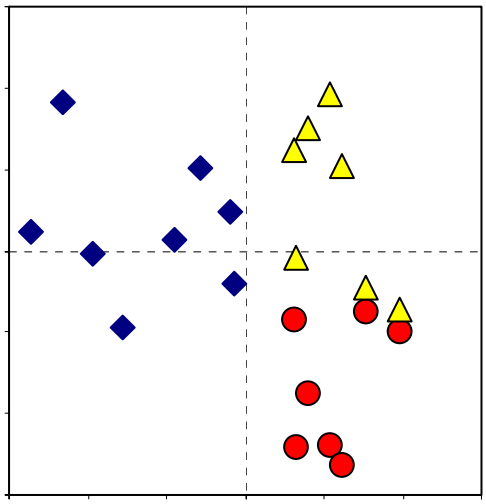
Multivariate Methoden in der Vegetationskunde
-
Kursskript

Michael Rudner & Hiltrud Brose
Fakultät für Biologie, Abt. Geobotanik

Oktober 2007
Albert-Ludwigs-Universität Freiburg

Inhaltsverzeichnis

EINLEITUNG



Ziel des Kurses:

Verständnis von Fachartikeln, Kenntnisse und Fertigkeiten in der Anwendung multivariater Methoden

Ziel des Skripts:

Kurze Übersicht über wesentliche Themen und Methoden der multivariaten Verfahren in der Vegetationskunde, Hinweise zur Umsetzung in R, komplementär zum R-Skript

Aufbau des Skripts

Das Skript handelt eine Auswahl wesentlicher Verfahren und Aspekte der multivariaten Methoden ab. Es stützt sich i.W. auf die im Literaturverzeichnis genannten Lehrbücher und Skripten. Jedes Thema ist auf einer Seite steckbriefartig zusammengefasst. Der Aufbau folgt soweit möglich einer einheitlichen Struktur:
Gebräuchliche Bezeichnung | Typische Grafik | Was passiert? | Was sagt das Ergebnis aus? | Eignung | Umsetzung in R | Wesentliche Angaben

Weitere Lehrmaterialien?

Zur Umsetzung der multivariaten Analysen in R gibt es ein Kursskript von Arne Saatkamp, Hiltrud Brose und Michael Rudner (Geobotanik, Universität Freiburg, 2007).
Derzeit werden E-Learning-Module entwickelt, die in WebKit an der Universität Freiburg umgesetzt werden. Sie werden Studierenden der Biologie über CampusOnline zur Verfügung stehen. In diesen Modulen werden ebenfalls wesentliche Aspekte und Verfahren der multivariaten Methoden behandelt. Sie sollen kursbegleitend zum Einsatz kommen, eignen sich aber auch zum Selbststudium.

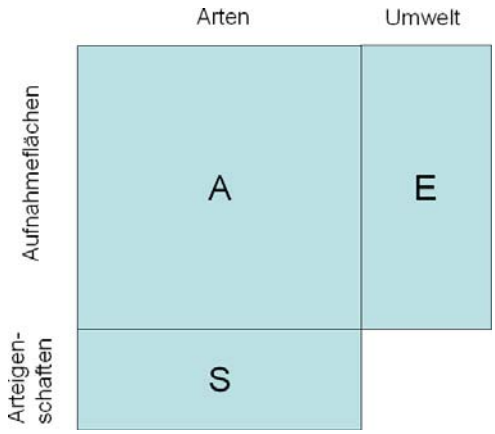
Eignung des Skripts

Das Skript ist als unterstützendes Material für den Kurs zu den multivariaten Methoden gedacht. Auch soll es eine rasche Orientierung bieten, um bei vorhandenen Kenntnissen den Einstieg ins Thema zu erleichtern, wenn Auswertungen eigener Daten anstehen. Für eingehendere Studien verweisen wir auf die Lehrbücher, die in den Literaturhinweisen genannt sind.

Umsetzung in R

In R sind mehrere Pakete für multivariate Statistik geeignet. Soweit möglich, werden Funktionen aus dem Paket `vegan` genutzt. Daneben sind die Pakete `ade4` und `cluster` relevant. Manche Funktionen sind auch im Standardpaket `stats` implementiert. Für einige Funktionen werden intern weitere Pakete aufgerufen (`MASS, akima, mgcv`).

MULTIVARIATE STATISTIK



Ziel:
Herausarbeiten von Mustern in der Vegetation und von Zusammenhängen zwischen Standort und Vegetation in unterschiedlichen Abstraktionsniveaus (Vegetationstypen, funktionale Pflanzentypen,…)
Zusammenfassen mehrerer Umweltvariablen zu einer oder wenigen synthetischen Variablen.

Was passiert?
In der Regel wird eine Distanzmatrix der Objekte (Aufnahmeflächen) erstellt. Auf dieser Grundlage wird dann entweder eine Klassifikation oder eine Ordination vorgenommen. Ordinationsverfahren erzeugen synthetische Achsen, die Gradienten im Datensatz repräsentieren. Zur Interpretation werden weitere Variablen, z.B. Umweltvariablen herangezogen - bei der direkten Gradientenanalyse werden diese in die Analyse integriert.

Wann sind multivariate Verfahren anzuwenden?
Multivariate Verfahren sind geeignet, Vegetationsmuster zu analysieren insbesondere, wenn Artengemeinschaften betrachtet werden. Sie eignen sich auch als Voruntersuchung, um die wesentlichen Faktoren für eine quantitative Analyse der Art-Habitatbeziehungen herauszufinden (siehe auch Tabelle 1).

Tab. 1: Übersicht zu statistischen Methoden in Abhängigkeit von der Fragestellung

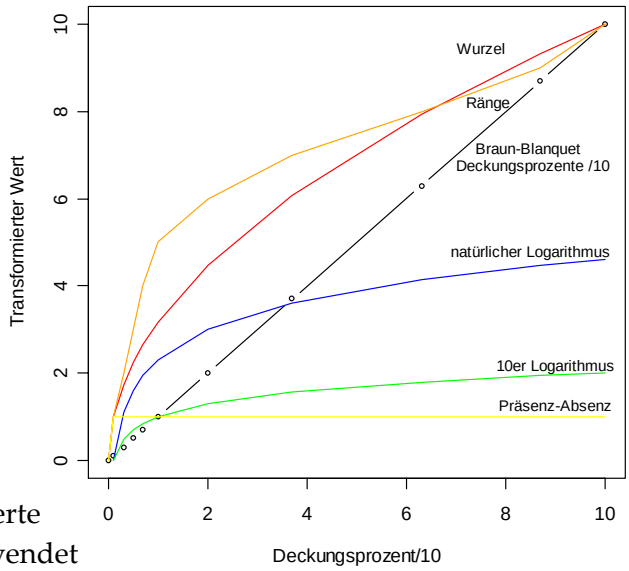
Orientierung	**Einzelart**	**Habitat**
Ziel	Vorhersage (Habitateignung, Vorkommen,…)	Beschreibung der Variabilität vorhandener Habitate, dann Positionieren der Zielart in diesem Raum
Analysemethoden	Logistische Regression Multiple Regression Diskriminanzanalyse	Extraktion primärer Gradienten durch Ordination oder Klassifikation
Vorteile	Bessere Prognosegüte	Bessere Beschreibung der allgemeinen Variabilität der Habitate Anwendbar für eine größere Anzahl von Arten
Nachteile	Keine Beschreibung der allgemeinen Variabilität der Habitate Nicht auf andere Arten übertragbar	Schlechte Vorhersagegüte für Einzelarten

