfS array with 3 sensors are better. The predictions of the calibration and the validation data are shown in figure 73 for the SPR setup (the neural networks had been optimized by the framework). For both setups, the framework selected time points spread over all sensors and over the complete process of sorption and desorption. An interesting point is the restriction of the available time points to a certain shorter time interval, which would correspond to faster measurements. The restriction of the data analysis for the SPR measurements to 240 seconds (120 seconds of sorption and 120 seconds of desorption) instead of 930 seconds hardly had a negative impact on the calibration, whereas the restriction to 120 seconds of sorption significantly decreased the calibration performance. Decreasing the analysis time of the RIfS array from 3470 seconds to 640 seconds deteriorated the prediction performance less than decreasing the analysis time from 640 seconds to 236 seconds, which corresponds to the sorption process only. This means that for both setups the measurement time can be drastically reduced without too high an impact on the prediction performance. Yet, it is important to consider not only the sorption

process but also the beginning of the desorption process. This is consistent with most variable selections by the frameworks in the previous sections (8.4.1, 9.1.2, 9.2.3, 9.2.4 and 9.3.2) with only variables selected directly after the beginning of exposure to analyte and directly after the end of exposure to analyte (see also discussion in section 10.3).

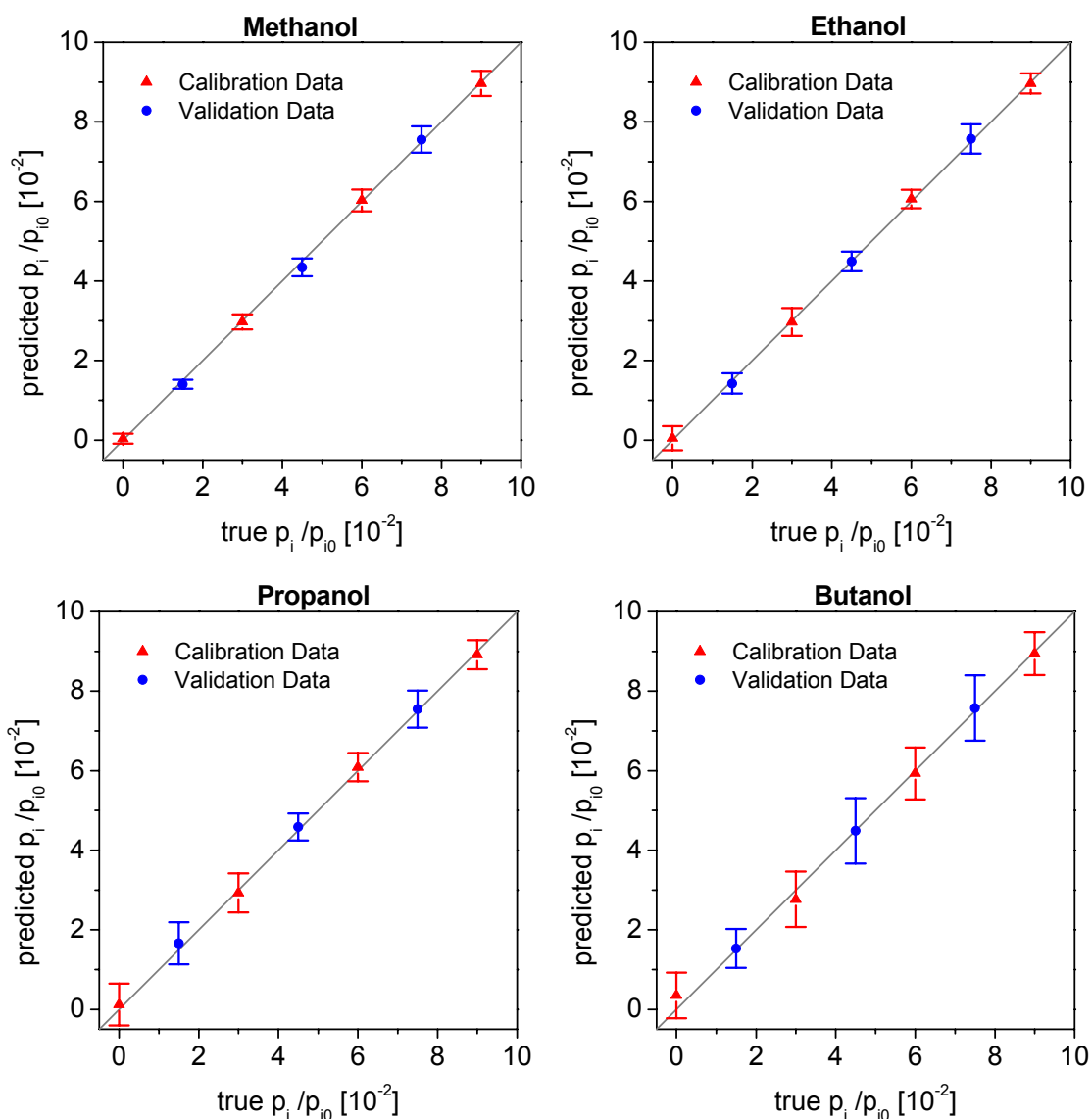


figure 73: True-predicted plots of the SPR setup using the optimized networks and the time points of the complete measurement time.

The effect of smoothing the sensor signals of the RIfS array was investigated for the single sensor responses of the two Makrolon layers. Similar to section 9.3.2, the thinner 95 nm layer benefited from the smoothing whereas the thicker 165 nm layer was adversely affected by smoothing. The single sensor predictions of the RIfS array are all worse than the single sensor predictions of the SPR setup even with nearly 4 times longer measurement times of the RIfS setup. As already discussed in section 9.3.1, the RIfS setup, which primarily detects changes

of the thickness d of the sensitive layer [260], is more affected by Makrolon mainly changing the refractive index n when exposed to analyte, than the SPR setup, which mainly detects changes of the refractive index n . Nevertheless, the comparison with the static sensor evaluation (only the highest sensor signals of each sensor), which is an undetermined system with 3 sensors for 4 analytes and which consequently shows very high prediction errors, demonstrates the potential of the time-resolved measurements also for the RIfS principle with many possibilities for further developments (last row of table 8).

Method	Meth.	Eth.	Prop.	But.	Mean
SPR, unoptimized	9.6	11.0	16.4	19.0	14.0
SPR, optimized	5.8	6.5	10.2	16.0	9.6
SPR, optimized, <240 s	5.4	4.5	12.7	20.4	10.7
SPR, optimized, <120 s	13.5	12.2	15.8	20.9	15.6
RIfS, 95 + 165 + PUT, unopt., smoothed	15.1	13.6	12.9	12.1	13.4
RIfS, 95 + 165 + PUT, opt., smoothed	10.1	10.0	8.2	6.5	7.3
RIfS, 95 + 165 + PUT, opt., sm., < 640 s	9.9	10.0	9.7	7.2	9.2
RIfS, 95 + 165 + PUT, opt., sm., < 236 s	14.1	12.8	12.2	9.1	12.1
RIfS, 95 optimized, raw	12.6	11.7	14.2	20.3	14.7
RIfS, 95 optimized, smoothed	12.5	10.5	12.6	18.1	13.4
RIfS, 165 optimized, raw	19.6	17.6	13.9	15.5	16.7
RIfS, 165 optimized, smoothed	22.1	19.6	15.8	18.1	18.9
RIfS, 95 + 165 + PUT, 1 static time point	43.8	43.9	41.1	22.5	37.8

table 8: *Relative RMSE of the validation data in % for the 4 analytes and the mean using different data analysis methods, different setups and different constraints of time points (95 = 95 nm Makrolon layer, 165 = 165 nm Makrolon layer, PUT = PUT layer; optimized = growing neural network framework, unoptimized = all time channels, smoothed / raw sensor signals).*

9.4.3. Conclusions

It was demonstrated that by the use of time-resolved measurements quaternary mixtures can be quantified, which were measured by sensor setups employing less sensors than analytes to be quantified. Thereby the 1-sensor SPR setup practically achieves the same results than the RIfS array using 3 sensors, whereas the evaluation of single RIfS sensors showed worse results than the SPR setup. The evaluation of single time points of the 3-sensor RIfS setup, which corresponds to the common static sensor evaluation, showed unacceptably bad results, as the data analysis was mathematically underdetermined (3 sensor responses for 4 analytes). This demonstrates how the principle of time-resolved measurements can help to reduce the number of sensors and thus the hardware costs. The principle allows the quantitative determination of systems, which would never have been quantified using static measurements. Similar to section 9.3, it was demonstrated that the influence of smoothing the time-resolved sensor responses is two-sided. Smoothing of very noisy sensor responses of thin sensitive layers improves the calibration whereas smoothing of sensor responses with a high signal to noise ratio adversely influences the calibration. Additionally, it was demonstrated once more that the frameworks introduced in this work help to improve the multivariate calibration of the sensor signals of all data sets investigated.

9.5. Quantification of the Refrigerants R22 and R134a in Mixtures: Part II

This section is an example how the calibration of old data sets can be improved by a time-resolved data analysis. The data set, which was recorded by the RIFS array with different polymers, was already introduced and investigated in chapter 3. During the development of the time-resolved measurements, this data set was investigated once more, since the sensor response of the Makrolon layer had also been recorded in a time-resolved mode (see section 4.5.1.2).

The parallel growing neural network framework (250 parallel runs for each analyte) was applied to the data set with all 23 variables. The framework selected 6 variables for an optimal prediction. Among these variables the 3 static sensor responses of HBP, PDMS and UE 2010 15% and additionally 3 time-resolved variables of Makrolon were selected. Using these variables instead of the 6 static sensor responses the test data were significantly better predicted with an improvement of 33% for the RMSE of R22 and 9% for the RMSE of R134a (see table 9). The parallel framework was also applied to the calibration data using only the time-resolved sensor signals of Makrolon. The predictions of the optimized neural nets with 18 variables are shown in table 9. Compared with the static sensor signals of 6 sensors, the predictions using the time-resolved sensor signals of the Makrolon sensor were by 12% better for R22 and by 68% worse for R134a. When comparing the mean error of the test data of this single sensor approach with the 15 combinations of the 2-sensor approach (see table 1 in section 3.4) it is astonishing that the time-resolved single sensor approach is better than most of the 2-sensor approaches (6th position).

Variables	Test R22	Test R134a	Test Mean
6 Static	0.00183	0.00630	0.00406
3 Static + 3 Time-resolved	0.00141	0.00575	0.00358
18 Time-resolved	0.00161	0.01061	0.00611

table 9: RMSE of the time-resolved data analysis compared with the static data analysis performed in chapter 3, see also table 1.

