Zeitschrift: Mitteilungen aus dem Gebiete der Lebensmitteluntersuchung und Hygiene = Travaux de chimie alimentaire et d'hygiène

Herausgeber: Bundesamt für Gesundheit

Band: 87 (1996)

Heft: 5

Artikel: Prognostischer Einsatz von multivariaten Techniken in der Weinanalytik = Prognostic use of multivariate techniques in wine analysis

Autor: Kaufmann, Anton

DOI: https://doi.org/10.5169/seals-982091

Nutzungsbedingungen

Die ETH-Bibliothek ist die Anbieterin der digitalisierten Zeitschriften auf E-Periodica. Sie besitzt keine Urheberrechte an den Zeitschriften und ist nicht verantwortlich für deren Inhalte. Die Rechte liegen in der Regel bei den Herausgebern beziehungsweise den externen Rechteinhabern. Das Veröffentlichen von Bildern in Print- und Online-Publikationen sowie auf Social Media-Kanälen oder Webseiten ist nur mit vorheriger Genehmigung der Rechteinhaber erlaubt. Mehr erfahren

Conditions d'utilisation

L'ETH Library est le fournisseur des revues numérisées. Elle ne détient aucun droit d'auteur sur les revues et n'est pas responsable de leur contenu. En règle générale, les droits sont détenus par les éditeurs ou les détenteurs de droits externes. La reproduction d'images dans des publications imprimées ou en ligne ainsi que sur des canaux de médias sociaux ou des sites web n'est autorisée qu'avec l'accord préalable des détenteurs des droits. En savoir plus

Terms of use

The ETH Library is the provider of the digitised journals. It does not own any copyrights to the journals and is not responsible for their content. The rights usually lie with the publishers or the external rights holders. Publishing images in print and online publications, as well as on social media channels or websites, is only permitted with the prior consent of the rights holders. Find out more

Download PDF: 02.07.2025

ETH-Bibliothek Zürich, E-Periodica, https://www.e-periodica.ch

Mitt. Gebiete Lebensm. Hyg. **87**, 604–622 (1996)
Eingegangen 15. Januar 1996. Angenommen 8. August 1996

Prognostischer Einsatz von multivariaten Techniken in der Weinanalytik

Prognostic Use of Multivariate Techniques in Wine Analysis

Key words: Wine, Multivariate statistic, Wine tasting, Chemometrics

Anton Kaufmann
Schuler Weine, Wein-Analytik, Seewen SZ

Einleitung

Die 4. Schweizer Meisterschaft im Weindegustieren ist eine Weindegustationsmeisterschaft, die von den Zeitschriften Vinum und NZZ im Zweijahresabstand organisiert wird. Der Wettkampf besteht aus drei Runden. Die Vorentscheidung wird «zu Hause» durchgeführt. Jeder Teilnehmer erhält ein Degustationspaket, bestehend aus 18 neutralen numerierten Probeflaschen. In der dazugehörigen Anleitung sind die 18 Weine detailliert beschrieben. Die Aufgabe des Teilnehmers besteht darin, die numerierten Fläschchen den in der Anleitung beschriebenen Weinen zuzuordnen. Aus den schriftlich eingegangenen Antworten werden die besten 10% der Teilnehmer für eine Zwischenrunde eingeladen. Maximal 50% der Zwischenrundenteilnehmer qualifizieren sich für die Endrunde, wo der Schweizer Meister erkoren wird. Im Gegensatz zu der Vorentscheidung finden die beiden folgenden Runden in einem öffentlichen Lokal statt.

Im Reglement der Schweizer Meisterschaft im Weindegustieren findet sich kein Absatz, der den Einsatz von technischen Hilfsmitteln reglementiert. Eine Bestreitung der Meisterschaft unter der ausschliesslichen Verwendung von Analysendaten und folgender statistischer Auswertung ist damit reglementskonform.

Verschiedene wissenschaftliche Arbeiten befassten sich mit der Bestimmung der geographischen Herkunft bzw. der Traubensorte von Weinen basierend auf analytisch chemischen Daten. Die bisher erfolgreichsten methodischen Ansätze lassen sich in vier verschiedene Gruppen einteilen.

1. Aromaanalytik. Insbesondere Terpene und gewisse heterocyklische Aromaten ermöglichen interessante Aussagen. Der Untersuchungsansatz basiert auf der Grundlage, dass jede Traubensorte entsprechend ihrem genetischen Aufbau unter-

schiedliche Stoffwechselprodukte bzw. stark unterschiedliche Mengen derselben produziert (1–6). Die Aromaanalytik ist methodisch aufwendig und erlaubt keinen hohen Probendurchsatz.
2. Spurenelemente. Dank dem Einsatz von ICP oder ICP-MS ist es möglich, praktisch das ganze Periodensystem analytisch abzudecken. Der methodische Ansatz basiert darauf, dass das geographische Vorkommen der einzelnen Elemente je nach Weinanbaugebiet sehr unterschiedlich ist (7–9). Die Interpretation wird aber stark erschwert, wenn Weinbehandlungsmittel, die Ionen abgeben, eingesetzt werden.
3. Isotopenanalyse. Das Deuterium-Wasserstoff-Verhältnis zeigt eine deutliche Abhängigkeit betreffend den geographischen Breitengrad. Isotope sind chemisch wohl identisch, der grosse Massenunterschied zwischen Wasserstoff und Deuterium hat aber einen messbaren Einfluss auf die Kinetik von enzymatisch gesteuerten Protonierungen. Neben Deuterium werden zunehmend auch Isotope von Sauerstoff, Kohlenstoff und schwereren Elementen gemessen (10–13). Die Methodik, besonders bei Einbezug schwererer Isotope, ist weitgehend fälschungsresistent. Die Aufarbeitung der Probe in Reinsträumen und der enorme apparative Aufwand verbieten aber vorläufig eine breitere Anwendung.
4. Routineanalysendaten. Routineanalysendaten, die bei der Weinuntersuchung anfallen, zeigen nur einen schwachen Zusammenhang mit der geographischen Herkunft bzw. Traubensorte. Mittels multivariater Statistik lassen sich aber durchaus ähnliche Fragestellungen beantworten, wie dies bei den oben erwähnten Methoden der Fall ist (14–16).

Aufgrund technischer Fortschritte sind die methodischen Ansätze 1–3 zwar «bezahlbarer» geworden. Eine geographische Zuordnung kann aber nur erfolgen, wenn eine gesicherte, breite Datenbank existiert. Gerade dieser Punkt ist leider selten erfüllt. Staatliche Institute, die die obigen Methodiken einsetzen, besitzen oft eine ihr Hoheitsgebiet abdeckende Datenbank. Weinanalysendaten ausserhalb dieses Gebietes sind aber meistens nur sehr spärlich vorhanden. Bei der 4. Schweizer Meisterschaft im Weindegustieren finden sich aber Weine wie ein Johannisberg aus dem Wallis, ein Cabernet Sauvignon aus Chile und ein Wein aus Ribera del Duero. Einer Teilnahme an einem solchen Wettbewerb ohne entsprechende Referenzmuster wäre kaum viel Erfolg beschieden. Primär aus diesem Grund ist die Zuordnung von Weinproben die Königsdomäne der Degustation geblieben. Unseres Wissens gab es bisher keinen erfolgreichen Einsatz von Analysendaten bei Degustationsmeisterschaften.

