

Mathematische Modelle in der chemometrischen Praxis

Eckhard Freuwört, Lauenau

Aus der Instrumentierung eines modernen chemischen Laboratoriums ist der Computer nicht mehr wegzudenken. Häufig stehen sogar mehrere solcher Geräte zur Verfügung, da sie zum Bestandteil von GC-, HPLC-, Titrationseinheiten usw. geworden sind. Es ist jedoch eher die Ausnahme als die Regel, daß alle Rechner gleichzeitig benutzt werden. Somit liegt es nahe, die Wirtschaftlichkeit der PC zu erhöhen, indem in diesen Zeiten der Nichtbenutzung chemometrische Arbeiten durchgeführt werden mit dem Ziel, Meßdaten besser zu interpretieren und statistisch zu sichern.

In diesem Kontext ist die Untersuchung etwaiger Zusammenhänge zwischen den Meßwerten von besonderer Bedeutung. Werden solche Zusammenhänge aufgefunden und lassen sie sich formalistisch beschreiben, so bildet dies die Grundlage eines mathematischen Modells. Letzteres wiederum erlaubt es, die untersuchten Zusammenhänge auf dem Rechner zu simulieren und so zu optimieren. Im Endeffekt führt dies dazu, daß zahlreiche Versuchsreihen überflüssig werden. Einerseits wird damit die Umwelt entlastet und andererseits erhöht sich die Effektivität des betreffenden Labors durch eindeutige Ergebnisse in kürzerer Zeit, was zur Kosteneinsparung führt. Dieser Beitrag beschreibt, wie sich Zusammenhänge auf chemometrischem Wege, d. h. mit mathematischen Verfahren, einfach und schnell untersuchen lassen und welche Vorteile der Anwender im Bereich der Chemie durch die Benutzung mathematischer Modelle hat.

Die Mathematik versteht Aussagen, welche die Beziehungen zwischen den Elementen vorgegebener Mengen festlegen, als Axiome. Aus derartigen Aussagen abgeleitete, widerspruchsfreie logische Schlüsse führen zu einer formalisierten Theorie, deren mathematisch formulierte Schlußregeln als formales System oder als mathematisches Modell bezeichnet werden. Das Modell stellt ein idealisiertes Abbild der Wirklichkeit dar, welches dieser umso näher kommt, je detaillierter es aufgebaut ist. Es bildet daher ein Hilfsmittel zum Verständnis der Wirklichkeit auf der Grundlage von vergleichenden Untersuchungen. Solche lassen sich sehr effektiv mit der chemometrischen Arbeitsweise durchführen.

Die Chemometrie (auch: Chemometrik, wobei die Korrektheit beider Fachbegriffe strittig ist) ist ein vergleichsweise junges Teilgebiet der Chemie, welches sich erst mit dem Aufkommen

der Computer entwickelte. Erst diese gestatten es, aufwendige mathematische Berechnungen im vertretbaren Rahmen durchzuführen. Chemometrie wird als ein chemisches Fachgebiet definiert, in dem formale Methoden der Mathematik, Statistik und Informatik angewendet, adaptiert und entwickelt werden mit dem Ziel, chemisch-experimentelle Arbeiten optimal zu gestalten sowie chemische Meßdaten computerunterstützt zu interpretieren.

Für den Chemiker ist auf diesem Gebiet vor allem die Entwicklung von Methoden und Software zur Interpretation chemischer Daten von Interesse. Besondere Bedeutung hat in der Chemie das Konzept der multivarianten Daten: Beispielsweise kann eine Gruppe von Proben durch mehrere Kenngrößen charakterisiert und können deren Zusammenhänge untersucht werden. Die Chemometrie bildet daher zusammen mit der naßchemischen, forensischen und automatisierten Analytik eine der vier grundlegenden und sich gegenseitig ergänzenden chemischen Arbeitsweisen. Sie setzt voraus, daß ein Computer und die geeignete mathematische Software zur Untersuchung chemischer Daten zur Verfügung stehen.

Allgemeines zur Chemometrie

Unter welchen Voraussetzungen läßt sich nun ein mathematisches Modell erstellen und wie werden mit ihm Rezepturen, Anlagenparameter u. ä. optimiert? Der Weg zur Erstellung eines mathematischen Modells folgt diesem ungefähren Schema: Versuchsplanung, Probenahme, Messung, Auffinden der korrespondierenden Größen, Auswertung durch Regressionsanalysen, Zusammenfassung mehrerer Funktionsgleichungen zum Modell und Optimie-

rung dieses Modells durch die Festlegung von Bereichsgrenzen.

Bereits bei der Versuchsplanung kann der Computer unterstützend eingesetzt werden. Je nach Problemstellung und vorhandener Laborausstattung ist dies auch bei der Probenahme (z. B. Berechnung von Probenahmeplänen) und bei der Messung (z. B. vermittels On-line-Meßverfahren) möglich. Zum Auffinden der miteinander korrespondierenden Größen und Variablen ist der Rechner dann schon unerlässlich und der Benutzer besonders gefordert – schließlich handelt es sich bei diesem Schritt um den arbeitsaufwendigsten Teil. Anhand von miteinander korrespondierenden Meßdaten lassen sich im allgemeinen Zusammenhänge sehr leicht verdeutlichen – einer unabhängigen Variablen (Meßwert oder Eingangsgröße) wird dabei eine entsprechende abhängige Variable (Resultat oder Ausgangsgröße) zugeordnet. Zur Darstellung solcher Zusammenhänge sind in der Praxis vier unterschiedliche Methoden gebräuchlich:

- Beschreibung der Zuordnung mit Worten
- Angabe einer Wertetabelle
- Graphische Darstellung der gegeneinander im Koordinatensystem aufgetragenen Meßwerte und
- Angabe einer Funktionsgleichung

Grundsätzlich ist das erstgenannte Verfahren immer anwendbar. Beinahe zwangsläufig erweist es sich jedoch als ungenau oder (und) unpräzise, da eine derartige verallgemeinernde Beschreibung letztlich bereits interpretiert und somit auf die gemessenen Daten verzichtet. Als Beispiele dieser Zuordnungsart sind Berichte und Notizen zu nennen.

