

Zeitschrift: Veröffentlichungen des Geobotanischen Institutes der Eidg. Tech. Hochschule, Stiftung Rübel, in Zürich

Herausgeber: Geobotanisches Institut, Stiftung Rübel (Zürich)

Band: 90 (1986)

Artikel: Analyse vegetationskundlicher Daten : Theorie und Einsatz statistischer Methoden = Analysis of vegetation data : theory and application of numerical methods

Autor: Wildi, Otto

DOI: <https://doi.org/10.5169/seals-308832>

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: 23.12.2025

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

Analyse vegetationskundlicher Daten

Theorie und Einsatz statistischer Methoden

Analysis of vegetation data

Theory and application of numerical methods

by Otto WILDI

1986

Inhalt

Vorwort	5
1. Eine Uebersicht	8
1.1 Besonderheiten statistischer Analysen	8
1.2 Das Untersuchungsobjekt: Die Stichprobe	11
1.3 Strukturmodelle von Stichproben	13
1.4 Regressionsmodelle	20
2. Versuchsplanung und Stichprobenverfahren	25
2.1 Problemstellungen	25
2.1.1 Untersuchung der Stichprobenstruktur	26
2.1.2 Zusammenhang von Pflanzengesellschaft und Standort	29
2.1.3 Zeitstudien	32
2.2 Stichprobenverfahren	33
2.2.1 Abgrenzung der Grundgesamtheit	34
2.2.2 Die Stratifizierung	36
2.2.3 Der Stichprobenplan	38
3. Skalierung und Transformation	43
3.1 Datentypen	44
3.1.1 Nominaldaten	44
3.1.2 Ordinaldaten (Rangdaten)	45
3.1.3 Metrische Daten	46
3.2 Skalierung von Artmächtigkeiten	47
3.3 Transformationen	50
3.3.1 Transformation von Einzelwerten	50
3.3.2 Vektortransformationen	54
3.4 Wirkung von Vektortransformationen auf Vegetationstabellen	60
4. Aehnlichkeitsmasse	63
4.1 Die Euklidsche Distanz	63
4.2 Die Sehnendistanz	67
4.3 Skalarprodukt und Kovarianz	68
4.4 Der Korrelationskoeffizient	72
4.5 Kontingenzmasse	73
4.6 Absolutwertfunktionen	80
4.7 Die Mahalanobis Distanz	82
4.8 Informationsmasse	85

5. Gruppierungsanalysen	96
5.1 Gruppenstruktur	96
5.2 Heuristische Verfahren	100
5.3 Teilungsverfahren	103
5.3.1 Assoziationsanalyse	103
5.3.2 Gridanalyse	105
5.4 Agglomerative Verfahren	109
5.4.1 Single Linkage Analysis	109
5.4.2 Complete Linkage Analysis	112
5.4.3 Average Linkage Analysis	114
5.4.4 Minimalvarianz-Analyse	115
5.4.5 Besonderheiten Agglomerativer Verfahren	117
6. Gewichtungsverfahren (Rangierung)	121
6.1 Feolis Methode	123
6.2 Rangierung nach erklärter Varianz	125
6.3 Rangierung nach Gruppenstruktur	131
6.4 Stressanalyse	138
7. Ordination	142
7.1 Vorbemerkungen	142
7.2 Prinzip der Ordinationsmethoden	144
7.3 Die Hauptkomponentenanalyse	146
7.4 Q- und D-Technik	157
7.5 Korrespondenzanalyse	161
7.6 Varianten der Hauptkomponentenanalyse	170
8. Analyse geordneter Vegetationstabellen	176
8.1 Die Konzentrationsanalyse	176
8.2 Vergleich zweier Vegetationstabellen	187
9. Zusammenhang zwischen Vegetation und Standort	193
9.1 Zielsetzung und Schwierigkeiten	193
9.2 Grafische Lösungen	196
9.3 Korrelation von Ordination und Standort	200
9.4 Die Diskriminanzanalyse	203
10. Diskussion und Schlussfolgerungen	214
10.1 Grundgesamtheit und Stichprobe	214
10.2 Methodische Entscheidungsschritte	215
10.3 Elemente einer Standardstrategie	216
Literatur	220

Vorwort

Statistische Analysemethoden haben sich im Laufe der vergangenen Jahre in der Vegetations- und Standortskunde rasch verbreitet. Zum einen ist dies auf die Entwicklung spezieller numerischer Verfahren sowie auf die Verfügbarkeit genügend leistungsfähiger Rechenautomaten und Programme zurückzuführen. Zum andern hat sich das Anwendungsgebiet der Vegetationskunde gewandelt. Die reine Beschreibung und Systematisierung der Pflanzendecke der Erde, die Synsystematik, verliert zusehends an Bedeutung, während die Analyse ökologischer Zustände und Veränderungen eine der wichtigsten Aufgaben geworden ist. Der Reproduzierbarkeit der Ergebnisse kommt dabei hohe Bedeutung zu. Eine naturwissenschaftlich einwandfreie Versuchsplanung und die Verwendung formaler, nachvollziehbarer Auswerteverfahren drängen sich auf. Letztere sind - speziell für Vegetationsanalysen - in rascher Entwicklung begriffen und mit der Verbreitung von Mikrocomputern wohl demnächst jedem Vegetationskundler zugänglich.

Der erfolgreiche Einsatz von Computerprogrammen erfordert die sorgfältige Abwägung aller Voraussetzungen, der Fragestellung ebenso wie der Besonderheiten der Datenstrukturen und der verfügbaren Methoden. Unter diesem Aspekt versucht die vorliegende Arbeit, für die Vegetationskunde allgemein gültige Lösungswege und Methoden vorzuschlagen. PIELOU (1984, S. 12) schreibt in ähnlichem Zusammenhang von Techniken, welche die Strukturen von Datensätzen zu untersuchen erlauben. Grösste Bedeutung haben dabei die Verfahren der multivariaten Statistik erlangt, die allerdings meist als geometrische Modelle verstanden werden. Auch der Begriff "numerische Methoden" umschreibt nach Auffassung einiger Autoren die Disziplin treffend. Die in dieser Arbeit dargestellten Auswerteverfahren lassen sich zwangslos unter "explorativer Statistik" einordnen. Aus diesem Grunde soll hier in der Regel von "statistischen" Auswertemethoden die Rede sein, wobei Statistik im weitesten Sinne als Disziplin formaler Datenanalyse zu verstehen ist.

Die Einführung statistischer Analysemethoden gestaltet sich in der Vegetationskunde schwieriger als erwartet. Bis zu Beginn der siebziger Jahre wurden praktisch alle Leistungen der Pflanzenökologie in traditioneller Arbeitsweise erbracht. Der Einsatz von Computern entsprang selten dem Wunsch, auf konsistenterere Methoden umzustellen. Im Vordergrund stand meistens das Bedürfnis, langwierige Analyseschritte zu automatisieren und zu beschleunigen. Automatisieren heisst aber, auf intuitives Vorgehen zu verzichten und Auswerteregeln eindeutig zu formulieren. Damit können sich erfahrungsgemäss nicht alle Pflanzenökologen befreunden. Dazu kommt, dass die vielerorts gelehrt "klassische" Statistik ihrer vielen Restriktionen wegen bei pflanzensoziologischen Problemen selten weiterhilft.

Zur strukturellen Analyse von Vegetationsdaten bieten sich viele multivariate Verfahren an. Diese sind vor allem in den Sozialwissenschaften gut eingeführt. Entsprechende Computerprogrammpakete sind an allen grossen Rechenzentren verfügbar. Deren Einsatz für Fragestellungen der Vegetationskunde hat sich als wenig erfolgreich erwiesen, treten doch namentlich in Vegetationstabellen sehr spezifische Datenstrukturen auf. Die Entwicklung fachgerechter Methoden hat sich in der Folge zu einer eigenen Disziplin entwickelt. Im englischen und französischen Sprachraum sind heute Lehrbücher vorhanden, welche mathematische Verfahren in der Vegetationskunde behandeln. Die Praxis zeigt aber, dass sich damit die Integration statistischer und vegetationskundlicher Denkweise in nur unbefriedigendem Masse vollzieht (GRABHERR 1985). Einer solchen Verbindung einen Schritt näher zu kommen, war die wesentliche Motivation zur vorliegenden Arbeit. Dabei wird die Vegetationsanalyse in ein konzeptionelles Umfeld gestellt, das sich teilweise von der traditionellen pflanzensoziologischen Sicht unterscheidet. Der Statistiker wiederum wird feststellen, dass die gewählte Systematik aus verschiedenen Gründen nicht immer der formalen Verwandtschaft der Methoden entsprechen kann.

Eine Voraussetzung für die erfolgreiche Anwendung statistischer Verfahren in Forschung und Praxis ist die Ver-

fügbarkeit geeigneter Computerprogramme. Dies hat den Autor veranlasst, ein eigenes Programmpaket zu entwickeln, welches einige der hier beschriebenen Verfahren zu realisieren erlaubt (WILDI und ORLOCI 1983). Programme spezieller Methoden finden sich in grosser Zahl bei ORLOCI (1978). TABORD ist ein weiteres Programm zum Ordnen von Vegetationstabellen (VAN DER MAAREL et al. 1978). ORDIFLEX, COMPCLUS (GAUCH 1977, 1979) sowie TWINSPLAN und DECORANA (HILL 1979a, 1979b) gehören zur Programmserie der Cornell University. Eine solche wurde auch an der Universität Trieste entwickelt (FEOLI et al. 1982). CLUSTAN ist ein grosses Paket von Gruppierungsalgorithmen (WISHART 1978) - um nur einige wichtigere zu erwähnen. Auf die zahlreichen Publikationen nicht vegetationskundlicher Art, welche Programme enthalten, wird in den entsprechenden Kapiteln verwiesen.

Bei der Abfassung dieser Arbeit habe ich von vielen Anregungen meiner Fachkollegen profitieren können. Ihnen allen möchte ich an dieser Stelle danken. Zu besonderem Dank verpflichtet bin ich den Herren H. R. Binz, dipl. Ing. ETH, Prof. Dr. G. Grabherr, A. Grünig, dipl. natw. ETH und Prof. Dr. W. Urfer, die das Manuskript aus der Sicht ihres Fachbereiches kritisch durchgesehen haben, sowie Herrn Prof. Dr. E. Landolt, der dessen Publikation ermöglicht hat.

Birmensdorf, Mai 1986

1. Eine Uebersicht

1.1 Besonderheiten statistischer Analysen

Die Botanik und insbesondere die Vegetationskunde sind Wissenschaften, die für die Anwendung von Analysen der explorativen Statistik bekannt sind. Viele ihrer Fachleute haben sich allerdings kaum je solcher Hilfsmittel bedient. In der Tat waren es stets kleinere Forschergruppen innerhalb verschiedenster vegetationskundlicher Schulen, welche die Anwendung statistischer Modelle auf vegetationskundliche Daten studierten und entwickelten. Viele Praktiker verwendeten bisher mangels leistungsfähiger Rechner subjektiv-manuelle Methoden der Datenverarbeitung, die mit der Zeit etwas vereinheitlicht wurden. ELLENBERGS (1956) Vorschriften zum Ordnen einer Vegetationstabelle sind ein Beispiel. Mit der Verfügbarkeit schneller Rechenautomaten hat sich nun die Situation geändert. Selbst aus der Sicht des reinen Anwenders lohnt es sich, nach Sinn und Leistungsfähigkeit statistischer Methoden zu fragen.

In allen nachfolgenden Ueberlegungen treten immer wieder die Begriffe Pflanzensoziologie und Pflanzenökologie auf. Deren Definition kann jedem einschlägigen Lehrbuch entnommen werden. Wesentlich ist, dass die Pflanzenökologie die Verbindung sucht zwischen Standortsdaten einerseits und pflanzensoziologischen Daten andererseits. Wenn gelegentlich von speziellen Datenstrukturen die Rede ist, so bezieht sich dies nur auf die letzteren. Standörtliche Daten dagegen unterscheiden sich von solchen anderer naturwissenschaftlicher Disziplinen kaum und lassen sich mit dem normalen statistischen Instrumentarium bearbeiten. Ein ausgesprochen fachspezifisches Problem bildet dagegen wiederum - immer aus der Sicht der Analyse - die Untersuchung des Zusammenhangs zwischen pflanzensoziologischen und standörtlichen Daten.

Wer erstmals statistische Analysemethoden einsetzt, wird feststellen, dass - im Gegensatz zu manuellen Methoden - die Resultate exakt reproduzierbar sind. Meist spielt es keine Rolle, in welcher Reihenfolge die Daten in den

Analyseprozess eingeschleust werden. Sofern die Methode genau beschrieben wird, ist die Verarbeitung für Drittpersonen bis ins Detail einsichtig. Ein weiterer Vorteil liegt in der hohen Effizienz. Wird ein multivariater Datensatz analysiert, also zum Beispiel eine Vegetationstabelle oder eine Tabelle von Standortsdaten, so bearbeiten gute Methoden alle Aufnahmen gleichzeitig und gleichberechtigt. Die Resultate widerspiegeln deshalb die Gesamtähnlichkeitsstruktur des Datensatzes. Dies wiederum erklärt den erheblichen rechnerischen Aufwand, mit dem solche Analysen verbunden sind. Dass dafür Computer eingesetzt werden müssen, mag einerseits als Nachteil gewertet werden. Andererseits hat dies den willkommenen Nebeneffekt, dass sich langwierige Arbeitsschritte wie das Umschreiben von Tabellen und das Zeichnen von Diagrammen automatisieren lassen. In der vegetationskundlichen Forschung wird davon zunehmend und mit bemerkenswertem Erfolg Gebrauch gemacht. Die neuere Fachliteratur bestätigt diese Feststellung (vgl. z.B. PODANI 1984). Sie wirft aber zugleich Fragen auf, mit denen sich auch der Leser dieser Ausführungen konfrontiert sieht.

Auffällig ist zunächst, dass es ausserordentlich viele verschiedene Methoden gibt und dass laufend neue entwickelt werden. Warum diese Vielfalt? Jeder Methode entspricht ein mehr oder weniger eigenständiges geometrisches oder statistisches Modell. Solche Modelle sind in fast beliebiger Zahl formulierbar. Der Vegetationskundler sucht für seine Daten ein Modell, das die Struktur seiner Befunde möglichst klar abzubilden vermag. Seine Analysen sind deshalb erfolgversprechender, wenn er unter vielen Modellen geschickt auszuwählen versteht. Aus diesem Grunde ist es auch nicht sinnvoll, im Interesse einer Vereinheitlichung ein einziges Verfahren zu propagieren. Vielmehr ist immer ein - natürlich überblickbares - Spektrum von Möglichkeiten in Erwägung zu ziehen.

In Anbetracht der Vielfalt der Methoden drängt sich auch die Frage auf, weshalb für die Pflanzenökologie spezielle Methoden erforderlich sind. Die Antwort ergibt sich leicht aus Erfahrung im Umgang mit solchen Daten. In der Praxis tritt

immer wieder ein ähnliches Spektrum von Datenstrukturen auf, welches eben für das Fachgebiet typisch zu sein scheint. In der Tat weisen Vegetationsdaten - im Vergleich zu Datensätzen anderer Wissenschaften (geografische, sozialwissenschaftliche, ökonomische Daten usw.) - einen hohen Standardisierungsgrad auf. Dies liegt an den allgemein gültigen Naturgesetzen, denen die Vegetation unterworfen ist, aber auch an der Verwendung weit verbreiteter, teilweise standardisierter Erhebungsverfahren (vgl. z.B. MUELLER-DOMBOIS und ELLENBERG 1974). Es ist auch als Vorteil zu werten, dass die Daten nicht bereits bestehenden Statistiken entstammen, sondern normalerweise auf einem räumlich und zeitlich nach wissenschaftlichen Kriterien festgelegten Stichprobenverfahren basieren. Insbesondere multivariate Verfahren arbeiten unter solchen Voraussetzungen optimal.

Inhomogenes Datenmaterial erfordert die Verwendung stark standardisierender Methoden. Der weit verbreitete Produktmomentkorrelationskoeffizient ist ein Beispiel dafür. Gerade bei diesem sind aus formalen Gründen fast immer Informationsverluste in Kauf zu nehmen. Bei der Analyse von Vegetationsdaten kann demgegenüber eine Auswahl wenig standardisierender und deshalb gut differenzierender Verfahren eingesetzt werden. Weil die Stichprobenplanung als Voraussetzung für ein gutes Gelingen eine entscheidende Rolle spielt, kann durch sie zugleich eine weitere Verbesserung der Effizienz vieler Methoden erzielt werden. Dies ist der einzige Weg, um die Reproduzierbarkeit numerischer Analysen voll auszunützen.

Schliesslich stellt sich die Frage, ob die beobachteten Datenstrukturen allgemeingültig oder bloss ein Abbild des einigermassen willkürlich gewählten Analyseverfahrens sind. Die letzte Vermutung kann wohl für strukturelle Details eines multivariaten Datensatzes zutreffen. Für die Interpretation ist es in diesem Falle besonders wichtig, die genaue Funktionsweise des Analyseverfahrens zu kennen. Hingegen zeigt die Erfahrung, dass die meisten Methoden eindeutige Zusammenhänge und Strukturen aufzudecken vermögen. Wer die Möglichkeit zur Variation seiner Methoden besitzt, verfügt

also zugleich über ein Instrument, um die Bestimmtheit seiner Befunde abzuschätzen.

Statistische Analysen können nicht isoliert betrachtet werden. Sie sind nur ein Schritt auf dem Weg zu wissenschaftlichen Erkenntnissen. Es ist nicht Absicht dieser Arbeit, auf alle methodischen Probleme einzugehen. So werden z.B. die pflanzensoziologischen Aufnahmemethoden nicht diskutiert. Die Planung von Stichprobenkonzepten ist jedoch ein so wichtiger Schritt, dass ihr ein eigenes Kapitel gewidmet ist. Doch zunächst sind einige einführende Überlegungen zu Datenstrukturen und Datenmodellen zu behandeln.

1.2 Das Untersuchungsobjekt: Die Stichprobe

Multivariate Analysenmethoden beschäftigen sich mit Stichproben. Jede Stichprobe besteht aus einer kleineren oder grösseren Anzahl von Individuen (Stichprobenelementen), welche sich mehr oder weniger voneinander unterscheiden. Beispiele sind die Schüler einer Schulkasse, die Bäume eines Waldstückes, die Häuser einer Siedlung usw. In der Pflanzensoziologie bildet oft die Vegetationstabelle die Stichprobe. Individuen sind floristische und ökologische Aufnahmen. Ein Beispiel zeigt Abb. 1.1. Wir haben hier einen multivariaten Datensatz vor uns. Im Gegensatz zu univariaten Datensätzen sind die Individuen grundsätzlich durch eine Vielzahl von Merkmalen (vgl. GAUCH 1982, S.4 und 5) charakterisiert. Bei Stichproben menschlicher Populationen wären analoge Merkmale Eigenschaften wie Körpergrösse, Haarfarbe, Alter usw., bei soziologischen Studien auch Bildung, Religion, Weltanschauung usw. In der Pflanzenökologie sind es Artvorkommen (Abundanz, Dominanz), Standortsfaktoren, Lebensformen, phänologischer Zustand und anderes mehr, welche die Individuen charakterisieren. Entscheidend an der eben dargestellten Sichtweise ist, dass Aufnahmen und Tabellen die fundamentalen Einheiten der Analyse bilden. Im Gegensatz etwa zur Auffassung von BRAUN-BLANQUET (1964) spielen Pflanzengesellschaften erst im Zusammenhang mit Ergebnissen eine wesentliche Rolle.

Stichprobe

Individuen

		1	2	3	4
		Arten	1	2	3
Merkmale	Aufnahmen	Arten			
	Aufnahmen	1			
	Aufnahmen	2			

Abb. 1.1 Beispiel einer Stichprobe, bestehend aus 4 Individuen, welche durch je 3 Merkmale charakterisiert sind.

Stichprobe 1 (Wald)

Stichprobe 2 (Wiese)

		1	2	3	
		Arten	1	2	3
Merkmale	Aufnahmen	Arten			
	Aufnahmen	1			
	Aufnahmen	2			
	Aufnahmen	3			
	Aufnahmen	4			

		1	2	3	
		Arten	1	2	3
Merkmale	Aufnahmen	Arten			
	Aufnahmen	1			
	Aufnahmen	2			
	Aufnahmen	3			
	Aufnahmen	4			

Abb. 1.2 Zwei klar unterscheidbare Stichproben. Sie werden sinnvollerweise separat analysiert.

Die Abgrenzung der Grundgesamtheit als Gegenstand einer Analyse ist ein wichtiger eigenständiger Schritt im Rahmen einer Untersuchung (vgl. Kap. 2.2.1). Manchmal ist eine einzige unstrukturierte Stichprobe Gegenstand einer Analyse. Bei umfangreicherem Forschungsprojekten sind kompliziertere Verhältnisse jedoch die Regel. Dabei kann es sich als sinnvoll erweisen, das gesamte Datenmaterial a priori in Stichproben zu unterteilen, welche einzeln analysiert werden. Dies zeigt Abb. 1.2. Die beiden Datensätze lassen keinen Zweifel offen, dass sie in jeder Hinsicht (d.h. bezüglich aller Merkmale) verschieden sind und deshalb ohne weiteres getrennt analysiert werden können. Es gibt aber öfters Situationen, bei denen die Aufteilung in verschiedene Stichproben nicht von vornherein gegeben ist. Dabei lassen wohl einige Merkmale eine bestimmte Unterteilung als sinnvoll erscheinen, andere widersprechen dem jedoch. Gleichzeitig mag es aber wenig sinnvoll sein, so zu handeln, als ob eine unstrukturierte Stichprobe vorläge. Die Untersuchung einer vermuteten Gruppenstruktur ist dann Gegenstand der Analyse. Im Beispiel von Abb. 1.3 liegt die Gruppierung dem vorgegebenen Merkmal "Exposition" zugrunde. Die daran anschliessende Analyse soll dann zeigen, inwiefern die Kombination der anderen Merkmale diese Vorgabe bestätigen.

1.3 Strukturmodelle von Stichproben

Ziel multivariater Analysen ist die Untersuchung der Struktur, dem Vorhandensein von Gesetzmässigkeiten innerhalb der Stichprobe. Man bildet zu diesem Zwecke geometrische Modelle der Daten. Alle diese Modelle zielen darauf ab, die Datenstruktur vereinfacht abzubilden. Der Preis, welcher für jede Vereinfachung zu bezahlen ist, besteht in einem Informationsverlust. Der Gewinn besteht darin, dass in der Stichprobe vorhandene Gesetzmässigkeiten klarer hervortreten als in den Originaldaten.

Soll nun die Struktur einer multivariaten Stichprobe untersucht werden, so geschieht dies durch den Vergleich ihrer Individuen oder ihrer Merkmale. Dafür stehen sehr unterschiedliche Methoden zur Verfügung. Im allgemeinen wird man

Aufnahmen	Stichprobe					
	Gruppe 1			Gruppe 2		
	1	2	3	4	5	6
Exposition	N	N	N	SE	SE	SE
Art 1						
Art 2						
Art 3						

Abb. 1.3 Eine Stichprobe, welche zum vornherein in zwei Gruppen unterteilt worden ist. Die Prüfung der Unterschiedlichkeit dieser Gruppen kann ein Analyseziel sein.

	Aufn. 1	Aufn. 2	Überein- stimmung s	Unter- schied d
Art 1			1	-
Art 2			-	1
Art 3			-	1
			$\Sigma s = 1$	$\Sigma d = 2$

Abb. 1.4 Vergleich zweier Aufnahmen. $\sum x_i$ ist ein Ähnlichkeitsmass, $\sum d_i$ ein Distanzmass (Unähnlichkeitsmass).

zunächst die Merkmale eines Individuenpaares einzeln vergleichen und die so erhaltenen Ergebnisse zusammenfassen. Abb. 1.4 zeigt zwei mögliche Vorgehensweisen: Je nach Definition des Vergleichskoeffizienten liegt am Schluss ein Mass für die Aehnlichkeit oder aber für die Unähnlichkeit (Distanz) der Individuen vor.

Die umfassende Analyse der Stichprobe erfordert nun eine Bestimmung der gesamten Stichprobenstruktur. Man kommt dabei nicht umhin, jedes Individuum mit jedem anderen zu vergleichen. Schon bei bescheidenen Stichprobengrössen ist daher eine grosse Zahl von Vergleichskoeffizienten zu berechnen. Abb. 1.5 zeigt ein einfaches Beispiel. Die Vegetationstabelle oben besteht aus vier Individuen. Unten ist die sich ergebende Stichprobenstruktur dargestellt. Es handelt sich um eine Aehnlichkeitsmatrix mit 16 Koeffizienten. Da jeder Koeffizient je zwei mal erscheint, ist nur etwas mehr als die Hälfte der Matrix zu berechnen. Die Zahl der paarweisen Vergleiche, M , ergibt sich nach der Formel

$$M = N(N+1)/2.$$

Darin ist N die Anzahl Individuen. Für die Tabelle in Abb. 1.5 erhalten wir:

$$M = 4(4+1)/2 = 10.$$

Der Aufwand zur Bestimmung der Stichprobenstruktur nimmt mit steigender Stichprobengrösse rasch zu. So ergibt sich für eine Vegetationstabelle mit 100 Aufnahmen

$$M = 100(100+1)/2 = 5050.$$

Die Aehnlichkeitsstruktur einer Stichprobe ist mit einer solchen Matrix abschliessend definiert. Sie ist aber ihrer Grösse wegen praktisch nicht überblickbar. Um sie zu analysieren, bedarf es spezieller Methoden, die in den späteren Kapiteln behandelt werden.

Für pflanzenökologische Untersuchungen bedeutungsvoll sind

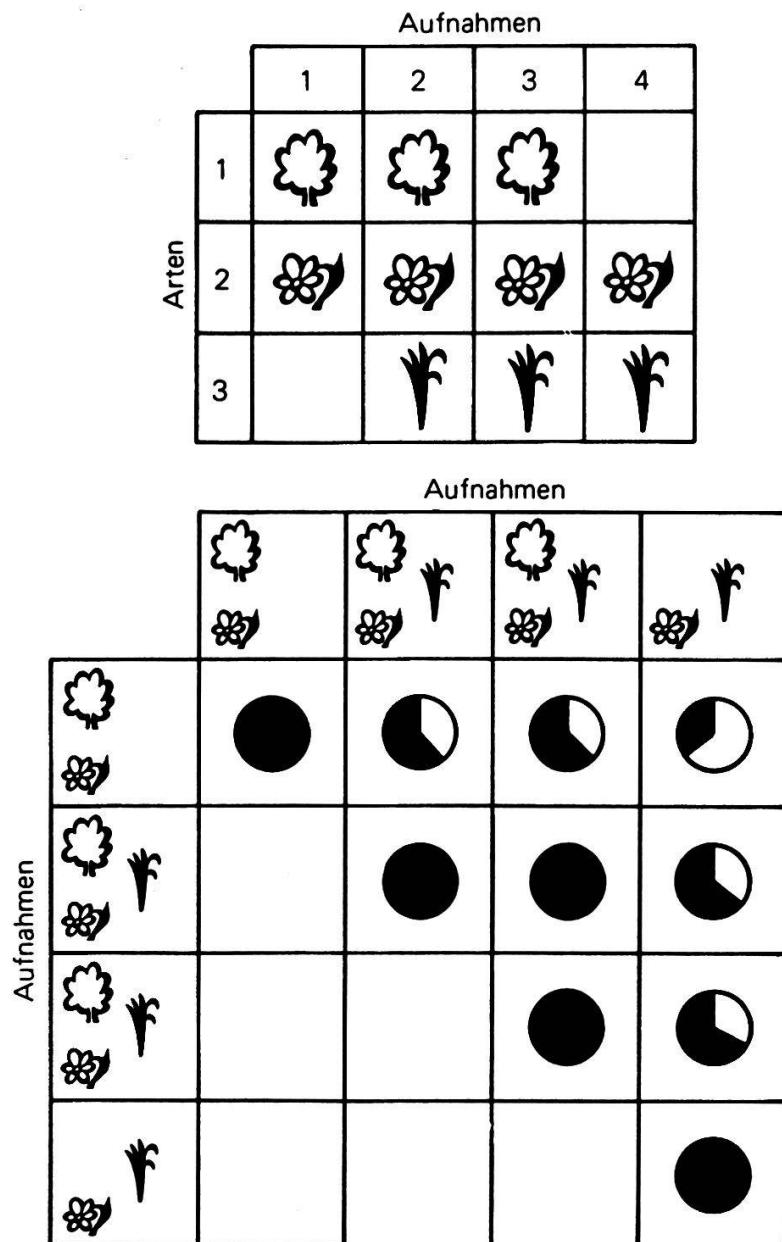


Abb. 1.5 Bestimmung der Ähnlichkeitsstruktur einer Stichprobe. Aus 4 Aufnahmen (oben) wird eine Ähnlichkeitsmatrix berechnet (unten). Die Größe der schwarzen Kreissektoren symbolisiert das Mass des Zusammenhangs zwischen den Aufnahmen.

Gradienten- und Gruppenstrukturen. Solche lassen sich recht gut aus einer ausgewerteten, geordneten Vegetationstabelle herauslesen. Abb. 1.6 zeigt einige einfache Beispiele. Ganz eindeutige Verhältnisse (A,C) treten in Wirklichkeit kaum auf. Die Fälle B und D sind eher typisch.

Ob sich eine Stichprobe in Gruppen aufteilen lässt, wird z.B. mit Hilfe sogenannter Gruppierungsanalysen (Clusteranalysen) untersucht (Kap.5). Manchmal werden deren Resultate in Form von Dendrogrammen präsentiert (Abb. 1.7). Darin sind auf der X-Achse die Individuen aufgetragen, während die Y-Achse die Unterschiedlichkeiten der Gruppen wiedergibt. Die Stärke der Gruppenstruktur ergibt sich aus dem Verhältnis der Ähnlichkeiten innerhalb der Gruppen zu jenen zwischen den Gruppen. In Abb. 1.7, A ist eine klar zweigeteilte Stichprobe dargestellt. Dendrogramm B zeigt dagegen eine schwach ausgeprägte Struktur. Die Unterschiede innerhalb der Gruppen sind hier fast ebenso gross wie diejenigen zwischen denselben.

Um die Gradientenstruktur einer Stichprobe darzustellen, greift man öfters zur Ordination. Die Individuen werden dabei in einem zwei- oder dreidimensionalen Koordinatensystem so dargestellt, dass sich ihre Ähnlichkeitsbeziehungen aus den gegenseitigen Abständen möglichst genau herauslesen lassen. Abb. 1.8 zeigt zwei Beispiele. Darin werden je zwei Merkmale zur zweidimensionalen Darstellung von 8 Individuen verwendet. Verfügt man über drei oder mehr Merkmale, so lassen sich in dieser Weise nur Teile der Gesamtinformation wiedergeben. Die Information ist stets an die Achsen gebunden. Die Beiträge der Achsen zur Gesamtinformation verhalten sich additiv. Trägt beispielsweise die erste Achse 50%, die zweite 30% und die dritte 20% der Information, so ist die maximale Effizienz einer zweidimensionalen Ordination gleich 80%.

Oft wird erwartet, dass Vegetationsaufnahmen, die im Felde einem einfachen Gradienten entstammen, in einer Ordination auf einer geraden Linie erscheinen müssen (GAUCH 1982). Dies ist nicht der Fall. Das Erscheinungsbild in Abb. 1.8 ist

A

		Aufnahmen			
		1	2	3	4
Arten	1				
	2				
	3				
	4				

B

		Aufnahmen			
		1	2	3	4
Arten	1				
	2				
	3				
	4				

C

		Aufnahmen			
		1	2	3	4
Arten	1				
	2				
	3				
	4				

D

		Aufnahmen			
		1	2	3	4
Arten	1				
	2				
	3				
	4				

Abb. 1.6 Unterschiedlich strukturierte Vegetationstabellen. Deutliche Gruppen (A), schwache Gruppen (B), deutlicher Gradient (C), schwacher Gradient (D).

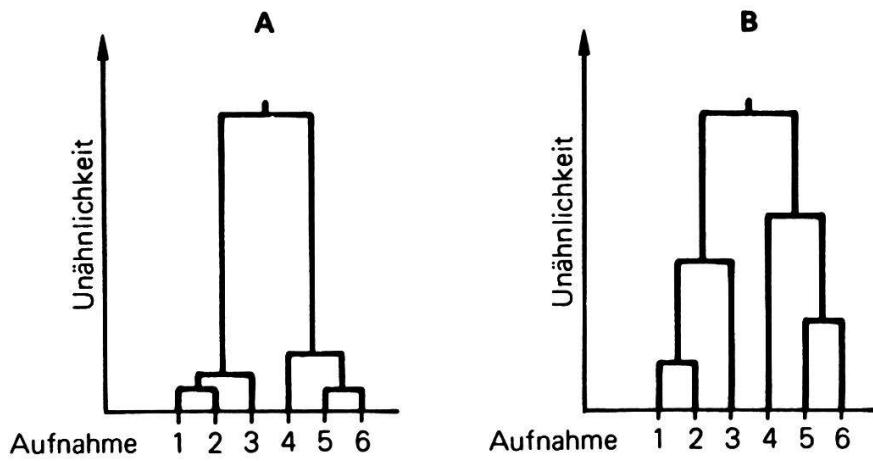


Abb. 1.7 Dendrogramme. Klare Gruppenstruktur mit zwei Gruppen (A), schwache Gruppenstruktur (B).

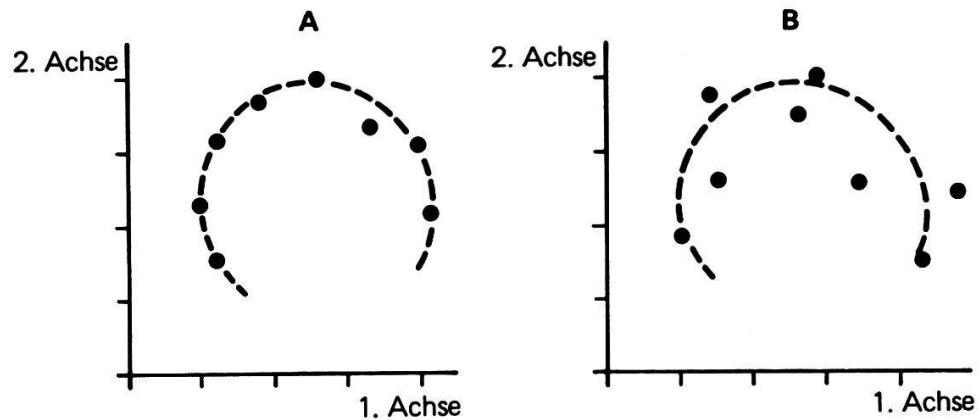


Abb. 1.8 Ordinationen. Klare Gradientenstruktur (A), schwache Gradientenstruktur (B). Umweltgradienten entstammende Vegetationsdaten zeigen in der Ordination oft die hier dargestellte Krümmung.

dagegen typisch, indem die Aufnahmen auf gekrümmten, oft fast kreisförmigen oder - im dreidimensionalen Raum - spiraligen Linien angeordnet sind. Die genauen Ursachen dafür werden in Kapitel 7 gezeigt. Weniger klare Gradienten ergeben Muster wie in Abb. 1.8, rechts, dargestellt. Die Aufnahmen sind hier bandförmig angeordnet. Strukturen dieser Art sind in der Praxis sehr häufig zu finden. Schliesslich können Ordinationen auch Gruppenstrukturen sehr differenziert wiedergeben. Beispiele finden sich in Kap. 5.

1.4 Regressionsmodelle

Es gibt vegetationskundliche Untersuchungen, die sich mit der Erforschung der Datenstruktur begnügen. Dank guter Kenntnis der Standortsansprüche der Arten können solche Resultate ökologisch interpretiert und nutzbar gemacht werden. Eine beliebte Möglichkeit bietet die Verwendung der Zeigerwerte der Arten nach LANDOLT (1977) oder nach ELLENBERG (1979). Man stützt sich in dieser Weise auf alt bekanntes Wissen ab. Pflanzenökologische Forschung umfasst aber auch die Messung von Standortsfaktoren. Ziel der Analyse des Zusammenhangs zwischen Vegetation und Standort ist die Bildung eines prädiktiven Modelles. Ein solches erlaubt die Voraussage des Standortes (auf Grund der Pflanzengesellschaft) oder der Pflanzengesellschaft (auf Grund des Standortes) oder aber eines Standortsfaktors auf Grund der Ausprägung eines andern.

Ein einfaches prädiktives Modell ist die lineare Regression. Sind die Ausgangsdaten als Punkteschwarm in einem zweidimensionalen Koordinatensystem darstellbar, so wird versucht, die Abhängigkeiten der Individuen durch eine Gerade zu beschreiben. Abb. 1.9 zeigt einen solchen Fall. Wird dieses Modell als gültig akzeptiert, so kann bei vorgegebener Neigung der zugehörige pH-Wert abgelesen werden. Die horizontalen Abstände der Punkte von der Regressionsgeraden zeigen, welche Fehler bei einer solchen Aussage toleriert werden müssen. Abb. 1.9, unten, zeigt ein anderes Modell, das dieselbe Datenstruktur genauer wiedergibt. Es handelt sich um eine nichtlineare Regression. Die Pfeile weisen auf

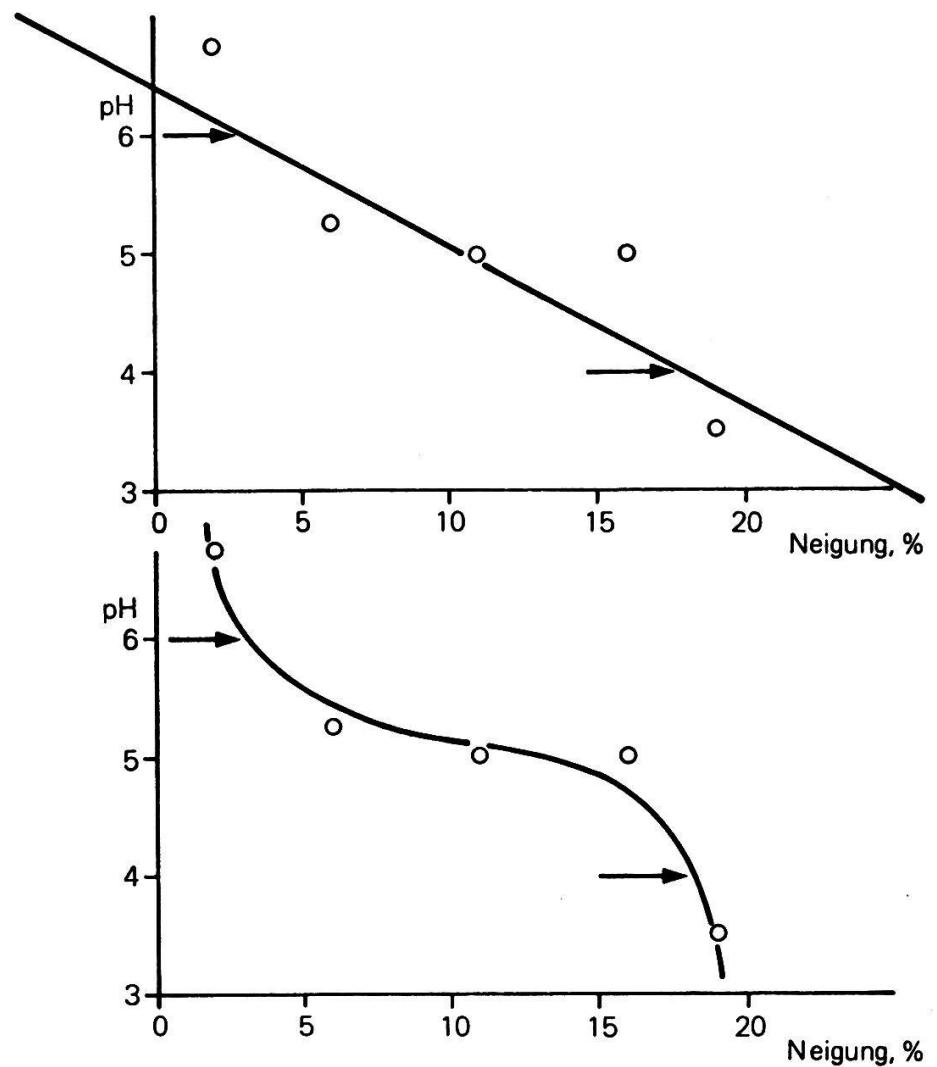


Abb. 1.9 Regressionsmodelle. Einfache, lineare Regression (oben), einfache nichtlineare Regression (unten).

Stellen hin, an welchen die Voraussage mit derjenigen des linearen Modelles identisch ist. Leider treten in der Pflanzenökologie selten Datensätze mit nur zwei Variablen auf, sodass sich mit der einfachen Regression die wenigsten Probleme bewältigen lassen.

Eine Möglichkeit, einen einfachen Vegetationsgradienten mit dem Standort zu verknüpfen, bietet die multiple Regressionsanalyse. In Abb. 1.10 werden zwei Faktoren, der pH-Wert und eine Mitteltemperatur, mit der Position von Vegetationsaufnahmen entlang eines Gradienten in Beziehung gesetzt. Die Lösung ist geometrisch darstellbar. Sie besteht aus einer Ebene, welche analog der Geraden einer einfachen Regressionsanalyse optimal einem Punkteschwarm (Aufnahmen) angepasst sein soll. Die eingezeichnete Aufnahme P demonstriert die Aussagemöglichkeiten des Modells: Werden ein pH-Wert und eine Temperatur vorgegeben, so lässt sich der zu erwartende Vegetationstyp (d.h. seine Entfernung vom Anfangspunkt des Gradienten) voraussagen.

Ein anderes, verwandtes Modell bietet die Möglichkeit, Stichproben mit Gruppenstruktur zu analysieren. In Abb. 1.11 wird diese diskriminanzanalytische Lösung dargestellt. Das besondere an der Situation ist, dass sich der Vegetationstyp weder aus dem pH-Wert allein noch aus der Temperatur voraussagen lässt. Beide Faktoren weisen innerhalb derselben Vegetationstypen gleichzeitig hohe und niedrige Werte auf. Wird jedoch eine günstige Kombination von Messwerten gefunden, so lassen sie sich eindeutig auseinander halten. Das Ergebnis lässt sich geometrisch als neue Achse,

$$y = f(\text{pH}, {}^\circ\text{C})$$

darstellen. Es kann aber auch verbal, mit Hilfe einer "Grammatik" (DALE 1980) ausgedrückt werden. Im vorliegenden Beispiel könnte man formulieren:

- Bei hoher Temperatur und gleichzeitig niedrigem pH-Wert tritt bevorzugt Typ A auf.

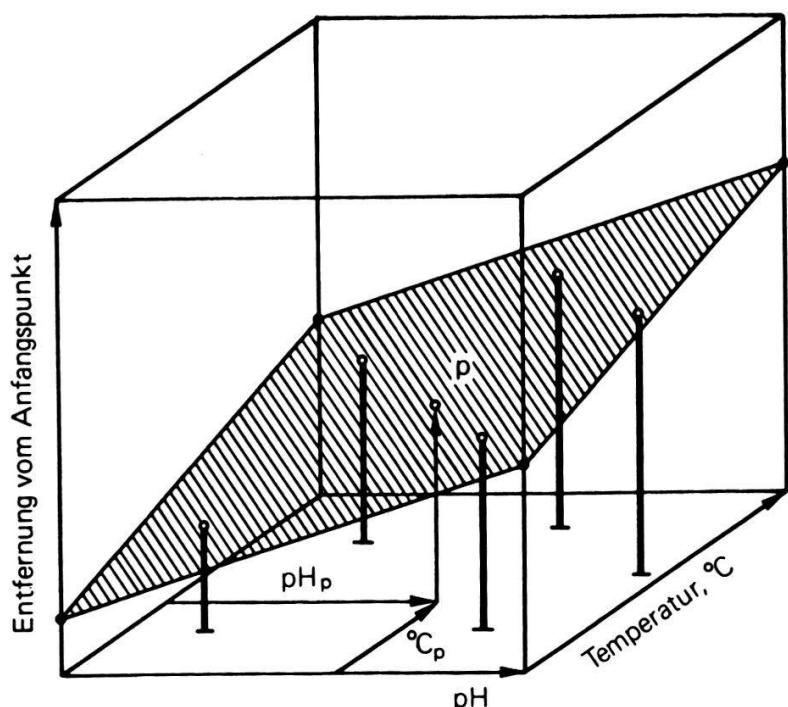


Abb. 1.10 Multiple Regression zwischen der Position von Aufnahmen entlang eines Transsektens und den Standortsfaktoren pH-Wert und Temperatur.

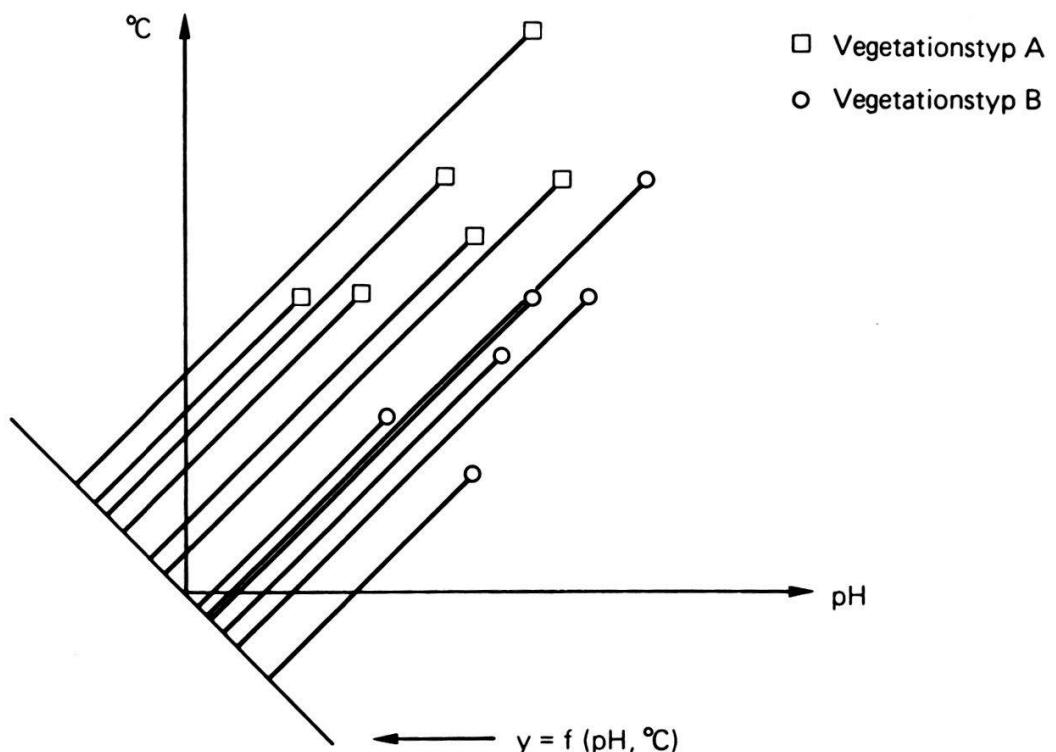


Abb. 1.11 Diskriminanzanalytische Trennung eines Punkteschwarmes mit Hilfe zweier Standortsfaktoren.

- Bei tiefer Temperatur und gleichzeitig tiefem pH- Wert lassen sich die Typen A und B nicht auseinanderhalten.
- usw.

Die Ableitung prädiktiver Modelle für pflanzenökologische Anwendungen ist eine recht schwierige Angelegenheit. Wie in Abb. 1.9 dargestellt, sind die wenigsten Strukturen linear. Im Falle der zweifachen Regression (Abb. 1.10) ist somit zu erwarten, dass die beste Lösung eine irgendwie gekrümmte Fläche und nicht eine Ebene sein wird. Werden zur Modellbildung eine grössere Zahl von Vegetationstypen verwendet, so bedarf die Darstellung der diskriminanzanalytischen Lösung einer Vielzahl von Dimensionen. Diese lassen sich nur schwer gleichzeitig veranschaulichen. Auch verbal ist eine solche Lösung schwierig zu fassen. Schliesslich werden in der Regel sowohl die Vegetation als auch der Standort durch mehrere Merkmale beschrieben (vgl. GAUCH 1982, S. 38). Es müssen also Modelle gefunden werden, die Beziehungen zwischen zwei multivariaten Datensätzen zu untersuchen erlauben. Dieses eigentliche Kernproblem der Vegetationkunde wird in Kapitel 9 näher erörtert.

2. Versuchsplanung und Stichprobenverfahren

Multivariate Verfahren bieten die Möglichkeit, aus komplexem Datenmaterial reproduzierbare Ergebnisse abzuleiten. Die Reproduzierbarkeit gilt insofern, als man von einem gegebenen Datensatz ausgehend durch Anwendung genau festgelegter Regeln immer wieder die gleichen Resultate erhält. Dieser Vorzug formaler Verfahren erfüllt aber seinen Zweck nur dann, wenn auch die Methode zur Gewinnung der Daten Reproduzierbarkeit gewährleistet. Die Schule Braun-Blanquet stellt z.B. keinen solchen Anspruch an die Datenerfassung. Sie appelliert vielmehr an die Beobachtungsgabe und Intuition des Forschers und verbindet diese mit zusätzlich einzuhaltenden Regeln (MUELLER-DOMBOIS und ELLENBERG 1974, S. 32f.). Nun wird aber Pflanzenökologie heute immer mehr zur Erforschung und Lösung von Umweltproblemen eingesetzt. Sie ist dabei gezwungen, von der Arbeitsweise des Bearbeiters unabhängige, nachprüfbare Resultate zu liefern. Die Versuchsplanung wird deshalb zum integrierenden Bestandteil der Methodik. Sie kann nicht unabhängig von den später anzuwendenden Analysemethoden betrachtet werden.

Bereits einleitend wurde festgestellt, dass die meisten der hier vorgestellten Analysemethoden rein geometrisch verstanden werden können. Die Versuchsplanung muss jedoch selbst dann statistischen Grundsätzen folgen, wenn auf Methoden der schliessenden Statistik verzichtet wird. Das Zufallsprinzip ermöglicht erst, bei Felderhebungen Nachvollziehbarkeit zu erreichen. GREEN (1979) hat diesen und weitere Grundsätze, die bei der biologischen Feldforschung zu berücksichtigen sind, in Form von zehn Regeln zusammengefasst (Tab. 2.1). Im Folgenden ist zu diskutieren, wie diese bei pflanzensoziologischen Untersuchungen zu interpretieren sind. Schliesslich sollen einige Stichprobenverfahren dargestellt werden, die sich zur vegetationskundlichen Datenerhebung eignen.

2.1 Problemstellungen

Nach GREEN (1979) soll die Problemstellung so klar und anschaulich dargestellt werden, dass sie einer Drittperson

jederzeit mit wenigen Worten erklärt werden kann. Einfachheit in der Formulierung kommt auch der späteren Analyse zugute. Komplizierte Fragestellungen erfordern zu ihrer Bearbeitung auch komplizierte Methoden. Eine Aufteilung des Gesamtziels (und später der Gesamtanalyse) in Teilziele (Teilanalysen) drängt sich in solchen Fällen auf. Um aber den Regeln von Green (Tab. 2.1) folgen zu können, sollte man sich über die Natur der Zielsetzung der Untersuchung Klarheit verschaffen. Die nachfolgenden Kapitel behandeln deshalb einige typische vegetationskundliche Fragestellungen.

2.1.1 Untersuchung der Stichprobenstruktur

Die Analyse der Stichprobenstruktur ist das klassische Forschungsfeld der Pflanzensoziologie (WESTHOFF und VAN DER MAAREL 1978). Ziel solcher Untersuchungen ist das Auffinden von Gruppen- und Gradientenstrukturen in Stichproben (POORE 1955, ORLOCI 1978) oder in Verbreitungsmustern einzelner Arten (PIELOU 1977, 1984, GREIG-SMITH 1982). Diese Problemstellung ist nicht zu verwechseln mit der Zuordnung neuer Vegetationsaufnahmen zu bereits beschriebenen Pflanzengesellschaften, bei welcher eine Erhebung nach dem Zufallsprinzip nicht notwendig ist. Wie die Suche nach Pflanzengesellschaften, so ist auch die Abgrenzung von Standortstypen als Analyse einer Stichprobenstruktur zu betrachten. Die Resultate können zur flächenhaften Darstellung vorgefundener Zusammenhänge in Form von Vegetationskarten verwendet werden. Im Folgenden soll besprochen werden, wie die Aufnahmemethode einer solchen Zielsetzung anzupassen ist.

Zwei besonders wichtige Aspekte eines pflanzensoziologischen Versuchsplanes sind die Bearbeitungsintensität und die Grösse der Stichprobenfläche (Aufnahmefläche). GREEN (1979) fordert als Regel 3 (Tab. 2.1) die Gleichberechtigung der zu untersuchenden Flächen, Typen oder Inhalte. Abb. 2.1 zeigt, wie dieses Prinzip realisiert werden kann. Ist die grobe Gliederung des Untersuchungsgebietes bekannt und sollen Teilflächen ihrer Grösse entsprechend gewichtet werden, so

Tabelle 2.1

Zehn Regeln für die Versuchsplanung (nach GREEN 1979).

1. Formuliere die Fragestellung mündlich und schriftlich.
2. Einzelmessungen sind nutzlos, Wiederholungen lassen sich dagegen einwandfrei interpretieren.
3. Pro Standort und Vegetation sollte wenn möglich eine gleich grosse Anzahl von Stichproben erhoben werden.
4. Man führe stets Messungen in Kontroll- oder Vergleichsflächen durch.
5. Man führe stets eine Voruntersuchung durch.
6. Die Erhebungsmerkmale sind möglichst gut der Fragestellung anzupassen. Die gesamte Untersuchung ist mit gleich bleibender Intensität und Genauigkeit durchzuführen.
7. Grosse und inhomogene Untersuchungsflächen sind zu stratifizieren. Die Zahl der Stichproben pro Teilfläche bestimmt deren Gewicht.
8. Die Grösse der Aufnahmefläche ist der Grösse, Dichte und räumlichen Verteilung der zu untersuchenden Organismen anzupassen. Die Anzahl der Aufnahmen, nicht deren Grösse entscheidet über die Genauigkeit der Untersuchung.
9. Es ist zu prüfen, ob die Wiederholungen normal verteilt sind. Andernfalls sind die Messresultate zu transformieren oder es sind verteilungsunabhängige Analysemethoden zu verwenden.
10. Man vertraue einem sauber erarbeiteten Resultat, auch wenn es unerwartet oder unerwünscht ist.

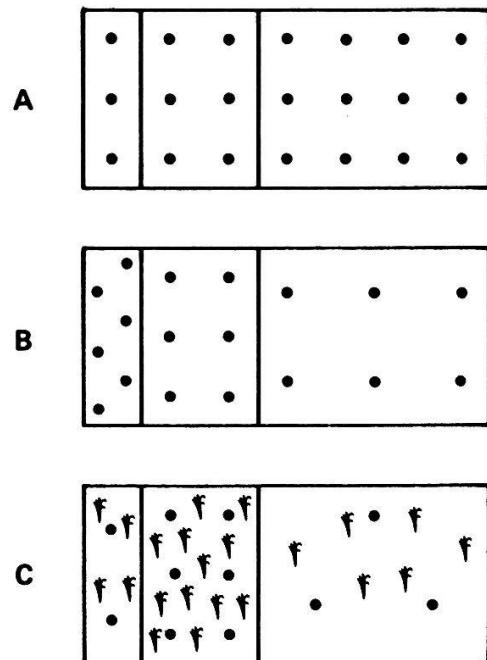


Abb. 2.1 Stichprobenpläne für Strukturuntersuchungen. Gleichberechtigung der Flächeneinheiten (A), der Typen (B), der Bedeutung (z.B. entsprechend der Anzahl vorhandener Individuen) (C). Die Stichproben, markiert durch Punkte, sind systematisch angeordnet.

kann Fall A zur Anwendung kommen (die Stichproben brauchen dazu nicht als systematisches Netz angeordnet zu sein wie in Abb. 2.1). Darin besitzt jede Flächeneinheit (z.B. jede Hektare) gleiches Gewicht. Sollen jedoch die Vegetationstypen gleichberechtigt miteinander verglichen werden, so ist Fall B anzustreben. Natürlich kann ein solcher Versuchsplan selten in einem einzigen Arbeitsgang realisiert werden. Sind weder Lage noch Ausdehnung der Vegetationstypen bekannt, so müssen diese in einem ersten Durchgang gesucht bzw. definiert werden. Dieser Vorgang wird Stratifizierung oder Schichtung genannt (vgl. Abschnitt 2.2.2). Aehnlich verhält es sich mit dem Fall C. Hier sollen Stichproben entsprechend ihrer Bedeutung (z.B. Produktivität) mit gleichem Gewicht untersucht werden. Zu diesem Zwecke müssen sie vorgängig gefunden, abgegrenzt und gemessen werden.