Wesentliche Richtungen

Klassifikation	Reduzieren auf eine Dimension, die Klassenzugehörigkeit
Indirekte Gradientenanalyse:	Projektion auf wenige Dimensionen (Ordinationsachsen)
Direkte Gradientenanalyse:	Projektion auf wenige Dimensionen, die durch Kombinationen der Umweltvariablen bestimmt sind (Kanonische Ordinationsachsen)

Kombinationen der Ergebnisse von Klassifikation und Ordination

TRANSFORMATION



Ziel:

- Verbesserung der Annahmen zu Verteilung, Linearität usw.
- Vergleichbarkeit der Einheiten unterschiedlicher Variablen
- Optimieren bzgl. der Distanzmaße
- Verringerung des Einflusses absoluter Mengen
- Verändern des Gewichts verbreiteter und seltener Arten

Typen:

Monotone Transformation wird auf alle Werte unabhängig von den anderen Werten angewendet und hat damit keine Änderung des Rangs zur Folge.
Relativierung wird zeilen- und spaltenweise angewendet und hängt von den anderen Werten ab (z.B. Mittel- oder Maximalwert). Als erster Schritt gehört zu den Transformationen auch das „**Code Replacement**“ dazu, das das Ersetzen der Abundanzklassen durch numerische Werte beschreibt.

Was passiert?

Die monotonen Transformationen zielen darauf ab, die hohen Abundanzwerte in ihrem Gewicht abzusenken, um die Analyse nicht zu stark auf Dominanz zu stützen, sondern die floristische Ähnlichkeit, die durch das gemeinsame Auftreten von Arten bestimmt ist, weiter in den Vordergrund zu rücken. Dies ist bei Präsenz-Absenz-Werten in vollem Maße gegeben. Normierungen führen dazu, den Wertebereich verschiedener Variablen anzupassen und diese somit gleich zu gewichten. Standardisierungen versuchen, die Verteilungen einer Normalverteilung anzunähern.

Monotone Transformationen

Wurzeltransformation, logarithmische Transformation (x'=log(x+b), da log(0) nicht definiert), Präsenz-Absenz-Transformation, Arcussinus-Wurzel-Transformation ($x' = 2/\pi \cdot \arcsin\left(\sqrt{x_{ij}}\right)$)

Relativierungen

- Normieren (Teilen durch das Maximum) -> Wertebereich wird auf [0;1] eingeschränkt
- Standardisieren -> Mittelwert 0 und Standardabweichung 1
- Ränge -> Abbilden artspezifischen Auftretens
- Doppelte Relativierung: Maximum der Art-Abundanz, Summe der Aufnahme

Eignung / Anforderungen an die Daten

Manchmal erfordern bestimmte Verfahren eine Transformation vorab oder aber sie ist in der Analyse implementiert. Beachten Sie in jedem Fall ob die Abundanzen etwa mit der Klassenmitte in Deckungsprozentwerten oder mit einer Rangstufe der Abundanzklasse vertreten sind (*Code Replacement*)

Befehle in R

Paket vegan: `decostand()` mit dem Parameter „`standardize`“, „`range`“, „`pa`“;
Funktionen: Wurzel `sqrt()`; Arcussinus-Wurzel `asin(sqrt(x/xmax))*2/pi`

DISTANZMAßE

Ziel:
Berechnung der Unähnlichkeit (oder Ähnlichkeit) zwischen den Objekten (z.B. Vegetationsaufnahmen) eines Datensatzes, paarweise für jedes Objektpaar. Die Berechnung einer Distanzmatrix ist der erste Schritt bei vielen multivariaten Methoden.

	1	2	3	4	5
1	0				
2	0,08	0			
3	0,25	0,32	0		
4	0,62	0,64	0,70	0	
5	0,55	0,58	0,64	0,18	0

Ähnlichkeitsmaße (S für „similarity“)
Jaccard-Ähnlichkeitsmaß: S = a/(a+b+c)
Sørensen-Ähnlichkeitsmaß: S = 2a/(2a+b+c)
(a ist die Zahl der Arten, die in beiden Aufnahmen vorkommen, b und c die Zahl der Arten, die in je einer der beiden Aufnahmen vorkommen. Die Zahl der Arten die in keiner der beiden Aufnahmen vorkommt (d), wird nicht berücksichtigt.)

→ Jaccard- und Sørensen-Ähnlichkeismaß berücksichtigen nur An- und Abwesenheit der Arten, sind also qualtitativ.
→ Umrechnung in ein Distanzmaß z.B. über D = 1 - S

Distanzmaße (D für „dissimilarity“)

Bray-Curtis-Distanz: $$D_{i,h} = \frac{\sum_{j=1}^{p} \left| a_{i,j} - a_{h,j} \right|}{\sum_{j=1}^{p} \left(a_{i,j} + a_{h,j} \right)}$$

→ Quantitative Form des Sørensen-Ähnlichkeitsmaß. Doppel-Nullen werden nicht berücksichtigt.

Euklidische Distanz: $$D_{i,h} = \sqrt{\sum_{j=1}^{p} \left(a_{i,j} - a_{h,j} \right)^2}$$

→ Beruht auf dem Satz des Pythagoras. Berücksichtigt Doppel-Nullen, und kann daher zu Problemen führen, wenn viele Doppel-Nullen im Datensatz auftauchen. Es sind verschiedene Abwandlungen der Euklidischen Distanz vorgeschlagen worden, um deren Nachteile zu umgehen, etwa die Chord-Distanz (Legendre & Legndre 1998, S. 278f.)

Manhattan-Distanz (= City-block-Distanz)**:** $$D_{i,h} = \sum_{j=1}^{p} \left| a_{i,j} - a_{h,j} \right|$$

→ Vorstellbar als Wegstrecke, wenn man mit einem Taxi durch Manhattan fahren würde. Berücksichtigt ebenfalls Doppel-Nullen.

Eignung / Anforderungen an die Daten
Die Wahl des Distanzmaßes richtet sich nach den vorliegenden Daten. Eine wichtige Frage ist, ob man ein qualitatives oder ein quantitatives Distanzmaß verwenden möchte. Ein weiterer wichtiger Punkt ist die Frage, wie das Distanzmaß mit Doppel-Nullen umgeht. Manche multivariate Verfahren erfordern die Anwendung eines bestimmten Distanzmaßes.

Befehle in R
Paket vegan: `vegdist()`, Paket stats: `dist()`

KLASSIFIKATION

Was passiert?

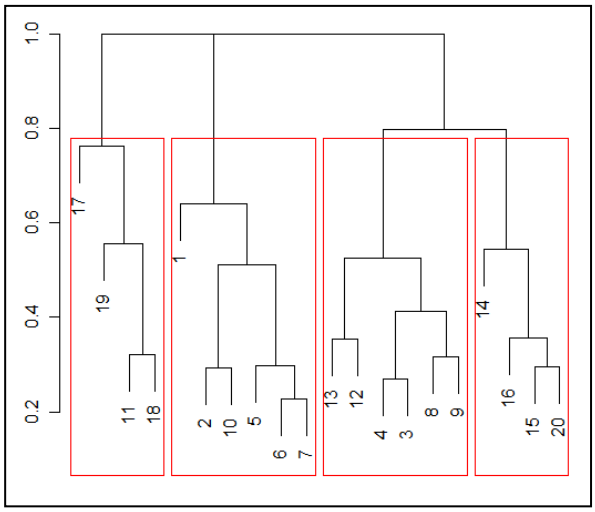
Objekte wie z.B. Vegetationsaufnahmen werden aufgrund ihrer Ähnlichkeit bzw.-Unähnlichkeit in Gruppen eingeteilt.
Es kann zwischen verschiedenen Klassifikationsmethoden unterschieden werden: Bei den **nicht-hierarchischen Methoden** werden Gruppen gebildet, ohne dass diese in Beziehung zueinander gesetzt werden, bei den **hierarchischen Methoden** dagegen wird als Ergebnis auch eine hierarchische Struktur angegeben. Die Gruppen lassen sich also zueinander in Beziehung setzen, so werden kleinere Gruppen zu größeren zusammengefügt, diese wiederum zu noch größeren usw.
Dabei gibt es zwei Möglichkeiten: Entweder werden am Anfang alle Objekte einzeln betrachtet, und dann schrittweise zu Gruppen zusammengefasst (**agglomerative Verfahren**), oder alle Objekte werden als zu einer einzigen großen Gruppe gehörend betrachtet, die dann nach und nach in kleinere Gruppen aufteilt wird (**divisive Verfahren**).
Hierbei gibt es zwei Arten, die Gruppe zu unterteilen: Man kann nur ein Merkmal berücksichtigen (**monothetische Verfahren**), oder bei der Unterteilung mehrere Merkmale berücksichtigen (**polythetische Verfahren**).

