

K. Danzer · H. Hobert · C. Fischbacher · K.-U. Jagemann

Chemometrik

Grundlagen und Anwendungen

Mit 206 Abbildungen



Springer

Prof. Dr. Klaus Danzer
Institut für Anorganische
und Analytische Chemie
Friedrich-Schiller-Universität Jena
Lessingstr. 8
07743 Jena
e-mail: klaus.danzer@uni-jena.de

Doz. Dr. Hartmut Hobert
Institut für Physikalische Chemie
Friedrich-Schiller-Universität Jena
Lessingstr. 10
07743 Jena
e-mail: chl@uni-jena.de

Dr. Christoph Fischbacher
Analytik Jena AG
Konrad-Zuse-Str. 1
07745 Jena
e-mail: c.fischbacher@analytik-jena.de

Dr. Kay-Uwe Jagemann
Brooks Automation GmbH
Goeschwitzer Str. 25
07745 Jena
e-mail: kuj@jjnet.de

Die Deutsche Bibliothek – CIP-Einheitsaufnahme
Chemometrik: Grundlagen und Anwendungen / von Klaus Danzer ... - Berlin; Heidelberg; New York;
Barcelona; Hongkong; London; Mailand; Paris; Singapur; Tokio: Springer, 2001
ISBN 978-3-642-62525-1 ISBN 978-3-642-56584-7 (eBook)
DOI 10.1007/978-3-642-56584-7

Dieses Werk ist urheberrechtlich geschützt. Die dadurch begründeten Rechte, insbesondere die der Übersetzung, des Nachdrucks, des Vortrags, der Entnahme von Abbildungen und Tabellen, der Funk- sendung, der Mikroverfilmung oder der Vervielfältigung auf anderen Wegen und der Speicherung in Datenverarbeitungsanlagen, bleiben, auch bei nur auszugsweiser Verwertung, vorbehalten. Eine Vervielfältigung dieses Werkes oder von Teilen dieses Werkes ist auch im Einzelfall nur in den Grenzen der gesetzlichen Bestimmungen des Urheberrechtsgesetzes der Bundesrepublik Deutschland vom 9. September 1965 in der jeweils geltenden Fassung zulässig. Sie ist grundsätzlich vergütungspflichtig. Zuwiderhandlungen unterliegen den Strafbestimmungen des Urheberrechtsgesetzes.

<http://www.springer.de>

© Springer-Verlag Berlin Heidelberg 2001
Ursprünglich erschienen bei Springer-Verlag Berlin Heidelberg New York 2001

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in diesem Werk berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, dass solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jeder- mann benutzt werden dürften.

Satz und Datenkonvertierung: Fotosatz-Service Köhler GmbH, 97084 Würzburg
Einband: de'blik, Berlin

Gedruckt auf säurefreiem Papier SPIN: 10097902 52/3020 ra – 5 4 3 2 1 0

Vorwort

Die Chemometrik wurde noch vor einigen Jahren von der Mehrzahl der Chemiker nicht als branchenübliches Werkzeug akzeptiert und angewendet. Sie wurde stattdessen als Domäne einiger weniger Insider betrachtet, und es wurde demzufolge ein großer Nachholbedarf in der Ausbildung von Chemikern festgestellt. Zu den erklärten Zielen des Arbeitskreises *Chemometrik und Labordatenverarbeitung* der Fachgruppe *Analytische Chemie* der *Gesellschaft Deutscher Chemiker* gehörte es deshalb, sich für die Einführung von Lehrveranstaltungen zur Chemometrik an möglichst vielen deutschen Hochschulen einzusetzen. Tatsächlich wurde dieses Ziel jedoch nur an einigen wenigen Universitäten, zu denen unter anderen Freiberg, Jena und Hannover gehören, erreicht.

Die Forderung nach Kenntnis chemometrischer Methoden und deren Beherrschung steht noch immer. Die im Rahmen analytischer oder allgemein chemisch-experimenteller Untersuchungen anfallenden Datenmengen nehmen an Umfang und Komplexität weiterhin zu und erfordern neue, nichtkonventionelle Herangehensweisen zu ihrer Auswertung und Beurteilung. Dennoch sind die Rufe nach einer umfassenden chemometrischen Grundausbildung im Rahmen des Chemiestudiums zurückhaltender geworden.