Als weltweit aktive Weinhandlung verfügen wir über eine Datenbank, die Weine aus praktisch jedem Weinanbaugebiet mit einer gewissen Bedeutung enthält. Vorversuche (16) zeigten, dass unsere Routineanalysendaten durchaus über ein gutes Diskriminationspotential verfügen. Es konnte dabei gezeigt werden, dass die Langzeitstabilität der Daten gesichert ist. Aus diesen Gründen wurde versucht, sich analytisch statistisch mit der Degustation zu messen. Es sei hier erwähnt, dass es sich dabei nicht um folgende triviale Fragestellung handelt: Analysiere 18 Flaschen mit Etiketten, dann 18 Fläschchen mit Nummern und ordne sie richtig zu. Es geht darum, neue Weine, die sich bisher noch nicht in der Datenbank befinden, richtig

zuzuordnen. Dies ist nur möglich, wenn die Datenbankweine die Informationen der Gesamtpopulation widerspiegeln.

Experimentelles

Probenmaterial

Unbekannte Weine

Die unbekannten Weine sind durch das Degustationsset gegeben. Es handelt sich um 18 Weine. Darunter befinden sich sechs Weissweine, drei Roséweine und neun Rotweine (Tabelle 1).

Um irgendwelche Beeinflussungen auszuschalten, wurde die Probenahme und Analyse nicht vom Autor durchgeführt. Die statistische Auswertung erfolgte so nur aufgrund der Datenbankeinträge. Sensorische Informationen wie der Geruch

Tabelle 1. Zuzuordnende Weine

Nr.	Weinname	Land	Region	Unterregion	Traube
			Weissweine		
1	«Filet d'Or» 1994 Mont-sur-Rolle	Schweiz	Waadt	La Côte	Chasselas
2	Fendant 1994 «Pentes Brûlées»	Schweiz	Wallis		Chasselas
3	Concannon 1992 Chardonnay Reserve	USA	Kalifornien	Livermore Valley	Chardonnay
4	Johannisberg 1994 de Chamoson	Schweiz	Wallis		Silvaner
5	Valencia 1994 Cavas Murviedro	Spanien	Valencia		Merseguera Moscatel
6	Ch. de Chantegrive Graves 1993	Frankreich	Bordeaux	Grave	Sémillion Sauvignon bl.
			Roséweine		
7	Rosé de Morges 1994 Domaine de Valmont	Schweiz	Waadt	La Côte	Gamay
8	Oeil de Perdrix du Valais 1994	Schweiz	Wallis		Pinot noir Gamay
9	Ch. d'Aigueville 1994	Frankreich	Côte du Rhône		Grenache Carignan

Nr.	Weinname	Land	Region	Unterregion	Traube
			Rotweine		
10	Dôle du Valais 1994	Schweiz	Wallis		Pinot noir Gamay
11	Nuits-Saint-Georges Barolet 1990	Frankreich	Burgund		Pinot noir
12	Châteauneuf-du-Pape 1993 Domaine de Beaurenard	Frankreich	Châteauneuf-du-Pape		Grenache Syrah usw.
13	Marqués del Puerto Rioja Crianza 1991	Spanien	Rioja		Tempranillo Granacha
14	Torreón de Paredes Cabernet Sauvignon 1990	Chile	Rengo		Cabernet Sauvignon
15	Ch. Yon-Figeac 1992 Saint-Emilion	Frankreich	Bordeaux	Saint-Emilion	Merlot Cabernet franc
16	Chianti Classico 1990 San Felice	Italien	Toscana	Chianti	Sangiovese Trebbiano
17	Barbera 1993 Marchesi Alfieri	Italien	Piemont		Barbera
18	Monte Vannos Crianza Ribera del Duero 1991	Spanien	Ribera del Duero		Tempranillo Merlot

beim Abfüllen der Probe oder die Farbe des Weines konnten damit nicht als Vorinformationen die statistische Auswertung beeinflussen.

Datenbankweine

Mehr als 6000 Weine aller Provenienzen.
Analysenmethode IC, LC und AAS siehe (16). Die Langzeitstabilität wird mittels Prüfmittelkontrolle (ISO 9002) sichergestellt.

Analysenparameter

pH
Trehalose, Glucose, Fructose, Glycerin, Alkohol
Essigsäure, Milchsäure, Chlorid, Galacturonsäure, Bernsteinsäure, Apfelsäure, Weinsäure, Sulfat, Phosphat, Citronensäure
Blei, Cadmium

Statistische Verfahren

Die Auswertung erfolgte auf einem 486 PC mittels Statistika 5.0. Es wurden ausschliesslich multivariate statistische Verfahren zur Auswertung herangezogen.

Zum Einsatz gelangten lineare Diskriminanzanalyse bzw. kanonische Korrelation im Mehrgruppenfall. Einzelne Problemstellungen wurden mittels nichtlinearer Diskriminanzanalyse betrachtet. Ebenfalls verwendet wurde die Hauptkomponentenanalyse mit nachfolgender linearer Diskriminanzanalyse.

Vorgehen bei der Auswertung

Die Weine wurden aufgrund der Farbe aufgeteilt in die drei Gruppen «Rot, Rosé und Weiss». Die Zuordnung erfolgte innerhalb dieser Gruppen. Da die geographische Herkunft gegeben ist, ist den unbekannten Weinen die aufgrund der Analysendaten wahrscheinlichste Herkunft zuzuordnen. Die Zuordnung erfolgt aufgrund zweier statistischer Kriterien:

Die Wahrscheinlichkeit, dass der fragliche Wein zur getesteten geographischen Herkunft gehört.

Die Mahalanobisdistanz von den entsprechenden Gruppenzentroiden.

Die Mahalanobisdistanz ist Entfernung eines fraglichen Datenpunktes vom Punkteschwarmmittel. Es handelt sich dabei nicht um eine euklidische Entfernung, sondern um ein Distanzmass, welches die Korrelation der Variablen untereinander berücksichtigt. Die Mahalanobisdistanz ist nicht auf den zwei- bzw. dreidimensionalen Raum beschränkt. Sie ist das geeignete Entfernungsmass, um multivariate Distanzen in höherdimensionalen Räumen zu beschreiben.

Auswertung der Daten

Roséweine

Da lediglich die Zuordnung von drei Weinen gefragt ist, handelt es sich wahrscheinlichkeitstheoretisch betrachtet um die einfachste der drei Fragestellungen. Es sind nur sechs verschiedene Antworten möglich. Folglich beträgt die Wahrscheinlichkeit eines Zufallschätzers 16%. Betreffend des Datenbankmaterials ist aber die Roséproblemstellung mit gewissen Schwierigkeiten behaftet. Die beiden Schweizer Rosés sind bezüglich der Herkunft und Traubensorte sehr ähnlich. Für alle Rosés insbesondere aber für den Côtes du Rhône waren nur sehr wenige Referenzmuster vorhanden. Es war deshalb notwendig, weisse und rote Muster der gleichen geographischen Herkunft als Referenzmuster zu verwenden. Dem Vergleich mit Rotweinen sollte ein höheres Gewicht beigemessen werden als den Weissweinen. Die Traubensorten für Rot- und Roséweine sind innerhalb eines Anbaugebietes meist identisch. Dies gilt aber nicht für die Weissweine. In einzelnen Parametern wie Galacturonsäure und Alkohol sind aber Rosés eher mit Weissweinen gleichzustellen. Es stellt sich die Frage, ob es angebrachter ist, eine Diskriminanzanalyse mittels den wenigen Roséweinen oder zusätzlich mit den Rot- und Weissweinen durchzuführen. Das Dreigruppenproblem lässt sich graphisch mit zwei Diskriminanzfaktoren «Roots» darstellen (siehe Abb. 1 und 2).