Mit dem zweiten Verfahren lassen sich immer nur einzelne und durch Messungen belegte Daten betrachten. Auch können sich Aussagen, welche aus den Meßdaten abgeleitet werden, lediglich auf die durch Messungen belegten Punkte beziehen. Ist bei diesem Verfahren die EDV zur Verwaltung der Daten einzusetzen, so wird ein elektronisches Arbeitsblatt, das Worksheet, benötigt (als Beispiele für die entsprechende Software seien Lotus 123, Excel, Symphony und Framework genannt). Eine solche Tabelle nimmt jeden einzelnen Meßwert auf und beansprucht demzufolge vergleichsweise viel Speicherplatz. Häufig erforderliche „Zwischenwerte“ fehlen.

Diesen Nachteil weist die dritte Methode nicht auf. Die graphische Darstellung erlaubt es, zusammenhängende

Werte intuitiv zu erfassen und die Art des Zusammenhangs auf einen Blick zu erkennen. Nachteilig ist hierbei, daß eine Weiterverarbeitung des (gezeichneten) Zusammenhangs auf elektronischem Wege nur sehr schwer zu realisieren ist. Der letztere Punkt ist jedoch die Voraussetzung für viele Auswertungen, zur Schaffung von Prozeßleitsystemen unter Verwendung von Steuerungen und Regelungen und zur Aufnahme von Know-how in eine Datenbank. Für alle diese Anforderungen kommt schließlich nur die Komprimierung der Meßwerte in einen funktionalen Zusammenhang oder anders ausgedrückt, in eine Funktionsgleichung, in Frage.

Das betreffende Verfahren wird als Regressionsanalyse bezeichnet und erfordert im allgemeinen auch den Einsatz des PC. Die Regressionsanalyse formt die Meßdaten zur Gleichung um. Kombiniert man daraufhin mehrere Gleichungen, welche auf dieselben unabhängigen Variablen zugreifen und als Resultate voneinander verschiedene abhängige Variablen ausgeben, so erhält man bereits ein erstes mathematisches Modell des untersuchten Zusammenhangs. Dieses Modell ist natürlich noch verbesserungsbedürftig, da u. a. Bereichsgrenzen zu berücksichtigen sind. Der „Lauf“ des (meist manuell optimierten) Modells mittels Computer führt daraufhin zu neuartigen Rezepturen, Produktionsparametern etc., welche noch im praktischen Versuch überprüft werden müssen. Der Vorteil dieser Methode ist darin zu sehen, daß wenig erfolgversprechende Versuche mit Hilfe des Rechners von vornherein ausgeschlossen werden und somit keine Kosten oder Umweltbelastungen verursachen.

EDV-Ausstattung

Statistische Auswertungen von Meßdaten sind typische Anwendungen der Chemometrie. Kennzahlen wie Mittelwert, Standardabweichung, Varianz oder Streuung lassen dabei zwar die Beschreibung einer Gesamtheit zu, jedoch sind Rückschlüsse auf etwaige Zusammenhänge zwischen korrespondierenden Größen kaum möglich. Aussagen über Zusammenhänge und darüber hinaus sogar Prognosen auf Basis der bekannten Meßdaten können jedoch abgegeben werden, wenn bestimmte Ausrüstungsanforderungen an die EDV erfüllt sind, so daß eine weitergehende

statistische Untersuchung der Meßdaten (beispielsweise die Korrelationsanalyse von Meßwert zu laufender Messungsnummer zwecks Auffindung systematischer Fehler) im Bereich des Möglichen liegt. Abbildung 1 weist auf die hierbei anwendbaren Verfahren hin. Resultierend werden dann die Anforderungen an die chemometrische EDV-Hardware primär durch die benutzte Software definiert.

Generell sind zur Erstellung mathematischer Modelle zwei Arten von Programmen erforderlich: eines zur „Übersetzung“ der empirisch bestimmten Daten in Funktionsgleichungen und eines zum „Verrechnen“ der Gleichungen miteinander. Daher wird zunächst einmal Software für die (multiple) Regressionsanalyse wie Curvefit, Regan, Discourv, Kwikstat, Lambda-Stat oder Simstat benötigt, die Meßdaten in funktionale Zusammenhänge umformt. Zusätzlich sind „Fit“-Programme wie Derive, Eureka, Maple V oder Mercury vonnöten, mit denen versucht wird, die Parameter für eine vorgegebene Gleichung so zu bestimmen, daß sich die resultierende Funktion optimal durch die eingegebenen Meßpunkte legt. Die Hardwareanforderungen resultieren nun aus der Softwarekombination – werden beispielsweise die Sharewareprogramme Simstat und Mercury kombiniert, so erhält man eine extrem preiswerte Lösung, welche bereits auf 286er AT mit 640 KByte RAM und Grafik ab EGA gut läuft und zudem noch zu gängiger kommerzieller Software wie Lotus 123 oder Harvard Graphics kompatibel ist. Die in diesem Beitrag aufgeführten Beispiele für Modelle benutzen durchweg die leicht verständliche Eureka/Mercury-Syntax, bei welcher geschweifte Klammern die Kommentare kennzeichnen und die formale Schreibweise die Anweisungen für den Computer enthält.