As a conclusion of this section it might be said that even for data sets, which were recorded without the intention of a time-resolved data analysis and without any optimization of the measurement parameters, the time-resolved data analysis can improve the calibration.

10. Results – Various Aspects of the Frameworks and Measurements

In this chapter, several minor issues concerning the frameworks and the time-resolved measurements are dealt with using the data of the previous chapter.

10.1. Single or Multiple Analyte Rankings

In principle, the variable selection of the frameworks can be performed for each analyte (response variable) separately or for all analytes together. For example, the variable selection by the parallel growing neural network was performed for each analyte separately in section 8.4.1 by using a separate ranking for both analytes, whereas the variable selection in the sections 9.1.2, 9.2.4, 9.3.2 and 9.3.3 was performed for all analytes together by using a combined ranking for all analytes together (addition of the individual rankings). In the second step of the iterative variable addition procedure, the combined ranking saves about n times the computing time for n analytes. Yet the question is, if a separate variable selection performs significantly better.

Thus, the ternary mixtures of methanol, ethanol and 1-propanol measured by SPR and evaluated by the parallel growing network framework (see section 9.2.4) were evaluated again using a separate ranking for each analyte and separate iterative addition steps (step 2 of the framework). The RMSE of the validation data, which are shown in table 10, are very similar and do not allow a clear decision which method to prefer.

Variable Selection	Meth.	Eth.	Prop.
Together	3.31	6.05	7.33
Separate	3.30	6.15	7.25

table 10: *Relative RMSE in % for the prediction of the validation data measured by SPR and evaluated by the parallel growing neural network framework.*

Evaluating the parallel growing network framework of the refrigerant data together instead of separately (compare with 4th row of table 4 in chapter 8) resulted in slightly worse predictions of 2.08% for the validation data of R22 instead of 2.04% and slightly better predictions for R134a with 2.53% instead of 2.61%. For other data sets, also no clear preference can be seen rendering the more computing intensive separate variable selection superfluous.

10.2. Stopping Criteria for the Parallel Frameworks

In the second step of the genetic algorithm framework and of the parallel growing neural network framework, variables are added to the model until the prediction errors evaluated by a subsampling procedure does not improve any more. A difficult question is, how the significance of this improvement should be judged (stopping criterion for the addition of variables). Many different approaches can be found in literature, which can be classified into several categories like significance tests or numerical comparisons, robust or non-robust tests, paired or non-paired tests and local or global error minima.

In this study, 6 different methods were used to determine the optimal number of variables. First of all, the simple numerical mean prediction errors of the subsampled test data were compared before and after the addition of a variable. Thereby the addition of variables was stopped when the first local minimum of the mean prediction error was found. A second approach calculates all mean prediction errors and uses the number of variables, which corresponds to the global minimum of the prediction errors. Commonly used methods to judge the improvement of predictions are based on statistical significance tests. An overview of the different tests can be found in [232]. The significance tests were implemented in the frameworks for stopping the addition of variables when the test determines the improvement of the predictions after the addition being not significant (see also figure 44 and figure 53). The most popular statistical test to compare the predictions of the subsampled test data before and after the addition of a variable is the Student T-test [254]. The T-test needs a normal distribution of the prediction errors (this can be checked by a Kolmogorov Smirnov test [38]) and thus is sensitive to outliers. A robust option for comparing the predictions of the test data subsets is the Kruskal Wallis Anova [262,263], which corresponds to the Man-Whitney U-test, as only two groups are compared. If the partitioning of the subsampling procedure is reproducible for each addition of a variable (this means that the same test subsets are predicted during each loop of the variable addition) paired significance tests can be used like the paired T-test for normally distributed prediction errors and the Wilcoxon signed rank test as its robust counterpart [103,264,265]. The different categories of the significance tests have different requirements, which can be summarized as follows. In contrast to robust tests, T-tests need a normal distribution of the prediction errors and thus are sensitive to biases and outliers whereas robust tests are less powerful in terms of detecting differences of the prediction abilities. The paired tests need the same partitioning of the data into calibration and test subsets for each loop of the variable addition step. In contrast to finding the global

minimum of the prediction errors, the implementation of the significance tests needs only as many loops as improvements of the prediction errors are observed.

The number of variables selected depends not only on the method, but also on the significance level of the statistical test, which was set to 5 % error probability for all tests. In principle, the different methods can be divided into 4 groups according to the number of variables selected. The T-test, the Kruskal Wallis Anova and the Wilcoxon signed rank test were most conservative in terms of selecting variables. These three tests selected the same small number of variables for all data sets under investigation in this work. The paired T-test selected in most cases some more variables followed by the criterion of the first local minimum of the prediction errors whereas the method of the global minimum of the prediction errors generally corresponded to more variables. All these methods are based on the prediction errors of the subsampled test data and not on an external data set. The question is how the prediction errors of the subsampled data correspond with the prediction errors of external validation data. The answer can be found in the so-called biasing [11], which means that when the same data are used for a model building process and for the variable selection process, the variable selection is biased towards selecting too many variables whereby the bias increases with a decreasing number of samples. As the subsampled data are used several times for both processes, the optimal method depends on the sample size. For large data sets, the global minimum of the prediction errors of the subsampled test data corresponds with the smallest prediction errors of external validation data whereas for smaller data sets methods that are more conservative correspond with the best errors of the validation data. This effect could be observed for all data sets under investigation. For the rather large refrigerant data set (441 samples for the calibration of only 2 analytes), the optimal method was the first local minimum criterion, whereas for the small quaternary mixtures (256 samples for the calibration of 4 analytes) and the ternary mixtures (245 samples for 3 analytes) the optimal method was the Kruskal Wallis Anova.

Although the selection of the stopping criterion influences the prediction ability of the frameworks, an investigation using all data sets of this work showed that the selection is less critical than supposed at first glance. Among all data sets, the highest difference of the prediction errors of external validation data was 0.4% when using different stopping criteria for the calibration data. The general recommendation of measuring as many samples as possible renders a sophisticated recommendation for a stopping criterion rather unnecessary, as in the case of not too small data sets, the local or global minimum criteria are adequate.

10.3. Optimization of the Measurements

Among the many parameters to be decided on and to be adjusted, the scanning speed of the time-resolved sensor responses has been an often-discussed subject during the measurements for this work. A slow scanning of the sensor responses over time results in a low number of time points allowing a calibration without a variable selection or at least allows significantly speeding up the variable selection procedures. On the other hand, a slow scanning of the sensor responses might miss the differences between the sensor responses of analytes, which show a very similar kinetics. To investigate this topic a little bit more in detail, fully connected neural networks were trained using the refrigerant data set whereby the number of time points was systematically reduced by using only each 2nd, each 3rd... time point. In table 11, the prediction errors are shown, which decrease with an increasing number of time points corresponding with an increasing scanning speed. This table also demonstrates that only a sophisticated variable selection procedure improves the performance of calibration and prediction (compared with table 3 and table 4).

Method	Calibration Data Set		Validation Data Set	
	R22	R134a	R22	R134a
Each Time Point	1.5	2.6	2.2	3.3
Each 2 nd Time Point	2.0	3.0	2.4	3.3
Each 3 rd Time Point	2.2	3.1	2.7	3.4
Each 4 th Time Point	2.4	3.2	2.8	3.5
Each 5 th Time Point	2.9	3.5	3.2	3.8
Each 10 th Time Point	4.5	3.7	4.9	4.1
Each 20 th Time Point	21.9	55.2	21.6	52.1

table 11: Relative RMSE in % for the prediction of the refrigerant data set by fully connected neural networks, which use each n^{th} time point simulating a slower scanning of the time-resolved sensor response.

Also the variable selection by the frameworks gives an indication of an optimal scanning speed for the time-resolved sensor responses. Practically for all variable selections by the frameworks of the previous chapters, many of the variables selected were adjacent in time. For example it is shown in figure 46 that 9 out of 12 time points within the time interval 67 s to 93 s are selected demonstrating that nearly all information of the selected interval is evaluated and that a further increase of the scanning speed might yield even more useful information.

The fact that variables are selected and used only within few intervals is also known in PLS and has been subject to some further developments of the PLS known as Interval Partial Least Squares (IPLS) [266]. It has often been stated that the collinearity of a certain number of variables stabilizes the predictions [41] whereby too high a number of collinear variables negatively affects the predictions (see also section 2.8).

For practically all selections of the variables by the frameworks (for example in the sections 8.4.1, 9.1.2, 9.2.3, 9.2.4 and 9.3.2), the variables are located directly after the beginning of exposure to analyte and directly after the end of exposure to analyte. This implies that not the complete measurement time is needed for the determination of the sample composition, but only a short interval of exposure and after that a short interval of analyte desorption. It also implies that the time of exposure to analyte can be reduced, which also results in a faster desorption of the analyte (like a synergetic effect) and consequently reduces the time needed between measurements. For this work, the time used for exposure to analyte and a subsequent recovery had been determined by visually inspecting the sensor responses of single analytes (like figure 24) and then by choosing the time interval, for which the shape of sensor responses significantly differ. For the routine analysis, the calibration should be repeated measuring only during the time intervals proposed by the frameworks, which will save time and money.

The number of measurements which have to be performed for a calibration is also a significant point, which has to be decided on when planning an experimental design. As the number of measurements for a full factorial design strongly increases with the number of analytes and the number of concentration levels (see equation (1)), the number of concentration levels for the calibration of ternary and quaternary mixtures was rather low compared with the binary mixtures of the refrigerants. The price to be paid for calibrating with a 4-level design (used for the calibration of the quaternary mixtures) instead of a 21-level design can be estimated by using only 16 calibration samples instead of 441 samples for the refrigerant data.

The mean relative RMSE of the validation for the non-optimized neural networks increases thereby from 2.7% for the 21-level design to 6.7% for the 4-level design. Thus, it is expected that the calibrations of the ternary and especially of the quaternary mixtures can be significantly improved by measuring more calibration samples.

The choice of the optimal thickness of the sensitive layer depends on several parameters, which are partly discussed in chapter 5 and in the results in more detail and which will be only summarized here. A thick layer means a slow kinetics of the analytes allowing the discrimination of very small and similar analytes. On the other side, big analytes need a very long time until a sensible sensor response can be observed resulting in long measurement times. Thin layers, which allow fast measurements can only be used in some setups due to a low signal to noise ratio, whereby a smoothing of noisy signals can improve the calibration (in contrast to smoothing the nearly noise-free signals of thick layers). Among the different devices, the SPR setup is most appropriate for time-resolved measurements using Makrolon, but needs the most complex equipment (like an exact constancy of the temperature). The 4λ setup is the smallest and cheapest device but is only fairly appropriate for Makrolon as sensitive layer, whereas the RIfS array setup can be found between the former two setups in respect to all concerns.