Die Situation hat sich nämlich in gewisser Weise gewandelt. Interessierte Chemiker sind mit der derzeitigen Generation von Personalcomputern sowie der kommerziell und über Internet verfügbaren Software auch ohne spezielle Ausbildung in der Lage, die Chemometrik für die Lösung ihrer speziellen Probleme nutzbringend anzuwenden. Bis zu einem bestimmten Grade ist die Chemometrik bereits zu einem alltäglichen Handwerkszeug vieler Chemiker geworden, die im Rahmen komplexer Aufgabenstellungen Messungen zu planen, durchzuführen oder zu überwachen und auszuwerten haben. Damit benutzt mancher Chemiker chemometrische Methoden, ohne diese explizit als solche zu kennen. Die nutzbringende und sinnvolle Anwendung von chemometrischen Verfahren setzt jedoch Kenntnisse der Möglichkeiten und Grenzen der verwendeten Methoden und Modelle voraus.

Mit diesem Buch sollen deshalb die Grundlagen der Chemometrik in einem Überblick dargelegt werden, der sowohl Fachleuten eine systematische Darstellung der wichtigsten Teilgebiete, Methoden und Anwendungen gibt, als auch Einsteigern eine Möglichkeit bietet, die Grundlagen und potentiellen Anwendungsfelder der Chemometrik kennenzulernen. Von den wahrscheinlichkeitstheoretischen Grundlagen chemischer Messungen auf der Basis der univariaten

und multivariaten Statistik bis hin zu Neuronalen Netzen und Genetischen Algorithmen als Methoden der künstlichen Intelligenz spannt sich der Bogen bis zu deren Anwendung in wichtigen Bereichen der experimentellen Chemie, insbesondere der Analytik. Dabei dient die logische Folge chemischer Messprozesse als Prinzip für die Ordnung und Anwendungsmöglichkeit chemometrischer Methoden und Techniken.

Die Schritte des Analytischen Prozesses, Probennahme, Proben- und Messvorbereitung, Messung, Kalibration und Auswertung sowie Dateninterpretation, finden sich in der Gliederung in entsprechenden Kapiteln wieder. Statistische Probennahmeprinzipien, Versuchsplanung und Optimierung chemischer Messverfahren werden in ihren theoretischen Grundlagen und Anwendungsaspekten ebenso behandelt wie Methoden der Signalverarbeitung, der Darstellung und Bearbeitung von Messwertreihen, insbesondere von Spektren, wobei Analogien zu Prinzipien der Bildverarbeitung, die in der Analytik ebenfalls zunehmend Bedeutung erhält, dargestellt werden.

Es gibt Analytiker, die es für eine bemerkenswerte Entwicklung halten, dass sich analytische Informationen messen, quantifizieren und in ihrer Effizienz und Rentabilität in Mark und Pfennigen ausdrücken lassen. Diese Aspekte werden bei der wachsenden Daten- und Informationsflut auch in der Chemie weiterhin an Bedeutung gewinnen. Informationstheoretische Grundlagen der Analytik wurden deshalb in einer angemessenen Breite berücksichtigt und mit Anwendungsbeispielen belegt.

Die wachsende Rolle von Prinzipien der Qualitätssicherung in vielen Bereichen der Chemie, insbesondere auch in der Analytik selbst, führt dazu, dass nicht nur die administrativen und formalen Aspekte, sondern auch die statistischen und chemometrischen Gesichtspunkte der Qualitätssicherung, und zwar auch innerhalb der Chemometrik und der Softwareentwicklung, immer breiteres Interesse gewinnen. Dem wurde durch ein eigenes Kapitel Rechnung getragen.

Durch ein ausgewogenes Verhältnis von theoretischen Grundlagen und praktischen Anwendungsbeispielen wurde versucht, dem Titel des Buches gerecht zu werden. Das Buch versucht, das heute sehr umfangreiche Gebiet der Chemometrik in einem umfassenden Überblick zu behandeln. Vollständigkeit war dabei verständlicherweise nicht möglich. Die erforderlichen Abstriche wurden jedoch nicht bei der Anschaulichkeit, der Darstellung von Zusammenhängen zwischen den Methoden und der Praxisrelevanz vorgenommen, sondern eher bei der theoretischen Herleitung mathematischer Sachverhalte, bei denen auf Standardwerke der Statistik und der numerischen Mathematik verwiesen werden kann.