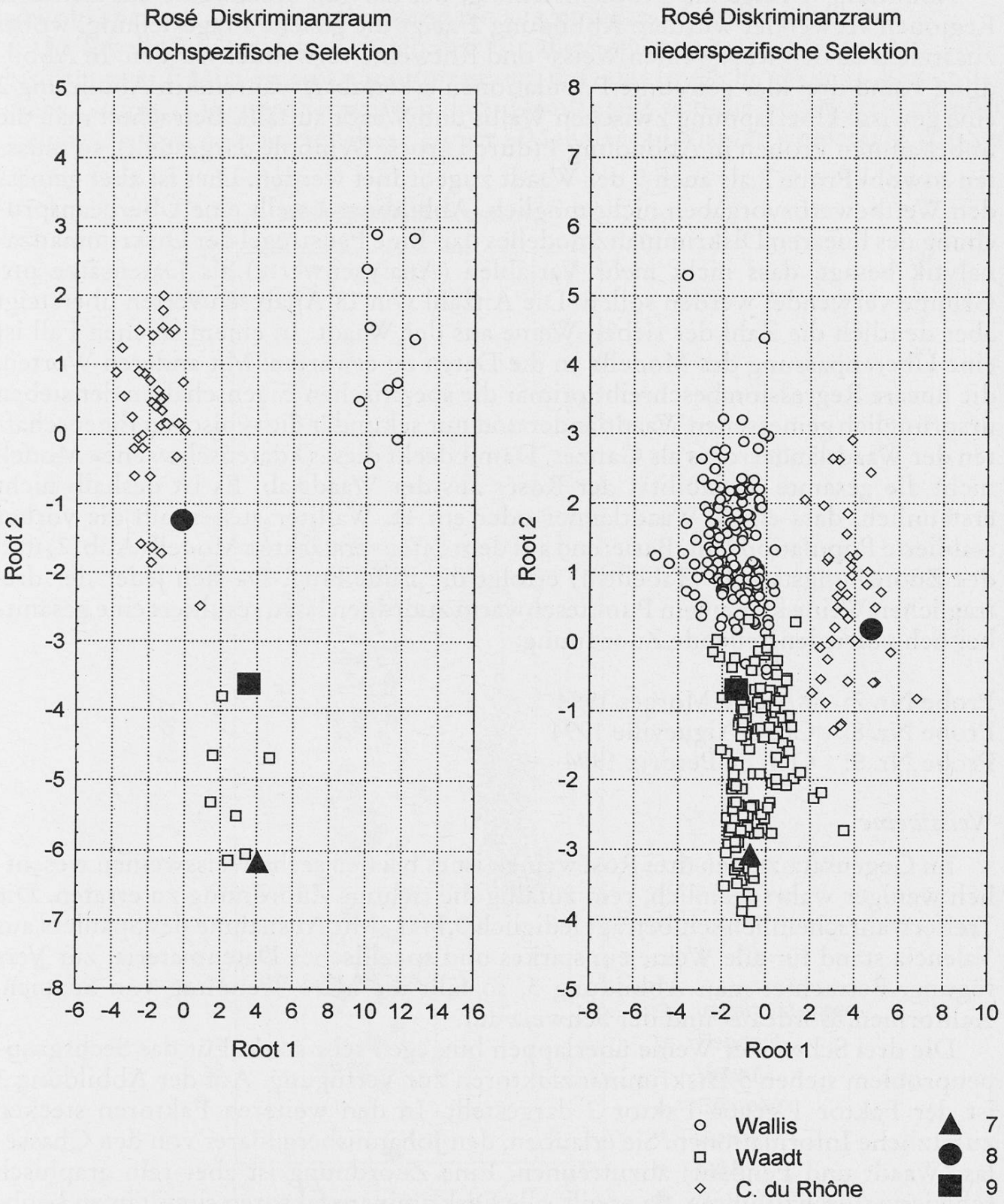

Abb. 1 und 2. Roséwein Diskriminanzraum. Die Zuordnung der drei Roséweine erfolgt mittels Datenbankweinen, welche betreffend Traubensorte und Herkunft möglichst identisch sind. Eine solche *hochspezifische Selektion* (Abb. 1) der Datenbankeinträge verkleinert die Anzahl der verfügbaren Referenzweine gegenüber einer niederspezifischen Selektion (Abb. 2).

Abbildung 1 zeigt die Problemstellung, bei der nur Roséweine aus den drei Regionen verwendet wurden. Abbildung 2 zeigt die gleiche Fragestellung, wobei zusätzlich zu den Roséweinen Weiss- und Rotweine verwendet wurden. In Abbildung 1 sind drei klar getrennte Populationen erkennbar, während in Abbildung 2 eine gewisse Überlappung zwischen Wallis und Waadt auffällt. Betrachtet man die unbekannten Proben in Abbildung 1 (durch grosse Symbole dargestellt), so müssten sowohl Probe 7 als auch 9 der Waadt zugeordnet werden. Dies ist aber gemäss den Wettbewerbsvorgaben nicht möglich. Abbildung 1 stellt eine Überbeanspruchung des linearen Diskriminanzmodelles dar. Eine Faustregel der Diskriminanzanalytik besagt, dass nicht mehr Variablen (Analysenwerte) als Datensätze pro Gruppe verwendet werden sollen. Die Anzahl von 18 Analysenwerten übersteigt aber deutlich die Zahl der sieben Weine aus der Waadt. In einem solchen Fall ist eine Überanpassung des Modells an die Daten zu erwarten. Mit anderen Worten, die lineare Regression beschreibt primär die spezifischen Eigenschaften der sieben ursprünglich gemessenen Waadtländer und nur sekundär die typischen Eigenschaften der Waadtländerrosés als Ganzes. Damit deckt dieses «datenschwache» Modell nicht die gesamte Variabilität der Rosés aus der Waadt ab. Es ist deshalb nicht erstaunlich, dass ein 8. Waadtländer oder ein 11. Walliser ausserhalb die vorher etablierte Population fällt. Basierend auf dem datenverstärkten Modell (Abb. 2) und der Zuordnungstabelle (Tabelle 1) erfolgt die Einteilung. Da sich jeder der drei fraglichen Weine klar einem Punkteschwarm zuordnen lässt, resultiert eine gesamtheitlich zufriedenstellende Zuordnung:

Probe Nr. 7: Rosé de Morges 1994
Probe Nr. 8: Ch. d'Aigueville 1994
Probe Nr. 9: Oeil de Perdrix 1994

Weissweine

Im Gegensatz zu den drei Roséweinen ist es bei den sechs Weissweinen wesentlich weniger wahrscheinlich, rein zufällig die richtige Zuordnung zu erraten. Die Trefferwahrscheinlichkeit beträgt lediglich 0,14%. Mit Ausnahme des Spaniers aus Valencia stand für alle Weine ein starkes und spezifisches Datenmaterial zur Verfügung. Betrachtet man Abbildung 3, so fällt die klare Trennung von Spanien, Kalifornien, Bordeaux und der Schweiz auf.

Die drei Schweizer Weine überlappen hingegen sehr stark. Für das Sechsgruppenproblem stehen 5 Diskriminanzfaktoren zur Verfügung. Auf der Abbildung 3 ist der Faktor 1 gegen Faktor 2 dargestellt. In den weiteren Faktoren stecken zusätzliche Informationen. Sie erlauben, den Johannisberg klarer von den Chasselas (Waadt und Fendant) abzutrennen. Eine Zuordnung ist aber rein graphisch schwierig durchzuführen, da es gilt, alle Diskriminanzfaktoren simultan zu beobachten. Es ist angebrachter, aufgrund der Mahalanobisdistanzen eine Einteilung vorzunehmen. Je kleiner die Mahalanobisdistanz (das multivariate Entfernungsmass), desto ähnlicher ist der fragliche Wein der zu vergleichenden Datenbankpopulation. In Tabelle 2 sind die Mahalanobisdistanzen von den 6 Gruppenzentroiden

aufgeführt. Eine überzeugende Zuordnung ist zu verzeichnen, wenn eine Zelle sowohl spalten- als auch zellenweise ein ausgeprägtes Minimum aufweist.