Wesentliche Überlegungen

Welches Klassifikationsverfahren bildet das Ergebnis der ökologischen Prozesse am besten ab, das in meinem Datensatz repräsentiert ist?

AGGLOMERATIVE KLASSIFIKATIONSVERFAHREN

Was passiert?



Bei agglomerativen Klassifikationsverfahren werden zunächst alle Aufnahmen (Objekte) einzeln betrachtet und dann schrittweise zu Gruppen (Clustern) zusammengefasst, die wiederum zu Gruppen höherer Ordnung zusammengefasst werden können. Diese Klassifikationen sind also hierarchisch.
Wie genau die Objekte zu Gruppen zusammengefasst werden, hängt vom gewählten Cluster-Algorithmus ab.
Als Ergebnis wird ein Dendrogramm der Objekte dargestellt.
Die Frage, in wie viele Gruppen sich der Datensatz teilen lässt, ist v. a. von der Fragestellung abhängig und wird meist auf der Basis externer Informationen (z.B. Vergleich mit pflanzensoziologischen Gruppen, Umweltvariablen) getroffen.

Was sagt das Ergebnis aus?

Das Dendrogramm spiegelt die Abstände der Aufnahmen (Objekte) zueinander wider. Die Länge der Äste ist ein Maß für die Unähnlichkeit der Objekte. In der Abbildung oben sind die Aufnahmen 4 und 20 ebenso unähnlich wie die Aufnahmen 4 und 14. Die Reihenfolge der Objekte spielt keine Rolle, vielmehr kann man sich ein Dendrogramm als frei um die Verzweigungspunkte bewegliches Mobile vorstellen.

Eignung / Anforderungen an die Daten

Es stellt sich immer die Frage, ob eine Klassifikation sinnvoll ist, ob im vorliegenden Datensatz eine klare Gruppenstruktur zu erwarten ist, oder ob eher graduelle Übergänge vorliegen.

Befehle in R

`hclust()`, Dendrogramm mit `plot()`, Gruppen einzeichnen mit `rect.hclust()`
Paket cluster: `agnes()`

Wesentliche Angaben

- Datensatz (welche Objekte, welche Eigenschaften, welcher Datentyp, welche Skala)
- Transformation des Datensatzes
- Wahl des Distanzmaßes
- Wahl des Cluster-Algorithmus
- Dendrogramm
- Skala, die angibt, wie das Dendrogramm skaliert wurde
- Interpretation des Dendrogramms (wie wurden die Gruppen gebildet?)

CLUSTER-ALGORITHMEN

Cluster-Algorithmen sind Rechenvorschriften, die angeben, nach welchen Regeln Aufnahmen (Objekte) zu Gruppen (Clustern) zusammengefasst werden. Bei allen Algorithmen werden zunächst die ähnlichsten Aufnahmepaare zusammengefasst.

Single linkage (nearest neighbour)
→ die Zuordnung zu einer Gruppe wird durch die beiden ähnlichsten Aufnahmen (Objekte) bestimmt.
Das single-linkage Verfahren neigt zur Kettenbildung („chaining“).

```
hclust(distanzmatrix,“single“)
```

Complete linkage (farthest neighbour)
→ die Zuordnung zu einer Gruppe wird durch die beiden unähnlichsten Aufnahmen (Objekte) bestimmt.
Beim complete-linkage werden klare Cluster ähnlicher Größe gebildet, auch wenn im Ausgangsdatensatz eher kontinuierliche Übergänge vorliegen.

```
hclust(distanzmatrix,“complete“)
```

Average linkage
→ die Zuordnung zu einer Gruppe wird durch den Mittelwert aller wechselseitigen Abstände (alle Objekte der einen zu allen Objekten der anderen Gruppe) bestimmt.

```
hclust(distanzmatrix,“average“)
```

Centroid linkage
→ die Zuordnung zu einer Gruppe wird durch den Abstand zwischen den Schwerpunkten zweier Gruppen bestimmt.
Beim centroid linkage treten gelegentlich reversals auf.

```
hclust(distanzmatrix,“centroid“)
```

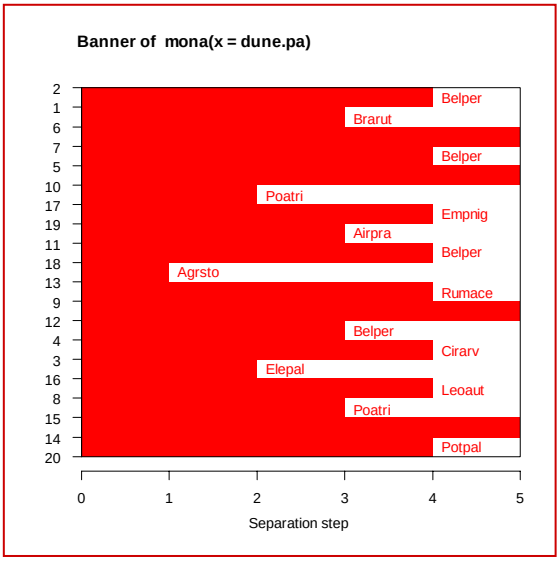
Minimum variance Methode (Wards-Methode)
→ Gruppen werden so gebildet, dass die Summe der Abweichungsquadrate innerhalb eines Clusters möglichst gering bleibt.
Im Dendrogramm werden die quadrierten Distanzen dargestellt, daher erscheinen die Gruppen klarer. Als Distanzmaß sollte die Euklidische Distanz verwendet werden.

```
hclust(distanzmatrix,“ward“)
```

DIVISIVE KLASSIFIKATIONSVERFAHREN

Divisive clustering

Monothetische und polythetische Verfahren



Was passiert?

Beginnend mit einem großen Cluster wird immer weiter aufgespaltet. Es wird jeweils ein geeignetes Kriterium (Art) zur Auftrennung gesucht. Zum Beispiel über eine χ^2-Statistik für alle Artenpaare. Die Art mit der höchsten Summe dieser Statistik wird zum Aufteilen in zwei Gruppen verwendet (mit und ohne die Indikatorart). Die Schritte werden für die einzelnen Gruppen wiederholt.

Was sagt das Ergebnis aus?

Im Bannerplot werden die Aufteilung und das Kriterium dazu dargestellt. Das Ergebnis kann meist auch im Dendrogramm dargestellt werden. Für beide Grafiken gilt, dass die Anordnung an jeder Verzweigung frei gedreht werden kann, ohne die Aussage zu verändern, d.h. nebeneinander liegende Aufnahmen sind nicht zwangsläufig ähnlicher als weiter enternt zu liegen kommende. Entscheidend ist immer, ob sie bei einem bestimmten Trennschritt zum gleichen Cluster gehören. Die Güte des Ergebnisses kann über den *Division Coefficient* beurteilt werden. Letzterer ist jedoch auch von der Größe des Datensatzes abhängig.