Versuchskonzepte

Nun sollte man sich jedoch keinesfalls der Illusion hingeben, daß sich ein sehr komplexer Zusammenhang mit Hilfe der Software auch immer modellieren läßt. Ein kompletter Herstellungsprozeß inklusive Anlagen, Personal, Marketing, Logistik etc. bildet für gewöhnlich ein komplexes dynamisches System und verhält sich aufgrund seiner Rückkopplungen indeterministisch, d. h. er ist nicht vorhersagbar. Beinahe

zwangsläufig treten innerhalb eines solchen chaotischen Systems jedoch (entsprechend der Ramsey-Theorie) regelmäßige Substrukturen auf, welche sich in Grenzen deterministisch verhalten. Man kann diese begrenzte Deterministik als Subsystem auffassen, welches ein mathematisch modellierbares und somit auch prognostizierbares Verhalten zeigt.

Um herauszubekommen, welche Einflüsse berücksichtigt werden müssen, um das Modell eines Subsystems zu erstellen, empfehlen sich Faktoren- und Blockversuche. Chemische Versuche werden üblicherweise derart geplant, daß sich bei der qualitativen Untersuchung von Versuch zu Versuch lediglich einer der beteiligten Parameter oder Faktoren verändert. Alle anderen Einflußgrößen werden dabei zwar mit erfaßt, jedoch konstant gehalten und können auf diese Weise keine Störungen verursachen. Man spricht in einem solchen Fall von einem „Ein-Faktoren-Versuch“. Die dabei notwendige Versuchsanzahl steigt exponentiell mit der Anzahl der zu untersuchenden Parameter an und errechnet sich nach

$$A = n^p - n$$

mit A = Versuchsanzahl und n = Anzahl der zu berücksichtigenden Faktoren oder Parameter.

Nach dieser herkömmlichen Versuchsmethodik wären zur Untersuchung eines durch nur vier Parameter beeinflussten Zusammenhangs bereits 252 Versuche notwendig – ein in vielen Fällen bereits nicht mehr zu vertretender Arbeitsaufwand. Bestehen zudem noch Wechselwirkungen zwischen den einzelnen Faktoren oder können konstant zu haltende Variablen nicht konstant gehalten werden, so wird der Versuch trotz des hohen Aufwandes ohne greifbares Ergebnis bleiben!

Die Vorgehensweise des Ein-Faktoren-Versuchs erweist sich daher meist als unzuverlässig. Sinnvoller ist es hingegen, wenn die Wirkung zweier Faktoren gleichzeitig untersucht wird. Man spricht in diesem Fall verallgemeinernd und nicht ganz korrekt von einem „Faktoren-Versuch“, obgleich eigentlich ein Zwei-Faktoren-Versuch gemeint ist. Die erforderliche Versuchsanzahl reduziert sich dabei zunächst einmal auf

$$A = 2^p$$

und rückt somit wieder in den Bereich des Machbaren. So verringern sich z. B. bei vier beteiligten Faktoren die 252 Versuche auf nur noch 16 Versuche, was einen immensen Zeitgewinn und daher hohe Effektivität bedeutet. Bei

dieser Vorgehensweise werden von Versuch zu Versuch immer zwei der beteiligten Parameter verändert. Bei geeigneter Parameterwahl läßt sich dabei vielleicht auf frühere Meßreihen zurückgreifen. Die Versuchsanzahl reduziert sich in derartigen Fällen nochmals deutlich.

Tabelle 1 zeigt die Planung eines solchen Zwei-Faktoren-Versuchs für drei beteiligte Parameter. Der Betrag „0“ weist dabei auf das Fehlen eines Faktors oder auf ein gewähltes Konzentrationsminimum hin, während der Betrag „1“ die Anwesenheit eines Faktors oder ein festgelegtes Konzentrationsmaximum kennzeichnet. Es ist aus dieser Art der Versuchsplanung unschwer erkennbar, daß dem rationelleren Zwei-Faktoren-Versuch das Zählen im Dualsystem, d. h. der binäre Zählalgorithmus zugrunde liegt. Dies prädestiniert den Computer geradezu zur Planung entsprechender Meßreihen. Eine entsprechende Hilfe bietet die Software Labtool 1. Zur Auswertung einer solchen Versuchsplanung sind nun die Differenzen aus denjenigen Versuchskombinationen zu bilden, bei denen sich von

Kombination zu Kombination immer nur der zu untersuchende Faktor verändert hat – dies setzt voraus, daß die numerische Reihenfolge der Versuche bei der Auswertung durchbrochen wird. Die Einflüsse des jeweils einzelnen Faktors 1a bis 1n, 2a bis 2n usw. errechnen sich dann nach

$$E = M - m$$

mit E = Einfluß des einzelnen Faktors/Versuch, M = Versuch mit Betragsmaximum/Faktor und m = Versuch mit Betragsminimum/Faktor.

Tabelle 2 zeigt die Auswertung für das oben angeführte Beispiel. Zur Bestimmung des Einflusses von Faktor 1 ist folglich der Mittelwert der Ergebnisse von E1a bis E1d zu bilden, für Faktor 2 wird der Mittelwert aus E2a bis E2d benötigt usw. Treten dabei innerhalb einer Reihe (z. B. von E1a bis E1d) größere Unterschiede zwischen den errechneten Differenzen auf, so deutet dies für gewöhnlich auf eine nicht-lineare Beeinflussung (z. B. Querempfindlichkeit) hin.

Derartige Auswertungen lassen sich bis hin zu etwa vier Faktoren noch recht gut manuell durchführen. Steigt die

Anzahl der beteiligten Parameter jedoch darüber hinaus an, so empfiehlt sich die Auswertung mit dem Computer. Die Basis einer derartigen Auswertung bildet dann die Matrixrechnung. Als Ergebnis erhält der Anwender damit nicht nur Aussagen über die qualitativen, sondern bereits über die quantitativen Zusammenhänge. Diese Art der Auswertung ist immer von Vorteil, wenn konstant zu haltende Parameter nicht konstant gehalten werden können.