In Tabelle 2 fällt Probe 1 durch ihre für alle Regionen extrem hohe Mahalanobisdistanz auf. Dies ist so zu interpretieren, dass der fragliche Wein zu keiner der sechs Gruppen zugeordnet werden darf. Damit stellt sich die Frage, welcher oder welche Analysenparameter diese enorme Mahalanobisdistanz verursacht. Bisher

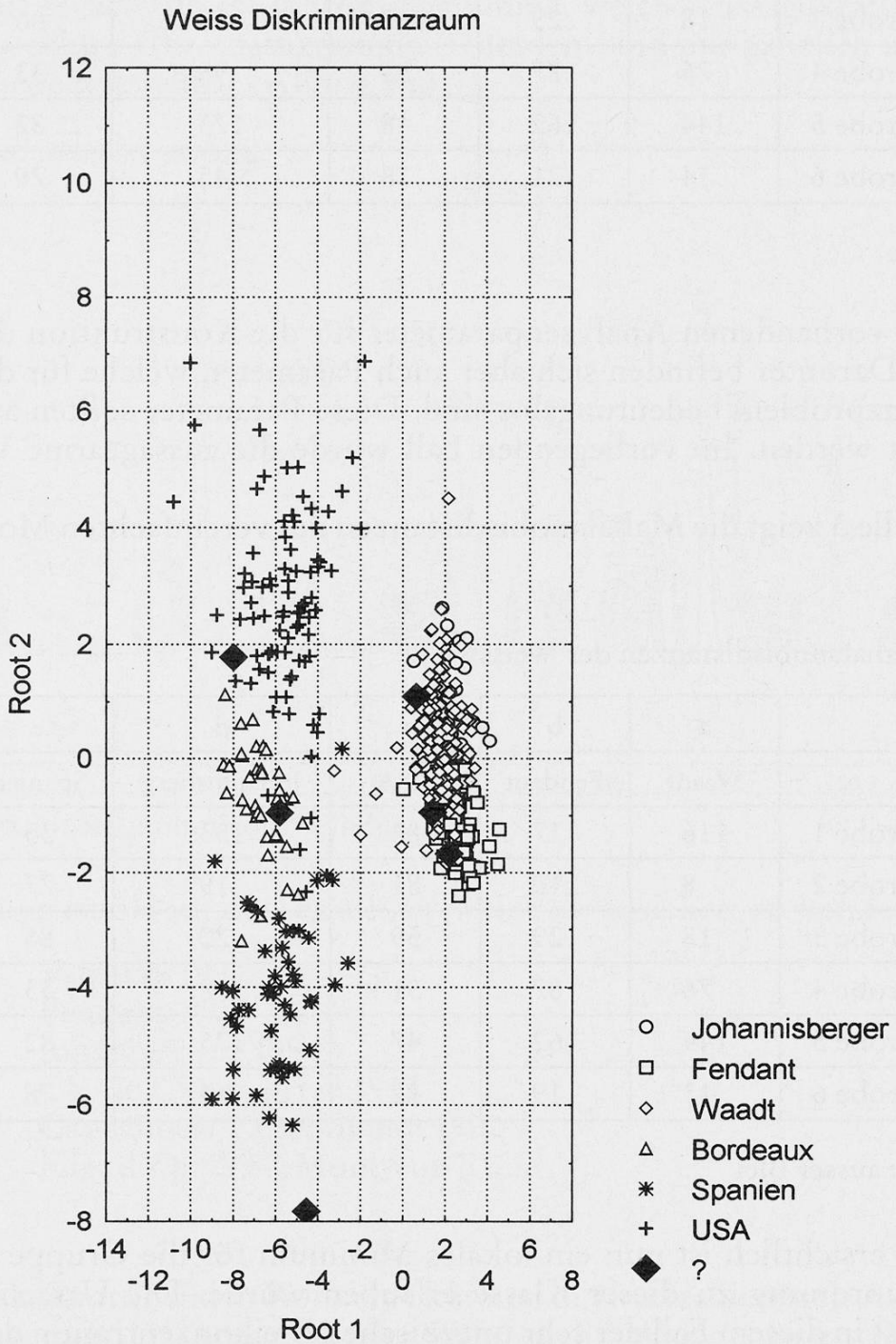

Abb. 3. Weisswein Diskriminanzraum. Deutlich ersichtlich ist die Trennung bezüglich Herkunft der Weissweine. Die Schweizer Weine sind hingegen nur schwach angetrennt (Fendant und Waadt gleiche Traubensorte).

Tabelle 2. Mahalanobisdistanzen der Weissweine

Code		a	b	c	d	e	f
Herkunft		Waadt	Fendant	USA	Johannisberg	Spanien	Bordeaux
	Probe 1	365	371	383	409	283	317
	Probe 2	10	12	89	21	80	88
	Probe 3	18	29	60	20	86	77
	Probe 4	76	87	33	98	33	15
	Probe 5	144	162	48	173	82	78
	Probe 6	14	21	78	45	79	90

Variabeln: alle

wurden alle vorhandenen Analysenparameter für die Konstruktion des Modelles verwendet. Darunter befinden sich aber auch Parameter, welche für das gegebene Diskriminanzproblem bedeutungslos sind. Diese Parameter sollten aus dem Modell entfernt werden. Im vorliegenden Fall wurde die aussagearme Variabel Blei entfernt.

Die Tabelle 3 zeigt die Mahalanobisdistanzen des vereinfachten Modelles (ohne Blei).

Tabelle 3. Mahalanobisdistanzen der Weissweine

Code		a	b	c	d	e	f
Herkunft		Waadt	Fendant	USA	Johannisberg	Spanien	Bordeaux
	Probe 1	116	117	86	153	33	68
	Probe 2	8	10	81	19	77	86
	Probe 3	18	29	59	20	86	77
	Probe 4	76	87	31	98	33	15
	Probe 5	144	162	47	173	82	78
	Probe 6	13	19	72	44	78	89

Variabeln: alle ausser Blei

Deutlich ersichtlich ist nun ein lokales Minimum für die Gruppe e (Spanien), was eine Zuordnung zu dieser Klasse erlauben würde. Die Ursache für dieses Phänomen ist in diesem Fall der sehr untypischen Bleikonzentration dieses Weines zuzuschreiben.

Der geogene Anteil von Blei in Wein ist gering. Eine weitaus grössere Menge stammt aus Messingarmaturen älterer Kellereinrichtungen. Deshalb ist die Verwen-

dung von Blei als Herkunftsparameter wenig sinnvoll oder, wie in diesem Fall, sogar gefährlich.

Wie zu erwarten war, stellt die Zuordnung der beiden Chasselas-Weine aus der Waadt und dem Wallis das grösste Problem dar. Die zur Verfügung stehenden Analysenwerte haben ein ungenügendes Diskriminanzpotential für diese Fragestellung. Dies ist eindrücklich ersichtlich, wenn man die Chromatogramme der beiden fraglichen Weine betrachtet. Bei den Chromatogrammen von Wein Nr. 2 und 6 handelt es sich um IC-Chromatogramme, welche aussagekräftiger sind als die Ionenausschluss-Chromatogramme. Trotzdem sind nur gänzlich unbedeutende Unterschiede feststellbar (Abb. 4 und 5).