Eignung / Anforderungen an die Daten

Als Distanzmaß wird die Euklidische Distanz oder die Manhattan-Distanz verwendet. Beim Einlesen einer Distanzmatrix können auch andere Maße, etwa die Bray-Curtis-Distanz, verwendet werden. Der Einfluss der ersten Teilung auf das Ergebnis ist sehr groß. Daher sollten unbedingt Ergebnisse mit unterschiedlichen Distanzmaßen oder Datentransformationen verglichen werden.

Befehle in R

Paket cluster: `mona()` mit oben beschriebenem Algorithmus, kann auch mit Distanzmatrix aufgerufen werden. Oder `diana()`das einen anderen Algorithmus verwendet (Wahl des größten Clusters, Wahl der unpassendsten Aufnahme, Verteilung der Aufnahmen auf altes Cluster und neue „Splittergruppe"...). Das Ergebnis kann mit `as.hclust()` zur weiteren Bearbeitung in das Format des Ergebnisses der Funktion `hclust()` umformatiert werden.

Wesentliche Angaben

Datensatz (welche Objekte, welche Eigenschaften, welcher Datentyp, welche Skala?)
Transformation des Datensatzes, Distanzmaß
Teilungskriterien, Begründung für die Zahl der berücksichtigten Cluster.
Interpretation des Dendrogramms oder des Bannerplots.
Begründung für die Wahl der Methode

HIERARCHISCH DIVISIVES CLUSTERN MIT INDIKATORARTEN TWINSPAN

Two Way Indicator Species Analysis

Kurzbezeichnung: TWINSPAN

Originaldatensatz				**TWINSPAN** (Cut levels 0,5,25,50)		
Fagus sylvatica	4 (63%)	3 (37%)		Fagu sylv1 Fagu sylv2 Fagu sylv3 Fagu sylv4	1 1 1 1	1 1 1 0
Quercus petraea	1 (5 %)	2 (15%)		Quer petr1 Quer petr2	1 0	1 1
Hieracium laevigatum	0 (0%)	+ (1%)		Hier laev 1	0	1

Was passiert?

Der Datensatz wird einer Ordination (*reciprocal averaging* vgl. CA) unterzogen. Nach einer Reskalierung wird am 0-Punkt der ersten Ordinationsachse geteilt. Für jedes neue Cluster wird diese Prozedur erneut angewendet, bis entweder die Gruppe so klein ist, dass sie nicht weiter unterteilt werden kann oder eine maximale Reihung von Teilungen in einem Cluster erreicht ist. Es handelt sich um ein polythetisches, ordinationsbasiertes Verfahren.

Was sagt das Ergebnis aus?

Das Ergebnis wird üblicherweise als Vegetationstabelle dargestellt. In dieser Tabelle sind die Unterteilungen ersichtlich. In einer Lösungsdatei werden die Trennungskriterien aufgelistet. Entgegen der üblichen, relativ freien Anordnung eines Dendrogramms (vgl. Mobile) werden bei TWINSPAN die Aufnahmen und Arten möglichst nach Ähnlichkeit platziert, um eine diagonale Anordnung der Tabelle zu erreichen.

Eignung / Anforderungen an die Daten

Das TWINSPAN-Verfahren ist insbesondere dann geeignet, einen Vegetationsdatensatz zu klassifizieren, wenn ein starker Gradient als treibende Kraft vermutet wird. Der Einfluss zweier Gradienten kann mit TWINSPAN nicht besonders gut herausgearbeitet werden.
Die Berücksichtigung quantitativer Werte erfolgt über die „*Pseudospecies*". Dazu müssen passend zum Datensatz, üblicherweise Vegetationsdaten in Prozentwerten, Schwellenwerte für die Unterteilung in Pseudoarten gesetzt werden.

Befehle in R

Bislang in R nicht verfügbar. Jedoch im Tabellensortierprogramm JUICE integriert und auch als freies Windows-Programm erhältlich..

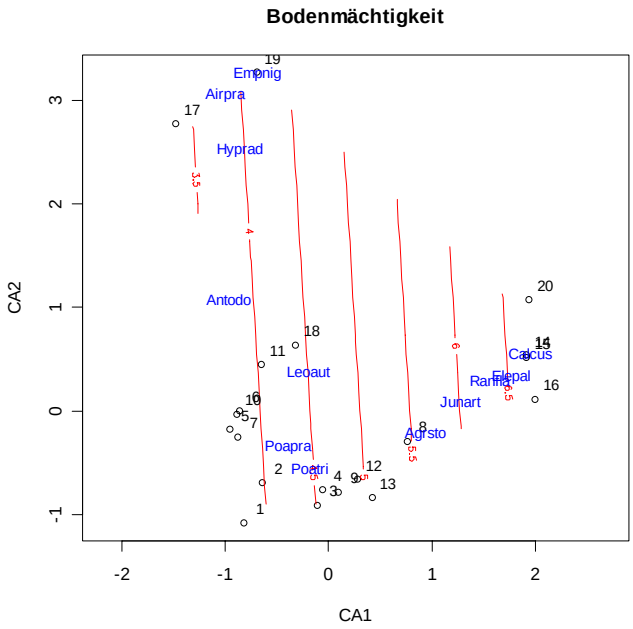
Wesentliche Angaben

Datensatz (welche Objekte, welche Eigenschaften, welcher Datentyp, welche Skala?)
Pseudospecies cut levels, minimum group size, maximum level of divisions
Teilungskriterien, Begründung für die Zahl der berücksichtigten Cluster.
Interpretation des Dendrogramms oder der Vegetationstabelle.
Begründung für die Wahl der Methode

INDIREKTE GRADIENTENANALYSE

Was passiert?

Es wird angestrebt, einen Datensatz so auf möglichst wenige Dimensionen (Achsen) zu projizieren, dass ein möglichst großer Anteil der Variabilität im Datensatz abgebildet wird und sozusagen die stärksten Muster sichtbar werden.



Was sagt das Ergebnis aus?

Eine erste Information über die Bedeutung der abgebildeten Muster gibt der Anteil an Varianz der gezeigten Dimensionen. Dieser ist auch vom Rauschen im Datensatz abhängig. Die Interpretation der Biplots ist abhängig vom verwendeten Verfahren (siehe dort).

Interpretationshilfen

Mit Overlays (Vektoren, Symbolgröße, -farbe und –form, Beschriftung, Isolinien) kann die Ausprägung einzelner Variablen (Umweltvariablen, Art-Abundanzen, Eigenschaften) an den Objekten (Aufnahmen, Proben,…) grafisch im Biplot dargestellt werden.

Wesentliche Überlegungen

Welche Transformation?	Gewichtung innerhalb der Variablen und Vergleichbarkeit der Variablen
Welches Distanzmaß?	Vermeidung des Doppelnull-Problems und Anforderungen der Verfahren
Reaktion der Arten auf den Gradienten?	Linear oder unimodal?
Welches Verfahren?	Polare Ordination, PCA, CA, DCA, NMDS?

Auswahl von Arten bzw. Eigenschaften im Biplot?

Skalierung des Ergebnisses?