Neben Labtool 1 bietet auch der Handel ein recht gutes Programm zur Faktorenversuchsplanung sowie -auswertung an. Es handelt sich dabei um das Programmpaket Factorial-X. Diese Software wurde im GWBasic-Quellcode erstellt, kompiliert und ist seit 1985 (leider nur in den USA) erhältlich. Neben der eigentlichen Versuchsplanung erstellt dieses Programm ein Meßprotokoll, in welches lediglich noch die Meßdaten des Anwenders einzutragen sind. Die Auswertung erfolgt dann automatisch anhand des Meßprotokolls und gibt prozentual die statistische Signifikanz der Einflüsse einzelner Faktoren inklusive ihrer Streuung an.



Abbildung 1: Zusammenhänge zwischen statistischen Verfahren

Faktoren:			Versuchs-Kombination:
A	B	C	
1	1	1	8
1	1	0	7
1	0	1	6
1	0	0	5
0	1	1	4
0	1	0	3
0	0	1	2
0	0	0	1

Tabelle 1: Faktorenversuch

E1a = M5 - m1	E2a = M3 - m1	E3a = M2 - m1
E1b = M6 - m2	E2b = M4 - m2	E3b = M4 - m3
E1c = M7 - m3	E2c = M7 - m5	E3c = M6 - m5
E1d = M8 - m4	E2d = M8 - m6	E3d = M8 - m7

Tabelle 2: Faktorenversuchsauswertung

Unser Service - Ihr Vorteil

CS - Chromatographie Service GmbH

Am Wehebach 26 - 52379 Langerwehe
Postfach 1207 - 52374 Langerwehe
Telefon 0 24 23 / 20 01 - Telefax 0 24 23 / 20 04

Analysis

Die Meßdaten selbst sollten generell über den gesamten zu untersuchenden Bereich verteilt sein. Als Optimum haben sich Nullpunkte, Einflußgrößen allein sowie miteinander kombinierte Komponenten zur Abdeckung des gesamten in Frage kommenden Bereichs erwiesen. Unter Verwendung derart verteilter Meßdaten eignet sich der Gauss-Jordan-Algorithmus zur Auswertung von Faktorexperimenten sowie zur Auswertung von Versuchen, bei denen es auf Grund von Sachzwängen nicht möglich ist, von mehreren zu erfassenden Parametern einige konstant zu halten.

Neben der Berechnung von Regressionskoeffizienten für Funktionsgleichungen muß die Software zusätzlich noch in der Lage sein, die Summe der Fehlerquadrate nach der Gaußschen Fehlerquadratmethode zu bestimmen. Dieser Wert gibt Hinweise auf die Genauigkeit der Analysis. Quantitativ brauchbare Ergebnisse erhält man im allgemeinen bei Fehlerquadratsummen <1 und noch bis etwa 10 bei eingeschränktem Gültigkeitsbereich. Ist hingegen eine eindeutige Lösung nicht möglich, so wird entweder kein Zusammenhang gefunden oder es liegt eine nur unzureichende Meßdatenanzahl vor. Bei der Verwendung von Meßwerten, welche den zu analysierenden Bereich nur unvollständig überdecken, ist ein solches Ergebnis ebenfalls möglich. Häufiger taucht dann jedoch eine Fehlerquadratsumme >10 anstelle der nicht eindeutigen Lösung auf. Diese besagt, daß das Ergebnis der Untersuchung nur noch qualitativ, d. h. zur Bestimmung von Tendenzen brauchbar ist.

Die Minimierung der Summe an Fehlerquadraten ist aufgrund geeigneter Parameterauswahl oder Meßwertumformung (z. B. Logarithmieren) nahezu immer möglich, so daß eine qualitative

A) Meßreihe zur Herzsüßlichkeit:

Herzsüßkohl (% V/V)	Andere (% V/V)	Sättigung (g/L)
20	80	0,01
50	50	0,3
70	30	2,5
90	10	22,1
100	0	66,5

B) Entsprechende Funktionsgleichung:

$$S = 0,001163 \cdot e^{0,11 \cdot A} \quad \text{mit}$$

S Herzsüßigkeit bei Sättigung (g/L)
A Anteil Herzsüßkohl am Lösungsmittel (% V/V)

Abbildung 2: Umformung von Meßwerten

Untersuchung bei einer großen Anzahl an zur Verfügung stehenden Variablen die Beschränkung auf das Wesentliche erlaubt. Eine besondere Bedeutung kommt der Fehlerquadratsumme noch zusätzlich bei der weiter unten beschriebenen Interpretation von Ergebnissen zu, welche vermittels eines mathematischen Modells errechnet wurden.

Mit den aus der qualitativen Untersuchung bestimmten, signifikanten Parametern wird die Meßreihe anschließend einer letzten – ggf. multiplen – Regressionsanalyse unterzogen. Damit erfolgt quasi die „Übersetzung“ der gemessenen Daten in eine Funktionsgleichung. Abbildung 2 zeigt ein entsprechendes Beispiel.

Eine solche Funktionsgleichung weist nun gleich mehrere Vorteile auf: Zum einen wird die Dokumentation vereinfacht, da die Gleichung alle gemessenen Daten repräsentiert und somit auf die Einzelwerte verzichtet werden kann und zum anderen ist die Meßreihe für den Computer verarbeitbar und der funktionale Zusammenhang optimierbar geworden.