Thus, no general recommendation except of a highest possible scanning speed of the sensor responses in combination with a variable selection and a highest possible number of calibration samples can be given for most parameters, as the optimal solution is determined by the analytes under investigation, by external conditions like the allowed time for each measurement, the demanded robustness of the devices and much more.

10.4. Robustness and Comparison with Martens' Uncertainty Test

In the field of calibration by the PLS, a method called Martens' Uncertainty Test [32,33,41] has gained increasing popularity during the last two years. This test uses a jackknifing procedure with many submodels to identify non-significant variables. Thereby a statistics is setup for the regression coefficients of all variables and then those variables are eliminated, which are identified as being non-significant according to this statistics. The genetic algorithm framework and the parallel growing network framework introduced in this work use a similar method of selecting only significant variables but work the other way round. Instead of eliminating variables, the frameworks add variables according to a ranking, which was established by many submodels in a previous step, until the prediction does not significantly improve determined by a subsampling process. The frameworks are generally more conservative in terms of selecting variables compared with Martens' Uncertainty Test. The significance determined by the subsampling process in the second step of the frameworks can also be used to access the uncertainties respectively the robustness of the predictions. Thereby the standard deviations for the predictions of the subsampled test data by the different submodels are calculated during the subsampling process [267]. For example, the uncertainties of the predictions of the refrigerant data for the parallel growing neural networks framework (4th row of table 4) were estimated as 0.17% for R22 and 0.14% for R134a in terms of subsampled standard deviations. For the evaluation of the same data by the genetic algorithm framework, the standard deviations are also low with 0.11% for R22 and 0.18% for R134a. The ternary mixtures of the alcohols measured by SPR showed uncertainties of 0.27% for methanol, 0.32% for ethanol and 0.34% for 1-propanol evaluated by the parallel growing neural network framework respectively 0.21% for ethanol, 0.25% for ethanol and 0.30% for 1-propanol evaluated by the genetic algorithm framework. Thus, the calibrations by the frameworks can be considered as generally being quite robust.

11. Summary and Outlook

In this work, time-resolved measurements and time-resolved data analyses of sensor responses have been introduced in the field of chemical sensing. The time-resolved measurements, which were systematically investigated for several sensor setups and for several analytes, can be regarded as a second major step in the field of sensor development for multianalyte determinations. The first step from selective sensors to cross-reactive sensor arrays had allowed the parallel quantifications of different analytes by the use of the signal patterns of one single array of sensors without the need of finding selective sensor materials for each analyte. This first step became very popular in the field of electronic noses during the 80s and 90s. The second step, the time-resolved evaluations of sensor signals in combination with suitable sensor coatings, combines the sensory principle with the chromatographic principle of separating analytes in space or time. This opens the door to a completely new dimension of information in chemical sensing. This work is the first extensive and systematic investigation of this second step for an improved and advanced quantitative determination of analytes in the field of chemical sensing.

The time-resolved measurements of this work are all based on the microporous polymer Makrolon as sensitive layer. This polycarbonate allows a kinetic separation of the analytes during the sorption and desorption on the basis of the size of analytes. It was shown using up to quaternary mixtures of the low molecular weight alcohols as analytes that small molecules can sorb very fast into the pores whereas the larger molecules sorb only slowly into the polymer. It was demonstrated that this specific sorption into the pores is a Langmuir type sorption, which is by far more important than the unspecific Henry type sorption of the analytes into the polymer matrix. An additional effect of an expansion of the pores during long-term exposure to analyte was observed. It was demonstrated that this effect could be exploited to measure bigger analytes by expanding the pores using other carrier gases than air. It was also shown that the variation of the thickness of the sensitive layer allows the tailoring of the sensitive layer to specific analytical questions.

The time-resolved measurements have been successfully used for three different sensor setups and for many multicomponent mixtures of the low alcohols and the refrigerants R22 and R134a. It was demonstrated that the time-resolved measurement principle can be applied to single sensor setups allowing the simultaneous quantification of several analytes and consequently rendering arrays of sensors unnecessary. It was furthermore shown that the

time-resolved measurement principle can also be applied to sensor arrays with the results of an improved calibration, a higher robustness, an increased flexibility to the number and properties of different analytes and a reduced number of sensors.

Generally speaking, the time-resolved measurement principle allows the reduction of the expenses for hardware at the cost of a more extensive and a more complicated data analysis. This leads over to the second objective of the work, the data analysis. It was shown that the best results for multianalyte quantifications are obtained when the measurements are performed with the highest possible scanning rate of the sensor responses and the highest possible number of measurements for calibration. The resulting increased number of input variables (the time points generated by the scanning of the sensor responses) and the nonlinear relationship between the sensor responses and the concentrations of the analytes put new challenges to the data analysis. It was demonstrated that most common methods for a multivariate data analysis like PLS, QPLS, INLR, CART, MARS and neural networks showed rather poor calibration results. Among these methods, the neural networks were most promising but had to struggle with the high number of correlated and redundant input variables resulting in improvable calibrations. The combination of variable selection methods and of neural networks, which is widely used in literature to solve the issue of too many redundant and correlated input variables, could not help due to the limited number of samples measured.

In order to find a calibration with the highest possible calibration and generalization ability three frameworks were innovated, implemented and optimized in this work. These frameworks use repeated runs of a combined variable selection and calibration with different subsets of the available data resulting in a very effective exploitation of the limited number of data. One framework is based on many parallel runs of genetic algorithms combined with neural networks, one framework bases on many parallel runs of growing neural networks and the third framework uses several runs of the growing neural networks in a loop. All three frameworks showed by far better calibrations than all common methods of multivariate calibration and than simple non-optimized neural networks for all data sets investigated. Additionally, the variable selection of these frameworks allowed an insight into the relationship between the time-resolved sensor responses and the concentrations of the analytes. The variable selection also suggested optimizations in terms of shorter measurements for several data sets. The variable selection quality of the parallel growing network framework could be confirmed by a brute force variable selection. The calibrations

and variable selections of all three frameworks were reproducible and were not disturbed by noise in the data. All three frameworks successfully solved the problems of too many variables for too few samples and the problems caused by the nonlinearities present in the data with practically no input needed by the analyst. Thus, all three frameworks showed excellent calibration and variable selection qualities whereby each framework has its own benefits. The genetic algorithm framework is the fastest framework whereas the parallel growing neural network framework shows a slightly better calibration. The loop-based growing neural network framework shows the best calibration performance as it allows building complicated yet sparse non-uniform neural networks. All three frameworks are not limited to time-resolved sensor data, but can be used for nearly any data when a powerful variable selection and calibration are needed and when the number of samples is limited. In the area of data-mining and pattern recognition, the application of these framework has also shown excellent results for data sets from medicinal chemistry [268].

Together, both main focuses of this work impressively demonstrate how the combination of an advanced measurement principle and of an intelligent data analysis can improve the results of measurements at reduced hardware costs. To prevent misunderstandings, an intelligent data analysis and an advanced measurement principle cannot help if a device provides bad or senseless data. However, the amount of information provided by a device can often be dramatically increased by using advanced measurement principles (like the time-resolved measurements of this work). Yet, it was also demonstrated in this work that additionally new intelligent methods of data analysis are needed, which are able to extract and use the valuable information out of the large pool of information provided by the advanced measurement principles (such as the frameworks introduced in this work). It was also shown that the results of the data analysis give feedback how the measurement principles, the measurement parameters and the devices can be optimized and improved. This demonstrates how the interconnection of the different parts of an analysis can improve the complete analysis in a synergetic effect.

Starting with this work further research can be performed in many fields of scientific research. Beginning with the sensitive layer, the principle of different-sized pores as size-sensitive recognition elements can be further investigated. For example it was shown in [269] that there are many other polymers with pores of different sizes like the Compimide 183 with a mean pore size of 0.038 nm^3 , the Polyimide PI2611 with a mean pore size of 0.058 nm^3 and the Polyimide PI2566 AL with a mean pore size of 0.13 nm^3 and many more. These polymers

allow extending the range of analytes to bigger and smaller molecules. The combination of these polymers in an array should result in a powerful setup for a size-selective discrimination of a broad range of analytes. Especially the extension of the SPR-device to an array setup seems to be very promising as the SPR setup demonstrated to be the most suited device for measurements using microporous polymers as sensitive layers. Furthermore, the principle of the time-resolved measurements is not limited to optical sensor devices but can be used for practically any arbitrary (sensor)-device like electronic noses, biosensors and many more as long as the sensor responses differ in the time domain. Thereby the recognition principle is not limited to size-selective recognitions but can be of any specific type that allows time-resolved discriminations. For example in the area of biosensing, different DNA with a different number of mismatches might be quantified simultaneously by differences of the DNA-DNA binding kinetics. Also, different antibodies might be discriminated on the basis of the kinetics, if the different antibodies show different adsorption kinetics due to different sizes of the FAB fragments. This allows single sensor applications for several selective and even cross-reactive analytes [270,271].

The combination of several sensors with different sensitive polymers for time-resolved measurements on a sensor array opens the door to second-order calibrations similar to GC-MS setups. Thereby the sensor signals represent the first order and the time represents the second order. Second-order calibrations allow the quantification of an analyte in the presence of unknown interferences, which is also known as second-order advantage. For example, the generalized rank annihilation method (GRAM) [272,273] can already work with a single standard addition to the prediction sample. Consequently, the extensive calibrations with experimental designs can be completely avoided resulting in dramatically reduced expenses for the calibration of specific analytes. Yet, further research has to be done concerning two topics. First of all, more polymers are needed, which allow time-resolved measurements and which show different chemical properties, as the second order advantage requires sufficient selectivity in both orders. Additionally, the second-order methods have to be further studied in respect to dealing with nonlinear relationships, as most of the up-to-date algorithms assume linear relationships in both orders.

An interesting approach similar to time-resolved measurements is the application of temperature-resolved measurements. Kato et al. [274] demonstrated that different analytes show different dynamic sensor responses if the sensor signal is recorded during a variation of the sensor temperature of tin oxide sensors. Mielle et al. [275] used a single tin oxide sensor

to discriminate 9 analytes measured at 6 different temperatures. These approaches are not limited to metal-oxide sensors but can also be used for polymer-based sensors. As long as the sorption kinetics of the various analytes depends in different ways on the temperature, the temperature-resolved measurements allow exploiting an additional information domain. A very interesting point is also the glass transition temperature of a polymer. Measurements below the glass transition temperature should show a more specific sorption behavior whereas measurements above the glass transition temperature should show a more unspecific sorption doubling the information provided by a sensor.