Dem Untersuchungsziel anzupassen ist auch das Ausmass der Aufnahmefläche. Nach Green (Tab. 2.1, Pt.8) ist dazu die Grösse, die Dichte und die Verbreitung der zu untersuchenden Organismen zu berücksichtigen. Solche Ueberlegungen haben in der Vegetationskunde zu Empfehlungen für die zu wählenden Aufnahmeflächen in den wichtigsten Vegetationsformationen geführt. Tab. 2.2 gibt eine solche Zusammenstellung wieder. Die Anwendung dieser allgemein akzeptierten Empfehlungen ist jedoch nicht immer zweckmässig. Wird eine Vegetationskarte hergestellt, so sollte die Aufnahmefläche vielmehr dem später gewünschten Detaillierungsgrad angepasst werden. Die kleinste in der Karte eben noch abgegrenzte Teilfläche sollte dann etwa der Aufnahmefläche entsprechen. Wird also eine Karte im Massstab 1:100'000 hergestellt, so sind die Angaben in Tab. 2.2 wohl meist zu klein. Obwohl das Problem der optimalen Aufnahmefläche Gegenstand vieler Veröffentlichungen ist, muss es nach wie vor als ungelöst betrachtet werden (vgl. PODANI 1984).

2.1.2 Zusammenhang von Pflanzengesellschaft und Standort

Bei pflanzenökologischer Grundlagenforschung begnügt man sich nicht mit der Erfassung der Vegetationsdecke in Form einer Karte. Das Ziel besteht vielmehr in der Erforschung

Tabelle 2.2

Grösse der Aufnahmeflächen für verschiedene Vegetationstypen
(nach MUELLER-DOMBOIS und ELLENBERG 1974, S.48).

Wälder	200-500m ²
Trockenrasen	50-100m ²
Heiden	10-25m ²
Heuwiesen	10-25m ²
Weiden	5-10m ²
Unkrautgesellschaften	25-100m ²
Moosgesellschaften	1-4m ²
Flechtengesellschaften	0.1-1m ²

Tabelle 2.3

Typen von Umweltstudien mit zeitlichen Veränderungen (nach GREEN 1979, s.72).

Hat die Standortsveränderung bereits stattgefunden?	NEIN		JA	
	•	•	•	•
Weiss man wann und wo?	JA	NEIN	JA	NEIN
•	•	•	•	•
Gibt es unbeeinflusste Referenzflächen?	JA	NEIN	•	•
•	•	•	•	•
Typ	1	2	3	4
				5

- 1- Optimale Ursache-Wirkung Studie ist durchführbar
- 2- Die Wirkung muss aus der zeitlichen Veränderung abgeschätzt werden
- 3- Ein Ueberwachungsprojekt hilft weiter
- 4- Die Wirkung muss aus räumlich auftretenden Unterschieden abgeschätzt werden
- 5- Guter Rat ist teuer

Ökologischer Ansprüche verschiedener Arten oder Pflanzengesellschaften, ihrer Empfindlichkeit gegenüber Umwelteinflüssen, ihrer Reaktion auf Nutzungen usw. Dies kann auf verschiedene Weise geschehen. Eine Möglichkeit besteht in der Anwendung der Zeigerwerte nach ELLENBERG (1979) oder nach LANDOLT (1977). Es ist offensichtlich, dass dabei autökologisch gesehen nur wenig neue Erkenntnisse zu erwarten sind. Ergiebiger ist es, wenn parallel zur Vegetation auch der Standort analysiert wird.

Am Anfang einer Untersuchung des Zusammenhangs von Vegetation und Standort kann eine Gruppierungsuntersuchung stehen (Kap. 5). In den so gefundenen Vegetationseinheiten können anschliessend gezielt Standortsmessungen durchgeführt werden. Gemäss Regel 2 von Green (Tab. 2.1) ist darauf zu achten, dass jeder Typ durch mehrere Messungen erfasst wird. Aus der Sicht der Analytik ist dieses Vorgehen zweckmässig, können doch die sich ergebenden Datensätze varianzanalytisch bearbeitet werden. Die Erfahrung zeigt aber, dass diskrete, sauber trennbare Typen in Vegetationsuntersuchungen sehr selten sind. Es muss also meist nach andern Ansätzen gesucht werden.

Ausgangspunkt der Untersuchungen kann auch der Standort sein. Greift man auf eine schon bestehende klimatische oder geologische Karte zurück, so können deren Einheiten vegetationskundlich untersucht werden. Auch dieses Vorgehen birgt den Nachteil, dass diskrete Standortstypen selten sind. Als unproblematisch und rationell erweisen sich beide Konzepte dann, wenn Vegetation und Standort gleichberechtigt erfasst werden (WILDI 1977). Deren Zusammenhang kann anschliessend durch eine multivariate Analyse untersucht werden. Wesentlich ist nur, dass die Zuordnung der Vegetationsdaten zu den Standortsdaten örtlich und zeitlich stimmt. Dies ist nach unserer Auffassung nur dann gegeben, wenn zu jeder Vegetationsaufnahme gleichzeitig eine Standortsanalyse vorliegt.

Ein bekanntes Versuchskonzept ist die direkte Gradientenanalyse (WHITTAKER 1967). Dabei werden entlang einer

Geländelinie (Transsekte) kontinuierlich ändernder Standortsfaktoren Stichproben erhoben. Die direkte Gradientenanalyse mit Hilfe von Transsektten hat jedoch den schwerwiegenden Nachteil, dass die Ergebnisse durch räumliche Autokorrelation beeinträchtigt werden. Man versteht darunter die Erscheinung, dass sich benachbarte Stichprobepunkte mehr gleichen, als auf Grund noch so detaillierter Standortsmessungen zu erwarten wäre. Lokale Besonderheiten verursachen dieses Phänomen. Meist beruhen sie auf zeitlich weit zurückliegenden, nicht mehr erkennbaren Ereignissen wie etwa Ueberflutungen oder Nutzungs-eingriffen. Weitere historische Zusammenhänge, die sich dem Forscher meist entziehen, sind die Einwanderung von Arten oder deren lokale, z.B. vegetative Ausbreitung. Der Schluss drängt sich auf, dass die Gradientenanalyse in ihrer bekannten Form nur Ergebnisse beschränkter Generalisierbarkeit liefern kann.

2.1.3 Zeitstudien

Bei Umweltstudien geht es öfters darum, einen durch Standortsveränderungen verursachten Wandel der Vegetation in Funktion der Zeit darzustellen. Gemäss Regel 4 in Tab. 2.1 ist dabei auch eine Referenzfläche vorzusehen, welche dem Wandel der Umwelt nicht unterworfen ist. Die Zeit ist jedoch nicht zu stoppen und sie bewirkt immer grössere oder kleinere Veränderungen, die nur schwer kontrollierbar sind. Je nach Problemlage ergeben sich daraus Schwierigkeiten unterschiedlicher Tragweite. GREEN (1979) unterscheidet fünf Fälle zunehmenden Schwierigkeitsgrades, die sich aus drei Randbedingungen ableiten lassen und die in der Pflanzenökologie allesamt aktuell sind (Tab. 2.3). Optimale Voraussetzungen (optimale Ursache-Wirkungstudie, Typ 1) ergeben sich, (i) wenn eine Standortsveränderung erst noch erwartet wird, (ii) wenn man weiss, wann und wo sie stattfinden wird sowie (iii) wenn es Referenzflächen gibt, die von der Veränderung sicher nicht betroffen sein werden. Die dritte dieser Bedingungen ist in Zeitstudien nicht wirklich zu erfüllen. Beim Typ 2 (Tab. 2.3) fehlen Referenzflächen, sodass die rein zeitlich verursachte

Veränderung von der standörtlich bedingten nicht sicher unterschieden werden kann. In Zeitstudien ist dies der günstigste Fall. Beim Typ 3 muss ein Ueberwachungsprojekt angelegt werden. Mit entsprechendem Aufwand an Stichproben muss hier erst noch herausgefunden werden, wann und wo die Veränderungen auftreten. Problemen vom Typ 4 ist allenfalls mit einer Analyse der Struktur der Pflanzengesellschaften beizukommen. Der Typ 5 entzieht sich naturwissenschaftlicher Ergründung weitgehend. Leider ist er in der Praxis gar nicht so selten vorzufinden.

2.2 Stichprobenverfahren

In diesem Kapitel werden einige statistische Stichprobenverfahren vorgestellt. Die klassische Schule Brau-Blanquet bedient sich solcher nicht. Wir sind uns im klaren, dass die Ansichten über deren Zweckmässigkeit auseinandergehen. So halten MUELLER-DOMBOIS und ELLENBERG (1974) Seite 36 fest: "Die Beispiele zeigen, dass Unterteilungen mit Hilfe objektiver Methoden nur mit einem sehr dichten Stichprobennetz entdeckt werden können, was für die meisten Anwendungen nicht praktikabel ist" (Uebersetzung). Im völligen Gegensatz dazu meinen DAGET und GODRON (1982), Seite 13: "Wenn die Stichprobenverteilung nicht probabilistisch ist, genügt es festzustellen und zur Kenntnis zu nehmen, dass alle weiteren Tests nichts beweisen können, was die untersuchte Gesamtpopulation betrifft" (Uebersetzung). Der von MUELLER-DOMBOIS und ELLENBERG (1974) erwähnte Schwachpunkt zufälliger Stichprobenverfahren ist für das dort zitierte Beispiel sicher begründet. Er lässt sich jedoch durch ein geeigneteres Stichprobenkonzept weitgehend vermeiden. Generell sind Aussagen älteren Datums über die Zweckmässigkeit probabilistischer Erhebungen im Lichte neuerer Erkenntnisse zu überprüfen. Nur nach dem Zufallsprinzip erhobene Stichproben ermöglichen es, die Fehlerhaftigkeit von Rückschlüssen auf die Grundgesamtheit zu quantifizieren (ZÖHRER 1980). Der ursprünglich als schwerwiegend erachtete Mehraufwand ist deshalb meist gerechtfertigt.

In jüngerer Zeit hat sich die Geographie vermehrt dem Entwurf räumlicher Stichprobenverfahren angenommen (z.B. OSTHEIDER 1978). Dabei wird vor allem darauf geachtet, dass das resultierende Datenmaterial für die Grundgesamtheit repräsentativ ist. In Anbetracht der Ausdehnung und Vielfalt der Untersuchungsobjekte ist es in der Pflanzensoziologie ebenso bedeutsam, dass die Stichprobenzahl (d.h. die Zahl der Aufnahmen) klein bleibt. Andernfalls ist die Realisierbarkeit vieler Projekte von vornherein in Frage gestellt. Einige neuere vegetationskundliche Untersuchungen verwenden statistische Stichprobenpläne mit gutem Erfolg (ORLOCI und STANEK 1979, GRABHERR 1985). Wir erläutern hier die als notwendig erachteten Planungsschritte anhand eines Beispiels, das sich an die Arbeit von GRUENIG (1977) anlehnt. Ziel seiner Untersuchungen war es, die Vegetations- und Standortsentwicklung (Sukzession) auf unterschiedlichen, neu angelegten Brutinseln im Staubereich eines Flusskraftwerkes zu analysieren.

2.2.1 Abgrenzung der Grundgesamtheit

In einem probabilistischen Stichprobenplan kann es leicht vorkommen, dass eine Aufnahmefläche an eine unerwünschte Stelle zu liegen kommt (z.B. auf einen Weg, in eine nicht zu bearbeitende Vegetationseinheit usw.). Mit der Abgrenzung der Grundgesamtheit werden solche Fehlplatzierungen vermieden. Gleichzeitig wird der sachliche und räumliche Gültigkeitsbereich der Untersuchung festgelegt. Der erst anschliessend zu entwerfende Stichprobenplan soll sicherstellen, dass die Resultate nicht nur für die Aufnahmeflächen selbst, sondern für die ganze Grundgesamtheit gelten werden.

GRUENIG (1977) stand als Versuchsgebiet ein einziges, neu geschaffenes Gewässer zur Verfügung. Die Ergebnisse seiner Arbeit besitzen somit auch nur für dieses Gültigkeit. Innerhalb des Gebietes standen mehrere, zumeist nierenförmig gestaltete Inseln zur Auswahl (Abb. 2.2). Vergleichbare, bereits voll entwickelte Inseln, die als Referenzflächen hätten dienen können, gab es nicht. Die Inseln selbst bestanden aus unterschiedlichen Rohböden, deren Einfluss auf

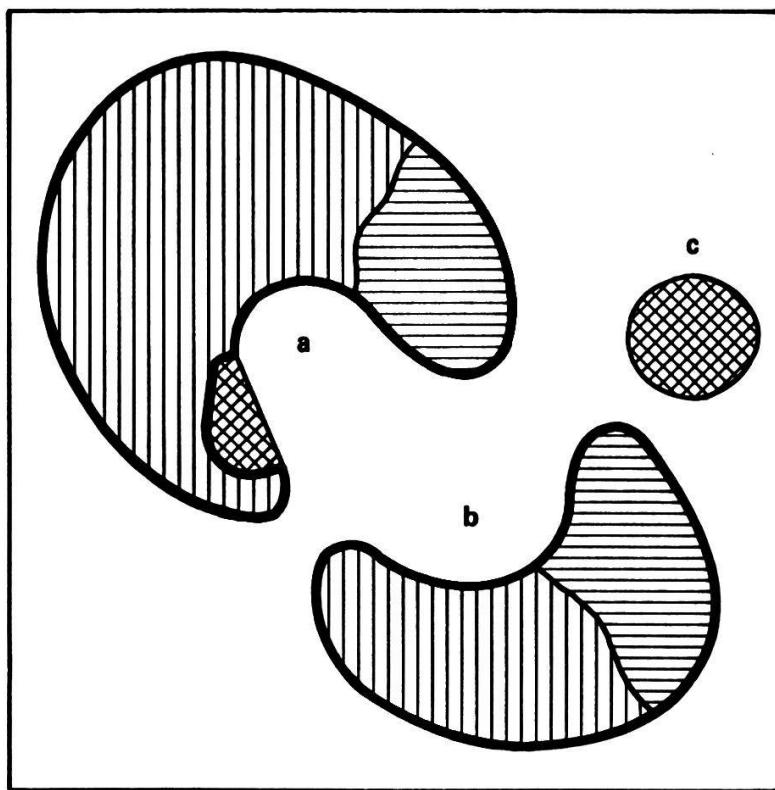


Abb. 2.2 Abgrenzung der Grundgesamtheit (in Anlehnung an GRÜNIG 1977). Humus: horizontal schraffiert. Kies: vertikal schraffiert. Blockwurfflächen (diagonal schraffiert) interessieren nicht, sodass der grösste Teil der Insel a und die ganze Insel b Gegenstand der Untersuchung werden (fette Umrandung). Insel c entfällt.

die Vegetationsentwicklung interessierte. In Abb. 2.2 wird vereinfachend angenommen, dass solche aus Grobkies, aus Humus oder aus Blockwurf zur Auswahl stehen. Nun werden alle Flächen ausgeschlossen, welche nicht Gegenstand der Untersuchung sein sollen. Da die Eigenschaften von Blockwurf nicht zu analysieren sind, entfällt Insel c sowie ein kleiner Teil der Insel a. Es kann vorkommen, dass von einem Untersuchungsgebiet weder gute Pläne noch Luftaufnahmen vorliegen. Auszuschliessende Flächen können dann erst beim Einmessen der Stichproben im Felde festgestellt werden. Wird das Ausschlussverfahren zum vornherein genau festgelegt, so tut dieses Vorgehen der Qualität des Stichprobenplanes kaum Abbruch. Um Störungen durch Randeinflüsse auszuschliessen, kann beispielsweise vereinbart werden, dass alle Stichproben, die näher als 5 Meter von einem Weg entfernt liegen, nicht zur Untersuchung gehören. Der Gültigkeitsbereich der letzteren reduziert sich dann auf Standorte abseits von Wegen.

2.2.2 Die Stratifizierung

Die Stratifizierung (auch Schichtung genannt) ist ein weiterer Planungsschritt, der den Gültigkeitsbereich der Untersuchung einschränkt (ZÖHRER 1980, PODANI 1984). Sie dient dazu, bereits vorhandenes Wissen in den Versuchsplan mit einzubeziehen und diesen damit effizienter zu gestalten. Am Beispiel der Inseln wird die Absicht klar. Hier soll ja der Einfluss verschiedener Substrate auf die Vegetationsentwicklung untersucht werden. Da nun der Aufbau der Inseln bekannt ist, sollen die Stichproben so angeordnet sein, dass jeder Substrattyp mit gleicher Intensität untersucht werden kann. Zu diesem Zwecke wird die Grundgesamtheit entsprechend unterteilt. Die Teilflächen werden Straten oder Schichten genannt. GRUENIG (1977) erwartete entscheidende Entwicklungsunterschiede auf Grund unterschiedlicher Wasserstände. Da diese durch ein Stauwehr gesteuert werden, waren sie schon bei Versuchsbeginn genau bekannt. Die Inselfläche konnte deshalb weiter nach Wasserstandsstufen stratifiziert werden. Abb. 2.3 zeigt eine Möglichkeit, bei welcher zwischen überflutbarer und nicht überflutbarer Zone unter-

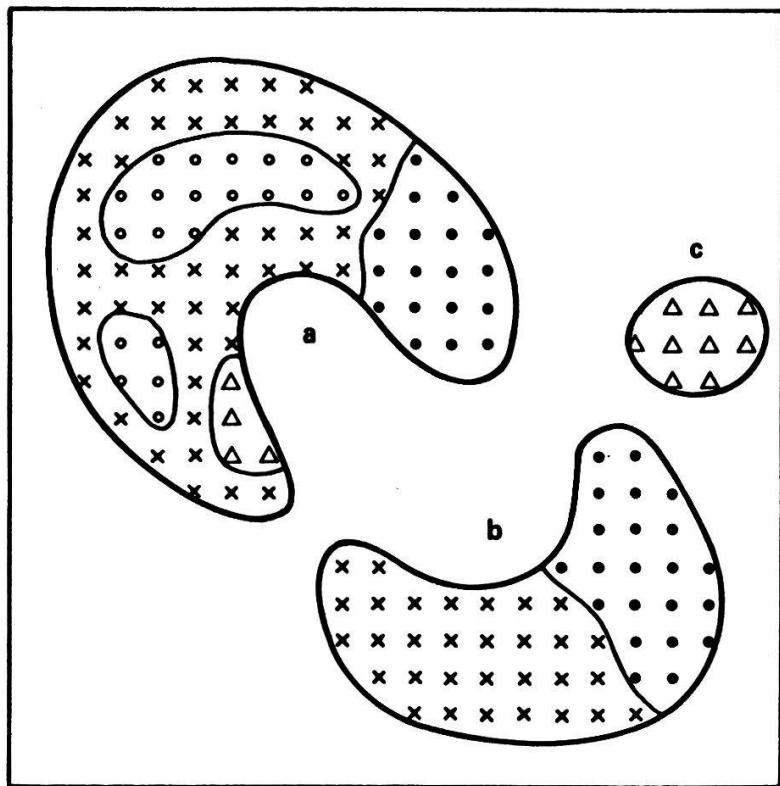


Abb. 2.3 Stratifizierung und Flächenbestimmung. Stratum 1: Kiesfläche ausserhalb der Ueberflutungszone (Kreise), 20 Flächeneinheiten. Stratum 2: Kiesfläche innerhalb der Ueberflutungszone (Kreuze), 80 Flächeneinheiten. Stratum 3: Humusfläche (Punkte), 40 Flächeneinheiten.

schieden wird. Zusammen mit den Substrattypen ergeben sich drei Straten, die teilweise auf verschiedenen Inseln liegen. Damit jede mit gleicher Intensität bearbeitet werden kann, müssen die (relativen) Flächenanteile bestimmt werden. In Abb. 2.3 wurde zu diesem Zwecke ein Raster über das Untersuchungsgebiet gelegt und die Punkte wurden ausgezählt. Die Flächen der Straten verhalten sich wie 1:4:2.

2.2.3 Der Stichprobenplan

Es sollen nun fünf Stichprobenpläne erörtert werden, wobei drei davon auf der soeben dargestellten Stratifizierung beruhen.

Als erste Möglichkeit bietet sich die **systematische, nicht stratifizierte Stichprobenerhebung** an (Abb. 2.4, oben). Sie hat die Eigenschaft, dass sich die Zahl der Stichproben etwa proportional zur Grösse der Teilflächen verhält. Tatsächlich wird unser Flächenverhältnis von 1:4:2 mit den Zahlen 5:21:9 annähernd erreicht. Dies ist im allgemeinen auch der Fall, wenn die Stichprobenzahl sehr klein ist. Die systematische Erhebung hat auch den Vorteil, dass die Einmessung der Probeflächen im Felde mit minimalem Aufwand erfolgen kann. Um die Zufälligkeit sicherzustellen, ist sowohl der Ursprung als auch die Richtung des Probenetzes probabilistisch zu ermitteln. Dieses erste Stichprobenkonzept versagt, wenn die Untersuchungsfläche periodisch auftretende Strukturen enthält (Bulten, Dünen usw.). Wenn dem so ist, muss auf Zufallsstichproben ausgewichen werden. Bei der systematischen Erfassung besteht auch die Gefahr, dass seltene Standortstypen in ungenügendem Masse vertreten sind. Sie fällt dann wegen des zu hohen Aufwandes ausser Betracht.

Sollen die Straten mit gleicher Intensität bearbeitet werden, so bietet sich die **systematische, stratifizierte Stichprobenerhebung** an (Abb. 2.4, unten). Die Netzweite ist hier stratenweise so festzulegen, dass sie proportional zur Wurzel der Stratumsfäche ist. Wir finden in unserem Beispiel ein solches von $1\frac{1}{72} : 4\frac{1}{72} : 2\frac{1}{72} = 1:2:1.414$. Mit Stichprobenzahlen von 20, 21 und 20 erreichen wir unsere Ab-

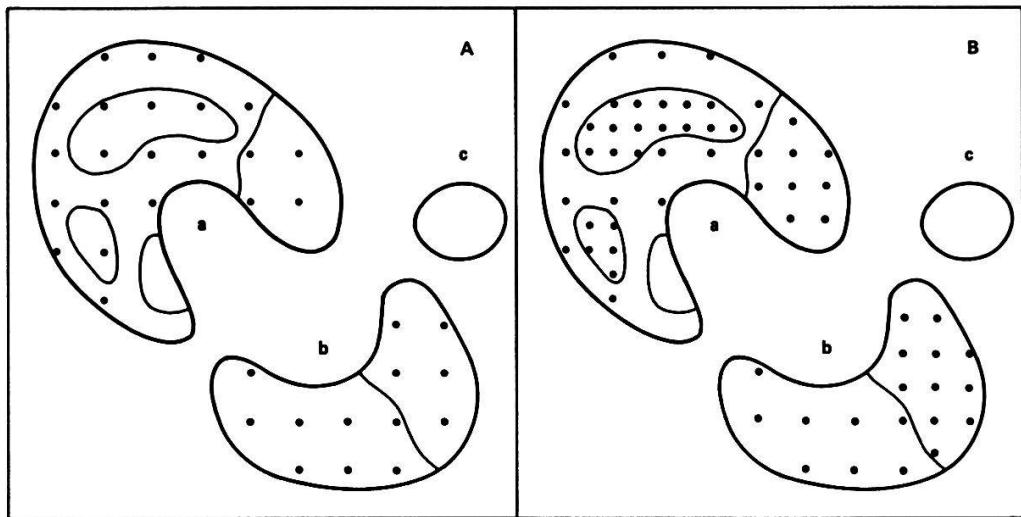


Abb. 2.4 Systematische Stichprobenpläne. A: Systematische, nicht stratifizierte Stichprobenerhebung (Stichprobenverhältnis 5:21:9). B: Systematische, stratifizierte Stichprobenerhebung (Stichprobenverhältnis 20:21:20).

sicht recht gut. Wie schon bei der nicht stratifizierten, systematischen Erhebung gilt, dass Ursprung und Richtung der Netze durch Zufall festzulegen sind. Auch hier ist wiederum zu prüfen, ob im Untersuchungsgebiet periodische Strukturen das Konzept stören.

Das allgemeinste Konzept ist in der nicht stratifizierten Zufallserhebung realisiert. Abb. 2.5, A, zeigt diese Lösung. Die Zahl der Stichproben innerhalb der Straten ist proportional zu deren Fläche, doch die zufallsbedingten Abweichungen sind generell wesentlich grösser als bei einer systematischen Erhebung (2:19:14 statt der zu erwartenden 5:20:10). Auffällig an einer echten Zufallsverteilung ist das herdenartige Auftreten vieler Punkte. Versuche, echte Zufälligkeit nach subjektiven Gesichtspunkten zu konstruieren, scheitern meistens. Bei der Bestimmung der Punktkoordinaten sind deshalb Zufallszahlen zu verwenden. Im vorliegenden Beispiel lieferte sie ein programmierbarer Taschenrechner.

Bei der stratifizierten Zufallserhebung liest man so lange Zufallskoordinaten aus einer Tabelle heraus, bis jedes Stratum seiner Fläche entsprechend repräsentiert ist (Abb. 2.5, B), oder bis alle Straten gleichmässig besetzt sind, d.h. jedes Stratum gleich repräsentiert ist (Abb. 2.5, C). Bei statistischen Tests, Varianz- oder Diskriminanzanalysen bietet dies erhebliche interpretatorische Vorteile. Für pflanzenökologische Untersuchungen ist deshalb die stratifizierte der unstratifizierten Stichprobenerhebung generell vorzuziehen.

In den hier vorgestellten Beispielen ist die Zahl der Stichproben entweder proportional zur Stratumsfäche oder sie ist in allen Straten gleich. Das sind jedoch nur die einfachsten Fälle. Befriedigen sie nicht, so sind andere Lösungen zu suchen. Solche werden z.B. bei DAGET und GODRON (1982), Seite 18 diskutiert. ZÖRER (1980) zeigt, wie sich in der Waldmesskunde die bei der Hochrechnung auftretenden Fehler für verschiedene Stichprobenpläne berechnen lassen. Seine Schlussfolgerungen bezüglich des Erhebungsaufwandes

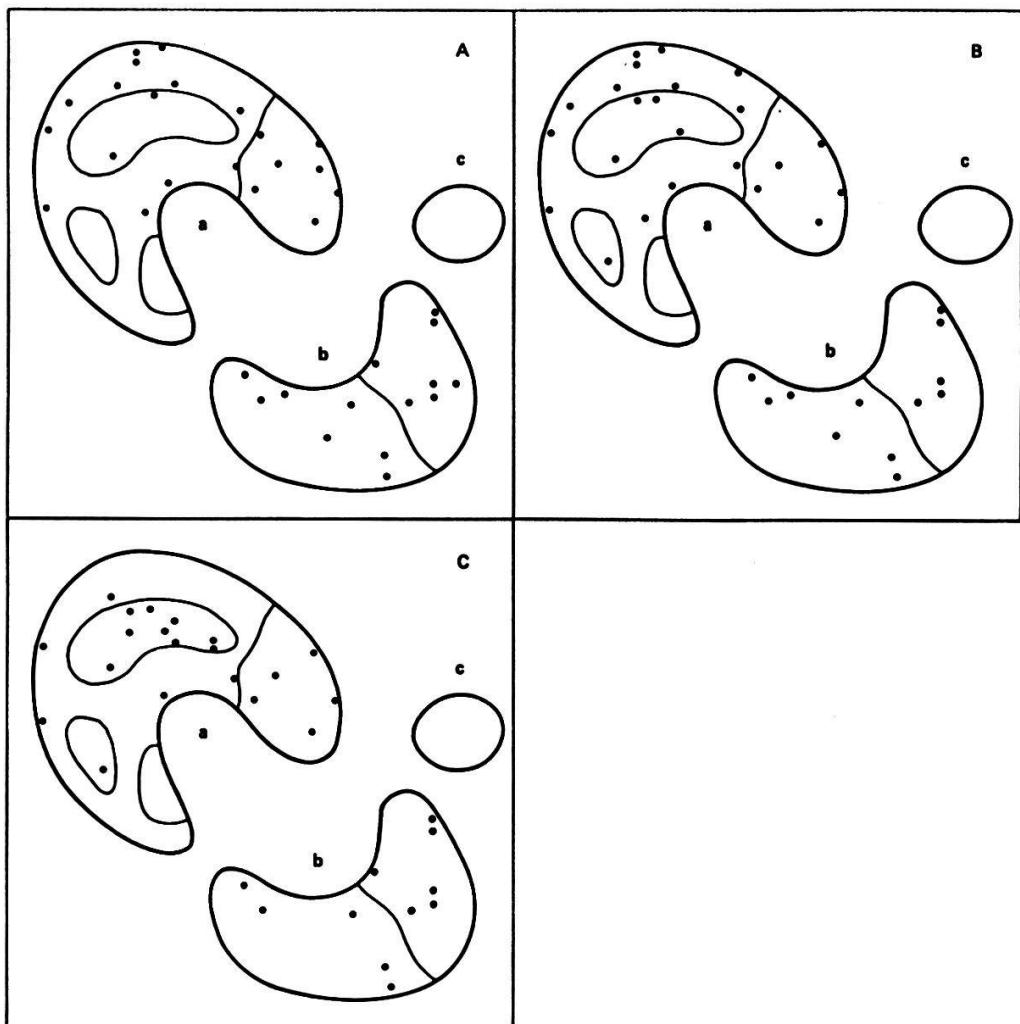


Abb. 2.5 Zufalls- Stichprobenpläne. A: Zufällige Stichprobenpläne (Stichprobenverhältnis 2:19:14). B: Stratifizierte, zufällige Plazierung mit Probenzahlen proportional zur Stratumfläche (Stichprobenverhältnis 5:20:10). C: Stratifizierte, zufällige Platzierung mit gleicher Probenzahl pro Stratum (je 10 Stichproben).

sind wohl auf viele vegetationskundliche Untersuchungen übertragbar.

Die Anwendung statistischer Stichprobenpläne ist heute im Zusammenhang mit kleinräumigen Untersuchungen und Experimenten üblich geworden. Für grossräumige Uebersichten sind uns hingegen nur wenige Beispiele bekannt (ORLOCI und STANEK 1979, GRABHERR 1985). Die Entwicklung neuer pflanzensoziologischer Systeme müsste aber auf einem solchen Vorgehen beruhen.

3. Skalierung und Transformation

In der Pflanzenökologie will man biologische Fakten mit abiotischen Umweltverhältnissen in Beziehung setzen. Es müssen also Daten sehr unterschiedlicher Form auf Zusammenhänge hin untersucht werden. Auf biologischer Seite fallen Informationen an über die artmässige Zusammensetzung von Aufnahmen, über Gruppenstrukturen, Unterschiede bezüglich Lebensformen, physiologischer Verhaltensweisen usw. Bei den Standortsfaktoren geht es um Temperatur-, Licht-, Wasserhaushalts- und Nährstoffmessungen, aber auch um qualitative Grössen wie Gesteinsart, Bodentyp usw. Man mag sich nun wundern, dass derart objektiv gewonnene Daten durch eine doch willkürlich gewählte Transformation verändert werden sollen. Dass aber solche Bedenken unbegründet sind, ist leicht zu zeigen. Jedes Messresultat ist bekanntlich vom verwendeten Messinstrument und dessen Skala abhängig. Verschiedene Temperaturskalen sind nur ein Beispiel:

Celsiusskala,	$^{\circ}\text{C}$:	0,	10,	20
Kelvinskala,	$^{\circ}\text{K}$:	273,	283,	293

Unter Verzicht auf statistische Ueberlegungen kann also festgestellt werden, dass Skalierungen und Transformationen vorzunehmen sind, um Daten miteinander vergleichbar zu machen. Das gilt immer dann, wenn unterschiedliche Methoden oder voneinander abweichende Skalen zur Erfassung der Merkmale verwendet wurden. In der Pflanzensoziologie ergibt sich diesbezüglich eine spezielle Situation. Die verschiedenen Pflanzenarten werden in der Regel mit der gleichen Schätz- oder Messskala quantifiziert. Deshalb sind sie als Merkmale in erster Näherung auch ohne Transformation durchaus vergleichbar. Von dieser Annahme wird je nach Untersuchungsgegenstand und -ziel graduell abgewichen. So sind in manchen Fällen häufige und seltene Arten durch Transformation vergleichbar zu machen. Noch allgemeiner betrachtet geht es darum, Quantitäten ihrer pflanzensoziologischen Bedeutung entsprechend zu gewichten.

3.1 Datentypen

Es bewährt sich, verschiedene Datentypen zu unterscheiden. Zahlreiche der später zu besprechenden Analysemethoden eignen sich nur für einen bestimmten Datentyp. Letzterer schränkt daher die Auswertemöglichkeiten ein. Das hat Konsequenzen, denn pflanzensoziologische Daten können je nach Auffassung verschiedenen Typen zugreordnet werden. Wir halten uns nachfolgend an die verhältnismässig einfache Gliederung von BAHRENBERG und GIESE (1975):

3.1.1 Nominaldaten

Nominaldaten sind rein qualitative, oft auch als kategoriell bezeichnete Daten. Beispiel ist der Standortsfaktor "Boden- typ", Tb. Er kann die Werte [(Podzol), (Braunerde), (Pseudogley)] annehmen. Echte Nominaldaten lassen nicht alle Operationen zu. Erlaubt sind die Aussagen:

$Tb_1 = Tb_2$ (Aufnahmen 1 und 2 besitzen gleichen Bodentyp)

$Tb_1 \neq Tb_2$ (Aufnahmen 1 und 2 besitzen verschiedenen Bodentyp)

Nicht erlaubt sind:

$Tb_1 > Tb_2$
 $Tb_1 < Tb_2$

Diese beiden Operationen sind auch nicht statthaft, wenn z.B. Tb_1 eine bessere Bodenentwicklung aufweist als Tb_2 . Will man einen solchen Sachverhalt ausdrücken, so ist für Tb ein anderer Datentyp zu postulieren (s. unten). Unerlaubt sind natürlich auch alle arithmetischen Grundoperationen, also

$$\begin{array}{l} \mathbf{Tb}_1 + \mathbf{Tb}_2 \\ \mathbf{Tb}_1 - \mathbf{Tb}_2 \\ \mathbf{Tb}_1 * \mathbf{Tb}_2 \\ \mathbf{Tb}_1 / \mathbf{Tb}_2 \end{array}$$

Ist man sich über die Art des Datentyps im unklaren, so kann sicherheitshalber davon ausgegangen werden, dass Nominaldaten vorliegen. Ein solch kritisches Beispiel sind die Himmelsrichtungen. Die Operationen " $=$ " und " \neq " sind dabei zulässig, formal auch " $<$ " und " $>$ ". Dabei besteht jedoch Mehrdeutigkeit bei 0° und 360° . Die Wahl des Ursprunges und die Messrichtung sind biologisch gesehen völlig willkürlich. Bei vielen Anwendungen ist es daher sinnvoller, das Merkmal zu diskretisieren und die sich ergebenden Klassen [N,E,S,W] nominal zu behandeln.

3.1.2 Ordinaldaten (Rangdaten)

Wie der Name andeutet, gibt es bei Ordinaldaten unter den Ausprägungen eines Merkmals eine Rangordnung. Ein Beispiel wären Grundwasserstände, Tg , welche charakterisiert werden durch die Werte [(tief), (mittel), (hoch)]. Erlaubt sind dabei Operationen wie

$$Tg_1 = Tg_2 \text{ (Gleiche Grundwasserstandskategorie)} \\ \text{für Aufnahme 1 und 2)}$$

$$Tg_1 \neq Tg_2 \\ Tg_1 > Tg_2 > Tg_3 \dots \text{ usw.}$$

Meist werden zur Bezeichnung der Merkmalsausprägungen nicht Buchstaben, sondern Rangzahlen verwendet, statt [A,B,C] also [1,2,3]. Trotzdem sind die arithmetischen Grundoperationen streng genommen nicht erlaubt. Gerade in der Pflanzensoziologie wird aber oft bewusst gegen diese Regel verstossen, um die Daten einem noch weiteren Spektrum von Methoden zugänglich zu machen. Dabei ist natürlich zu überlegen, ob die so eingeführten Fehler toleriert werden können. Das Grundwasserbeispiel zeigt die Problematik. Es gelte:

Grundwasserstands-kategorie	A	B	C
Bereich in cm unter Bodenoberfläche	>100	20-100	0-20

$$Tg_1 = 15 = C \text{ (Rang 3)}$$

$$Tg_2 = 200 = A \text{ (Rang 1)}$$

Die Berechnung des Mittelwertes erfordert die Verwendung arithmetischer Grundoperationen:

$$\bar{Tg} = 1/2(Tg_1 + Tg_2) = 1/2(15+200) = 107.5 \text{ (Typ A)}$$

Mit den Rangzahlen wird ein falsches Resultat errechnet:

$$\bar{Tg} = 1/2(1+3) = 2 \text{ (Typ B)}$$

Diese Überlegungen gelten z.B. für die Artmächtigkeiten, die in der Vegetationskunde am weitesten verbreiteten Ordinaldaten. Die Symbole der Schule Braun-Blanquet, $[r, +, 1, 2, 3, 4, 5]$, lassen sich problemlos durch die Rangzahlen $[1, 2, 3, 4, 5, 6, 7]$ ersetzen. Weitergehende Transformationen dieser Skala sollen in den folgenden Kapiteln behandelt werden.

3.1.3 Metrische Daten

Metrische Daten sind Messdaten, die einen Ursprung (Nullpunkt) besitzen. Grundsätzlich gilt, dass alle arithmetischen Grundoperationen erlaubt sind. Es ist jedoch nicht immer leicht, den Nachweis zu erbringen, dass tatsächlich metrische Daten vorliegen. Schwierigkeiten ergeben sich insbesondere bei abgeleiteten Größen, z.B. bei den später zu behandelnden Ähnlichkeitsmassen. Bei diesen ist zu prüfen, ob die Dreiecksungleichung erfüllt ist. Sind drei Werte gegeben, welche die Abstände dreier Punkte A, B, C im Raum messen, so ist zu prüfen, ob

$$d(A,C) + d(B,C) \geq d(A,B)$$

Diese Formel besagt, dass die Summe der Abstände zweier beliebiger Punkte gleich oder grösser sein muss, als der dritte Abstand. Wir betrachten zwei Beispiele:

Dreieck	1	2
$d(A,B)$	3	3
$d(A,C)$	2	1
$d(B,C)$	2	5

Die Prüfung ergibt für Dreieck 1

$$2 + 2 \geq 3, \quad 2 + 3 \geq 2, \quad 2 + 3 \geq 2.$$

Damit ist die Dreiecksungleichung erfüllt. Bei Beispiel 2 hingegen ist dies nicht der Fall, denn $3 + 1 < 5$! Es muss aber betont werden, dass viele auf metrischen Daten basierende Verfahren auch mit nicht metrischen durchführbar sind. Dabei muss ein öfters schwer abschätzbarer Fehler in Kauf genommen werden. Die Ordinationsmethode nach BRAY und CURTIS (1957) ist ein solches Beispiel.

3.2 Skalierung von Artmächtigkeiten

Genereller formuliert wird hier die Umwandlung von Datentypen behandelt. Die Substitution traditioneller Artmächtigkeitssymbole durch Zahlenwerte spielt in der Pflanzensoziologie seit der Einführung numerischer Methoden eine sehr wichtige Rolle. Auf diese Weise wird es erst möglich, mit traditionellen Methoden erhobene Vegetationsdaten durch metrisch arbeitende Verfahren zu analysieren.

Schon im vorhergehenden Kapitel wurde darauf hingewiesen, dass die Dominanz- und Abundanzskalen trotz der gleichzeitigen Verwendung von Buchstaben und Zahlen nichts anderes sind als Rangskalen. Je nach Auffassung variieren

darin die Anzahl der Klassen, sowie die Klassengrenzen. Verschiedene Möglichkeiten sind bei MUELLER-DOMBOIS und ELLENBERG (1974) aufgeführt (Skalen von Domin Krajina, Braun-Blanquet, Daubenmire usw.). Weite Verbreitung hat neuerdings die Skala von LONDO (1975) erfahren.

Dem Problem des Ersatzes der verschiedenen Symbole durch Zahlenwerte hat VAN DER MAAREL (1979) eine längere Abhandlung gewidmet. Er stellt fest, dass sich viele Vorschläge auf einige Grundsätzliche Lösungen reduzieren lassen. Die einfachste Lösung besteht darin, statt des Codes seinen jeweiligen Rang innerhalb der Skala zu verwenden (Tab 3.1). Damit liegen nun Ordinaldaten vor, welche aber noch immer nicht allen numerischen Verfahren zugänglich sind. Echt metrische Daten wären eigentlich die ursprünglichen Deckungswerte in Prozenten. Man kann also den Code durch den Mittelwert seiner Deckungsklasse ersetzen. Tab. 3.1 zeigt dies für die erweiterte Skala von Braun-Blanquet. Bei letzterer sind niedrige Deckungswerte quantitativ nicht genau definiert, sodass mit behelfsmässigen Werten gearbeitet werden muss.

Rein gefühlsmässig wird meist angenommen, dass die eigentlichen Deckungsprozente nicht ihrer ökologischen Bedeutung entsprechen. Ueberlegungen wie die folgende werden angestellt: Ob eine Art mit recht grosser oder aber kleiner Deckung auftritt ist weniger wichtig als die Frage, ob sie überhaupt vorkommt. Sollen Vegetationsdaten so ausgewertet werden, dass die resultierenden Gruppen- und Gradientenstrukturen standörtlichen Unterschieden entsprechen, so muss für die Skalierung eine Lösung gefunden werden, die einerseits der Präsenz oder Absenz einer Art hohe Bedeutung beimisst, andererseits aber auch die Deckungswerte nicht ganz vernachlässigt. Dieser Idee entspringt die Skala von Braun-Blanquet, die als transformierte Deckungsskala verstanden werden kann. VAN DER MAAREL (1979) hat gezeigt, dass sich durch eine einfache Transformation der Rangzahlen x viele praktisch verwendbare Gewichtungen erzielen lassen. Die transformierten y -Werte ergeben sich nach der Formel

Tabelle 3.1

Potenztransformation der Skala Braun-Blanquet nach VAN DER MAAREL (1979).

Skala Braun- Bl.	Deck- ung, %	Ordinal- skala x	Transformation $y = x^w$					
			w=0	w=0.25	w=0.5	w=1	w=2	w=4
leer	0.0	0	0.0	0.00	0.00	0	0	0
r	()	1	1.0	1.00	1.00	1	1	1
+	0.1	2	1.0	1.19	1.41	2	4	16
1	5.0	3	1.0	1.32	1.73	3	9	81
2m		4	1.0	1.41	2.00	4	16	256
2 2a	17.5	5	1.0	1.50	2.24	5	25	625
2b		6	1.0	1.57	2.45	6	36	1296
3	37.5	7	1.0	1.63	2.65	7	49	2401
4	62.5	8	1.0	1.68	2.83	8	64	4096
5	87.5	9	1.0	1.73	3.00	9	81	6561

$$y = x^w$$

Der Exponent w kann je nach der angestrebten Lösung variiert werden. Die Wirkung von w kann leicht aus Tab. 3.1 ersehen werden. In Abweichung von der Skala Braun-Blanquet verwendet van der Maarel eine feinere Unterteilung der Artmächtigkeit 2, sodass insgesamt 10 Klassen entstehen.

Tab. 3.1 berücksichtigt einige praktisch wichtige Werte für den Exponenten w . Mit $w = 0$ erhält man eine Binärskala, das heisst reine Präsenz-Absenz Daten. Wohl die meisten Pflanzensoziologen interpretieren die Skala Braun-Blanquet so, wie sie sich durch eine Transformation der Rangwerte mit einem w von 0.25 bis 0.5 ergäbe. Mit $w = 1$ reproduziert man die ursprüngliche Rangskala. Bei $w = 2$ erhält man mit geringfügigen Abweichungen die ursprünglichen Deckungswerte der Arten. Bei $w < 2$ erhalten hohe Deckungswerte noch mehr Gewicht im Vergleich zu den niedrigen. Solcherart durchgeföhrte Analysen berücksichtigen speziell die dominanten Arten.

An dieser Stelle ist anzumerken, dass geordnete Vegetations-tabellen oft nach ihrem rein optischen Erscheinungsbild beurteilt werden. Symbole einerseits und Leerstellen andererseits sollten darin möglichst konzentriert auftreten. Die Erfahrungen bestätigen die Vermutung, dass sich solche Ergebnisse am besten mit Präsenz-Absenz Transformation erzielen lassen.

3.3 Transformationen

3.3.1 Transformation von Einzelwerten

Im vorhergehenden Kapitel wurde gezeigt, wie sich die verschiedenen Möglichkeiten der Umwandlung von Artmächtigkeiten in Zahlenwerte auf unterschiedliche Transformation der Rangzahlen zurückführen lassen. Bei der Analyse von Standorts-faktoren tritt ein ähnliches Problem auf. Hier geht es darum, die bei der Messung im Felde oder im Labor angewandte

Skala den Erfordernissen der Problemstellung und der Analyse anzupassen. Das ist immer dann angezeigt, wenn verschiedene Standortsfaktoren untereinander sowie mit der Vegetation korreliert werden sollen. Es kommt aber auch vor, dass mit Skalen gemessen wird, die bereits zweckmäßig transformiert sind. pH-Werte sind ein solches Beispiel. Die Mehrzahl europäischer Pflanzengesellschaften kommt auf Substraten mit dem pH-Bereich von etwa 3.5 bis 7.5 vor. Der Versuch, anhand eines Beispiels für transformierte und untransformierte Daten einen Mittelwert zu berechnen, ergibt folgendes Resultat:

Aufnahme	1	2	3	4	Summe	Mittel
pH-Werte	4.0	5.0	4.0	7.0	20.0	5.0
Konz. (H^+)	10^{-4}	10^{-5}	10^{-4}	10^{-7}	$2.101 \cdot 10^{-4}$	$5.25 \cdot 10^{-5}$

Als Mittelwert der untransformierten H^+ -Konzentration erhält man also $5.25 \cdot 10^{-5}$ oder umgerechnet pH 4.28. Dieses Ergebnis liegt nahe der unteren Grenze der in unserem Beispiel angenommenen Werte. Transformiert ergibt sich aber pH 5.0, was wohl eher als sinnvoll angenommen werden kann.

Ob eine Transformation angezeigt ist, lässt sich auch auf etwas formalerem Wege abklären. Dazu betrachtet man die Verteilung der Messwerte. Sind die Daten metrisch, so müssen sie zunächst in Klassen eingeteilt werden. Der Wahl der Klassengrenzen kommt dabei eine wesentliche Bedeutung zu. Dieses Thema wird in vielen Lehrbüchern der Statistik ausführlich behandelt (vgl. z.B. BARTEL 1974). Abb. 3.1 stellt anhand eines Beispiels zwei verschiedene Verteilungen dar. Beide Diagramme entstanden aus 11 pH-Werten (hier nicht gezeigt) mit den Klassengrenzen 3.5, 4.5, 5.5, 6.5, 7.5 und 8.5. Im Diagramm links liegt eine fast symmetrische Verteilung vor. Der Median (häufigster Wert) liegt bei pH 6.0, der Mittelwert (arithmetisches Mittel) wie leicht nachzuprüfen ist bei pH 6.1. Der Mittelwert liefert also auch einen guten Näherungswert für den Median, was ebenfalls auf die gute Symmetrie hinweist. Anders im schiefen Fall rechts. Hier liegt der Median bei pH 5.0, das arithmetische Mittel bei pH

5.64. Eine Transformation soll nun die Symmetrie wesentlich verbessern.

Häufig angewandte Transformationen sind die Wurzeltransformation

$$x' = x^{1/2}$$

sowie die Logarithmustransformation

$$x' = \log x, \text{ bzw. } x' = \log(x+1)$$

Die Wurzeltransformation zeigt eine deutlich schwächere Wirkung als die Logarithmustransformation. Da die Wurzel negativer Messwerte imaginär ist, muss deren Auftreten durch eine Verschiebung des Ursprunges vermieden werden. Dabei ist jeder Messwert um einen konstanten Betrag derart zu erhöhen, dass der kleinste auftretende Wert grösser als Null ist. Aehnlich liegen die Probleme bei der Logarithmustransformation. Logarithmen negativer Werte sind bekanntlich undefiniert, sodass nur positive Messungen berücksichtigt werden können.

Um die Wirkung beider Transformationen zu demonstrieren, ist das Säulendiagramm der Abb. 3.1, rechts, mit wurzel- und logarithmustransformierter x-Achse in Abb. 3.2 verändert dargestellt. Die Zwischenräume der Klassen sind dabei verschieden breit geworden. Aus der Wurzeltransformation resultiert noch immer eine leicht schiefe Verteilung. Die Logarithmustransformation hingegen ergibt ein fast perfektes Bild der Symmetrie.

Der Zusammenhang vieler Standortsfaktoren ist nicht linear. In den seltensten Fällen linear ist gar derjenige zwischen Vegetation und Standort. Die meisten zur Zeit verfügbaren Analysemethoden beruhen jedoch auf der Annahme, dass die Korrelationen annähernd linear seien (vgl. z.B. FEWSTER und ORLOCI 1983). Wo dies nicht zutrifft, kann nun versucht werden, mit transformierten Messwerten zu arbeiten. Ob damit im einzelnen Falle die gewünschte Wirkung erzielt wird, kann

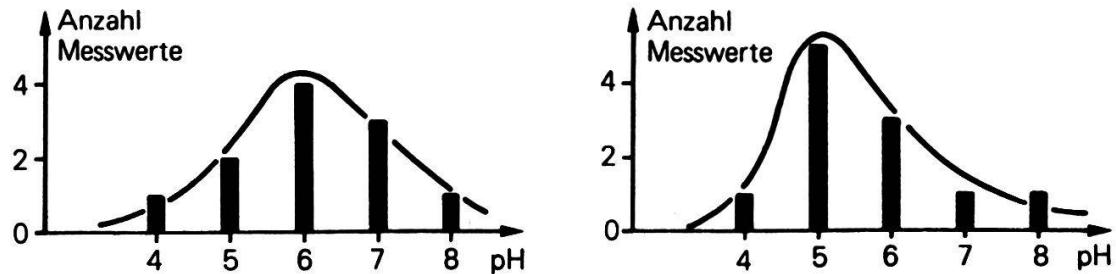


Abb. 3.1 Beispiel einer angenähert symmetrischen Verteilung eines Standortsfaktors (links), sowie einer schießen Verteilung (rechts).

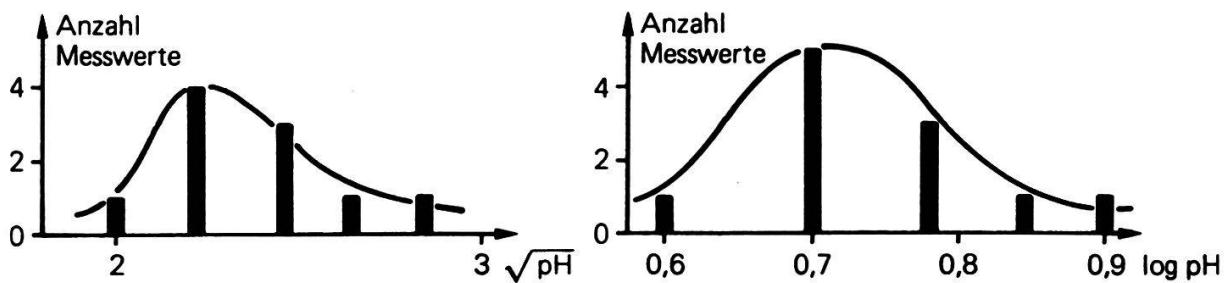


Abb. 3.2 Histogramme der transformierten Messwerte, welche zur Konstruktion der Abb. 3.1, rechts verwendet wurden. Histogramm links: Wurzeltransformation. Histogramm rechts: Logarithmustransformation.

grafisch leicht überprüft werden. Tab. 3.2 und Abb. 3.3 geben ein Beispiel. Innerhalb einer Aufnahmeserie entlang eines Vegetationsgradienten seien die pH-Werte und der Grundwasserstand x in cm unter der Bodenoberfläche gemessen worden. Um den Zusammenhang zwischen den gemessenen Standortsfaktoren zu prüfen, wurden die Wertepaare in Abb. 3.3 in einem rechtwinkligen Koordinatensystem gegeneinander aufgetragen. Eine willkürlich durch den Punkteschwarm gezogene Linie weist auf den engen, jedoch nicht linearen Zusammenhang zwischen pH- Wert und Grundwasserstand hin. In Abb. 3.3, rechts, werden dagegen die Logarithmen der Grundwasserstände verwendet. Sie weisen eine fast perfekte lineare Korrelation mit den pH-Werten auf. Eine Mehrheit der nachfolgend behandelten Methoden würde die Daten in solcherart transformierter Form wirkungsvoller analysieren.

3.3.2 Vektortransformationen

Mit der Transformation der einzelnen Messwerte kann die Häufigkeitsverteilung der Standortsfaktoren, der Arten und eventuell auch der Aufnahmen verändert werden. Oft ist es jedoch nötig, verschiedene Mess- und Zählskalen einander anzupassen. Man erreicht damit, dass jeder Wert einer Vegetations- oder Standortstabelle bezüglich seiner Grösse und Streuung mit jedem andern direkt vergleichbar wird. Dazu ist es unerlässlich, ganze Zeilen oder Spalten (Vektoren) einer Vegetationstabelle zu transformieren.

Eine erste Möglichkeit bietet die **Bereichsanpassung**. Zuerst bestimmt man den grössten und den kleinsten Wert eines Vektors (d.h. einer Aufnahme, einer Art). Anschliessend wird das so gefundene Minimum von jedem Einzelwert subtrahiert und das Ergebnis durch die Differenz der Extremwerte dividiert:

$$x'_i = \frac{x_i - \min(x)}{\max(x) - \min(x)}$$

für jedes Element i . Die transformierten Werte x'_i haben den

Tabelle 3.2

Logarithmustransformation von Standortsfaktoren

Aufn.	1	2	3	4	5	6	7	8	9	10
pH	4.8	5.1	5.2	5.3	5.3	5.2	4.9	5.4	4.7	5.5
Grundwasserstand x	35	22	15	15	13	18	26	11	45	9
log x	1.54	1.34	1.18	1.18	1.11	1.26	1.41	1.04	1.65	0.95

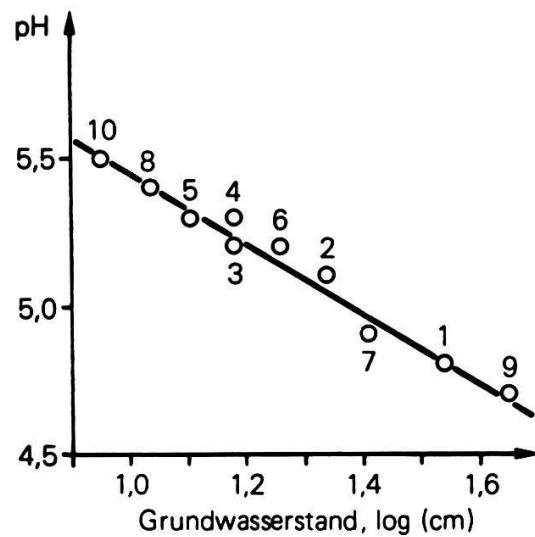
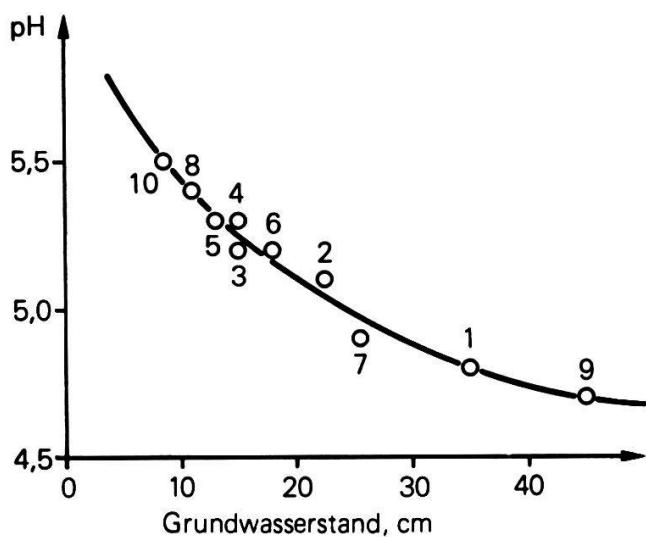


Abb. 3.3 Direkte Ordination der Aufnahmen in Tabelle 3.2.:
Untransformiert (links), transformiert (rechts).