Zumeist hat man es in der meßtechnischen Praxis jedoch mit multivariaten Daten zu tun, d. h. aus den gleichen Eingangsgrößen lassen sich mit differierenden Gleichungen auch unterschiedliche Resultate (z. B. die Konzentrationen mehrerer Reaktionspartner) er-

rechnen. Faßt man nun diese verschiedenen Funktionsgleichungen zu einem Gleichungssystem zusammen, so erhält man bereits ein erstes und noch recht einfaches mathematisches Modell des untersuchten Sachverhalts.

Modellaufbau

Obleich das Gleichungssystem bereits ein mathematisches Modell darstellt, sind hiermit noch keine Optimierungen von beispielsweise Rezepturen oder Produktionsparametern möglich, denn das Gleichungssystem allein ist auf dem PC nicht lauffähig. Es muß daher der – manuell durchzuführende – Arbeitsgang der Modellprogrammierung folgen, wobei die Gültigkeitsbereiche der Untersuchung, Eingangs- und eventuell Kenngrößen sowie die Definition der benutzten Variablen das Gleichungssystem ergänzen. Diese Arbeiten lassen sich mit den Editorfunktionen, über welche die obengenannten Fit-Programme verfügen, bewerkstelligen. Sollen hingegen zusätzlich umfangreiche Kommentierungen vorgenommen werden, so ist es sinnvoll, dies mit einem kompatiblen Texteditor zu tun. Es empfiehlt sich dabei die Beschränkung auf das ASCII-Format, so daß Sharewareprogramme wie Qedit Advanced, Writer usw. nützlich sein können.

Das so erweiterte Modell ist daraufhin in bezug auf seine Lauffähigkeit vervollständigt und kann vermittels des Fit-Programms für weitere Berechnungen herangezogen werden. Die so erzeugten Resultate bilden kybernetisches Wissen – da das mathematische Modell jedoch ein idealisiertes Abbild der Realität ist, kann dies auch für das kybernetische Wissen angenommen werden, sofern die Kontrolle durch den Lauf des Modells gute Übereinstimmung mit den zugrunde liegenden Meßdaten aufweist.

(Wird fortgesetzt)

Computer- und Software-Handhabung * Schule und Ausbildung * Autorensystem * Literaturverwaltung * Datenbanken * Rechenprogramme * Chemische Analytik * Chromatographie * Organische Chemie * Molecular Modelling * Biotechnologie * Projektmanagement * Computer Aided Creativity.

Das sind die Gebiete unseres Katalogs September '93. Fordern Sie ihn kostenlos an.

UMSCHAU SOFTWARE

UMSCHAU ZEITSCHRIFTENVERLAG Breidenstein GmbH

Postfach 11 02 62 60037 Frankfurt am Main
Tel. (069) 26 00-605 Hotline (02304) 8 18 54

Mathematische Modelle in der chemometrischen Praxis (Teil 2)

Eckhard Freuwört, Lauenau

Grundsätzlich ist zwischen fünf Arten von Modellen zu unterscheiden:

- Modelle zur exakten und direkten Berechnung von Resultaten
- Modelle zur exakten und indirekten Berechnung von Resultaten durch Auflösen des Gleichungssystems
- Modelle zur exakten und indirekten Berechnung von Resultaten unter Verwendung der Trial-and-error-Methode
- Modelle mit iterativer Lösung durch Intervallschachtelung
- Modelle mit iterativer Lösung durch Minimierung oder Maximierung von Variablen nach dem Simplexverfahren

Im einfachsten Fall erfolgt der Lauf direkt und exakt, d. h. bei diesem Modelltyp werden alle Eingangsgrößen eingegeben und anhand der Gleichungen das Resultat ermittelt. Abbildung 3 zeigt hierzu ein Beispiel, bei dem die Dichte und die Leitfähigkeit einer wäßrigen Lösung von Al in NaOH aus den Substanzkonzentrationen bestimmt werden. Dieser Modelltyp findet bevorzugt bei der Messungsauswertung und in der On-line-Meßtechnik Verwendung.

Häufig sind nur einige der Eingangsgrößen, dafür jedoch auch gleichzeitig die Resultate bekannt. In einem solchen Fall kann das Modell exakt berechnet

werden, wenn sich das Gleichungssystem hinsichtlich einer (oder mehrerer) Eingangsgrößen auflösen läßt. Abbildung 4 zeigt das Modell für eine entsprechende Mischungskreuzrechnung. Die Umstellung des Gleichungssystems wird dabei automatisch vom Fit-Programm durchgeführt, so daß lediglich die nur zum Teil bekannten Eingangsgrößen und das bekannte Resultat eingegeben werden müssen. Dieser Modelltyp eignet sich immer dann, wenn ein Zurückrechnen vonnöten ist, d. h. wenn Eingangsdaten rekonstruiert werden sollen.

Daneben treten – besonders in Gegenwart von Funktionen mit logistischem Verlauf – Fälle auf, in denen eine exakte Lösung des Gleichungssystems nicht immer möglich ist, da das Modell nicht schlüssig nach einer Variablen aufgelöst werden kann. Obgleich es sich aufgrund des Aufwandes um eine unbefriedigende Arbeitsweise handelt, muß man sich dann mit Hilfe der Trial-and-error-Methode an die „richtigen“ Werte herantasten. Der Funktionstyp $Y=K/(1+A \cdot e^{-(B \cdot X)})$ ist ein typisches Beispiel für solche Trial-and-error-Berechnungen. Setzt sich die Variable X in dieser Funktion nämlich aus weiteren Variablen zusammen, so kann die Gleichung

nicht nach diesen zusätzlichen Variablen aufgelöst werden – auch dann nicht, wenn die Beträge aller anderen Variablen bekannt sind. Es bleibt also letztlich nur das Ausprobieren. Dabei soll nicht verschwiegen werden, daß sich eine solche Näherung vermittelt einer Schleifenkonstruktion in einer fast beliebigen Programmiersprache eleganter durchführen läßt. Sofern X jedoch einen Meßwert repräsentiert, wird dieser Modelltyp bevorzugt, um bestimmte Eingangsgrößen durch Zurückrechnen zu ermitteln.