In summary, it may be said that this work once more demonstrates that not the lack of information is the limit for chemical sensing but the frontier of scientific research, which makes this information available and understandable for the analyst and this frontier is moving from day to day opening the doors to new possibilities in scientific research.

12. References

- [1] B.K. Lavin and J. Workman, *Anal. Chem.*, 74 (2002) 2763-2770.
- [2] D.H. Doehlert, *Appl. Statist.*, 19 (1970) 231-239.
- [3] R.H. Meyers and D.C. Montgomery, *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, Wiley-Interscience, New York, 1995.
- [4] L. Erikson, E. Johansson and C. Wikström, *Chemom. Intell. Lab. Syst.*, 43 (1998) 1-24.
- [5] M.E. Johnson and C.J. Nachtsheim, *Technometrics*, 25 (1983) 271-277.
- [6] V.V. Federov, *Theory of Optimal Experiments*, Academic Press, New York, 1972.
- [7] P.C. Jurs, G.A. Bakken and H.E. McClelland, *Chem. Rev.*, 100 (2000), 2649-2678.
- [8] F. Despagne and D. L. Massart, *Analyst*, 123 (1998) 157R-178R.
- [9] I.H. Witten and E. Frank, *Data Mining*, Morgan Kaufmann Publishers, San Francisco, 1999.
- [10] B.M. Smith and P.J. Gemperline, *Anal. Chim. Acta*, 423 (2000), 167-177.
- [11] A.M. Kupinski and M.L. Giger, *Med. Phys.*, 26 (1999) 2176-2182.
- [12] D. Broadhurst, R. Goodacre, A. Jones, J.J. Rowland and D.B. Kell, *Anal. Chim. Acta*, 348 (1997) 71-86.
- [13] M. Clementi, S. Clementi, M. Fornaciari, F. Orlandi and B. Romano, *J. Chemometrics*, 15 (2001) 397-404.
- [14] H. A. Martens and P. Dardenne, *Chemom. Intell. Lab. Syst.*, 44 (1998) 91-121.
- [15] T. Naes, *Chemom. Intell. Lab. Syst.*, 5 (1989) 155-168.
- [16] A. Höskudsson, *Chemom. Intell. Lab. Syst.*, 32 (1996) 37-55.

- [17] J. Shao, *J. Am. Stat. Assoc.*, 88 (1993) 486-494.
- [18] B.M. Smith and P.J. Gmperline, *J. Chemometrics*, 16 (2002) 241-246.
- [19] R.R. Picard and R.D. Cook, *J. Am. Stat. Assoc.*, 79 (1984) 575-583.
- [20] M. Forina, G. Drava, R. Boggia, S. Lanteri and P. Conti, *Anal. Chim. Acta*, 295 (1994) 109-118.
- [21] M. Baroni, S. Clementi, G. Cruciani, G. Costantino and D. Riganelli, *J. Chemometrics*, 6 (1992) 347-356.
- [22] Q.-S. Xu and Y.-Z. Liang, *Chemom. Intell. Lab. Syst.*, 56 (2001) 1-11.
- [23] R.W. Kennard and L.A. Stone, *Technometrics*, 11 (1996) 137-148.
- [24] B. Bourguignon, P.F. de Aguiar, K. Thorré and D.L. Massart, *J. Chromatogr. Sci.*, 32 (1994) 144-152.
- [25] D.R. Rogers and A. J. Hopfinger, *J. Chem. Intell. Lab. Syst.*, 9 (1994) 854-866.
- [26] J. Zupan, M. Novic and I. Ruisánchez, *J. Chem. Intell. Lab. Syst.*, 38 (1997) 1-23.
- [27] D. Svozil and J.G.K. Sevcik, *J. Chem Inf. Comput. Sci.*, 37 (1997) 338-342.
- [28] F.R. Rapp, *Kalibrierung und Validierung von Sensorarrays unter Einsatz neuronaler Netze und genetischer Algorithmen*, Master Thesis, University of Tübingen, 1996.
- [29] N.M. Faber, X.-H. Song and P.K. Hopke, *Sample-specific standard error of prediction for partial least squares regression*, *Trac – Trend. Anal. Chem.*, in press.
- [30] Z.P. Chen, Y.Z. Liang, J.H. Jiang, Y. Li, J.Y. Qian and R.Q. Yu, *J. Chemometrics*, 13 (1999) 15-30.
- [31] H.T. Eastman and W.J. Krzanowski, *Technometrics*, 24 (1982) 73-77.
- [32] H. Martens and M. Martens, *Food Qual. Prefer.*, 11 (2000) 5-16.

- [33] H. Martens, M. Hoy, F. Westad, B. Folkenberg and M. Martens, *Chemom. Intell. Lab. Syst.*, 58 (2001) 151-170.
- [34] N.M. Faber, L.M.C. Buydens and G. Kateman, *Anal. Chim. Acta*, 296 (1994) 1-20.
- [35] B.K. Dable and K.S. Booksh, *J. Chemometrics*, 15 (2001) 591-613.
- [36] R. Kramer, *Chemometric Techniques for Quantitative Analysis*, Marcel Dekker Inc., New York, 1998.
- [37] H. Wold, *Nonlinear Iterative Partial Least Squares (NIPALS) Modeling - Some Current Developments*, in: P. R. Krishnajah (Ed.), *Multivariate Analysis*, Academic Press, New York 1973.
- [38] G. Baffi, E. Martin, J. Morris, *Chemom. Intell. Lab. Syst.*, 61 (2002) 151-165.
- [39] H. Martens and T. Naes, *Multivariate Calibration*, John Wiley & Sons Inc., New York, 1989.
- [40] H. Martens and H. Martens, *Multivariate Analysis of Quality - An Introduction*, John Wiley & Sons Inc., New York, 2000.
- [41] K. H. Esbensen, *Multivariate Data Analysis – In Practice*, 5th edn., Camo Process AS, Norway, 2001.
- [42] A.D. Shaw, A. di Camillo, G. Vlahov, A. Jones, G. Bianchi, J. Rowland and D.B. Kell, *Anal. Chim. Acta*, 348 (1997) 357-374.
- [43] W. Lindberg, J-A. Persson and S. Wold, *Anal. Chem.*, 55 (1983) 643-648.
- [44] D.M. Haaland and E.V. Thomas, *Anal. Chem.*, 60 (1988) 1193-1202.
- [45] P. Geladi and B.R. Kowalski, *Anal. Chim. Acta*, 185 (1986) 1-17.
- [46] J. Zupan and J. Gasteiger, *Anal. Chim. Acta*, 248 (1991) 1-30.
- [47] G. Kateman, *Chemom. Intell. Lab. Syst.*, 19 (1993) 135-142.
- [48] B.K. Lavine, *Anal. Chem.*, 70 (1998) 209R-228R.

- [49] B.K. Lavine, *Anal. Chem.*, 72 (2000) 91R-97R.
- [50] J. Zupan and J. Gasteiger, *Neural networks in chemistry and drug design*, 2nd edn., Wiley-VCH, Weinheim, 1999.
- [51] M. Bos, A. Bos and W.E. van der Linden, *Analyst*, 118 (1993) 323-328.
- [52] T. Blank and S.D. Brown, *Anal. Chem.*, 65 (1993) 3081-3089.
- [53] P. J. Gemperline and J.R. Long, *Anal. Chem.*, 63 (1991) 2313-2323.
- [54] C. Borggaard and H.H. Thodberg, *Anal. Chem.*, 64 (1992) 545-551.
- [55] F. Dieterle, D. Nopper and G. Gauglitz, *Fresen. J. Anal. Chem.* 370 (2001) 723-730.
- [56] D. Patterson, *Artificial neural networks, theory and applications*, Prentice Hall Inc., Upper Saddle River, 1996.
- [57] J. Principe, N. Euliano, W. Lefebvre, *Neural and adaptive systems: Fundamentals through simulations*, John Wiley & Sons Inc., New York, 2000.
- [58] S. Kaykin, *Neural networks a comprehensive foundation*, Prentice Hall Inc., Upper Saddle River, 1999.
- [59] A. Zell, *Simulation Neuronaler Netzwerke*, Addison-Wesley, Bonn, Paris, 1994.
- [60] S.W. Morre, J.W. Gardner, E.L. Hines, W. Göpel and U. Weimar, *Sens. Actuators B*, 15-16 (1993) 344-348.
- [61] D. E. Rumelhart and J. L. McClelland, *Parallel distributed Processing: Explorations in the Microstructure of Cognition*, Vol. 1, MIT Press, Cambridge, 1986.
- [62] D.E. Rumelhart, G.E. Hinton and R.J. Williams, *Learning internal representations by error propagation*, in: J.A. Anderson and E. Rosenfeld (eds.), *Neurocomputing: Foundations of Research*, MIT Press, Cambridge, 1988.
- [63] R.A. Jacobs, *Neural Networks*, 1 (1988) 295-307.
- [64] T. Tollenaere, *Neural Networks*, 3 (1990) 561-573.

- [65] M. Riedmiller and H. Braun, Rprop - A fast adaptive learning algorithm, Technical report, University of Karlsruhe, 1992.
- [66] M. Riedmiller and H. Braun, A direct adaptive method for faster backpropagation learning: The RPROP algorithm, in: ICNN-93, IEEE Int. Conf. on Neural Networks, San Francisco, 1993.
- [67] A. Zell, SNNS Stuttgart Neural Network Simulator
<http://www-ra.informatik.uni-tuebingen.de/SNNS/>, 2002.
- [68] E. Richards, C. Bessant and S. Saini, Chemom. Intell. Lab. Syst., 61 (2002) 35-49.
- [69] T. Udelhoven and B. Schutt, Chemom. Intell. Lab. Syst., 51 (2000) 9-22.
- [70] R. Fletcher, Practical Methods of Optimisation: Unconstrained Optimisation, Vol. 1, John Wiley & Sons Inc., New York, 1980.
- [71] M.F. Moller, Neural Networks, 6 (1993) 525-533.
- [72] D. Montana and L. Davis, Training feedforward neural networks using genetic algorithms, in: Proceedings of the 11th International Joint Conference on Artificial Intelligence, Morgan Kaufman, San Mateo, 1989.
- [73] A Homaifar, S. Guan, Training weights of neural networks by genetic algorithms, in: Proceedings of the 2nd IASTED, Anaheim, 1990.
- [74] W.S. Sarle, Proceedings of the 27th Symposium on the Interface of Computing Science and Statistics, 1995, 352-360.
- [75] K. Hornik, Neural Networks, 6 (1993) 251-257.
- [76] J.F. Kolen and J.B. Pollack, in: R.P. Lippmann, J.E. Moody and D.S. Touretzky, Advances in Neural Information Processing Systems, Vol. 3, Morgan Kaufmann, San Mateo, 1991.
- [77] A. Esposito, M. Marinaro, D. Oricchio and S. Scarpetta, Neural Networks, 13 (2000) 651-665.