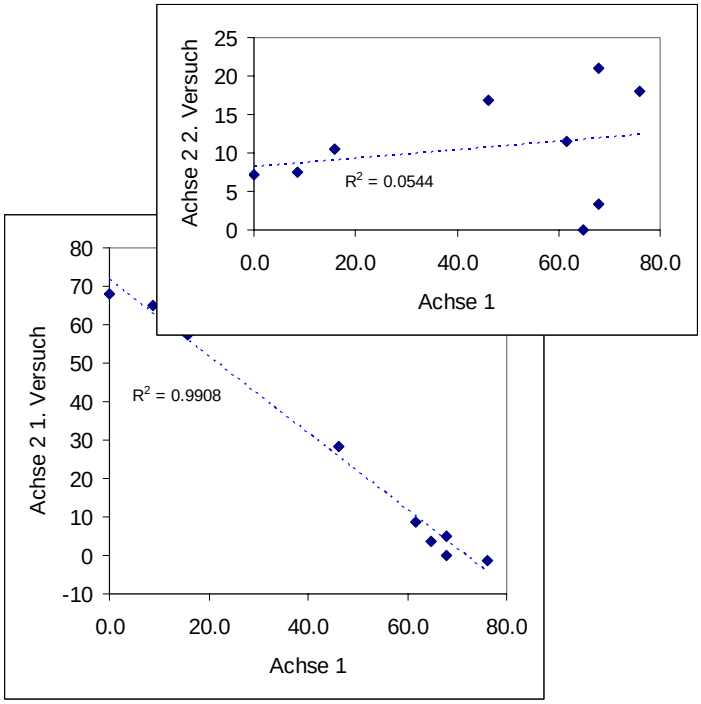
Wie viele Achsen werden interpretiert?

Sollen Ergebnisse einer Klassifikation und der Ordination kombiniert werden?

POLARE ORDINATION PO

Bray-Curtis ordination

Kurzbezeichnung: PO



Was passiert?
Aufspannen der ersten Ordinationsachse zwischen den beiden Objekten mit der größten Distanz zueinander. (Besser: Aufnahme mit der größten Varianz in der Distanz als 1. Endpunkt wählen). Berechnung der Achsenwerte für die anderen Objekte nach der Formel $(D^2+D1^2-D2^2)/2D$. Suche nach einem weiteren Objektpaar, das möglichst weit voneinander entfernt sein soll, um die 2. Achse aufzuspannen. Diese soll jedoch so wenig wie möglich mit der 1. Achse korreliert sein. I.d.R. sind mehrere Versuche notwendig.

Was sagt das Ergebnis aus?
In der Ordinationsgrafik werden nur die Aufnahmen (Objekte) dargestellt. Ggf. können wie bei NMDS die gewichteten Mittel der Arten (Zentroide) in der Grafik dargestellt werden. Die relative Lage der Objekte zueinander spiegelt die Distanzen wider.

Eignung / Anforderungen an die Daten
Die Ordination wird aufgrund einer Distanzmatrix erstellt. Dafür kann ein Maß verwendet werden, das Probleme wie z.B. das Doppelnull-Problem umschifft. Darüber hinaus gibt es keine speziellen Anforderungen. Die Methode ist leistungsfähig, wird aber wenig verwendet.

Befehle in R
Bray-Curtis-Ordination ist derzeit in R nicht implementiert.
(Nähere Beschreibung in McCune, Grace & Urban 2002, S. 143ff.)

Wesentliche Angaben
Datensatz (welche Objekte, welche Eigenschaften, welcher Datentyp, welche Skala?)
Transformation des Datensatzes
Distanzmaß
Gesamtvarianz, Anteil der Varianz, die auf den Achsen abgebildet ist,
Korrelation der Achsen untereinander
Welche Achsen werden interpretiert?
Interpretation der Ordinationsplots, Angaben von Interpretationshilfen (z.B. Overlays)
Begründung für die Wahl der Methode

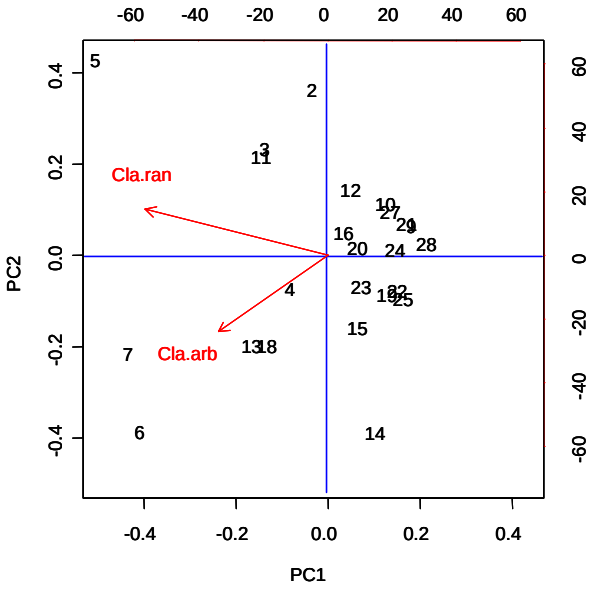
HAUPTKOMPONENTENANALYSE PCA

Principal components analysis

Kurzbezeichnung: PCA

Was passiert?

Rotation der mehrdimensionalen Datenwolke und Anordnung der ersten Ordinationsachse (1. Hauptkomponente) in Richtung der größten Ausdehnung der Punktwolke und damit Abbildung eines maximalen Varianzanteils auf der ersten Achse
Extraktion weiterer Achsen nach dem gleichen Prinzip mit der Bedingung, dass die Achsen voneinander linear unabhängig sind



Was sagt das Ergebnis aus?

Im Biplot werden Aufnahmen (Objekte) und Arten (Eigenschaften) zugleich dargestellt. In den Aufnahmen nimmt die Abundanz der Arten in Richtung des Pfeils zu (für Vergleich von Aufnahmen Senkrechte auf Pfeil fällen)
Güte des Ergebnisses kann über den Anteil durch die Achsen erklärter Varianz bemessen werden.

Eignung / Anforderungen an die Daten

PCA fußt auf der Annahme linearer Zusammenhänge. Das heißt, sie kann nur monotone Zunahmen oder Abnahmen der Arten entlang eines Umweltgradienten abbilden. Als Distanzmaß wird die Euklidische Distanz verwendet – in Datensätzen mit vielen gemeinsamen Nullstellen kann das problematisch sein. Liegt ein sehr starker Gradient zu Grunde, kann der Hufeisen-Effekt auftreten, d.h. die Aufnahmen werden in einem Bogen angeordnet und die Enden sind nach innen gebogen. Streng genommen müssen die Variablen multinormal verteilt sein. Es sollen mehr Objekte als Eigenschaften vorhanden sein.

Befehle in R

Paket vegan: `rda()` mit nur einer Tabelle. Paket stats: `princomp()`. Paket ade4: `dudi.pca()`

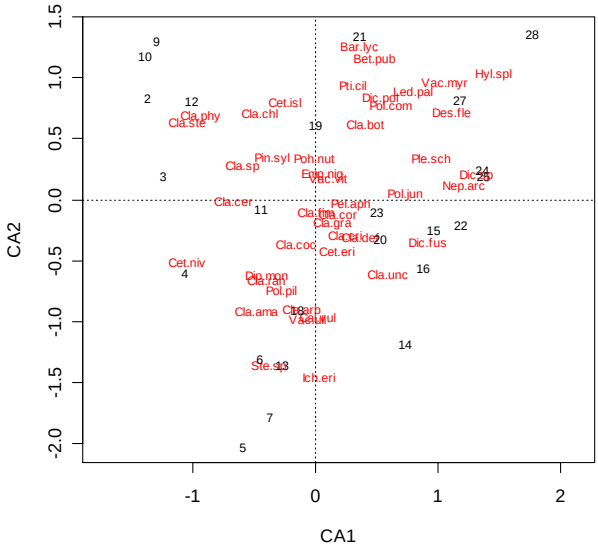
Wesentliche Angaben

Datensatz (welche Objekte, welche Eigenschaften, welcher Datentyp, welche Skala?)
Transformation des Datensatzes
Verwendung einer Varianz-Kovarianz-Matrix oder einer Korrelationsmatrix
Gesamtvarianz, Anteil der Varianz, die auf den Achsen abgebildet ist
Welche Achsen werden interpretiert?
Ggf. Rotation
Interpretation der Ordinationsplots, Angaben von Interpretationshilfen (z.B. Overlays)
Begründung für die Wahl der Methode

KORRESPONDENZANALYSE CA

Correspondence analysis

Kurzbezeichnung: CA



Was passiert?

