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

Kapitel: 3: Skalierung und Transformation

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: 14.07.2025

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

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,	^{o}C:	0,	10,	20
Kelvinskala,	^{o}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 "Bodentyp", 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

$Tb_1 + Tb_2$
$Tb_1 - Tb_2$
$Tb_1 * Tb_2$
Tb_1 / Tb_2

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 "≠" sind dabei zulässig, formal auch "<" und ">". Dabei besteht jedoch Mehrdeutigkeit bei 0^o und 360^o. 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

$Tg_1 = Tg_2$ (Gleiche Grundwasserstandskategorie für Aufnahme 1 und 2)
$Tg_1 \neq Tg_2$
$Tg_1 > Tg_2 > Tg_3$ ... 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

$Tg_1 = 15 = C$ (Rang 3)
$Tg_2 = 200 = A$ (Rang 1)

Die Berechnung des Mittelwertes erfordert die Verwendung arithmetrischer Grundoperationen:

$$\bar{T}g = 1/2(Tg_1+Tg_2) = 1/2(15+200) = 107.5 \quad (Typ\ A)$$

Mit den Rangzahlen wird ein falsches Resultat errechnet:

$$\bar{T}g = 1/2(1+3) = 2 \quad (Typ\ B)$$

Diese Ueberlegungen 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 Aehnlichkeitsmassen. 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

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

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

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

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.	Deckung, %	Ordinalskala 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 durchgeführte Analysen berücksichtigen speziell die dominanten Arten.

An dieser Stelle ist anzumerken, dass geordnete Vegetationstabellen 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 Standortsfaktoren 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}	$2.101*10^{-4}$	$5.25*10^{-5}$

Als Mittelwert der untransformierten H^+-Konzentration erhält man also $5.25*10^{-5}$ 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

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

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

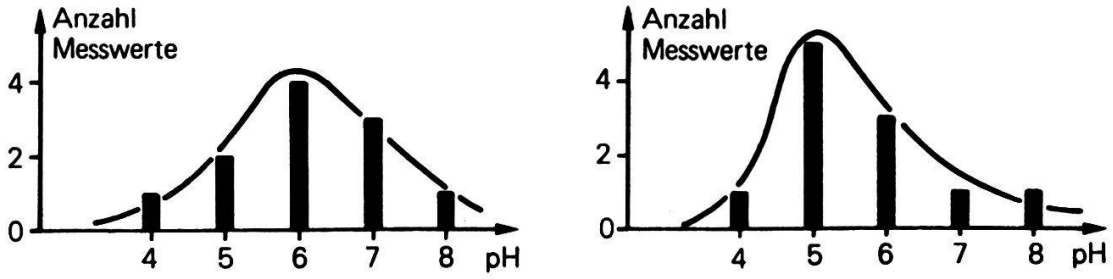


Abb. 3.1 Beispiel einer angenähert symmetrischen Verteilung eines Standortsfaktors (links), sowie einer schiefen Verteilung (rechts).

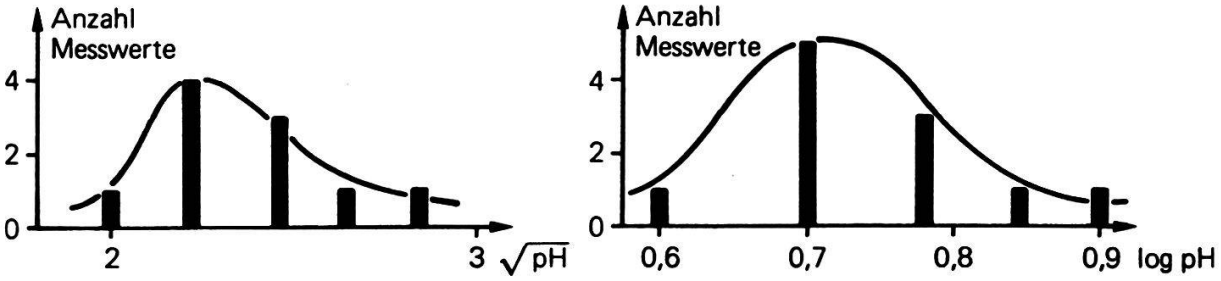


Abb. 3.2 Histogramme der transformierten Messwerte, welche zur Konstruktion der Abb. 3.1, rechts verwendet wurden. Histogramm links: Wurzeltransformation. Histogramm rechts: Logarithmustransformation.

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
Grundwasser-stand 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