- [78] A. Weigend, in: M.C. Mozer, P. Smolensky, D.S. Touretzky, J.L. Elman, and A.S. Weigend (eds.), *Proceedings of the 1993 Connectionist Models Summer School*, Lawrence Erlbaum Associates, Hillsdale, 1994.
- [79] M.L. Astion and P. Wilding, *Arch. Pathol. Lab. Med.*, 116 (1992) 995-1001.
- [80] C. Macleod and G.M. Maxwell, *Artif. Intell. Rev.*, 16 (2001) 201-224.
- [81] A.J. Miller, *Subset selection in regression*, Chapman & Hall, London, 1990.
- [82] M.B. Seasholtz and B. Kowalski, *Anal. Chim. Acta*, 277 (1993) 165-177.
- [83] D.J. Livingstone and D.T. Manallack, *J. Med. Chem.*, 36 (1993) 65-70.
- [84] L. Xu and W.-J. Zhang, *Anal. Chim. Acta*, 446 (2001) 477-483.
- [85] S. Chatterjee and B. Proce, *Regression Analysis by Example*, John Wiley & Sons Inc., New York, 1977.
- [86] E. Kaskavelis, E. Martin, P. Jonathan and J. Morris, *J. Chemometrics*, 15 (2001) 665-683.
- [87] M. Novic and M. Vracko, *Chemom. Intell. Lab. Syst.*, 59 (2001) 33-44.
- [88] D.L. Selwood, D.J. Livingstone, J.C. Comley, B.A. O'Dowd, A.T. Hudson, P. Jackson, K.S. Jandu, V.S. Rose and J.N. Stables, *J. Med. Chem.*, 33 (1990) 136-142.
- [89] D.L. Selwood, D.J. Livingstone, J.C. Comley, B.A. O'Dowd, A.T. Hudson, P. Jackson, K.S. Jandu, V.S. Rose, and J.N. Stables, *J. Med. Chem.*, 33 (1990) 136-142.
- [90] M.C.U. Araújo, T.C.B. Saldanha, R.K.H. Galvao, T. Yoneyma, H.C. Chame and V. Visani, *Chemom. Intell. Lab. Syst.*, 57 (2001) 65-73.
- [91] U. Hörchner and J.H. Kalivas, *Anal. Chim. Acta*, 311 (1995) 1-13.
- [92] C.B. Lucasius, M.L.M. Bekers and G. Kateman, *Anal. Chim. Acta*, 286 (1994) 135-153.
- [93] B.E. Mitchell and P.C. Jurs, *J. Chem. Inf. Comput. Sci.*, 37 (1997) 538-547.

- [94] H.E. McClelland and P.C. Jurs, *J. Chem. Inf. Comput. Sci.*, 40 (2000) 967-975.
- [95] M.D. Wessel, P.C. Jurs, J.W. Tolan and S.M. Muskal, *J. Chem. Inf. Comput. Sci.*, 38 (1998) 726-735.
- [96] B.E. Mattioni and P.C. Jurs, *J. Chem. Inf. Comput. Sci.*, 42 (2002) 232-240.
- [97] R. Leardi, *J. Chemometrics*, 15 (2001) 559-569.
- [98] D. Jouan-Rimbaud, D.L. Massart, R. Leardi and O.E. de Noord, *Anal. Chem.*, 67 (1995) 4295-4301.
- [99] R. Leardi and A.L. González, *Chemom. Intell. Lab. Syst.*, 41 (1998) 195-207.
- [100] K. Hasegawa, Y. Miyashita and K. Funatsu, *J. Chem. Inf. Comput. Sci.*, 37 (1997) 306-310.
- [101] S.S. So and M. Karplus, *J. Med. Chem.*, 39 (1996) 5246-5256.
- [102] H. Handels, T. Roß, J. Kreuzsch, H.H. Wolff and S.J. Pöppel, *Art. Intell. Med.*, 16 (1999), 283-297.
- [103] H. Yoshida, R. Leardi, K. Funatsu and K. Varmuza, *Anal. Chim. Acta*, 446 (2001) 485-494.
- [104] R. Cela, E. Martínez and A.M. Carro, *Chemom. Intell. Lab. Syst.*, 57 (2001) 75-92.
- [105] C. Borggaard, *Spectrosc. Eur.*, 6 (1994) 24-27.
- [106] B. Walczak, W. Wegscheider, *Anal. Chim. Acta*, 283 (1993) 508-517.
- [107] S. Ventura, M. Silva, D. Pérez-Bendito and C. Hervás, *J. Chem. Inf. Comput. Sci.*, 37 (1997) 287-291.
- [108] C. W. Brown and S. Lo, *Anal. Chem.*, 70 (1998) 2983-2990.
- [109] R.J. Poppi and D.L. Massart, *Anal. Chim. Acta*, 375 (1998) 187-195.

- [110] F. Gao, M. Li, F. Wang, B. Wang and P. Yue, *Ind. Eng. Chem. Res.*, 38 (1999) 4330-4336.
- [111] Neurogenetic Optimizer 2.6, Biocomp Systems Inc., Redmond, 1998.
- [112] J.H. Jiang, J.H. Wang, X.H. Song and R.Q. Yu, *J. Chemometr.*, 10 (1996) 253-267.
- [113] T. Ash, *Connection Sci.*, 1 (1989) 365-375.
- [114] A. Yasri and D. Hartsough, *J. Chem. Inf. Comput. Sci.*, 41 (2001) 1218-1227.
- [115] G. Chakraborty, *IEEE Conference on Neural Networks*, 2 (1995) 1116-1120.
- [116] J. Hertz, A. Krotgh and R.G. Palmer, *Introduction to the Theory of Neural Computation*, Addison-Wesley, Redwood City, 1991.
- [117] S.E. Fahlman and C. Lebiere, The cascade correlation architecture, in: D.S. Touretzky, *Advances in Neural Information Processing Systems II*, Morgan Kaufmann, San Mateo, 1990.
- [118] M. Mitchell, *An introduction to genetic algorithms*, MIT Press, Massachusetts, 1998.
- [119] G. F. Miller, P.M. Todd and S.U. Hegde, Designing neural networks using genetic algorithms, in: J.D. Shaffer, *Proceedings of the Third International Conference on Genetic Algorithms*, Morgan Kaufmann, San Mateo, 1989.
- [120] J.D. Schaffer, R.A. Caruana and L.J. Eshelman, *Pysica D*, 42 (1990) 244-248.
- [121] H. Kitano, *Complex Systems*, 4 (1990) 461-476.
- [122] R. Reed, *Neural Networks*, 4 (1993) 740-747
- [123] S. Haykin, *Neural networks a comprehensive foundation*, Prentice Hall Inc., Upper Saddle River, 1999.
- [124] F. Despagne and D.L. Massart, *Chemom. Intell. Lab. Syst.*, 40 (1998) 145-163.
- [125] V.V. Vinod and S. Ghose, *Neurocomputing*, 10 (1996) 55-69.

- [126] S.S. So and M. Karplus, *J. Med. Chem.*, 39 (1996) 1521-1530.
- [127] W. J. Krzanowski, *Principles of Multivariate Analysis – A User's Perspective*, Oxford University Press, Oxford, 1988.
- [128] R.J. Brook and G. C. Arnold, *Applied regression analysis and experimental design*, Marcel Dekker Inc., New York, 1985.
- [129] T. Kimura, K. Hasegawa and K. Funatsu, *J. Chem. Inf. Comput. Sci.*, 38 (1998) 276-282.
- [130] U. Depczynski, V.J. Frost and K. Molt, *Anal. Chim. Acta*, 420 (2000) 217-227.
- [131] R. Meusinger and R. Moros, *Chemom. Intell. Lab. Syst.*, 46 (1999) 67-78.
- [132] M. Clark and R.D. Cramer, *Quant. Struct. Act. Relat.*, 12 (1993) 9-12.
- [133] S. Kirkpatrick, C.D. Gelatt and M.P. Vecchi, *Science*, 220 (1982) 671-680.
- [134] M.C. Mozer and P. Smolensky, Skeletonization: A technique for trimming the fat from a network via relevance assessment, in: D.S. Touretzky (ed.), *Advances in Neural Information Processing Systems 1*, Morgan Kaufmann, San Mateo, 1989.
- [135] F. Despagne and D.L. Massart, *Chemom. Intell. Lab. Syst.*, 40 (1998) 145-163.
- [136] T. Schreiner, *Ausdünnungsverfahren für Neuronale Netze*, Master Thesis, University of Stuttgart, 1994.
- [137] B. Hassibi and D.G. Stork, Second order derivatives for network pruning, in: S.J. Hanson, J.D. Cowan and C.L. Giles (eds.), *Advances in Neural Information Processing Systems 5*, Morgan Kaufmann, San Mateo, 1993.
- [138] Y. LeCun, J.S. Denker and S.A. Solla, Optimal Brain Damage, in: D.S. Touretzky (ed.), *Advances in Neural Information Processing Systems 2*, Morgan Kaufmann, San Mateo, 1990.
- [139] W.J. Egan, S.M. Angel and S.L. Morgan, *J. Chemometrics*, 15 (2001) 29-48.
- [140] T.T. Bachmann and R.D. Schmid, *Anal. Chim. Acta*, 401 (1999), 95-103.