Achsen werden zugleich im Arten-Raum und im Aufnahmen-Raum rotiert mit dem Ziel eine größtmögliche Entsprechung zu finden. Der ursprüngliche Algorithmus war *reciprocal averaging*, der aktuelle Algorithmus ist *singular value decomposition* (svd).

Was sagt das Ergebnis aus?

Im Biplot werden Aufnahmen (Objekte) und Arten (Eigenschaften) zugleich dargestellt. Die Zentroide der Arten werden als Punkte dargestellt. Für den relativen Vergleich der Aufnahmen sind die Distanzen zum jeweiligen Zentroid relevant. Die Güte des Ergebnisses kann über den Anteil durch die Achsen erklärter Inertia („Varianz“) bemessen werden.

Eignung / Anforderungen an die Daten

CA fußt auf der Annahme unimodaler Zusammenhänge. Das heißt, sie kann Zunahmen und Abnahmen der Arten entlang eines Umweltgradienten abbilden. Lineare Zusammenhänge können damit ebenfalls abgebildet werden. Als Distanzmaß wird die Chi²-Distanz verwendet – auch hier existiert das Doppelnull-Problem. Liegt ein sehr starker Gradient zu Grunde, kann der Bogen-Effekt auftreten, d.h. die Aufnahmen werden in einem Bogen angeordnet und die Enden sind jedoch nicht nach innen gebogen.

Befehle in R

Paket vegan: `cca()` mit nur einer Tabelle. Paket ade4: `dudi.coa()`

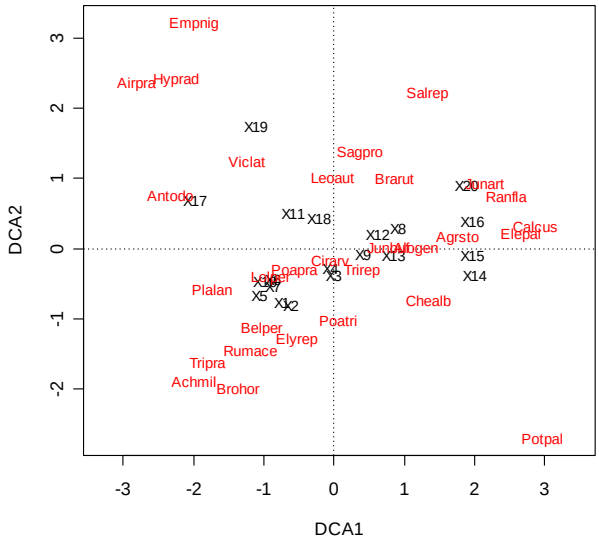
Wesentliche Angaben

Datensatz (welche Objekte, welche Eigenschaften, welcher Datentyp, welche Skala?)
Transformation des Datensatzes
Summe der Inertia, Anteil der Inertia, der auf den Achsen abgebildet ist
Welche Achsen werden interpretiert? Ggf. Rotation
Interpretation der Ordinationsplots
Begründung für die Wahl der Methode (Ist Chi²-Distanzmaß adäquat?)
Interpretation der Ordinationsplots, Angaben von Interpretationshilfen (z.B. Overlays)
Begründung für die Wahl der Methode

ENTZERRTE KORRESPONDENZANALYSE DCA

Detrended correspondence analysis

Kurzbezeichnung: DCA



Was passiert?

Nach einer CA wird mit dem Ziel, den *arch-effect* zu beheben, die erste Achse in Segmente aufgeteilt und diese so parallel zur 2. Achse verschoben, dass der Mittelwert der Aufnahmekoordinaten auf der 2. Achse im Segment 0 ergibt. Dazu werden die Achsen neu skaliert.

Was sagt das Ergebnis aus?

Im Biplot werden Aufnahmen (Objekte) und Arten (Eigenschaften) zugleich dargestellt. Die Zentroide der Arten werden als Punkte dargestellt. Eine Einheit auf der Ordinationsachse entspricht einem *half-change* der Artenzusammensetzung und gilt als Maß für die β-Diversität des Datensatzes. Über 4 Einheiten wird also einmal das Arteninventar ausgetauscht.

Eignung / Anforderungen an die Daten

Die Anforderungen sind die gleichen wie für eine CA. Diese fußt auf der Annahme unimodaler Zusammenhänge. Das heißt, sie kann Zunahmen und Abnahmen der Arten entlang eines Umweltgradienten abbilden. Lineare Zusammenhänge können damit ebenfalls abgebildet werden. Als Distanzmaß wird die Chi²-Distanz verwendet. Bei erfolgreicher Entzerrung erhält man eine unabhängig von der 1. Achse interpretierbare 2. Achse.

Befehle in R

Paket vegan: `cca()` mit nur einer Tabelle. Paket ade4: `dudi.coa()`

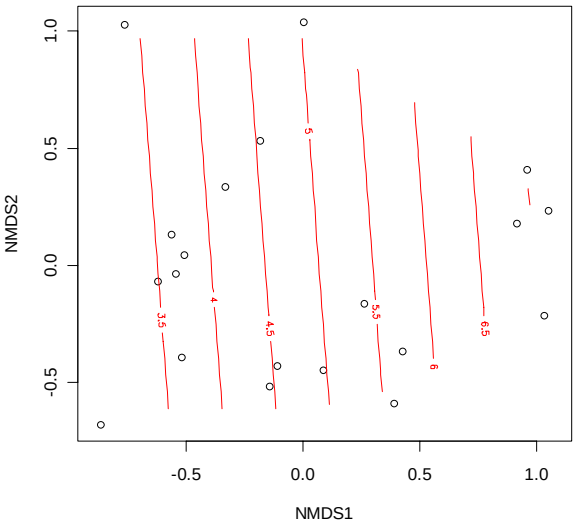
Wesentliche Angaben

Datensatz (welche Objekte, welche Eigenschaften, welcher Datentyp, welche Skala?)
Transformation des Datensatzes
Summe der Inertia, Anteil der Inertia, die auf den Achsen abgebildet ist
Welche Achsen werden interpretiert? Ggf. Rotation
Begründung für die Wahl der Methode (Ist Chi²-Distanzmaß adäquat? Warum *detrending*?)
Interpretation der Ordinationsplots, Angaben von Interpretationshilfen (z.B. Overlays)

NICHTMETRISCHES MEHRDIMENSIONALES SKALIEREN NMDS

Nonmetric multidimensional scaling

Kurzbezeichnung: NMDS oder NMS



Was passiert?

Iterative Suche nach der besten Anordnung von n Objekten (Aufnahmen) in k Dimensionen (Achsen), um den Stress der k-dimensionalen Konfiguration zu minimieren. Stress bezeichnet die Abweichung der Monotonie im Verhältnis der Distanzen im Datensatz und im k-dimensionalen Raum. Das Ergebnis hängt von der Anzahl der Dimensionen ab.

Was sagt das Ergebnis aus?