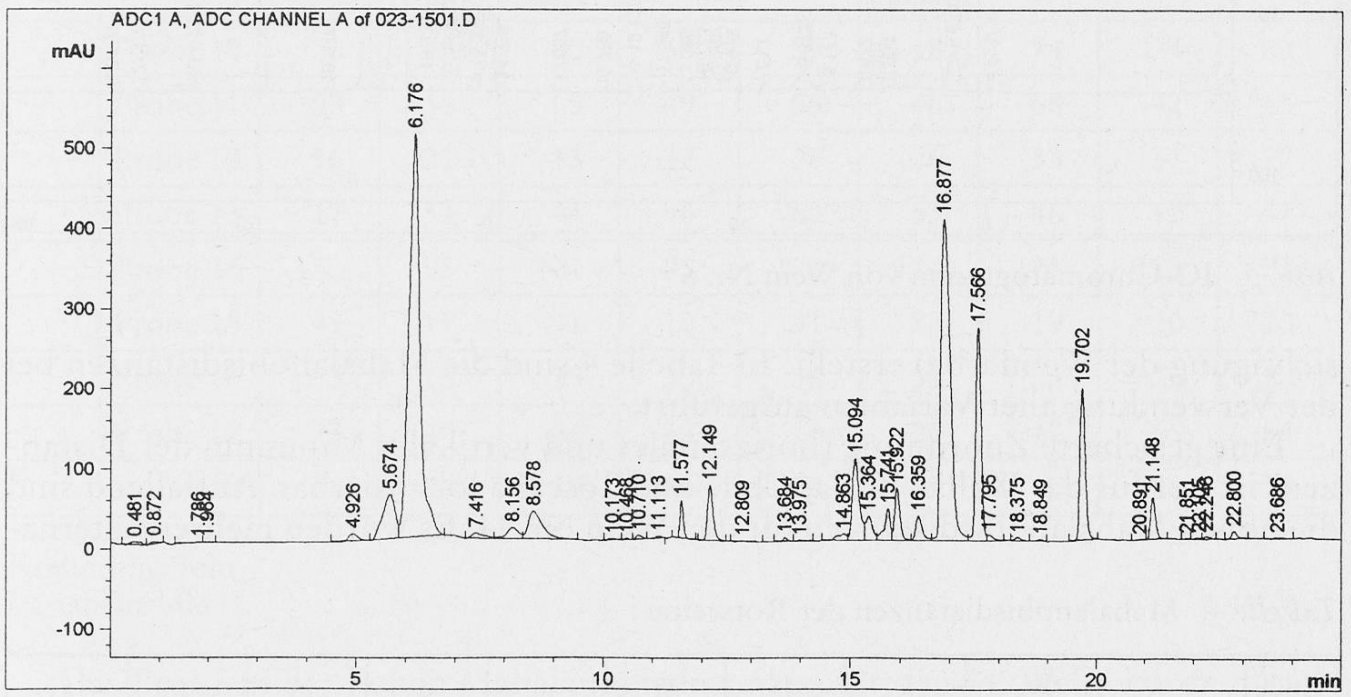

Abb. 4. IC-Chromatogramm von Wein Nr. 2

Die Zuordnung ist bis auf die Schweizer Weine als gesichert zu betrachten:

Probe Nr. 1: Valencia 1994
Probe Nr. 2: Fendant 1994
Probe Nr. 3: Johannisberg 1994
Probe Nr. 4: Ch. de Chantegrive 1993
Probe Nr. 5: Cancannon Chardonnay 1992
Probe Nr. 6: Filet d'Or 1994, Mont-sur-Rolle

Rotweine

Die neun zuzuordnenden Rotweine können in 9! verschiedenen Anordnungen zugeordnet werden. Das entspricht der Wahrscheinlichkeit einer zufällig richtigen Anordnung von lediglich 0,0003%. Für alle Regionen mit Ausnahme von Ribera del Duero waren genügend Datenbankeinträge vorhanden. Als erster Schritt wurde die Zuordnung ausschliesslich aufgrund geographischer Kriterien (ohne Berück-

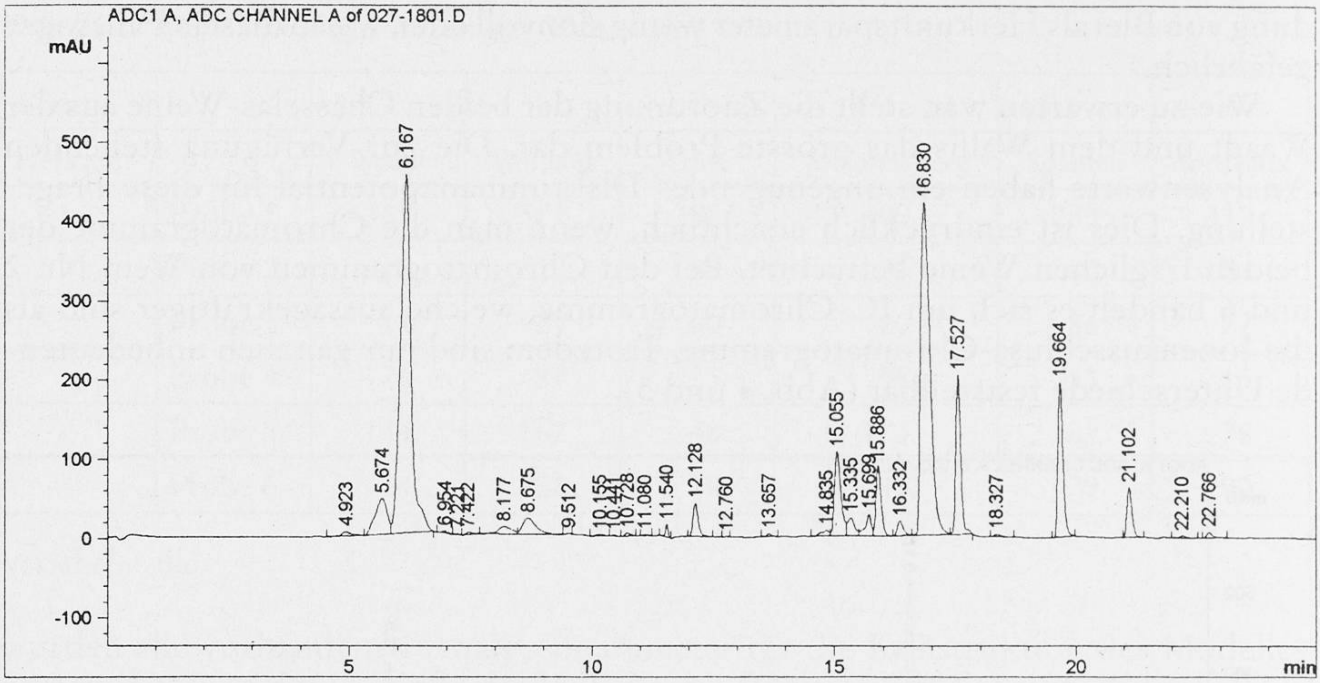

Abb. 5. IC-Chromatogramm von Wein Nr. 6

sichtigung der Weinfarbe) erstellt. In Tabelle 4 sind die Mahalanobisdistanzen bei der Verwendung aller Variablen aufgeführt.

Eine gesicherte Zuordnung (horizontales und vertikales Minimum der Distanzen) ist nur für das Wallis, Rioja, Chile und Toscana lokalisierbar. Auffallend sind die hohen Mahalanobisdistanzen für den Wein Nr. 14. Es wurden mehrere alterna-

Tabelle 4. Mahalanobisdistanzen der Rotweine

Code		j	k	l	m	n	o	p	q	r
Herkunft		Wallis	Burgund	Côte du Rhone	Rioja	Chile	Bordeaux	Toscana	Piemont	Spanien
	Probe 10	25	34	44	44	40	28	56	48	48
	Probe 11	73	38	38	39	41	35	46	31	33
	Probe 12	49	22	24	36	23	25	32	21	25
	Probe 13	16	27	32	40	46	39	36	28	50
	Probe 14	156	106	114	97	123	115	130	123	89
	Probe 15	43	15	22	8	24	20	17	12	15
	Probe 16	45	12	24	24	26	24	20	17	28
	Probe 17	56	33	33	37	14	19	32	34	36
	Probe 18	52	13	20	23	23	20	10	14	23

Kodierung: grob
Variabeln: alle

tive Betrachtungen gewählt. So wurde z. B. die Codierung verfeinert. Für Chile wurden nur chilenische Cabernet Sauvignons berücksichtigt. Im Piemont gelangten lediglich Barberaweine in die Berechnung, Côte du Rhone wurde durch Châteauneuf-du-Pape enger eingekreist (Tabelle 5).