Bereich 0 bis 1. Tabelle 3.3 gibt ein Rechenbeispiel. Die Aufnahmen 1 und 2 (Tab. 3.3, A) bilden die Ausgangsdaten, in B bis E sind sie transformiert. Aus dem Beispiel (Tab. 3.3, B) ist auch die grösste Schwäche der Bereichsanpassung erkennbar, indem ein einziger sehr grosser oder sehr kleiner Wert das Resultat entscheidend zu verändern vermag. Die 6 in Aufnahme 2 ist ein solcher Fall. Sie bringt Art 2 auf den transformierten Wert 0.2, während Art 3 der Aufnahme 1 bei gleichem Ausgangswert auf 0.5 kommt. Die Bereichsanpassung ist daher eine mit Vorsicht zu verwendende Transformation.

Eine zweckmässigere Lösung besteht in der Angleichung der Summen der Vektorelemente. Dazu wird jeder Wert durch die Vektorsumme dividiert:

$$x'_i = x_i / \sum_i x_i$$

Die Wirkung auf das bereits besprochene Beispiel ist in Tabelle 3.3, Teil C, ersichtlich. Eine solche Transformation wurde zum Beispiel von WHITTAKER (1952) für Vegetationsdaten vorgeschlagen. Er verwendet sie im Zusammenhang mit einem speziellen Aehnlichkeitsmass, der Absolutwertfunktion (Kap. 4). In der Tat ist es sinnvoll, Vektortransformationen so zu wählen, dass sie dem Konzept später zu verwendender Aehnlichkeitsmasse entsprechen. Whittakers Mass ist nichts anderes als die Summe der Differenzen zwischen den Artmächtigkeiten zweier Aufnahmen. Dementsprechend ist es plausibel, die Transformation ebenfalls auf den arithmetischen Operationen erster Ordnung beruhen zu lassen. In Kapitel 4 wird zu zeigen sein, dass die Summe der Quadrate der Elemente eine grösse Bedeutung erlangt hat als die Summe der nicht quadrierten Elemente. Die entsprechende Transformation lautet dann

$$x'_i = x_i / (\sum_i x_i^2)^{1/2}.$$

Damit wird nicht die Summe, wohl aber die Länge der Vektoren (Arten, Aufnahmen) vereinheitlicht - im Sinne des Satzes von Pythagoras:

Tabelle 3.3

Wirkung verschiedener Transformationen auf die Ausgangsdaten

A: Bereichsanpassung (B), Angleichung der Summen (C), Normalisierung (D), Standardisierung (E).

A

Art	1	2	3	4	5	xmin	xmax	xmax-xmin	Σx	Σx^2	$[\Sigma x^2]^{1/2}$
Aufn. 1	3	1	2	1	3	1	3	2	10	24	4.90
Aufn. 2	1	2	1	1	6	1	6	5	11	43	6.48

B

Art	1	2	3	4	5
Aufn. 1	1.0	0	0.5	0	1.0
Aufn. 2	0	0.2	0	0	1.0

C

Art	1	2	3	4	5
Aufn. 1	0.33	0.10	0.20	0.10	0.33
Aufn. 2	0.091	0.181	0.091	0.091	0.545

D

Art	1	2	3	4	5
Aufn. 1	0.612	0.204	0.408	0.204	0.612
Aufn. 2	0.152	0.304	0.152	0.152	0.914

E

Art	1	2	3	4	5
Aufn. 1	1.118	-1.118	0.0	-1.118	1.118
Aufn. 2	-0.618	-0.103	-0.618	-0.618	1.959

$$(\sum_i x_i'^2)^{1/2} = 1.$$

Die Wirkung anhand unseres schon verwendeten Beispiels ist wiederum in Tabelle 3.3 (Teil D) dargestellt. Betrachten wir nun Art 4, welche in beiden Aufnahmen den Ausgangswert 1 aufweist. Ihr zufolge erfahren bei dieser **Normalisierung** genannten Transformation alle Werte artenärmer Aufnahmen eine ausgeprägte Mehrgewichtung. Seltene Arten gewinnen darin im Vergleich zu den Ursprungswerten ausgeprägt an Bedeutung.

Die soweit erwähnten Vektortransformationen haben eine Angleichung der Skalierungen zum Ziel. Gerade Standortsfaktoren unterscheiden sich nicht nur in ihrer Masseinheit, sondern oft ganz erheblich in ihrer Streuung. Trifft dies zu, so ist ein sinnvoller Vergleich zweier Vektoren nur nach einer **Standardisierung** möglich. Auf diese greift man letztlich auch zurück, wenn man sich des weit verbreiteten Korrelationskoeffizienten bedient (vgl. Kapitel 4). Standardisierte Vektoren zeichnen sich durch einen Mittelwert von 0 und eine Standardabweichung von 1 aus. Die Formel lautet:

$$x_i' = \frac{x_i - \bar{x}}{s}$$

Jedes Element x_i wird um den Mittelwert \bar{x} vermindert und durch die Streuung dividiert. Die Streuung s , bzw. deren Quadrat s^2 (die Varianz), ergibt sich aus der Summe der quadrierten Differenz jedes Elementes vom Mittelwert, dividiert durch die Anzahl der Elemente n minus 1:

$$s^2 = 1/(n-1) \sum_i (x_i - \bar{x})^2$$

Diese Formel ist zwar einfach, doch sie ist für die Berechnung sehr unrationell. Direkter geht das in der folgenden Art:

$$s^2 = \frac{\sum x^2 - (\sum x)^2 / n}{n-1}$$

Betrachten wir wieder Tabelle 3.3, so ergibt sich für die Aufnahmen 1 und 2

$$s_1^2 = \frac{24 - 100 / 5}{4} = 1.0 ; s_1 = 1.000$$

$$s_2^2 = \frac{43 - 121 / 5}{4} = 4.7 ; s_2 = 2.168$$

Als Mittelwerte erhält man

$$\bar{x}_1 = 2$$

$$\bar{x}_2 = 2.2$$

Nun lassen sich die standardisierten Aufnahmen sehr leicht errechnen (Tabelle 3.3, E). Die hauptsächlichste Wirkung der Standardisierung im Falle dieses Beispiels ist offensichtlich. Die ursprünglich sehr unterschiedlichen Streuungen sind identisch geworden! Damit wird auch klar, warum die Standardisierung bei vegetationskundlichen Daten mit Vorsicht anzuwenden ist. Die Streuung einer Art ist eines ihrer wesentlichen Charakteristika. Gerade dieses geht aber bei der Standardisierung verloren.

Damit sind natürlich die Transformationsmöglichkeiten noch lange nicht erschöpft. Einige weitere sind speziell im Hinblick auf bestimmte Analysemethoden entwickelt worden. Die Transformationsmethode für die Korrespondenzanalyse ist ein Beispiel. Sie wird im Kapitel 8 erläutert.

3.4 Wirkung von Vektortransformationen auf Vegetationstabellen

Die Veränderung, welche eine Vegetationstabelle durch Transformationen erfährt, hängt ausschliesslich von den Ausgangsdaten ab. Weil bei pflanzensoziologischen Aufnahmen häufig ähnliche Datenstrukturen auftreten, ist es möglich, über die Wirkung der Transformationen generelle Aussagen zu machen. In Abb. 3.4 (A) sind einige typische Phänomene beispielhaft vereinigt. So findet man denn immer wieder besonders reichhaltige (Nr. 4), aber auch artenarme Aufnahmen (Nr. 1, 2). Bei den Arten weisen einige grosse Streuungen auf (Nr. 1), andere dagegen eine kleine (Nr. 2).

In den Diagrammen B und C der Abb. 3.4 ist die Auswirkung der Normalisierung der Aufnahmen beziehungsweise der Arten dargestellt. Es ist zu beachten, dass die Skalen für die grafischen Darstellungen willkürlich gewählt sind. Einer Beurteilung zugänglich sind also nur die Relationen zwischen den Abundanzwerten innerhalb der einzelnen Diagramme. Wir finden folgende, auch rein logisch ableitbare Regeln bestätigt:

1. Die Vektortransformation erbringt die geforderte Angleichung der Zeilen oder Spalten. Die schwach besetzten werden verstärkt (Aufn. 1 in B, Art 4 in C), dominante verlieren an Gewicht (Aufn. 4 in B, Art 1 in C).
2. Die Normalisierung der Aufnahmen führt zu einer verstärkten Gewichtung häufiger Arten (Art 1 in B), Normalisierung der Arten zu einer weiteren Verstärkung artenreicher Aufnahmen (Aufn. 4 in C).
3. Die Normalisierung der Aufnahmen kann die Streuungsverhältnisse innerhalb der Arten vollständig verändern (z.B. Art 2 in B). Dasselbe gilt sinngemäss für die Normalisierung der Arten.

Schliesslich findet sich in Abb. 3.4 auch die Wirkung der Standardisierung der Aufnahmen (D) und der Arten (E). Man kann folgende Phänomene festhalten:

1. Die bei der Normalisierung festgestellten Tendenzen gelten auch für die Standardisierung.

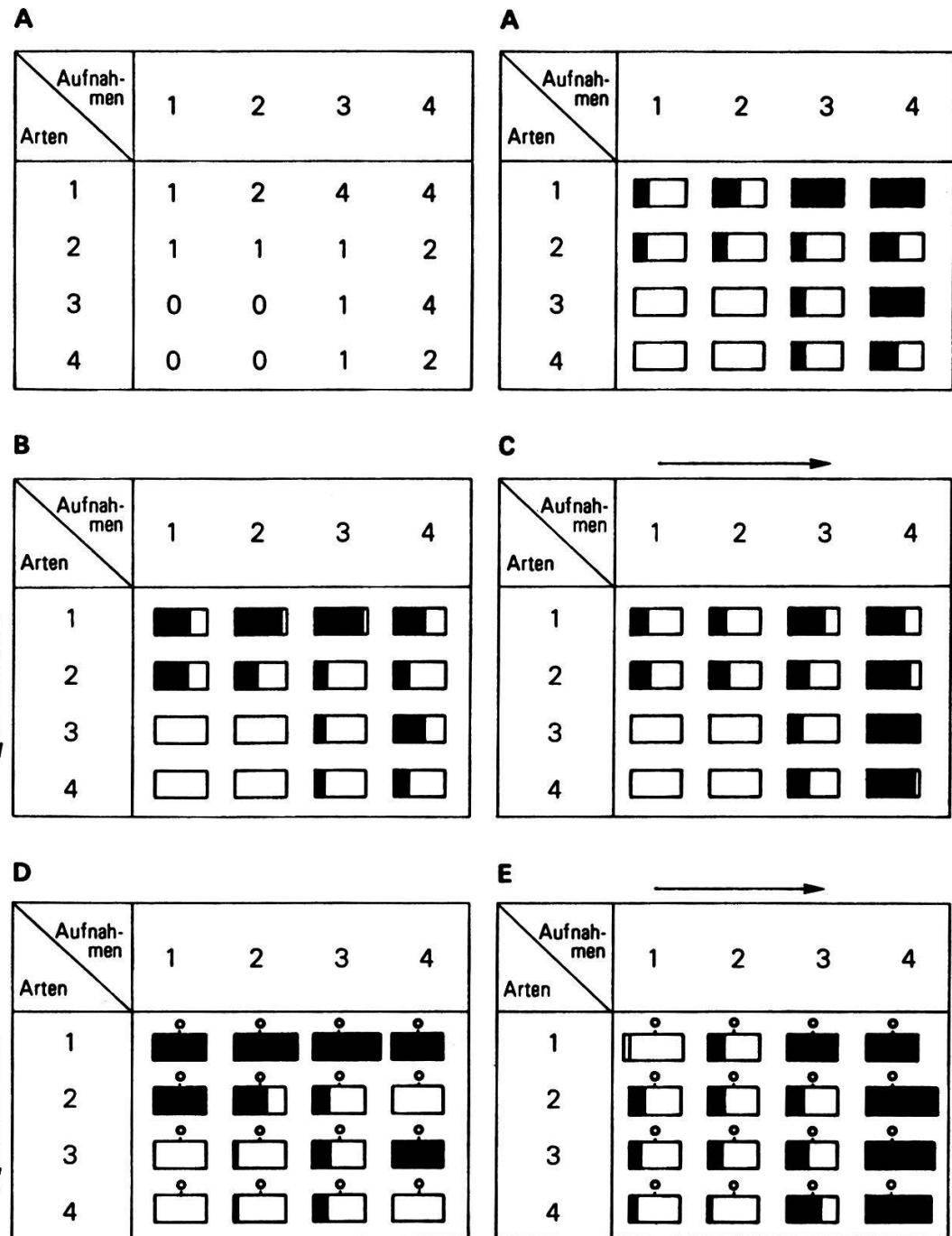


Abb. 3.4 Grafische Darstellung der Wirkung verschiedener Vektortransformationen auf eine Vegetationstabelle A. B: Normalisierung der Aufnahmen, C: der Arten. D: Standardisierung der Aufnahmen, E: der Arten. Die Skalierung der Grafiken ist willkürlich gewählt. Bei D und E treten auch negative Werte auf. Die Pfeile markieren die Richtung der Transformation.

2. Die Daten verändern sich noch stärker als bei der ersten, wenn deutliche Streuungsunterschiede vorliegen, die alsdann ausgeglichen werden. Aufnahme 4 in D ist ein Beispiel.

Es ist offensichtlich, dass die gezeigten Vektortransformationen bei pflanzensoziologischen Daten tiefgreifende Veränderungen der Ähnlichkeitsstruktur bewirken. Normalisierung und Standardisierung der Aufnahmen helfen, bei Artenarmut die doch vorhandenen qualitativen Unterschiede (oft etwas zu ausgeprägt) hervorzuheben. Werden die Artvektoren in dieser Weise transformiert, so verändern sich unter durchschnittlichen Umständen die Ähnlichkeitsverhältnisse der Aufnahmen drastisch und seltene Arten prägen die Ergebnisse weitgehend (Beispiel E in Abb. 3.4).

4. Aehnlichkeitsmasse

Im vorigen Kapitel wurde gezeigt, dass die Wahl der Transformationsmethode einer Entscheidung bedarf, welche sowohl von der Datenstruktur als auch vom Untersuchungsziel abhängt. Analog verhält es sich mit der Anwendung der verschiedenen Aehnlichkeitsmasse, die zur Bestimmung des Zusammenhangs zweier Aufnahmen oder zweier Arten beizuziehen sind. Je nach Standpunkt des Vegetationskundlers ergeben sich unterschiedliche Betrachtungsweisen. Viele verschiedene mathematische Formulierungen können diesen Rechnung tragen. Im folgenden wird eine beschränkte Auswahl teils bekannter, häufig verwendeter, teils für Vegetations- und Standortsdaten bewährter Aehnlichkeitsmasse vorgestellt und diskutiert.

4.1 Die Euklidsche Distanz

Die Euklidsche Distanz als Mass für die Unähnlichkeit zweier Vegetationsaufnahmen oder zweier Arten ist geometrisch leicht interpretierbar, mithin anschaulich und in ihren Eigenschaften ohne weiteres überblickbar. Aus diesen Gründen ist sie trotz einiger noch zu erwähnender Nachteile schon sehr oft verwendet worden (vgl. z.B. KUHN 1983). Betrachten wir zunächst den Vergleich zweier Aufnahmen. Dazu werden die Arten als räumliche Achsen aufgefasst. Im einfachsten Falle besitzt jede Aufnahme zwei Arten:

	Aufnahmne A	Aufnahmne B
Art 1	2	1
Art 2	2	3

Mit diesen Werten lassen sich die Aufnahmen als Punkte im zweidimensionalen Raum darstellen (Abb. 4.1). Die Distanz (Unähnlichkeit) ergibt sich nach dem pythagoräischen Lehrsatz als

$$d(A, B) = \left[\sum_i (x_{Ai} - x_{Bi})^2 \right]^{1/2}.$$

Im vorliegenden Beispiel mit $i = 2$ Arten ergibt das:

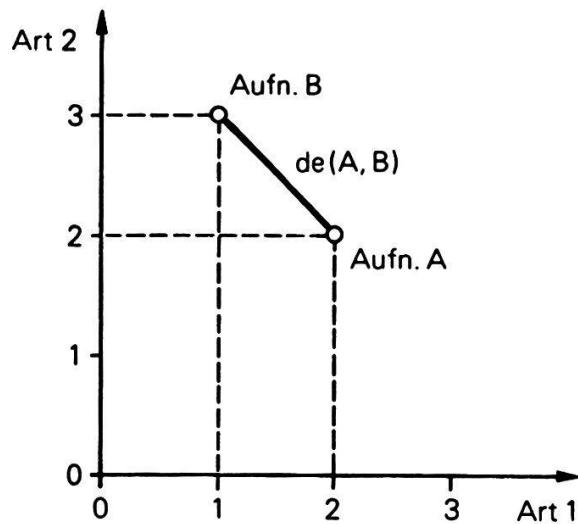


Abb. 4.1 Geometrische Bestimmung der Euklidischen Distanz zwischen zwei Aufnahmen A und B mit je 2 Arten.

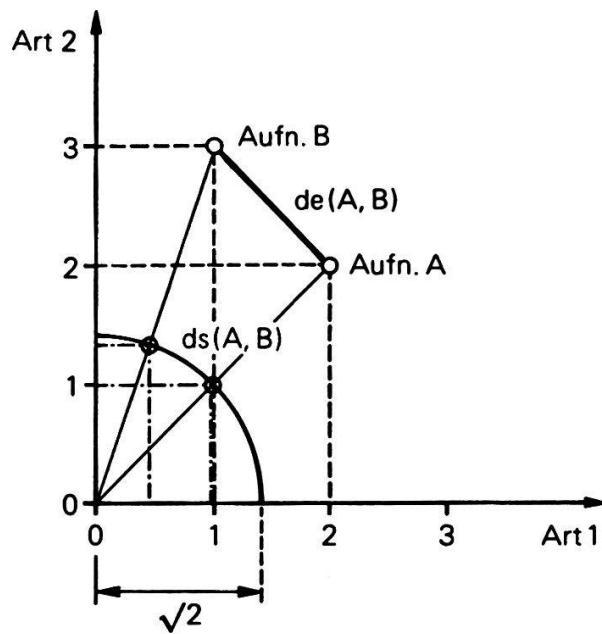


Abb. 4.2 Vergleich der Euklidischen Distanz de und der Sehnendistanz ds . Die ursprünglichen Abundanzen sind punktiert, die durch die Sehnendistanz implizit transformierten sind strichpunktiert.

$$d(A, B) = [(2 - 1)^2 + (2 - 3)^2]^{1/2} = 2^{1/2} = 1.41 .$$

Die Euklidische Distanz beruht also grundsätzlich auf der Annahme, dass die Artmächtigkeiten als metrische Angaben aufgefasst werden können. Sie funktioniert aber ebenso für Präsenz-Absenz Daten. Für fehlende Arten wird dann mit Gewicht 0, für anwesende mit 1 gerechnet. Dabei tritt nun eine bedeutende Schwäche dieses sehr einfachen Masses besonders augenfällig in Erscheinung. In pflanzensoziologischen Tabellen kommt es sehr häufig vor, dass eine Art sowohl in einer Aufnahme A als auch in Aufnahme B fehlt. Diese Art trägt also nicht zur Unterschiedlichkeit von A und B bei. Artenarme Aufnahmen weisen deshalb meist eine kleinere Euklidische Distanz auf als artenreiche und erscheinen daher oft ungerechtfertigterweise als relativ ähnlich. ORLOCI (1978), S. 46, demonstriert einen Fall, in dem zwei Aufnahmen mit gleichem Artspektrum unterschiedlicher erscheinen als solche mit verschiedener floristischer Zusammensetzung (Tab. 4.1, A). Tatsächlich findet man

$$d(1, 3) = [(0 - 0)^2 + (1 - 4)^2 + (1 - 4)^2]^{1/2} = 18^{1/2} = 4.24$$

Die floristisch total verschiedenen Aufnahmen 1 und 2 stehen sich wesentlich näher:

$$d(1, 2) = [(0 - 1)^2 + (1 - 1)^2 + (1 - 0)^2]^{1/2} = 3^{1/2} = 1.73$$

Als Konsequenz kann man festhalten, dass die Euklidische Distanz nur dann zu empfehlen ist, wenn die Artenvielfalt der Aufnahmen nicht allzu unterschiedlich ist. Für den Vergleich des gemeinsamen Auftretens von Arten eignet sie sich wohl in den seltensten Fällen, da sich Arten bezüglich ihrer Häufigkeit (absolute Stetigkeit) in der Regel zu sehr unterscheiden.

Tabelle 4.1

Ausgangsdaten als Beispiel zur Berechnung der Euklidschen- und der Sehnendistanz (nach ORLOCI 1978). Rohdaten (A), normalisierte Aufnahmen (B).

A

	Aufnahme 1	Aufnahme 2	Aufnahme 3
Art 1	0	1	0
Art 2	1	0	4
Art 3	1	0	4

B

	Aufnahme 1	Aufnahme 2	Aufnahme 3
Art 1	0	1	0
Art 2	0.707	0	0.707
Art 3	0.707	0	0.707

4.2 Die Sehnendistanz

Die Sehnendistanz wird von ORLOCI (1978) vorgeschlagen, um den eben erwähnten Nachteil der Euklidschen Distanz zu eliminieren. Implizit werden die Aufnahmen normalisiert, das heisst auf Einheitslänge (1.0) gebracht. Quantitative Unterschiede in den Artmächtigkeiten, welche zu den im vorigen Kapitel erwähnten, unerwünschten Aehnlichkeitsbeziehungen geführt haben, werden ausgeglichen. Die Einheitslängentransformation ist geometrisch darstellbar. In Abb. 4.2 sind die ursprünglichen Verhältnisse zwischen den untransformierten Arten (punktiert) und diejenigen nach der Transformation (strichpunktiert) dargestellt. Jede neue Artmächtigkeit x'_i ergibt sich aus der ursprünglichen durch Division mit der Länge des Aufnahmevektors La :

$$x'_i = x_i / La$$

Die Länge des Aufnahmevektors berechnet sich mit

$$La = (\sum x_i^2)^{1/2}.$$

Man kann diese Transformationsvorschrift in die Formel für die Euklidsche Distanz einsetzen und erhält dann nach einigen Umformungen eine neue Formel für die direkte Berechnung der Sehnendistanz:

$$ds(A, B) = [2(1 - \sum_i x_{Ai} x_{Bi} / (\sum_i x_{Ai} \sum_i x_{Bi}))^{1/2}]^{1/2}.$$

Darin ist x_{Ai} die Artmächtigkeit der Art i in der Aufnahme A. Dem Rechengang ist bei schrittweisem Vorgehen leichter zu folgen. Transformiert man das Beispiel in Tab. 4.1, A, wie besprochen, so erhält man die Werte in 4.1, B. Die Anwendung der Formel für die Euklidsche Distanz ergibt

$$ds(1, 3) = [(0-0)^2 + (0.707-0.707)^2 + (0.707-0.707)^2]^{1/2} = 0.0$$

$$ds(1, 2) = [(0-1)^2 + (0.707-0.0)^2 + (0.707-0.0)^2]^{1/2} = 1.414$$

Die quantitativen Unterschiede zwischen den Aufnahmen 1 und

3 fallen damit nicht mehr ins Gewicht, dafür wird dem als wichtig empfundenen Artenspektrum Rechnung getragen. Im Gegensatz zur Euklidschen Distanz gibt es für die maximale Unterschiedlichkeit einen festen Höchstwert, nämlich $2^{1/2}$ oder 1.414. Das Beispiel zeigt auch gleich einen Nachteil der Sehnendistanz, indem artenreiche Aufnahmen (Nr. 1 und 3) im allgemeinen schlecht, artenarme (Nr. 2) jedoch eher gut differenziert werden. Je nach Zielsetzung der Analyse kann diese Eigenschaft erwünscht oder unerwünscht sein. Günstig wirkt sich die Sehnendistanz aus, wenn Aufnahmen ähnlicher Diversität verglichen werden sollen. Auch bei der Untersuchung von Aufnahmen schwer vergleichbarer Skalierung, z.B. bei gemischten Daten aus der Literatur und aus Originaluntersuchungen, kann die Sehnendistanz die nötige Anpassung bringen. Dagegen führt der Vergleich von Arten oft zu unbefriedigenden Ergebnissen: Die Sehnendistanz bewirkt eine Ueberkorrektur der Nachteile der Euklidschen Distanz. Die Folge ist, dass seltene Arten gut, abundante schlecht differenziert werden.

4.3 Skalarprodukt und Kovarianz

Diese beiden Masse für die Aehnlichkeit sollen hier gemeinsam behandelt werden. Im Gegensatz zu den Distanzmassen führt eine gute Uebereinstimmung von Arten oder Aufnahmen zu hohen positiven, Unterschiedlichkeit zu negativen Werten. Die auftretenden Höchst- und Tiefstwerte besitzen keine festen Grenzen, was die Interpretation erschwert. Beide Masse enthalten auch eine Transformation der Daten: Sie zentrieren die Vektoren (Das Skalarprodukt ist auch auf unzentrierte Daten anwendbar. Es besitzt dann dieselben Vorteile wie die Euklidsche Distanz). Im Falle des Vergleiches von Aufnahmen sind alle Artmächtigkeiten um ihren Mittelwert in der betreffenden Aufnahme zu vermindern:

$$x'_i = x_i - \bar{x}, \quad \bar{x} = \frac{1}{p} \left(\sum_i x_i \right).$$

Dabei ist p die Zahl aller in der betreffenden Vegetationstabelle auftretenden Arten. Ein Beispiel zeigt Tabelle 4.2. Links enthält sie die rohen Ausgangsdaten, rechts die

Tabelle 4.2

Ausgangsdaten für die Berechnung von Skalarprodukt und Kovarianz. Rohdaten (A), zentrierte Daten (B).

A

	Aufn 1	Aufn 2
Art 1	0	0
Art 2	0	1
Art 3	1	0
Art 4	1	0
$\bar{x} =$	1/2	1/4

B

	Aufn 1	Aufn 2
Art 1	-1/2	-1/4
Art 2	-1/2	3/4
Art 3	1/2	-1/4
Art 4	1/2	-1/4
$\bar{x} =$	0	0

Tabelle 4.3

Wirkung der Zentrierung auf Arten unterschiedlicher Dominanz. Rohdaten (A), zentrierte Daten (B).

A

Aufn.	1	2	3	4	\bar{x}
Art 1	3	3	4	2	3
Art 2	1	1	2	0	1

B

Aufn.	1	2	3	4	\bar{x}
Art 1	0	0	1	-1	0
Art 2	0	0	1	-1	0

über die Aufnahmen zentrierten. Das Skalarprodukt entspricht sodann einfach der Summe der Produkte der Mächtigkeit jeder Art:

$$S(A, B) = \sum_i x' A_i x' B_i .$$

Für die Werte in Tabelle 4.2 erhalten wir

$$\begin{aligned} S(A, B) &= (-1/2 * -1/4) + (-1/2 * 3/4) + (1/2 * -1/4) + (1/2 * -1/4) \\ &= 1/8 - 3/8 - 1/8 - 1/8 = -1/2 \end{aligned}$$

Um die Kovarianz zu erhalten, ist das Ergebnis noch durch $p-1$ zu dividieren:

$$C(A, B) = \frac{1}{p-1} S(A, B) = \frac{1}{p-1} \sum_i x' A_i x' B_i$$

Da sich die Zahl p innerhalb eines Datensatzes nicht ändert, sind $S(A, B)$ und $C(A, B)$ stets proportional. Verwendet man sie z.B. für eine Gruppierungsanalyse, so werden die Ergebnisse identisch ausfallen.

Wie schon bei den Distanzmassen, so ergeben sich auch bei Skalarprodukt und Kovarianz Schwierigkeiten mit dem in der Pflanzensoziologie typischen Nebeneinander von häufigen und seltenen Arten. Fast leere (d.h. mit vielen Nullen besetzte) Vektorpaare besitzen immer eine kleine Kovarianz. Gut besetzte Vektorpaare weisen dagegen eine grösse Streubreite auf, sowohl in positiver wie in negativer Richtung. Sie werden daher durch nachfolgende Analysen besser differenziert.

Ansonsten sind Euklidsche Distanz und Kovarianz recht verschiedene Konzepte. Die Zentrierung, welche letztere impliziert, bringt unter Umständen grosse quantitative Verschiebungen mit sich. Tab. 4.3 verdeutlicht diesen Verhalt. Die beiden unterschiedlich gut vertretenen Arten werden nach der Transformation genau gleich bewertet. Tab. 4.4 zeigt ein Beispiel, in welchem der Unterschied zwischen Euklid-

Tabelle 4.4

Beispiel zur Unterschiedlichkeit von Euklidscher Distanz und Skalarprodukt der zentrierten Daten. Erstere beträgt für die Aufnahmen beider Tabellen 1.414. Das Skalarprodukt beträgt in Tabelle A -1/4, in Tabelle B 0.0.

<u>A</u>		<u>B</u>			
Aufn.	1	2	Aufn.	1	2
Art 1	1	0	Art 1	1	0
Art 2	0	0	Art 2	0	0
Art 3	0	0	Art 3	1	1
Art 4	0	1	Art 4	0	1
$\bar{x} =$	1/4	1/4	$\bar{x} =$	1/2	1/2

Tabelle 4.5

Ausgangsdaten als Beispiel zur Berechnung des Korrelationskoeffizienten zwischen zwei Aufnahmen 1 und 2. A: Rohdaten. B: Zentrierte Daten. C: Standardisierte Daten.

<u>A</u>		<u>B</u>		<u>C</u>	
Aufn.	1	2	Aufn.	1	2
Art 1	1	1	Art 1	-1	-2
Art 2	2	1	Art 2	0	-2
Art 3	3	7	Art 3	1	4
$\bar{x} =$	2	3	$(\Sigma x^2)^{1/2} = 1.41$	4.89	
$s =$					
$\bar{x} =$	0	0			

scher Distanz einerseits und Skalarprodukt und Kovarianz andererseits voll zum Tragen kommt. In den Aufnahmen der Tabelle A fehlt Art 3, während sie in B vorkommt. Die Euklidische Distanz beträgt in beiden Fällen 1.414. Für das Skalarprodukt erhält man in Tabelle A -1/4 (Gegenläufigkeit), in Tabelle B 0.0 (Unabhängigkeit). Es reagiert damit auf die unterschiedlichen Varianzverhältnisse.

4.4 Der Korrelationskoeffizient

Der Korrelationskoeffizient ist das bekannteste und in Verfahren der schliessenden Statistik wohl am häufigsten verwendete Ähnlichkeitsmass (vgl. z.B. GAENSSLER und SCHUBÖ, 1973, S. 15ff.). Seine Anwendung in der Analyse von Vegetationsdaten kann jedoch zu überraschenden, meist enttäuschenden Ergebnissen führen. Die Untersuchung seiner Eigenschaften zeigt einige Gründe auf.

Zunächst liegt dem Korrelationskoeffizienten die Formel zur Berechnung der Kovarianz zugrunde. Er hat auch deren im vorigen Kapitel gezeigte Nachteile. Zusätzlich unterwirft er die Datenvektoren einer weiteren Transformation, der z-Transformation oder Standardisierung. Dabei werden alle Elemente um den Mittelwert ihres Vektors vermindert und durch die Standardabweichung dividiert:

$$z = (x - \bar{x}) / s ,$$

wobei z der neue, standardisierte Wert ist. In Tabelle 4.5 ist die Wirkung dieser Transformation anhand eines Beispiels dargestellt. Die Verhältnisse zwischen den Artmächtigkeiten ändern sich gegenüber den ursprünglichen Daten grundsätzlich. Insbesondere ist zu beachten, dass die Streuungen der Aufnahmen vereinheitlicht werden. Damit fällt ein für sie als wesentlich erachtetes Charakteristikum ausser Betracht. Insbesondere Tabelle 4.5, C, zeigt, dass die Gewichtung der Abundanzen nach einer Standardisierung wohl selten den ursprünglichen Absichten eines Vegetationskundlers entsprechen dürfte. Daher befriedigt auch eine Aussage über

die Struktur von Vegetationstabellen selten, sofern sie auf Grund des Korrelationskoeffizienten ermittelt wurde. Anders verhält es sich mit Standortsdaten. Sind diese durch Verwendung unterschiedlicher Messskalen gewonnen worden, z.B. wenn pH-Werte und Grundwasserstände vorliegen, so gewährleistet die Standardisierung - durch die Anwendung des Korrelationskoeffizienten - die gewünschte Vergleichbarkeit weitgehend.

Nach der Standardisierung der Daten ergibt sich der Korrelationskoeffizient durch die Anwendung der Formel für die Kovarianz. Für die Aufnahmen in Tab. 4.5 findet man:

$$\begin{aligned} r^2 &= 1/3 [(-0.707 * -0.408) + (0.0 * -0.408) + (0.707 * 0.816)] \\ &= 0.228 \\ r &= 0.537 \end{aligned}$$

Das Resultat weist auf einen mässig positiven Zusammenhang der Aufnahmen 1 und 2 hin. Da r zwischen +1 und -1 liegen kann, treten bei Gegenläufigkeit (z.B. wenn zwei Aufnahmen keine gemeinsamen Arten aufweisen) natürlich auch negative Werte auf.

Die oben angewandte Art der Berechnung des Korrelationskoeffizienten ist in der Praxis recht umständlich. Der Vollständigkeit halber sei noch eine zweckmässigere Formel angeführt:

$$r^2 = - \frac{1}{p} \frac{\sum (x_A - \bar{x}_A)(x_B - \bar{x}_B)}{\left[\sum (x_A - \bar{x}_A)^2 \right]^{1/2} \cdot \left[\sum (x_B - \bar{x}_B)^2 \right]^{1/2}}$$

A und B sind die zu vergleichenden Aufnahmen, p die Anzahl Arten. Damit lässt sich nun r direkt aus den Rohdaten errechnen.

4.5 Kontingenzmasse

In dieser Gruppe finden sich die frühesten Formeln zur Berechnung der Ähnlichkeiten von Aufnahmen, nähmlich diejenigen von JACCARD (1901) und SOERENSEN (1948). Varian-

ten von diesen sind in grosser Zahl beschrieben worden und sie tauchen ihrer Beliebtheit wegen auch in neueren Lehrbüchern immer wieder auf (MUELLER-DOMBOIS und ELLENBERG 1974, ORLOCI 1978, GAUCH 1982 usw.). Ausgangspunkt zur Berechnung der Aehnlichkeiten sind ursprünglich reine Präsenz-Absenz Daten, welche in einer Kontingenztafel auszuwerten sind:

Aufn. A	+	-	
Aufn. B			
+	a	b	a+b
-	c	d	c+d
	a+c	b+d	n

Im Feld a wird die Anzahl der Arten eingetragen, die sowohl in A wie auch in B vorkommen. In b werden die Arten gezählt, wenn sie in A fehlen, in B aber vorkommen. In c finden sich Arten, die in B fehlen und in A vorkommen. Feld d schliesslich enthält die Anzahl Arten, welche in beiden Aufnahmen fehlen. In der letzten Kolonne bzw. Spalte findet man die sogenannten Randsummen $a+c$, $b+d$, $a+b$, $c+d$ sowie die Gesamtfrequenz n . Mit dieser Notation lassen sich sowohl metrische als auch nichtmetrische Masse formulieren. Welcher Fall vorliegt, ist auf formalem Wege abzuklären: Wird die Dreiecksungleichung nicht verletzt, so ist das Mass metrisch (Kap. 3.1.3). Für die meisten Funktionen sind diese Eigenschaften direkt bei LEGENDRE und LEGENDRE (1979, Bd II, S. 31f) nachzuschlagen. Die nicht metrischen unter ihnen haben dabei den Vorzug, dass sie gegenüber Leerstellen mehr oder weniger unempfindlich sind. Verzerrungen, wie sie bei der Euklidschen Distanz besprochen wurden, können dadurch vermieden werden. Für die Anwendung verschiedener Methoden bedeutet dies jedoch eine Einschränkung, wie später noch zu zeigen ist.

Die Berechnung der Felder a, b, c und d wird in Tabelle 4.6 anhand eines Beispiels gezeigt. Ausgehend von 4 Aufnahmen

Tabelle 4.6

Messung des Zusammenhangs zweier Arten einer Vegetationstabelle (A) mit Hilfe einer Kontingenztabelle (B). "+" bedeutet, dass die betreffende Art vorkommt, "-" dass sie fehlt.

A

	Aufn. A	Aufn. B	Aufn. C	Aufn. D
Art 1	1	1	1	0
Art 2	0	1	0	0

B

Art 1	+	-	
Art 2			
+	$a=1$	$b=0$	$a+b=1$
-	$c=2$	$d=1$	$c+d=3$
	$a+c=3$	$b+d=1$	$n=4$

soll darin das Zusammenfallen von 2 Arten ermittelt werden können. Die Gesamtfrequenz n entspricht zwangsläufig der Summe der Aufnahmen, d.h. 4. Alle für die Bestimmung der Ähnlichkeit benötigten Parameter sind damit bekannt.

Da nun hier eine echte Kontingenztafel vorliegt, drängt sich als erstes ein aus der schliessenden Statistik bekanntes Mass zur Bestimmung des Zusammenhangs von A und B geradezu auf. Werden nämlich die Randsummen $a+c$, $b+c$, $a+b$ und $c+d$ (Tabelle 4.6) als invariante, naturgegebene Größen angenommen (was nach PIELOU (1977) in Wirklichkeit durchaus nicht immer zutreffen muss), so lässt sich ein sogenanntes Chiquaret χ^2 formulieren. Diese statistische Testgrösse gibt an, inwiefern die Inhalte der Felder a bis d von zufällig zu erwartenden Werten abweichen. Sie berechnet sich als Summe der quadrierten Differenzen zwischen Erwartungswerten E und beobachteten Werten O , dividiert durch die Erwartungswerte jedes Feldes i :

$$\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i}$$

Als Erwartungswert für das Feld a nimmt man das Produkt der Randsummen der Art A, also $(a+c)$ und $(a+b)$ und misst dieses mit der Gesamtfrequenz:

$$E(a) = \frac{(a+c)(a+b)}{N}$$

Die beobachtete Frequenz ist natürlich der Wert a selbst dividiert durch N . Durch einfache Ableitung (vgl. z.B MUELLER-DOMBOIS und ELLENBERG 1974, S. 238) lässt sich die Formel für die ganze Tafel finden:

$$\chi^2 = \frac{N((ad-bc)^2)}{(a+b)(a+c)(c+d)(b+d)}$$

Diese Grösse ist unter bestimmten Voraussetzungen χ^2 -verteilt (vgl. BARTEL 1972, S. 71). Bedienen wir uns ausnahmsweise der schliessenden Statistik, so kann einer χ^2 -Tabelle (Anz. Freiheitsgrade = 1) entnommen werden, ob die Unterschiedlichkeit im Auftreten von Art 1 und 2 bis auf normale Abweichungen rein zufälliger Natur sind. Im Beispiel von Tabelle 4.6 erhält man

$$\chi^2 = \frac{4((1-0)^2)}{1 * 3 * 1 * 3} = \frac{4}{9} = 0.44$$

Bei einer Irrtumswahrscheinlichkeit von 5% findet man in entsprechenden Tabellen einen Zufallshöchstwert von 3.84, womit die Unterschiedlichkeit der beiden Arten erwartungsgemäss nicht gesichert werden kann. Selbstverständlich ist diese Aussage mit Vorbehalten zu interpretieren, wurden doch die Voraussetzungen für einen χ^2 -Test nicht überprüft. Sie kann aber - nun wieder im Sinne der explorativen Statistik - als Interpretationshilfe ohne Bedenken verwendet werden.

Die Inhalte der Kontingenztafel können nun zur Formulierung weiterer Ähnlichkeitsmasse verwendet werden. Ausgehend vom Chiadrat diskutiert PIELOU (1977, S. 208ff.) eingehend den **V- Wert von YULE (1912):**

$$V = \frac{ad - bc}{[(a+b)(a+c)(c+d)(b+d)]^{1/2}}$$

Es wird auch der Nachweis erbracht, dass es sich um einen Korrelationskoeffizienten handelt mit den Eigenschaften

- $V = 1$ für vollständigen Zusammenhang
- $V = 0$ für fehlenden Zusammenhang
- $V = -1$ für Gegenläufigkeit

In unserem Beispiel (Tabelle 4.6) erhalten wir

$$v = (1-0)/9^{1/2} = 1/9^{1/2} = 0.33$$

Die eben besprochenen Ähnlichkeitsmasse haben für pflanzensoziologische Anwendungen den Vorteil, dass anstelle der Rohdaten mit Erwartungswerten operiert wird. Sie interpretieren eine Zunahme der Leerstellen (d.h. von Arten, die in beiden Aufnahmen fehlen) so, dass das gemeinsame Vorkommen zweier Arten weniger wahrscheinlich werde und bewerten es daher höher. Diese Art der Betrachtung ist in der meisten vegetationskundlichen Anwendungen sinnvoll - sofern der Datensatz eine gewisse Größe besitzt und einigermassen homogen ist.

Wahrscheinlich das erste in der Pflanzensoziologie bekannt gewordene Ähnlichkeitsmass ist der **Koeffizient von Jaccard** (JACCARD 1901). Er berechnet sich nach der Formel

$$SJ = a / (a+b+c) .$$

Für das Beispiel in Tab. 4.6 erhält man

$$SJ = 1 / (1+0+2) = 0.33$$

Als Grenzwerte sind möglich:

$$\begin{aligned} \max(SJ) &= 1 \text{ (A und B identisch)} \\ \min(SJ) &= 0 \text{ (A und B ohne gemeinsame Arten)} . \end{aligned}$$

Dem Koeffizient von Jaccard verwandt ist der **Koeffizient von Soerensen** (SOERENSEN 1948). Er birgt jedoch eine etwas stärkere Gewichtung gemeinsamer Arten in sich:

$$SS = 2a / (2a+b+c) .$$

Sein Wertebereich erstreckt sich ebenfalls von 0 bis 1. Für das Beispiel in Tab. 4.6 erhält man

$$SS = 2 / (2+0+2) = 0.5 .$$

Bei teilweise übereinstimmender Artengarnitur ergeben sich also etwas höhere Ähnlichkeitswerte als bei Jaccard.

Beide Koeffizienten sind unbeeinflusst von fehlenden Arten sowie vom Total n. Werden sie zur Analyse einer Vegetations-tabelle eingesetzt, so ändern die Nenner von Aufnahmepaar zu Aufnahmepaar. Das gemeinsame Vorkommen der Arten wird mithin mit ständig änderndem Massstab gemessen, sodass die resultierende Ähnlichkeitsstruktur nicht-metrisch ist. Mit SJ und SS werden aber immer wieder gute Erfahrungen gemacht; wohl deshalb, weil sie der oft vorkommenden Heterogenität der Daten entgegenkommen.

Die bisher betrachteten Koeffizienten berücksichtigen nur die An- oder Abwesenheit einer Art. Es besteht jedoch öfters das Bedürfnis, auch die Artmächtigkeit in die Berechnungen mit einzubeziehen. Um dies bei Kontingenzmassen verwirklichen zu können, wurden verschiedenste Formeln vorgeschlagen (vgl. z.B. die Uebersicht bei MUELLER-DOMBOIS und ELLENBERG 1974). ELLENBERG (1956) verwendet in Anlehnung an Jaccard

$$SE = \frac{1/2 \sum a_i}{\sum b_i + \sum c_i + 1/2(\sum a_i)} * 100$$

Darin ist $\sum a_i$ die Summe der Elemente der Zelle a in der Kontingenztafel (d.h. alle gemeinsamen Arten), $\sum b_i$ die Summe der Elemente in b und $\sum c_i$ diejenige in Zelle c.

Van der Maarels Koeffizient SM (VAN DER MAAREL et al. 1978) entspricht ebenfalls demjenigen von Jaccard, modifiziert für quantitative Daten:

$$SM = \frac{\sum_i x_{Ai} x_{Bi}}{\sum_i x_{Ai}^2 + \sum_i x_{Bi}^2 - \sum_i x_{Ai} x_{Bi}}$$

Das folgende Beispiel zeigt die Berechnung dieser beiden Koeffizienten:

Art	1	2	3	4
Aufn. A	1	2	3	0
Aufn. B	0	1	2	3

$$SE = \frac{1/2(2+1+3+2)}{1+3+1/2(2+1+3+2)} * 100 = \frac{4}{1+3+4} * 100 = 50$$

$$SM = \frac{(1*0)+(2*1)+(3*2)+(0*3)}{(1+4+9)+(1+4+9)-[(1*0)+(2*1)+(3*2)+(0*3)]}$$

$$= \frac{8}{14+14-8} = 0.4$$

Werden in den Ausgangsdaten alle von Null verschiedenen Werte durch 1 ersetzt, so ergibt sich für SE 50(%), für SM 0.5 und für SJ (Jaccard) ebenfalls 0.5. Konzeptionell erbringen die Koeffizienten Ellenbergs und van der Maarels keine neuen Aspekte.

Damit sind die Möglichkeiten der Aehnlichkeitsbestimmung auf Grund der Kontingenztafel bei weitem nicht erschöpft. Eine Uebersicht findet sich bei LEGENDRE und LEGENDRE (1979), wo in Band II, Seite 32/33 nicht weniger als 14 verschiedene Koeffizienten aufgeführt sind.

4.6 Absolutwertfunktionen

Aus dem Bestreben heraus, den Rechenaufwand für die Aehnlichkeitsbestimmung möglichst klein zu halten, wird gelegentlich die folgende, einfache Absolutwertfunktion verwendet:

$$A(A, B) = \sum_i ABS(x_{Ai} - x_{Bi}) .$$

Meistens wird diese Funktion als Alternative zur Euklidschen Distanz gesehen. Für ihre geometrische Interpretation sei auf Abb. 4.1 verwiesen. Die Euklidsche Distanz zwischen A und B entspricht dort der direkten Verbindung zwischen den beiden Punkten. Die Absolutwertfunktion misst dagegen ausschliesslich entlang der horizontalen und der vertikalen Achse und addiert die Beträge ($A(A,B)=2$), weshalb auch von City-Block- oder Manhattan-Distanz gesprochen wird. Gegenüber der Euklidschen Distanz ergeben sich kaum Vorteile. Es treten auch die bei ORLOCI (1978) erwähnten Fälle auf, bei welchen Aufnahmen mit identischem Artenspektrum als sehr unterschiedlich bewertet werden (Kap. 4.1). WHITTAKER (1952) skaliert deshalb die Aufnahmen innerhalb seines Koeffizienten W :

$$W(A,B) = \sum_i \text{ABS} \left(\frac{x_{Ai}}{Q_A} - \frac{x_{Bi}}{Q_B} \right), \text{ wobei}$$
$$Q_A = \sum_i x_{Ai} \quad \text{und} \quad Q_B = \sum_i x_{Bi}.$$

Die Werte für $W(A,B)$ sind dadurch auf den Bereich 0 bis 2 begrenzt. Betrachten wir unser aus dem letzten Kapitel bekanntes Beispiel:

Art	1	2	3	4
Aufn. A	1	2	3	0
Aufn. B	0	1	2	3

Für Whittakers Koeffizient ist $Q_A = Q_B = 6$ und

$$W(A,B) = \text{ABS}((1/6)-(0/6)) + \text{ABS}((2/6)-(1/6)) + \text{ABS}((3/6)-(2/6)) + \text{ABS}((0/6)-(3/6)) = 1.$$

ORLOCI (1978) erwähnt, dass $W(A,B)$ auf die Darstellung einfacher Vegetationsgradienten linearisierende Wirkung hat, was natürlich auf die Bereichsanpassung zurückzuführen ist (und auch von der Sehnendistanz gesagt werden könnte).

4.7 Die Mahalanobis Distanz

Alle bisher besprochenen Ähnlichkeitsmasse sind von der Korrelation der Arten beeinflusst. Für viele Anwendungen ist diese Eigenschaft durchaus erwünscht. Pflanzenarten, die auf einen Standortsgradienten fast gleich reagieren, gehen dabei in die Berechnung der Aufnahmähnlichkeit mit gleichem Gewicht ein wie solche, die eine ganz andere Eigenschaft der Standortes wiederspiegeln. Dem kann bei Bedarf abgeholfen werden. An einem einfachen Beispiel aus ORLOCI (1978) soll dies erläutert werden (Tabelle 4.7, A). Darin ist sofort ersichtlich, dass Art 1 und 2 hoch korrelieren (weil sie möglicherweise ähnliche Standortsansprüche besitzen). Art 3 ist dagegen annähernd unabhängig. Die Korrelationsmatrix bestätigt diese Befunde (Tabelle 4.7, B). Die Aufnahmen 1 bis 5 entstammen offensichtlich zwei verschiedenen Vegetationsgradienten, wobei der eine mit den Arten 1 und 2 besser vertreten ist als der andere mit Art 3. Dieses Ungleichgewicht soll die Mahalanobisdistanz ausgleichen. In Abbildung 4.3 wird das Vorgehen anhand eines zweidimensionalen Falles erläutert. Man erinnert sich, dass zur Berechnung der Euklidschen Distanz jede Art mit gleichem Gewicht in die Berechnung eingeht. In Abb. 4.3, A, erhält man

$$D^2(A, B) = (2-1)^2 + (1-2)^2 = 2.$$

Darin wird stillschweigend angenommen, dass die Arten 1 und 2 voneinander unabhängig sind und das Koordinatensystem deshalb rechtwinklig sei. Dies trifft hier eigentlich nicht zu. Den allgemeinen Fall zeigt Abb. 4.3 B. Darin korrelieren Art 1 und 2 positiv, was geometrisch durch schiefwinklige Koordinaten dargestellt wird. Die Distanz $DM(A, B)$ vermindert sich gegenüber der Euklidschen Distanz. Die Berechnung ergibt sich direkt aus den Cosinussatz:

$$DM^2(A, B) = [(2-1)^2 + (1-2)^2] - 2(2-1)(1-2) \cos \alpha$$

Der Winkel α ist Ausdruck der Korrelation der Arten 1 und 2. Den benötigten Zusammenhang bringt die Formel

Tabelle 4.7

Beispiel zur Berechnung der Mahalanobisdistanz (A) mit der Korrelationsmatrix der Arten (B).

A

	Aufn.1	Aufn.2	Aufn.3	Aufn.4	Aufn.5
Art 1	2	5	2	1	0
Art 2	3	4	1	0	0
Art 3	0	1	4	3	1

B

	Art 1	Art 2	Art 3
Art 1	1.000	0.883	-0.163
Art 2		1.000	-0.536
Art 3			1.000

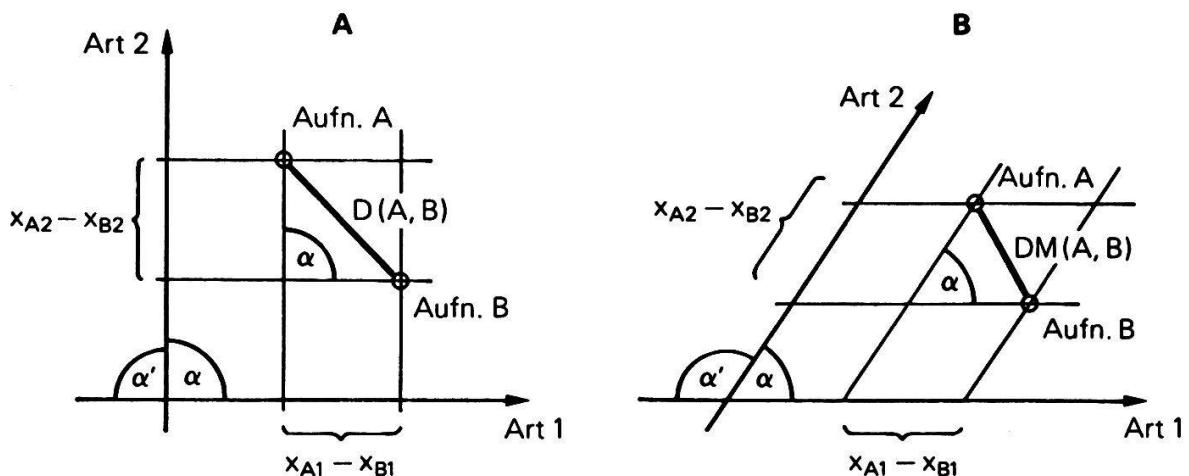


Abb. 4.3 Berechnung der Euklidschen Distanz (A) und der Mahalanobisdistanz (B).

$$r(1,2) = -\cos \alpha' = -\cos (180^\circ - \alpha)$$

Es ist leicht zu prüfen, dass sich $DM(A,B)$ auf $D(A,B)$ reduziert, wenn $r(1,2)$ null wird, die Arten 1 und 2 also von einander unabhängige Information tragen. In der allgemeinen Form, d.h. bei mehr als zwei Arten, berechnet sich die Mahalanobis Distanz wie folgt:

$$DM^2(A,B) = \sum_i (x_{Ai} - x_{Bi})^2 - 2 * \sum_{j,k} (x_{Aj} - x_{Bj})(x_{Ak} - x_{Bk})$$

Ist p die Anzahl der Arten, so laufen die Summationen wie folgt:

$$\begin{aligned} i &= 1, \dots, p \\ j &= 1, \dots, p-1 \\ k &= 1, \dots, p \end{aligned}$$

Für die beiden ersten Aufnahmen in Tab. 4.7 erhält man demzufolge:

$$\begin{aligned} DM^2(1,2) &= (2-5)^2 + (3-4)^2 + (0-1)^2 \\ &\quad - 2[(2-5)(3-4)(-0.883) \\ &\quad + (2-5)(0-1)(0.163) \\ &\quad + (3-4)(0-1)(0.536)] = 3.775^2 \end{aligned}$$

Das Beispiel zeigt, dass bei der Berechnung der Mahalanobisdistanz laufend auf die Korrelationsmatrix zugegriffen werden muss, was bei der Verwendung von Rechenautomaten einen hohen Speicherbedarf erfordert und lange Rechenzeiten verursacht. Für die Praxis empfiehlt sich ein anderes Vorgehen. Anstelle von Artvektoren können die Faktorenwerte einer Hauptkomponentenanalyse verwendet werden, die immer unkorreliert sind (Kap. 7). Eine nachfolgend gerechnete Distanz ist dann ebenfalls frei vom Einfluss der Korrelation der Arten.

Die Mahalanobisdistanz hat nur einen Sinn im Kontext einer grösseren Vegetationstabelle. Ihr Vorteil gegenüber der Euklidischen Distanz ist meist unbedeutend, wird doch der nicht-linearen Struktur der Daten nicht Rechnung getragen.

4.8 Informationsmasse

Die Mehrheit bisher besprochener Ähnlichkeitsmasse wurde für metrische Daten entwickelt. Ihre Anwendung auf Artmächtigkeitsskalen ist nur mit Kunstgriffen möglich (Abschnitt 3.2). Echte Nominaldaten sind jedoch typisch für vegetationskundliche Untersuchungen. So wird bei Waldaufnahmen häufig Unterschieden zwischen Krautschicht, Strauchschicht und Baumschicht. Solche Unterschiede lassen sich metrisch nicht treffend beschreiben. Hingegen kann das Spektrum der Merkmale durch die Entropiefunktion von SHANNON (1948) erfasst werden. Damit kommen die Methoden der Informationstheorie zur Anwendung. Über deren Grundzüge orientiert z. B. ein umfassender Anhang bei RENYI (1962). Die nachfolgenden Ausführungen halten sich an die Nomenklatur von PIELOU (1977) und ORLOCI (1978).

Gegeben seien also Stichproben (z.B. von einzelnen Arten), die in Klassen (z.B. Artmächtigkeitsklassen) mit den relativen Häufigkeiten p_1, p_2, \dots, p_n aufgeteilt werden. Die mittlere Entropie für die Aufnahme h ergibt sich dann als

$$H(P_h) = - \sum_j p_{hj} \ln p_{hj}, \quad j=1, \dots, n.$$

Wie PIELOU (1977) ausführt, kann $H(P_h)$ als Informationsgehalt der Aufnahme h aufgefasst werden. Die Entropie eignet sich z.B. als Maß für die Diversität derselben. Genauer gefasst ist $H(P_h)$ ein Ausdruck für die Ungewissheit über die Zusammensetzung von h . Anhand von Tabelle 4.8 kann dies gezeigt werden. Zu diesem Zwecke ist die Formel für die Entropie so umzuschreiben, dass statt der relativen Häufigkeiten p_j die absoluten innerhalb der Klassen j , f_j verwendet werden können. Wir setzen

$$p_{hj} = f_{hj} / f_h.$$

p_{hj} ist die relative Häufigkeit für die Klasse j in Aufnahme h . f_h steht für die Summe der Häufigkeiten in allen Klassen, also für die Gesamtzahl der Individuen der Aufnahme h . In Tabelle 4.8 ist zum Beispiel $p_{12} = f_{12} / f_2 = 9/9 = 1$.