- [141] T.T. Bachmann, B. Leca, F. Vilatte, J.L.Marty, D. Fournier and R.D. Schmid, *Biosens. Bioelectron.*, 15 (2000) 193-201.
- [142] A. Krogh and J. Hertz, A simple weight decay can improve generalization, in: J.E. Moody, S.J. Hanson, and R.P. Lippmann (eds.), *Advances in Neural Processing System 4*, Morgan Kaufmann, 1992 .
- [143] R. Reed, *IEEE T. Neural Networ.*, 4 (1993), 740-747.
- [144] J. Sietsma and R. Dow, *Neural Networks*, 4 (1991), 67-79.
- [145] R.B. Boozarjomehry and W.Y. Svrcek, *Comput. Chem. Eng.*, 25 (2001) 1075-1088.
- [146] H. Braun and T. Ragg, ENZO, *Evolution of Neural Networks, User Manual and Implementation Guide*, Technical Report 21/96, University of Karlsruhe, 1996.
- [147] H. Braun and J. Weisbrod, *Evolving neural feedforward networks*, in R.F. Albrecht (ed.), *Proceedings of the International Conference on Artificial Neural Nets and Genetic Algorithms*, Springer, Wien, 1993.
- [148] <http://www-ra.informatik.uni-tuebingen.de/downloads/SNNS/ENZO.tar.gz>.
- [149] T. Ragg and S. Gutjahr, *Building High Performan Classifiers by Integrating Bayesian Learning, Mutual Information and Committee Techniques – A Case Study in Time Series Prediction*, in: W. Gerstner, A. Germond, M. Hasler, J.-D. Nicoud (eds.), *Artificial Neural Networks - ICANN97*, Springer-Verlag, Wien, 1997.
- [150] T. Ragg, W. Menzel, W. Baum, and M. Wigbers, *Predicting Sales Rates for Thousands of Retail Traders*, in: D. Tsaptsinos: *Proceedings of the International Conference on Engineering Applications of Neural Networks*, Kingston, England, 2000.
- [151] D. Hörnel and T. Ragg, *A Connectionist Model for the Evolution of Styles of Harmonization*, in: *Proceedings of the 1996 International Conference on Music Perception and Cognition*, Montreal, Canada 1996.
- [152] R.B. Boozarjomehry and W.Y. Svrcek, *Comput. Chem. Eng.*, 25 (2001) 1075-1088.

- [153] C. Hervàs, *J. Chem. Inf. Comput. Sci.*, 40 (2000) 724-731.
- [154] C.G. Zampronio, J.J.R. Rohwedder and R.J. Poppi, *Chemom. Intell. Lab. Syst.*, 62 (2002) 17-24.
- [155] Measured by Carmen Betsch, Institute of Physical and Theoretical Chemistry, University of Tübingen, 2000.
- [156] J. Huang and K.H. Esbensen, *Chemom. Intell. Lab. Syst.*, 57 (2001) 37-56.
- [157] G. Maier, *Angew. Chem.*, 110 (1998) 3128-3143.
- [158] G. Dlubek, A. Clarke, H. Fretwell, S.B. Dugdale and M.A. Alam, *Phys. Status Solidi A*, 157 (1996) 351-364.
- [159] R. Buchhold, A. Nakladal, G. Gerlach, M. Herold, G. Gauglitz, K. Sahre and K.J. Eichhorn, *Thin Solid Films*, 350 (1999) 178-185.
- [160] P.C. Painter and M. M. Coleman, *Fundamentals of Polymer Science – An Introductory Text*, 2nd edition, Technomic Publishing Company, Lancaster, 1997.
- [161] F. Dieterle, G. Belge and G. Gauglitz, *Anal. Bioanal. Chem.*, 374 (2002) 858-867.
- [162] J.W. Grate, A. Snow, D.S. Ballantine, H. Wohltjen, M.H. Abraham, R.A. McGill and P. Sasson, *Anal. Chem.*, 60 (1988) 869-875.
- [163] J.W. Grate, M. Klusty, R.A. McGill, M.H. Abraham, G. Whiting and J. Andonian-Haftvan, *Anal. Chem.*, 64 (1992) 610-624.
- [164] J.W. Grate, S.J. Patrash and M.H. Abraham, *Anal. Chem.*, 67 (1995) 2162-2169.
- [165] D. Leipert, F. Rathgeb, M. Herold, J. Mack, G. Gauglitz and G. Jung, *Anal. Chim. Acta*, 392 (1999) 213-221.
- [166] F. Rathgeb, D. Reichl, M. Herold, O. Mader, T. Mutschler and G. Gauglitz, *Fresenius J. Anal. Chem.*, 368 (2000) 192-195.
- [167] W.R. Vieth and J.M. Howell, *J. Membr. Sci.*, 1 (1976) 177-220.

- [168] H. Yokouchi, M. Matsuguchi, Y. Sadaoka and Y. Sakai, *Sens. Mater.*, 8 (1996) 69-78.
- [169] F. Rathgeb and G. Gauglitz, *Anal. Chim. Acta*, 372 (1998) 333-340.
- [170] G. Allen and J.C. Bevington, *Comprehensive Polymer Science*, Pergamon Press, Oxford, 1989.
- [171] B. Kieser, *Charakterisierung sensitiver Schichten zur Selektivitätssteigerung optischer Chemosensoren*, Ph.D. Thesis, University of Tübingen, 2002.
- [172] E. Kretschmann and H. Raether, *Z. Naturforsch.*, 23a (1968) 2135-2136.
- [173] B. Kieser, F. Dieterle and G. Gauglitz, *Anal. Chem.*, 74 (2002) 4781-4787.
- [174] B. Kieser, D. Pauluth and G. Gauglitz, *Anal. Chim. Acta*, 434 (2001) 231-237.
- [175] G. Gauglitz and W. Nahm, *Fresenius Z. Anal. Chem.*, 341 (1991) 279-283.
- [176] G. Kraus and G. Gauglitz, *Fresenius J. Anal. Chem.*, 344 (1992) 153-157.
- [177] F. Rathgeb, *Charakterisierung von Wechselwirkungsprozessen in sensitiven Schichten*, Ph.D. Thesis, University of Tübingen, 1999.
- [178] D. Reichl, R. Krage, C. Krummel and G. Gauglitz, *Appl. Spectrosc.*, 54 (2000) 583-586.
- [179] BRD Deutsches Patent- und Markenamt, *Offenlegungsschrift DE 198 30 727 A1*, 1999.
- [180] M.J. Molina and F.S. Rowland, *Nature*, 249 (1974) 810-812.
- [181] M.J. Molina and F.S. Rowland, *J. Phys. Chem.*, 79 (1975) 667-669.
- [182] F.S. Rowland and M.J. Molina, *J. Phys. Chem.*, 80 (1976) 2049-2056.
- [183] United States Environmental Protection Agency, Office of Air and Radiation, <http://www.epa.gov/oar/caa/contents.html>, 2002.
- [184] Umweltbundesamt, *Bundesanzeiger*, 245 (1995) 12994.

- [185] United States Environmental Protection Agency, Ozone Protection Regulations: Substitutes for Ozone-Depleting Substances, <http://www.epa.gov/ozone/title6/snap/snap.html>, 2000.
- [186] Umweltbundesamt, Bundesanzeiger, 91 (1999) 7939.
- [187] Alternative Fluorocarbons Acceptability Study, <http://www.afeas.org>, 2002.
- [188] S. Busche, F. Dieterle, B. Kieser, and G. Gauglitz, *Sensors Act. B*, 89 (2003) 192-198.
- [189] Measured by Stefan Busche, Institute of Physical and Theoretical Chemistry, University of Tübingen, 2002.
- [190] B. Kieser, F. Dieterle and G. Gauglitz, *Anal. Chem.*, 74 (2002) 4781-4787.
- [191] Measured by Dr. Birgit Kieser, Institute of Physical and Theoretical Chemistry, University of Tübingen, 2001.
- [192] I.V. Tetko and A.E.P. Villa, *Neural Networks*, 10 (1997) 1361-1374.
- [193] Measured by Dr. Birgit Kieser, Institute of Physical and Theoretical Chemistry, University of Tübingen, 2001.
- [194] F. Dieterle, B. Kieser and G. Gauglitz, *Chemom. Intell. Lab. Syst.*, 65 (2003) 67-81.
- [195] Measured by Georg Belge (RIFS) and Dr. Birgit Kieser (SPR), Institute of Physical and Theoretical Chemistry, University of Tübingen, 2001.
- [196] Measured by Maura Kasper (RIFS) and Stefan Busche (SPR), Institute of Physical and Theoretical Chemistry, University of Tübingen, 2002.
- [197] J. Park, W.A. Groves and E.T. Zellers, *Anal. Chem.*, 71 (1999) 3877-3886.
- [198] P.C. Jurs, G.A. Bakken and H.E. McClelland, *Chem. Rev.*, 100 (2000) 2649-2678.
- [199] K.J. Albert, N.S. Lewis, C.L. Schauer, G.A. Sotzing, S.E. Stitzel, T.P. Vaid and D.R. Walt, *Chem. Rev.*, 100 (2000) 2595-2626.
- [200] W.B. Carey, *Trends Anal. Chem.*, 13 (1994) 210-218.

- [201] J. Seemann, F.R. Rapp, A. Zell and G. Gauglitz, *Fresenius J. Anal. Chem.*, 359 (1997) 100 – 106.
- [202] A. Hierlemann, U. Weimar, G. Kraus, G. Gauglitz and W. Göpel, *Sensor Mater.*, 7 (1995) 179 – 189.
- [203] D.R. Walt, T. Dickinson, J. White, J. Kauer, S. Johnson, H. Engelhardt, J. Sutter and P. Jurs, *Biosens. Bioelectron.*, 13 (1998) 697-699.
- [204] H.M. Yan, G. Kraus and G. Gauglitz, *Anal. Chim. Acta.*, 312 (1995) 1-8.
- [205] H.P. Beck and C. Wiegand, *Fresenius J. Anal. Chem.*, 351 (1995) 701-707.
- [206] J. White, J.S. Krauer, T.A. Dickinson and D.R. Walt, *Anal. Chem.*, 68 (1996) 2191-2202.
- [207] M. Slama, C. Zaborosch, D. Wienke and F. Spener, *Anal. Chem.*, 68 (1996) 3845-3850.
- [208] O. Hernández, A.I. Jiménez, F. Jiménez and J.J. Arias, *Anal. Chim. Acta*, 310 (1995) 53-61.
- [209] B.W. Saunders, D.V. Thiel and A. Mackay-Sim, *Analyst*, 120 (1995) 1013-1018.
- [210] S.R. Johnson, J.M. Sutter, H.L. Engelhardt, P.C. Jurs, J. White, J.S. Kauer, T.A. Dickinson and D.R. Walt, *Anal. Chem.*, 69 (1997) 4641-4648.
- [211] J.M. Sutter and P.C. Jurs, *Anal. Chem.*, 69 (1997) 856-862.
- [212] C. Hervás and S. Ventura, *J. Chem. Inf. Coput. Sci.*, 38 (1998) 1119-1124.
- [213] K. Kato, Y. Kato, K. Takamatsu, T. Udaka, T. Nakahara, Y. Matsuura and K. Yoshikawa, *Sens. Actuators B.*, 71 (2000) 192-196.
- [214] V. Plegge, M. Slama, B. Süsselbeck, D. Wienke, F. Spener, M. Knoll and M. Zaborosch, *Anal. Chem.*, 72 (2000) 2937-2942.
- [215] I. Sugimoto, M. Nakamura and H. Kuwano, *Sens. Actuators B*, 10 (1993) 117-122.