Tabelle 5. Mahalanobisdistanzen der Rotweine

Code		j	k	l	m	n	o	p	q	r
Herkunft		Wallis Dôle	Burgund P. Noir	Châteaun. du Pape	Rioja	Chile Cab. Sauv.	Bordeaux Saint-Emilion	Toscana Chianti	Piemont Barbera	Spanien Nordwest
	Probe 10	26	45	70	56	59	48	74	74	62
	Probe 11	98	56	65	49	56	46	68	42	45
	Probe 12	56	21	33	37	28	26	33	31	27
	Probe 13	17	34	44	46	62	53	46	42	57
	Probe 14	150	95	101	88	120	114	125	113	77
	Probe 15	49	17	41	10	31	28	19	20	20
	Probe 16	55	15	39	30	36	30	23	25	36
	Probe 17	73	45	66	48	14	25	42	66	54
	Probe 18	58	14	27	26	28	23	10	26	29

Kodierung: fein
Variabeln: alle

Das Problem der hohen Mahalanobisdistanzen für den 14. Wein wurde mittels Variabelnentfernung (Rückwärtselimination) angegangen (siehe Tabelle 6). Analog dem spanischen Weisswein wurden der diskriminanzanalytisch aussageschwächste Parameter aus dem Modell entfernt. Dieser Vorgang wird mehrfach wiederholt. Der Abbruch der Rückwärtselimination erfolgt vor dem Entstehen einer signifikante Modellverschlechterung.

Neben den reduzierten Mahalanobisdistanzen ist nun der Burgunderwein (Probe Nr. 16) in Tabelle 6 sowohl horizontal als auch vertikal klar zuordnungsbar. Die Mahalanobisdistanzen der vier verbleibenden Weine (Nr. 10, 11, 12 und 14) zeigen aber nach wie vor widersprüchliche Informationen. Trotz verschiedenster Kodierungen, Variabeleliminationen usw. konnte dieses Problem nicht zur gänzlichen Zufriedenheit gelöst werden. Die Restzuordnung erfolgt somit weniger aufgrund der Wahrscheinlichkeiten als der Minimierung der Unwahrscheinlichkeiten.

In Tabelle 7 sind die vier verbliebenen «Problemweine» repräsentiert. Bisher wurde mittels linearer Diskriminanzanalytik gearbeitet. Nichtlineare Zusammenhänge können damit nur unzufriedenstellend modelliert werden. Deshalb wurde für jede Variabel nun zusätzlich deren logarithmierter Wert mit in die Berechnung einbezogen. Aufgrund der damit verdoppelten Variablenanzahl müssen genügend

Tabelle 6. Mahalanobisdistanzen der Rotweine

Code		j	k	l	m	n	o	p	q	r
Herkunft		Wallis Dôle	Burgund P. Noir	Châteaun. du Pape	Rioja	Chile Cab. Sauv.	Bordeaux Saint-Emilion	Toscana Chianti	Piemont Barbera	Spanien Nordwest
	Probe 10	14	36	60	49	54	42	65	58	56
	Probe 11	84	44	51	39	43	36	51	27	35
	Probe 12	47	14	25	25	18	12	25	25	20
	Probe 13	8	21	35	28	49	33	38	29	39
	Probe 14	100	44	41	34	63	59	60	53	23
	Probe 15	45	15	39	6	28	22	18	17	17
	Probe 16	46	9	30	19	25	14	14	21	26
	Probe 17	65	39	62	43	11	21	40	58	48
	Probe 18	56	12	25	22	23	19	8	24	23

Kodierung: fein
Variabeln: Rückwärtselimination (11 verbleibende Variabeln)

Tabelle 7. Mahalanobisdistanzen der verbleibenden «Problemfälle»

Code		l	o	q	r
Herkunft		Côte du Rhone	Bordeaux	Piemont	Spanien
	Probe 10	30,14	17,76	40,25	30,67
	Probe 11	24,36	25,19	15,19	11,93
	Probe 12	13,23	18,71	16,65	16,45
	Probe 14	48,88	44,69	60,14	23,04

Kodierung: grob
Variabeln: alle sowie log, Rückwärtselimination

Datensätze zur Verfügung stehen, um eine Überanpassung an die vorgegebenen Gruppen zu verhindern. Aus diesem Grund wurde eine wenig spezifische Gruppencodierung gewählt (z. B. Bordeaux anstelle Saint-Emilion). Das Viergruppenproblem wurde zusätzlich mittels Rückwärtselimination angegangen, um nun aufgrund der kleineren Gruppenzahl mehr an Störvariablen entfernen zu können. Es wurden 13 Variablen im endgültigen Modell belassen.

Die Zuordnung ist damit gesamthaft für fünf Weine gesichert, für zwei (Bordeaux und Côte du Rhône) zufriedenstellend und für zwei (Ribera del Duero und Barbera) zweifelhaft.

Probe Nr. 10: Ch. Yon-Figeac 1992
Probe Nr. 11: Barbera 1993
Probe Nr. 12: Châteauneuf-du-Pape 1993
Probe Nr. 13: Dôle du Valais 1994
Probe Nr. 14: Monte Vannos 1991
Probe Nr. 15: Marqués del Puerto 1991
Probe Nr. 16: Nuits-Saint-Georges 1990
Probe Nr. 17: Torréon de Paredes 1990
Probe Nr. 18: Chianti Classico

Diskussion

Diese Arbeit inklusive der Diskussion wurde vor Auflösung der Vorrunde zur Publikation eingereicht. Die endgültige Beurteilung der Zuordnungsgüte wird deshalb im Nachtrag behandelt.

Wie aus den aufgeführten Daten ersichtlich ist, konnten viele Weine statistisch eindeutig zugeordnet werden. Bei ein bis zwei Problempaaren ist aber «wissenschaftliches Roulett» der treffendere Ausdruck. Es sei hier noch einmal anzumerken, dass diese Zuordnung lediglich aufgrund von Routineanalysendaten erfolgte, Daten, denen man solches Diskriminationspotential nicht unbedingt zusprechen würde. Zweifelsohne würden zusätzliche Analysenparameter die obigen Fragestellungen noch deutlicher beantworten können.

Wie eingangs erwähnt, ist die Referenzdatenbank von grösster Bedeutung. Je höher die Anzahl der Datensätze aus den verschiedenen Anbaugebieten mit ihren verschiedenen Traubensorten ist, desto besser wird die resultierende Zuordnung ausfallen. Es wurde in diesem Zusammenhang aber mehr als einmal die Erfahrung gemacht, dass es erfolgversprechender ist, viele mittelspezifische Datensätze zu verwenden als wenige hochspezifische. Als Beispiel möge hier die Roséproblemstellung mit den wenigen Datensätzen dienen.

