

# 1 Einleitung

In den vergangenen Jahren hat die Nah-Infrarot (NIR)–Spektroskopie im industriellen Umfeld zunehmend an Bedeutung gewonnen. Vor allem Anwender aus der chemischen, der pharmazeutischen und der Lebensmittelindustrie setzen die NIR–Technologie als schnelle Analysentechnik ein. Die qualitative und quantitative Analyse von Stoffen und Stoffgemischen — z.B. von Rohstoffen beim Wareneingang und von fertigen Produkten beim Warenausgang — stellen in diesem Umfeld die wesentlichen Anwendungsgebiete dar. Hier wird die NIR–Spektroskopie vor allem zur Überwachung vorgegebener Spezifikationen eingesetzt.

Noch betriebnäher ist die Online–Prozeßkontrolle, die quantitative Aussagen liefern muß, um einen chemischen Prozeß zu steuern und adäquat auf Änderungen und Störungen reagieren zu können. Auch hier wird neben der Messung klassischer Parameter wie Druck, Temperatur, pH–Wert etc. zunehmend die NIR–Spektrometrie eingesetzt.

Die Vorzüge der NIR–Spektroskopie liegen in den extrem kurzen Meßzeiten, der guten Automatisierbarkeit und der einfachen Handhabung. Die NIR–Technik ersetzt zunehmend zeit- und materialintensive Routine–Analytik und trägt so zu einer Verbesserung der Produktivität in Unternehmen bei.

Je nach Art und Konsistenz der Proben kann zur Aufnahme von NIR–Spektren auf verschiedene Meßanordnungen zurückgegriffen werden. Sie basieren auf den Prinzipien der Transmission, Transflexion und der diffusen Reflexion. Küvetten aus robusten Materialien mit gut handhabbaren Schichtdicken führen neben einer ganzen Palette von Sonden und Remote–Sensoren mit Lichtleiterkopplung zu einer unkomplizierten Anwendung.

Eine aufwendige Präparation der Proben ist in der Regel nicht notwendig. Die erhaltenen Spektren müssen nicht an wenigen und auf Anhieb ersichtlichen diskreten Wellenlängen die zu erfassende Eigenschaft einer Probe wiedergeben. Es genügt, wenn das Spektrum durch letztere nur geringfügig (jedoch reproduzierbar!) in seiner Kontur beeinflußt wird. Die erforderliche Selektivität wird dann durch die Berechnung eines geeigneten Kalibrationsmodells erreicht. Dadurch ist es sogar möglich, verpackte Proben direkt durch den Boden von Glasgefäßen oder dünne Folien hindurch zu analysieren. So kommen die Proben nicht mit den Meßeinrichtungen in Kontakt, und es besteht keine Kontaminationsgefahr — ein Aspekt, der insbeson-

dere in der pharmazeutischen Industrie eine wesentliche Rolle spielt.

Bei der Erstellung einer NIR–Applikation ist die Kalibration, d.h. die Aufstellung eines mathematischen Zusammenhangs (*Kalibrationsmodell*) zwischen den Spektren von Standards und der zu kalibrierenden Eigenschaft, der aufwendigste Teil. Hierzu sind Standards in ausreichender Zahl erforderlich, deren Zusammensetzung im Hinblick auf die zu kalibrierenden Komponenten mit einem geeigneten Referenzverfahren bestimmt wurden. Diese Standards müssen so ausgewählt werden, daß die später bei den Analysenproben zu erwartende Varianz der Eigenschaften im Kalibrationsmodell abgedeckt ist. Nach Aufnahme der NIR–Spektren der Standards kann die Modellierung der erhaltenen Daten mit Hilfe chemometrischer Methoden beginnen.

Unter Chemometrie wird allgemein diejenige chemische Disziplin verstanden, die neben anderen mathematischen Methoden vor allem die Statistik nutzt, um

1. optimale Meßprozeduren und Experimente zu entwerfen und
2. bei der Analyse chemischer Daten das Maximum an relevanter chemischer Information zu extrahieren [1].

Da die Chemometrie eine weitgefächerte Disziplin mit vielen unterschiedlichen Einsatzmöglichkeiten ist, wird in dieser Arbeit für die chemometrische Analyse spektraler Daten der Begriff *Spektrochemometrie* verwendet.

Zur Analyse spektraler Daten werden verschiedene multivariate Verfahren eingesetzt, wobei die faktoranalytischen Verfahren *principal component regression* (PCR) und *partial least squares* (PLS) dominieren [2]. Bei diesen handelt es sich um Vollspektrenmethoden, d.h. es werden im allgemeinen die an allen Wellenlängen des Spektrums gemessenen Absorptionswerte berücksichtigt. Die Selektivität wird durch die nachträgliche Auswahl geeigneter Faktoren erreicht.