- [216] R.P. Podgorsek, H. Franke, J. Woods and S. Morrill, *Sens. Actuators B*, 51 (1998) 146-151.
- [217] C. Delpha, M. Lumbreras and M. Siadat, *Sens. Actuators B*, 80 (2001) 59-67.
- [218] G. Dlubek, A. Clarke, H. Fretwell, S.B. Dugdale and M.A. Alam, *Phys. Status Solidi A*, 157 (1996) 351-364.
- [219] D. Schrader and Y. Jean, *Positron and Positronium Chemistry*, Elsevier, Amsterdam, 1988.
- [220] C.S. Coughlin, K.A. Mauritz and R.F. Storey, *Macromolecules*, 24 (1992) 1526-1534.
- [221] L. Palacio, P. Prádanos, J.I. Calvo and A. Hernández, *Thin Solid Films*, 348 (1999) 23-29.
- [222] T. Mutschler, B. Kieser, R. Frank and G. Gauglitz, *Anal. Bioanal. Chem.*, 374 (2002) 658-664.
- [223] D.W. Breck, *Zeolithe Molecular Sieves*, John Wiley & Sons, New York, 1974.
- [224] A.M. Halpern, E.D. Glendening, *J. Mol. Struct. Theochem*, 365 (1996) 9-12.
- [225] F. Rathgeb, *Charakterisierung von Wechselwirkungsprozessen in sensitiven Schichten*, Ph. D. Thesis, University of Tübingen, 1999.
- [226] F. Rathgeb and G. Gauglitz, *Anal. Chim. Acta*, 372 (1998) 333-340.
- [227] S. Busche, *Optimierung von sensitiven Schichten zur Charakterisierung der Freone R22 und R134a durch Oberflächenplasmonenresonanz*, Master Thesis, University of Tübingen, 2002.
- [228] G. Baffi, E.B. Martin and A.J. Morris, *Chemom. Intell. Lab. Syst.*, 52 (2000) 5-22.
- [229] T. Naes, T. Issaksson and B.R. Kowalski, *Anal. Chem.*, 62 (1990) 664-673.
- [230] M.K. Phelan, C.H. Barlow, J.J. Kelly, T.M. Jinguji and J.B. Callis, *Anal. Chem.*, 61 (1989) 1419-1424.

- [231] V. Centner, O.E. de Noord and D.L. Massart, *Anal. Chim. Acta*, 376 (1998) 153-168.
- [232] J.N. Miller and J.C. Miller, *Statistics and Chemometrics in Analytical Chemistry*, 4th edn., Pearson Education Limited, Edinburgh Gate, 2000.
- [233] F.S. Swed and C. Eisehart, *Ann. Math. Statist.*, 14 (1943) 66.
- [234] S. Gourvéneec, D.L. Massart and D.N. Rutledge, *Chemom. Intell. Lab. Syst.*, 61 (2002) 51-61.
- [235] *S-Plus 6 for Windows, Guide to Statistics, Vol. 1*, Insightful Corporation, Seattle, 2001.
- [236] N.R. Drapper and H. Smith, *Applied Regression Analysis*, 2nd edn., Wiley, New York, 1981.
- [237] A. Berglund and S. Wold, *J. Chemometrics*, 11 (1997) 141-156.
- [238] A. Berglund and S. Wold, *J. Chemometrics*, 13 (1999) 461-471.
- [239] B.G.M. Vandeginiste, D.L. Massart, L.M.C. Buydens, S. De Jong, P.J. Lewi and J. Smeyers-Verbeke, *Handbook of Chemometrics and Qualimetrics: Part B*, Elsevier, Amsterdam, 1998.
- [240] S. Wold, N. Kettaneh-Wold and B. Skagerberg, *Chemom. Intell. Lab. Syst.*, 7 (1989) 53-65.
- [241] K. Hasegawa, T. Kimura and K. Funatsu, *Quant. Struct. Act. Relat.*, 16 (1997) 219-223.
- [242] T. Li, H. Mei and P. Cong, *Chemom. Intell. Lab. Syst.*, 45 (1999) 177-184.
- [243] H. Yoshida and K. Funatsu, *J. Chem. Inf. Comput. Sci.*, 37 (1997) 1115-1121.
- [244] T. Kimura, Y. Miyashita, K. Funatsu and S. Sasaki, *J. Chem. Inf. Comput. Sci.*, 36 (1996) 185-189.
- [245] G. Baffi, E.B. Martin, A.J. Morris, *Comput. Chem. Eng.*, 23 (1999) 395-411.

- [246] Q. Ding, G.W. Small and M.A. Arnold, *Appl. Spectrosc.*, 53 (1999) 402-414.
- [247] F.L. Dickert, O. Hayden and M.E. Zenkel, *Anal. Chem.*, 71 (1999) 1338-1341.
- [248] L. Breiman, J.H. Friedman, R.A. Olshen and C.J. Stone, *Classification and Regression Trees*, Chapman & Hall, New York, 1984.
- [249] D. Steinberg and P. Colla, *CART-Classification and Regression Trees*, Salford Systems, San Diego, 1997.
- [250] A. Berglund, N. Kettaneh, L.-L. Uppgard, S. Wold, N. Bendwell and D.R. Cameron, *J. Chemometrics*, 15 (2001) 321-336.
- [251] J.R. Quinlan, *Mach. Learn.*, 1 (1986) 81-106.
- [252] J.H. Friedman, *Ann. Stat.*, 19 (1991) 1-141.
- [253] Mars 2.0, Salford Systems, San Diego, 2000.
- [254] M.J. Arcos, M.C. Ortiz, B. Villahoz and L. A. Sarabia, *Anal. Chim. Acta*, 339 (1997) 63-77.
- [255] D. Jouan-Rimbaud, D.L. Massart and O.E. de Noord, *Chemom. Intell. Lab. Syst.*, 35 (1996) 213-220.
- [256] R. Leardi, *J. Chemom.*, 8 (1994) 65-79.
- [257] J.R.M. Smits, W.J. Melssen, L.M.C. Buydens and G. Kateman, *Chemom. Intell. Lab. Syst.*, 22 (1992), 165-189.
- [258] D. Svozil, V. Kvasnicka and J. Pospichal, *Chemom. Intell. Lab. Syst.*, 39 (1997) 43-62.
- [259] K. Hornik, M. Stinchcombe and H. White, *Neural Networks*, 4 (1989) 359-366.
- [260] K. Hehl, *Untersuchung dünner polymerer Filme und deren Wechselwirkung mit Analyten mit spektraler Ellipsometrie*, Ph.D. Thesis, University of Tübingen, 1998.

- [261] D. Reichl, Aufbau, Charakterisierung und Optimierung eines optischen Sensorsystems zur reflektometrischen Interferenzspektroskopie mit mehrfarbigen Leuchtdioden, Ph. D. thesis, University of Tübingen, 2000.
- [262] A.H.C. van Kampen and L.M.C. Buydens, *Computers Chem.*, 21 (1997) 153-160.
- [263] W.H. Kruskal and W.A. Wallis, *J. Am. Stat. Assoc.*, 47 (1952) 583-621.
- [264] A. Herrero and M.C. Ortiz, *Anal. Chim. Acta*, 378 (1999) 245-259.
- [265] M. Hollander and D.A. Wolfe, *Nonparametric Statistical Methods*, Wiley, Chichester, 1973.
- [266] L. Norgaard, A. Saudland, J. Wagner, J.P. Nielsen, L. Munck and S.B. Engelsen, *Appl. Spectrosc.*, 54 (2000) 413-419.
- [267] F. Dieterle, S. Busche and G. Gauglitz, *Anal. Chim. Acta*, in press.
- [268] F. Dieterle, S. Müller-Hagedorn, H.M. Liebich and G. Gauglitz, in preparation.
- [269] R. Buchhold, *Bimorphe Gassensoren*, Dresden University Press, Dresden, 1999.
- [270] J. Piehler, Thermodynamische und kinetische Charakterisierung biomolekularer Erkennung mit direkten optischen Transducern, Ph. D. Thesis, University of Tübingen, 1997.
- [271] S. Reder, F. Dieterle, H. Jansen, S. Alcock and G. Gauglitz, Multi-analyte assay for triazines using cross-reactive antibodies and neural networks, *Biosens. Bioelectron.*, submitted.
- [272] C.-N. Ho, G.D. Christian and E.R. Davidson, *Anal. Chem.*, 50 (1978) 1108-1113.
- [273] E. Sánchez and B.R. Kowalski, *Anal. Chem.*, 58 (1986) 496-499.
- [274] K. Kato, Y. Kato, K. Takamatsu, T. Udaka, T. Nakahara, Y. Matsuura and K. Yoshikawa, *Sens. Actuators B*, 71 (2000) 192-196.
- [275] P. Mielle and F. Marquis, *Sens. Actuators B*, 76 (2001) 470-476.

13. Publications

Parts of this work and transfers of the methods introduced in this work to other areas of analytical chemistry have been published in literature:

13.1. Publications

- F. Dieterle, D. Noper and G. Gauglitz, Quantification of butanol and ethanol in aqueous phases by reflectometric interference spectroscopy – different approaches to multivariate calibration, *Fresenius J. Anal. Chem.*, 370 (2001): 723-730.
- B. Kieser, F. Dieterle and G. Gauglitz: Discrimination of methanol and ethanol vapors by the use of a single optical sensor with a microporous sensitive layer, *Anal. Chem.*, 74 (2002): 4781-4787.
- F. Dieterle, G. Belge, C. Betsch and G. Gauglitz: Quantification of the refrigerants R22 and R134a in mixtures by means of different polymers and reflectometric interference spectroscopy, *Anal. Bioanal. Chem.*, 371 (2002): 858-867.
- G. Belge, F. Dieterle and G. Gauglitz: Parallel detection of VOCs in gas phases based on a compact sensor system, *Proceedings of the OPTO 2002, Erfurt, 2002.*
- G. Belge, F. Dieterle, C. Betsch and G. Gauglitz: Kompakte optische Sensorsysteme zur Detektion der Freone R22 und R134a – Ein entscheidender Beitrag zum Kühlmittelrecycling, in J. D. Basel, G. Gerlach: *Sensoren im Fokus neuer Anwendungen*, w.e.b., Dresden, 2002, ISBN 3-935712-5.
- F. Dieterle, B. Kieser and G. Gauglitz: Genetic algorithms and neural networks for the quantitative analysis of ternary mixtures using a single sensor set-up, *Chemometr. Intell. Lab. Syst.*, 65 (2003): 67-81.
- S. Busche, F. Dieterle, B. Kieser and G. Gauglitz: Quantification of binary mixtures of the freones R22 and R134a by surface plasmon resonance, *Sens. Actuators B*, 89 (2003) 192-198.
- F. Dieterle, S. Busche and G. Gauglitz: Growing neural networks for a multivariate calibration and variable selection of time-resolved measurements, *Anal. Chim. Acta*, in press.