Tabelle 4.8

Beispiel zur Berechnung der Entropie. "0" bedeutet, dass die betreffende Art fehlt, "-" bedeutet, dass sie nicht untersucht wurde.

Art j	Aufnahme h	1	2	3
1		3	9	3
2		3	0	3
3		3	0	3
4		-	-	0
5		-	-	0
$f_{h.} =$		9	9	9

Man erhält für die Entropie

$$\begin{aligned} H(P_h) &= -\sum f_{hj} / f_h \ln f_{hj} / f_h \\ &= \ln f_h - 1/f_h \sum f_{hj} \ln f_{hj} \\ &= 1/f_h (f_h \ln f_h - \sum f_{hj} \ln f_{hj}) \end{aligned}$$

Die Entropie ist ein relatives Mass (mit der Kolonnensumme f_h als Massstab). Für die weiteren Betrachtungen verwenden wir nun aber deren f_h -faches:

$$I(F_h) = H(P_h) f_h = f_h \ln f_h - \sum f_{hj} \ln f_{hj}$$

$I(F_h)$ bezeichnen wir als Informationsgehalt von h . Er eignet sich ebenfalls als Mass für die Diversität der Aufnahme h . Dessen Eigenschaften sollen anhand der drei Aufnahmen in Tab. 4.8 gezeigt werden:

1. Die Diversität wird maximal, wenn die Individuen der Aufnahme gleichmässig auf alle Arten verteilt sind. Aufnahme 1 zeigt diesen Fall: $I(F_1) = 9 \ln 9 - (3 \ln 3 + 3 \ln 3 + 3 \ln 3) = 9.89$. Das Minimum wird erreicht, wenn alle Individuen einer einzigen Art angehören, nämlich $I(F_2) = 9 \ln 9 - (9 \ln 9 + 0 \ln 0 + 0 \ln 0) = 0$. Erinnern wir uns an die Definition von $I(F_h)$ als Mass für die Ungewissheit über die Zusammensetzung einer Aufnahme: Wird bei Aufnahme 2 ein beliebiges Individuum aus der Gesamtstichprobe herausgegriffen, so kann mit absoluter Gewissheit gesagt werden, dass es der Art 1 angehört. Die Unsicherheit der Aussage ist gleich null. Im Falle der Aufnahme 1 jedoch ist die Wahrscheinlichkeit für eine Angehörigkeit zu jeder der Arten genau gleich, nämlich $1/3$. Die Ungewissheit über die Artzugehörigkeit ist also maximal.

2. Zusätzliche Arten, welche keine Individuen enthalten, tragen nicht zur Diversität bei. In Tabelle 4.8 ist also $I(F_1) = I(F_3)$.

3. Welcher Natur die Merkmale sind, ist unerheblich. Artmächtigkeit, Lebensform oder Schichtzugehörigkeit sind

gleichermaßen zulässig und erhalten gleiches Gewicht. Eine Transformation, wie sie für metrische Ähnlichkeitsmasse notwendig ist, entfällt.

4. Werden die Merkmale nach zwei verschiedenen Kriterien K und L in Klassen eingeteilt (z.B. in K, Arten und L, Schichtzugehörigkeit), so ist die gesamte Diversität gleich der Summe der Diversitäten K und L. In Tabelle 4.9, A, ist ein einschlägiger Fall konstruiert. Wir finden folgende Informationswerte:

$$I(F_{hK}) = 9 \ln 9 - (3 \ln 3 + 3 \ln 3 + 3 \ln 3) \\ = 19.77 - (3.3 + 3.3 + 3.3) = 9.88$$

$$I(F_{hL}) = 9 \ln 9 - (3 \ln 3 + 3 \ln 3 + 3 \ln 3) = 9.88$$

$$I(F_{hKL}) = 9 \ln 9 - 9(1 \ln 1) = 9 \ln 9 = 19.77$$

Innerhalb von Rundungsfehlern bestätigt sich, dass die Diversitäten addiert werden dürfen:

$$I(F_{hKL}) = I(F_{hK}) + I(F_{hL}).$$

Im Falle, dass K und L teilweise abhängig sind, ist die Gesamtdiversität kleiner als die Diversitäten K plus L:

$$I(F_{hKL}) < I(F_{hK}) + I(F_{hL}).$$

Einen Fall vollständiger Abhängigkeit zeigt Tabelle 4.9, B. Darin gilt natürlich, dass die Diversität sowohl durch das Artsspektrum wie auch durch die Schichtzugehörigkeit vollständig beschrieben werden kann:

$$I(F_{hKL}) = I(F_{hK}) = I(F_{hL}) = 9.88$$

Die bisher besprochenen Funktionen der Infomationstheorie können nun leicht verwendet werden, um Masse für den Zusammenhang von Vegetationsaufnahmen zu formulieren. Anhand des Beispiels in Tabelle 4.10 sollen verschiedene Möglichkeiten gezeigt werden. Ausgegangen wird von den beiden Aufnahmen

Tabelle 4.9

Diversitätsberechnungen. Beispiel vollständiger Unabhängigkeit von Merkmalen (A) sowie von vollständiger Abhängigkeit (B).

A

K=	L=	Kraut-schicht	Strauch-schicht	Baum-schicht	f_{hK}
Art 1		1	1	1	3
Art 2		1	1	1	3
Art 3		1	1	1	3
f_{hL}		3	3	3	$f_{h.} = 9$

B

K=	L=	Kraut-schicht	Strauch-schicht	Baum-schicht	f_{hK}
Art 1		3	0	0	3
Art 2		0	3	0	3
Art 3		0	0	3	3
f_{hL}		3	3	3	$f_{h.} = 9$

Tabelle 4.10

Berechnung des Zusammenhangs zwischen zwei Aufnahmen (A), Häufigkeit der Symbole (B) und Kontingenztafel (C).

A

Art	Aufn.	
	1	2
1	.	.
2	+	+
3	.	+
4	1	.
5	.	1
6	2	1
7	.	.
8	1	1
9	1	2
10	+	+

B

Aufn.	Klassen				$f_h.$
	.	+	1	2	
1	4	2	3	1	10
2	3	3	3	1	10

C

Klassen in Aufn. 2	Klassen in Aufn. 1				F_2
	.	+	1	2	
.	2	0	1	0	$f_{21} = 3$
+	1	2	0	0	$f_{22} = 3$
1	1	0	1	1	$f_{23} = 3$
2	0	0	1	0	$f_{24} = 1$
F_1	f_{11}	f_{12}	f_{13}	f_{14}	$f_{1.} = f_{2.} = 10$
	=4	=2	=3	=1	

in Tabelle 4.10, A. Darin sind die Artmächtigkeiten mit Hilfe von Symbolen notiert. Es liegen also Nominaldaten vor, die auch als solche behandelt werden sollen. Ohne auf Zusammenhänge zwischen den Aufnahmen einzugehen, lässt sich eine Statistik der auftretenden Klassen (Symbole) erstellen (Tabelle 4.10, B). Um nun Beziehungen darzustellen, ist eine Kontingenztafel zu konstruieren mit den Zelleninhalten $f_{1s,2t}$. Jede Zelle addiert die Anzahl Fälle, in welchen Aufnahme 1 bezüglich einer bestimmten Art das Symbol s, Aufnahme 2 das Symbol t enthält. Das leicht nachzuvollziehende Ergebnis zeigt Tabelle 4.10, C. Es ist offensichtlich, dass Aufnahme 1 und 2 teilweise abhängig sind. Schematisch ist dies in Abb. 4.4 dargestellt. Jeder Kreis entspricht dem Informationsgehalt einer Aufnahme. Es überlappen die gemeinsamen Informationsanteile, welche die Abhängigkeiten darstellen. Anhand von Tabelle 4.10, C und Abb. 4.4 lässt sich diese Abhängigkeit in folgender Weise fassen:

1. Information der Randverteilungen:

$$I(F_1) = f_{1.} \ln f_{1.} - \sum_t f_{1t} \ln f_{1t}$$

$$I(F_2) = f_{2.} \ln f_{2.} - \sum_j f_{2j} \ln f_{2j}$$

Die Summation erfolgt darin über alle Elemente der jeweiligen Randwerte. Die beiden Randverteilungen haben mit dem Zusammenhang der Aufnahmen nichts zu tun, sondern beschreiben deren Informationsgehalt (bezüglich der Häufigkeiten vorkommenden Symbole) separat. Dies ist in Abb. 4.4, A, dargestellt.

2. Gesamtinformation der Kontingenztafel:

$$I(F_1, F_2) = f_{1.,2.} \ln f_{1.,2.} - \sum_s \sum_t f_{1s,2t} \ln f_{1s,2t}$$

Hier erfolgt die Summation über alle Elemente innerhalb der Kontingenztafel. Grafisch ist der Fall in Abb. 4.4, B, dargestellt. Sind nun unsere beiden Aufnahmen unabhängig, so kann die Gesamtinformation auch aus den Randverteilungen berechnet werden:

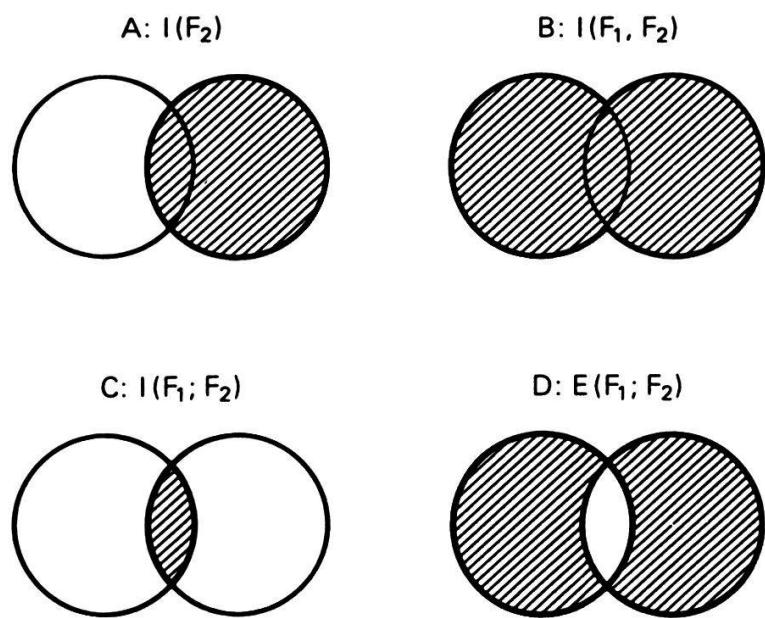


Abb. 4.4 Grafische Darstellung verschiedener Informationsmasse: Information einer Randverteilung (A), Gesamtinformation (B), Wechselseitige Information (C) und Spezifische Information (D).

$$I(F_1, F_2) = I(F_1) + I(F_2)$$

Grafisch dargestellt dürften sich dann die beiden Kreise in Abb. 4.4 nicht überschneiden. Sind hingegen beide Aufnahmen genau gleich, so müssen sich die Kreise decken. Die Information einer Randverteilung (einer Aufnahme) entspricht dann zugleich der Gesamtinformation:

$$I(F_1, F_2) = I(F_1) = I(F_2) .$$

Die Gesamtinformation ist also bereits ein Mass für die Uebereinstimmung der beiden Aufnahmen.

3. Wechselseitige Information:

$$I(F_1; F_2) = I(F_1) + I(F_2) - I(F_1, F_2)$$

Anhand der Abb. 4.4, C, ist leicht einzusehen, dass es sich hier um den beiden Aufnahmen gemeinsamen Teil der Information handelt. Bei vollständiger Unabhängigkeit ergibt sich natürlich der Grenzwert 0:

$$I(F_1; F_2)_{\min} = I(F_1) + I(F_2) - I(F_1) - I(F_2) = 0$$

Bei Identität der Aufnahmen ist die wechselseitige Information gleich derjenigen einer einzelnen Aufnahme:

$$I(F_1; F_2)_{\max} = I(F_1) + I(F_2) - I(F_1) = I(F_2) = I(F_1) .$$

Damit qualifiziert sie sich als Aehnlichkeitsmass mit dem Bereich

$$0 \leq I(F_1; F_2) \leq I(F_1), I(F_2) .$$

4. Spezifische Information

$$E(F_1; F_2) = I(F_1, F_2) - I(F_1; F_2)$$

Es handelt sich um die Information, welche jede Aufnahme für

sich allein trägt (Abb. 4.4, D). Bei Unabhängigkeit gilt, dass sie der Summe der Informationen beider Aufnahmen entspricht:

$$E(F_1; F_2)_{\max} = I(F_1) + I(F_2).$$

Bei volliger Identität wird sie dagegen 0:

$$E(F_1; F_2)_{\min} = I(F_1) - I(F_2) = 0.$$

Mit der spezifischen Information verfügt man somit über ein Unähnlichkeitsmass mit den Grenzwerten

$$I(F_1) + I(F_2) \geq E(F_1; F_2) \geq 0.$$

5. Rajskis Divergenzkoeffizient:

Um ein relatives Mass zu erhalten, schlug RAJSKI (1961) einen Koeffizienten vor, der den Anteil der spezifischen Information an der Gesamtinformation misst:

$$d(F_1; F_2) = \frac{E(F_1; F_2)}{I(F_1, F_2)} = 1 - \frac{I(F_1; F_2)}{I(F_1, F_2)}$$

Die Grenzwerte von 0 (Identität) und 1 (Unabhängigkeit) sind somit von der Gesamtinformation unbeeinflusst.

Auf das Beispiel in Tabelle 4.10 angewandt ergeben die hier besprochenen Masse folgende Werte:

1. Information der Randverteilungen:

$$I(F_1) = 10 \ln 10 - (4 \ln 4 + 2 \ln 2 + 3 \ln 3 + 1 \ln 1) = 12.798$$
$$I(F_2) = 10 \ln 10 - (3(3 \ln 3) + 1 \ln 1) = 13.138$$

2. Gesamtinformation:

$$I(F_1, F_2) = 10 \ln 10 - (2(2 \ln 2) + 6(1 \ln 1)) = 20.253$$

3. Wechselseitige Information:

$$I(F_1; F_2) = 12.798 + 13.138 - 20.253 = 5.683$$

4. Spezifische Information:

$$E(F_1; F_2) = 20.253 - 5.683 = 14.57$$

5. Rajskis Koeffizient:

$$d(F_1; F_2) = 14.57 / 20.253 = 0.719$$

Die hier gezeigten Informationsmasse behandeln die Daten so, als wären sie nominal. Beurteilt wird die Zufälligkeit des Zusammentreffens zweier Symbole wie "+" oder "1" der Skala Braun-Blanquet. Das ist ein ausserordentlich strenges Vergleichskriterium. Wie sich am Beispiel der Tabelle 4.10 leicht nachprüfen lässt, können Vegetationsaufnahmen als völlig zusammenhangslos eingestuft werden, wenn sie zwar lauter gemeinsame Arten, jedoch unterschiedliche Artmächtigkeiten aufweisen. Informationsmasse sollten eher angewandt werden, wenn mit echt qualitativen Kriterien gearbeitet wird, also Schichtzugehörigkeit, Wuchsform oder Phänologie, aber auch bei Präsenz- Absenz Daten.

Die Informationsmasse bieten eine beachtliche Auswahl von Möglichkeiten für die Beschreibung der Ähnlichkeit von Aufnahmen. Dabei fällt sofort ihre Verwandtschaft zu den Koeffizienten der Kontingenztafel auf (Abschnitt 4.7), wo die Felder a und d Gemeinsamkeiten, b und c Divergenzen repräsentieren. FEOLI et al. (1984) weisen denn auch nach, dass sich die meisten Informationsmasse im Falle von Präsenz- Absenz Daten auf das Chiquadrat oder ein Derivat davon reduzieren. Für Datensätze mit komplizierter Merkmalsstruktur (z.B. Arten und Lebensformen gemischt) bieten sich die Informationsmasse als ideale Lösung an. Beispiele für Anwendungen in grösserem Rahmen gibt es kaum und es ist in dieser Hinsicht noch einige Pionierarbeit zu leisten.

5. Gruppierungsanalysen

Bei den Gruppierungsanalysen geht es darum, Stichproben, also z.B. eine Reihe von Vegetationsaufnahmen, in Gruppen möglichst ähnlicher Individuen zu unterteilen. Ist eine solche Analyse erfolgreich, so kann in der Folge anstatt mit den Eigenschaften der Individuen mit denjenigen der viel weniger zahlreichen Gruppen weitergearbeitet werden. Dies erleichtert das Erkennen von Zusammenhängen und Gesetzmäßigkeiten innerhalb eines Datensatzes. In der Vegetationskunde gibt es Theorien, welche von der Existenz organismusähnlicher Pflanzengesellschaften und mithin natürlicher Gruppen ausgehen. Analysen dienen dabei der Erkennung "echter" Pflanzengesellschaften. Andere Auffassungen gehen dahin, die Vegetationsdecke der Erde als "Kontinuum" zu betrachten (GLEASON 1926, 1939). Logisch gefolgt müssten Gruppierungsanalysen bevorzugte Methoden der Anhänger diskreter Pflanzengesellschaften sein. Wie nun aber im folgenden zu zeigen ist, können mit geeigneten Methoden auch gradientenhafte Datenstrukturen analysiert werden. Die Resultate sind in jedem Falle vom Datensatz einerseits, vom Gruppierungsalgorithmus andererseits abhängig.

5.1 Gruppenstruktur

Zunächst sollen verschiedene Möglichkeiten von Gruppenstrukturen erörtert werden. Der einfachste Fall ist in Abb. 5.1 dargestellt. Diese zweidimensionale Struktur ist kontinuierlich und einigermassen linear. Die gesamte Wolke von Punkten kann sinnvollerweise nur als eine einzige natürliche Gruppe aufgefasst werden. Während gewisse Gruppierungsmethoden tatsächlich zu diesem Ergebnis kommen, lassen andere eine weitergehende Unterteilung zu (gestrichelte Trennlinien in Abb. 5.1). Dabei kann von natürlichen Gruppen nicht die Rede sein. Eine willkürliche Unterteilung einer an sich kompakten Punktwolke kann aber sinnvoll und notwendig sein, wenn eine gradientenhafte Ähnlichkeitsstruktur vorliegt.

Ein wichtiger Spezialfall kontinuierlicher Strukturen ist die mehrdimensionale Normalverteilung. Um eine solche erken-

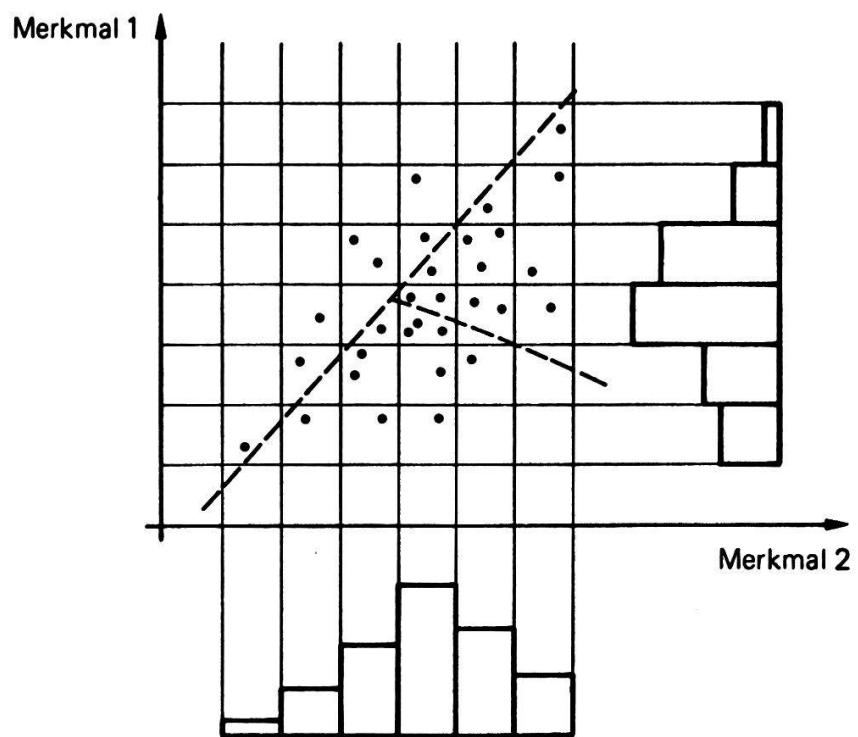


Abb. 5.1 Angenähert multivariat-normalverteilte Stichprobe.
Die Balkendiagramme zeigen die Häufigkeitsverteilungen
bezüglich der Merkmale 1 und 2. Gestrichelt ist eine
Möglichkeit zur Unterteilung angedeutet.

nen zu können, werden alle Achsen in Klassen unterteilt und die Anzahl der Individuen pro Klasse in Balkendiagrammen dargestellt (Abb. 5.1). Die so entstehenden Häufigkeitsverteilungen müssen bezüglich jeder Dimension durch eine Glockenkurve angenähert werden können. In Vegetationstabellen tritt eine solche Struktur fast ausschliesslich bei Artvergleichen auf, beim Vergleich von Aufnahmen dagegen fast nie. Aus diesem Grunde sind zur Gruppierung von Arten und Aufnahmen gelegentlich verschiedene Methoden zu verwenden.

Recht häufig sind Strukturen, wie sie in Abb. 5.2, A, dargestellt sind (vgl. auch BARTEL 1974, S. 83). Sie sind kontinuierlich, oft ausgesprochen länglich und gekrümmt. Bei deren Analyse sind all jene Methoden ungeeignet, welche speziell mehrdimensional normalverteilte Gruppen auseinanderhalten können. Wirkungsvoller sind hier Methoden, die mehr oder weniger willkürliche, eher kompakte Gruppierungen herbeiführen.

Abb. 5.2, B, zeigt den Fall disjunkter "natürlicher" Gruppen. Methoden, die solche Strukturen aufzudecken vermögen, gibt es zahlreiche, doch können viele davon durch intermediäre Individuen (dünner Pfeil in Abb. 5.2, B) gestört werden. Aberrante Individuen (dicker Pfeil) lassen sich dagegen leicht aufspüren und werden in der Regel als eigene Gruppe betrachtet.

Damit ist das Spektrum in der Pflanzensoziologie auftretender Strukturen noch nicht erschöpft. Fast immer zeigt es sich, dass irgendwelche Kombinationen der in Abb. 5.2, A und B, gezeigten Fälle vorliegen. Abb. 5.2, C, ist ein Beispiel. In der Tat gibt es Methoden, die auch diese Konfiguration als aus drei Gruppen bestehend erkennt.

Die hier gezeigten Beispiele lassen die Vermutung aufkommen, dass die visuelle Gruppierung einer Stichprobe viel rascher und sicherer sein könnte, als eine numerische. SPÄTH (1977) erwähnt, dass dies für ein- und zweidimensionale Strukturen zutrifft. Liegen jedoch drei und mehr Dimensionen vor, so werden numerische Verfahren rasch überlegen. Bei

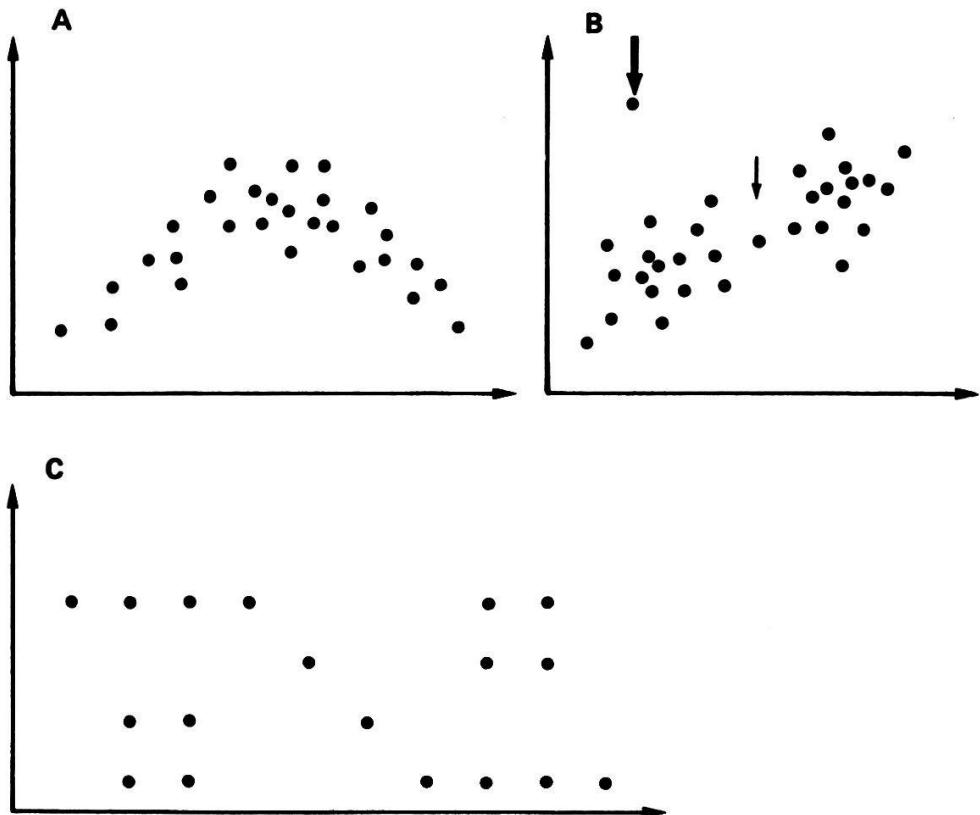


Abb. 5.2 Verschiedene Gruppenstrukturen. A: Langgezogene, gekrümmte Gruppe. B: Zwei getrennte Gruppen mit intermediärem (dünner Pfeil) und aberrantem Individuum (fetter Pfeil). C: Zwei quadratische und eine langgezogene Gruppe.

vegetationskundlichen Anwendungen mit oft hundert und mehr Dimensionen - nämlich Arten - bilden sie den einzigen Weg zu nachvollziehbaren Ergebnissen.

5.2 Heuristische Verfahren

Heuristische Verfahren erfüllen die Qualitätsanforderungen nicht, welche normalerweise an multivariate Verfahren gestellt werden. In der Regel basieren sie nicht auf einer vollständigen Ähnlichkeitsmatrix. Damit kann, streng genommen, die wirkliche Gruppenstruktur einer Stichprobe gar nicht vollständig erfasst werden. Vorteile sind dagegen ein geringer Rechen- und Speicheraufwand. Tausende von Vegetationsaufnahmen lassen sich rasch und leicht provisorisch gruppieren. Beispiel eines iterativen Verfahrens dieser Art ist TABORD (VAN DER MAAREL et al. 1978), welches Vegetationsstabellen grösseren Ausmasses zu strukturieren erlaubt.

Einen typischen, ausgesprochen einfachen Vertreter heuristischer Verfahren finden wir bei ANDERBERG (1973) und SPÄTH (1977). Er wird als LEADER-Algorithmus bezeichnet, lehnt sich stark an ein intuitives Vorgehen an und dient bei VAN DER MAAREL et al. (1978) als erstes grobes Ordnungsverfahren. Als Ähnlichkeitsmass wird oft die Euklidsche Distanz verwendet. Es gelte folgende Notation:

i i = 1,...,n ist die momentan zu verarbeitende Aufnahme bei einem Total von n;

rho ist eine vom Benutzer zu definierende Distanz. Sie begrenzt die in einer Gruppe auftretenden Unterschiede zwischen den Aufnahmen;

NMAX ist die maximal erlaubte Anzahl Gruppen.

Die LEADER-Methode verläuft sodann wie folgt:

1. Die erste Aufnahme des Datensatzes ($i=1$) wird erste Aufnahme (Leitaufnahme) der ersten Gruppe.
2. i wird um 1 erhöht. Die nächstfolgende Aufnahme i wird verarbeitet (im ersten Durchgang ist $i = 2$). Dazu wird die Euklidsche Distanz zu den ersten Aufnahmen (Leitaufnahmen) der bereits bestehenden Gruppen berechnet.
3. Aufnahme i wird der ersten Gruppe zugeordnet, zu deren Leitaufnahme die Distanz kleiner ist als ρ . Falls dies gelingt, wird mit 2. weitergefahrt.
4. Ist die Distanz von i zu allen Leitaufnahmen grösser als ρ , so wird i zur Leitaufnahme einer neuen Gruppe. Das Verfahren wird mit 2. fortgesetzt.
5. Uebersteigt die Zahl der Gruppen N_{MAX} , so werden die noch nicht verarbeiteten Aufnahmen nicht klassifiziert. ρ sollte etwas vergrössert werden. Der ganze Gruppierungsprozess ist zu wiederholen (1.), bis alle Aufnahmen klassifiziert sind.

Der Ablauf lässt sich noch vereinfachen, indem die Zahl der Gruppen nicht festgelegt wird. Schwerwiegendster Nachteil der Methode ist, dass das Resultat von der Reihenfolge der Eingabe der Aufnahmen abhängt. Das illustrieren die Beispiele in Abb. 5.3. ρ sei gleich 2. Im Falle A sei die Reihenfolge der Aufnahmen $(1,2,3,4,5)$. Die erste in den Prozess eingeschleuste Aufnahme wird zur Leitaufnahme der ersten Gruppe. Aufnahme 2 besitzt Distanz $d = 1$ zu dieser und kommt damit ebenfalls in Gruppe 1. Aufnahme 3 besitzt $d = 5^{1/2} = 2.24$ zu Aufnahme 1 und wird Leitaufnahme der Gruppe 2. Aufnahme 4 wird letzterer zugeordnet und Aufnahme 5 schliesslich bildet eine selbständige Gruppe 3. Im Falle B soll versuchsweise die Reihenfolge der Aufnahmen so geändert werden, dass Nummer 3 an erster Stelle steht. Es gilt also $3,1,2,4,5$. Leitaufnahmen werden zunächst 3 und 1. Aufnahme 2 wird der Leitaufnahme der Gruppe 1 zugeordnet, also Aufnahme 3. Auch Aufnahme 4 gehört zu Gruppe 1, während Aufnahme 5 eine neue Gruppe 3 bildet.

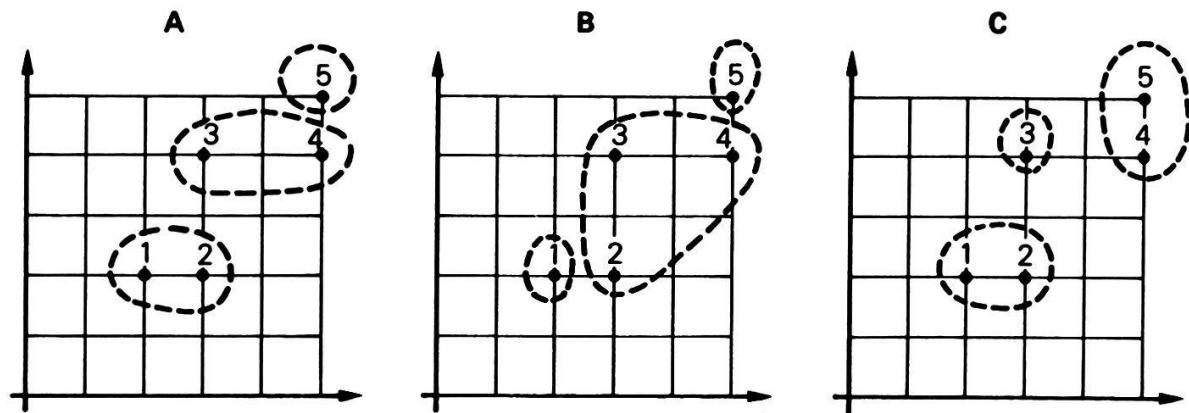


Abb. 5.3 Verschiedene Lösungen des Leader-Algorithmus. A und B ergeben sich durch unterschiedliche Reihenfolge in der Verarbeitung der Aufnahmen. Lösung C resultiert, wenn der maximal zulässige Durchmesser der Gruppen (rho) reduziert wird.

Tabelle 5.1

Beispiel für die Durchführung der Assoziationsanalyse (A). Aufbau der Kontingenztafel zur Berechnung von Chiquaret zwischen den Arten 1 und 2 (B).

Art	Aufnahme				
	1	2	3	4	5
1	1	1	0	0	0
2	1	1	1	0	0
3	1	1	1	1	0
4	0	0	0	1	1

Art 1	Art 2		$a+b=3$
	+	-	
+	$a=2$	$b=1$	$a+b=3$
-	$c=0$	$d=2$	$c+d=2$
	$a+c=2$	$b+d=3$	$N = 5$

Das Resultat der Analyse ist stark von der Wahl von rho abhangig. Setzt man z.B. rho = 1.5, so erhält man die visuell leicht als optimal erkennbare Lösung mit den Gruppen (1,2), (3), (4,5). Für das LEADER-Verfahren gibt es zahlreiche Verbesserungsvorschläge und Alternativen (ANDERBERG 1973), doch sind die Resultate immer von der ursprünglichen Reihenfolge der Aufnahmen abhängig. Heuristische Verfahren werden ihrer Einfachheit wegen eingesetzt, um sehr grosse vegetationskundliche Datensätze zu strukturieren. Das Beispiel in Abb. 5.3 zeigt deren Grenzen auf. Eine provisorische Gliederung ist aber doch zu erzielen. Sie kann in einem anschliessenden Schritt korrigiert und optimiert werden (z.B. im Programm CLUSLA, LOUPPEN und VAN DER MAAREL (1979)).

5.3 Teilungsverfahren

5.3.1 Assoziationsanalyse

Bei Teilungsverfahren wird versucht, für die Gesamtstichprobe von Individuen eine möglichst sinnvolle Unterteilung zu finden. Als typischer Vertreter soll zuerst die Assoziationsanalyse dargestellt werden, die von WILLIAMS und LAMBERT (1959) publiziert und später mehrfach variiert wurde. Es handelt sich um ein monothetisches Verfahren, d.h. jede Unterteilung wird nur aufgrund der Gegenwart oder des Fehlens einer einzelnen Art durchgeführt. Immerhin wird zur Auswahl dieser Art die Gesamtähnlichkeitsstruktur der Vegetationstabelle berücksichtigt. Ausgangspunkt bildet eine Ähnlichkeitsmatrix S der Arten mit den Elementen

$$S_{ij} = x_{ij}^2 / N.$$

Das Chiquaret berechnet sich wie in Kapitel 4.5 gezeigt. N ist die Anzahl Aufnahmen der Vegetationstabelle, die Division durch N hat daher auf das Ergebnis keinen Einfluss. Anhand von Tabelle 5.1, A, soll die Berechnung gezeigt werden. Für die Arten 1 und 2 konstruiert man die Kontingenztafel der Tabelle 5.1, B. Daraus ergibt sich

$$\chi^2(1,2) = \frac{(ad - bc)^2 N}{(a+b)(a+c)(c+d)(b+d)} = \frac{(4 - 0)^2 5}{3*2*2*3} = \frac{80}{36} = 2.22$$

In der Originalversion der Assoziationsanalyse wird die Unabhängigkeit der Arten - unbesehen von Voraussetzungen der schliessenden Statistik - anhand einer χ^2 -Tabelle auf Signifikanz getestet (Anzahl Freiheitsgrade $df = 1$). Da N in unserem Beispiel sehr klein ist, müsste χ^2 dazu erst korrigiert werden. Entsprechende Formeln finden sich bei MÜLLER-DOMBOIS und ELLENBERG (1974) und PIELOU (1977). Nicht signifikante Werte werden meist durch null ersetzt. Als Element der S - Matrix erhalten wir zum Beispiel

$$S(1,2) = \chi^2(1,2)/N = 2.22/5 = 0.44$$

In gleicher Weise werden nun die andern Elemente berechnet. Man setzt die Werte der Diagonalen gleich null und erhält

$$S = \begin{matrix} 0 & 0.44 & 0.166 & 0.44 \\ 0.44 & 0 & 0.375 & 1.00 \\ 0.166 & 0.375 & 0 & 0.375 \\ 0.44 & 1.00 & 0.375 & 0 \end{matrix}$$

Wie in Kapitel 4.5 erwähnt, lässt sich χ^2 in relativierter Form als Korrelationskoeffizient für die Kontingenztafel verstehen (V - Wert). Um nun diejenige Art zu finden, die den grössten gemeinsamen Zusammenhang mit allen andern Arten aufweist, müssen nur die Kolonnen (oder Zeilen) in S aufaddiert zu werden. Wir erhalten

$$S_{.j} = 1.046, 1.815, 0.916, 1.815$$

Das maximale Chiquadrat weisen Art 2 und 4 auf. Anhand ihrer Gegenwart oder Abwesenheit in den Aufnahmen wird Tabelle 5.1, A, unterteilt in die Gruppen (1,2,3) und (4,5). Das Verfahren geht weiter, indem beide der neu gebildeten Gruppen in derselben Weise analysiert werden. Da Art 2 sicher zu keiner weiteren Unterteilung zu verwenden ist,

kann sie (und im vorliegenden Fall auch Art 4) weggelassen werden. Die beiden neuen S-Matrizen, welche zur Suche der nächsten zwei Trennarten dienen, haben deshalb immer eine um 1 verminderte Dimension. Beim vorliegenden Beispiel ist es jedoch sinnvoll, die Analyse hier abzubrechen.

Die Gruppen der Assoziationsanalyse sind streng hierarchisch gegliedert. Weil sie auf eindeutigen Trennarten beruhen, lassen sie sich sehr leicht identifizieren. Für eine einfache Unterteilung wird eine einzige Art benötigt, für vier Gruppen mindestens zwei Arten, für acht Gruppen mindestens drei Arten usw. Da aber eine einzige Art über die Gruppenzugehörigkeit entscheidet, ist das Verfahren stark von Zufälligkeiten abhängig. Wie zu erwarten ist, befriedigt die so entstehende Struktur pflanzensoziologisch kaum, krasse Fehlklassifikationen wegen unerwartet auftretender oder abwesender Arten sind die Regel. Günstiger liegen die Verhältnisse, wenn statt der Artmächtigkeiten Koordinaten aus Hauptkomponentenanalysen verwendet werden (NOY MEIR 1973). Auf diese Weise entfällt jedoch die Möglichkeit einfacher Identifikation. Um pflanzensoziologische Einheiten trotzdem mit Hilfe weniger Arten charakterisieren zu können, bieten sich Rangierungsmethoden an (vgl. dazu Kap. 6 sowie HILL (1979b)).

Die Assoziationsanalyse entspricht ihrer Idee nach dem Konzept der Charakter- und Differentialarten des Systems Braun-Blanquet. Wohl aus diesem Grunde hat sie einen bemerkenswerten Bekanntheitsgrad erlangt.

5.3.2 Gridanalyse

Die Gridanalyse versucht, die hauptsächlichsten Nachteile der Assoziationsanalyse zu vermeiden (WILDI 1979). Aus diesem Grunde soll nicht eine, sondern es sollen mehrere Dimensionen gleichzeitig zur Abgrenzung von Gruppen beigezogen werden. Gesucht werden echte Gruppen, welche einer lokalen Anhäufung ähnlicher Individuen entsprechen (Noda im Sinne von POORE 1955). Wir verfolgen den Algorithmus anhand von Abb. 5.4. Der Übersichtlichkeit halber wird nur ein

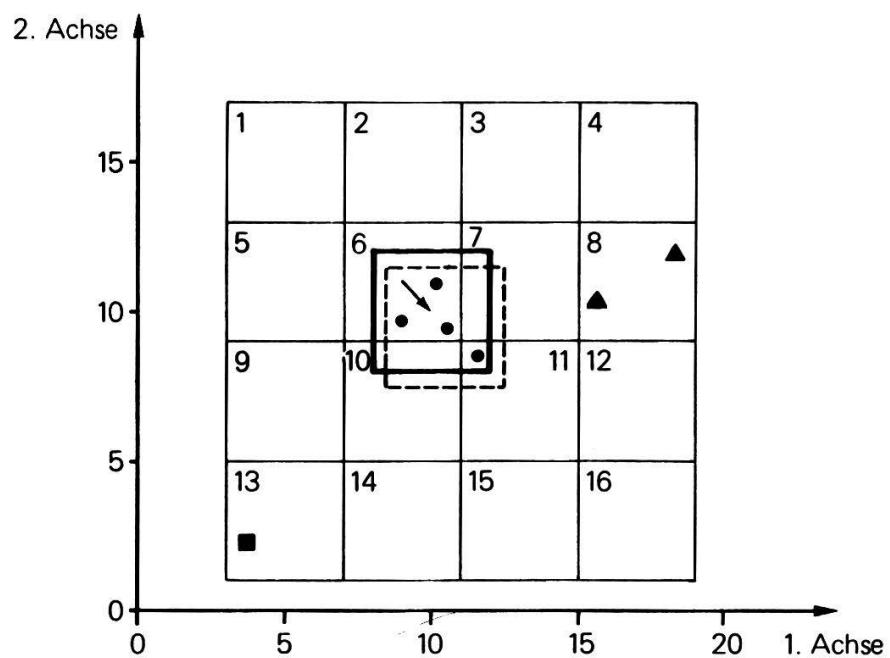


Abb. 5.4 GRID-Analyse im zweidimensionalen Fall.

zweidimensionaler Fall dargestellt.

x_{ij} sei die Koordinate bezüglich der i -ten Achse ($i = 1, \dots, p$; Achsen können Arten, Standortsmessungen, Ordinationsachsen usw. sein) der Aufnahme j , $j = 1, \dots, n$. Es sind folgende Operationen durchzuführen:

1. Für jede Dimension ist der Bereich δ_i festzulegen.

$$\delta_1 = \max(x_{1j}) - \min(x_{1j}), \quad j = 1, \dots, n$$

$$\delta_2 = \max(x_{2j}) - \min(x_{2j}), \quad j = 1, \dots, n$$

In Abb. 5.4 ist $\delta_1 = 15$, $\delta_2 = 10$.

2. Bestimme $\delta_{\max} = \max(\delta_1, \dots, \delta_p)$. In Abb. 5.4 ist $\delta_{\max} = \delta_1 = 15$.

3. Lege ein p -dimensionales Netz mit der Seitenlänge von mindestens δ_{\max} über den Aufnahmeraum, so dass alle Individuen innerhalb des Netzes liegen. Die Auflösungskraft wird durch die Anzahl Unterteilungen jeder Dimension bestimmt ($m = 4$ in Abb. 5.4).

4. In den entstandenen $m^p = 16$ Zellen werden die Individuen gezählt. Wir erhalten

Zelle	Individuen
6	3
8	2
11	1
13	1

5. Die individuenreichste, noch nicht verarbeitete Zelle (Nr. 6) wird geprüft, ob sie ein lokales Dichtemaximum enthält, also Zentrum einer echten Gruppe darstellt. Dazu wird kontrolliert, ob nicht eine der Nachbarzellen (2,5,7,10) bereits ein Zentrum enthält. Im Beispiel bleibt Zelle Nr. 6 Zentrum der ersten Gruppe.

6. Weil das p-dimensionale Netz willkürlich in den Raum gelegt wurde, wird nun die genauere Lage der Gruppe 1 gesucht. Dazu ist die aktuelle Zelle in Richtung des Schwerpunkts der in ihr enthaltenen Individuen zu verschieben (Pfeil, fett ausgezogenes Quadrat). Enthält die verschobene Zelle mehr Individuen, so wird weiter geschoben. In Abb. 5.4 kommt ein Individuum in Zelle 11 dazu, so dass das erste Gruppenzentrum durch 4 Individuen repräsentiert wird.
7. Nun wird die Zelle mit der nächstniedrigeren Anzahl Individuen verarbeitet (Schritte 5 und 6). Dies wäre nach dem ersten Durchgang Zelle 8.

Der Prozess wird so lange fortgesetzt, bis alle Zellen abgearbeitet sind. Einzelindividuen können den Status einer unabhängigen Gruppe erhalten, oder aber dem nächstgelegenen Gruppenzentrum zugeordnet werden. In Abb. 5.4 erhalten wir somit 3 Gruppen, erkennbar an den unterschiedlichen Symbolen.

Die GRID-Analyse liefert natürliche, nicht hierarchische Gruppen. Dies allerdings nur, wenn die Auflösung zweckmäßig gewählt wird. Man sollte daher zuerst mit einer niedrigen Zellenzahl (d.h. niedriger Auflösung) beginnen. Steigt die Gruppenzahl mit stetig wachsender Zellenzahl sprunghaft an, so sind vermutlich echte Gruppen halbiert worden. Im weiteren ist zu bemerken, dass die Zahl der Dimensionen aus Gründen der Übersichtlichkeit nicht zu gross gewählt werden sollte. Das Verfahren hat sich bewährt, wenn statt einfacher Artmächtigkeiten aus Hauptkomponentenanalysen stammende Koordinaten verwendet werden (WILDI 1979).

Die GRID-Analyse eignet sich zum Auffinden diskreter Gruppen. Sie hat gegenüber anderen Verfahren den Vorteil, dass lokale Verdichtungen von Punkten als Gruppenzentren interpretiert werden, sodass intermediäre Punkte (vgl. Abb. 5.2) nicht störend wirken. Ist der zu untersuchende Datensatz genügend gross, so können Gruppen fast beliebiger Form entdeckt werden. Beschränkungen ergeben sich aus der prak-

tisch begrenzten Anzahl von Dimensionen, die sich noch sinnvoll bearbeiten lassen. Die GRID-Analyse ist deshalb nicht geeignet, um strukturelle Details zu untersuchen.

5.4 Agglomerative Verfahren

Bei dieser Kategorie von Gruppierungsverfahren werden schrittweise Individuen - später Gruppen von Individuen - zu neuen Gruppen zusammengeschlossen. Dabei können Dendrogramme gebildet werden. Diese dienen der übersichtlichen Darstellung von Resultaten hierarchischer Gruppierungsmethoden divisiver oder agglomerativer Art. Die Sachverhalte lassen sich anhand der einfachsten agglomerativen Methode demonstrieren, nämlich der Single Linkage Analysis (ANDERBERG 1973).

5.4.1 Single Linkage Analysis

Das Prinzip der Single Linkage Analysis lässt sich am univariaten Fall verfolgen. Abb. 5.5, A zeigt ein Beispiel. Jede der 4 Aufnahmen wird charakterisiert durch ein einziges Merkmal, entsprechend den folgenden Werten:

Merkmal	Aufn.	1	2	3	4
1		2	4	7	8

Im ersten Schritt muss nun eine Ähnlichkeitsmatrix berechnet werden. Die Euklidische Distanz hat gegenüber andern Massen den Vorteil, dass sie direkt aus Abb. 5.5, A, herausgelesen werden kann. Für $d_{1,2}$ erhält man 2, für $d_{2,3} = 3$, $d_{3,4} = 1$ usw. Die Gesamtähnlichkeitsstruktur ergibt:

$$D = \begin{matrix} & 0 & 2 & 5 & 6 \\ 0 & & 2 & 0 & 3 & 4 \\ 2 & & 5 & 3 & 0 & 1 \\ 5 & & 6 & 4 & 1 & 0 \end{matrix}$$

Die Gruppierung erfolgt nach einer einzigen Vorschrift: Es sind stets jene 2 Gruppen zu einer neuen Gruppe zusammen-

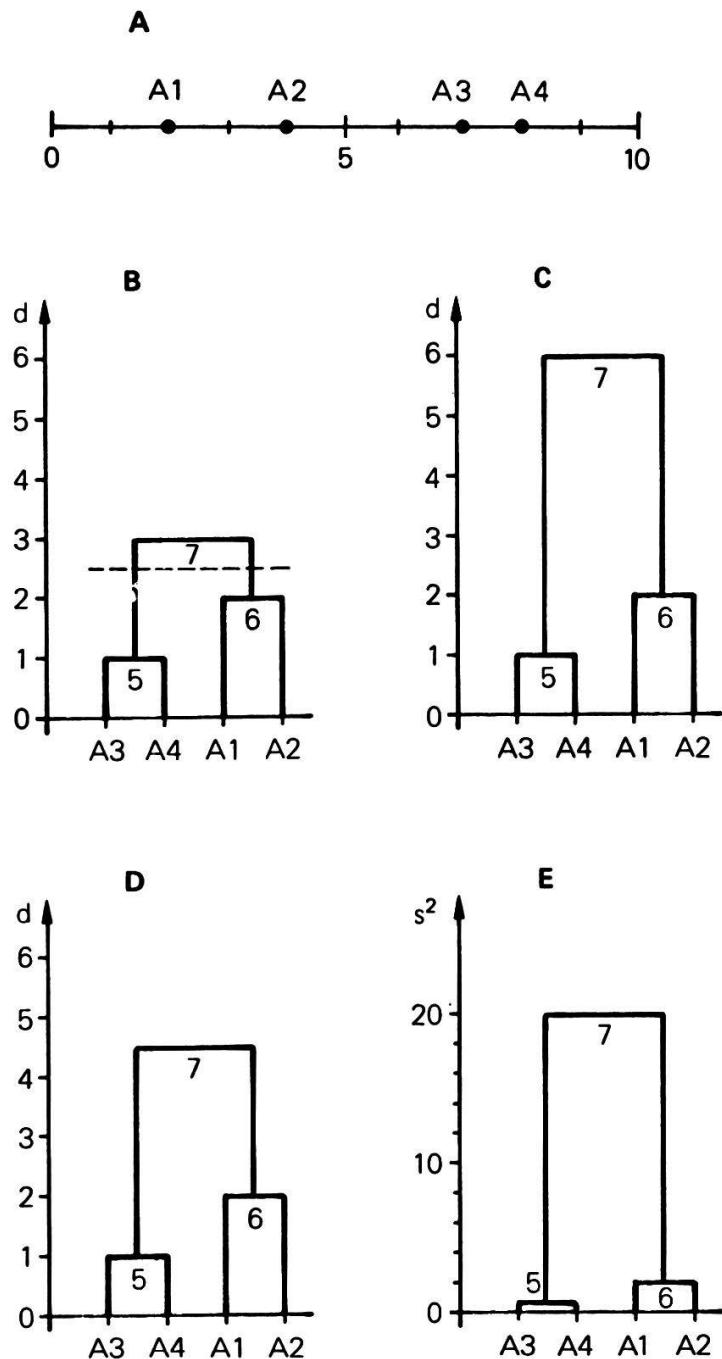


Abb. 5.5 Vier univariat charakterisierte Punkte (A) mit verschiedenen Methoden gruppiert: Single Linkage Analysis (B), Complete Linkage Analysis (C), Average Linkage Analysis (D) und Minimalvarianzanalyse (E).

zufassen, welche sich bezüglich ihrer **ähnlichsten** Individuen am nächsten stehen. Zu Beginn der Analyse finden wir in Abb. 5.5 4 Gruppen, die aus je einem einzigen Individuum bestehen. Der grafischen Darstellung, aber auch der Distanzmatrix entnehmen wir, dass A3 und A4 am ähnlichsten sind und mit $d_{3,4} = 1.0$ zu einer neuen Gruppe zusammengeschlossen werden können. Diese soll zweckmässigerweise die Nummer 5 erhalten. Im Dendrogramm (Abb. 5.5, B) sind die beiden Aufnahmen als erste aufgezeichnet und durch einen Bügel der Höhe $d = 1$ miteinander verbunden. Der zweite Zusammenschluss gestaltet sich komplizierter. Zu prüfen sind nun noch 3 Distanzen, nämlich $d_{1,2}$, $d_{2,5}$ und $d_{1,5}$. $d_{1,2}$ kann sofort der Distanzmatrix entnommen werden. Für $d_{1,5}$ gilt bei der Single Linkage Analysis, dass der Wert von $d_{1,3} = 5$ gewählt werden muss, da A3 der Gruppe 1 näher steht als A4. Entsprechend gilt $d_{2,5} = d_{2,3} = 3$. Zusammengefasst erhält man:

$$\begin{aligned}d_{1,2} &= 2 \\d_{1,5} &= 5 \\d_{2,5} &= 3.\end{aligned}$$

Damit werden A1 und A2 als nächste Gruppe, Nr. 6, auf dem Niveau $d = 2$ zusammengeschlossen. Nun müssen noch Gruppe 5 und 6 zusammengeschlossen werden. Um das Niveau des Zusammenschlusses zu finden, ist die gesamte Distanzmatrix zu durchsuchen. Für $d_{5,6}$ kommen folgende Werte in Frage:

$$\begin{aligned}d_{1,3} &= 5 \\d_{1,4} &= 6 \\d_{2,3} &= 3 \\d_{2,4} &= 4\end{aligned}$$

Als nächststehende Nachbarn der beiden Gruppen qualifizieren sich A2 und A3 mit $d_{2,3} = 3$. Auf diesem Niveau wird in Abb. 5.5, B die neue Gruppe 7 gebildet. Damit ist die Analyse abgeschlossen. Ihr Resultat ist ein Dendrogramm, welches über die Gruppenstruktur der Stichprobe (A1,A2,A3,A4) Aufschluss gibt. Meist besteht das Ziel der Analyse darin, eine bestimmte Anzahl von Gruppen, sagen wir 2, zu generieren. Zu diesem Zwecke ist das Dendrogramm zu zerschneiden, und zwar

zwischen $d = 2$ und $d = 3$ (gestrichelte Linie in Abb. 5.5, B). Es resultieren (A3,A4) und (A1,A2) als Gruppen.

Die Anwendungsmöglichkeiten der Single Linkage Analysis sollen beim Vergleich verschiedener Methoden erörtert werden (Kap. 5.4.5). Hier ist noch zu vermerken, dass es zahlreiche Varianten und Erweiterungen gibt. JANCEY (1974) schlägt eine Methode vor, bei welcher die Anzahl resultierender Gruppen vorzugeben ist. Kann aufgrund der Stichprobenstruktur eine Lösung mit natürlichen Gruppen gefunden werden, so erfolgt die Unterteilung des Dendrogrammes automatisch. Andernfalls wird die Zahl der Gruppen durch den Algorithmus selbst verändert.

Eine herausragende Rolle spielt die Single Linkage Analysis in der Geographie, und zwar vor allem im zweidimensionalen Fall zur Lösung des "Nächster Nachbar"-Problems. Abb. 5.6 zeigt eine Karte mit 5 Ortschaften. Diese sind so miteinander zu verbinden, dass

1. jeder Punkt mindestens einmal verbunden wird;
2. keine Schleifen auftreten;
3. die Summe aller Verbindungsstrecken minimal ist.

Nach GOWER und ROSS (1969) liefert die Single Linkage Analysis unmittelbar die Lösung. Das so entstehende Gebilde (Abb. 5.6) heisst Minimalbaum. Es kann derart unterteilt werden, dass bei pflanzensoziologischen Datensätzen ausgesprochene Gradientenstrukturen aufzufinden sind (KUHN 1983).

5.4.2 Complete Linkage Analysis

Die Complete Linkage Analysis ergibt sich durch ganz geringe Änderung des Single Linkage Algorithmus. Die Vorschrift lautet: Es sind stets jene 2 Gruppen zu einer einzigen, neuen Gruppe zusammenzufassen, welche sich bezüglich ihrer unähnlichsten Individuen am nächsten stehen. Am Beispiel der Abb. 5.5, C, soll der Ablauf verfolgt werden. Der erste Schritt (Gruppe 5) verläuft dabei genau gleich wie bei der

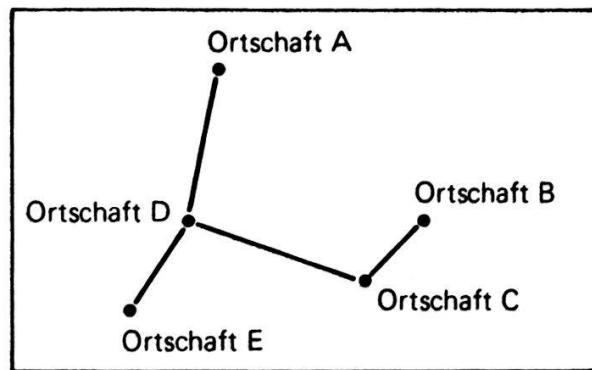


Abb. 5.6 Lösung des Problems "nächster Nachbar" im zweidimensionalen (geografischen) Falle.

Single Linkage Analysis, was sich auch im Dendrogramm niederschlägt. Auch im zweiten Schritt sind wieder die 3 Distanzen $d_{1,2}$, $d_{1,5}$ und $d_{2,5}$ zu prüfen. $d_{1,2}$ kann direkt an der Distanzmatrix abgelesen werden. Für $d_{1,5}$ ist im Gegensatz zur Single Linkage Analysis der Maximalabstand massgebend, nämlich $d_{1,4} = 6$. Entsprechend gilt neu $d_{2,5} = d_{2,4} = 4$. Zusammenfassend gilt:

$$\begin{aligned}d_{1,2} &= 2 \\d_{1,5} &= 6 \\d_{2,5} &= 4\end{aligned}$$

A1 und A2 werden zur Gruppe 6 zusammengeschlossen, und zwar auf dem Niveau $d = 2$. Für die Bildung der Gruppe 7 müssen wiederum die Maximalabstände gesucht werden. Wir finden

$$\begin{aligned}d_{5,6} &= \max (d_{1,3}, d_{1,4}, d_{2,3}, d_{2,4}) \\&= \max (5, 6, 3, 4) = d_{1,4} = 6\end{aligned}$$

Dieser Wert wird zur Vollendung des Dendrogrammes in Abb. 5.5, C, verwendet.