Abbildung 5 zeigt ein derartiges (Näherungs-) Modell, welches zunächst einmal das Resultat direkt und exakt bestimmt. Erst durch das Einsetzen verschiedener Eingangsgrößen kann das errechnete Resultat manuell mit einer Sollwertvorgabe verglichen werden. Die Eingangsgrößen sind daraufhin solange zu variieren, bis Resultat und Vorgabe übereinstimmen.

Neben dem Arbeitsaufwand ist die Problematik der Trial-and-error-Lösung darin zu sehen, daß man nie sicher sein kann, die optimalen Beträge der zu bestimmenden Variablen auch gefunden zu haben. Iterative Bestimmungen hingegen weisen diese Nachteile nicht auf. Bei diesem Modelltyp wird innerhalb definierter Bereichsgrenzen nach der Gauß'schen Methode zur Minimierung von Fehlerquadraten versucht, optimal passende Beträge für die Variablen zu finden. Häufig geschieht dies unter Anwendung von Intervallschachtelungsverfahren wie der Hénon'schen Intervallschachtelung, für deren Anwendung die Abbildung 6 ein Beispiel zeigt.

Die Verwendung eines solchen iterativ arbeitenden Modells kommt der Generierung von Rezepturen und Prozeßparametern durch den Computer selbst bereits sehr nahe, wobei selbstverständlich auch in diesem Falle Funktionsgleichungen auf der Basis von empirischen Messungen die Grundlage bilden. Allerdings müssen bei weitem nicht alle Beträge der benutzten Variablen vorgegeben werden. Es reicht vielmehr aus, etwa jeweils ein Viertel der Variablen als Sollwert und als Bereichsdefinition (z. B. in Form eines Mindestbetrages) festzusetzen. Dabei erfordert die Modellprogrammierung jedoch einiges an Erfahrung, um brauchbare Resultate zu erzielen. In besonderem Maße gilt dies für die Auswahl der vorzuziehenden Sollwerte und Bereichsgrenzen.

Gleiches gilt für einen weiteren Modelltyp, mit welchem die Lösung ebenfalls auf iterativem Wege ermittelt wird. Es handelt sich dabei um Modelle, mit

Rechenprogramme der UMSCHAU SOFTWARE

Paket Rechenprogramme

Nützliche Mathematik
DM 115,-

STATCAL

Statistik für Analytiker
DM 431,25 *

OPAL

Funktionen und Grafik
DM 483,- *

REGAN

Regressionsanalyserechner
DM 92,-

STAREG

Statistik und Regression
DM 316,25 *

MEGRAX

Meßwert- und Funktionsverläufe
DM 333,50 *

Lassen Sie sich informieren. * Demodiskette erhältlich: DM 23,-
Katalog und Langbeschreibungen: (069) 2600-605, Fax 2600-619
Hotline für technische Fragen: (02304) 81854, Fax 83271

UMSCHAU SOFTWARE

Umschau Zeitschriftenverlag
Breidenstein GmbH
Stuttgarter Straße 18 - 24
60239 Frankfurt am Main

```

(L Leitfähigkeit (mS/cm) bei 20°C und α 2,185/K )
{
(B Dichte (g/mL) bei 20°C )
(MM1 NaOH-Konzentration (g/L) ± 7% Abweichung im Bereich 2-6 g/L
(AL Anteil an gelöstem Al (g/L) ± 24% Abweichung im Bereich 0,1-3 g/L
{
----> Berechnet werden L und B <----
}
MM1= (Eingabe)
AL= (Eingabe)
L=4,523*MM1-1,992*AL+0,752 (Berechnung)
B=1,334E-3*MM1+0,337E-3*AL+0,9973 (Berechnung)
    
```

Abbildung 3: Modell mit direkter und exakter Lösung

```

{KST Konzentration der starken Komponente (SG/G) }
{KMI Konzentration der Mischung (SG/G) }
{KVK Konzentration der Verdünnung (SG/G) }
{MST Masse der starken Komponente (z. B. g, Kg, etc.) }
{MMI Masse der Mischung (z. B. g, Kg, etc.) }
{MVK Masse der Verdünnung (z. B. g, Kg, etc.) }
{
----> Berechnet werden KST und MVK <----
}
MST= (Eingabe)
KMI= (Eingabe)
KVK= (Eingabe)
MMI= (Eingabe)
MST=MMI*(KMI-KVK)/(KST-KVK) (Berechnung)
MVK=MMI*(KST-KMI)/(KST-KVK) (Berechnung)
    
```

Abbildung 4: Modell mit Lösung durch Gleichungsumstellung

```

(L Materialhelligkeit L* nach CIE-Lab-Norm )
{
bei der alkalischen Beizung von Aluminium }
(T Verweilzeit im Beizprozeß (s) im Bereich )
{
von 5 bis 50 Sekunden bei konstanter )
{
Temperatur }
{
Empirischer Grenzwert der Materialhelligkeit }
{
----> Berechnet wird L <----
}
T= (Eingabe)
K=87,3 (Konstante)
L=K/(1+0,1718*EXP(-0,05377*T)) (Berechnung)
    
```

