

## 8 Zusammenfassung

Die automatisierte Erstellung und Optimierung von Kalibrationen bildet den anwendungstechnisch höchst aktuellen Hintergrund der vorliegenden Arbeit. Der analytisch orientierte Anwender der IR / NIR-Spektrometrie fragt heute mehr denn je nach einfachen Meß- und Auswerteverfahren. Er hat selten die Zeit, sich bei der Erstellung von Kalibrationen mit Fragestellungen der Chemometrie detailliert auseinanderzusetzen. Dies bedeutet, daß von der verwendeten Anwendungssoftware auf dem Gebiet der Kalibrationsoptimierung ein entsprechend hoher Automatisierungsgrad gefordert wird.

Wie läßt sich die Automatisierung einer so komplexen Aufgabe wie der einer spektrochemometrischen Kalibration erreichen? Der wichtigste und zeitintensivste Schritt in der Verwendung faktoranalytischer Verfahren zur Erzeugung von Kalibrationen auf Basis der *Principal Component Regression* (PCR) und ihrer Optimierung ist die Auswahl geeigneter Faktorkombinationen für die Modelle. In diesem Schritt müssen spektral relevante Informationen von störenden Artefakten separiert werden.

Hier gibt es drei Wege zur Automatisierung: Der erste besteht darin, die Frage nach der Auswahl geeigneter Variablen streng kombinatorisch anzugehen. Dies bedeutet, daß unter *allen* Kombinationsmöglichkeiten der Faktoren diejenige herausgesucht wird, die gemäß einem wohldefinierten Kriterium die beste Kalibration ergibt. Diese Vorgehensweise ist zwar effektiv, da sie immer das gesuchte Optimum liefert, aber nicht effizient, weil sie gerade bei komplexen Kalibrationsaufgaben wegen der zu hohen Rechenzeiten praktisch kaum durchführbar ist.

Deterministische Verfahren wie die *Forward-Stepwise Variable Selection* und die *Backward-Stepwise Variable Selection* stellen den zweiten Weg zur Lösung dieser Fragestellung dar. Beide Verfahren führen wesentlich schneller zu einer Lösung als die rein kombinatorische Vorgehensweise. Sie optimieren ein Modell ausgehend von einer bestimmten Startkombination durch sukzessives Weglassen oder Hinzunehmen einzelner Faktoren. Der Nachteil dieser Methoden besteht darin, daß abhängig vom Startpunkt nur die Optima in der nächsten Nachbarschaft gefunden werden, und unter Umständen eine bessere globale Lösung unentdeckt bleibt.

Eine Alternative zu diesen Methoden bieten Genetische Algorithmen (GAs), die zu den stochastischen Verfahren zählen. Sie berechnen nicht alle Kombinationsmöglichkeiten, sondern finden in einem adaptiven Prozeß gute Lösungen des Optimierungsproblems. Für die Untersuchungen zur Faktor-Selektion wurde ein

Genetischer Algorithmus entworfen und programmiert, dessen Fitnessfunktion speziell auf das beschriebene Optimierungsproblem zugeschnitten ist. Die Entwicklung erfolgte in Zusammenarbeit mit U. Depczynski vom Institut für Angewandte Mathematik und Statistik der Universität Hohenheim. Der GA beinhaltet neben den klassischen Operatoren und Auswahlverfahren ein Migrationsmodell, das den Austausch von Informationen zwischen einzelnen Populationen ermöglicht. Dieses Migrationsmodell und verschiedene Hybridisierungen führen zu einer hohen Robustheit des Verfahrens und ermöglichen den problemlosen Einsatz in kommerzieller chemometrischer Software. Der GA wird in der vorliegenden Arbeit erfolgreich auf die automatische Faktor–Selektion und die Optimierung von PCR–Kalibrationen angewendet. Durch seinen Einsatz wird ein Anwender bei der Erstellung neuer Kalibrationen durch die automatische Optimierung von Modellen unterstützt und zeitlich erheblich entlastet. Die Leistungsfähigkeit und Zuverlässigkeit des Verfahrens wird an realen Datensätzen aus den klassischen Anwendungsbereichen der NIR–Spektrometrie überprüft. Hierzu standen unter anderem Spektren von Weizenkörnern aus der Lebensmittelindustrie sowie von Olfen–Tabletten aus der pharmazeutischen Industrie zur Verfügung.