5.4.3 Average Linkage Analysis

Im Vergleich zu den eben beschriebenen Methoden handelt es sich dabei um eine mittlere, gemässigte Lösung. Statt der Maximal- oder Minimaldistanz zwischen Individuen verschiedener Gruppen, wird als Kriterium für den Zusammenschluss die mittlere Distanz gewählt. So beträgt dann der massgebende Abstand zwischen den Gruppen 2 (A2) und 5 (A3, A4) $d_{2,5} = 3.5$, wie aus Abb. 5.5, A sofort ersichtlich wird. Alle übrigen Operationen sind mit denjenigen der schon beschriebenen Verfahren identisch. Für den Leser, der die Analyse nachvollziehen will, wird in Abb. 5.5, D das resultierende Dendrogramm gegeben.

5.4.4 Minimalvarianz-Analyse

Im Gegensatz zu den bisher besprochenen Methoden beruht die Minimalvarianz - Analyse (ORLOCI 1967) auf den Streuungsverhältnissen der Gruppen. Der Zusammenschluss bestehender Gruppen zu grösseren, neuen, erfolgt stets so, dass die gruppeninterne Varianz möglichst wenig zunimmt. Das Verfahren beruht mithin auf den Konzepten der Varianzanalyse.

Wir beginnen die Betrachtung mit der Definition der gruppeninternen Varianz Q_g (vergleiche dazu auch die Ausführungen in Kapitel 6.3). Diese ist gleich der Summe der quadrierten Abstände jedes Gruppenindividuums zum Gruppenzentrum:

$$Q_g = \sum_{i=1}^p \left(\sum_{j=1}^{n_g} (x_{ij} - \bar{x}_i)^2 \right)$$

Darin ist x_{ij} die Koordinate (Artmächtigkeit) der Art i in der Aufnahme j , n_g die Anzahl Aufnahmen der Gruppe g , \bar{x}_i der Mittelwert aller Arten in g und p die Anzahl Arten. Q_g lässt sich rascher berechnen aus der Matrix der quadrierten Distanzen:

$$Q_g = \frac{1}{n_g} \sum_{i < j} d_{ij}^2$$

Den formalen Nachweis für diesen Zusammenhang zeigt z.B. PIELOU (1977), S. 319 f. Σ für $i < j$ bedeutet, dass alle $n_g * (n_g - 1) / 2$ Elemente der D^2 Matrix, welche sich auf die Individuen der Gruppe g beziehen, summiert werden. Verwenden wir wieder das Beispiel aus Abb. 5.5, so müssen die Distanzen zuerst quadriert werden:

$$D^2 = \begin{matrix} & 0 & 4 & 25 & 36 \\ 0 & & & & \\ 4 & & & & \\ 25 & & & & \\ 36 & & & & \end{matrix} \begin{matrix} 0 & 4 & 25 & 36 \\ 4 & 0 & 9 & 16 \\ 25 & 9 & 0 & 1 \\ 36 & 16 & 1 & 0 \end{matrix}$$

Die linke untere Hälfte der Matrix ist zu vernachlässigen, sodass jedes Element nur einfach gezählt wird. Als Beispiel erhalten wir für

$$Q_{2,3,4} = 1/3 (9+16+1) = 26/3 = 8 \frac{2}{3}.$$

$Q_{2,3,4}$ ist die interne Varianz der Gruppe (2,3,4). ORLOCIs (1967) Kriterium zur Fusion zweier Gruppen A und B lautet nun, dass die Zunahme der Varianz $Q(A,B)$ minimal sein soll, wobei gilt

$$Q(A,B) = Q(A+B) - Q(A) - Q(B).$$

$Q(A+B)$ ist die Varianz der neu zu bildenden Gruppe, $Q(A)$ und $Q(B)$ sind diejenigen der alten Gruppen. Damit ist $Q(A,B)$ jener Betrag, um welchen die Varianz beim Zusammenschluss von A und B vermehrt wird.

Für den ersten Zusammenschluss brauchen bloss die Elemente der D^2 -Matrix nach dem kleinsten Wert abgesucht zu werden. Als Minimum qualifiziert sich

$$Q(3,4) = 1/2 (1) = 0,5 = Q(5).$$

Für den nächsten Zusammenschluss müssen zuerst alle $Q(A,B)$ -Werte berechnet werden. Man erhält:

$$Q(1,2) = 1/2 (4) = 2$$

$$\begin{aligned} Q(1,5) &= \frac{1}{n_1+n_5} (d_{1,3}^2 + d_{1,4}^2 + d_{3,4}^2) - Q(1) - Q(5) \\ &= 1/3 (25+36+1) - 0 - 1/2 = 20 \frac{2}{3} - 1/2 = 20 \frac{1}{6} \end{aligned}$$

$$Q(2,5) = 1/3 (9+16+1) - 0 - 1/2 = 8 \frac{2}{3} - 1/2 = 8 \frac{1}{6}$$

In der Matrix-Schreibweise gilt

$$Q = \begin{matrix} 0 & 2 & 20 \ 1/6 \\ 2 & 0 & 8 \ 1/6 \\ 20 \ 1/6 & 8 \ 1/6 & 0 \end{matrix}$$

Als neue Gruppe 6 qualifizieren sich die Individuen (1,2) mit $Q = 2$. Schliesslich ist für Gruppe 7 $Q(5,6)$ zu berechnen:

$$Q(5,6) = 1/4 (4+25+36+9+16+1) - 1/2 - 2 = 20 \ 1/4$$

Werden die eben gefundenen Q-Werte auf der y-Achse aufgetragen, so erhält man das Dendrogramm in Abb. 5.5, E.

5.4.5 Besonderheiten agglomerativer Verfahren

Beim Vergleich der hier gezeigten Methoden anhand des kleinen 4-Punkte Beispiels (Abb. 5.5) ist bemerkenswert, dass das Resultat stets gleich ausfällt. In der Tat kann generell gesagt werden, dass die zu erwartenden Unterschiede bei den meisten Datenstrukturen klein sind. In der Anfangsphase funktionieren alle vier besprochenen Methoden gleich: Zuerst werden die nächsten Nachbarn zu Zweiergruppen zusammengefasst. Dies ist auch der konzeptionell schwächste Teil agglomerativer Verfahren. Die Lage einzelner Punkte, welche stets gewissen Zufälligkeiten unterworfen ist, entscheidet wesentlich über das Resultat. Erst bei steigender Individuenzahl pro Gruppe treten die Unterschiedenheiten deutlicher hervor. Die wesentlichsten Unterschiede betreffen:

a) Die Tendenz zur Kettenbildung

Es handelt sich um eine typische Eigenschaft der Single Linkage Analysis. Gradienten bildende, beliebig lange Reihen von Aufnahmen können als eigenständige Gruppen erkannt werden. Gerade gegenteilig verhält sich die Complete Linkage Analysis. Bei ihr wachsen die Gruppen fast gleichförmig um die zu Beginn gefundenen Zentren. Sie unterteilt langgestreckte Reihen in mehrere, gedrungene Gruppen und eignet sich damit besser für Daten, die aus natürlichen Gruppen

zusammengesetzt sind. Abb. 5.7 illustriert solche Verhältnisse. Sie zeigt den (künstlich konstruierten Fall) einer Stichprobe von drei Gruppen, wobei eine davon eine lange Kette bildet. Die einzige Methode, welche diese Struktur erfolgreich aufdeckt, ist die Single Linkage Analysis. Complete Linkage, Average Linkage und Minimalvarianz-Analyse sind dazu nicht in der Lage.

Die Vermutung liegt nahe, die Single-Linkage Analysis könnte sich zur Strukturierung von Daten einer Gradientenanalyse eignen. Es ist aber zu beachten, dass in der Praxis selten einfache Punkteketten auftreten. Sobald kompliziertere Konfigurationen vorliegen, lassen sich sphärische oder elliptische Gruppen leichter interpretieren.

b) Reaktion auf Einzelindividuen

Wie oben dargelegt, bestimmt in der Anfangsphase stets die Lage einzelner Individuen den Gang der Analyse. Erst bei fortschreitendem Gruppierungsprozess unterscheiden sich die Methoden. Single und Complete Linkage Analysis verwenden als Kriterium für Zusammenschlüsse bloss die Distanz zu einem einzigen Individuum einer Gruppe. Beide Methoden sind deshalb wenig robust. Average Linkage Analysis und das Minimalvarianzverfahren berücksichtigen dagegen die Lage aller Individuen einer Gruppe. Sie tragen mithin der gesamten Gruppenstruktur Rechnung.

c) Berücksichtigung der Gruppengrösse

Die einzige der hier erwähnten Methoden, welche auch die Anzahl der Punkte einer Gruppe als Zusammenschlusskriterium verwendet, ist die Minimalvarianzanalyse. Abb. 5.8 zeigt zwei Fälle, die von der Average Linkage Analyse gleich interpretiert werden. Wir nehmen an, dass die beiden Gruppen A und B zusammengeschlossen werden sollen. Die alten Gruppenzentren liegen im rechten und linken Beispiel gleich weit auseinander. In der Minimalvarianzanalyse ist jedoch nicht dieser Abstand massgebend, sondern die Varianz der neuen Gruppe (A,B). Da die Lage des Zentrums von der Anzahl Individuen in A und B abhängig ist, liegt es rechts näher bei der grösseren Gruppe, A. In der Minimalvarianzanalyse wird

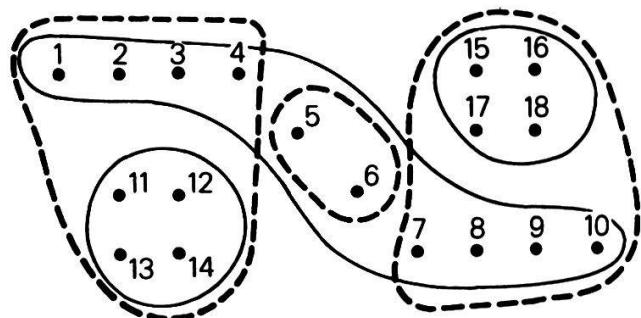


Abb. 5.7 Unterschiede zwischen verschiedenen agglomerativen Clusterverfahren am Beispiel der Ähnlichkeitsstruktur von Abb. 5.2, C: Gruppenbildung von Single Linkage Analysis (geschlossene Linie), Gruppenbildung von Complete Linkage Analysis (gestrichelt).

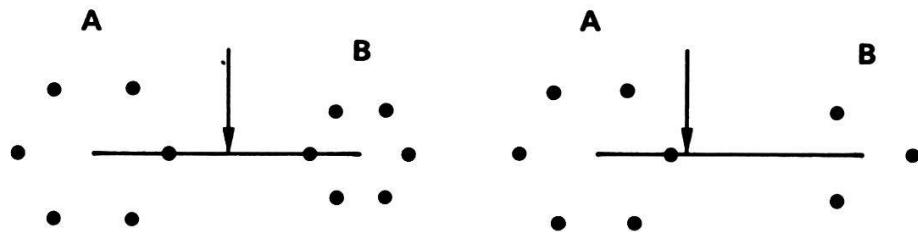


Abb. 5.8 Durch Average Linkage- und Minimalvarianzanalyse verschieden behandelte Fälle eines Gruppenzusammenschlusses. Für die Average Linkage Analyse sind die Zusammenschlüsse links und rechts gleich günstig. Aus der Sicht der Minimalvarianzanalyse erhält derjenige rechts den Vorzug.

der Zusammenschluss rechts wegen der kleineren Varianz die günstigere Lösung sein. Das ist für die meisten Anwendungen sinnvoll. Je mehr Punkte an einer Gruppe beteiligt sind, desto eher kann davon ausgegangen werden, dass es sich nicht um einen Artefakt handelt und dass sie deshalb eigenständig ist. Erwartet man dagegen in einem Datensatz Ausreisser und möchte diese sicher von echten Gruppen abgetrennt haben, so ist die Complete Linkage Analyse angemessen.

6. Gewichtungsverfahren (Rangierung)

Der Gewichtung von Merkmalen kommt in der Vegetationskunde besondere Bedeutung zu. Vegetationsaufnahmen besitzen die Besonderheit, dass sie durch sehr viele Merkmale (Arten) charakterisiert sind. Ist die Zahl der Merkmale gar grösser als diejenige der Aufnahmen, so kann ein Datensatz überbestimmt sein. In einem solchen Falle lässt sich jede beliebige Gruppen- oder Gradientenstruktur anhand ausgewählter Merkmale begründen. Auf die traditionelle Tabellenarbeit übertragen (ELLENBERG 1956) bedeutet dies, dass sich dieselbe Vegetationstabelle in verschiedenster Weise durchaus sinnvoll ordnen lässt. Es ist daher schon aus formalen Ueberlegungen heraus erstrebenswert, Aufnahmen mithilfe möglichst weniger Arten oder Standortsfaktoren charakterisieren zu können. Dies erfordert aber meistens eine Gewichtung der Merkmale.

Eine solche kann vorgenommen werden, um Zeigerarten zu finden. Letztere dienen dann einerseits der Klassifikation der Aufnahmen. Andererseits erlauben sie die Zuordnung neuer Aufnahmen zu beschriebenen Pflanzengesellschaften (Identifikation). Schliesslich dienen sie dem Aufbau von Vegetationsschlüsseln zur pflanzensoziologischen Kartierung im Felde.

ORLOCI (1978) schlägt eine in der Praxis bislang kaum realisierte Anwendung der Gewichtungsverfahren vor. Seine Ueberlegungen zielen auf eine Minimalisierung der Feldarbeit bei vegetationskundlichen Kartierungen ab. Traditionellerweise werden in Vegetationsaufnahmen aus verschiedenen Gründen alle Arten von Gefässpflanzen berücksichtigt (POORE 1962, MUELLER-DOMBOIS und ELLENBERG 1974), was aus der Sicht der Grundlagenforschung auch sinnvoll ist. Namentlich bei den heute aktuellen Anwendungen der Vegetationskunde im Bereich der Umweltforschung ist dieses Vorgehen aber nicht zwingend. Der Aufwand für die Feldarbeiten fällt geringer aus, wenn nur Zeigerarten erhoben werden müssen. Diese sind aber bei echten Neuuntersuchungen nicht hinreichend bekannt. Mit dem folgenden Vorgehen könnte das Problem im Rahmen einer Vorun-

tersuchung gelöst werden:

1. Man macht einige Vegetationsaufnahmen im zukünftigen Untersuchungsgebiet (z.B. 20).
2. Durch ein Rangierungs- oder Gewichtungsverfahren werden die besten Zeigerarten innerhalb des Datensatzes bestimmt.
3. Es werden weitere Aufnahmen zugefügt (z.B. 10). Eine erneute Rangierung ergibt wahrscheinlich einige Änderungen in der Gewichtung der Arten.
4. Die Erweiterung des Datensatzes ist so lange durchzuführen, bis sich in der Gewichtung kaum noch Verschiebungen ergeben. Die Arten hohen Gewichtes (Zeigerarten) beschreiben die Ähnlichkeitsstruktur der gesamten zu untersuchenden Vegetation hinreichend.

Wird nach Abschluss dieser Vorabklärung eine Kartierung in Angriff genommen, so brauchen keine vollständigen Aufnahmen mehr durchgeführt zu werden. Es kann unmittelbar mit der flächendeckenden Erhebung auf Grund der gefundenen Zeigerarten begonnen werden. Selbstverständlich kann dieses Vorgehen nur dann zum Erfolg führen, wenn eine zweckmäßige Auswahl der Aufnahmeflächen getroffen wurde, sodass die Erhebung für das Untersuchungsobjekt repräsentativ ist.

Man kann grundsätzlich zwei Typen von Gewichtungsverfahren unterscheiden. Im einen Fall wird die Gesamtstichprobe untersucht, ohne dass ihre Gruppenstruktur bereits bekannt wäre. Dieser Weg führt in der Regel über die Analyse einer Ähnlichkeitsmatrix. Anders liegen die Verhältnisse, wenn von bekannten Pflanzengesellschaften ausgegangen wird. Hier stellt sich die Frage, welche Arten die besten Zeiger für eine vorgegebene Gruppenstruktur sind. Wir werden im folgenden Vertreter beider Typen diskutieren.

6.1 Feolis Methode

Diese einfache und leicht verständliche Methode soll einführend dargestellt werden (vgl. FEOLI 1973). Grundsätzlich basiert sie auf Präsenz und Absenz der Arten. Sie kann auf folgende zwei Annahmen zurückgeführt werden:

1. Eine Art ist dann ein guter Zeiger, wenn die Aufnahmen, in denen sie auftritt, möglichst ähnlich sind.
2. Aufnahmen, in denen sie vorkommt, sollen im Vergleich zu solchen, in denen sie fehlt, möglichst verschieden sein.

Das Prinzip der Methode zeigt Tab. 6.1, A bis C. In Tab. 6.1, A, sind die Rohdaten des nachfolgenden Beispiels zu finden. Zunächst muss eine Ähnlichkeitsmatrix S der Aufnahmen verfügbar gemacht werden. In Tab. 6.1, B und C ist eine solche dargestellt. Verwendet wird van der Maarels Koeffizient, doch ist die Methode auch mit jedem anderen Mass zu realisieren. Feoli bezeichnet nun die Ähnlichkeitskoeffizienten zwischen Aufnahmen, in denen die zu gewichtende Art vorkommt, mit SI. SZ werden jene genannt, welche Aufnahmen mit und solche ohne die betreffende Art vergleichen. Für die Berechnung des Gewichts der Art, SF, sind die Mittelwerte massgebend:

$$SF = 1 - \frac{\bar{S}Z}{\bar{S}I}$$

Tab. 6.1, B illustriert das Vorgehen für die Art 1. Diese kommt nur in den Aufnahmen 1 und 2 vor, die eine Ähnlichkeit von 0.8 aufweisen. Für SZ berücksichtigt man alle Koeffizienten, welche die Aufnahmen 1 und 2 mit 3, 4 und 5 vergleichen. Diese sind punktiert eingeraumt. Wir erhalten

$$SF_1 = 1 - \frac{(0.143+0.25)/6}{0.8} = 1 - \frac{0.066}{0.8} = 0.9175$$

Da Art 2 genau das gleiche Vorkommen hat, gilt dieses

Tabelle 6.1

Feolis Artrangierung. Vegetationstabelle (A), Ähnlichkeitsmatrix der Aufnahmen (B,C). Fall B illustriert die Berechnung des Gewichtes für Art 1 oder 2, Fall C für die Art 3.

A

Art	Aufn. 1	2	3	4	5
1	2	1			
2	1	1			
3	1	1	1		
4		1	1	1	
5			1	2	

B

Aufnahme	1	2	3	4	5
1	1	0.8	0.143	0	0
2		1	0.25	0	0
3			1	0.33	0.17
4				1	0.75
5					1

C

Aufnahme	1	2	3	4	5
1	1	0.8	0.143	0	0
2		1	0.25	0	0
3			1	0.33	0.17
4				1	0.75
5					1

Gewicht auch für sie. Art 3 hingegen kommt in den Aufnahmen 1, 2 und 3 vor (Tab. 6.1, C). Wir erhalten

$$SF_3 = 1 - \frac{(0.33+0.17)/6}{(0.8+0.143+0.25)/3} = 1 - \frac{0.083}{0.397} = 0.79$$

Ihre Zeigereigenschaft ist somit deutlich schlechter als jene der Arten 1 und 2.

Analog werden nun die Gewichte der Arten 4 und 5 berechnet. Schliesslich erhält man folgende Rangordnung:

Art	SF-Wert	Rang
1	0.9175	1
2	0.9175	1
5	0.889	2
4	0.843	3
3	0.79	4

Anhand dieser Ergebnisse lässt sich leicht erkennen, welche Arten beispielsweise für die Zusammenstellung eines Vegetationsschlüssels gut geeignet sind (z.B. 1, 2 und 5). Zur Datenreduktion eignet sich jedoch das Verfahren weniger. Aus den Arten 1 und 2 wird deutlich, weshalb dem so ist. Beide besitzen hohe SF-Werte. Wird auf eine von beiden verzichtet, so bleibt die Differenzierung zwischen Aufnahmen (1,2) und (3,4,5) dank der verbleibenden Art erhalten. Trotz hohen SF-Wertes würde somit der Verzicht auf Art 2 die Gruppenstruktur der Tabelle nicht zerstören. Will man den Informationsverlust durch Weglassen von Arten untersuchen, so bedarf es komplizierterer Verfahren.

6.2 Rangierung nach erklärter Varianz

Dieses von ORLOCI (1973) unter dem Namen RANK vorgeschlagene Verfahren ist dann angezeigt, wenn ein äusserstes Minimum an Zeigerarten gesucht wird. Die Analyse sehr grosser Tabellen

gestaltet sich einfacher, falls es gelingt, die Anzahl Arten auf einen Bruchteil zu reduzieren. Natürlich darf dabei die ursprüngliche Ähnlichkeitsstruktur der Aufnahmen nicht allzu sehr verändert werden. Eine erfolgreiche Anwendung dieses Verfahrens zeigen FEOLI und LAGONEGRO (1982). Sie revidierten die Klassifikation italienischer Buchenwälder anhand von 602 Aufnahmen. Die ursprünglich 475 Arten wurden mit Hilfe des RANK-Verfahrens vorgängig auf 108 reduziert, was die Analyse entscheidend vereinfachte.

Im Gegensatz zu Feolis Verfahren geht Orlocis RANK schrittweise vor. Das Prinzip ist folgendes (Tab. 6.2): Zu Beginn jedes Schrittes wird eine provisorische Gewichtung der Arten gesucht. Als Gewichtungskriterium sind verschiedene Masse denkbar. Würde z.B. die Summe der Abundanzwerte verwendet, so erhielte Art 1 Rang 1 (dass das RANK-Verfahren ein anderes Kriterium verwendet, ist aus Tab. 6.2 zu schliessen). Ausgehend von diesem Zwischenresultat ist nun der Datensatz zu reduzieren. Dazu muss die gesamte Information, welche die Art mit Rang 1 trägt, aus der Tabelle entfernt werden. Die Information der Art 2 müsste dabei fast verschwinden, da sie mit derjenigen von Art 1 annähernd identisch ist. Der nächste Schritt besteht in einer erneuten provisorischen Gewichtung der verbleibenden Arten. Art 4 dürfte dabei den höchsten Wert erhalten und mit dem definitiven Rang 2 ausgeschieden werden. Doch wenden wir uns nun dem eigentlichen RANK-Algorithmus zu.

Das Verfahren beginnt mit der Berechnung einer Ähnlichkeitsmatrix der Arten. Dabei kann ein beliebiges Mass ausgewählt werden. Verwendet man den Korrelationskoeffizienten, so erhält man für die $p = 4$ Arten in Tab. 6.2

$$R = \begin{matrix} & 1.0 & 0.85 & -0.91 & -0.64 \\ 1.0 & & & & \\ 0.85 & & 1.0 & -0.71 & -0.43 \\ -0.91 & & -0.71 & 1.0 & 0.91 \\ -0.64 & & -0.43 & 0.91 & 1.0 \end{matrix}$$

Wie erwartet, heben sich als mögliche Gruppen (1,2) und (3,4) mit hohen Korrelationen von 0.85 und 0.91 ab. Nun wird

Tabelle 6.2

Einfache Vegetationstabelle zur Rangierung der Arten. Die Ränge ergeben sich aufgrund Orlocis RANK.

Aufnahme	1	2	3	4	Rang
Art 1	2	2	1		4.
Art 2	2	1	1		2.
Art 3			1	1	1.
Art 4			2	1	3.

eine provisorische Rangordnung gesucht. Massgebend ist die Summe der Korrelationen jeder Art mit allen andern, also

$$SS_i = \sum_{h=1}^p r_{hi}^2 / r_{ii}$$

Darin ist SS_i die Summe der total p quadrierten Ähnlichkeitswerte, welche die Art i betreffen. r_{ii} ist der Diagonalwert, der im Falle der Korrelationsmatrix 1.0 beträgt und damit vernachlässigt werden kann. Für Art 1 finden sich alle Ähnlichkeitswerte r_{hi} in der ersten Zeile (oder in der ersten Kolonne) von R . Somit ergibt sich für sie

$$SS_1 = ((1.0)^2 + (0.85)^2 + (-0.91)^2 + (-0.64)^2) / 1 = 2.95$$

Entsprechend erhält man für die übrigen Arten

$$\begin{aligned} SS_2 &= 2.41 \\ SS_3 &= 3.14 \\ SS_4 &= 2.41 \end{aligned}$$

Die Art mit der höchsten Summe, Art 3, erhält den definitiven Rang 1. Bevor die Bedeutung dieses Ergebnisses genauer untersucht wird, soll aber der Algorithmus weiter verfolgt werden. Zu diesem Zwecke ist der Einfluss der Art 3 auf die Korrelationen zu eliminieren. Die Elemente der neuen, reduzierten Matrix R berechnen sich nach der Regel

$$r'_{hi} = r_{hi} - y_{hm} y_{im} .$$

Symbol m steht für die zu entfernende Art, hier also Art 3. Die Formel gilt für die Korrelation jeder Art h und i . Bei den y -Werten handelt es sich um Elemente der Korrelationsmatrix, denn sie sind folgendermassen definiert:

$$y_{hm} = r_{hm} / (r_{mm})^{1/2} , \text{ sowie } y_{im} = r_{im} / (r_{mm})^{1/2} .$$

r_{mm} ist in unserem Falle die Korrelation von Art 3 mit sich

selbst, also 1.0. Mit dem ersten Element beginnend, ergibt die Reduktion:

$$\begin{aligned}r'_{11} &= 1 - (-0.91)^2 = 0.17 \quad (h=1, i=1) \\r'_{12} &= 0.85 - (-0.71 \cdot -0.91) = 0.20 \quad (h=1, i=2) \\r'_{13} &= -0.91 - (1.0 \cdot -0.91) = 0 \quad (h=1, i=3) \\r'_{14} &= -0.64 - (-0.91 \cdot +0.91) = 0.19 \quad (h=1, i=4)\end{aligned}$$

Diese Operation ist für alle r_{hi} -Werte vorzunehmen. Damit verschwindet jener Anteil der Korrelation aus der Matrix, der durch Art 3 verursacht wird. Zeile und Spalte 3 von R müssen nun zwangsläufig mit Nullen besetzt sein:

$$\begin{matrix} & 0.17 & 0.20 & 0 & 0.19 \\R = & 0.20 & 0.50 & 0 & 0.21 \\ & 0 & 0 & 0 & 0 \\ & 0.19 & 0.21 & 0 & 0.17 \end{matrix}$$

Die erneute provisorische Rangordnung ergibt:

$$\begin{aligned}SS_1 &= (0.17^2 + 0.20^2 + 0.19^2) / 0.17 = 0.61 \\SS_2 &= (0.20^2 + 0.50^2 + 0.21^2) / 0.50 = 0.67 \\SS_4 &= (0.19 + 0.21 + 0.17) / 0.17 = 0.64\end{aligned}$$

Art 2 erreicht nun den maximalen Wert und erhält den endgültigen Rang 2. Die Ähnlichkeitsmatrix kann damit weiter reduziert werden, wobei Zeile und Kolonne 2 und 3 mit Nullen besetzt sind:

$$\begin{matrix} & 0.14 & 0 & 0 & 0.16 \\R' = & 0 & 0 & 0 & 0 \\ & 0 & 0 & 0 & 0 \\ & 0.16 & 0 & 0 & 0.14 \end{matrix}$$

Den Rangierungsprozess mit dieser Präzision weiterzuführen, lohnt sich nicht, da nunmehr immer grössere Rundungsfehler auftreten! Mit den Arten 3 und 2 sind aber offensichtlich zwei typische Vertreter der Tabelle 6.2 gefunden worden. Die Korrelationssummen SS bilden ausserdem ein Mass für deren Güte als Zeigerwert. Dieser ist jedoch leichter zu interpre-

tieren, wenn er am maximal möglichen Wert gemessen wird, wie er bei einer absolut idealen Zeigerart auftreten würde. Solche Maximalkorrelationen finden sich in der Diagonalen der Matrix, wo jede Art mit sich selbst verglichen wird. Der Maximalwert beträgt

$$SS_{\max} = \sum_{i=1}^p r_{ii}^2$$

Symbol p steht für die Anzahl Arten. In unserem Beispiel ist $SS_{\max} = 4$. Damit ergeben sich folgende Anteile erklärter Korrelation:

Art 3: $\frac{3.14}{4.0} * 100 = 78.5\%$

Art 2: $\frac{0.67}{4.0} * 100 = 17\%$

Die Artkombination (Art 3, Art 2) erklärt damit $78.5\% + 17\% = 95.5\%$ der gesamten Korrelation. Dies bestätigt die oft gemachte Erfahrung, dass schon sehr wenige Arten die Struktur einer Vegetationstabelle weitgehend darzustellen vermögen (vgl. WILDI 1979).

Rechnet man RANK mit grösserer Genauigkeit für alle Arten, so erhält man

Rang	Art	Erklärte Varianz in %
1	3	78.5
2	2	17.0
3	4	4.5
4	1	0

Die Information, welche die Art 1 trägt, ist also völlig redundant, sodass eine Reduktion auf drei Arten verlustfrei erfolgen kann!

Die Interpretation der Ergebnisse ist mit einigen Schwierigkeiten verbunden. Zunächst ist zu beachten, dass Arten aufeinanderfolgenden Ranges immer sehr unterschiedliches Vorkommen aufweisen, wie hier 3 und 2, 2 und 4 usw. Sodann kann es ohne weiteres geschehen, dass an sich gute Zeigerarten (hier z.B. Art 1) wegfallen, weil ihre Varianz (oder Korrelation) bereits durch andere Arten vollständig erklärt wird. Das hat oft den Nachteil, dass im Feld gut sichtbare und leicht bestimmmbare Arten aus dem Datensatz entfernt werden. Um dem vorzubeugen, wurde ein interaktives RANK Verfahren entwickelt, bei welchem der Experimentator aus einer kleinen Auswahl günstig gelegener Arten diejenige auswählt, welche definitiv zu rangieren ist (WILDI 1984). Damit steht dem Pflanzensoziologen ein Instrument zur Verfügung, das sehr rasch einen guten Einblick in die Ähnlichkeitsverhältnisse komplexer Datensätze vermittelt. Aus früher erwähnten Gründen sollte die Analyse nicht auf dem Korrelationskoeffizienten, sondern auf dem Skalarprodukt oder der Kovarianz basieren. Sehr gut interpretierbare Ergebnisse erzielt man mit den Kontingenzmassen von Jaccard, Soerensen bzw. van der Maarel. Allerdings muss dabei der Rechenvorgang wegen fehlender Metrik frühzeitig abgebrochen werden.

6.3 Rangierung nach Gruppenstruktur

Bei allen bisher behandelten Rangierungsverfahren wurde versucht, die Arten oder Standortsfaktoren vorgängig einer weiterführenden Analyse zu gewichten. Dabei wurde stets die gesamte Information innerhalb der Ähnlichkeitsmatrix verwendet. Hat man sich jedoch einmal zu einer Klassifikation entschlossen, so hat man auch auf einen Teil der Gesamtinformation des ursprünglichen Datensatzes verzichtet. Statt einzelner Aufnahmen stehen jetzt nur noch Gruppen im Zentrum des Interesses. Es lohnt sich, die einzelnen Arten oder Standortsfaktoren erneut zu gewichten, und zwar nach der Zuverlässigkeit, mit welcher sie auf eine Gruppenzugehörigkeit hinweisen. JANCEY (1979) schlägt dazu ein Verfahren vor, welches auf der Varianzanalyse basiert.

Die Varianzanalyse beruht auf dem Prinzip, dass sich die Varianz einer gruppierten Stichprobe wie folgt aufteilen lässt:

$$\begin{array}{l} \text{Totalvarianz} \\ \text{der Stichprobe} \end{array} = \begin{array}{l} \text{Varianz} \\ \text{innerhalb} \\ \text{der Gruppen} \end{array} + \begin{array}{l} \text{Varianz} \\ \text{zwischen} \\ \text{den Gruppen} \end{array}$$

Bekanntlich berechnen sich die Varianzen als Summen von Abweichungsquadraten (Abschnitt 4.3). Da obige Beziehungen auch für die nicht quadrierten Abweichungen gelten, lässt sich der Sachverhalt leicht grafisch veranschaulichen (Abb. 6.1).

Ausgangspunkt bilde eine Stichprobe, bestehend aus drei Gruppen zu je drei Individuen. Die Indices in den nachfolgenden Formeln bedeuten:

$$\begin{array}{ll} i \text{ für die laufende Aufnahme, } & i = 1, \dots, n \\ j \text{ für die laufende Art, } & j = 1, \dots, p \\ g \text{ für die laufende Gruppe, } & g = 1, \dots, m \\ t_g \text{ für die laufende Aufnahme} \\ \text{in der Gruppe } g, & t_g = 1, \dots, m_g \end{array}$$

Man berechnet zuerst die Mittelwerte \bar{x}_j für jede Art j . Für die Summe der Abweichungsquadrat ST_j^2 ergibt sich dann

$$ST_j^2 = \sum_{i=1}^n (x_{ji} - \bar{x}_j)^2$$

für jede Art j . Für die Varianz innerhalb der Gruppen wird analog vorgegangen, wobei die Abweichungsquadrat aller m Gruppen noch zu summieren sind:

$$SI_j^2 = \sum_{g=1}^m \sum_{t_g=1}^{m_g} (x_{j t_g} - \bar{x}_g)^2$$

Darin ist g die Gruppennummer, m_g die Anzahl Individuen in

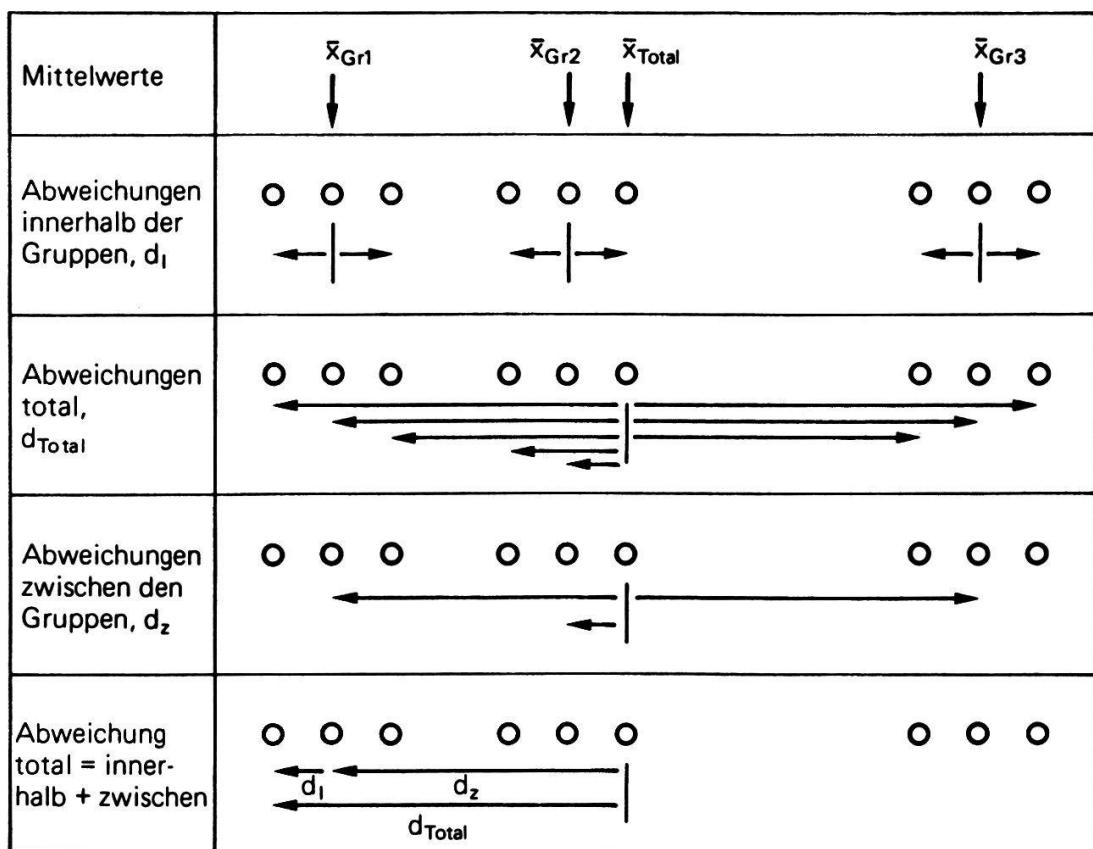


Abb. 6.1 Beziehungen zwischen der Gesamtabweichung d_{Total} , der Abweichung innerhalb einer Gruppe d_I und der Abweichung zwischen den Gruppen d_z . Der Gesamtmittelwert ist \bar{x}_{Total} , der Mittelwert der Gruppe g ist $\bar{x}_{Gr\ g}$.

Gruppe g und x_{jt} der Messwert (Abundanz) der Art j in der Aufnahme t der Gruppe g . Die Summe der Abweichungsquadrate zwischen den Gruppen beträgt

$$s_{Zj}^2 = \sum_{g=1}^m m_g (\bar{x}_g - \bar{x}_j)^2.$$

Man ersieht daraus, dass die Abweichungen zwischen dem Gesamtmittelwert und den Gruppenmittelwerten jeweils mit der Gruppengrösse m_g zu multiplizieren sind.

Am folgenden Beispiel soll nun demonstriert werden, dass sich die verschiedenen Summen der Abweichungsquadrate als Messgrösse für den Zeigerwert von Arten eignen. Daten und Resultate sind zusammengefasst in Tab. 6.3. Art 1 ist so gleichmässig auf alle Gruppen verteilt, dass sie als Zeigerart sicher nicht in Frage kommt.

Genau gegenteilig verhält sich Art 3, bei welcher die gruppeninternen Abweichungen klein, diejenigen zwischen den Gruppen dafür gross sind. Art 3 ist somit ein optimaler Zeiger - zumindest für Gruppe 1 gegenüber 2 und 3. Art 2 schliesslich steht zwischen den beiden Extremen.

Man könnte nun der Idee verfallen, die Summe der Abweichungsquadrate zwischen den Gruppen direkt als Mass für die Güte des Zeigerwertes zu verwenden. In Realität treten aber nicht so einfache Fälle auf wie in Tab. 6.3. Insbesondere unterscheidet sich die Gesamtvarianz ST_j^2 von Art zu Art. Stichproben- und Gruppengrösse beeinflussen ausserdem die Zuverlässigkeit der Ergebnisse. JANCEY (1979) standardisiert deshalb zuerst die Artvektoren j ,

$$x'_{ji} = (x_{ji} - \bar{x}_j) / s_j.$$

x'_{ji} ist der transformierte Artwert (Abundanz) der Art j in der Aufnahme i . s_j ist die Standardabweichung der Art j . Um die verminderte Zuverlässigkeit des Ergebnisses bei kleinen Gruppen auszugleichen, wird darin durch $n-1$ und nicht wie in

Tabelle 6.3.

Rangierung von Arten auf Grund einer Klassifikation. A: Rohwerte. B: Standardisierte Artvektoren. C: Gruppeninterne Statistik. D: Summen der Abweichungsquadrate, Varianzen und F-Werte sowie die resultierende Rangfolge.

<u>A</u>	Aufn.	1	2	3	4	5	6	7	8	9	\bar{x}_j	$\Sigma(x_{ji} - \bar{x}_j)^2$	s_j
	Art 1	3	-	-	2	-	-	1	-	-	0.666	10	1.118
	Art 2	3	2	-	1	-	-	-	-	-	0.666	10	1.118
	Art 3	3	2	1	-	-	-	-	-	-	0.666	10	1.118

<u>B</u>	Aufn.	1	2	3	4	5	6	7	8	9	ST_j^2
	Art 1	2.09	-0.6	-0.6	1.19	-0.6	-0.6	0.3	-0.6	-0.6	8.0
	Art 2	2.09	1.19	-0.6	0.3	-0.6	-0.6	-0.6	-0.6	-0.6	8.0
	Art 3	2.09	1.19	0.3	-0.6	-0.6	-0.6	-0.6	-0.6	-0.6	8.0

<u>C</u>	\bar{x}_1	ST_j^2	\bar{x}_2	ST_j^2	\bar{x}_3	ST_j^2	
	Art 1	0.30	4.82	0	2.14	-0.3	0.54
	Art 2	0.89	3.75	-0.3	0.54	-0.6	0
	Art 3	1.19	1.6	-0.6	0	-0.6	0

<u>D</u>	SI_j^2	SZ_j^2	ST_j^2	VI	VZ	F	Rang
	Art 1	7.50	0.50	8.0	1.25	0.25	0.20
	Art 2	4.29	3.71	8.0	0.715	1.86	2.60
	Art 3	1.60	6.40	8.0	0.266	3.20	12.03

Kap. 3 durch n dividiert:

$$s_j = \left[\frac{1}{n-1} \sum_i (x_{ji} - \bar{x}_j)^2 \right]^{1/2}$$

In Tab. 6.3, A lässt sich leicht nachrechnen, dass die Mittelwerte \bar{x}_j aller Arten gleich 0.666 sind. Weiter beträgt die Summe der Abweichungsquadrate von diesem Mittelwert je genau 10.0. Somit ist auch s_j überall gleich, nämlich 1.118. In Tab. 6.3, B, ist dieselbe Tabelle in transformierter Form wiedergegeben. Das erste Element berechnet sich wie folgt:

$$x'_{ji} = (3-0.666)/1.118 = 2.09$$

Nun wird für die transformierten Werte in Abb. 6.3, B, die Summe der Abweichungsquadrate berechnet. Da der Mittelwert aller Arten nun null ist, brauchen nur die Elemente quadriert und addiert zu werden:

$$ST_j^2 = \sum_{i=1}^n x'_{ji}^2.$$

Die Ergebnisse betragen überall 8.0 (was $n-1$ entspricht!). Nun kann zur Berechnung der gruppeninternen Summen der Abweichungsquadrate geschritten werden (Tab. 6.3, C). Für die erste Art in der ersten Gruppe erhält man den Mittelwert

$$\bar{x}_1 = 1/3 (2.09-0.6-0.6) = 0.30$$

Die Summe der Abweichungsquadrate ergibt

$$\begin{aligned} ST_{11}^2 &= (2.09-0.3)^2 + (-0.6-0.3)^2 + (-0.6-0.3)^2 \\ &= 4.824 \end{aligned}$$

Wir kommen nun auf die in Tab. 6.2 dargestellten Beziehungen zurück und sind in der Lage, die Summe der Abweichungsquadrate innerhalb aller Gruppen zu berechnen. Für die erste Art findet man (Tab. 6.3, D):

$$SI_{11}^2 = 4.82 + 2.14 + 0.54 = 7.50$$

Die Summe der Abweichungsquadrate zwischen den Gruppen, SZ_j^2 wird aus den Gruppenmittelwerten berechnet, für Art 1 also 0.30, 0, -0.3. In Tab. 6.3, D ist auch die totale Summe der Abweichungsquadrate aus Tab. 6.3, B übernommen worden. Es bestätigt sich deren Zerlegbarkeit in SI_j^2 und SZ_j^2 :

$$ST_j^2 = SI_j^2 + SZ_j^2$$

Für Art 1 gilt:

$$8.0 = 7.50 + 0.50.$$

Das Verhältnis zwischen SI_j^2 und SZ_j^2 ist tatsächlich charakteristisch für die Differenzierungskraft jeder Art, wie leicht in Tab. 6.3, D nachzuprüfen ist. JANCEY (1979) verwendet nun aber statt der Summen der Abweichungsquadrate die Varianzen. Damit wird ein eventueller Unterschied in der Grösse der Gruppen ausgeglichen. Sie berechnen sich gemäss der Theorie der Varianzanalyse wie folgt (vgl. z.B. GAENSSLER und SCHUBÖ 1973):

$$Vz = \frac{1}{m-1} SZ_j^2$$

$$Vi = \frac{1}{n-m} SI_j^2$$

Darin ist n wiederum die Anzahl Aufnahmen, m die Anzahl Gruppen. Die Ergebnisse sind in Tab. 6.3, D dargestellt. Das endgültige Kriterium zur Rangierung der Arten ist nun der F-Wert der Varianzanalyse. Er ist nichts anderes als das Verhältnis der Varianzen zwischen und innerhalb der Gruppen:

$$F = \frac{VZ}{VI}$$

Grosse F-Werte deuten auf eine grosse Trennkraft der Arten hin, kleine auf ein unspezifisches Vorkommen der Arten innerhalb der Vegetationstabelle. Dementsprechend erhält Art 3 in Tab. 6.3 Rang 1. Die F-Werte sind der schliessenden Statistik gemäss F-verteilt mit $df = 2$ und 6 Freiheitsgraden. Anhand einer F-Tabelle kann jeder Wert auf Ueberzufälligkeit geprüft werden (BARTEL 1972, S. 187). Für eine Irrtumswahrscheinlichkeit von 5% finden wir

$$F(df = 2,6; p = 0.05) = 5.14$$

Damit ist der F-Wert der Art 3 signifikant von Null verschieden, diejenigen der Arten 1 und 2 nicht. Da auch hier die Voraussetzungen für die Durchführung des F-Testes nicht geklärt wurden, ist diese Aussage als reine Interpretationshilfe zu betrachten.

Mit dem dargelegten Beispiel wird die Trennkraft der Arten bezüglich aller drei Gruppen bestimmt. JANCEY (1979) weist darauf hin, dass die Rechnung auch nur für einen Teil einer Tabelle durchgeführt werden kann, also z.B. für die Gruppen 1 und 2. Eine mögliche Anwendung der Methode besteht nun darin, die ursprüngliche Vegetationstabelle auf die trennenden Arten zu reduzieren, womit man einen Vegetationsschlüssel erhält. Aus der Literatur sind leider noch keine grösseren Anwendungen bekannt.

6.4 Stressanalyse

Ein mögliches Ziel von Rangierungsanalysen besteht in der Reduktion der Artenzahl. Dies verbessert die Uebersicht über die Daten und erleichtert rechenintensive Analysen. Mit der Methode RANK haben wir bereits ein Verfahren erörtert, welches für eine reduzierte Vegetationstabelle geeignete Ar-

ten direkt bezeichnet. Wieviele Arten letztlich weggelassen werden können, bleibt eine offene Frage. Sicher ist, dass die Datenstruktur durch die Auswahl möglichst wenig Änderungen erfahren sollte. Die Stressanalyse zeigt einen Weg auf, durch Artenreduktion auftretende Verzerrungen unter Kontrolle zu halten (ORLOCI 1978).

Als Beispiel verwenden wir die Vegetationstabelle in Tab. 6.2. Die Arten sollen dabei in der Reihenfolge verwendet werden, wie dies das Ergebnis der RANK-Analyse als zweckmäßig erscheinen lässt. Zur Ermittlung der Aufnahmenstruktur wird eine Matrix der Euklidschen Distanzen zwischen den Aufnahmen gerechnet. Werden alle vier Arten berücksichtigt, so erhalten wir

$$D(1,2,3,4) = \begin{matrix} 0.00 & 1.00 & 2.65 & 3.16 \\ & 0.00 & 2.45 & 2.65 \\ & & 0.00 & 1.73 \\ & & & 0.00 \end{matrix}$$

Der untere Teil der Matrix braucht nicht geschrieben zu werden, da die selben d-Werte auftreten wie oben. Lässt man nun die letztrangierte Art (Nr. 1) weg, so werden die Distanzen generell kleiner. Wir berechnen also die neue Matrix und erhalten:

$$D(2,3,4) = \begin{matrix} 0 & 1.00 & 2.45 & 2.45 \\ & 0 & 2.24 & 1.73 \\ & & 0 & 1.41 \\ & & & 0 \end{matrix}$$

Da nur die relativen Veränderungen der Distanzen interessieren, wird zwischen allen von null verschiedenen $N(N-1)/2$ Distanzen der Korrelationskoeffizient $p(DP;Dp)$ berechnet. Im obigen Beispiel gilt

1.00	1.00
2.65	2.45
D(1,2,3,4)=DP=3.16	D(2,3,4)=Dp=2.45
2.45	2.24
2.65	1.73
1.73	1.41

$p(DP;Dp) = 0.903$

Nun wird auch noch Art 4 weggelassen. Die Aufnahmen weisen jetzt folgende Distanzen auf:

0	1.00	1.41	2.24
D(2,3) =	0	1.00	1.41
	0	1.00	
		0	

Wir berechnen wiederum die Korrelation mit der Ähnlichkeitsstruktur des vollständigen Datensatzes und erhalten

1.00	1.00
2.65	1.41
D(1,2,3,4) = 3.16	D(2,3) = 2.24
2.45	1.00
2.65	1.41
1.73	1.00

Dass die Korrelation weiter gesunken ist, entspricht durchaus den Erwartungen. Es kommt aber auch vor, dass die Struktur mit einem Minimum an Arten hervorragend reproduziert wird. In unserem Beispiel ist dies der Fall. Nur noch auf der Art 3 basierend erhält man

1.00	0.00
2.65	1.00
D(1,2,3,4) = 3.16	D(3) = 1.00
2.45	1.00
2.65	1.00
1.73	0.00

Dass Art 3 alleine die Struktur so gut zu reproduzieren ver-

mag, erstaunt weiter nicht. Sie differenziert klar zwischen den beiden Gruppen (1,2) und (3,4) in der Tabelle und qualifizierte sich auch als beste Differentialart bei RANK. Zusammenfassend lässt sich der Schluss ziehen, dass die Reduktion der Artenzahl die Datenstruktur in unerwarteter Weise verändern kann. Eine Untersuchung mit Hilfe der Stressanalyse ist daher empfehlenswert.

7. Ordination

7.1 Vorbemerkungen

Die Schule Braun-Blanquet kennt die Ordination als Analysemethode nicht. Sie bedient sich direkter Schlussfolgerungen im Felde und der sogenannten Tabellenarbeit (ELLENBERG 1956, MUELLER-DOMBOIS und ELLENBERG 1974). Letztere ist ein Klassifikationsprozess innerhalb der Arten und innerhalb der Aufnahmen. Die auf reinen Vegetationsdaten beruhende Ordination ("indirekte Ordination", WHITTAKER 1967) wurde unseres Wissens durch GOODALL (1954) eingeführt. Er bediente sich der Hauptkomponentenanalyse, einer für damalige Verhältnisse sehr aufwendigen Methode. BRAY und CURTIS (1957) publizierten wenig später ein neu entwickeltes Ordinationsverfahren, das in der Folge unter dem Namen Polare Ordination weite Verbreitung fand. Zahlreiche konzeptionelle Fehler führten dazu, dass bald Revisionen vorgeschlagen wurden, welche dessen grösste Schwächen beheben sollten (ORLOCI 1966, SWAN et al. 1969, VAN DER MAAREL 1969). Leider ist jede dieser Verbesserungen auch mit einem zusätzlichen Rechenaufwand verbunden. Bei der Konzeption des vorliegenden Kapitels wurde geprüft, ob sich die Methode Bray und Curtis als grundlegendes, leicht diskutierbares Näherungsverfahren verwenden liesse. Zahl und Schwere der dabei in Kauf zu nehmenden Fehler sind aber nicht akzeptabel, sodass davon Abstand genommen werden muss. Stattdessen wird zuerst die Hauptkomponentenanalyse erörtert. Deren Durchführung ist auch bei kleineren Datensätzen rechenintensiv. Auf Beweisführungen für die Lösungswege wird erneut verzichtet und auf die einschlägige Literatur verwiesen. Der Nachvollzug des Rechenganges und die Interpretation der Ergebnisse dürften es einem Pflanzenökologen ermöglichen, Nutzen und Grenzen der Ordination unter den in der Vegetationskunde gegebenen Umständen zu erkennen.

Eine weitere Bemerkung gilt der Vielfalt der Methoden. Sie ist, sofern man auf die publizierten Namen abstellt, recht gross (ORLOCI 1978). Wie später zu zeigen ist, bleiben die Funktionen aller Ordinationsmethoden stets dieselben: Es

geht um die Abbildung einer vieldimensionalen Ähnlichkeitsstruktur in möglichst wenig Dimensionen. Im Gegensatz etwa zu den Gruppierungsverfahren findet man, dass Ordinationsmethoden, welche sich nur durch das verwendete Ähnlichkeitsmaß unterscheiden, öfters unterschiedliche Namen tragen.

Gute Ordinationsmethoden sind mit einem hohen Rechenaufwand verbunden. Wohl bei keiner andern Gruppe von Methoden erscheinen in der pflanzensoziologischen Literatur so viele Arbeiten, welche sich anhand echter oder konstruierter Daten dem Vergleich und der Bewertung verschiedener Algorithmen annehmen (DEL MORAL 1980, GAUCH 1982). Dabei wird das Resultat in der Regel an den Erwartungen des Experimentators gemessen. Diese sind leider oft unrealistisch, wie z.B. FEOLI und FEOLI-CHIAPELLA (1980) treffend nachweisen.

Schliesslich muss betont werden, dass gute Ordinationen mehr leisten als nur eine grafische Darstellung von Beziehungen zwischen Arten oder Aufnahmen. Eines der fundamentalsten Probleme der Analyse ökologischer Daten besteht darin, dass zahlreiche gleichzeitig messbare Faktoren in unterschiedlicher Masse miteinander korreliert sind. Damit ist ihre separate Interpretation nicht sinnvoll. Methoden, wie z.B. die Hauptkomponentenanalyse, liefern eine Lösung zur Entkopplung simultan wirkender Faktoren. Eine pflanzenökologische Analyse, die diesen Schritt umgeht, ist deshalb kaum denkbar.

In der pflanzenökologischen Literatur wird zwischen direkter und indirekter Ordination unterschieden (WHITTAKER 1967, GAUCH 1982). Bei der direkten Ordination geht es darum, Aufnahmen in einem zweidimensionalen Koordinatensystem so darzustellen, dass ihre Ähnlichkeiten bezüglich zweier Standortsfaktoren ersichtlich werden. Als Achsen dienen deshalb Messwerte wie Höhenlage, pH, Bodenfeuchte usw. Auch ein Oekogramm ist in diesem Sinne eine Ordination (vgl. ELLENBERG 1978), worin statt einzelner Aufnahmen die Standortsbereiche ganzer Pflanzengesellschaften aufgezeichnet werden. Das vorliegende Kapitel beschäftigt sich jedoch

ausschliesslich mit der indirekten Ordination. Deren Prinzip wird im folgenden erläutert.

7.2 Prinzip der Ordinationsmethoden

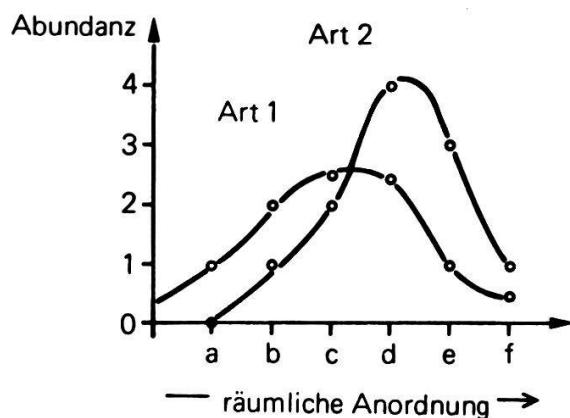
Das Prinzip der Ordination kann an einem Beispiel veranschaulicht werden, das einen Vegetationsgradienten darstellt. Abb. 7.1, A, zeigt zunächst die konstruierten Daten als Vegetationstabelle, der Einfachheit halber mit nur zwei Arten. Nehmen wir an, dass die Aufnahmen in gleichen Abständen entlang eines Transsektes erhoben wurden, so lassen sich die Ergebnisse wie in Abb. 7.1, B, darstellen. Daraus wird klar ersichtlich, dass es sich um einen einfachen, unverzweigten Vegetationsgradienten handelt. Eine solche Gradientenstruktur eignet sich naturgemäß gut zur Darstellung als Ordination. Letztere ist in Abb. 7.1, C, zu sehen. Als Achsen werden die beiden Pflanzenarten, als Koordinaten deren Artmächtigkeiten verwendet. Dabei zeigt es sich, dass die Abfolge der Aufnahmen genau derjenigen im Felde entspricht, die Anordnung jedoch fast kreisförmig ist. Dieses Muster ist für Vegetationsgradienten typisch (FEOLI und FEOLI-CHIAPELLA 1980). Es muss betont werden, dass Abb. 7.1, C, sich durch einfache Darstellung der Daten in Abb. 7.1, A, und ohne Umrechnung ergibt. Oval- und kreisförmige Ähnlichkeitsmuster werden fälschlicherweise immer wieder als durch die Hauptkomponentenanalyse verursacht gedeutet. Letztere ist deshalb - zu unrecht - vielerorts wenig beliebt (VAN DER MAAREL 1980, GAUCH 1982). Dass sie die Ähnlichkeitsverhältnisse perfekt abzubilden erlaubt, wurde jedoch mehrfach nachgewiesen und wird auch im folgenden an einem Beispiel gezeigt.