Abbildung 5: Modell mit Trial-and-error-Lösung

```

(DV Sättigungslöslichkeit des DSM-Polyesters HYDOLITE 9193W (g/L))
(SMA Sättigungslöslichkeit des Styrol-Maleinmaleinmethylen-Copolymer)
{
Monsanto-Scorpeet 520 (g/L)
}
{EP Sättigungslöslichkeit des Shell-Epoxydharzes Epikote 1031 (g/L)}
(DMF Mischungsanteil N,N-Dimethylformamid (% V/V))
(MEK Mischungsanteil Diäthan-2 (% V/V))
(ME Mischungsanteil 2-Methoxyethanol (% V/V))
(EO Mischungsanteil 2-Ethoxyethanol (% V/V))
(SOAS S aller sonstigen Lösungsmittel in der Mischung (% V/V))
{
----> Berechnet werden DMF, ME, EO, ME, SMA und EP <----
}
SOAS= (Eingabe Lösungsmittelwert)
MEK= (Eingabe Lösungsmittelwert)
EP= (Eingabe Harzbereich)
DMF= (Eingabe Harzbereich)
SMA= (Eingabe Harzbereich)
S=DMF+EO+MEK+DMF+EO (Konstante Bereichsdeklaration)
EO=EO+EO (Konstante Bereichsdeklaration)
MD=100-SOAS-MEK-DMF-EO (Berechnung)
(E=3,5)*DMF+1,61*MEK+0,54*SOAS+27,81 (Berechnung)
SMA=1,816*MD+0,7119*EO+0,4460 (Berechnung)
EP=6,64*MD+13*DMF+6,24*MEK+3*SOAS (Berechnung)
    
```

Abbildung 6: Modell mit iterativer Lösung innerhalb definierter Bereichsgrenzen

```

A) Empirische Meßreihe:
B) Garaus errechnete Funktionsgleichung:
C) Optimierungsprogramm:
D) Lösung bzw. Rezeptur:
    
```

B (g/L)	D (g/L)	C (SG/G)
186	100	1,2
146	50	2,4
312	100	1,2
186	300	0,7

$C = 2/B + 31,22 + 1/D * 109,93$

```

(E Amino-cis-N-Methylenphosphonsäure (g/L))
(F N-Methylolchloracetamid-Derivat (g/L))
(C Signifikant wirksame Isolierung des
wäßrigen Konservierungsmittels (% G/G))
{
----> Berechnet wird C <----
}
G= (Bereichseingabe)
H= (Bereichseingabe)
H=0,550 (Konstantendeklaration)
C=(1/H)*31,22+(1/D)*109,93 (Berechnung)
MINIMIZE C (Berechnung)
    
```

Variables:

B = +191,2065675069286

D = +358,793432493744

C = 1,0831,22+1,09*109,93

= +0,469566690426246642

Largest residual is 6,7E-10

Method: Iterative

Abbildung 7: Modell mit iterativer Lösung durch Minimierung oder Maximierung

H. J. E. Loewenthal E. Zass
Der clevere Organiker
Leitfaden zum Erfolg in der Synthese

Johann Ambrosius Barth

Loewenthal/Zass
Der clevere Organiker
Leitfaden zum Erfolg in der Synthese

Von H. J. E. Loewenthal und Engelbert Zass.
1993. VIII, 230 S., 80 Abb., 14 Tab.
Gebunden DM/sFr 48,- öS 375,- ISBN 3-335-00360-8

Der clevere Organiker schließt eine Lücke zwischen dem, was Studierende in Vorlesungen und Seminaren gelernt haben, und dem, was sie für die tägliche praktische Arbeit im Labor wissen müssen. Es werden Hinweise zur Literaturrecherche (einschließlich computer-gestützter Nachforschungen) sowie praktische Hinweise für die Arbeit im Labor gegeben.

(Originaltitel: „A Guide for the Perplexed Organic Experimentalist“, Wiley & Sons)

BESTELLCOUPON

Bitte bestellen Sie bei Ihrem Buchhändler oder direkt beim Verlag
Exp. Loewenthal/Zass, Der clevere Organiker
DM/sFr 48,- öS 375,- ISBN 3-335-00360-8

Anschaff. _____
Unterschrift _____

denen Minimierungen oder Maximierungen durchgeführt werden. Vermittels des Simplexverfahrens wird das Gleichungssystem unter Berücksichtigung der angegebenen Grenzen hinsichtlich einer einzigen Variablen optimiert. Dies geschieht, indem im Falle der Minimierung eine Variable den kleinstmöglichen und alle Randbedingungen noch erfüllenden Betrag und im Falle der Maximierungen den größtmöglichen Betrag erhält. Der Computer generiert auf diese Weise selbständig z. B. eine optimale Zubereitungsrezeptur. Abbildung 7 zeigt ein entsprechendes Beispiel. Auch bei solchen Modellen ist Fingerspitzengefühl (oder Übung) notwendig, um alle Bereichsgrenzen so vorzugeben, daß unsinnige Resultate (wie beispielsweise negative Konzentrationsangaben) von vornherein ausgeschlossen sind.

Modellanwendung

Im Anschluß an die Programmierung eines mathematischen Modells folgt logischerweise dessen Lauf, mit welchem eine reale Möglichkeit vorab auf dem Computer simuliert wird. Dazu müssen zunächst die zur Verfügung stehende Meßwerte für die Eingangsvariablen eingegeben werden. Daraufhin versucht die Software, das Gleichungssystem zu lösen und gibt die Resultate aus. Auswertungen werden auf diese Art bei Modellen mit direkter und exakter Lösung sowie bei Modellen mit Gleichungsumstellung enorm vereinfacht. Auch ist es problemlos möglich, bei bekannten Resultaten unbekannte Eingangswerte zu rekonstruieren, so daß die Wiederholungsmessung häufig überflüssig wird. Mit Trial-and-error-Modellen kann darüber hinaus versucht werden, abweichende Betriebsbedingungen für einen Prozeß oder für eine Zubereitung aufzufinden.