Es ist ebenfalls wichtig, den Einsatz der zur Verwendung stehenden Variablen gründlich zu überdenken. Nicht die grösstmögliche Anzahl an Variablen produziert die aussagekräftigsten Zuordnungen. Es müssen diejenigen Variablen sein, die relevant sind für die Problemstellung. Variablen, die für die gegebene Problemstellung lediglich ein Rauschen darstellen, sind zu eliminieren. Statistikpakete erlauben den Einsatz der Rückwärtselimination, die vorzuziehende Eliminationstechnik. Die Elimination im Mehrgruppenfall ist aber nicht ganz problemlos. Der Analysenwert a mag für sieben von neun Gruppentrennungen lediglich Rauschen darstellen, er kann hingegen für die Trennung der zwei Restgruppen hoch signifikant sein. In diesem Fall sollten die fraglichen Restgruppen erneut mittels Rückwärtselimination diskriminiert werden.

Die Datenanalyse wurde primär mit linearer Diskriminanzanalyse durchgeführt. Versuche mittels Hauptkomponentenanalyse bzw. Hauptkomponentenanalyse und nachfolgender Diskriminanzanalyse lieferten weniger überzeugende Resultate. Die Hauptkomponentenanalyse produzierte wohl Cluster, diese standen aber nur wenig im Zusammenhang mit der geographischen Fragestellung. Eine nachfolgende lineare Diskriminanzanalyse produzierte die gewünschten Resultate. Die resultierende Trennung war aber der einfachen Diskriminanzanalyse unterlegen. Basierend auf der Theorie wäre eher der umgekehrte Fall zu erwarten. Die einfache Diskriminanzanalytik, die auf der linearen Regression beruht, sieht einen Fehlerterm für die Y Variabel, die Gruppenzuordnung vor. Die X Ausgangsvariablen werden aber als fehlerfrei angenommen. Dies ist sicherlich nicht korrekt, wenn man an die inhärenten Messfehler der Analytik denkt.

Die Hauptkomponentenanalyse hingegen ermöglicht den Bau von Modellen, die die Fehlervarianz der Analysenwerte eliminieren können. Zudem haben starke Korrelationen der Analysenwerte untereinander keinen negativen Einfluss. Hohe Korrelationen dieser Werte können bei der linearen Diskriminanzanalyse hingegen instabile Modelle produzieren. Die Regressionsebenen im multidimensionalen Raum können dann aufgrund einzelner Daten «kippen». Der Hauptunterschied zwischen der linearen Diskriminanzanalyse und der Hauptkomponentenanalyse liegt jedoch darin, dass die erstere eine zielgerichtete geführte Methode ist. Die Hauptkomponentenanalyse hingegen deckt objektiv vorhandene Gruppen auf. Es wäre unrealistisch anzunehmen, dass eine vom Benützer ungeführte Technik wie die Hauptkomponentenanalyse automatisch ohne irgendwelche Vorgaben die gewünschte Gruppentrennung produziert. Neben der geographischen Trennung wäre eine Trennung nach Traubensorte, Weinalter usw. ebenso logisch. Die lineare Diskriminanzanalyse hingegen «zwingt» die Daten in die vorgegebene gewünschte Klassifizierung. Es ist dabei aber zu beachten, dass nicht eine Überanpassung erfolgt. Solche Überanpassungen wurden bei der Verwendung von iterativen nichtlinearen Regressionsverfahren wiederholt beobachtet.

Mit diesen Techniken konnten Modelle gebildet werden, die sich nicht nur an der Problemstellung, sondern zugleich auch den Messfehlern anpassten. Leider können diese Modelle nicht zwischen den beiden Effekten unterscheiden. So erzeugte die logistische Regression mittels einer «maximum likelihood loss function» und dem quasi-Newton Minimalisierungs-Algorithmus hervorragende Gruppentrennungen. Dies war selbst bei den oben genannten Problempaaren zu beobachten. Das Chasselas-Problem wurde aber so gelöst, dass beide Weine mit einer sehr hohen Zuordnungsgüte dem Wallis zugeordnet wurden. Da die logistische Regression nur die Zustände 1 und 0 kennt, werden auch Ausreisser in diese «Zwangsjacke» gesteckt. Eine Mahalanobisdistanz, die in der Diskriminanzanalyse fremde Muster erkenntlich macht, existiert in der logistischen Regression nicht. Es muss bei diesen Problemzuordnungen davon ausgegangen werden, dass aus den gegebenen Analysendaten nicht mehr an Information herausextrahiert werden kann. Weitere Fragen können wahrscheinlich nur aus zusätzlichen Analysedaten erhalten werden, nicht aber bei der Verwendung von noch komplexeren mathematischen Modellen. Im Zeitalter der Hochleistungs-PC's ist die Versuchung aber

gross, komplexere statistische Prozeduren anstelle einer anspruchsvolleren Analytik einzusetzen.

Schlussfolgerung

Dieser Vergleich «Maschine gegen Mensch» soll nicht zu Interpretationen führen, dass die Degustation dank Analytik und multivariater Statistik bald überflüssig wird. Das Erkennen der Weinherkunft ist nicht die eigentliche Aufgabe der Sensorik. Vielmehr wurde der Degustation gesellschaftlich diese Rolle zugeschoben oder aufgedrängt. Die Aufgabe der Sensorik sollte das Erkennen von Typizitäten und Fehlern und nicht das Bestimmen von geographischen Breiten- und Längengraden sein. Auf diesem Gebiet muss die Degustation nicht vor der analytischen Konkurrenz bangen. Einen sensorischen Fehlton stellt das oben vorgestellte Verfahren nicht fest. Das mag wohl eine Stärke der statistischen Analytik bei der Herkunftsbestimmung darstellen, es ist aber ihre Schwäche, wenn es um konsumentenbezogene hedonistische Fragestellungen geht. Es mag Bedenken auslösen, wenn eine «Maschine» eine eigentlich urmenschliche Fähigkeit imitieren kann. Dem sei aber angemerkt, dass die Weinherkunftsbestimmungsmaschine damit nicht erfunden wurde. Ein erfolgreiches Arbeiten erfordert eine solide Analytik, Vertrautheit mit den statistischen Techniken und «last but not least» fundierte Kenntnisse auf dem Gebiet des Rebbaus, der Vinifikation und der Weingeographie. Gute Sinnesorgane mögen für die Weinerkennung mittels Degustation wichtiger sein als ein fundiertes Weinwissen. Für den analytisch statistischen Ansatz ist aber dieses Weinwissen essentiell. Wohl lassen sich gewisse Aufgaben EDV-seitig automatisieren. Die angepasste Kodierung der Datensätze erfolgt aber durch den Anwender.

Er muss entscheiden, ob die zur Verfügung stehenden Datensätze ausreichen. Sollte das nicht der Fall sein, so müssen diese Sätze durch möglichst artgleiche aus anderen Regionen «verstärkt» werden. Diese zusätzlichen Sätze sind primär aufgrund der Traubensorte und Vinifikationstechnik auszulesen. Daneben muss der Analytiker mit Weinanalysendaten vertraut sein, um Extremwerte erkennen und diese aus der Auswertung eliminieren zu können.

Die multivariate Auswertung von Analysendaten ist ein hocheffizientes Fehleraufdeckungsinstrument. Mit ihr wurden mehr als einmal in unserem Labor unkorrekte Angaben bzw. Analysendaten in der Datenbank entdeckt. Diese Datensätze fallen wohl univariat nicht auf, multivariat produzieren sie aber eine enorme Mahalanobisdistanz. In diesem Sinne dürfte diese Technik die überlegendste Form der Qualitätssicherung von Analysendaten darstellen.

Auch wenn die Volksmeinung die Häufigkeit von Panschereien und Falschdeklarationen zweifelsohne weit überschätzt, braucht es doch Instrumente, um solche Praktiken zum Verschwinden zu bringen. Die multivariate Auswertung von Routineanalysendaten dürfte aufgrund ihres Aufwand-Ertrag-Verhältnisses und ihrer hohen Fälschungsresistenz das geeignete Instrument darstellen.