Im Gegensatz zu deterministischen Optimierungsverfahren liefern GAs nicht nur ein einzelnes Ergebnis, sondern in jedem Schritt des Optimierungsprozesses eine Ergebnismenge (Population). In der Population existiert eine Fülle von Lösungen (Faktor–Selektionen), die sich oft nur wenig voneinander unterscheiden. Bei der Optimierung von Kalibrationsmodellen ist es daher notwendig, diese Faktor–Selektionen hinsichtlich ihrer Güte zu bewerten und miteinander zu vergleichen. Diesem Problem wurde in der vorliegenden Arbeit besondere Aufmerksamkeit gewidmet. Aus den Untersuchungen resultierten eine Reihe von Fitnessfunktionen, die diese Problematik widerspiegeln und hinsichtlich ihrer Leistungsfähigkeit beurteilt werden.

Zur Bewertung von Kalibrationsmodellen ist es in der Spektrochemometrie gängige Praxis, die Standardabweichung der Vorhersagen von Kalibrations- und Validationsstandards als Gütekriterium heranzuziehen. Es wird gezeigt, daß die Verwendung eines Verfahrens, das auf der Varianz dieser Standardabweichungen basiert, die Möglichkeit zu einer pragmatischen Unterscheidung von Kalibrationsmodellen bietet. Durch Berücksichtigung der relativen Varianz lassen sich Intervallgrenzen um die Standardabweichung für den Vorhersagefehler definieren, so daß zwei Kalibrationsmodelle dann als unterscheidbar angesehen werden können, wenn sich diese Intervallgrenzen nicht überschneiden. Mit Hilfe des Varianz–Kriteriums konnten Kalibrationen ausgewählt werden, die im Hinblick auf zwei Punkte optimiert waren, nämlich hohe Genauigkeit (geringer Vorhersagefehler) und die Berücksichtigung möglichst weniger Faktoren (Robustheit).

Eine weitere Möglichkeit zur Bewertung von Kalibrationsmodellen ist die Berechnung des *Net Analyte Signals*. Diese relativ neue Methode erlaubt die univariate

Darstellung des Zusammenhangs zwischen Probenvarianz und spektraler Varianz. Der in der Analytik wichtige Parameter der Empfindlichkeit ist dadurch auch für multivariate Kalibrationen zugänglich und ermöglicht eine weitergehende Differenzierung zwischen Kalibrationsmodellen.

Unter Einbeziehung dieser Vergleichskriterien und verschiedener darüberhinausgehender Kenngrößen wie dem Durbin–Watson–Test wurde durch entsprechende Fitnessfunktionen mit dem Genetischen Algorithmus eine abschließende, universelle Lösung für die Problemstellung der Faktor–Selektion in der PCR erreicht.

Über die in dieser Arbeit beschriebenen Anwendungen hinaus ergeben sich im Rahmen der automatischen Kalibrationsoptimierung eine Reihe von weiteren Einsatzmöglichkeiten des Genetischen Algorithmus.

Eine Anwendung, die in der Literatur einen breiten Raum einnimmt, ist die Selektion von Wellenzahlen. Die Problematik ist hier ähnlich gelagert wie bei der Auswahl von Faktoren in großen Datensätzen. Auch hier macht die große Zahl denkbarer Kombinationen die vollständige Berechnung aller Kombinationsmöglichkeiten praktisch unmöglich und den Einsatz von Genetischen Algorithmen sinnvoll.

Eine weitergehende Automatisierung faktoranalytischer Verfahren kann durch eine Änderung der Art, in der Datensätze dargestellt werden, erfolgen. Zum Beispiel kann auf Basis einer Wavelet–Zerlegung sowohl eine verbesserte Vorbehandlung der Spektren (Pretreatment), im Sinne einer Kontrastverstärkung und Glättung, als auch eine eindeutigere Selektion spektraler Merkmale ermöglicht werden. Das damit verbundene Optimierungsproblem der Selektion geeigneter Wavelet–Koeffizienten konnte bereits durch den in dieser Arbeit vorgestellten Genetischen Algorithmus gelöst werden [81].

Die Kalibration von Spektrenbibliotheken zum Zwecke der Identitätskontrolle ist ein weiteres potentiell Anwendungsfeld Genetischer Algorithmen. Auch hier stellen sich analog zu quantitativen Kalibrationen die Fragen der Feature–Selektion und des optimalen Pretreatments. Auf Grund der Erkenntnisse der vorliegenden Arbeit und aus Untersuchungsergebnissen im Arbeitskreis ist davon auszugehen, daß auch für diesen Problemkreis Genetische Algorithmen ein effizientes Werkzeug sind. Der Weg zur Realisierung und vor allem die Frage nach geeigneten Fitnessfunktionen ist noch offen, so daß sich hier ein interessantes Feld für zukünftige Forschungen anbietet.