Mit diesem Buch sollen gleichermaßen interessierte Studenten und Lehrkräfte an Universitäten und Hochschulen, als auch Praktiker angesprochen werden, die chemometrische Methoden bereits in der einen oder anderen Form nutzen oder aber in ihrer Arbeit anzuwenden beabsichtigen. Die Verfahren selbst stehen, wie erwähnt, als Software zur Verfügung. Eine gewisse Orientierungshilfe zu Einsatzmöglichkeiten und Anwendungsvorzügen wird hier gegeben, ohne dass dies als ausgesprochenes Ziel des Buches anzusehen ist.

Danken möchten wir einer Reihe von Fachkollegen, mit denen als Weggefährten auf dem Gebiete der Chemometrik im Rahmen von Begegnungen oder gar längeren Zusammenarbeiten mancher fruchtbarer Erfahrungsaustausch stattfand. Da dies erschöpfend nicht möglich ist, seien hier nur hervorgehoben die leider viel zu früh verstorbenen Klaus Doerffel, Merseburg/Leipzig, und Karel Eckschlager, Prag, sowie Günter Ehrlich, Dresden, und Günter Henrion. Ganz besonderer Dank gebührt Dietrich Wienke, Nijmegen, der in seiner Jenaer Zeit dieses Buch mit konzipierte, es dann aber nicht mit ausführen konnte.

Den Partnern im Springer-Verlag danken wir für die konstruktive Zusammenarbeit in allen Bearbeitungsphasen des Buches.

Jena, März 2001

Klaus Danzer
Hartmut Hobert
Christoph Fischbacher
Kay-Uwe Jagemann

Inhaltsverzeichnis

1	Gegenstand der Chemometrik	1
1.1	Entwicklung der Chemometrik	1
1.2	Gegenstand und Aufgabenbereiche der Chemometrik	2
1.3	Prinzipien und Teilgebiete der Chemometrik	5
1.4	Chemometrik im Internet	8
	Literatur	9
2	Chemische Messungen	12
2.1	Allgemeine Grundlagen	12
2.2	Dimensionalität chemischer Messungen	16
2.3	Wahrscheinlichkeitstheoretische und statistische Grundlagen	20
2.3.1	Messwerthäufigkeiten und -verteilungen	20
2.3.2	Höherdimensionale Verteilungen von Zufallsgrößen	26
2.3.3	Bedingte Wahrscheinlichkeiten und Verteilungen	27
2.3.4	Statistische Parameter: Mittelwerte und Streuungsmaße	29
2.3.5	Vertrauensbereiche	32
2.4	Abhängigkeitsuntersuchungen: Korrelation und Regression	33
2.4.1	Korrelationsanalyse	35
2.4.2	Regressionsanalyse	37
2.4.3	Fehlergrößen der linearen Regression	40
2.4.4	Robuste Regression	44
2.5	Informationstheoretische Grundlagen	45
2.5.1	Informationsgehalt von Messergebnissen	47
2.5.2	Quantitative Messergebnisse unter dem Aspekt der Präzision	52
2.5.3	Quantitative Messergebnisse unter dem Aspekt der Richtigkeit	56
2.5.4	Signal-Rausch-Verhältnis und Informationsgehalt	59
2.5.5	Die Informationsmenge von Mehrkomponentenanalysen	61
2.5.6	Redundanz von Analyseergebnissen	64
2.5.7	Informationsmenge und Informationsleistung von zeit- und orts aufgelösten chemischen Informationen	68
2.6	Systemtheoretische Aspekte chemischer Messungen	70