Die Ordination in Abb. 7.1 ist leicht interpretierbar und bedarf für praktische Anwendungen kaum weiterer Bearbeitung. Geht man jedoch bei komplizierten Daten genau gleich vor, so ergeben sich zwei Probleme, für welche z.B. die Hauptkomponentenanalyse eine Lösung bietet. Das erste Problem ergibt sich aus der Anzahl der Dimensionen. Eine unverzerrte Darstellung der Ähnlichkeitsverhältnisse setzt voraus, dass ebensoviele Achsen berücksichtigt werden, wie der Datensatz

A

	Aufnahme					
	a	b	c	d	e	f
1	1	2	2,5	2,5	1	0,5
2	0	1	2	4	3	1

B



C

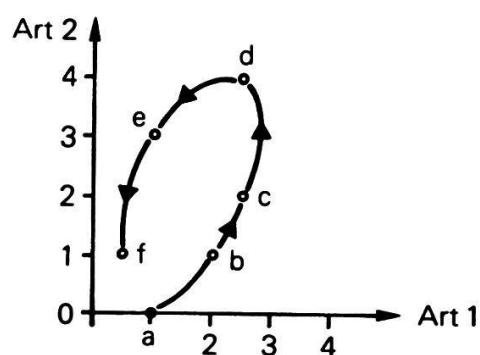


Abb. 7.1 Ein Vegetationsgradient a bis f, der von zwei Arten gebildet wird. Er ist dargestellt als Vegetationstabelle (A), als Transsekte (B) und als Ordination (C).

Arten enthält. Da aber mehr als drei Dimensionen grafisch nur schwer gleichzeitig darstellbar sind, ergibt sich die Forderung nach einer Verminderung der Achsen. Es ist klar, dass nicht jede Art gleichermassen guten Einblick in die Datenstruktur gewährt. Die meisten Ordinationsverfahren bestimmen die Achsen so, dass die Aehnlichkeitsverhältnisse möglichst unverändert wiedergegeben werden. Dabei fallen optimale Achsen selten genau mit einer einzelnen Art zusammen, sondern liegen irgendwo zwischen mehreren, aussagekräftigen Arten.

Das zweite Problem liegt in der Beziehung zwischen den Achsen. Die Punkte in Abb. 7.1, C sind deutlich korreliert. Die zweidimensionale, rechtwinklige Darstellung ist einigermassen irreführend. Art 2 enthält teilweise die gleiche Information wie Art 1 und ist somit als unabhängige Dimension ungeeignet. Die Hauptkomponentenverfahren bestimmen z.B. die Achsen so, dass sie völlig unkorreliert sind. Die resultierenden Koordinaten eignen sich daher besser zur Konstruktion rechtwinkliger Ordinationen.

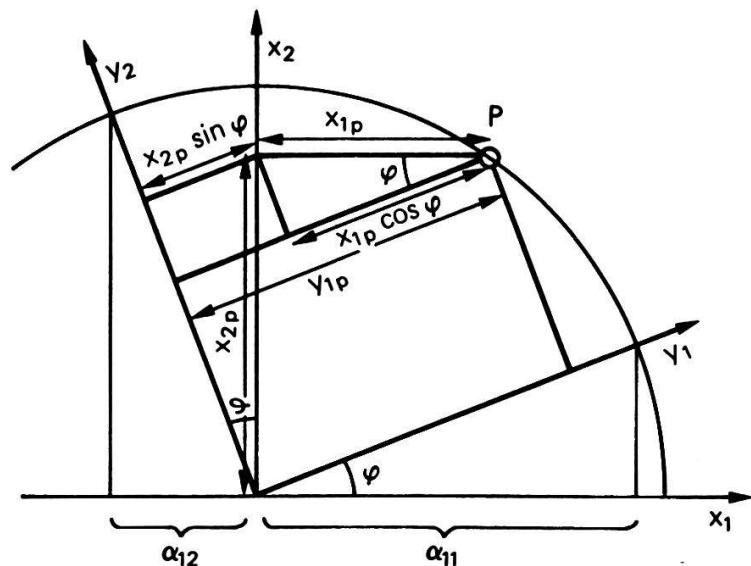
7.3 Die Hauptkomponentenanalyse

Die Aehnlichkeitsverhältnisse zwischen Aufnahmen sind durch die Daten, eventuelle Transformationen und das verwendete Aehnlichkeitsmass abschliessend festgelegt. Aufgabe der Hauptkomponentenanalyse ist es, von den ursprünglichen Variablen (Arten) ausgehend neue, unkorrelierte Achsen zu finden. Der Weg dazu führt ausschliesslich über eine Koordinatentransformation, die einigen speziellen Bedingungen zu genügen hat.

Zum besseren Verständnis der Vorgänge betrachten wir im folgenden zuerst das Ergebnis einer Koordinatentransformation. Davon ausgehend wird der Lösungsweg erörtert. In Abb. 7.2 ist das allgemeine geometrische Problem dargestellt.

Das ursprüngliche, rechtwinklige System (mit "Achsen") ist um einen bestimmten Winkel, φ , gekippt. Ein gegebener Punkt P soll seine Lage im Raum nicht verändern. Im alten System

A



B

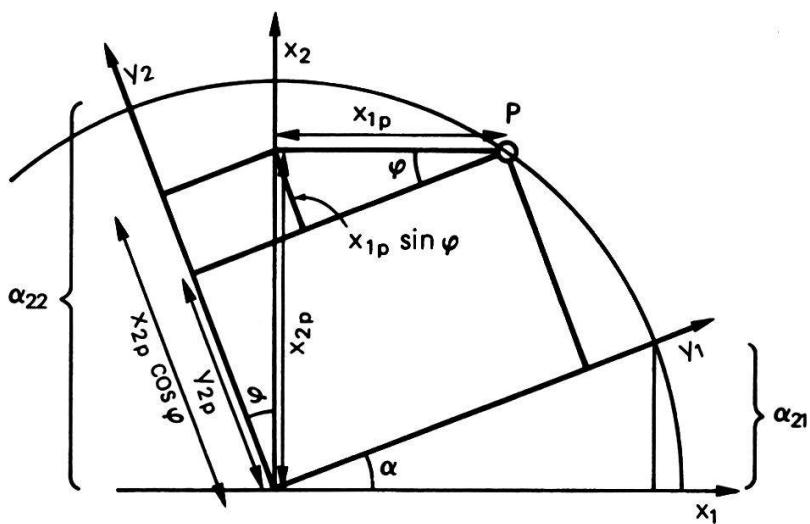


Abb. 7.2 Drehung eines Koordinatensystems mit den Achsen x_1 , x_2 . In A wird mit y_{1p} die erste neue, in B mit y_{2p} die zweite neue Koordinate des Punktes P bestimmt.

hat er die Koordinaten x_{1p} , x_{2p} . Diese sind im neuen System (mit "Faktoren") nicht mehr gültig, sondern sind durch y_{1p} , y_{2p} zu ersetzen. In Abb. 7.2, A ist dargestellt, wie die neuen Koordinaten aus den alten hergeleitet werden. Die Berechnung von y_{1p} zeigt Abb. 7.2, A:

$$y_{1p} = x_{1p} \cos \varphi + x_{2p} \sin \varphi$$

Die neue Koordinate setzt sich also aus zwei Teilstrecken (Summanden) zusammen. Dabei spielt natürlich auch der Drehwinkel φ eine Rolle. Um y_{1p} interpretieren zu können, muss die Bedeutung von $\cos \varphi$ und $\sin \varphi$ verstanden werden. In Abb. 7.2, A sind diese mit α_{11} und α_{12} bezeichnet. α_{11} erreicht den Maximalwert 1, wenn Achse 1 und Faktor 1 zusammenfallen. Der Minimalwert von 0 ergibt sich, wenn Achse 1 und Faktor 1 rechtwinklig zueinander stehen, also unabhängig sind. Die Vermutung, dass α_{11} ein Korrelationskoeffizient ist, lässt sich trigonometrisch leicht nachweisen (BATSCHET 1982). Es ist auch leicht einzusehen, dass α_{12} die Korrelation von Achse 1 mit Faktor 2 ist. Die Berechnung der neuen Koordinate y_{1p} lässt sich deshalb folgendermassen formulieren:

$$y_{1p} = x_{1p}^* \cdot (\text{Korrelation 1. Faktor mit 1. Achse}) + x_{2p}^* \cdot (\text{Korrelation 2. Faktor mit 1. Achse})$$

Analog gestaltet sich nun die Berechnung der zweiten Koordinaten (Abb. 7.2, B). In obiger Schreibweise ergibt sich

$$y_{2p} = x_{1p}^* \cdot (\text{Korrelation 1. Faktor mit 2. Achse}) + x_{2p}^* \cdot (\text{Korrelation 2. Faktor mit 2. Achse})$$

Die a -Werte spielen bei der Interpretation der Hauptkomponentenanalyse eine zentrale Rolle. Im Falle vegetationskundlicher Daten entsprechen die ursprünglichen Achsen den Pflanzenarten. Die Matrix α , Matrix der Faktorenladungen genannt, hat folgenden Aufbau:

	Faktor 1	Faktor 2	Faktor 3...	Faktor p
Art 1	α_{11}	α_{12}	α_{13}	\dots
Art 2	α_{21}	α_{22}	α_{23}	\dots
Art 3	α_{31}	α_{32}	α_{33}	\dots
.
Art p	α_{p1}	α_{p2}	α_{p3}	\dots

Darin ist z.B. α_{32} die Korrelation der Art 3 mit der neuen Ordinationsachse 2. Es geht nun darum, die α -Matrix zu berechnen. Der Rechengang soll anhand des Beispiels in Abb. 7.1 nachvollzogen werden. Dieses umfasst 6 Aufnahmen und nur 2 Arten, sodass der Arbeitsaufwand bescheiden bleibt. Wird dagegen mit drei, vier oder mehr Dimensionen gearbeitet, so empfiehlt sich die Verwendung eines Computerprogrammes.

Gegeben ist also die Vegetationstabelle A:

Art	Aufnahme						Σx	\bar{x}
	1	2	3	4	5	6		
1	1	2	2.5	2.5	1	0.5	9.5	1.58
2	0	1	2	4	3	1	11	1.83

Diese Daten müssen für eine normale Hauptkomponentenanalyse bezüglich der Arten zentriert werden (Es gibt auch eine Variante der Hauptkomponentenanalyse, die auf die Zentrierung verzichtet!). Dazu ist von jeder Artmächtigkeit der Mittelwert des jeweiligen Artvektors zu subtrahieren. Die neue Tabelle, X, sieht wie folgt aus:

Art	Aufnahme						Σx^2	\bar{x}
	1	2	3	4	5	6		
1	-0.58	0.42	0.92	0.92	-0.58	-1.08	3.71	0
2	-1.83	-0.83	0.17	2.17	1.17	-0.83	10.83	0

Abb. 7.3 A,B, vergleicht die Ordinationen der Vegetationsabellen A und X. Es ergibt sich wie erwartet nur eine

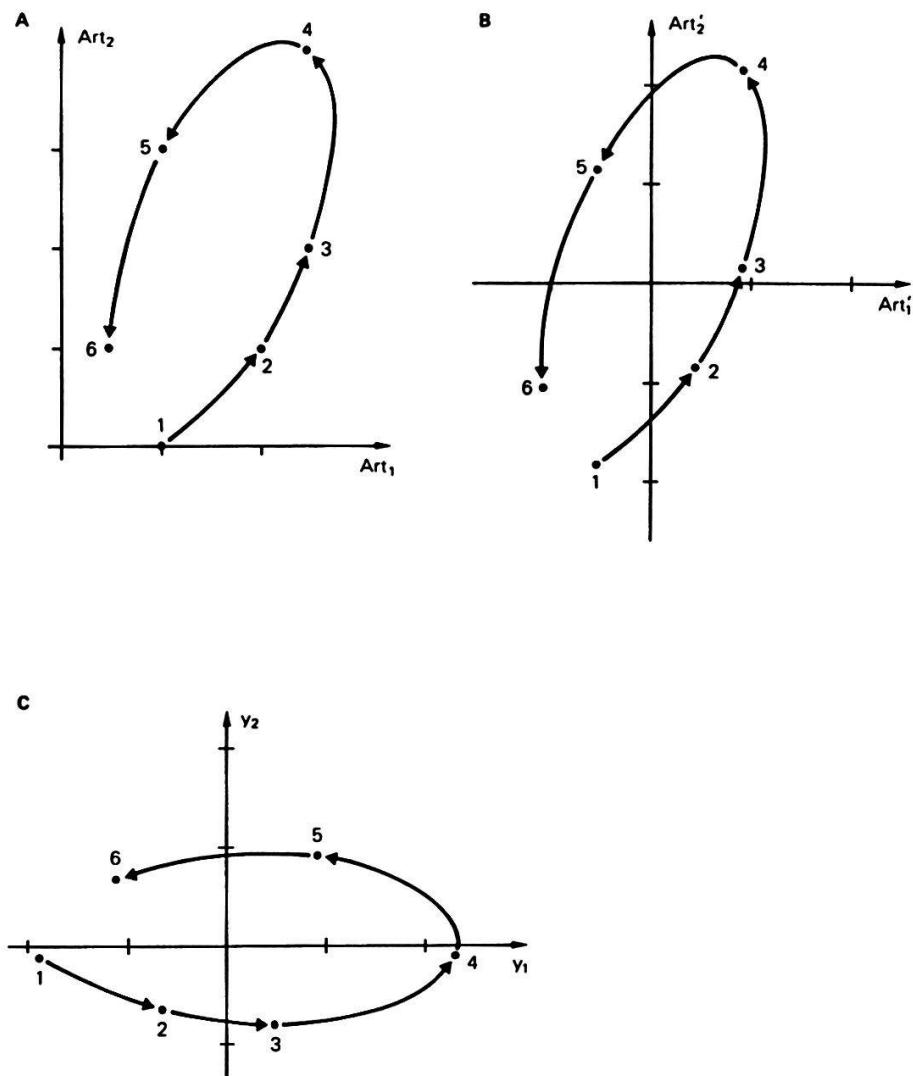


Abb. 7.3 Ordination von 6 Aufnahmen mit den Ausgangsdaten (A), den zentrierten Daten (B) und den Faktorenwerten einer Hauptkomponentenanalyse (C).

Verschiebung des Achsenkreuzes, ohne dass an den Ähnlichkeitsverhältnissen irgend etwas verändert würde. Als nächstes wird die Ähnlichkeitsmatrix der Arten der Vegetationstabelle A berechnet. Als Ähnlichkeitsmaß dient das Skalarprodukt, doch könnten ebensogut die Kovarianz oder der Korrelationskoeffizient verwendet werden. Wir erhalten

$$S = \begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} = \begin{bmatrix} 3.71 & 3.08 \\ 3.08 & 10.80 \end{bmatrix} .$$

Darin ist 3.71 das Skalarprodukt (die $n-1$ fache Varianz) der Art 1, 10.8 dasjenige der Art 2. Die beiden Arten besitzen auch einen deutlich positiven Zusammenhang von 3.08.

Für die weiteren Betrachtungen machen wir einen Sprung und studieren zunächst die Streuungsverhältnisse des Endresultates. Nach Beendigung der Hauptkomponentenanalyse werden die Arten durch Faktoren ersetzt sein. Die neue Tabelle Y wird so aussehen:

Faktor	Aufnahme					
	1	2	3	4	5	6
1	-1.92	-0.635	0.476	2.35	0.889	-1.16
2	-0.09	-0.68	-0.800	-0.10	0.95	0.72

In Abb. 7.3, C, ist die daraus konstruierbare Ordination dargestellt. Beim genauen Vergleich mit Abb. 7.3, A und B ergibt sich, dass die Beziehungen zwischen den Aufnahmen gleich geblieben sind. Es ist nur eine Veränderung festzustellen: Das Achsenystem wurde gedreht (oder, was dasselbe ist, der Punkteschwarm wurde gedreht). Zweck dieser Drehung ist es, die Achse 1 (Faktor 1) so zu legen, dass sie den Punkteschwarm möglichst gut auflöst. Ferner sind jetzt die Achsen 1 und 2 unkorreliert, was sich in der Vegetationstabelle Y leicht nachrechnen lässt. Die Faktorenladungen, welche zu diesem Resultat führten, sollen nun berechnet werden. Die Lösung erfolgt in zwei Schritten:

1. Berechnung der Skalarprodukte der neuen Koordinaten Y (Eigenwertbestimmung).
2. Berechnung der Faktorenladungen α (Eigenvektorbestimmung).

Aufgrund der Zielsetzung der Hauptkomponentenanalyse kann schon jetzt gesagt werden, dass die Ähnlichkeitsmatrix der neuen Ordinationskoordinaten wie folgt aussieht:

$$\begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}$$

Darin steht λ_1 für das Skalarprodukt der ersten Achse, λ_2 für dasjenige aller Werte der zweiten Achse. λ_1 wird als erster, λ_2 als zweiter Eigenwert bezeichnet (als Symbol dafür wird in der Literatur meist das griechische Lambda verwendet). λ_1 soll dabei so gross wie möglich sein, um die Auflösung des 1. Faktors maximal zu halten. Alle Elemente ausserhalb der Diagonalen sind gleich null, weil ja die Faktoren unkorreliert sein müssen. Eine weitere Einsicht hilft bei der Bestimmung der λ -Werte (Eigenwerte): Da die Ähnlichkeitsstruktur unverändert bleibt, ändert sich auch die Summe der Skalarprodukte nicht. Es gilt also

$$\lambda_1 + \lambda_2 = s_{11} + s_{22} = 3.71 + 10.8 = 14.51$$

Um die gesuchten Werte für λ zu erhalten, bedient man sich der Differentialrechnung (vgl. BATSCHELET 1982). Man findet, dass die sogenannte Determinante der Ähnlichkeitsmatrix null wird, wenn von jedem Diagonalelement die gesuchten Eigenwerte subtrahiert werden:

$$\begin{vmatrix} 3.71 - \lambda & 3.08 \\ 3.08 & 10.8 - \lambda \end{vmatrix} = 0$$

Der Vollständigkeit halber sei auch noch die Matrixschreibweise dieser Gleichung angeführt. Sie erscheint in allen einschlägigen Lehrbüchern und lautet $S - \lambda I = 0$. Darin ist I eine Einheitsmatrix, d.h. eine Matrix, deren Dia-

gonale mit 1, die übrigen Positionen mit 0 besetzt sind. Diese Gleichung ist nun nach λ zu lösen. Nach den Regeln der Matrizenrechnung kann sie wie folgt ausmultipliziert werden (BATSCHET 1982):

$$(3.71 - \lambda)(10.8 - \lambda) - (3.08)(3.08) = 0 \text{ oder}$$

$$\lambda^2 - 6.2 \lambda + 9.2 = 0$$

Die quadratische Gleichung wird nach bekannter Vorschrift

$$\lambda = [-b \pm (b^2 - 4ac)^{1/2}] / 2a$$

gelöst. Wir erhalten

$$\begin{aligned}\lambda_1 &= (14.51 + (14.51^2 - (4)(30.58))^{1/2}) / 2 = 11.95 \\ \lambda_2 &= (14.51 - (14.51^2 - (4)(30.58))^{1/2}) / 2 = 2.56\end{aligned}$$

Es kann gezeigt werden, dass 11.95 tatsächlich der maximal mögliche Wert für das Skalarprodukt des ersten Faktors ist, während λ_2 das Skalarprodukt des zweiten Faktors darstellt (BATSCHET 1982). Auch ergibt sich mit $\lambda_1 + \lambda_2 = 14.51$ das Skalarprodukt der Ausgangsmatrix S.

Nun schreiten wir zur Bestimmung der Faktorenladungen α . Wir beginnen mit dem ersten Faktor und suchen α_{11} und α_{21} . Diese sind aus der alten und der neuen Kovarianzmatrix direkt bestimmbar (BATSCHET 1982). Es gilt nämlich folgende Beziehung:

$$\begin{bmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{bmatrix} \begin{bmatrix} \alpha_{11} \\ \alpha_{21} \end{bmatrix} = \lambda_1 \begin{bmatrix} \alpha_{11} \\ \alpha_{21} \end{bmatrix}.$$

Setzen wir die schon bekannten Werte ein, so erhalten wir:

$$\begin{bmatrix} 3.71 & 3.08 \\ 3.08 & 10.8 \end{bmatrix} \begin{bmatrix} \alpha_{11} \\ \alpha_{21} \end{bmatrix} = 11.95 \begin{bmatrix} \alpha_{11} \\ \alpha_{21} \end{bmatrix}.$$

Nach den Regeln der Matrixrechnung ausmultipliziert ergeben sich zwei Gleichungen mit den Unbekannten α_{11} und α_{21} :

$$\begin{aligned} 3.71 \alpha_{11} + 3.08 \alpha_{21} &= 11.95 \alpha_{11} \\ 3.08 \alpha_{11} + 10.8 \alpha_{21} &= 11.95 \alpha_{21} \end{aligned}$$

Die Auflösung ist nicht ganz trivial. Zunächst finden wir:

$$\begin{aligned} -8.24 \alpha_{11} + 3.08 \alpha_{21} &= 0 \\ 3.08 \alpha_{11} - 1.15 \alpha_{21} &= 0 \end{aligned}$$

Durch Addition der Gleichungen ergibt sich

$$-5.16 \alpha_{11} + 1.93 \alpha_{21} = 0 .$$

Daraus berechnet sich das Verhältnis von α_{11} zu α_{21} :

$$\alpha_{11} = \frac{1.93}{5.16} \alpha_{21} = 0.374 \alpha_{21}$$

Substituiert man z.B. α_{21}^{11} in einer der obigen Gleichungen, so erhält man nur die triviale Lösung $\alpha_{11}^{11} = \alpha_{21}^{21} = 0$. Aus unendlich vielen Möglichkeiten wählen wir eine aus, welche die Bedingung $\alpha_{11} = 0.374 \alpha_{21}$ erfüllt. Wir setzen kurzerhand und willkürlich

$$\alpha_{21} = 1$$

$$\alpha_{11} = \frac{1.93}{5.16} \alpha_{21} = \frac{1.93}{5.16} = 0.374$$

Die a-Werte haben die besondere Eigenschaft, dass ihre absolute Grösse gar nicht eindeutig zu bestimmen ist. Festgelegt ist nur das Größenverhältnis α_{11}/α_{21} . Es ist üblich, den Vektor

$$\begin{bmatrix} \alpha_{11} \\ \alpha_{21} \end{bmatrix} = \begin{bmatrix} 0.374 \\ 1.0 \end{bmatrix}$$

zu normalisieren. Dazu ist, wie früher gezeigt, jedes Element durch die euklidsche Länge des Vektors zu dividieren:

$$\begin{bmatrix} \frac{0.374}{\sqrt{((0.374)^2 + 1^2)^{1/2}}} \\ \frac{1}{\sqrt{((0.374)^2 + 1^2)^{1/2}}} \end{bmatrix} = \begin{bmatrix} 0.350 \\ 0.937 \end{bmatrix}$$

Nun setzt man die Gleichungen für die Bestimmung von α_{12} und α_{22} auf:

$$\begin{aligned} 3.71 \alpha_{12} + 3.08 \alpha_{22} &= 2.56 \alpha_{12} \\ 3.08 \alpha_{12} + 10.8 \alpha_{22} &= 2.56 \alpha_{22} \end{aligned}$$

Es bleibt dem Leser überlassen, die Rechnung durchzuführen. Die normalisierten Ergebnisse lauten:

$$\begin{aligned} \alpha_{12} &= 0.937 \\ \alpha_{22} &= 0.350 \end{aligned}$$

Nachdem nun alle Faktorenladungen bekannt sind, schreiten wir zur Berechnung der neuen Koordinaten Y. Sie folgt der bereits untersuchten Regel

$$y_{1p} = x_{1p} \alpha_{11} + x_{2p} \alpha_{21}$$

Die aktuellen Werte sind:

$$\begin{aligned} y_{11} &= (-0.58)(0.35) + (-1.83)(0.937) = -1.92 \\ y_{12} &= (0.42)(0.35) + (-0.83)(0.937) = -0.63 \\ y_{13} &= (0.92)(0.35) + (0.17)(0.937) = 0.48 \\ \text{usw.} & \end{aligned}$$

Damit ist die Hauptkomponentenanalyse abgeschlossen. Für die Interpretation fassen wir die Resultate zusammen. Wir erhalten folgende Eigenwerte:

$$\begin{aligned}\lambda_1 &= 11.95 = 82 \% \\ \lambda_2 &= 2.56 = 18 \% \\ \lambda_1 + \lambda_2 &= 14.51 = 100\%\end{aligned}$$

Die aus den neuen Koordinaten zu konstruierende Ordination (Abb. 7.3) weist entlang der 1. Achse 82 % der Unterschiedlichkeiten (Varianz) der Aufnahme aus, die 2. Achse nur 18 %.

Die Faktorenladungen sehen wie folgt aus:

Art Faktor	1	2
1	0.35	0.937
2	-0.937	0.35

Der neue Faktor 1 steht mit einem Korrelationskoeffizienten von 0.937 der alten Art 2 nahe. Entsprechend liegt Faktor 2 in der Nähe der alten Art 1, ist aber gegenläufig zu dieser (-0.937).

Schliesslich erhielten wir die neuen Faktorenwerte:

Faktor	Aufnahme					
	1	2	3	4	5	6
1	-1.92	-0.635	0.476	2.35	0.889	-1.16
2	-0.09	-0.68	-0.80	-0.10	0.95	0.72

Ohne die Ordination zu konsultieren, können aus der Tabelle Aufnahmen herausgesucht werden, die besonders ähnlich sind. Dies ist dann der Fall, wenn alle Faktorenwerte (vor allem aber die ersten) etwa gleiche Grösse besitzen. In unserem Beispiel liegen jedoch alle Punkte deutlich auseinander, wie wir aus der bereits gezeichneten Ordination wissen (Abb. 7.3).

Eine Besonderheit pflanzensoziologischer Anwendungen der Hauptkomponentenanalyse ist, dass mit sehr vielen Variablen (Arten) gerechnet wird. Selbst nach einer Datenreduktion bleiben noch 50, 100 oder mehr Eigenwerte zu bestimmen. Es sind daher lange Rechenzeiten zu erwarten. Für die Interpretation stellt sich die Frage, wie viele Dimensionen überhaupt betrachtet werden sollen. Dazu gibt es Regeln, die für pflanzensoziologische Tabellen erfahrungsgemäss unzweckmässig sind (UEBERLA 1971). Im allgemeinen zeigt es sich, dass bei Tabellen der Grössenordnung 100 mal 100 mindestens 4, höchstens 6 bis 8 Dimensionen interpretiert werden sollten. Die damit erklärte Varianz kann um 15 bis 30% liegen. Mit diesem doch bescheidenen Anteil wird es aber meist möglich sein, einen guten Einblick in die Ähnlichkeitsstruktur der Aufnahmen zu erhalten. Der Rest geht auf Konto "Rauschen".

7.4 Q- und D-Technik

Die bereits gezeigte Hauptkomponentenanalyse wird auch als R-Analyse (R-Technik) bezeichnet. Der Begriff Technik besagt, dass es sich nicht eigentlich um verschiedene Methoden handelt, sondern lediglich um einen anderen Wege zum selben Resultat.

Im vorigen Kapitel berechneten wir mit Hilfe der R-Technik Koordinaten für 6 Aufnahmen. Dazu benützten wir eine zweidimensionale Ähnlichkeitsmatrix der Arten. Stellen wir uns nun die Aufgabe, eine Ordination der Arten zu berechnen, so muss bei der R-Technik die sechsdimensionale Ähnlichkeitsmatrix der Aufnahmen analysiert werden. Das ist recht aufwendig, da insgesamt 6 Eigenwerte und 6 Eigenvektoren zu berechnen sind. Tun wir das anhand unseres Beispiels mit Hilfe eines Computerprogrammes, so sehen wir, dass fünf Eigenwerte gleich null sind: Die Ähnlichkeitsverhältnisse zweier Arten lassen sich nämlich zwangslos auf einer einzigen Achse darstellen. Der Rechenaufwand zur Bestimmung des einzigen brauchbaren Eigenwertes wäre also übermäßig hoch.

Die Q-Technik bietet in diesem Fall die Möglichkeit, die Koordinaten der Arten direkt aus einer zweidimensionalen Artmatrix zu bestimmen. Identische Ergebnisse sind aber immer auch über eine R-Analyse zu erzielen.

Die Q-Technik wird so begonnen, als ob eine R-Aehnlichkeitsmatrix der Aufnahmen erstellt werden müsste. Dazu sind letztere zu zentrieren. Erinnern wir uns der Rohdaten A:

Aufn.	1	2	3	4	5	6
Art 1	1	2	2.5	2.5	1	0.5
Art 2	0	1	2	4	3	1

Die Mittelwerte der Aufnahmen \bar{x}_i , betragen

$$(0.5 \quad 1.5 \quad 2.25 \quad 3.25 \quad 2 \quad 0.75)$$

Nun wird von jeder Artmächtigkeit der Mittelwert der jeweiligen Aufnahme subtrahiert und es resultiert die zentrierte Tabelle B:

Aufn.	1	2	3	4	5	6
Art 1	0.5	0.5	-0.25	-0.75	-1	-0.25
Art 2	-0.5	-0.5	0.25	0.75	1	0.25

Obwohl nun also die Aufnahmen zentriert sind, wird das Skalarprodukt der Arten berechnet, die sogenannte Q- Matrix. Wir erhalten für die ersten beiden Elemente:

$$q_{11} = 0.5^2 + 0.5^2 - 0.25^2 - 0.75^2 - 1^2 - 0.25^2 = 2.1875$$

$$q_{12} = (0.5)(-0.5) + (0.5)(-0.5) + (-0.25) + (0.75)(0.75) + (-1)(1) + (-0.25)(0.25) = -2.1875$$

Die gesamte Matrix ergibt

$$Q = \begin{matrix} 2.1875 & -2.1875 \\ -2.1875 & 2.1875 \end{matrix}$$

Nun bestimmt man die Eigenwerte wie bei der R-Technik. Für die ausgeschriebene Determinante erhält man:

$$(2.1875 - \lambda)(2.1875 - \lambda) - (-2.1875)(-2.1875) = 0$$

Nach den Multiplikationen ergibt sich

$$\lambda^2 - 4.375\lambda = 0$$

Wir dividieren durch λ und erhalten die einzige von null verschiedene Lösung

$$\lambda_1 = 4.375$$

Die Eigenvektoren sind unter diesen Umständen rasch zu bestimmen. Zu lösen ist die Gleichung

$$\begin{bmatrix} 2.1875 & -2.1875 \\ -2.1875 & 2.1875 \end{bmatrix} \begin{bmatrix} \alpha_{11} \\ \alpha_{21} \end{bmatrix} = 4.375 \begin{bmatrix} \alpha_{11} \\ \alpha_{21} \end{bmatrix}$$

Daraus ergeben sich die zwei Gleichungen

$$\begin{aligned} 2.1875\alpha_{11} - 2.1875\alpha_{21} &= 4.375\alpha_{11} \\ -2.1875\alpha_{11} + 2.1875\alpha_{21} &= 4.375\alpha_{21} \end{aligned}$$

Setzt man $\alpha_{11} = 1$ und löst die obere Gleichung nach α_{21} auf, so ergibt sich

$$\begin{aligned} -2.1875\alpha_{21} &= 2.1875 \\ \alpha_{21} &= -1 \end{aligned}$$

Auch mit $\lambda_2 = 0$ kann ein Eigenvektor bestimmt werden. Es ist wiederum zweckmäßig, $\alpha_{12} = 1$ zu setzen. Für α_{22} erhält man dann -1. Zusammengefasst ergibt sich für die Faktorenlösungen

$$\begin{bmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -1 & -1 \end{bmatrix}$$

Normalisiert betragen sie

$$\begin{bmatrix} 0.707 & 0.707 \\ -0.707 & -0.707 \end{bmatrix}$$

Im Gegensatz zur R-Technik ist nun noch jeder Vektor der Grösse des Eigenwertes anzupassen:

$$\alpha_{11}^2 + \alpha_{21}^2 = \lambda_1 = 4.375$$

Wir multiplizieren alle aktuellen Werte mit $\lambda^{1/2}$ und erhalten die gesuchten Koordinaten

$$\alpha'_{11} = y_{11} = 1.479$$

$$\alpha'_{21} = y_{21} = 1.479$$

Für die zweite Achse betragen die Koordinaten

$$y_{12} = y_{22} = 0.$$

Das Beispiel hat deutlich gemacht, dass sich mit der Q-Technik unter Umständen der Rechenaufwand beträchtlich reduzieren lässt. Gelegentlich kommt es auch vor, dass statt einer Q-Matrix bereits eine Distanzmatrix verfügbar ist. OR-LOCI (1973, 1978) hat gezeigt, dass sich aus dieser, nach entsprechender Transformation, dieselben Eigenwerte und Eigenvektoren extrahieren lassen. Auch unser Beispiel eignet sich für die D-Technik. Ob die Ausgangsdaten zentriert sind oder nicht, spielt dabei keine Rolle. Wir erhalten:

$$D^2 = \begin{bmatrix} 0 & 8.75 \\ -8.75 & 0 \end{bmatrix}$$

Die d^2 -Distanzen sind nun in $-0.5d^2$ - Ähnlichkeitskoeffizienten umzuwandeln:

in spezifischer Weise transformiert. Schliesslich wird eine kombinierte R- und Q-Hauptkomponentenanalyse durchgeführt. Aus grundsätzlichen Überlegungen heraus dürften streng genommen nur Zähldaten (Frequenzen) in dieser Weise analysiert werden. Dass trotzdem Messdaten verwendet werden, ist in der Praxis die Regel.

Der Rechengang könnte nun ohne weiteres am Beispiel des vorigen Kapitels gezeigt werden. Dies wäre mit Nachteilen verbunden, wie in Kapitel 7.6 noch zu sehen sein wird. Deshalb bedienen wir uns eines Datensatzes mit 4 Aufnahmen und 3 Arten, der trotz seiner Kleinheit für Vegetationsaufnahmen typisch ist.

Diese Vegetationstabelle (Tabelle 7.1) wird als Kontingenztafel aufgefasst. Wir verwenden Notierungen, wie sie in Tab. 7.1, A, dargestellt sind. $f_{..}$ steht z.B. für die Summe aller Artmächtigkeiten der Aufnahme 2, $f_{.3}$ für die Summe innerhalb der Art 3, usw. Die Rohdaten, in Tabelle 7.1, B, dargestellt, werden nun schrittweise transformiert, und zwar zunächst zu relativen Häufigkeiten bezüglich der Gesamtsumme $f_{..}$:

$$p_{ij} = f_{ij} / f_{..}$$

Im Zahlenbeispiel erhält man dadurch eine neue Matrix P, wie sie aus Tab. 7.2, A, ersichtlich ist.

Nun wird die Annahme getroffen, dass die Randsummen $f_{.j}$ und $f_{i.}$ bzw. $p_{.j}$ und $p_{i.}$ für den gesamten Datensatz charakteristisch sind. Jede Artmächtigkeit ist deshalb nur bedeutsam, wenn sie von der charakteristischen (d.h. zu erwartenden) Häufigkeit abweicht. Es sind also Erwartungswerte zu berechnen. Diese ergeben sich aus der relativen Häufigkeit einer Art (bezüglich des Gesamtdatensatzes) und der relativen Häufigkeit der Individuen innerhalb einer Aufnahme:

$$e_{ij} = (p_{.j})(p_{i.})$$

Für die erste Artmächtigkeit erhalten wir

$$-0.5D^2 = \begin{bmatrix} 0 & -4.375 \\ -4.375 & 0 \end{bmatrix}$$

Die Eigenwerte berechnen sich nach bewährtem Schema:

$$(0-\lambda)(0-\lambda) - 4.375^2 = 0 \\ \lambda^2 = 4.375^2 \\ \lambda = 4.375$$

Das Ergebnis ist identisch mit demjenigen der Q-Analyse. Führt man die Analyse weiter, ergeben sich auch gleiche Eigenvektoren und schliesslich die gewünschten Koordinaten.

In der Praxis spielt die Q-Technik eine untergeordnete Rolle, denn es gibt Algorithmen, die nur die grössten Eigenwerte extrahieren und daher auch in ungünstigen Fällen einigermassen effizient arbeiten. Die D-Technik erweist sich als interessante Alternative, wenn die übrigen Untersuchungen ohnehin auf einer Distanzmatrix basieren.

7.5 Korrespondenzanalyse

Die Korrespondenzanalyse ist ein Hauptkomponentenverfahren, das sich bei Pflanzenökologen grosser Beliebtheit erfreut. Insbesondere im französischen Sprachbereich hat sie sich als Standardmethode eingebürgert (BENZECRI 1969, PIELOU 1977). Die angelsächsische Welt ist durch die Arbeiten von HILL (1974, 1979a) auf ihre Nützlichkeit aufmerksam geworden. Während bei der gewöhnlichen Hauptkomponentenanalyse zwei Rechengänge erforderlich sind um Aufnahmen und Arten zu ordinieren, tut die Korrespondenzanalyse beides gleichzeitig. Mehr noch: Die beiden Koordinatensätze sind gleich skaliert, sodass sie übereinander projiziert gemeinsam interpretiert werden dürfen. Der Zusammenhang zwischen Arten und Aufnahmen ist so besonders leicht erkennbar.

Formal unterscheidet sich die Korrespondenzanalyse nur wenig von den Hauptkomponentenanalysen: Zunächst werden die Daten

$$e_{11} = (0.143)(0.286) = 0.041$$

Zusammengefasst ergeben sich die Erwartungswerte in Tab. 7.2, B. Die höchsten Erwartungswerte erzielt Aufnahme 2, die ja auch die reichhaltigste ist. Schliesslich ist die Abweichung der effektiven Häufigkeiten p_{ij} von den erwarteten e_{ij} zu berechnen:

$$d_{ij} = p_{ij} - e_{ij}$$

Für das erste Element erhalten wir

$$d_{11} = 0.143 - 0.041 = 0.102$$

Die vollständige D-Matrix ist in Tab. 7.2, C dargestellt. Auf Grund dieser Abweichungen von den erwarteten, relativen Häufigkeiten wird nun weitergerechnet. Es fällt sofort auf, dass sowohl die Artvektoren wie auch die Aufnahmen den Mittelwert null besitzen, d.h. dass sie bereits zentriert sind. Erinnern wir uns der Hauptkomponentenanalyse, so könnte so gleich die R- oder die Q-Technik angewandt werden. In der Tat werden beide gleichzeitig durchgeführt. Wir beginnen mit der Matrix der Skalarprodukte der Arten (mir der Matrix der Aufnahmen ergäbe sich genau dasselbe Resultat!). Wir erhalten für den Vergleich der Arten 1 und 2

$$s_{12} = (0.102)(-0.061) + (0.020)(-0.041) + (-0.081)(0.020) + (-0.041)(0.082) = 0.012$$

Die gesamte Matrix lautet:

$$S = \begin{matrix} 0.0192 & -0.0121 & -0.0071 \\ -0.0121 & 0.0125 & -0.0004 \\ -0.0071 & -0.0004 & 0.0075 \end{matrix}$$

Wie bei der Hauptkomponentenanalyse werden die Eigenwerte aus der Determinantengleichung bestimmt:

Tabelle 7.1

Vegetationstabelle, die mit Hilfe der Korrespondenzanalyse untersucht werden soll. Formale Darstellung als Kontingenztafel (A) und Zahlenbeispiel (B).

<u>A</u>	Art/Aufn.	j=1	j=2	j=3	j=4	Σ
$F =$	i=1	f_{11}	f_{12}	f_{13}	f_{14}	$f_{1.}$
	i=2	f_{21}	f_{22}	f_{23}	f_{24}	$f_{2.}$
	i=3	f_{31}	f_{32}	f_{33}	f_{34}	$f_{3.}$
	Σ	$f_{.1}$	$f_{.2}$	$f_{.3}$	$f_{.4}$	$f_{..}$

<u>B</u>	Art/Aufn.	1	2	3	4	Σ Arten
	1	1	1			2
	2		1	1	1	3
	3		1	1		2
	Σ Aufn.	1	3	2	1	Total = 7

Tabelle 7.2

Abgeleitete Tabellen des Zahlenbeispiels in Tab. 7.1, B.
Relative Häufigkeiten (A), Erwartungswerte (B), Abweichungen
von den Erwartungswerten (C).

<u>A</u>	Art	Aufn.	1	2	3	4	Σ
$P =$	1		0.143	0.143	0.0	0.0	0.286
	2		0.0	0.143	0.143	0.143	0.429
	3		0.0	0.143	0.143	0.0	0.286
	Σ		0.143	0.429	0.286	0.143	1.0

<u>B</u>	Art	Aufn.	1	2	3	4	Σ
$E =$	1		0.041	0.123	0.081	0.041	0.286
	2		0.061	0.184	0.123	0.061	0.429
	3		0.041	0.123	0.081	0.041	0.286
	Σ		0.143	0.429	0.286	0.143	1.0

<u>C</u>	Art	Aufn.	1	2	3	4	Σ
$D =$	1		0.102	0.020	-0.081	-0.041	0
	2		-0.061	-0.041	0.020	0.082	0
	3		-0.041	0.020	0.061	-0.041	0
	Σ		0	0	0	0	0

$$\begin{vmatrix} 0.0192 - \lambda & -0.0121 & -0.0071 \\ -0.0121 & 0.0125 - \lambda & -0.0004 \\ -0.0071 & -0.0004 & 0.0075 - \lambda \end{vmatrix} = 0$$

Multipliziert man diese Determinante aus, so erhält man eine kubische Gleichung dritten Grades. Schneller geht es mit einem Computerprogramm und wir finden

$$\begin{aligned}\lambda_1 &= 0.0297 \\ \lambda_2 &= 0.0094 \\ \lambda_3 &= 0\end{aligned}$$

Nun ist für jeden Eigenwert ein Eigenvektor α (Faktorenladungen) zu finden. Das Gleichungssystem

$$\begin{bmatrix} 0.0192 & -0.0121 & -0.0071 \\ -0.0121 & 0.0125 & -0.0004 \\ -0.0071 & -0.0004 & 0.0075 \end{bmatrix} \begin{bmatrix} \alpha_{11} \\ \alpha_{21} \\ \alpha_{31} \end{bmatrix} = 0.0297 \begin{bmatrix} \alpha_{11} \\ \alpha_{21} \\ \alpha_{31} \end{bmatrix}$$

ist zu lösen. Normalisiert lauten die Ergebnisse

$$\begin{aligned}\alpha_{11} &= -0.79670 \\ \alpha_{21} &= 0.55311 \\ \alpha_{31} &= 0.24359\end{aligned}$$

Analoge Berechnungen für $\lambda_2 = 0.0094$ ergeben

$$\begin{aligned}\alpha_{12} &= 0.17871 \\ \alpha_{22} &= 0.60061 \\ \alpha_{32} &= -0.77931\end{aligned}$$

Die Ermittlung des dritten Eigenvektors ist auch möglich:

$$\begin{aligned}\alpha_{13} &= 0.57735 \\ \alpha_{23} &= 0.57735 \\ \alpha_{33} &= 0.57735\end{aligned}$$

Nun geht es um die Bestimmung der neuen Aufnahmekoordinaten. Es gilt hier zu beachten, dass die alten Koordinaten nicht

unsere Rohwerte sind, sondern die transformierten Daten, die wir in der D-Matrix finden (Tab. 7.2, C):

Art	Aufnahme			
	1	2	3	4
1	0.102	0.020	-0.081	-0.041
2	-0.061	-0.041	0.020	0.082
3	-0.041	0.020	0.061	-0.041

$$= \begin{matrix} d_{11} & d_{12} & d_{13} & d_{14} \\ d_{21} & d_{22} & d_{23} & d_{24} \\ d_{31} & d_{32} & d_{33} & d_{34} \end{matrix}$$

Das bewährte Rezept für die neuen Koordinaten Y lautet:

$$\begin{aligned}
 y_{1.\text{Achse},1.\text{Aufn.}} &= d_{11} \alpha_{11} + d_{21} \alpha_{21} + d_{31} \alpha_{31} + d_{41} \alpha_{41} \\
 &= (0.102)(-0.79670) + (-0.061)(0.55311) + (-0.041)(0.24359) \\
 &= -0.12510
 \end{aligned}$$

$$\begin{aligned}
 y_{1.\text{Achse},2.\text{Aufn.}} &= d_{12} \alpha_{11} + d_{22} \alpha_{21} + d_{32} \alpha_{31} + d_{42} \alpha_{42} \\
 &= (0.020)(-0.79670) + (-0.041)(0.55311) + (0.020)(0.24359) \\
 &= -0.03386 \\
 &\text{usw.}
 \end{aligned}$$

Für alle Aufnahmen und Achsen ergibt sich

Achse	Aufnahme			
	1	2	3	4
1	-0.12510	-0.03386	0.09124	0.06773
2	0.01327	-0.03677	-0.05004	0.07354
3	0.0	0.0	0.0	0.0

Der letzte Schritt der Korrespondenzanalyse ist die Ableitung der Koordinaten der Arten, die wir Z nennen wollen. Auch dieses Vorgehen ist nicht neu. Wir brauchen nämlich die Faktorenladungen α wie in der Q-Technik der Hauptkomponentenanalyse bloss mit den Eigenwerten λ zu skalieren:

$$Z = \begin{matrix} -0.7967 \lambda_1^{1/2} & 0.17871 \lambda_2^{1/2} & -0.577 \lambda_3^{1/2} \\ 0.5531 \lambda_1^{1/2} & 0.60061 \lambda_2^{1/2} & -0.577 \lambda_3^{1/2} \\ 0.2436 \lambda_1^{1/2} & -0.77931 \lambda_2^{1/2} & -0.577 \lambda_3^{1/2} \end{matrix}$$

Ausmultipliziert finden wir

$$Z = \begin{matrix} -0.13732 & 0.01736 & 0.0 \\ 0.09534 & 0.05836 & 0.0 \\ 0.04198 & -0.07572 & 0.0 \end{matrix}$$

Somit ist die Korrespondenzanalyse abgeschlossen. Die Y- und Z-Werte dienen der Ordination der Aufnahmen und Arten (Abb. 7.4). In unserem sehr einfachen Beispiel stehen sich verschiedene Punkte sehr nahe, was leicht zu interpretieren ist. Wir finden:

- Für Aufnahme 1 ist Art 1 typisch.
- Für Aufnahme 3 ist Art 3 typisch.
- Für Aufnahme 4 ist Art 2 typisch.

Das Resultat leuchtet sofort ein, wenn man es anhand der Rohdaten überprüft. Man findet, dass die Aufnahmen 1 bis 3 einen floristischen Gradienten repräsentieren, in den sich Aufnahme 4 nicht zwangslässig einpassen lässt (Pfeil in Abb. 7.4). Es bestätigt sich, dass Aufnahme 2 nicht durch eine einzelne Art differenziert werden kann. Der Erklärungsgrad der beiden Ordinationsachsen ist ebenfalls klar:

$$\begin{aligned} \sum \lambda_i &= 0.0297 + 0.0094 = 0.0391 = 100\% \\ \lambda_1 &= 75.8\% \\ \lambda_2 &= 24.2\% \end{aligned}$$

Die x-Achse erklärt also 75.8% der Varianz zwischen den Ordinationspunkten, die y-Achse 24.2%. Die gesamte Punktekonfiguration ist in zwei Dimensionen verlustfrei darstellbar.

Die doppelte Zentrierung der Daten bringt es mit sich, dass

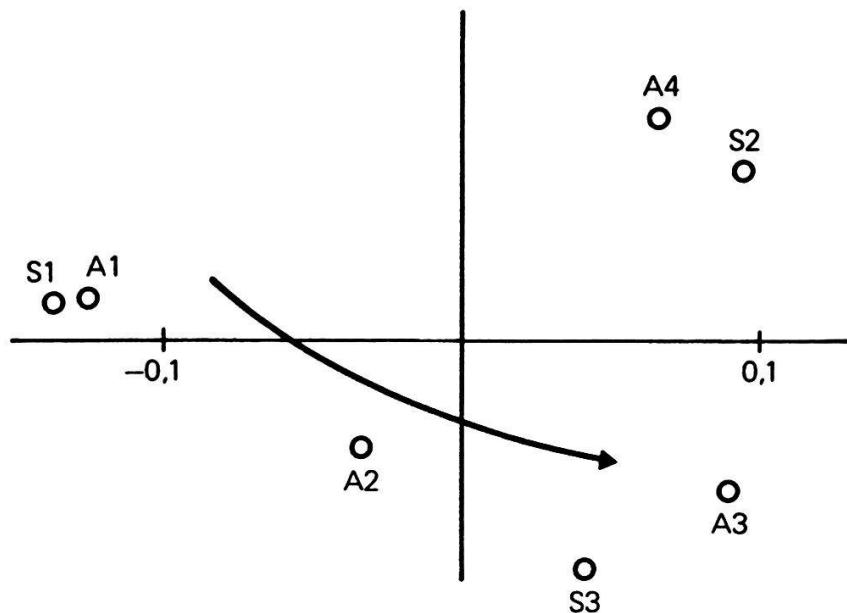


Abb. 7.4 Ordination als Resultat einer Korrespondenzanalyse.
Mit A bezeichnete Punkte stehen für Aufnahmen, mit S bezeichnete für Arten.

die Ähnlichkeitsverhältnisse, mit denen die Korrespondenzanalyse operiert, deutlich von denjenigen der Rohdaten abweichen. Ein durchschnittlicher floristischer Gradient etwa präsentiert sich als Ordination nach einer Hauptkomponentenanalyse meist schlaufen- oder (dreidimensional) schraubenförmig (vgl. Abb. 7.3). Aus der Korrespondenzanalyse resultieren dagegen sanft bogenförmige Punktwolken, weshalb auch von einem Hufeiseneffekt gesprochen wird. HILL (1979a) entwickelte mit seiner Detrended Correspondence Analysis (DCA) ein Verfahren, das leicht gekrümmte Strukturen auf eine Achse projiziert. In vielen Fällen erzielt er damit eine lineare Darstellung eines Vegetationsgradienten.

7.6 Varianten der Hauptkomponentenanalyse

Mit dem bisher Behandelten sind die Hauptkomponentenverfahren bei weitem nicht erschöpft. Die Vorschläge für Änderungen lassen sich unterteilen nach dem Analyseschritt, in welchem sie von der Hauptkomponentenanalyse abweichen:

Eine erste Gruppe von Verfahren bedient sich ausschliesslich einer besonderen Transformation der Rohdaten. Dem praktisch arbeitenden Pflanzenökologen bietet sich damit ein Spektrum leicht durchschaubarer, echter Alternativen. Wir werden im Folgenden zwei typische Beispiele genauer untersuchen.

Eine zweite Gruppe setzt bei den Ähnlichkeitsmassen an. Statt Skalarprodukt, Varianz oder Korrelationskoeffizient wird z.B. ein halb- oder nichtmetrisches Mass verwendet. Wir betrachten im folgenden den Effekt von VAN DER MAARELS (1979) similarity ratio. GOWER (1966, 1967) zeigt dagegen einen Weg, wie ein beliebiges, nicht metrisches Mass für eine metrisch einwandfreie Ordination verwendet wird. Seine "Principle Coordinate Analysis" ist z.B. bei PIELOU (1977) im Detail beschrieben. Kennt man dagegen nur die Vegetationstypen der Aufnahmen und will ohne Berücksichtigung des Artenspektrums eine Ordination herleiten, so bietet sich die "Principle Axis Analysis" an (GOWER 1966, ORLOCI 1978).

Eine dritte Gruppe schliesslich versucht, die Nachteile der Linearität gewöhnlicher Hauptkomponentenanalysen zu umgehen. KRUSKAL (1964) erlaubt bei seiner "nichtmetrischen multidimensionalen Skalierung" die Verwendung beliebiger Ähnlichkeitsmasse und die Ordination versucht bloss, die relativen Positionen der Aufnahmen zu reproduzieren. Die Methode soll hier nicht weiter diskutiert werden.

In Anbetracht der Besonderheiten vegetationskundlicher Daten lohnt es sich, den Einflussmöglichkeiten mittels verschiedener Transformationen besondere Beachtung zu schenken. Bei der Hauptkomponentenanalyse sehen wir, dass Ordinationen typischer Vegetationsgradienten zu schlaufen-, kreis- oder spiralförmigen Strukturen führen können (Abb. 7.5 A). Dies ändert die früher besprochene Einheitslängentransformation der Aufnahmen. Ausgehend von den Rohdaten bestimmen wir die euklidische Länge der Aufnahmeverektoren:

Art i	Aufnahme j					
	1	2	3	4	5	6
1	1	2	2.5	2.5	1	0.5
2	0	1	2	4	3	1
$l_j = [\sum a_{ij}^2]^{1/2}$	1	2.24	3.20	4.72	3.16	1.12

Durch Division jeder Artmächtigkeit mit der Länge des Aufnahmeverektors erhalten wir die transformierten Werte a' :

$$a'_{ij} = \frac{a_{ij}}{l_j}$$

Dies ergibt die neue Tabelle:

Art	Aufnahme					
	1	2	3	4	5	6
1	1	0.89	0.78	0.53	0.32	0.45
2	0	0.45	0.625	0.85	0.95	0.89

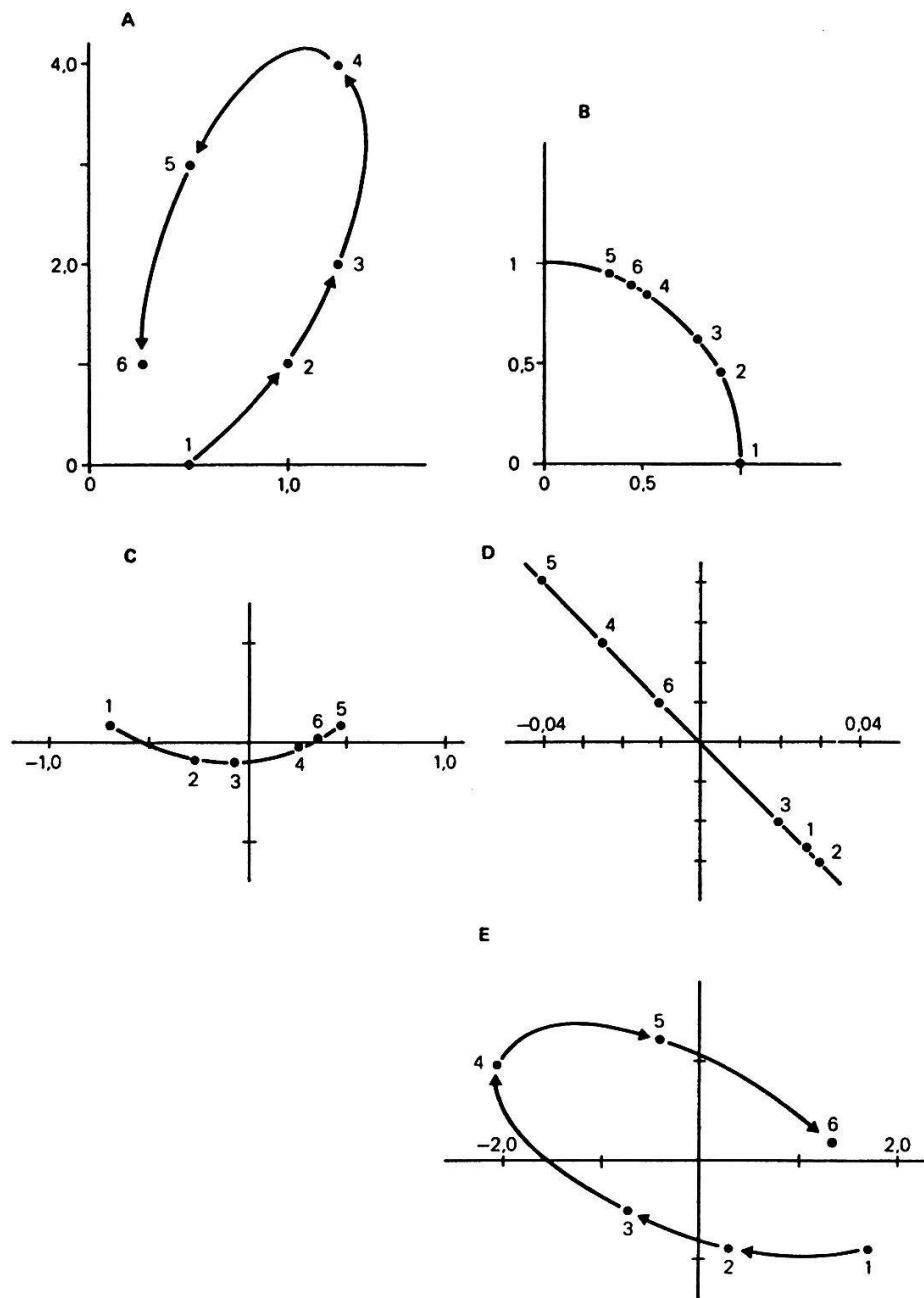


Abb. 7.5 Ordination als Resultat eines verschieden transformierten Datensatzes. A: Rohdaten. B: Aufnahmen auf Einheitslänge transformiert. C: Hauptkomponentenanalyse der auf Einheitslänge transformierten Aufnahmen. D: Transformation für eine Korrespondenzanalyse. E: Hauptkomponentenanalyse mit van der Maarels similarity ratio.