Nachtrag

Die Auflösung der Resultate der 4. Schweizerischen Meisterschaft im Weindegustieren ergab die maximal erreichbare Punktzahl. Alle 18 Weine konnten aufgrund der Analysendaten und der statistischen Auswertung korrekt zugeordnet werden. Damit wurde, soweit uns Informationen vorliegen, erstmalig ein offizieller Degustationswettkampf mittels eines nichtsensorischen Ansatzes erfolgreich bestritten. Aufgrund der fehlerfreien Zuordnung resultiert die Qualifikation für die nächste Ausscheidungsrunde. In dieser zweiten Runde ist die Degustationszeit limitiert. Die zur Verfügung stehende Zeit reicht damit nicht für die Analyse und folgende statistische Auswertung. Da der Anlass zusammen mit anderen Kandidaten in einem öffentlichen Lokal stattfindet, ist eine Teilnahme mittels Analysengeräten und EDV nicht angebracht.

Wie bereits eingangs erwähnt, soll dieses für den ersten Anlauf doch nicht selbstverständliche Resultat nicht so interpretiert werden, dass die Degustation damit an Bedeutung verloren hätte. Viel eher soll der Vorwurf entkräftigt werden, dass die Analytik lediglich trockene, technische und kaum interpretierbare Zahlenlisten ausdruckt. Diese Zahlen enthalten wie gezeigt erstaunlich viel an echten Informationen. Die Analysenwerte werden lediglich in einer Form präsentiert, die sie auch für den Fachmann wenig interpretierbar machen. Betrachtet man den zeitlichen und finanziellen Aufwand, der sich hinter der Erstellung einer solchen «Analysenliste» versteckt, so erstaunt es, dass nicht mehr Aufwand betrieben wird, die Daten auch intelligent zu präsentieren. Obwohl Analytik für Trennung steht, sollte sich der Ehrgeiz des Analytikers nicht auf möglichst tiefe Nachweisgrenzen und hohe Wiederfindungsraten beschränken. In der Zukunft wird sich der Analytiker vermehrt von der Technik lösen und seine gemessenen Einzelwerte zu einem sinnvollen Ganzen zusammenfügen müssen. Es darf nicht das Ziel sein, ein komplexes Lebens- oder Genussmittel mittels möglichst vielen unzusammenhängenden Zahlen zu beschreiben. Diese Zahlen und deren Interaktion untereinander müssen produktrelevant interpretiert als Antworten auf eine eingangs gestellte Frage das Labor verlassen.

Zusammenfassung

Eine offizielle Weindegustationsmeisterschaft wurde erstmals ausschliesslich aufgrund von Analysendaten und deren statistischer Auswertung erfolgreich bestritten. Alle 18 Weine konnten korrekt zugeordnet werden. Bei den Mustern handelte es sich um 6 Weine aus der Schweiz, 10 aus Europa und 2 aus Übersee.

Diese Aufgabe wurde mittels 18 verschiedenen Routineanalyseparametern (organische und anorganische Anionen, Kohlenhydrate, Blei und Cadmium) sowie multivariaten statistischen Techniken (Diskriminanzanalyse) gelöst. Zur Verfügung stand eine Datenbank von mehr als 6000 Weinen.

Résumé

Un concours officiel de dégustation de vin a été éffectué avec succès pour la première fois que sur la base de données d'analyse et d'exploitation statistique. Tous les 18 vins ont été classifiés correctement. Parmi les échantillons il y avaient 6 vins de Suisse, 10 vins d'Europe et 2 vins de pays d'outre-mer.

Le problème a été résolu à l'aide de données (anions organique et inorganique, hydrates de carbone, plomb et cadmium) ainsi que de techniques multivariante et statistique (analyse des discriminants). La banque de données à nôtre disposition contenaient plus de 6000 vins.

Summary

An official Wine-tasting Championship was for the first time successfully contested by the sole use of analytical data and statistical techniques. All 18 wines were correctly classified. Among the samples were 6 Swiss, 10 European and 2 Oversea wines.

This task was preformed based on routine analytical data (organic and inorganic anions, carbohydrates, lead and cadmium) and the use of multivariate statistical techniques (discriminant analysis). The basis was a set of data containing more than 6000 wines.

Literatur

1. *Rapp, A., Suckrau, I.* und *Versini, G.:* Untersuchungen des Trauben- und Weinaromas. Z. Lebensm.-Unters.-Forsch. **197,** 249–254 (1993).
2. *Noble, A.C., Flath, R.A.* and *Forrey, R.R.:* Wine headspace analysis. Reproducibility and application to varietal classification. J. Agric. Food Chem. **28,** 346–353 (1980).
3. *Schreier, F.* und *Drawert, F.:* Anwendung der multiplen Diskriminanzanalyse zur Differenzierung von Rebsorten an Hand der quantitativen Verteilung flüchtiger Weininhaltsstoffe. Mitt. Klosterneuburg. **22,** 225–234 (1976).
4. *Rapp, A., Hastrich, H.* und *Engel, L.:* Gaschromatographische Untersuchungen über die Aromastoffe von Weinbeeren. Vitis **15,** 29–36 (1976).
5. *Rapp, A.* und *Hastrich, H.:* Gaschromatographische Untersuchungen über die Aromastoffe von Weinbeeren. Vitis **17,** 288–298 (1978).
6. *Rapp, A.* und *Güntert, A.:* Beitrag zur Charakterisierung des Weines der Rebsorte Weisser Riesling. Vitis **24,** 139–150 (1985).
7. *Siegmund, H.* und *Bächmann, K.:* Anwendung der numerischen Taxonomie für die Klassifizierung von Weinen. Z. Lebensm.-Unters.-Forsch. **166,** 298–303 (1978).
8. *Boreszéki, J.* und *Koltay, L.:* Untersuchung der Mineralstoffzusammensetzung von Weinen aus Transdanubien und ihre Klassifikation nach Weingegenden. Z. Lebensm.-Unters.-Forsch. **177,** 15–18 (1983).
9. *Eschnauer, H.:* Spurenelemente und Ultra-Spurenelemente in Wein. Naturwissenschaften. **73,** 281–290 (1986).
10. *Kalinowski, H.:* Der Schrecken der Weinpanscher. Chemie in unserer Zeit. **5,** 162–171 (1988).
11. *Holbach, B.* und *Förstel, H.:* Das Verhältnis der Stabilisotopen ^{18}O und ^{16}O zur Beurteilung von Auslandwein. Z. Lebensm.-Unters.-Forsch. **198,** 223–229 (1994).

12. *Eschnauer, H.* und *Hölzl, S.:* Isotopensignaturen schwerer Elemente als Parameter zur Charakterisierung von Weinen: Isotopen-Vinogramme. Vitic. Enol. Sci. **49 (3),** 125–129 (1994).
13. *Martin, P.* and *Zhang, B.:* The use of trace element data to complement stable isotope methods in the characterization of grape musts. Am. J. Enol. Vitic. **45,** 79–85 (1994).
14. *Etiévant, P.* and *Schlich, P.:* Varietal and geographic classification of French red wines in term of major acids. J. Sci. Food Agric. **46,** 421–438 (1989).
15. *Larrechi, M.S.* and *Rius, F.X.:* Multivariate data analysis applied to the definition of two Catalan viticultural regions. Z. Lebensm.-Unters.-Forsch. **185,** 181–184 (1978).
16. *Kaufmann, A.:* Ansätze zur Bestimmung der Weinprovinienz mittels Routineanalysendaten. Mitt. Gebiete Lebensm. Hyg. **85,** 566–578 (1994).

Anton Kaufmann
Schuler Weine
Wein-Analytik
CH-6423 Seewen