Im Plot wird die relative Lage der Aufnahmen (Objekte) zueinander dargestellt. Arten können in einem weiteren Schritt über die gewichteten Mittel ihrer Abundanzen (*weighted averages*) im Plot positioniert werden. Zur Interpretation des Zusammenhangs mit Umweltvariablen empfehlen sich Overlay-Verfahren. Die Güte des Ergebnisses kann über den Anteil durch die Achsen erklärter Varianz bemessen werden.

Eignung / Anforderungen an die Daten

Keine Annahme linearer Zusammenhänge. Kein Doppelnull-Problem, da Distanzmaß frei wählbar und Umwandlung in Distanzränge.

Befehle in R

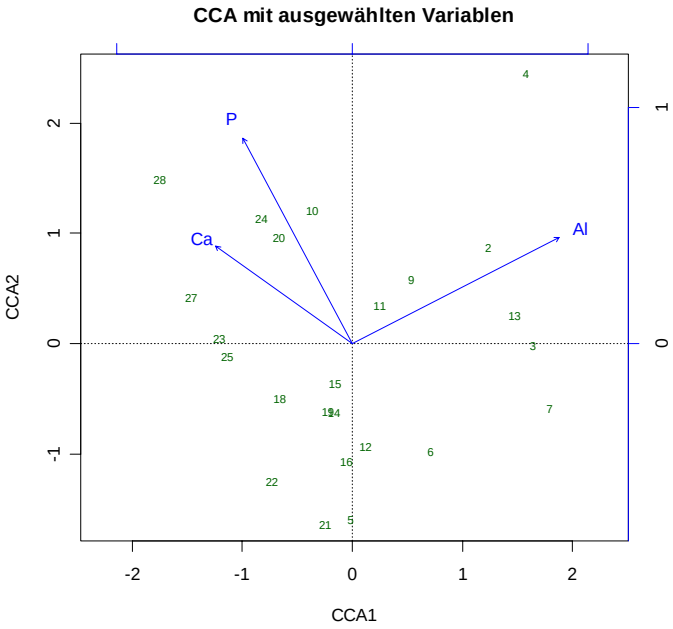
Paket vegan: `metaMDS()` automatischer Aufruf mehrer Läufe.
Paket MASS: `isoMDS()` einzelne Aufrufe (wird von metaMDS aufgerufen).
Für Overlays im Paket vegan: `ordisurf(), envfit()` oder Eigenschaften der Symbole

Wesentliche Angaben

Datensatz (welche Objekte, welche Eigenschaften, welcher Datentyp, welche Skala?)
Transformation des Datensatzes, verwendetes Distanzmaß
Verwendete Software / Algorithmus
Startwerte zufällig oder gesetzt, Anzahl der Läufe
Bestimmung der Dimensionen des Datensatzes, Anzahl der Dimensionen in der Ordination
Ergebnis des Monte-Carlo-Tests (Wahrscheinlichkeit, Ergebnis per Zufall zu erreichen)
Anzahl der Iterationen bis zum Endergebnis, Beurteilung der Stabilität des Ergebnisses
Anteil der Varianz, die auf den Achsen abgebildet ist (r^2 zwischen Distanz im Ordinationsraum und im Originaldatensatz)
Interpretation der Ordinationsplots, Angaben von Interpretationshilfen (z.B. Overlays)
Begründung für die Wahl der Methode

DIREKTE GRADIENTENANALYSE[1]

Was passiert?



Die durch Linearkombinationen der Umweltvariablen erklärbaren Anteile der Art-Abundanzen werden so auf synthetischen (kanonischen oder *constrained*) Achsen abgebildet, dass ein größtmöglicher Anteil der Variabilität auf den ersten Achsen dargestellt werden kann. Damit werden Muster im Datensatz, für die die ausgewählten Umweltvariablen verantwortlich sind, erkennbar.

Was sagt das Ergebnis aus?

Zunächst ist zu sehen, welcher Anteil der Gesamtvarianz durch die Kombination der ausgewählten Umweltvariablen abgebildet werden kann (*constrained inertia*). Eine weitere Information über die Bedeutung der abgebildeten Muster gibt der Anteil an Varianz der gezeigten Dimensionen bezogen auf die *constrained inertia*. Rauschen im Datensatz wird auf die Restvarianz *unconstrained inertia* verlagert. Daher liegen die erklärten Varianzen pro Achse vergleichsweise hoch. Die Auswahl der Umweltvariablen beeinflusst das Ergebnis maßgeblich. Sie sollte hypothesengeleitet erfolgen, denn das beste statistische Ergebnis bringt nur dann einen Fortschritt, wenn es eine gute ökologische Erklärung dafür gibt.

Interpretationshilfen

Der Triplot kann mit Overlays (Symbolgröße, -farbe und –form, Beschriftung, Isolinien), die Ausprägung einzelner Variablen (Umweltvariablen, Art-Abundanzen, Eigenschaften) an den Objekten (Aufnahmen, Proben,…) darstellen, ergänzt werden.

Wesentliche Überlegungen

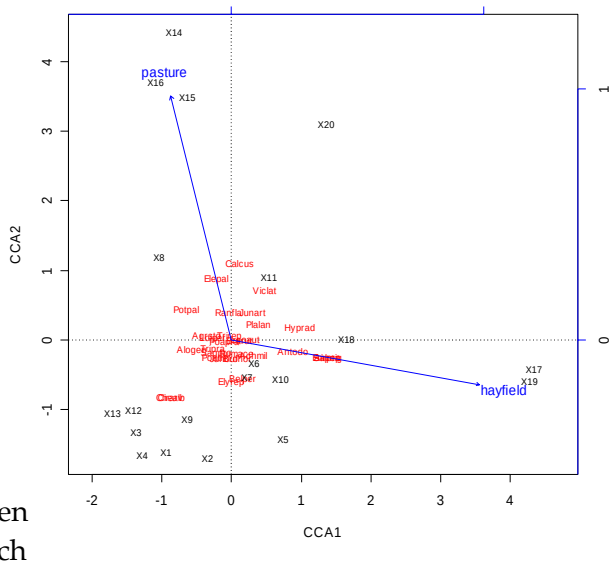
Welche Transformation?	Gewichtung innerhalb der Variablen und Vergleichbarkeit der Variablen
Welches Distanzmaß?	Vermeidung des Doppelnull-Problems und Anforderungen der Verfahren
Reaktion der Arten auf den Gradienten?	Linear oder unimodal?
Welches Verfahren?	RDA, CCA, DCCA?
Welche Umweltvariablen?	Hypothesengeleitete oder schrittweise Auswahl

Auswahl von Arten und Eigenschaften im Triplot?
Skalierung des Ergebnisses?
Wie viele Achsen werden interpretiert?

[1] Im Skript wird nur CCA behandelt. RDA ist ein Derivat der CCA, das analog zur PCA gehört. Die DCCA, analog zur DCA ist meist nicht notwendig.

KANONISCHE KORRESPONDENZANALYSE CCA

Canonical correspondence analysis / constrained correspondence analysis



Kurzbezeichnung: CCA

Was passiert?

Zunächst wird eine CA gerechnet. Die Aufnahmekoordinaten (WA) werden dann als abhängige Variablen in einer multiplen linearen Regression mit den Umweltvariablen verwendet. Mit den Modellwerten (LC) wird wieder eine CA gerechnet usw. bis ein stabiles Ergebnis erreicht ist.

Was sagt das Ergebnis aus?