- F. Dieterle, S. Müller-Hagedorn, H.M. Liebich and G. Gauglitz: Urinary nucleosides as potential tumor markers evaluated by learning vector quantization, *Artif. Intell. Med.*, in press.
- G. Gauglitz, S. Reder, F. Dieterle and I. Coille, Indicator assays and biosensors of estrogens, in: *EDCs in waste water, surface water, and sludge*, Springer, in press.
- S. Reder, F. Dieterle, H. Jansen, S. Alcock and G. Gauglitz: Multi-analyte assays for triazines using cross-reactive antibodies and neural networks, *Biosens. Bioelectron.*, in press.
- F. Dieterle, S. Reder and G. Gauglitz: A biosensor for estrone, estradiol and ethinylestradiol based on cross-reactive antibodies, *Anal. Chem.*, submitted.
- M. Kasper, S. Busche, F. Dieterle and G. Gauglitz: Comparison of RfS and SPR for the quantification of quaternary mixtures of the low alcohols using time-resolved measurements, in preparation.
- M. Vollprecht, F. Dieterle and G. Gauglitz: Quantification of quaternary mixtures of the low alcohols in water – time-resolved measurements for the reduction of the number of sensors, in preparation.
- F. Dieterle, M. Kasper and G. Gauglitz: Quantification of the refrigerants R22, R32 and R143a in mixtures using RfS and time-resolved measurements, in preparation.
- S. Reder, F. Dieterle and G. Gauglitz: Multivariate data analysis performed by neural networks in the field of immunoassays, in preparation.

13.2. Posters

- S. Reder, F. Dieterle, C. Barzen, I. Coille, R. Abuknesha and G. Gauglitz: Immunoassay for estrogens in the environment: Multi-analyte detection based on fluorescence, Fourth Workshop on Biosensors and Biological Techniques in Environmental Analysis, Maò, Menorca, 1999.
- F. Dieterle, S. Reder and G. Gauglitz: Immunoassay for estrogens in the environment based on fluorescence and neural networks, Poster, Fourth International Conference on Environmetrics and Chemometrics, Las Vegas, 2000.

- S. Reder, F. Dieterle, C. Barzen, I. Coille, R. Abuknesha and G. Gauglitz: Immunosensor zur Multi-Analyt-Detektion von endokrin wirksamen Substanzen im Abwasser, 2. Biosensor Symposium 2001, Tübingen, 2001.
- G. Belge, F. Dieterle and G. Gauglitz: Diskriminierung von Gasen mit dünnen mikroporösen Polymerfilmen zur Steigerung der Selektivität von Sensorarrays, Anakon 2001, Konstanz, 2001.
- S. Busche, F. Dieterle and G. Gauglitz: Quantification of binary mixtures of the freones R22 and R134a by surface plasmon resonance, Eurotrode VI, Manchester 2002.
- F. Dieterle, C. Betsch, B. Kieser, G. Belge, S. Busche, M. Harbeck and G. Gauglitz: Simultaneous quantification of mixtures of volatile organic compounds by dynamic single sensor set-ups, IMCS 2002, Boston, 2002.
- G. Belge, F. Dieterle and G. Gauglitz: Vergleich zweier Sensorsysteme in der optischen Sensorik zur Quantifizierung ternärer Gemische mit chemometrischen Methoden, Anakon 2003, Konstanz, 2003.
- S. Busche, M. Kasper, F. Dieterle and G. Gauglitz: Quantifizierung von homologen Alkoholgemischen in der Gasphase durch einen polymerbasierten Oberflächenplasmonensensor, Anakon 2003, Konstanz, 2003.
- M. Vollprecht, F. Dieterle and G. Gauglitz: Zeitaufgelöste Messungen zur Quantifizierung von Multianalytgemischen in Wasser – Ein neuer Weg zur Reduktion der Sensoranzahl, Anakon 2003, Konstanz, 2003.
- M. Vollprecht, F. Dieterle and G. Gauglitz: Time-resolved swelling/deswelling behaviour of thin polymer films - a new approach in sensor application, 2nd International Symposium on Reactive Polymers in Inhomogeneous Systems, in Melts and at Interfaces, Dresden, 2003.

14. Acknowledgments

First of all, it is a pleasure to thank Prof. Dr. Günter Gauglitz for giving me the opportunity to do my Ph.D. at the Institute of Physical and Theoretical Chemistry (IPTC) at the University of Tübingen. The excellent working conditions, the topic of my research and the opportunities to participate in several international conferences and meetings were all made available by Prof. Gauglitz forming the basis of this work.

I wish to thank PD Dr. Udo Weimar from the University of Tübingen for his willingness to serve as second reviewer for this thesis.

I would like to express my thanks to Prof. Dr. Johann Gasteiger from the University of Erlangen for reviewing this thesis as third referee.

This work would not exist without many colleagues working with the different devices and generating all data analyzed in this thesis. I would like to thank my colleagues Dr. Birgit Kieser and Stefan Busche for the measurements by the SPR setup. I want to express my thanks to my colleagues Carmen Betsch, Maura Kasper, Georg Belge, Dirk Nopper and Matthias Vollprecht for the measurements using different RIFS setups. I want to say thank you to Dr. Sabine Reder who performed uncountable measurements using a TIRF setup. I would like to thank my coworkers Sevinal Altundal, Julia Fischer, Birgit Gaiser, Britt Schuster, Markus Hofmann and Michael Zatloukal for their tireless help and support.

I am deeply indebted to all my colleagues of the Gauglitz group for their help, the lively discussions and the great ambience. It was a pleasure to work in the "Verfügungsgebäude" thanks to Dr. Udo Weimar and his group who established an excellent working atmosphere.

I would like to express my thanks to Fred Reiner Rapp who made his powerful yet easy to use programs "Nemo" and "Solves" available. He never got tired of implementing all more or less useful features of my wish lists. I would like to thank Jürgen Mitschele for proof reading the spelling of this work.

The EC and the BMBF are acknowledged for the financial support (projects "AWACSS" and "VERN").

Last but not least, I would like to thank my family, in particular my parents, who enabled my studies and who gave me the support necessary during the last years.

15. Appendix

The following software was used for the analyses and implementations performed in this work:

- **Borland Delphi 6.0 Enterprise**, Borland Software Corporation, Scotts Valley CA, USA.
- **CART 4.0**, Salford Systems, San Diego CA, USA.
- **Enzo 1.0**, Heinrich Braun, Institute for Logic, Complexity and Deduction Systems, University of Karlsruhe, Karlsruhe, Germany.
- **Excel XP**, Microsoft Corporation, Redmond WA, USA.
- **Genetic Server 1.1**, Neurodimension Inc., Gainesville FL, USA.
- **JNNS 1.0**, Prof. Dr. Andreas Zell, Wilhelm-Schickard Institute, University of Tübingen, Tübingen, Germany.
- **MARS 2.0**, Salford Systems, San Diego CA, USA.
- **Mathematica 4.2**, Wolfram Research Inc., Champaign IL, USA.
- **Nemo 1.23**, Prof. Dr. Andreas Zell, Wilhelm-Schickard Institute, University of Tübingen, Tübingen, Germany.
- **Neuroolutions Custom Developer 4.1**, Neurodimension Inc., Gainesville FL, USA.
- **SNNS 4.2**, Prof. Dr. Andreas Zell, Wilhelm-Schickard Institute, University of Tübingen, Tübingen, Germany.
- **Solves 2002/08**, Prof. Dr. Andreas Zell, Wilhelm-Schickard Institute, University of Tübingen, Tübingen, Germany.
- **S-Plus 6.0 Pro**, Insightful Corporation, Seattle WA, Germany.
- **Systat 10.0**, Systat Software Inc., Richmond CA, USA.
- **Tablecurve 4.01**, Systat Software Inc., Richmond CA, USA.
- **The Unscrambler 7.6**, Camo ASA, Oslo, Norway.
- **Weka 3.36**, Ian H Witten and Eibe Frank, University of Waikato, New Zealand.

Meine akademischen Lehrer waren:

K. Albert, E. Bayer, D. Christen, H. Eckstein, G. Gauglitz, F. Gönenwein, W. Göpel, G. Häfelinger, H. P. Hagenmaier, M. Hanak, V. Hoffmann, W. Jäger, G. Jung, S. Kemmler-Sack, B. Koppenhöfer, K.-A. Kovar, D. Krug, N. Kuhn, E. Lindner, M. E. Maier, H. A. Mayer, H.-J. Meyer, U. Nagel, H. Oberhammer, D. Oelkrug, H. Pauschmann, H. Pommer, G. Reinhardt, V. Schurig, E. Schweda, F. F. Seelig, B. Speiser, H. Stegmann, J. Strähle, W. Voelter, K.-P. Zeller, C. Ziegler

Lebenslauf

Name	Frank Jochen Dieterle
Geburtsdatum	8. Oktober 1973
Geburtsort	Tübingen
Nationalität	deutsch
Familienstand	ledig
Ausbildung:	
1981-1984	Grundschule in Tübingen
1984-1993	Geschwister-Scholl-Schule in Tübingen
11. Mai 1993	Abitur
1993-2000	Chemiestudium an der Eberhard-Karls-Universität Tübingen
12. Oktober 1995	Diplomvorprüfungen
1999-2000	Diplomarbeit "Multivariate Analysen zur Mehrkomponentenbestimmung" in der Arbeitsgruppe Prof. Dr. G. Gauglitz
Seit Januar 2000	Wissenschaftlicher Angestellter am Institut für Physikalische und Theoretische Chemie der Universität Tübingen
Februar 2000 - Juli 2003	Promotion "Multianalyte quantifications by means of an integration of artificial neural networks, genetic algorithms and chemometrics for time-resolved analytical data" bei Prof. Dr. G. Gauglitz