Aus diesen Werten ist die in Abb. 7.5, B gezeigte Ordination konstruiert. Die Einheitslängentransformation hat alle Aufnahmen auf ein Kreissegment projiziert! Die Ähnlichkeitsverhältnisse haben sich gegenüber den Rohdaten deutlich verändert. In der neuen Ordination sind nur noch die relativen Anteile der Arten einer Aufnahme für Vergleiche massgebend. Die Aufnahmen 4, 5 und 6 sind sich deshalb sehr nahe gekommen - sie besitzen alle hohe Werte bezüglich Art 2 und niedrige Werte bezüglich Art 1.

Aus Abbildung 7.5, A wird erneut klar, dass sich die Rohdaten nicht sinnvoll auf nur einer linearen Ordinationsachse darstellen lassen. Mit Abb. 7.5, B kommen wir einer solchen Möglichkeit schon näher. Unterziehen wir deshalb die transformierten Daten einer Hauptkomponentenanalyse. Wir erhalten

Faktor	Aufnahme					
	1	2	3	4	5	6
1	-0.706	-0.283	-0.073	0.256	0.463	0.342
2	0.098	-0.081	-0.095	-0.025	0.088	0.015

Das Resultat ist in Abb. 7.5, C als Ordination dargestellt. Der Kreisbogen ist nun so gedreht, dass er in Richtung der ersten Achse verläuft. Die zweite Achse ist für die Differenzierung der Aufnahmen kaum von Bedeutung. Diese Tatsache lässt sich auch an den Eigenwerten ablesen. Man findet nämlich

$$\begin{aligned}\lambda_1 &= 0.981 = 96.7\% \\ \lambda_2 &= 0.034 = 3.3\%\end{aligned}$$

Eine je nach Datenstruktur noch tiefergreifendere Transformation haben wir bereits kennengelernt. Es sind die Abweichungen von den Erwartungswerten, welche in die Korrespondenzanalyse eingehen. Wie dort gezeigt, beginnen wir mit der Bildung der Randsummen der Aufnahmen und der Arten:

Art	Aufnahme						Σ Arten
	1	2	3	4	5	6	
1	1	2	2.5	2.5	1	0.5	9.5
2	0	1	2	4	3	1	11
Σ Aufn.	1	3	4.5	6.5	4	1.5	$\Sigma = 20.5$

Wir vollziehen nun alle bei der Korrespondenzanalyse gezeigten Operationen und erhalten

Art	Aufnahme					
	1	2	3	4	5	6
1	0.026	0.030	0.020	-0.025	-0.041	-0.010
2	-0.026	-0.030	-0.020	0.025	0.041	0.010

Abb. 7.5, D zeigt die entsprechende Ordination. Die Transformation hat die Aufnahmen auf eine Gerade projiziert! Generell reduziert nämlich die Korrespondenzanalyse die Dimensionalität um 1. Im Kapitel 7.5 musste aus diesem Grunde ein dreidimensionales Beispiel gewählt werden, sodass als Resultat zumindest noch eine zweidimensionale Ordination möglich war. Zu beachten ist die Reihenfolge der Aufnahmen auf der Ordinationsgerade. Sie dokumentiert, dass sich die Aehnlichkeitsverhältnisse grundlegend von denjenigen in den Rohdaten unterscheiden.

Im Kapitel Aehnlichkeitsmasse wurde gezeigt, dass jene öfters versteckt Transformationen beinhaltet. Verwenden wir z.B. van der Maarels Koeffizient anstelle des Skalarproduktes, so werden die Daten bezüglich ihrer Aehnlichkeit anders beurteilt. Der Versuch ist in Abb. 7.5, E dargestellt. Wie leicht zu sehen ist, bleiben hier die Aehnlichkeitsverhältnisse der Rohdaten erhalten, doch liegen die Achsen anders als bei der normalen Hauptkomponentenanalyse. Van der Maarels similarity ratio ist ja kein metrisches Aehnlichkeitsmass. Die Hauptkomponentenanalyse interpretiert es jedoch metrisch. Man führt also formal

gesehen einen nicht kontrollierbaren, die Analyse meist nicht weiter störenden Fehler ein, der zu einem anderen Resultat führt. Ob ein solches Vorgehen akzeptabel ist, bleibe dem Urteil des Anwenders überlassen.

8. Analyse geordneter Vegetationstabellen

Ein erstes Ziel der Analyse von Vegetationsdaten ist oft die Bildung von Aufnahmen- und Artengruppen. Wir haben früher gesehen, dass der Weg dazu über numerische Gruppierungsanalysen oder über eine subjektive Klassifikation führen kann. Art und Qualität des Resultates hängen in jedem Falle von der Wahl der Methode ab. Der eigentliche Sinn geordneter Vegetationstabellen liegt darin, einerseits Einblick in die Fülle der Rohdaten zu gewinnen, andererseits den Erfolg des Auswertemodells unmittelbar vor Augen zu haben. Dabei ist zu beachten, dass eine rechteckige Tabelle nur beschränkt geeignet ist, komplizierte Aehnlichkeitsstrukturen wiederzugeben. So widerspricht denn die Annahme, Arten müssten sich zu diskreten Gruppen zusammenfassen lassen, den gängigen Vorstellungen über Selektionsmechanismen. Zudem ist zu beachten, dass auch der geübte Fachspezialist kaum beurteilen kann, ob eine Tabelle wirklich optimal ausgewertet ist. Der Wunsch liegt daher nahe, die Qualität einer Lösung irgendwie zu messen. Mit Hilfe eines geeigneten Verfahrens ist man zumindest imstande, verschiedene Gruppierungslösungen desselben Datensatzes miteinander zu vergleichen. Von den Gruppierungsanalysen wissen wir, dass jede formal abgeleitete Klassifikation optimal ist - aber eben nur bezüglich der gewählten Analysevorschrift. Es ist deshalb nötig, für die Güte einer Vegetationstabelle ein einheitliches Beurteilungskriterium zu finden. Eine Möglichkeit schufen FEOLI und ORLOCI (1979) mit der Konzentrationsanalyse.

8.1 Die Konzentrationsanalyse

Die Konzentrationsanalyse basiert auf der Präsenz oder Ab-
senz der Arten. Eine Artengruppe gilt dann als perfekt, wenn sie eine Aufnahmegruppe eindeutig zu identifizieren erlaubt. Umgekehrt ist eine Aufnahmegruppe dann ideal zusammengesetzt, wenn sie eine Gruppe eindeutiger Differentialarten besitzt. In Tab. 8.1 ist dieses Prinzip illustriert. Jede Tabelle (A bis C) umfasst vier Aufnahmen und vier Arten, die in je zwei Gruppen unterteilt sind. Es sind damit

Tabelle 8.1

Vegetationstabellen maximaler (A), mittlerer (B) und minimaler Konzentration (C).

A

Arten	Aufnahmen	
	Gr 1	Gr 2
Gr 1	1 + 2 1	
Gr 2		1 + 3 4

B

Arten	Aufnahmen	
	Gr 1	Gr 2
Gr 1	1 + 1 2	+
Gr 2		1 1 2

C

Arten	Aufnahmen	
	Gr 1	Gr 2
Gr 1	2 +	+
		1
Gr 2	1	4 2
		3

vier "Zellen" entstanden. Tabelle 8.1, A, ist optimal strukturiert. Aufgrund des Auftretens einer Art kann eindeutig auf die Aufnahmegruppe geschlossen werden. Alle von null verschiedenen Abundanzwerte konzentrieren sich auf zwei Zellen (daher auch der Begriff "Konzentrationsanalyse"). Tabelle 8.1, B, zeigt einen weniger günstigen Fall. Die beiden Zellen der Aufnahmegruppe 1 weisen maximale Konzentration auf, diejenigen der Gruppe 2 nicht ganz. Der ungünstigste Fall liegt bei Tabelle C vor. Hier sind die Abundanzen bezüglich aller Zellen völlig diffus verteilt, die Konzentration ist gleich null.

In der Praxis kommt es selten vor, dass alle Gruppen aus gleich vielen Individuen zusammengesetzt sind. Ob eine Gruppe gross oder klein ist, sagt jedoch normalerweise wenig aus über deren Bedeutung. In der Konzentrationsanalyse wird allen Art- und Aufnahmegruppen, also allen Zellen der Vegetationstabelle, gleiches Gewicht beigemessen. Wir verfolgen die Analyse anhand eines etwas komplizierteren Beispiels, das wir WILDI und ORLOCI (1983) entnehmen (Tab. 8.2). Es handelt sich um eine Vegetationstabelle von 10 Aufnahmen mit 19 Arten, welche einen Gradienten innerhalb eines Buchen-Laubmischwaldes beschreiben. Die Gruppenstruktur beruht auf einer Minimalvarianz-Analyse.

Im ersten Schritt der Konzentrationsanalyse wird eine Kontingenztafel hergestellt. Zu diesem Zwecke zählt man alle von null verschiedenen Abundanzwerte f_{ij} in den neun Zellen der Vegetationstabelle und erhält

Artengruppe	Aufnahmegruppe			Σ	
	1	2	3		
F =	1	4	9	26	39
	2	20	18	12	50
	3	14	0	0	14
Σ	38	27	38	103	

Nun werden alle Zellengrößen bestimmt. Die erste Ar-

Tabelle 8.2

Geordnete Vegetationstabelle (aus WILDI und ORLOCI 1983).

Releve Group no.	1	1	1	1	2	2	2	3	3	3
Releve no.	4		2	1	3	2		1		
	9	2	6	5	8	9	9	7	4	0
pH	4	4	4	5	6	5	5	6	6	6

Altitude (m a. s. l.)	8	4	8	0	0	6	2	0	2	5
	5	4	4	4	4	5	5	4	5	5
	5	5	2	8	0	8	0	5	0	6
	0	0	0	0	0	0	0	0	0	0
Slope, deg.	1	1	1	1	1	1	1	1	1	1
	5	0	4	8	2	6	2	2	0	4

	0	0	5	0	5	0	5	5	5	0
Ranunculus ficaria	1					4	1	2		
Primula elatior	1					2	+	2		
Allium ursinum	1					+	4	2		
Arum maculatum	1					+	1			
Fraxinus excelsior	1				+	+	2	3	3	
Acer pseudoplatanus	1				+	1	+	4	1	2
Sambucus nigra	1				+		1	+	2	
Lamium galeobdolon	1	+	+			+	+	2	2	1
Lonicera xylosteum	1	+		+			+	1	1	+
Eurhynchium striatum	2	+	1		+	+	1	+	+	+
Galium odoratum	2	+	+		2	1	+	1	+	+
Carex sylvatica	2	+	+	1	2	+	+	2	+	
Oxalis acetosella	2	+	+	+	1	1	3	1	2	
Fagus sylvatica	2	2	3	4	4	5	4	5	1	+
Quercus petraea	2	4	3	2	3	+	1	+		
Vaccinium myrtillus	3	+	2	1						
Sambucus racemosa	3	+	2	1						
Polytrichum formosum	3	3	+	1	3					
Luzula nemorosa	3	2	2	+	1					

tengruppe besteht aus 9 Arten, die erste Aufnahmegruppe aus 4 Aufnahmen. Die erste Zelle hat damit eine Felderzahl von 36. Für die gesamte Tabelle ergibt sich

$$z = \begin{matrix} 36 & 27 & 27 \\ 24 & 18 & 18 \\ 16 & 12 & 12 \end{matrix}$$

Die Summe aller Elemente der Tabelle beträgt 190, die mittlere Zellengröße \bar{z} ist gleich 21.1. Sodann gilt es, die Inhalte der Kontingenztabelle F der mittleren Zellengröße anzupassen. Die Elemente der neuen Matrix A sind korrigierte f_{ij} -Werte:

$$a_{ij} = f_{ij} \frac{\bar{z}}{z_{ij}}$$

Demzufolge erhält man für die erste korrigierte Zelle

$$a_{11} = 4 \frac{21.1}{36} = 2.34$$

Die gesamte, auf mittlere Zellengröße gebrachte Kontingenztafel ergibt

Artgruppe	Aufnahmegruppe			Σ	
	1	2	3		
A =	1	2.34	7.03	20.3	29.67
	2	17.58	21.1	14.06	52.74
	3	18.46	0	0	18.46
Σ		38.38	28.13	34.36	100.87

Weil im vorliegenden Beispiel die Zellengrößen nicht allzu unterschiedlich sind, bewegen sich die a_{ij} -Werte in den Größenordnungen der ursprünglichen f_{ij} -Werte.

A ist also formal betrachtet noch immer eine gewöhnliche Kontingenztafel, allerdings mit gebrochenen Zahlen. Die Konzentrationsanalyse lehnt sich von nun an eng an die Korrespondenzanalyse an (vgl. Kap. 7.5). Gearbeitet wird also nicht mit den a_{ij} -Werten, sondern mit Abweichungen von Erwartungswerten. Wie bei der Korrespondenzanalyse besprochen, muss dazu einerseits mit relativen Häufigkeiten gerechnet werden, andererseits sind die Erwartungswerte irgendwie zu definieren. Für beide Schritte verfolgt die Konzentrationsanalyse ein eigenes Konzept. Zur Berechnung der relativen Häufigkeiten werden die a_{ij} -Werte nicht wie bei der Korrespondenzanalyse am gesamten Tabellentotal $a_{..}$ gemessen, sondern an der maximal möglichen Frequenz innerhalb jeder Zelle i,j . Betrachten wir das erste Element a_{11} genauer. Wären alle Artvorkommen der Artengruppe 1 auf Zelle 1,1 konzentriert, so ergäbe sich eine Häufigkeit von 29.67. Konzentriert man andererseits alle Vorkommen der Aufnahmegruppe 1 auf diese Zelle, so erhält man 38.38. Als möglichen Maximalwert bedient man sich des geometrischen Mittels und gewichtet a_{11} mit dessen Hilfe wie folgt:

$$b_{11} = 2.34 / [(29.67)(38.38)]^{1/2} = 0.069$$

Verwenden wir für die Zeilensummen die Symbolik a_i und für die Spaltensummen a_j , so lautet diese Berechnungsformel

$$b_{ij} = a_{ij} / (a_i \cdot a_j)^{1/2}$$

Für die gesamte Matrix erhält man

$$B = \begin{matrix} 0.069 & 0.243 & 0.636 \\ 0.391 & 0.548 & 0.330 \\ 0.694 & 0.0 & 0.0 \end{matrix}$$

Als Erwartungswerte werden die schon besprochenen maximal möglichen Häufigkeiten jeder Zelle genommen und mit Hilfe einer Division durch das Gesamttotal $a_{..}$ der Skalierung von Matrix B angepasst.

$$e_{ij} = \frac{(a_{i.} a_{.j})^{1/2}}{a_{..}}$$

Für das erste Element erhält man

$$e_{11} = \frac{[(29.67)(38.38)]^{1/2}}{100.87} = 0.334$$

Die Matrix der Erwartungswerte ergibt

$$E = \begin{matrix} 0.334 & 0.286 & 0.316 \\ 0.446 & 0.382 & 0.422 \\ 0.263 & 0.226 & 0.250 \end{matrix}$$

Durch einfache Subtraktion erhält man die Abweichungen von den Erwartungswerten:

$$B - E = \begin{matrix} -0.265 & -0.043 & 0.320 \\ -0.055 & 0.166 & -0.092 \\ 0.431 & -0.226 & -0.250 \end{matrix}$$

Nun folgt eine Eigenwertanalyse (vgl. Kap. 7) mit dem Ziel, sowohl für Arten- als auch für Aufnahmegruppen Ordinationskoordinaten zu erhalten. Man benötigt dazu eine Ähnlichkeitsmatrix (der Artengruppen oder der Aufnahmegruppen). Berechnet man das Skalarprodukt zwischen den Zeilen von $B - E$ so erhält man

$$S = \begin{matrix} 0.174 & -0.0218 & -0.1839 \\ -0.0218 & 0.0390 & -0.0383 \\ -0.1839 & -0.0383 & 0.2980 \end{matrix}$$

Die Eigenwerte dieser Matrix - mit einem Computerprogramm berechnet - betragen

$$\begin{aligned}\lambda_1 &= 0.4309 \\ \lambda_2 &= 0.0801 \\ \lambda_3 &= 0\end{aligned}$$

Wie bei der Hauptkomponentenanalyse gezeigt, werden daraus die Eigenvektoren α berechnet. Sie sind wiederum zu normalisieren und man erhält

$$\alpha = \begin{matrix} -0.5788 & 0.6089 & 0.5425 \\ -0.474 & -0.6892 & 0.7230 \\ 0.841 & 0.3928 & 0.4277 \end{matrix}$$

Die erste Spalte dieser Matrix bezieht sich auf die erste noch zu berechnende Ordinationsachse, die zweite auf die zweite Achse usw. Wie schon bei der Hauptkomponentenanalyse besprochen, enthalten die Eigenvektoren die Korrelationen zwischen den ursprünglichen Daten und den neuen Ordinationskoordinaten: So besagt das Element unten links in α , dass Artengruppe 3 mit 0.841 eine erhebliche Korrelation zur ersten Ordinationsachse aufweisen wird.

Nun geht es noch darum, die Eigenvektoren so zu transformieren, dass sie auch als Ordinationskoordinaten für Artengruppen verwendet werden können. Dies geschieht so, dass die anschliessend zu berechnenden Koordinaten der Aufnahmegruppen gleich skaliert sind. Das Prinzip ist bereits aus der Korrespondenzanalyse (Kap. 7.5) bekannt, nur haben FEOLI und ORLOCI (1979) eine etwas andersartige Gewichtung der α -Werte gewählt. Für die Artengruppen ergeben sich die Koordinatenwerte wie folgt:

$$x_{ij} = \alpha_{ij} \frac{a_i}{\sqrt{a_i}}$$

Für den ersten Koordinatenwert ergibt dies

$$x_{11} = \alpha_{11} \frac{a_{11}^{1/2}}{a_{11}} = -0.5788 \frac{100.87^{1/2}}{29.67} = -1.0669$$

Die Werte der ganzen Matrix betragen

$$X = \begin{bmatrix} -1.0669 & 1.1224 & 0.0 \\ -0.0655 & -0.9532 & 0.0 \\ 1.9033 & 0.9183 & 0.0 \end{bmatrix}$$

Die Punktekonfiguration ist also in zwei Dimensionen darstellbar.

Die Koordinaten der Aufnahmegruppen ergeben sich als Linear-kombination der Matrix X und der Angangsdaten A. Sie werden mit der zugehörigen Spaltensumme a_j und - wie beim D-Algorithmus der Hauptkomponentenanalyse - mit der Wurzel aus dem zugehörigen Eigenwert λ_j gewichtet:

$$y_{ij} = [\sum_i a_{ij} x_{ij}] / a_{.j} \lambda_j^{1/2}$$

Für die erste Koordinate erhält man

$$y_{11} = [(2.34)(-1.0669) + (17.58)(-0.0655) + (18.46)(1.9032)] / 38.38 * 0.4309^{1/2} = 1.249$$

Für alle Elemente ergibt sich

$$Y = \begin{bmatrix} 1.249 & 0.260 & 0.0 \\ -0.481 & -1.535 & 0.0 \\ -1.001 & 0.966 & 0.0 \end{bmatrix}$$

Die Interpretation erfolgt genau gleich wie bei der Korrespondenzanalyse. Die Lage der Aufnahme- und Artengruppen in der Ordination (Abb. 8.1) widerspiegelt in jeder Hinsicht die Beziehungen der Gruppe in der Vegetationstabelle. Für Aufnahmegruppe 3 ist die Artengruppe 1 ein guter Indikator, für Aufnahmegruppe 2 Artengruppe 2 und

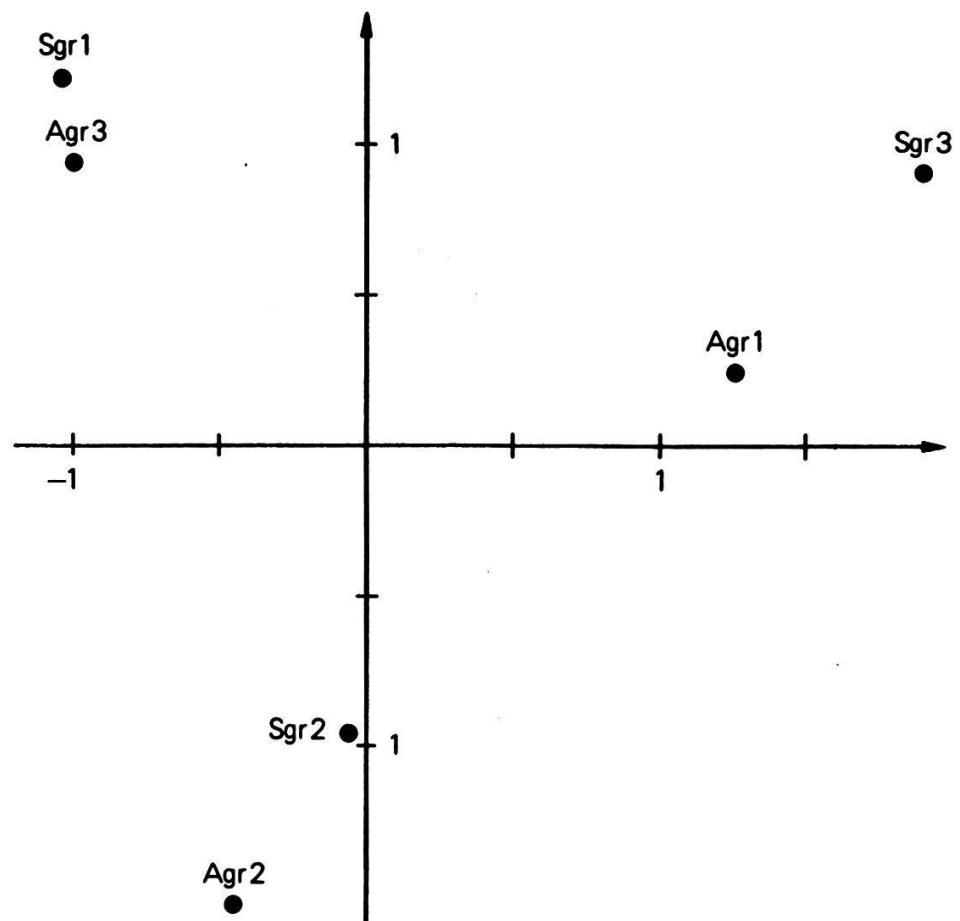


Abb. 8.1 Ordination der Aufnahmen- und Artengruppen (Tab. 8.2) mit den Koordinaten der Konzentrationsanalyse. Agr sind Gruppen von Aufnahmen, Sgr sind Gruppen von Arten.

für Aufnahmegruppe 1 Artengruppe 3. Dabei besteht die engste Beziehung zwischen Aufnahmegruppe 3 und Artengruppe 1. Dies entspricht auch der gefühlsmässigen Interpretation der Daten in Tab. 8.2.

Für die Güte des Zusammenhangs zwischen Arten- und Aufnahmegruppen kann auch ein Messwert definiert werden. Wir erinnern uns, dass die Konzentrationsanalyse auf Abweichungen von Erwartungswerten beruht. Letztere geben die Verhältnisse in einer völlig unstrukturierten Tabelle wieder. Je grösser also die Abweichungen gesamhaft ausfallen, desto klarer ist die Gruppenstruktur. Die Berechnung ergibt keinerlei Probleme, denn die relativen Abweichungen sind in der Matrix B - E schon vorhanden. Deren Quadratsumme entspricht einem Chiquadrat. Dank der Multiplikation mit der Gesamtsumme a.. bezieht sich Chiquadrat auf die absoluten Frequenzen der Tabelle:

$$\begin{aligned}\chi^2 &= a \cdot \sum_{i \cdot j} (B - E)_{ij}^2 \\ &= 100.87 ((-0.265)^2 + (-0.043)^2 + (0.320)^2 \\ &\quad + (-0.055)^2 + (0.166)^2 + (-0.092)^2 \\ &\quad + (0.43)^2 + (-0.226)^2 + (-0.25)^2) \\ &= 51.59\end{aligned}$$

Das Chiquadrat ist natürlich von der Grösse der Tabelle abhängig. Man kann es auf den Bereich 0 bis 1 skalieren, indem man den mittleren quadratischen Kontingenzkoeffizienten C berechnet:

$$C = \chi^2 / [a \cdot (q-1)]$$

Darin ist q die Dimension der Kontingenztabelle, d.h. der kleinere Wert aus der Anzahl der Artengruppen und der Anzahl der Aufnahmegruppen. Wir erhalten für unsere Tabelle

$$C = 51.59 / (100.87)(2.0) = 0.256$$

Das relativierte Mass C kann nun als Vergleichswert für andere Klassifikationslösungen der selben oder auch anderer

Vegetationstabellen verwendet werden.

Die hier erzielte Konzentration von 0.256 weist erfahrungsgemäss auf eine wohlstrukturierte Tabelle hin. Werte zwischen 0.1 und 0.2 sind die Regel. Je umfassender eine Vegetationstabelle ist, desto kleiner wird der Kontingenzkoeffizient; die fehlende Linearität macht sich bemerkbar.

Die Konzentrationsanalyse liefert mit dem mittleren quadratischen Kontingenzkoeffizienten ein Ordnungsmass, das von der Grösse der Tabelle unabhängig ist. Nur erklärt er natürlich nicht, ob eine schwach ausgeprägte Gruppenstruktur eine Eigenheit der Daten ist oder ob bloss ein unzweckmässiges Gruppierungsverfahren gewählt worden ist. Wenn verschiedene Klassifikationslösungen desselben Datensatzes verglichen werden, so führt die bessere zu einem höheren Chi-quadrat. Entscheidend für die Güte einer Tabelle ist aber auch die Zahl der gebildeten Gruppen. Tab. 8.3 zeigt, dass mit einer feineren Gliederung gelegentlich eine bessere Konzentration erreicht wird. Die Bildung zu vieler Gruppen ist jedoch mit der Zielsetzung der Untersuchung öfters unvereinbar. Der Kontingenzkoeffizient allein genügt daher für einen wertenden Vergleich in der Regel nicht. Eine vorgegebene Anzahl Vegetationseinheiten kann aus praktischen Erwägungen wichtiger sein als eine optimale Konzentration der Abundanzwerte. Wir erörtern nachfolgend die Frage nach der Güte der Ordnung einer Tabelle nicht weiter.

8.2 Vergleich zweier Vegetationstabellen

Oft stellt sich die Frage, welches die Beziehungen der Aufnahmegruppen verschieden geordneter Tabellen zueinander sind. Ein Problem dieser Art tritt immer dann auf, wenn eine pflanzensoziologische Klassifikationen revidiert wird. Wir verfolgen eine Möglichkeit der Untersuchung anhand des Beispiels in Tab. 8.2. Nehmen wir der Einfachheit halber an, dass die Abfolge der Aufnahmen gleich bleibt, dass aber neu vier anstelle von drei Gruppen gebildet werden. Die alte Klassifikation lautet:

Tabelle 8.3

Die Unterteilung dieser Tabelle in drei Aufnahmegruppen ergibt eine bessere Konzentration als eine solche in zwei Gruppen.

Art	Aufnahme					
	1	2	3	4	5	6
1	1	1				
2	1	1				
3			1	1	1	1
4			1	1	1	1
5					1	1

Gruppe 1	Gruppe 2	Gruppe 3
(49 2 6 25)	(18 9 39)	(27 4 19)

Neu finden wir:

Gruppe 1	Gruppe 2	Gruppe 3	Gruppe 4
(49 2)	(6 25 18)	(9 79)	(27 4 10)

Nun stellen wir die alten und neuen Zuordnungsverhältnisse in einer Kontingenztafel (Tab. 8.4, A) zusammen. Zur weiteren Verarbeitung sind nur noch die Frequenzen massgebend (Tab. 8.4, B). Die Tafel kann nun weiter analysiert werden. Die Uebereinstimmung der beiden Klassifikationen soll mit der Konzentrationsanalyse untersucht werden. Damit wird wiederum von der Grösse der Gruppen abgesehen. Möchte man dagegen grossen Gruppen mehr Gewicht verleihen, so könnte zur Korrespondenzanalyse gegriffen werden. Wir führen jedoch genau die im Kapitel 8.1 dargestellten Operationen durch und finden folgende Resultate:

$$\chi^2 = 16.822$$

$$\text{Kontingenzkoeffizient } C = 0.841$$

Ordinationskoordinaten:

Aufnahmegruppen		Achse 1	Achse 2
alt			
1		-0.622	-1.216
2		-0.622	1.140
3		1.607	0.0

Tabelle 8.4

Vergleich zweier Klassifikationslösungen der Vegetationstabelle in Tab. 8.2. Zuordnung der Aufnahmen (A), Kontingenztafel (B).

A

		Alte Tabelle, Gruppe		
Revidierte Tabelle, Gr.		1	2	3
1	49	2		
2	6	25	18	
3			9	39
4				27 4 10

B

		Alte Tabelle, Gruppe			Σ
Revidierte Tabelle, Gr.		1	2	3	
1		2			2
2		2	1		3
3			2		2
4				3	3
Σ		4	3	3	10

Aufnahmegruppen neu	Achse 1	Achse 2
1	-0.622	-1.464
2	-0.622	-0.329
3	-0.622	1.373
4	1.607	0.0

Die Eigenwerte von 1.0, 0.83 und 0.0 weisen darauf hin, dass die Beziehung zwischen den beiden Klassifikationen in zwei Dimensionen darstellbar ist. Die entsprechende Ordination findet sich in Abb. 8.2. Aus ihr lässt sich leicht herauslesen, dass die neue Gruppe 2 in ihrer Zusammensetzung zwischen den alten Gruppen 1 und 2 liegt, dass die neue Gruppe 4 exakt mit der alten Gruppe 3 zusammenfällt, usw.

Der Kontingenzkoeffizient von 0.841 ist sehr hoch. Es ist zu beachten, dass die hier analysierte Datenstruktur ganz anders geartet ist, als wenn einzelne Vegetationstabellen analysiert werden. Werte von 0.5 und mehr sind bei der Untersuchung alternativer Klassifikationen durchaus normal.

Mit der soeben gezeigten Methode sind die Vergleichsmöglichkeiten für Datenstrukturen natürlich noch nicht erschöpft. So berücksichtigt die Konzentrationsanalyse die interne Struktur der Gruppen nicht. Auch Gradienten sind schwierig zu untersuchen. Sind detailliertere Vergleiche wünschenswert, so kann auf die Dendrogramme zurückgegriffen werden. Deren Aufbau lässt sich durch eine sogenannte Topologiematrix beschreiben (SOKAL und ROHLF 1962, SNEATH und SOKAL 1973). Verschiedene Topologiematrizen sind wiederum einem formalen Vergleich zugänglich. Wir verzichten hier auf eine weitergehende Erörterung dieser Methoden.

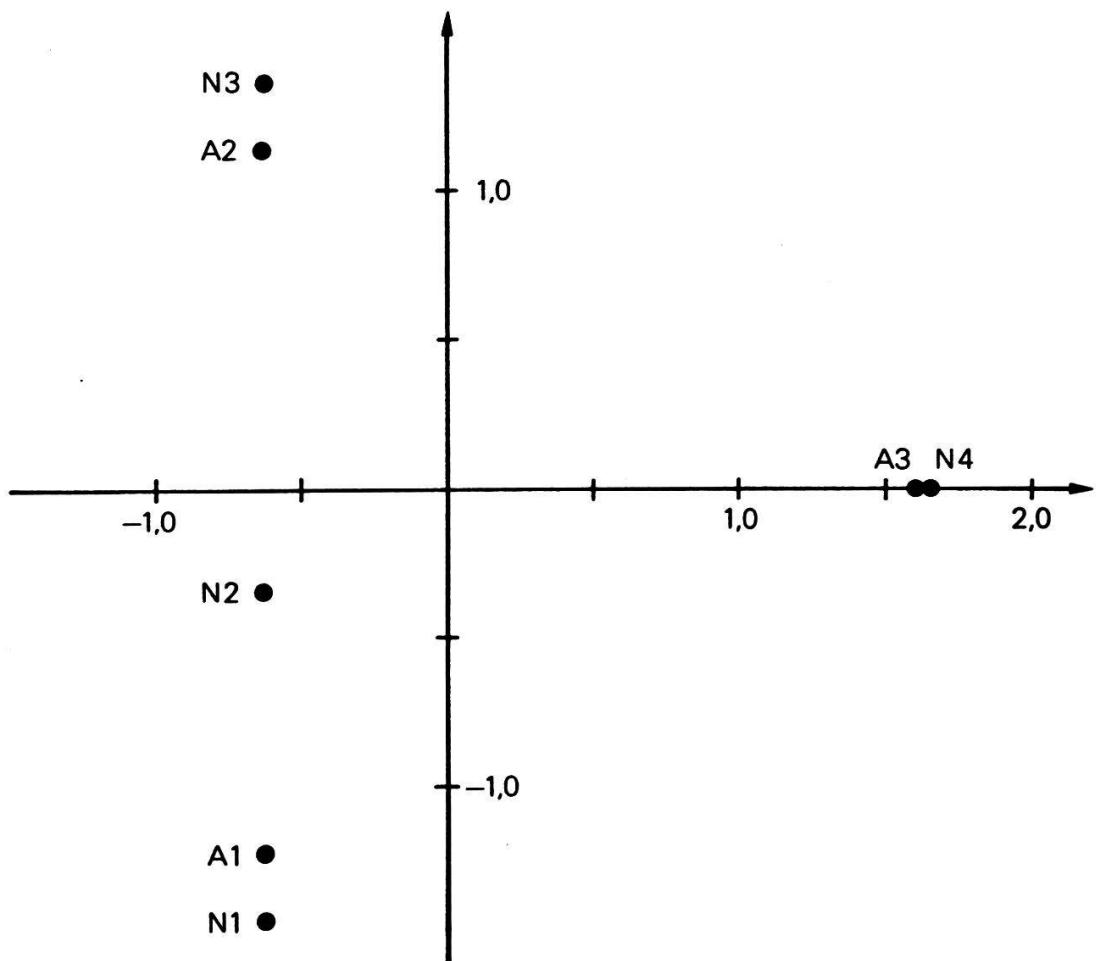


Abb. 8.2 Vergleich alternativer Klassifikationen mit Hilfe der Konzentrationsanalyse. A: Gruppen der alten Klassifikation. N: Gruppen der neuen Klassifikation.

9. Zusammenhang zwischen Vegetation und Standort

9.1 Zielsetzung und Schwierigkeiten

Mit der Analyse des Zusammenhangs zwischen Vegetation und Standort sind wir bei der schwierigsten und vielleicht wichtigsten Aufgabe der Pflanzenökologie angelangt (vgl. z.B. POORE 1955, S. 246). Dabei werden Lösungen gesucht auf Fragen, welche gleichermaßen von wissenschaftlichem und umweltpolitischem Interesse sind. Unsere Umwelt ist heute tiefgreifenden Veränderungen und Störungen ausgesetzt. Eingriffe können direkt auf die Vegetation erfolgen. Entwaldung, Schnitt, Weidegang oder die vollständige Entfernung der Pflanzendecke sind Beispiele. Ebenso häufig aber wird primär der Standort verändert. Grossflächig erfolgt dies bei landwirtschaftlicher Düngung, bei der Eutrophierung der Flüsse und Seen durch Abwässer oder durch saure Niederschläge. In jedem Falle geht die Aufgabe der Pflanzenökologie dahin, Veränderungen und mithin zukünftige Zustände vorauszusagen. Je nach der Art des Eingriffes sollen Aussagen nach folgenden Schemata gemacht werden:

1. Falls Standort A umgewandelt wird in Standort B, so folgt, dass Vegetation C sich umwandelt in Vegetation D.
2. Falls Vegetation C umgewandelt wird in Vegetation D, so folgt, dass Standort A sich umwandelt in Standort B.

Als Spezialfall eines Standortfaktors soll die Zeit mit in die Betrachtungen eingeschlossen sein. Voraussetzung für die postulierten Voraussagen ist die formale, im statistischen Sinne deskriptive Bestimmung des Zusammenhangs zwischen Vegetation und Standort. Natürlich kann man methodisch erheblich weiter gehen und die dynamischen Prozesse numerisch nachzubilden versuchen. Davon soll hier abgesehen werden.

Formal ergibt sich die Aufgabe, die durch mehrere Variablen charakterisierte Vegetation mit mehreren Standortsvariablen

in Beziehung zu setzen. Gegeben sind also zwei multivariate Datensätze. Der Zusammenhang zwischen einem Vektor des einen Satzes (z.B. einer Art) und einem Vektor des andern Satzes (einem Standortsfaktor) kann durch eine normale Korrelationsanalyse gefunden werden. Für die multivariate Fragestellung muss zur Kanonischen Korrelationsanalyse gegriffen werden (PIELOU 1977). Deren Ergebnisse sind nicht sehr einfach zu interpretieren. GITTINS (1985) hat ihr eine umfassende Darstellung mit Anwendungsbeispielen gewidmet. Es bleibt abzuwarten, ob sie sich damit in der Ökologie doch noch durchsetzen kann. Zwei Haupthindernisse erschweren ihren Einsatz in der Pflanzenökologie: Zum einen ist es die grosse Zahl der Variablen (nämlich Arten), die zur Beschreibung der Vegetation nötig sind. Zum andern ist es die meist ausgeprägte Nichtlinearität in der Beziehung zwischen den beiden Datensätzen.

Zur Illustration des Linearitätsproblems verwenden wir die Vegetationsdaten des Kapitels 7.3 und ergänzen diese mit einem Standortsfaktor, pH. Wir erhalten folgende Tabelle:

Aufn.	A	B	C	D	E	F
Art 1	1	2	2.5	2.5	1	0.5
Art 2	0	1	2	4	3	1
pH	4.0	4.2	4.3	4.8	5.2	5.7

Es wäre hier naheliegend, mittels der einfachen Korrelationsanalyse den Zusammenhang der Arten 1 und 2 mit dem pH zu untersuchen. In Abb. 9.1, A, sind die pH-Werte in Funktion der Abundanzwerte von Art 1 aufgetragen. Es ist offensichtlich, dass eine lineare Regression versagen muss: Sowohl bei pH 4.0 als auch beim Höchstwert von 5.7 erreicht Art 1 ein Minimum. Damit lässt sich aus der Häufigkeit ihres Auftretens innerhalb einer Probefläche nicht eindeutig ein pH-Wert voraussagen. Außerdem ist es wünschenswert, nicht eine einzige, sondern mehrere Arten gleichzeitig zu berücksichtigen. Man könnte also statt einer einzelnen Art die Faktorenwerte einer Komponentenanalyse mit dem pH

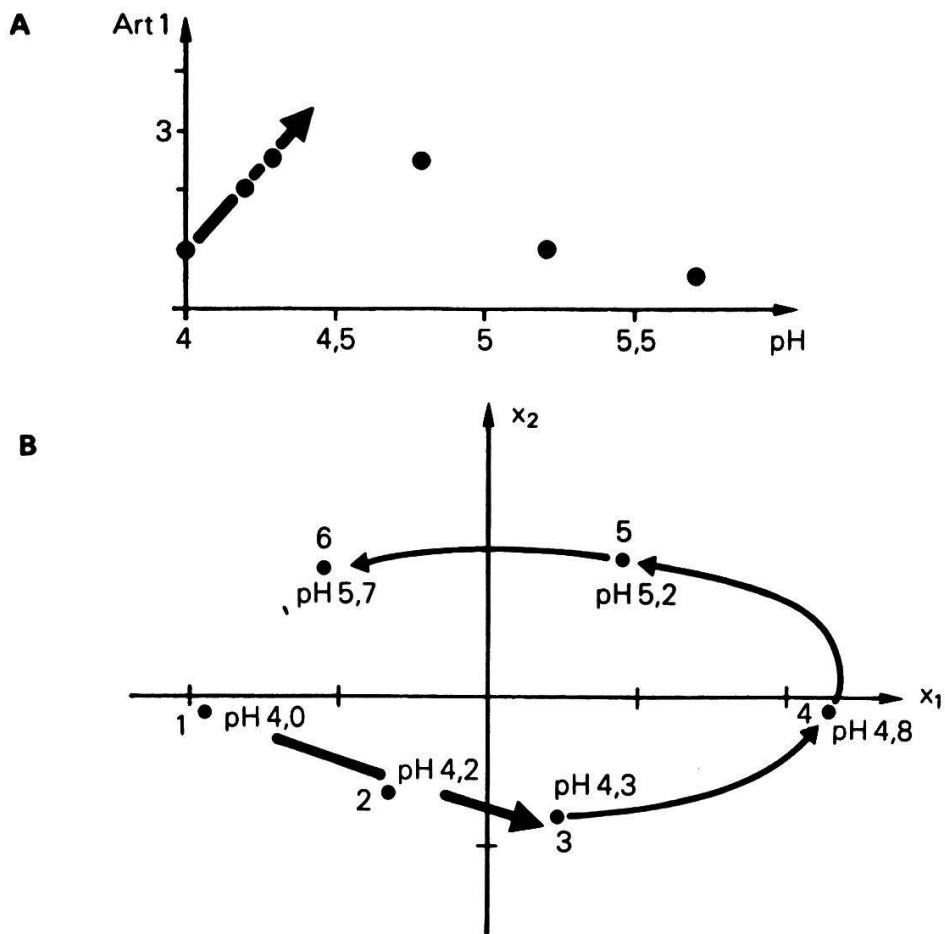


Abb. 9.1 Bestimmung der Beziehung zwischen dem Standortsfaktor pH und der Abundanz einer Pflanzenart (A), bzw. den Achsen einer Ordination (B). Fette Pfeile zeigen Bereiche, innerhalb welcher die Zusammenhänge annähernd linear sind.

korrelieren. Ein solcher Versuch ist in Abb. 9.1, B, dargestellt. Doch die nichtlineare Ähnlichkeitsstruktur der Aufnahmen lässt die Regressionsanalyse auch hier scheitern. Bei den ersten Punkten des Gradienten steigt der pH von links nach rechts, in der zweiten Hälfte von rechts nach links an. Die fett ausgezogenen Bereiche in Abb. 9.1, A und B weisen auf eine mögliche Lösung hin: Wird ausschliesslich in einem engen ökologischen Bereich gearbeitet, so stellt die lineare Regression tatsächlich eine brauchbare Näherungslösung dar. Leider tritt dieser Fall in der Praxis eher selten auf.

Günstiger stehen die Erfolgsaussichten einer Analyse, wenn die Daten bereits gruppiert sind. Es braucht blos geprüft zu werden, ob die in der Vegetation gefundenen Gruppen sich auch standörtlich unterscheiden. Doch auch hier sind die Resultate kritisch zu interpretieren. Analysiert wird ja nicht die Struktur der Vegetation selbst, sondern das Resultat einer Gruppierungsanalyse. Dieses kann wiederum unterschiedlich treffend sein, je nach gewählter Methode.

Ziel einer Gruppierungsanalyse ist das Erkennen von Vegetationseinheiten, welche Umweltfaktoren möglichst genau vorauszusagen erlauben. Man wird also stets versuchen, eine geeignete Gruppierungsmethode zu finden (Kap. 5). Es kann aber sein, dass eine gute Klassifikationslösung unerwartet schlecht mit dem Standort übereinstimmt. In Abb. 9.2 ist ein solcher Fall dargestellt. Die beiden strukturell sauber getrennten Gruppen überlappen bezüglich pH deutlich. Eine solche Situation wurde von FEOLI und FEOLI-CHIAPELLA (1980) an reellen Daten nachgewiesen. Wir ersehen daraus, dass immer auch die Kombination der angewandten Methoden, nicht nur die einzelne Methode für sich kritisch auszuwählen ist.

9.2 Grafische Lösungen

Nicht immer muss auf eine rechnerische Lösung zurückgegriffen werden. Man kann ein Strukturmodell (Ordination, Dendrogramm) anhand standörtlicher Messungen direkt interpretieren. Im Prinzip wird in Abb. 9.1 so vorgegangen. Die

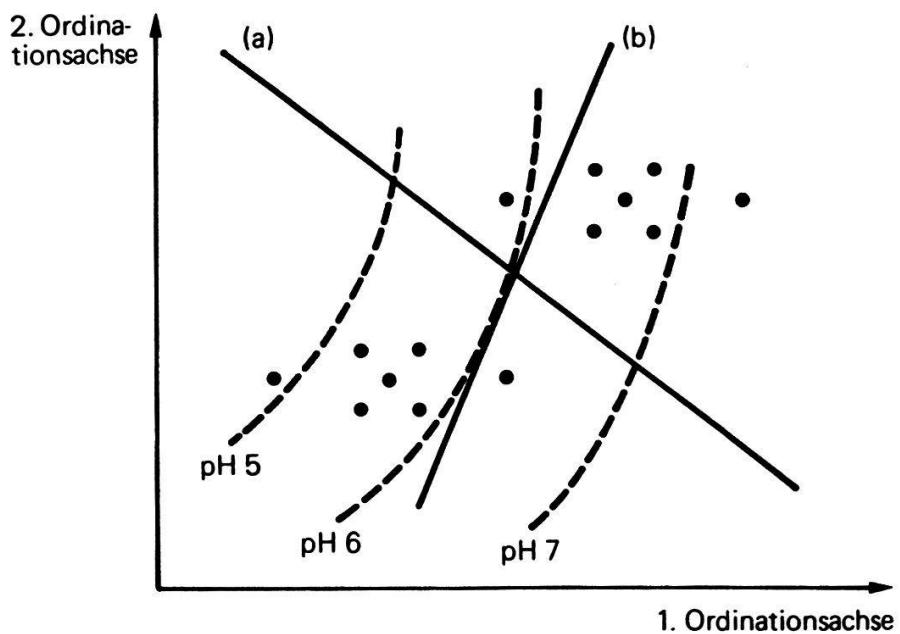


Abb. 9.2 Optimale Auftrennung einer Ordination in zwei Aufnahmegruppen auf Grund der Ähnlichkeiten der Aufnahmen (a) und des pH-Wertes (b).

Tabelle 9.1

Versuchsdaten zur Untersuchung einer Zeitreihe.

	1. Jahr			2. Jahr			3. Jahr		
Aufnahmefläche Nr.	1	2	3	1	2	3	1	2	3
Art 1	1	1		1	1		2		
Art 2	1	2	1	2	1	1	2	1	1
Art 3		1	2		2	2	1	3	3
Art 4			1			2		1	3

dort gezeigte Ordination beruht ausschliesslich auf Vegetationsdaten (vgl. Kap. 8). Nachträglich werden die Aufnahmen mit gemessenen pH-Werten beschriftet. Damit lässt sich die gefundene Datenstruktur auf einfachste Weise ökologisch interpretieren. Im allgemeinen ist die Zahl der Aufnahmen bei pflanzenökologischen Untersuchungen so hoch, dass nicht mehr jeder Punkt der Ordination einzeln beurteilt werden kann. Dann ist es zweckmässig, Punkteschwärme innerhalb eines Standortbereiches zu umfahren und zu beschriften (Abb. 9.3). Dass sich die Bereiche meist überschneiden, beeinträchtigt die Uebersichtlichkeit kaum. Die Methode ist denn auch recht beliebt und Beispiele finden sich vielerorts (vgl. GREEN 1979, S. 215, CLYMO 1980, KOMARKOVA 1980).

Die Zeit kann als Spezialfall eines Standortfaktors betrachtet werden. Alle Aufnahmen einer Zeitreihe werden dazu in einem Datensatz vereinigt. In einer Ordination lassen sich in der Folge nicht nur die standörtlich bedingten Differenzierungen, sondern auch der zeitliche Trend darstellen. Das folgende Beispiel lehnt sich der Arbeit VAN DER MAARELS (1969) an. Wir gehen von Tabelle 9.1 aus. Die Population besteht aus 9 gleichberechtigten Individuen (Aufnahmen). Um die Analyse überblickbar zu halten, sollen die Daten nicht transformiert werden. Es wird eine Hauptkomponentenanalyse gerechnet. Im vorliegenden Beispiel basiert sie auf van der Maarels Koeffizient als Ähnlichkeitsmass. Man erhält als Eigenwerte $\lambda_1 = 2.10$, $\lambda_2 = 1.25$, $\lambda_3 = 0.377$, und $\lambda_4 = 0.277$. Bereits zwei Dimensionen reproduzieren 83.7% der gesamten Varianz. Die Aufnahmen, welche derselben Probefläche entstammen, werden in der Reihenfolge der Erhebungszeitpunkte durch Pfeile verbunden. Bei der sehr einfachen Datenstruktur in Abb. 9.4 erkennt man leicht Entwicklungen in zwei verschiedene Richtungen. Sie werden durch die Aufnahmen 1 einerseits sowie 2 und 3 andererseits repräsentiert.

Als Ordination dargestellt (Abb. 9.4) sind die Veränderungen rascher und unter komplexeren Verhältnissen stets leichter erkennbar als in einer Vegetationstabelle. Man erkennt, dass sich die Aufnahmefläche 1 dem Anfangszustand der Fläche 2 annähert. Fläche 2 entwickelt sich in Richtung Fläche 3 und

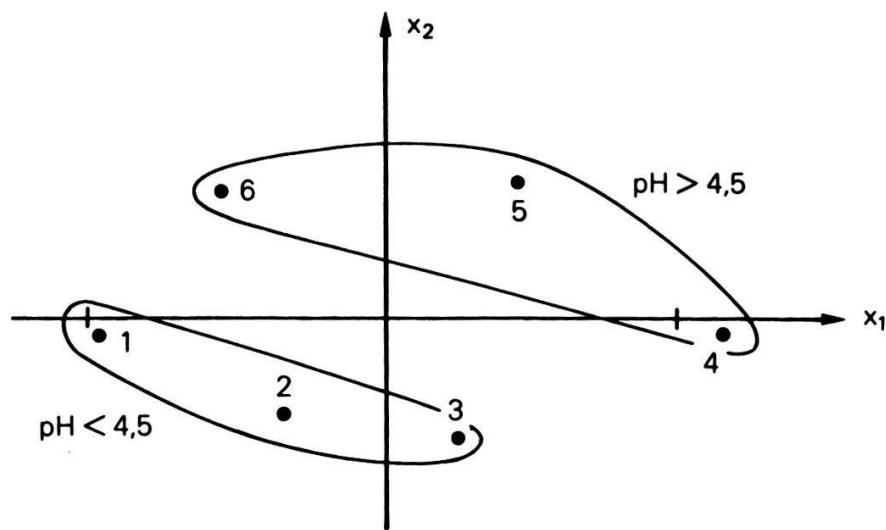


Abb. 9.3 Standörtliche Interpretation der Ordination aus Abb. 9.1. Aufnahmen innerhalb festgelegter Standortsbereiche werden umrandet und beschriftet.

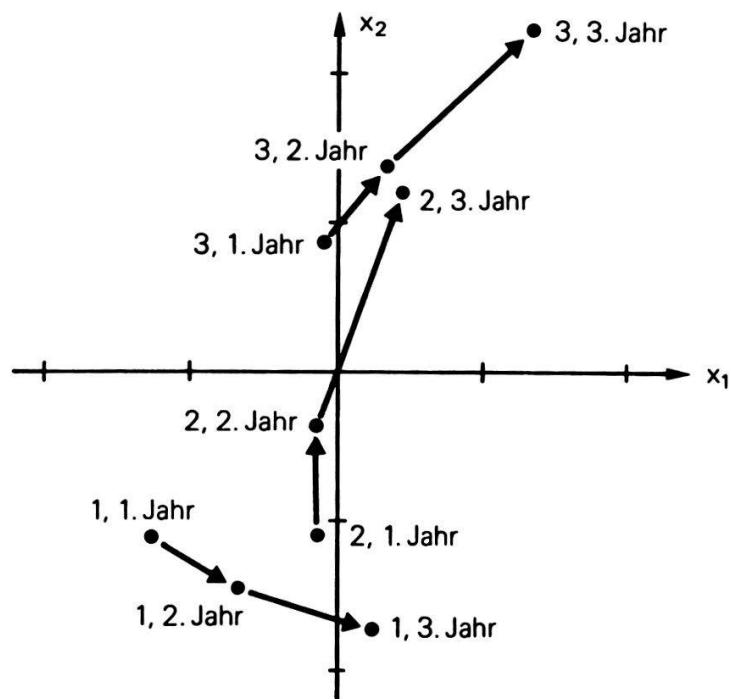


Abb. 9.4 Ordination der Daten einer Sukzessionsuntersuchung über drei Jahre (Tab. 9.1). Die Pfeile zeigen die Entwicklungsrichtung der Aufnahmeflächen.

erreicht als Endzustand etwa deren Zusammensetzung während des zweiten Jahres.

Aehnliche Möglichkeiten ergeben sich, wenn aus verschiedenen Jahren stammende Aufnahmen gemischt und gruppiert werden. Liegt eine deutliche Gruppenstruktur vor, so kann die "Wanderung" jeder sich verändernden Aufnahme durch ein Dendrogramm durch grafisches Hervorheben desselben verdeutlicht werden.

9.3 Korrelation von Ordination und Standort

Bereits in Abb. 9.1 wurde gezeigt, dass der Zusammenhang von Ordinationsachsen und Standortsmessungen mit der linearen Korrelationsanalyse nur in sehr einfachen Fällen erfassbar ist. Ordinationsverfahren wie die Hauptkomponenten- oder die Korrespondenzanalyse führen nur zum Ziel, wenn die Aehnlichkeitsstrukturen einigermassen kontinuierlich und linear sind. Wesentlich einfacher ist die Aufgabe, wenn die Individuen und die Merkmale zu Gruppen zusammengefasst werden können. Dies ist in Abb. 8.2, Abschnitt 8.2 der Fall. Die dort gezeigte Konzentrationsanalyse liefert auch Koordinaten für die Gruppen. Wir bilden also die Gruppenmittelwerte der verschiedenen Standortsfaktoren (Tab. 9.2). Nun werden die Standortsfaktoren mit der ersten Ordinationsachse korreliert und man erhält

$$\begin{aligned} r(x, \text{pH}) &= -0.977 \\ r(x, \text{Höhe}) &= -0.991 \\ r(x, \text{Neigung}) &= 0.959 \end{aligned}$$

Die hohen Korrelationswerte weisen zunächst auf eine Ueber-einstimmung zwischen der Gruppenstruktur und den Standortsfaktoren hin. Rein deskriptiv betrachtet bestätigen die Korrelationskoeffizienten die Zusammenhänge, die sich leicht aus Abb. 8.2 herauslesen lassen. Sehr vertrauenswürdig ist die auf nur drei Referenzgruppen basierende Aussage nicht. Einem Statistikbuch (BARTEL 1974) entnehmen wir, dass der Zufalls-Höchstwert im obigen Beispiel bei einer Irrtumswahrscheinlichkeit von 5% ganze 0.997 beträgt! Die soeben

berechneten Korrelationen sind offenbar mit Vorsicht zu interpretieren. Die Methode wird erst dann sinnvoll, wenn die Zahl der Gruppen grösser ist als drei. Natürlich müssten nun auch die Korrelationen der Standortsmessungen mit der zweiten und dritten Ordinationsachse überprüft werden. Ob bestehende Zusammenhänge auf diese Weise gefunden werden, bleibt ungewiss. Die Aehnlichkeitsstruktur der Aufnahme- gruppen ist wie diejenige der ungruppierten Aufnahmen mehr oder weniger nichtlinear!