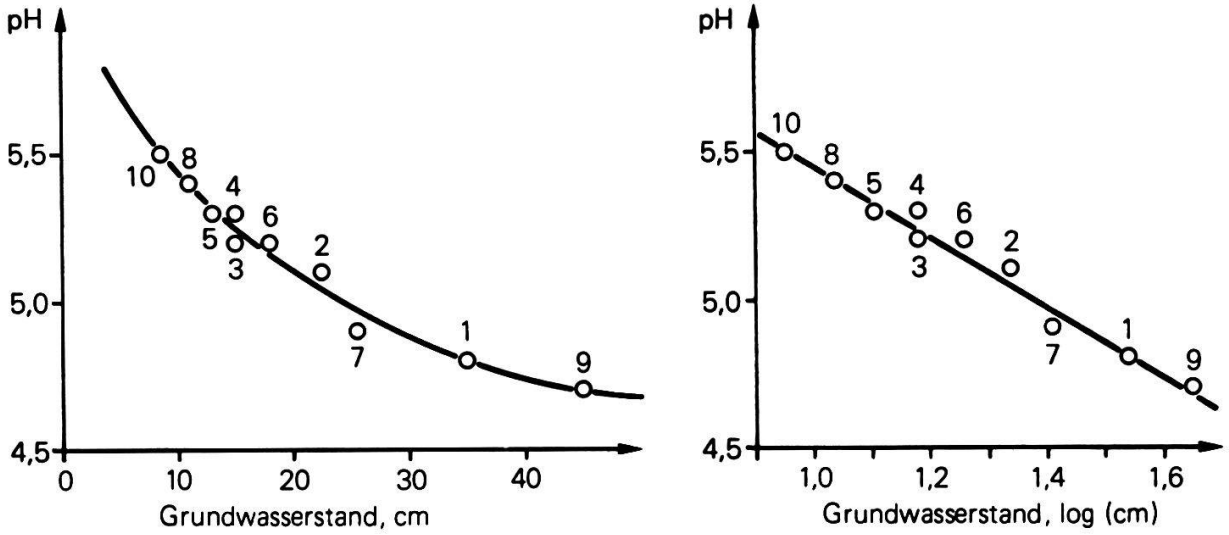


Abb. 3.3 Direkte Ordination der Aufnahmen in Tabelle 3.2.: Untransformiert (links), transformiert (rechts).

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 grössere 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 artenarmer 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:

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

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.

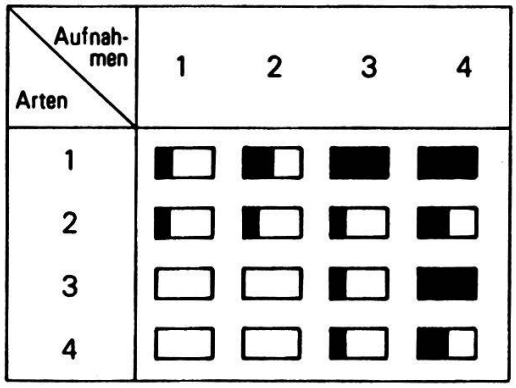
Schliesslich findet sich in Abb. 3.4 auch die Wirkung der Standardisierung der Aufnahmen (D) und der Arten (E). Man kann folgende Phänomene festhalten:

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

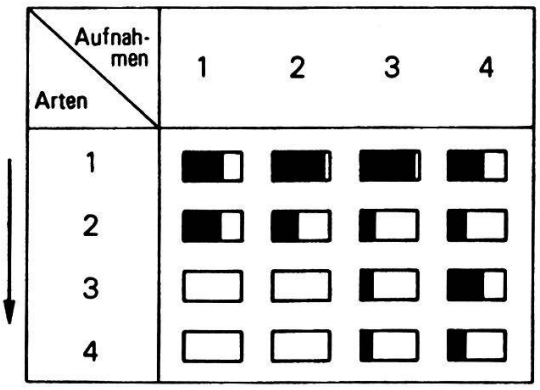
A

Arten \ Aufnahmen	1	2	3	4
1	1	2	4	4
2	1	1	1	2
3	0	0	1	4
4	0	0	1	2

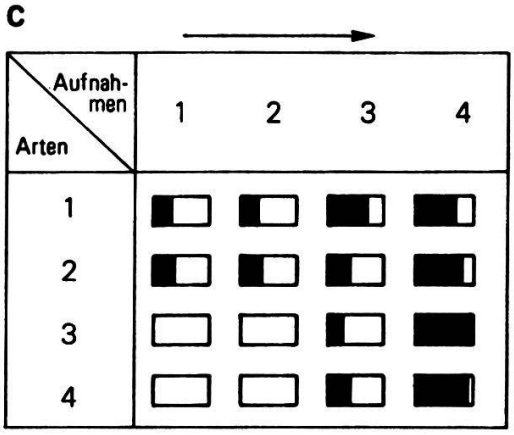
A



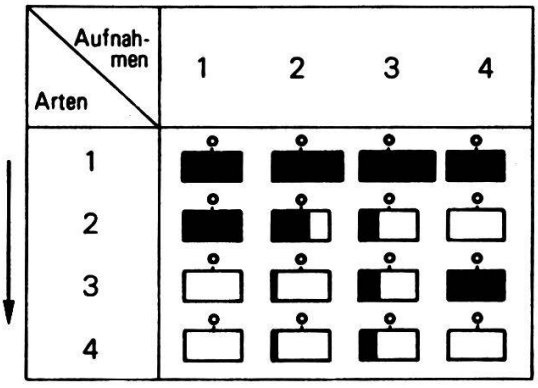
B



C



D



E

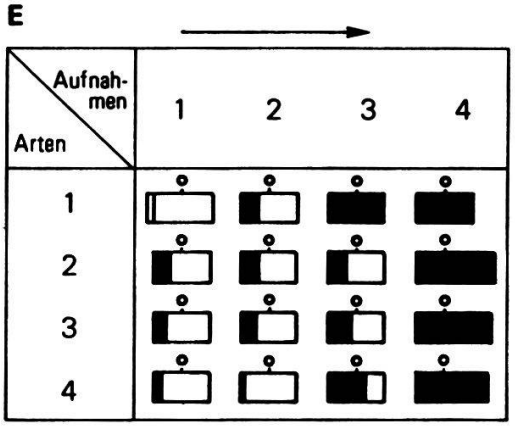


Abb. 3.4 Grafische Darstellung der Wirkung verschiedener Vektortransformationen auf eine Vegetationstabelle A. B: Normalisierung der Aufnahmen, C: der Arten. D: Standardisierung der Aufnahmen, E: der Arten. Die Skalierung der Grafiken ist willkürlich gewählt. Bei D und E treten auch negative Werte auf.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

2. Die Daten verändern sich noch stärker als bei der ersteren, 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 Aehnlichkeitsstruktur 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 Aehnlichkeitsverhältnisse der Aufnahmen drastisch und seltene Arten prägen die Ergebnisse weitgehend (Beispiel E in Abb. 3.4).