2.6.1	Verbesserung des Signal-Rausch-Verhältnisses	72
2.6.2	Verbesserung der Signalauflösung	74
2.7	Von der univariaten zur multivariaten Statistik	77
	Literatur	78
3	Multivariate Datenanalyse	81
3.1	Daten und Datenräume	83
3.1.1	Variablentypen	83
3.1.2	Datenmatrix	84
3.1.3	Lineare Modelle	89
3.1.4	Varianz-Kovarianz- und Korrelationsmatrix	89
3.1.5	Multivariate Distanzmaße	92
3.1.6	Vorgehensweise bei der Datenanalyse	96
3.2	Datenerkundung und -aufbereitung	98
3.2.1	Explorative Datenanalyse und Datenvisualisierung	98
3.2.2	Datenaufbereitung	103
3.3	Faktoren- und Hauptkomponentenanalyse	105
3.4	Clusteranalyse	111
3.4.1	Hierarchische Clusteranalyse	112
3.4.2	Nichthierarchische Verfahren	116
3.5	Klassifikationsverfahren	117
3.5.1	Lineare Diskriminanzanalyse	121
3.5.2	Methode der k -nächsten Nachbarn	123
3.5.3	Weitere Methoden	124
3.6	Regression und Modellierung	125
3.6.1	Multiple lineare Regression	126
3.6.2	Hauptkomponentenregression	127
3.7	Softwareaspekte	128
3.7.1	Datenformate	128
3.7.2	Datenanalysesoftware	128
	Literatur	131
4	Probennahme	133
4.1	Repräsentanz von Proben	133
4.2	Inhomogene Untersuchungsobjekte	136
4.3	Repräsentative Probenanzahl	138
4.4	Experimentelle Ermittlung des Probennahmefehlers	140
4.5	Zeitabhängige Probennahme	146
4.6	Geostatistische Probennahmemodelle	151
	Literatur	153

5	Planung und Optimierung chemischer Experimente und Messungen	156
5.1	Statistische Versuchsplanung	156
5.1.1	Vollständige Versuchspläne	157
5.1.2	Unvollständige Versuchspläne	159
5.1.3	Auswertung von Faktorplänen	162
5.1.4	Anwendungen der SVP	163
5.2	Optimierungsverfahren	166
5.2.1	Response-Surface-Methode	168
5.2.2	Box-Wilson-Optimierung	168
5.2.3	Sequentielle Simplex-Optimierung	169
5.2.4	Globale Optimierungsverfahren	171
	Literatur	175
6	Signal- und Bildverarbeitung	176
6.1	Charakteristik von Signalen	176
6.1.1	Signalentstehung	176
6.1.2	Signaltypen	177
6.1.3	Prinzipien der Signalverarbeitung	179
6.2	Fourier-Transformation	180
6.2.1	Fourier-Integral	180
6.2.2	Eigenschaften von FT-Paaren	183
6.2.3	Diskrete Fourier-Transformation	185
6.2.4	Zweidimensionale Fourier-Transformation	187
6.2.5	Wichtige Fourier-Korrespondenzen	188
6.2.6	Faltung und Faltungstheorem	191
6.3	Elemente der Systemtheorie	195
6.3.1	Wechselwirkung von Signalen mit Systemen	195
6.3.2	Systemanalyse	198
6.3.3	Signalbegrenzung und Frequenzauflösung	199
6.3.4	Diskretisierung und Frequenzbereich	201
6.4	Digitales Filtern/Konvolution	204
6.4.1	Signalglättung	205
6.4.2	Signalkorrekturen	209
6.4.3	Gradientenverstärkung	210
6.4.4	Dekonvolution	213
6.5	Korrelation und Leistungsspektren	216
6.5.1	Autokorrelation und spektrale Leistungsdichte	216
6.5.2	Kreuzkorrelation und spektrale Kreuzleistungsdichte	218
6.5.3	Systemcharakterisierung durch Rauschen	219
6.6	Wavelets: ein neues Werkzeug für die Signalanalyse	220
6.6.1	Charakterisierung instationärer Signale	220
6.6.2	Was ist Wavelet-Analyse?	224