Gerade der letztgenannte Punkt ist im Bereich der Chemie von eminenter Wichtigkeit, denn nur allzuoft ist ein zu analysierendes Problem vielschichtig und wird daher von einer Vielzahl an Variablen beeinflusst. Man spricht in diesem Zusammenhang auch von „multidimensionalen Systemen“. So kann es beispielsweise vorkommen, daß ein Produktionsprozeß von Faktoren wie Temperaturen, Konzentrationen, Behandlungszeiten und Anzahl oder Art der Prozeßschritte abhängt. Meist wird dann versucht, eine vorgegebene Pro-

duktqualität durch die Konstanthaltung von empirisch ermittelten Erfahrungswerten zu gewährleisten. Dabei handelt es sich jedoch nur um einen einzigen aus einer ganzen Anzahl von möglichen Betriebsbereichen, und es ist keineswegs sichergestellt, ob nicht in einem anderen, abweichenden Betriebsbereich ökologisch und ökonomisch günstiger produziert oder gearbeitet werden kann. Da für derartige Betriebsbereiche jedoch im allgemeinen auch keine Erfahrungswerte vorliegen, werden sie meist für schlichtweg unmöglich gehalten.

Doch genau hier zeigen die iterativ arbeitenden Modelle ihre wahre Stärke. So lassen sich Versuche durch die Variation der Grenzwerte und der Eingangsdaten bereits am Bildschirm ausschließen oder als geeignet beurteilen oder aber abweichende Betriebsbereiche durch die Minimierungs- und Maximierungsverfahren auffinden. Zubereitungsrezepturen beispielsweise enthalten meist eine Reihe von Komponenten, von denen jede einzelne eine bestimmte Wirkung bei der Anwendung des Produktes ausüben soll. Dabei wird normalerweise nicht berücksichtigt, daß auch Nebenwirkungen auftreten, welche zusammen u. U. eine der enthaltenen Komponenten ersetzen oder deren Wirkung verstärken können (Synergetik). Als Beispiel hierfür mag Glykol dienen – primär läßt sich diese Substanz als Frostschutzmittel einsetzen und sekundär wirkt sie darüber hinaus noch als Lösemittel und stabilisiert Zubereitungen gegen zu rasche Verdunstung. Das Modell beachtet solche Tatsachen jedoch und kann daher zu einer Vereinfachung der Zubereitungszusammensetzung führen. Zudem kann auf einfache Weise das Zusammenwirken verschiedener Produkte (z. B. Film und Entwickler) vorab getestet werden, so daß es durchaus möglich ist, etwas von der Flexibilität, Stabilität und Belastbarkeit, welche biologischen Systemen zu eigen ist, auch in chemische Systeme einzubauen.

Die Resultate eines Modell-Laufs können als Parameter für Rezepturen, Betriebsbereiche etc. (z. B. Trocknungsoptimierung von Beschichtungen, Lackentwicklung, Produktfunktionalität, Lösungsstabilität) benutzt werden und sind noch im Kontrollversuch zu überprüfen. Hierbei ist die Fehlerquadratsumme des Resultates zu beachten. Ist diese gering, so errechnet der PC auf der Basis des bekannten Wissens die optimierten Bedingungen selbständig und die Übertragung der Ergebnisse in die Praxis ist problemlos möglich. An-

ders sieht es aus, wenn ein Resultat mit einer hohen Fehlerquadratsumme behaftet ist. Eine solche weist für gewöhnlich darauf hin, daß die geforderten Sollwerte mit den vorliegenden Fakten nicht erreichbar sind.

Allerdings müssen hier zwei Ausnahmefälle berücksichtigt werden. Zum einen ist es manchmal möglich, mit nur geringfügig variierten Sollwerten dennoch ein akzeptables Resultat zu erreichen. Zum anderen ist es mitunter nützlich, das Praxis-Handling bei der Herstellung eines Produktes zu ändern (abweichende Lösebedingungen wie Temperatur, Zeit, pH-Wert, Rührerenergie etc.). Unter solchen Bedingungen ist es etwa in der Hälfte der Fälle mit Fehlerquadratsummen bis ca. 30 noch möglich, die Resultate des Modell-Laufs in die Praxis zu übertragen. Schließlich erlaubt es die Kombination verschiedener Modelle noch, ganze Prozeßbläufe auf dem Rechner zu simulieren oder zu entwerfen.

Fazit

Die Chemometrie ergänzt sehr effektiv die bisher üblichen chemischen Arbeitsweisen. Der dazu als Hardwarekomponente erforderliche Rechner findet sich heute ohnehin in fast jedem Labor. Eine geeignete Softwareauswahl ermöglicht es dann, mit nur sehr geringen Zusatzkosten Meßwerte statistisch so aufzubereiten, daß weitergehende Prognosen für überschaubare, deterministische Teilbereiche möglich sind. Solche Prognosen basieren auf den Ergebnissen von mathematischen Modellen, welche selbst wiederum durch die statistisch aufbereiteten Meßdaten repräsentiert werden. Mathematische Modelle in der chemometrischen Praxis dienen somit u. a. der Auswertung, der elektronischen Weiterverarbeitung von Daten, der Simulation von Versuchen und Prozeßschritten sowie der Optimierung und Generierung von Parametern im technischen Bereich. Durch diese Arbeiten am PC lassen sich zahllose praktische Versuche ersetzen. Dies reduziert Kosten und Personalaufwand und ist somit ökonomisch sinnvoll. Zusätzlich entfällt die Umweltbelastung durch verbrauchte Chemikalien. Darüber hinaus können mit den dargestellten Verfahren mitunter noch Prozeßschritte optimiert werden, so daß es zu weiteren ökologischen und ökonomischen Vorteilen kommt.