Die sich im Zusammenhang mit der NIR–Spektrochemometrie ergebenden Problemstellungen der Principal Component Regression bilden den thematischen Schwerpunkt der vorliegenden Arbeit. Die PCR stellt ein leistungsfähiges Verfahren zur quantitativen Analyse spektraler Daten dar und bietet gleichzeitig die Möglichkeit zur Datenreduktion. Durch das faktoranalytische Verfahren werden die spektralen Daten in eine orthogonale Basis von Eigenvektoren und Eigenspektren, die sogenannten latenten Variablen oder (Faktoren), zerlegt. Durch gezielte Auswahl der Faktoren können diejenigen spektralen Merkmale (Features) erfaßt werden, die am besten mit der zu kalibrierenden Probeneigenschaft korrelieren. Mit diesem Selektionsschritt geht gleichzeitig eine deutliche Reduktion der in die Kalibration eingehenden Datenmenge einher.

Im Rahmen dieser Arbeit wird die Auswahl geeigneter Faktorkombinationen für NIR–spektroskopische Kalibrationen auf Basis stochastischer Verfahren durch-

geführt. Stochastische Optimierungsverfahren sind zufallsgesteuerte Methoden zur systematischen Optimierung eines Systems unter Berücksichtigung eines definierten Optimierungskriteriums. Sie benötigen bis auf die Notwendigkeit der Existenz einer Lösung keine Vorannahmen über das zu optimierende System und sind daher äußerst flexibel und universell einsetzbar. Im Bereich der kombinatorischen Optimierungsprobleme haben sich *Genetische Algorithmen* (GAs) in einem breiten Anwendungsfeld durchgesetzt und stellen eine gute Alternative zu klassischen deterministischen Verfahren dar.

### Was sind Genetische Algorithmen?

Genetische Algorithmen sind Such-Algorithmen, die auf Mechanismen der natürlichen Evolution und Selektion und der biologischen Genetik basieren. Sie kombinieren das darwinistische Prinzip der Selektion durch *survival of the fittest* mit dem strukturierten, aber durch Zufälle gesteuerten, Austausch von Information. Dieses Prinzip wurde in dieser Arbeit auf Kalibrationsmodelle angewandt, für die es galt, eine bestimmte optimale Kombination von Faktoren zu finden. Bei der Anwendung Genetischer Algorithmen werden die zu optimierenden Parameter in einer dem Schemata-Theorem von *J. Holland* entsprechenden Weise kodiert [3]. Für die in dieser Arbeit untersuchten kombinatorischen Problemstellungen ergibt sich diese Kodierung durch eine Folge von Nullen und Einsen. Durch diese *Bit-Strings* (*Chromosomen*) werden einzelne Kombinationen der zu optimierenden Parameter repräsentiert. Jedes Chromosom stellt dabei eine potentielle Lösung des Optimierungsproblems dar, deren Güte durch ein festzulegendes Kriterium (*Fitnessfunktion*) bewertet werden muß. Eine Menge (*Population*) von Chromosomen bildet eine *Generation*.

Genetische Algorithmen beginnen ihren Optimierungsprozeß nicht an einem einzelnen Punkt, sondern aus einer Anfangspopulation heraus. Durch Prozesse der *Selektion*, *Rekombination* und *Mutation* entwickeln sich die Populationen von Generation zu Generation weiter. Dabei benötigt der Genetische Algorithmus keine Information über das zu optimierende System oder dessen Randbedingungen. Er optimiert vielmehr ein festzulegendes Kriterium, das die Güte der Chromosomen definiert und als *Fitnessfunktion* bezeichnet wird. Genetische Algorithmen sind also adaptive Such-Algorithmen, die unter Berücksichtigung der Fitness Chromosomen verändern und selektieren und dadurch Parameterkombinationen optimieren. Diese hohe Adaptivität ist es auch, welche die Genetischen Algorithmen zu den Systemen der *computational intelligence* (CI) zählen lassen.

Im weiteren wird in Kapitel 2 zunächst ein Überblick über die historische Entwicklung Genetischer Algorithmen gegeben, die dann um prägnante Anwendungsbeispiele aus Technik, Computer-Science und Chemie erweitert wird. Daran schließen sich Ausführungen zu den mathematischen Grundlagen der Spektrochemometrie und der Modellierung von Daten an. In diesem Kontext wird speziell auf die Problematik der Optimierung von Kalibrationsmodellen mit Hilfe klassischer, automatisierter Verfah-

ren eingegangen. Darüber hinaus findet sich hier eine Beschreibung des Aufbaus und des Funktionsprinzips Genetischer Algorithmen.

Der Ergebnisteil gliedert sich in zwei Kapitel: Er beginnt in Kapitel 6 mit Untersuchungen zur Kalibrationsoptimierung durch Selektion von Faktoren. Speziell wird hier die Problematik des Einsatzes bestimmter Selektionsverfahren und Fitnessfunktionen an praktischen Beispielen dargestellt. In Kapitel 7 liegt der Schwerpunkt auf den Überlegungen und Entwicklungsergebnissen zum Genetischen Algorithmus selbst. In den einzelnen Abschnitten wird hierbei detailliert auf Entwicklungsschritte eingegangen, die maßgeblich zur hohen Effizienz und Robustheit des im Rahmen des Projektes entwickelten Genetischen Algorithmus geführt haben.