6.6.3	Diskrete Wavelet-Transformation (DWT)	227
6.6.4	Anwendungen	229
6.7	Datenreduktion und Interpolation	230
6.7.1	Datenreduktion	230
6.7.2	Bildsegmentierung	233
6.7.3	Interpolation	234
6.7.4	Interpolation mit Splinesfunktionen	236
6.8	Bilddarstellung	238
6.8.1	Bilder als Datensätze	238
6.8.2	Histogramme	239
6.8.3	Profile und Konturen	241
	Literatur	241
7	Kalibration	243
7.1	Analytische Informationen und Messergebnisse	243
7.2	Kalibration quantitativer Analysenverfahren	247
7.3	Experimentelle Kalibration	249
7.3.1	Verfahren der linearen Kalibration	251
7.3.2	Fehlergrößen der linearen Kalibration und der Auswertung	253
7.3.3	Validierung empirischer Kalibrationen	255
7.4	Kalibration durch Standardaddition	259
7.5	Dreidimensionale Kalibration	260
7.6	Nichtlineare Kalibration	263
7.7	Mehrkomponenten-Kalibration	266
7.8	Multivariate Kalibration	268
7.8.1	Klassische Kalibration	270
7.8.2	Inverse Kalibration	273
7.9	Kalibration mit Künstlichen Neuronalen Netzen	279
7.9.1	Künstliche Neuronale Netze	280
7.9.2	Mehrschichtige Perceptrons (MLP)	281
7.9.3	Radial Basis Function (RBF)-Netze	284
	Literatur	286
8	Auswertung analytischer Messungen	289
8.1	Auswertung qualitativer Analysen und Identifikationen	289
8.2	Auswertung quantitativer Zusammenhänge mit absoluten und definitiven Analysenverfahren	292
8.3	Auswertung relativer Analysenverfahren nach experimenteller Kalibration	294
8.4	Statistische Bewertung von Analyseergebnissen	295

8.4.1	Signifikanzprüfungen	295
8.4.2	Tests für Messreihen	297
8.4.3	Vergleich von Messreihen	298
8.4.4	Vergleich mehrerer Mittelwerte: Einfache Varianzanalyse . . .	300
8.4.5	Mehrfache Varianzanalyse	304
8.4.6	Vertrauensbereiche	307
8.4.7	Mess- und Ergebnisunsicherheit	311
8.5	Selektivität und Spezifität	313
8.6	Nachweis-, Erfassungs- und Bestimmungsgrenze	318
	Literatur	323
9	Klassifikation und Interpretation von Mess- und Analysendaten	324
9.1	Anwendungsgebiete	324
9.1.1	Anwendungen nichtüberwachter Methoden	326
9.1.2	Anwendung überwachter Methoden	326
9.2	Anwendungsbeispiele der multivariaten Datenanalyse	327
9.2.1	Lebensmittelanalytik	327
9.2.2	Weinanalytik	329
9.2.3	Chemische Industrie	332
9.2.4	Klinische und forensische Analytik	334
	Literatur	335
10	Spektrenauswertung	337
10.1	Bibliothekssuche und Spektrenbibliotheken	338
10.1.1	Prinzipien	338
10.1.2	Realisierung der Bibliothekssuche	339
10.1.3	Methodische Aspekte der Bibliothekssuche	341
10.2	Spektreninterpretation und Strukturermittlung	344
10.2.1	Spektren-Struktur-Relationen	344
10.2.2	Expertensysteme zur Strukturermittlung	346
10.2.3	Lineare Lernmaschine (Entwicklung von Diskriminanzfiltern)	348
10.2.4	Spektreninterpretation mit Neuronalen Netzen	350
10.3	Spektrenresolution	352
10.3.1	Spektrale Resolution einfacher Gemische	353
10.3.2	Bandentrennung und Spektrensimulation	356
10.4	Spektren-Eigenschafts-Korrelationen	360
	Literatur	362

11 Qualitätssicherung	365
11.1 Grundlagen und Prinzipien	365
11.2 Validierung von Analysenverfahren	368
11.3 Statistische Qualitätskontrolle	369
11.3.1 Statistische Qualitätskriterien	370
11.3.2 Attributprüfung	372
11.3.3 Sequenzanalyse	372
11.3.4 Qualitätsregelkarten	374
11.3.5 Ringversuche	379
11.4 Labor-Informationsmanagement-Systeme (LIMS)	383
11.5 Qualitätssicherung von Software	385
11.5.1 Definition und Aufgabenstellung von Software	385
11.5.2 Qualitätsmerkmale von Software	386
11.5.3 Qualitätssicherung bei der Softwareentwicklung	387
11.5.4 Softwarevalidierung	388
Literatur	389
Sachverzeichnis	393