Die Erkenntnis, dass Aufnahmen durch viele Ordinations- methoden als Kreise oder Spiralen dargestellt werden, machten sich FEOLI und FEOLI-CHIAPELLA (1980) zu Nutze. Weiss man mit einiger Sicherheit, dass die Vegetations- aufnahmen einem einfachen Gradienten entstammen, so lässt sich deren Position auf dem Ordinationskreis mit einem Winkelmaß gut beschreiben. Solche Voraussetzungen sind im hier verwendeten Beispiel in extremer Weise gegeben (Abb. 9.3, 9.5). Zunächst muss jedoch eine Ursprungsrichtung gefunden werden, von welcher aus die Winkel zu messen sind. In unserem Beispiel (Abb. 9.5) liegt diese zwischen den Aufnahmen 6 und 1, also dem Anfangs- und dem Endpunkt des Gradienten. Wir verwenden der Einfachheit halber die x-Achse als Ursprungsrichtung und messen nun die Winkel im mathematisch positiven Sinne. Man kann sie berechnen oder direkt Abb. 9.5 entnehmen und findet dann folgenden Zusammenhang:

Aufnahmen	1	2	3	4	5	6
Positionswinkel α	3	44	120	179	225	322
pH	4.0	4.2	4.3	4.8	5.2	5.7

Die Korrelation ergibt:

$$r(\alpha, \text{pH}) = 0.979$$

Man kann sich auch hier fragen, ob dieser Wert über dem zu erwartenden Zufallshöchstwert liegt. Die Zahl der Frei-

Tabelle 9.2

Korrelation von Ordinationskoordinaten mit Standortsfaktoren von Aufnahmegruppen.

Aufnahmegruppe	1	2	3
x-Koordinate aus Konzentrationsanalyse	1.249	-0.481	-1.001
pH	4.75	5.60	6.23
Höhe über Meer	475	493.3	503.3
Neigung in Grad	11.9	7.0	2.33

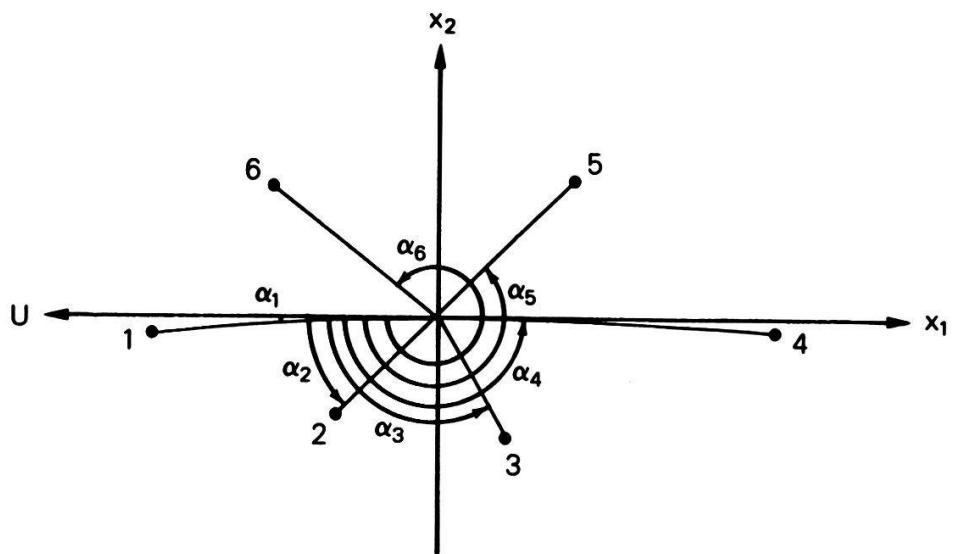


Abb. 9.5 Bestimmung des Winkels zwischen der Ursprungsrichtung U und den Aufnahmen in einer zweidimensionalen Ordination.

heitsgrade beträgt im obigen Beispiel $df = 5$. Einem Statistikbuch entnehmen wir den Zufallshöchstwert bei einer Irrtumswahrscheinlichkeit von 5% (BARTEL 1972) und finden ein r von 0.754. Unser $r(\alpha, pH)$ ist sogar noch signifikant von Null verschieden bei einer Irrtumswahrscheinlichkeit von 1%, denn der Zufallshöchstwert beträgt dann 0.874.

9.4 Die Diskriminanzanalyse

Zur Bestimmung des Zusammenhangs zwischen Vegetationseinheiten und Standortsfaktoren eignet sich die Diskriminanzanalyse. Sie erfordert Ausgangsdaten in einer Form, wie sie bei pflanzenökologischen Untersuchungen häufig anfallen. Von jeder Aufnahme ist die Gruppenzugehörigkeit vorzugeben. Man bedient sich zu deren Bestimmung beispielsweise einer Gruppierungsanalyse. Die Standortsdaten müssen hingegen ordinal oder metrisch vorliegen. Geeignet sind also pH-Werte, Konzentrationen, Temperaturen, die Höhenlage usw. Das Resultat der Diskriminanzanalyse ist als Ordination darstellbar. Deshalb kommen dem Anwender Erfahrungen aus der Interpretation von Hauptkomponentenanalysen sehr zu statten. Die Nützlichkeit der Diskriminanzanalyse zeigt sich auch an der wachsenden Zahl praktischer Anwendungen (DEL MORAL 1975, WILDI 1977, FEOLI 1979). Darstellungen der Methode finden sich bei COOLEY und LOHNES 1971, MORRISON 1976, PIELOU (1977), SUCHARD-FICHER et al. (1982) und anderen. Aehnlich wie bei den Hauptkomponentenanalysen weichen die Auffassungen bezüglich der durchzuführenden Transformationen der Ausgangsdaten voneinander ab. Im Falle pflanzenökologischer Daten muss davon ausgegangen werden, dass Messwerte mit sehr unterschiedlichen Skalierungen vorliegen. Eine Normalisierung oder Standardisierung der Standortsvektoren ist deshalb stets angezeigt.

Wir fassen im folgenden die Diskriminanzanalyse als geometrisches Verfahren auf. Wie bei der Hauptkomponentenanalyse wird für einen mehrdimensionalen Punkteschwarm eine günstige Projektion gesucht. Die Ausgangskoordinaten sind die Standortsmessungen. Die neuen Diskriminanzachsen

(meist Diskriminanzfunktionen genannt) sollen nun so zu liegen kommen, dass die Aufnahmegruppen möglichst klar getrennt erscheinen. In Abb. 9.6 ist das in Kap. 1.3, Abb. 1.11 gezeigte Beispiel nochmals wiedergegeben. Es ist leicht erkennbar, dass weder die x-Achse noch die y-Achse allein die beiden Gruppen zu trennen vermag. Die erste (und einzige) Diskriminanzfunktion hingegen erlaubt eine klare Zuordnung aller Individuen aufgrund ihrer neuen Koordinaten. Wo immer eine solcherart verdeckte Gruppenstruktur vorliegt, hilft die Diskriminanzanalyse weiter. Abb. 9.6 zeigt ausserdem, dass die erste Achse der Hauptkomponentenanalyse hier das erwünschte Ergebnis nicht erbringen kann.

Anhand eines übersichtlichen Beispiels sollen Rechengang und Wirkungsweise dargestellt werden. Wie schon bei der Hauptkomponentenanalyse verzichten wir auf Beweisführungen und verweisen auf statistische Fachliteratur (z.B. MORRISON 1976). In Abb. 9.7 ist ein Datenbeispiel dargestellt, bei welchem die Diskriminanzanalyse ein aufschlussreiches Resultat verspricht. Gegeben sind drei Gruppen von Aufnahmen, welche im dreidimensionalen Raum liegen. Die Achsen sind mit A1, A2 und A3 bezeichnet. Wir nehmen an, dass es sich dabei um Standortsfaktoren handelt wie pH, Einstrahlungssumme oder Jahresniederschlag. Die Individuen der Gruppen liegen je auf einem Achsenkreuz, welches der Übersichtlichkeit wegen innerhalb eines Würfels liegend dargestellt ist. Kein einziger Standortsfaktor x, y oder z ist geeignet, die Gruppen eindeutig zu trennen. Auch in zwei Dimensionen (Abb. 9.8, A) ist dies nicht möglich. Hingegen erahnt man, dass bei einer speziellen Blickrichtung schiefwinklig zu den drei Achsen alle drei Gruppen getrennt erscheinen könnten. In der Tat liesse sich die Gesamtpopulation so projizieren wie in Abb. 9.8, B. Dies ist denn auch das im folgenden gesuchte Endresultat.

Wie bei den Ordinationsmethoden dargestellt, bilden Artmächtigkeiten oder Standortsfaktoren den Ausgangspunkt der Analysen. Sie sind mit Hilfe von Faktorenladungen so zu transformieren, dass die gewünschte Projektionsrichtung erreicht wird. Bei der Hauptkomponentenanalyse ging es darum,

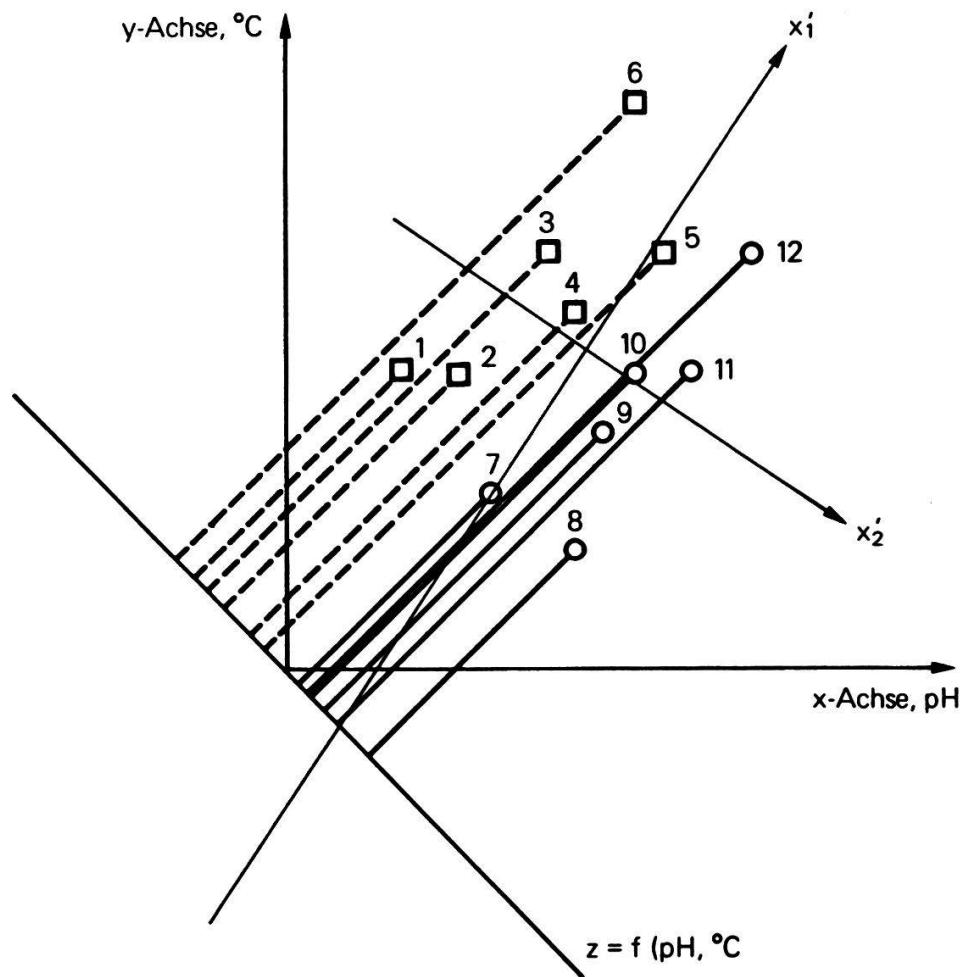


Abb. 9.6 Diskriminanzfunktion $z = f(\text{pH}, ^{\circ}\text{C})$ für zwei Gruppen einer Stichprobe. Zum Vergleich sind die beiden Achsen (x'_1, x'_2) einer Hauptkomponentenanalyse eingetragen.

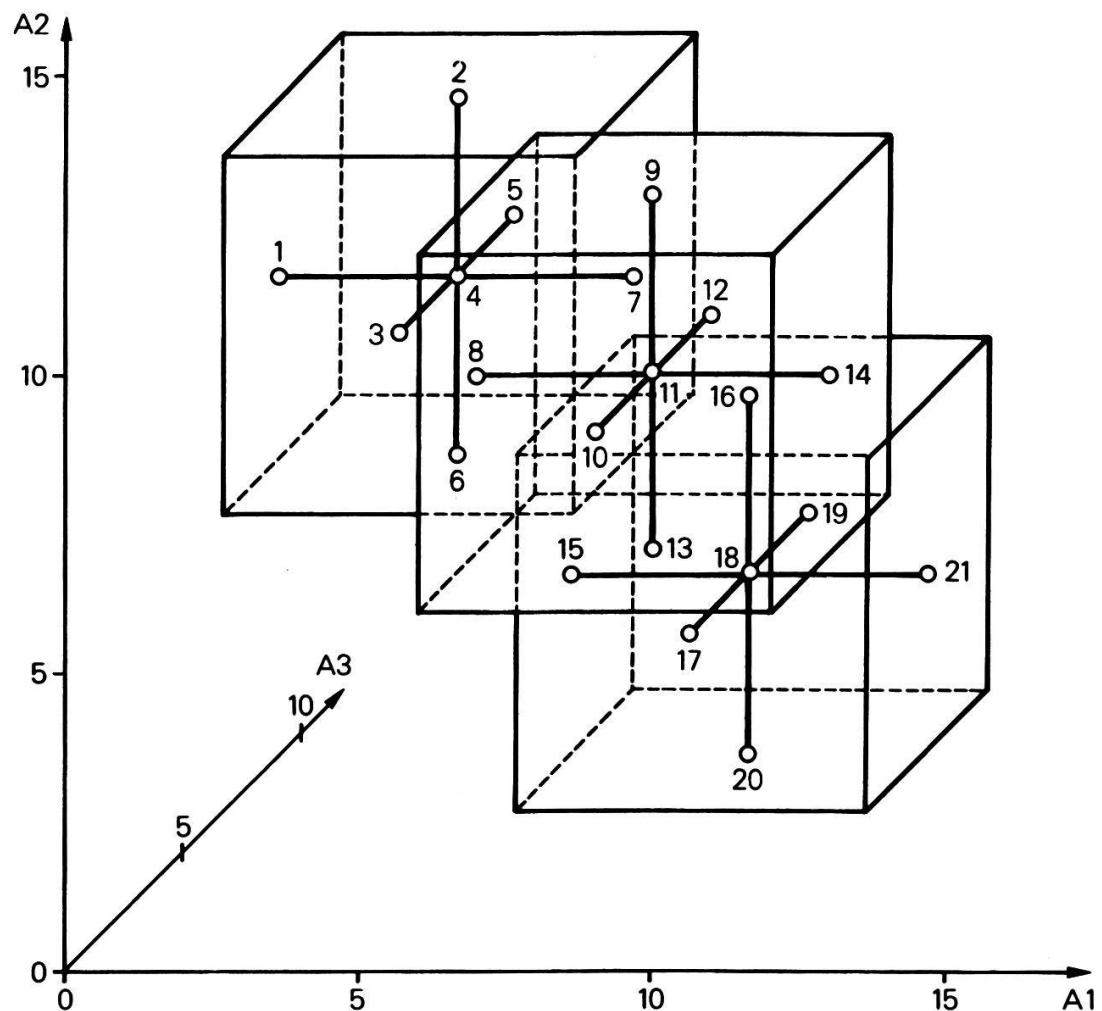


Abb. 9.7 Dreidimensionaler Punkteschwarm, bestehend aus drei durch Achsenkreuze verbundene Punktegruppen. Aus dieser Perspektive überschneiden sich die Gruppen. Die Würfel sollen die Lage der Punkte im Raum verdeutlichen.

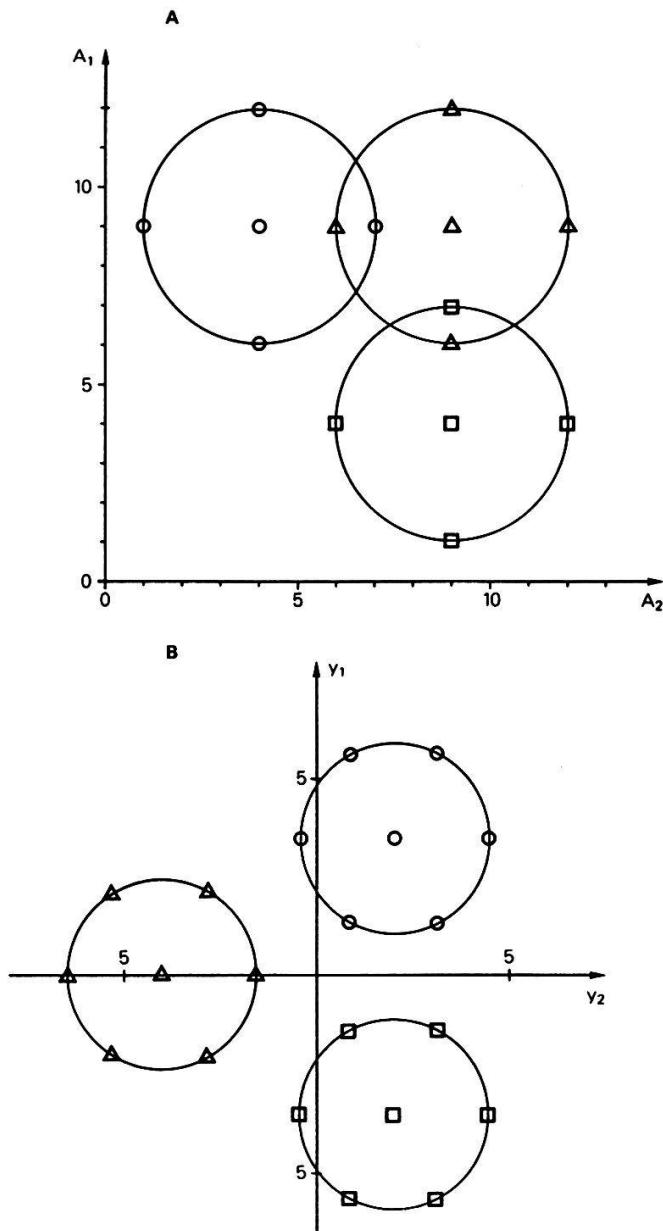


Abb. 9.8 Projektion des Punkteschwarmes in Abb. 9.7 in den Dimensionen a_1 und a_2 (A). Die Diskriminanzanalyse ist in der Lage, zwei neue Achsen y_1 und y_2 so zu finden, dass sich die drei Gruppen nicht überschneiden (B).

die Varianz der neuen Koordinaten auf der ersten Achse zu maximieren. Die Diskriminanzanalyse hingegen versucht, die Varianz zwischen den Gruppen möglichst gross erscheinen zu lassen. Das Problem der Trennung von Gruppen kennen wir bereits von der Varianzanalyse her (Kap. 6.3). Dabei gilt ja

$$\text{Totalvarianz der Stichprobe} = \frac{\text{Varianz innerhalb der Gruppen}}{\text{Varianz zwischen den Gruppen.}}$$

Neu ist nur, dass bei der Diskriminanzanalyse mehrere Merkmale gleichzeitig berücksichtigt werden. Es sind somit stets Varianz- Kovarianzmatrizen zu analysieren. Liest man aus Abb. 9.7 alle Koordinaten heraus, so erhält man die Datenmatrix A (Tabelle 9.3). Wie schon bei der Hauptkomponentenanalyse, so sollen auch hier die Kolonnen zentriert werden. Dazu ist von jeder Koordinate der Mittelwert zu subtrahieren:

$$x_{ij} = a_{ij} - \bar{a}_j$$

Die neuen, transformierten x_{ij} - Werte finden sich ebenfalls in Tabelle 9.3. Nun ist die Ähnlichkeitsstruktur der zentrierten Standortsfaktoren x_1, x_2 und x_3 zu bestimmen. Als Ähnlichkeitsmaß wählen wir der Einfachheit halber das Skalarprodukt. Wir bezeichnen die so entstehende Matrix mit T (für total). Bei zwei beliebigen Standortsfaktoren n und m gilt:

$$t_{nm} = \sum_{i=1}^N x_{im} x_{in}$$

N ist die Gesamtzahl der Aufnahmen. Für das Element t_{12} erhält man somit als Beispiel

$$t_{12} = (-6.33*1.66) + (-3.33*4.66) + \dots + (4.66*-1.333) \\ = -58.33$$

Die gesamte Matrix ergibt

Tabelle 9.3

Datenmatrix für das Rechenbeispiel in Abb. 9.6.

Aufnahme	Gruppe	Standortsfaktor			Zentrierte Daten		
		A1	A2	A3	X1	X2	X3
1	1	1	9	9	-6.33	1.66	1.66
2	1	4	12	9	-3.33	4.66	1.66
3	1	4	9	6	-3.33	1.66	-1.33
4	1	4	9	9	-3.33	1.66	1.66
5	1	4	9	12	-3.33	1.66	4.66
6	1	4	6	9	-3.33	-3.33	1.66
7	1	7	9	9	-0.33	1.66	1.66
8	2	6	9	4	-1.33	1.66	-3.33
9	2	9	12	4	1.66	4.66	-3.33
10	2	9	9	1	1.66	1.66	-6.33
11	2	9	9	4	1.66	1.66	-3.33
12	2	9	9	7	1.66	1.66	-0.33
13	2	9	6	4	1.66	-1.33	-3.33
14	2	12	9	4	4.66	1.66	-3.33
15	3	6	4	9	-1.33	-3.33	1.66
16	3	9	7	9	1.66	-3.33	1.66
17	3	9	4	6	1.66	-3.33	-1.33
18	3	9	4	9	1.66	-3.33	1.66
19	3	9	4	12	1.66	-3.33	4.66
20	3	9	1	9	1.66	-6.33	1.66
21	3	12	4	9	4.66	-3.33	1.66
$\bar{x} =$		7.33	7.33	7.33	0	0	0

$$T = \begin{matrix} 170.66 & -58.33 & -58.33 \\ -58.33 & 170.66 & -58.33 \\ -58.33 & -58.33 & 170.66 \end{matrix}$$

Die Gleichförmigkeit verschiedener Elemente ergibt sich aus dem gewählten Beispiel: Alle Gruppen in Abb. 9.7 sind ja gleich gross und haben die gleiche Form. Im folgenden Schritt sind die Ähnlichkeitsverhältnisse innerhalb der einzelnen Gruppen zu bestimmen. Das Vorgehen ist genau gleich wie für die Gesamtstichprobe: Die Daten werden (innerhalb der Gruppen!) zentriert und die Matrix der Skalarprodukte W_1 (für "within") berechnet. Man erhält für die erste Gruppe:

$$W_1 = \begin{matrix} 18.0 & 0 & 0 \\ 0 & 18.0 & 0 \\ 0 & 0 & 18.0 \end{matrix}$$

Es ist leicht einzusehen, dass die Resultate für die Gruppen 2 und 3, also W_2 und W_3 , identisch ausfallen müssen. Aus der Summe der drei Matrizen ergeben sich schliesslich die totalen Abweichungsquadrate innerhalb der Gruppen:

$$W = W_1 + W_2 + W_3 = \begin{matrix} 54.0 & 0 & 0 \\ 0 & 54.0 & 0 \\ 0 & 0 & 54.0 \end{matrix}$$

Nun sollen also die neuen Koordinaten gesucht werden, welche die Punktewolken optimal auflösen. Der Weg führt wie bei der Hauptkomponentenanalyse über ein Eigenwertproblem. Zunächst sind geeignete Eigenwerte λ zu suchen. Sie ergeben sich durch Auflösung der Determinantengleichung

$$[W^{-1} (T-W) - \lambda I] = 0.$$

Das Problem unterscheidet sich von demjenigen der Hauptkomponentenanalyse nur insofern, als zunächst der Ausdruck $W^{-1} (T-W)$ zu berechnen ist. $T-W$ soll mit A bezeichnet werden und wir erhalten

$$\begin{aligned}
 A &= T - W = \begin{bmatrix} 170.66 & -58.33 & -58.33 \\ -58.33 & 170.66 & -58.33 \\ -58.33 & -58.33 & 170.66 \end{bmatrix} - \begin{bmatrix} 54.0 & 0 & 0 \\ 0 & 54.0 & 0 \\ 0 & 0 & 54.0 \end{bmatrix} \\
 &= \begin{bmatrix} 116.66 & -58.33 & -58.33 \\ -58.33 & 116.66 & -58.33 \\ -58.33 & -58.33 & 116.66 \end{bmatrix}
 \end{aligned}$$

W^{-1} ist die Inverse von W , d.h. die Lösung der Gleichung $WW^{-1} = I$. Die Bedeutung der Matrixinversion ist z.B. bei BATSCHET (1980) erklärt. Dort findet sich auch ein Lösungsweg (S. 518-520). Folgt man den beschriebenen Rechenregeln, so scheitert man mit unserem Beispiel: Die Determinante der Matrix W ist Null, die Inverse W^{-1} ist undefiniert. Glücklicherweise wurde von Mathematikern ein sicherer Weg zur Bestimmung der Eigenstruktur von $W^{-1} A$ gefunden. Die einfachen, aber rechenintensiven Operationen findet man bei COOLEY und LOHNES (1971), S. 192 bis 194, zusammen mit einem Computerprogramm. Wir verzichten deshalb auf die detaillierte Darstellung des Rechenganges. Für unser Beispiel lauten die Eigenwerte

$$\lambda_1 = 3.24 ; \lambda_2 = 3.24$$

Die Eigenvektoren α (alpha) betragen

Alte Koordinate	Diskriminanzachse	
	1	2
1	-0.408	-0.707
2	-0.408	0.707
3	0.817	0.0

Die Interpretation der Eigenvektoren erfolgt genau gleich wie bei der Hauptkomponentenanalyse. Demzufolge ist $\alpha_{11} = -0.408$ die Korrelation (der Cosinus) der alten Achse 1 mit der Diskriminanzachse 1. Diese ist mit $\alpha_{13} = 0.817$ ähnlich orientiert wie die alte Achse 3. Die Diskriminanzachse 2 steht in 45° zu den alten Achsen 1 und 2 (Cosinus $45^\circ =$

Tabelle 9.4

Koordinaten auf den Dikriminanzachsen

Aufnahme	Gruppe	Koordinaten	
		Y1	Y2
1	1	3.266	5.657
2	1	0.817	5.657
3	1	-0.408	3.536
4	1	2.041	3.536
5	1	4.491	3.536
6	1	3.266	1.414
7	1	0.817	1.414
8	2	-2.858	2.121
9	2	-5.307	2.121
10	2	-6.532	0.000
11	2	-4.082	0.000
12	2	-1.633	0.000
13	2	-2.858	-2.121
14	2	-5.307	-2.121
15	3	3.266	-1.141
16	3	0.817	-1.141
17	3	-0.408	-3.536
18	3	2.041	-3.536
19	3	4.491	-3.536
20	3	3.266	-5.657
21	3	0.817	-5.657

0.707). Zur Achse 3 steht sie rechtwinklig. Dass dem so ist, zeigt auch Abb. 9.8, B. Nun sind noch die dort verwendeten Koordinaten Y zu bestimmen. Dies geschieht wiederum wie bei der Haupkomponentenanalyse. Für jede Dimension p gilt

$$y_{1p} = x_{1p} \alpha_{11} + x_{2p} \alpha_{21} + x_{3p} \alpha_{31}$$

Die aktuellen Werte (Tabelle 9.4) sind

$$\begin{aligned} y_{11} &= (-0.408)(-6.33) + (-0.408)(1.667) + (0.817)(1.667) \\ &= 3.26 \end{aligned}$$

$$\begin{aligned} y_{12} &= (-0.707)(-6.33) + (0.707)(1.667) + (0.0)(1.667) \\ &= 5.65 \end{aligned}$$

Alle weiteren Koordinatenwerte sind in Tabelle 9.4 wiedergegeben. Wie in Abb. 9.8, B, gezeigt, präsentieren sich die drei Punkteschwärme in zwei Dimensionen klar getrennt. Sind die ursprünglichen Koordinaten Standortsfaktoren, so können wir jetzt festhalten, dass diese die - aufgrund floristischer Angaben - vorgegebenen Gruppen einzeln nicht zu unterscheiden erlauben. In Kombination charakterisieren sie jedoch jede Gruppe eindeutig.

10. Diskussion und Schlussfolgerungen

Vegetationskunde gehört zu den sogenannten "weichen" Wissenschaften. Sie beschäftigt sich mit Gegebenheiten, die sehr komplex, in der Regel schwer erkenn- und nachvollziehbar sind. Viele Störfaktoren verursachen Streuung in den Daten und erschweren das Erkennen einfacher Gesetzmässigkeiten. Ziel der statistischen Analyse ist es, interpretierbare Strukturen von blossem "Rauschen" zu trennen. Man kann nun davon ausgehen, dass die in den vorhergehenden Kapiteln dargestellten Lösungswege und Methoden bei vegetationskundlichen Daten eine solche Trennung herbeiführen können. Beweisbar ist diese Annahme nicht. Misserfolge sind zu erwarten. In der Literatur tauchen sie kaum auf, denn negative Ergebnisse werden selten publiziert. Die Vielfalt der statistischen Methoden lässt es als sehr wahrscheinlich erscheinen, dass Fehler im Auswertekonzept für viele Fehlversuche verantwortlich sind. Selbst bei Verwendung der besten Computerprogramme ist es möglich, klare Strukturen zu übersehen und richtige Ergebnisse als falsch zu betrachten. Erfolgsentscheidend ist insbesondere nicht nur die Auswertemethode, sondern der ganze Analyseweg (WILDI und ORLOCI 1983, S.10). Bei seiner Konzeption müssen die grundlegenden Probleme der Vegetationskunde berücksichtigt werden, wie zum Teil in den vorhergehenden Kapiteln angedeutet wurde. Wenn es schon schwierig ist, die Einsatzmöglichkeiten einzelner Methoden in verallgemeinerter Form darzustellen, so gilt dies noch viel mehr für den gesamten Auswertegang. Einige Grundsätze sollen gleichwohl nachfolgend diskutiert werden.

10.1 Grundgesamtheit und Stichprobe

Dem Problem der Stichprobenerhebung wurde in der Vegetationskunde bislang wenig Beachtung geschenkt (Kap. 2). Das liegt unter anderem daran, dass das Ziel vieler Untersuchungen in der Ermittlung qualitativer Zustände liegt. Im Gegensatz zu land- und forstwirtschaftlichen Fragestellungen, die oft dynamische Elemente enthalten, spielt die Interpolier- und Extrapolierbarkeit der Messungen, die sehr

stark auf dem Stichprobenkonzept beruht (ZÖHRER 1982), eine untergeordnete Rolle. Es ist klar, dass jede Stichprobe eine Schätzung erlaubt für die Grundgesamtheit, der sie entnommen wurde. Der Fehler dieser Schätzung ist jedoch nur quantifizierbar, wenn die Stichprobe mit Hilfe eines Zufallsstichprobenplanes gewonnen wurde. Solche Ueberlegungen sind bei explorativer Datenanalyse nicht zwingend. PIELOU (1984) betrachtet konsequenterweise ihre Datensätze als Gegenstand der Untersuchung. Ob es sich dabei um die eigentliche Grundgesamtheit oder um eine Stichprobe handelt, hinterfragt sie nicht. Und in der Tat bleibt die explorative Analyse der Daten in beiden Fällen genau dieselbe.

Wenn nun in Kapitel 2 die Ansicht vertreten wird, es sollte mit Zufallsstichproben gearbeitet werden, so liegt der Grund nicht in den Erfordernissen der Analyse, sondern in der besseren Interpretierbarkeit der Ergebnisse. Die Vegetationsdecke ist von einer enormen Variabilität gekennzeichnet. Bei subjektiver Auswahl der Stichprobeflächen wird meist versucht, ihr aus dem Wege zu gehen. Die Ergebnisse werden damit klarer, verlieren aber an Allgemeingültigkeit. In Zufallsstichproben erscheinen anstelle von Gruppen von Aufnahmen oft Gradienten. Letztere sind schwieriger zu beschreiben und weniger leicht umzusetzen für praktische Anwendungen. Besonders Ordinationen decken das Fehlen echter Unstetigkeiten unweigerlich auf. Die Erkenntnis, dass eine Stichprobe kaum eine Gruppenstruktur aufweist und mit viel Rauschen behaftet ist, kann aber nicht als negativ beurteilt werden. Es ist vielmehr zu prüfen, ob der Befund nicht den wirklichen Verhältnissen entspricht - ganz im Sinne von GREEN (1979), S. 63f, der feststellt, dass man einem sauber erarbeiteten Ergebnis Vertrauen entgegenbringen soll.

10.2 Methodische Entscheidungsschritte

Die explorative Analyse von Daten ist ein Experimentieren. Das Experiment soll aber nicht planlos verlaufen, sondern einer noch zu diskutierenden Strategie entsprechen. Von Experimenten im Labor und von Felduntersuchungen wissen wir,

dass sie auf zahlreichen, willkürlichen Entscheidungen beruhen. Entscheidungen schränken den Gültigkeitsbereich der Ergebnisse ein. Sie sind in jegliche Interpretation als zusätzliche Randbedingung mit einzubeziehen. Beispiele sind die Wahl des Untersuchungsgegenstandes, des Untersuchungszeitpunktes, der Messparameter und der Messmethode, usw. Bei der explorativen Datenanalyse geht dieser Prozess weiter, denn in der Regel gibt es kein ganz bestimmtes statistisches Modell, auf welches die Untersuchung zugeschnitten ist. Der Gliederung dieser Schrift entsprechend kann aber wiederum schrittweise vorgegangen werden. Es sind namentlich Entscheide zu fällen über die Skalierung, die Transformation, die Wahl des Ähnlichkeitsmasses, die Methode zur Analyse der Ähnlichkeitsstruktur, die Art der Weiterverarbeitung der Ergebnisse usw. Was resultiert, sind Hypothesen - zur Entstehung und Funktionsweise eines Vegetationsmusters, zur Ursache und zum Verlauf einer Veränderung, zur Möglichkeit der Steuerung usw. Die Methoden der explorativen Statistik helfen deshalb, Hypothesen zu formulieren. Liegen letztere vor, so kann eigentlich erst abgeklärt werden, ob sie mit Hilfe neuer Experimente oder Untersuchungen statistisch getestet werden können und sollen. Das wird in einigen Fällen sinnvoll sein, in anderen nicht. Überlegungen zum Sinn statistischer Tests in der Feldbiologie diskutiert GREEN (1979) eingehend. Wir gehen hier davon aus, dass der Engpass zum Verständnis ökologischer Prozesse beim Finden von Hypothesen hinreichenden Erklärungsgrades liegt.

10.3 Elemente einer Standardstrategie

Nehmen wir an, dass ein Datensatz zu analysieren ist, der eine Lebensgemeinschaft beschreibt. Sowohl in der Pflanzensoziologie, wie auch in der Faunistik fallen dabei zumeist Zähldaten oder Deckungswerte an. Der erste Schritt besteht nun darin, eine geeignete Skalierung zu finden. Besonders Individuenzahlen schwanken im allgemeinen in einem so grossen Bereich, dass die Vergleichbarkeit verschiedener Arten mit einem der üblichen Standardisierungsverfahren nicht zu erreichen ist. Es ist deshalb eine geeignete (ordinale) Skalierung zu finden, die sogar von Art zu Art verschieden

sein kann. Die sich später ergebenden Resultate basieren natürlich vollständig auf der neuen Skala. Dieser Schritt ist sogar oft der kritischste in der ganzen Analyse. Besonders bei Deckungswerten ist es sinnvoll, den gesamten Auswertegang mit unterschiedlichen Skalen durchzuspielen. Im einen Fall soll vor allem die Gegenwart der Arten berücksichtigt werden (Präsenz-Absenz Skala), im andern Falle vor allem deren Quantitäten. Die sich ergebenden Unterschiede zeigen, welches Vorgehen sinnvoll ist und ob allenfalls eine Zwischenlösung gewählt werden muss.

Neu skalierte Daten weiter zu transformieren, ist durchaus denkbar. Es ist zu überlegen, ob sie als Zähldaten interpretiert werden können, so dass eine Weiterverarbeitung mit einem Kontingenzmass oder im Sinne der Korrespondenzanalyse möglich ist. Bereichsanpassung, Normalisierung und Standardisierung eignen sich allenfalls für den Vergleich von Aufnahmeverektoren. Sie sind mit Bedacht zu verwenden für Artvektoren, die in der Regel viele Nullen aufweisen. Auch hier lohnt es sich, im Zweifelsfalle mit und ohne Standardisierung zu rechnen - womit bei Verwendung zwei verschiedener Skalierungen bereits vier Analysesequenzen durchzuspielen sind.

Ein dritter Schritt ist die Wahl des Vergleichsmasses. In vielen Fällen wird sie diktiert durch die später anzuwendende Analysemethode. Zu beachten ist, dass verschiedene Vergleichsmasse eine Transformation in sich bergen, was einen ungewollten Informationsverlust zur Folge haben kann. Beliebt sind einfache Masse, wie die Euklidsche Distanz oder eine Absolutwertfunktion. Beide sind für die Analyse von Lebensgemeinschaften unbedenklich, sofern eine sinnvolle Skalierung gewählt wurde. Dagegen erscheint die Verwendung des Korrelationskoeffizienten meist ungerechtfertigt, denn er bringt eine oft ungewollte Angleichung der Streuungsverhältnisse mit sich.

Eine der charakteristischen Eigenschaften eines multidimensionalen Datensatz ist dessen Dimensionalität. Diese sollte in einem vierten Schritt geklärt werden. Die Haupt-

komponentenverfahren bringen zwar nur eine lineare Lösung des Problemes, die aber meist gute Hinweise auf die Komplexität der Stichprobe liefert. Lebensgemeinschaften können im allgemeinen nur unter Berücksichtigung mehrerer Dimensionen befriedigend beschrieben werden. Typisch ist, dass auch sechs bis zehn Achsen oft nur 30 bis 40% der Gesamtvarianz erklären. Die Erfahrung lehrt, dass dieser an sich bescheidene Anteil genügt, um die wichtigsten Strukturen und Gesetzmässigkeiten erkennen zu können. Eine Ordination der Aufnahmen zeigt auch gleich, ob eine Gruppenstruktur vorliegt. Sofort erkennbar sind Ausreisser oder deutlich abgetrennte Teile der Stichprobe. Sie können die Analyse dominieren und alle andern Strukturen verwischen. Bei Eindeutigkeit sind sie aus der Stichprobe zu entfernen (im Sinne einer Stratifizierung) und separat auszuwerten.

Hauptkomponentenverfahren geben auch Auskunft über die Bedeutung der Arten in einer Ordination der Aufnahmen. Eine Alternative dazu stellen Gewichtungsverfahren (Rangierungen) dar, die meist leichter zu interpretieren sind. Eine sehr sorgfältige Interpretation der Ergebnisse einer Artenrangierung lohnt sich nicht zuletzt deshalb, weil hier noch am ehesten Fehlgriffe in der Methodenwahl erkannt werden können. Bei unzweckmässig skalierten Daten können Arten ein hohes Gewicht erhalten, die ausnahmsweise massenhaftes Vorkommen aufweisen - eine in der Natur ganz normale Erscheinung. Gegebenenfalls ist das Auswertekonzept anzupassen.

Die Suche nach Gruppen hat als fünfter Schritt praktische Bedeutung, denn viele Anwendungen erfolgen auf der Basis von generalisierenden Typen von Lebensgemeinschaften. Der alleinige Einsatz von Gruppierungsverfahren ist möglich, die Ergebnisse lassen sich aber im allgemeinen nicht genügend gut interpretieren. Insbesondere sollte der Gruppierungsalgorithmus der aus der Ordination ersichtlichen Ähnlichkeitsstruktur angepasst sein. Liegt eine diffuse Punktwolke vor, so generiert ein Teilungsverfahren die gewünschten Gruppen - welche dann aber künstlicher Natur sind. Agglomerative Verfahren liefern immer ein Ergebnis. Es

bleibt aber meist einem weiteren Analyseschritt vorbehalten, herauszufinden, ob die Gruppenunterteilungen Unstetigkeiten in den Daten entsprechen. Methoden wie die GRID-Analyse lassen eine beliebige Teilung der Stichprobe nicht zu. Ihrem Verhalten kann entnommen werden, ob allenfalls echte Gruppen vorliegen. Zusammenfassend kann festgestellt werden, dass auch beim Einsatz von Gruppierungsverfahren oft zweckmässigerweise Versuche mit unterschiedlichen Methoden durchgeführt werden sollten.

Nach einer sorgfältigen Interpretation der soweit angefallenen Ergebnisse umfasst der sechste Schritt die Weiterverarbeitung der Resultate. Diese liegen ja in der Regel wieder als Daten vor, nämlich als Ordinationskoordinaten, Gruppenmittelwerte und Streuungen usw. Das Spektrum der in Frage kommenden Methoden erstreckt sich über die bereits dargelegten Möglichkeiten hinaus auf spezielle Methoden wie die Konzentrations-, die Diskriminanzanalyse und viele andere. In komplexen Fällen wird man nicht umhin kommen, die verfügbaren Hypothesen zum Entwurf eines mathematischen Simulationsmodells zu verwenden, um zu einer Ueberprüfung der Ergebnisse zu gelangen.

Die meisten Analysemethoden ergeben informativere Ergebnisse, wenn mit kleinen Datensätzen gearbeitet wird. Es ist deshalb wohl immer vorteilhaft, wenn auf Grund einer ersten Analyse eine begründbare Unterteilung des Datensatzes vorgenommen werden kann. Für jeden Datenteil beginnt dann die Analyse wieder von vorn. Viele strukturelle Details werden damit leichter erkennbar sein.

Literatur

- ANDERBERG R.M., 1973: Cluster analysis for applications. Acad. Press, New York San Francisco London. 359 S.
- BAHRENBERG G. und GIESE E., 1975: Statistische Methoden und ihre Anwendung in der Geographie. Teubner, Stuttgart. 308 S.
- BATSCHELET E., 1980: Einführung in die Mathematik für Biologen. Springer Verlag. Berlin Heidelberg New York. 557 S.
- BARTEL H., 1974: Statistik I für Psychologen, Pädagogen und Sozialwissenschaftler. 2. Aufl. Fischer, Stuttgart. 161 S.
- BARTEL H., 1972: Statistik II für Psychologen, Pädagogen und Sozialwissenschaftler. Fischer, Stuttgart. 203 S.
- BENZECRI J.-P., 1969: Statistical analysis as a tool to make patterns emerge from data. In: WATANABE S. (Hrsg.), Methodologies of pattern recognition. Acad. Press, New York, 35-60.
- BRAUN-BLANQUET J., 1964: Pflanzensoziologie. Grundzüge der Vegetationskunde. (3. Aufl.). Springer, Wien. 865 S.
- BRAY J.R. und CURTIS J.T., 1957: An Ordination of the Upland Forest Communities of Southern Wisconsin. Ecological Monographs, 27, 325-349.
- CLYMO R.S. , 1980: Preliminary survey of the peat-bog Hummell Knowe Moss using various numerical methods. Vegetatio 42, 129-148.
- COOLEY W.W. und LOHNES P.R., 1971: Multivariate data analysis. John Wiley Sons, New York Chichester Brisbane Toronto Singapore. 346 S.
- DAGET PH. und GODRON M., 1982: Analyse fréquentielle de l'écologie des espèces dans les communautés. Masson, Paris New York Barcelone Milan Mexico Rio de Janeiro. 163 S.
- DALE M.B., 1980: A syntactic basis of classification. Vegetatio 42, 93-98.
- DE LAGARDE J., 1983: Initiation à l'analyse des données. Dunod, Paris. 158 S.
- DEL MORAL R., 1975: Vegetation clustering by means of ISO-DATA: revision by multiple discriminant analysis. Vegetatio 29, 179-190.
- DEL MORAL R., 1980: On selecting indirect ordination methods. Vegetatio 42, 75-84.

- ELLENBERG H. , 1956: Aufgaben und Methoden der Vegetationskunde. In: WALTER H. (Hrsg.), Einführung in die Phytologie 6, 1, 136 S.
- ELLENBERG H., 1978: Vegetation Mitteleuropas mit den Alpen in ökologischer Sicht. (2. Aufl.), Ulmer, Stuttgart. 981 S.
- ELLENBERG H., 1979: Zeigerwerte der Gefässpflanzen Mitteleuropas. (2. Aufl.), Scripta Geobotanica 9, 97 S.
- FEOLI E. , 1973: Un indice che stima il peso dei caratteri per classificazioni monotetiche. Giornale Botanico Italiano 107, 263-268.
- FEOLI E. und ORLOCI L., 1979: Analysis of concentration and detection of underlying factors in structured tables. Vegetatio 40 (1), 49-54.
- FEOLI E. und FEOLI-CHIAPELLA L., 1980: Evaluation of Ordination Methods through Simulated Coenoclines: Some Comments. Vegetatio 42, 35-41.
- FEOLI E., LAGONEGRO M., und ZAMPAR A., 1982: Classificazione e ordinamento della vegetazione. Metodi e programmi di calcolo. Consiglio Nazionale delle Ricerche, AQ/5/35, Udine. 192 S.
- FEOLI E. und LAGONEGRO M., 1982: Syntaxonomical analysis of beech woods in the Apennines (Italy) using the program package IAHOPA. Vegetatio 50, 129-173.
- FEOLI E., LAGONEGRO M. und ORLOCI L., 1984: Information Analysis of Vegetation Data. Junk, The Hague. 143 S.
- GAENSSLER H. und SCHUBOE W., 1973: Einfache und komplexe statistische Analyse. UTB 274, Reinhardt, München. 326 S.
- GAUCH H.G., 1977: ORDIFLEX - a flexible computer program for four ordination techniques: Weighted Averages, Polar Ordination, Principal Components Analysis, and Reciprocal Averaging, release B. Ithaca, New York: Cornell University.
- GAUCH H.G., 1979: COMPCLUS - a FORTRAN program for rapid clustering of large data sets. Ithaca, New York, Cornell University.
- GAUCH H.G., 1982: Multivariate analysis in community ecology. Cambridge Univ. Press. 298 S.
- GLEASON H.A., 1926: The individualistic concept of the plant association. Bull.Torrey Bot.Club 53, 7- 26.
- GLEASON H.A., 1939: The individualistic concept of the plant association. Amer. Midl. Nat. 21, 92-110.

- GITTINS R., 1985: Canonical analysis; a review with applications in ecology. Springer, Berlin Heidelberg New York. 351 S.
- GOODALL D.W., 1954: Objective methods for the classification of vegetation. III. An essay in the use of factor analysis. Australian J.of Botany 2, 304-324.
- GOWER J.C., 1966: Some distance properties of latent root and vector methods used in multivariate analysis. Biometrika 53, 325-338.
- GOWER J.C., 1967: Multivariate analysis and multidimensional geometry. The Statist. 17, 13-28.
- GOWER J.C. und ROSS G.J.S., 1969: Minimum spanning tree and single linkage cluster analysis. Appl.Statist. 18, 54-64.
- GRABHERR G., 1985: Numerische Klassifikation und Ordination in der alpinen Vegetationsökologie als Beitrag zur Verknüpfung moderner "Computermethoden" mit der pflanzensoziologischen Tradition. Tüxenia 5, 181-190.
- GREEN R.H., 1979: Sampling design and statistical methods for environmental Biologists. John Wiley and Sons, New York Chichester Brisbane Toronto. 257 S.
- GREIG-SMITH P. , 1982: Quantitative plant ecology. (3. Aufl.). Studies in Ecology. Wiley-Interscience, New York. 359 S.
- GRUENIG A., 1977: Die Vegetationsentwicklung im Flachseegebiet. Jahresber. 1977, Stiftung Reusstal, 16-23.
- HILL M.O., 1974: Correspondence analysis: A neglected multivariate method. Appl.Statist. 23, 340- 354.
- HILL M.O., 1979a: DECORANA - a FORTRAN program for detrended correspondence analysis and reciprocal averaging. Ithaca, New York: Cornell University. 52 S.
- HILL M.O., 1979b: TWINSPLAN - a FORTRAN program for arranging multivariate data in an ordered two way table by classification of individuals and attributes. Ithaca, New York: Cornell University. 48 S.
- HUBER H., 1979: Ueber den Gebrauch mathematisch-statistischer Methoden in der Taxonomie. Ber.Schweiz. Bot.Ges. 89(3/4), 227-250.
- JACCARD P., 1901: Etude comparative de la distribution florale dans une portion des Alpes et du Jura. Bull. Soc.Vaud.Sc.Nat. 37, 547-579.

- JANCEY R.C.**, 1974: Algorithm for detection of discontinuities in data sets. *Vegetatio* **29**, 131-133.
- JANCEY R.C.**, 1979: Species ordering on a variance criterion. *Vegetatio* **39**, 59-63.
- KOMARKOVA V.**, 1980: Classification and ordination in the Indian Peaks area, Colorado Rocky Mountains. *Vegetatio* **42**, 149-163.
- KRUSKAL J.B.**, 1964: Nonmetric multidimensional scaling: A numerical method. *Psychometrika* **29**, 115-129.
- KUHN N.**, 1983: VEGTAB, ein Computer-Programm als Hilfe zur tabellarischen Vegetationsgliederung. *Tuexenia* **3**, 499-522.
- LANDOLT E.**, 1977: Oekologische Zeigerwerte zur Schweizer Flora. Veröff. Geobot. Inst. ETH, Stiftung Rübel, **64**. 208 S.
- LEGENDRE L. und LEGENDRE P.**, 1979: Ecologie numérique. Tome I. Le traitement multiple des données écologiques. Masson, Paris New York Barcelone Milan. 197 S.
- LEGENDRE L. und LEGENDRE P.**, 1979: Ecologie numérique. Tome II. La structure des données écologiques. Masson, Paris New York Barcelone Milan. 254 S.
- LINDER A. und BERCHTOLD W.**, 1982: Statistische Methoden III. Multivariate Verfahren. UTB 1189, Birkhäuser, Basel Boston Stuttgart. 218 S.
- LONDO G.**, 1975: Dezimalkala für die vegetationskundliche Aufnahme von Dauerquadraten. In: SCHMID W. (Hrsg.), Sukzessionsforschung. Cramer, Vaduz. 613-617.
- LOUPPEN J.M.W. und VAN DER MAAREL E.**, 1979: CLUSLA: A computer program for the clustering of large phytosociological data sets. *Vegetatio* **40**, 107-114.
- MORRISON D.F.**, 1976: Multivariate statistical methods. (2. Aufl.). McGraw-Hill Book Company, New York. 415 S.
- MUELLER-DOMBOIS D. und ELLENBERG H.**, 1974: Aims and Methods of Vegetation Ecology. John Wiley Sons, New York Chichester Brisbane Toronto. 547 S.
- NOY-MEIR I.**, 1973: Divisive polythetic classification of vegetation data by optimized division on ordination components. *J. Ecol.* **61**, 753-760.
- OBERGFOELL F.J.**, 1984: Trittbelaetzung auf Halbtrockenrasen im Ballungsraum Stuttgart und Möglichkeiten der Renaturierung. *Diss. Bot.* **76**. 169 S.

- ORLOCI L., 1966: Geometric models in ecology. *J.Ecol.* **54**, 193-215.
- ORLOCI L., 1973: Ranking characters by a dispersion criterion. *Nature* **244**, 371-373.
- ORLOCI L., 1976: Ranking species by an information criterion. *J.Ecol.* **64**, 417-419.
- ORLOCI L., 1978: Multivariate analysis in vegetation research. (2. Aufl.). Junk, The Hague. 451 S.
- ORLOCI L. und STANEK W., 1979: Vegetation survey of the Alaskan Highway, Yukon Territory: types and gradients. *Vegetatio* **41**, 1-56.
- OSTHEIDER M., 1978: Räumliche Stichprobenverfahren. *Veröffentlichungen der Geographischen Kommission der Schweizerischen Naturforschenden Gesellschaft* **5**, 29-54.
- PIELOU E.C., 1977: Mathematical ecology. (2. Aufl.). John Wiley and Sons, New York London Sidney Toronto. 385 S.
- PIELOU E.C., 1984: The interpretation of ecological data. John Wiley and Sons, New York. 263 S.
- PODANI J., 1984: Spatial processes in the analysis of vegetation: Theory and review. *Acta Bot. Hungarica* **30**, 75-118.
- POORE M.E.D., 1955: The use of phytosociological methods in ecological investigations. I-III. *J.Ecol.* **43**, 226-244, 245-269, 606-651.
- POORE M.E.D., 1962: The method of successive approximation in descriptive ecology. *Advances in Ecological Research* **1**, 35-68.
- RAJSKI C., 1961: Entropy and metric spaces. In: CHERRY C. (Hrsg.), *Information Theory*. London, Butterworths, 41-45.
- RENYI A., 1962: Wahrscheinlichkeitsrechnung mit einem Anhang über Informationstheorie. VEB Deutscher Verlag der Wissenschaften, Berlin. 547 S.
- SHANNON E., 1948: A mathematical theory of communication. *Bell Syst.Techn.J.* **27**, 379-423, 623-653.
- SNEATH, P.H.A. und SOKAL R.R., 1973: *Numerical Taxonomy*. Freeman, San Francisco.
- SOERENSEN T., 1948: A method of establishing groups of equal amplitude in plant sociology based on similarity of species content. *Biol.Skr., K.Danske Vidensk.Selsk.* **5**, 1-34.
- SOKAL R.R. und ROHLF F.J., 1962: The comparison of dendograms by objective methods. *Taxon* **11**, 33-40.

- STUART A., 1962: Basic Ideas of Scientific Sampling. Charles Griffin Co. Ltd., London. 99 S.
- SUCHARD-FICHER C., BACKHAUS K., HUMME U., LOHRBERG W., PLINKE W. und SCHREINER W., 1982: Multivariate Analysemethoden. (2. Aufl.). Springer Verlag Berlin Heidelberg New York. 346 S.
- SPAETH H., 1977: Cluster-Analyse-Algorithmen zur Objektklassifizierung und Datenreduktion. (2. Aufl.). Oldenbourg, München Wien. 217 S.
- SWAN J.M.A., DIX R.L. und WEHRHAHN C.F., 1969: An ordination technique based on the best possible stand-defined axes and its application to vegetation analysis. *Ecology* 50, 206-212.
- UEBERLA K., 1971: Faktorenanalyse. Springer, Berlin. 400 S.
- ULLMANN H., 1971: Biomasse und Dominanzgesellschaften in einem Hochmoor. *Ber.Dtsch.Bot.Ges.* 84, 637-647.
- VAN DER MAAREL E., 1969: On the use of ordination models in phytosociology. *Vegetatio* 19, 21-46.
- VAN DER MAAREL E., 1979: Transformation of cover-abundance values in phytosociology and its effects on community similarity. *Vegetatio* 39, 97-114.
- VAN DER MAAREL E., 1980: On the interpretability of ordination diagrams. *Vegetatio* 42, 43-45.
- VAN DER MAAREL E., JANSSEN J.G.M. und LOUPPEN J.M.W. , 1978: Tabord, a program for structuring phytosociological tables. *Vegetatio* 38, 143-156.
- WESTHOFF V. und VAN DER MAAREL E., 1978: The Braun-Blanquet approach. (2. Aufl.). In: WHITTAKER R.H. (Hrsg.) Classification of plant communities. Junk, The Hague. 287-399.
- WHITTAKER R.H., 1952: A study of summer foliage insect communities in the Great Smoky Mountains. *Ecol.Monogr.* 26, 1-80.
- WHITTAKER R.H., 1967: Gradient Analysis of Vegetation. *Biol.Rev.* 42, 207-264.
- WHITTAKER R. H., 1978: Direct gradient analysis. In: WHITTAKER R.H. (Hrsg.), Ordination of Plant Communities. Junk, The Hague. 7-50.
- WILDI O., 1977: Beschreibung exzentrischer Hochmoore mit Hilfe quantitativer Methoden. *Veröff.Geobot.Inst.ETH, Stiftung Rübel*, 60. 128 S.
- WILDI O., 1979: GRID - A space density analysis for recognition of nodes in vegetation samples. *Vegetatio* 41, 95-100.

- WILDI O.**, 1984: Species selection by interactive ranking. *Vegetatio* 56, 161 - 166.
- WILDI O. und ORLOCI L.**, 1983: Management and multivariate analysis of vegetation data. (2. überarb. Aufl.). Eidg.Anst.forstl.Versuchswes., Ber. 215. 139 S.
- WILLIAMS W.T.**, 1959: Multivariate methods in plant ecology. I. Association-analysis in plant communities. *J.Ecol.* 47, 83-101.
- WISHART D.**, 1978: Clustan users manual. (3. Aufl.). Edinburgh: Edinburgh University.
- YULE G. U.**, 1912: On the methods of measuring association between two attributes. *J.R.Stat.Soc.* 75, 579-642.
- ZOEHRRER F.**, 1980: Forstinventur. Parey, Hamburg Berlin. Pareys Studentexte 26. 207 S.