Im Triplot werden Aufnahmen (Objekte) und Arten (Eigenschaften) sowie die Umweltfaktoren zugleich dargestellt. Die Zentroide der Arten werden als Punkte dargestellt. Für den relativen Vergleich der Aufnahmen sind die Distanzen zum jeweiligen Zentroid relevant. Die Pfeile zeigen an, in welcher Richtung die entsprechende Umweltvariable zunimmt. Die Länge des Pfeils ist ein Maß für die Stärke des Einflusses der jeweiligen Umweltvariablen. Die Güte des Ergebnisses kann über den Anteil durch die Achsen erklärter Inertia („Varianz“) bemessen werden. Dabei ist weiterhin der Anteil an Inertia, der durch die Umweltvariablen (*constrained variables*) erklärt wird, zu beachten. Das Ergebnis sollte auch mit dem Ergebnis einer Ordination ohne Umweltvariablen (*unconstrained ordination*, CA) verglichen werden

Eignung / Anforderungen an die Daten

Die CCA impliziert die Annahme linearer Zusammenhänge zwischen den Umweltvariablen und der Ausprägung der Art-Abundanzen. Die im Ansatz verwendete CA fußt auf der Annahme unimodaler Zusammenhänge (vgl. CA). Als Distanzmaß wird die Chi²-Distanz verwendet – auch hier existiert das Doppelnull-Problem. Werden nur wenige Umweltvariablen verwendet, kann der Bogen-Effekt (*arch-effect*) sehr wahrscheinlich vermieden werden. Bei der Berücksichtigung sehr vieler Umweltvariablen sind die Zwänge (*constraints*) weniger stark und das Odinationsergebnis nähert sich dem der CA an.

Befehle in R

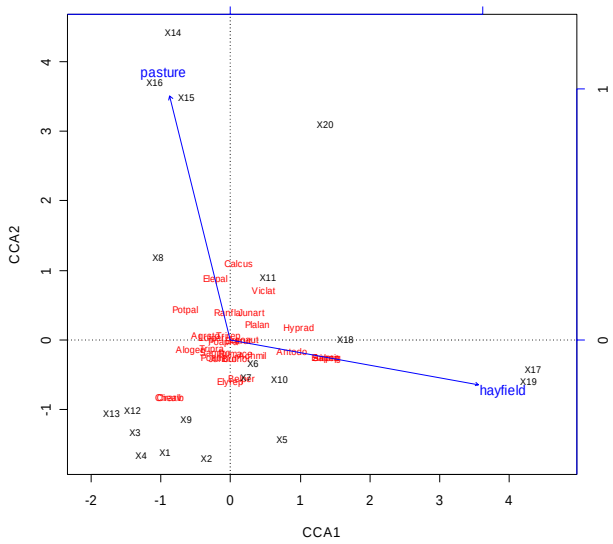
Paket vegan: `cca()` mit einer Vegetationstabelle und einer Tabelle der Umweltvariablen, bzw. der Angabe einer Formel `cca(Vegetations.Daten~Var1+Var2,Umwelt.Daten)`

Wesentliche Angaben

Datensatz (welche Objekte, welche Eigenschaften, welcher Datentyp, welche Skala?)
Transformation des Datensatzes
Summe der Inertia, Verhältnis *constrained/unconstrained* Inertia, Anteil der Inertia, die auf den Achsen abgebildet ist; welche Achsen werden interpretiert?
Begründung für die Wahl der Methode und die Wahl der Umweltvariablen
Interpretation der Ordinationsplots

PARTIELLE ORDINATION

Partial ordination



Was passiert?

In der partiellen Ordination wird der Einfluss einer oder mehrerer Umweltvariablen, der sog. Kovariablen, im Vorfeld der Analyse aus dem Datensatz entfernt und nur die Restvarianz wiederum in Abhängigkeit anderer Umweltvariablen untersucht. Es handelt sich also um einen speziellen Fall von direkter Gradientenanalyse.

Was sagt das Ergebnis aus?

Das Ergebnis ist genauso zu interpretieren wie das Ergebnis einer einfachen direkten Ordination (z.B. CCA) mit dem Unterschied, dass der Einfluss ein(ig)er Umweltvariablen entfernt wurde. Bezüglich der Verteilung der Inertia sind nun die folgenden Fraktionen zu beachten: Gesamtvariabilität (*total inertia*), Anteil der Kovariablen (*conditional inertia*), Anteil der Umweltvariablen (*constrained inertia*) und Restvarianz (*unconstrained inertia*). Die Güte des Ergebnisses kann über den Anteil durch die Achsen erklärter Inertia („Varianz“) bemessen werden. Dabei ist weiterhin der Anteil an Inertia, der durch die Umweltvariablen (*constrained variables*) erklärt wird, zu beachten.

Eignung / Anforderungen an die Daten

Die partielle Ordination wird verwendet bei schrittweiser Variablenauswahl, um die Verbesserung des Modells bei Hinzunahme einer weiteren Variablen zu testen. Weiterhin ist sie sinnvoll, um den Einfluss von Zeit oder Raum wechselseitig aus der Analyse zu nehmen oder andere Einflüsse wie etwa die Höhenlage herauszunehmen , um sich dem Einfluss der anderen Variablen zu widmen, ohne dass die Effekte verdeckt werden. Wesentlicher Ansatz für die Varianzaufteilung. Anforderungen an die Daten wie bei CCA.

Befehle in R

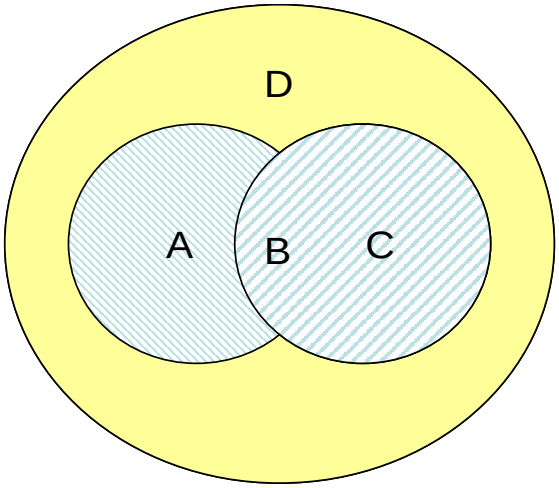
Paket vegan: `cca()` oder `rda()`mit der Angabe einer Formel und den Kovariablen als `Condition`. `cca(Daten.Vegetation~Var1+Var2+Condition(Var3+Var4),Daten.Umwelt)`

Wesentliche Angaben

Datensatz (welche Objekte, welche Eigenschaften, welcher Datentyp, welche Skala?)
Transformation des Datensatzes
Summe der Inertia, Verhältnis *conditional/constrained/unconstrained* Inertia, Anteil der Inertia, die auf den Achsen abgebildet ist; welche Achsen werden interpretiert?
Begründung für die Wahl der Methode, die Wahl der Umweltvariablen und die Wahl der Kovariablen, evtl. Vergleich mit Analyse ohne Kovariablen
Interpretation der Ordinationsplots

VARIANZAUFTEILUNG

Variation partitioning



Was passiert?

Durch die mehrfache, z.T. partielle Ordination eines Datensatzes wird ermittelt, welcher Anteil der Gesamtvariabilität im Datensatz (Inertia D) durch einzelne (Gruppen von) Umweltvariablen abgebildet werden kann (A+C, B+C) und welchen Anteil diese Variablen gemeinsam (C) und jeweils nur für sich (A, B) erklären.

Wie wird vorgegangen?

Es wird zunächst eine direkte Ordination mit allen zu betrachteten Umweltvariablen gerechnet (Ergebnis: A+B+C). Dann wird eine partielle Ordination mit der Variablengruppe 1 als Kovariablen und der Variablengruppe 2 als Umweltvariablen gerechnet (Ergebnis: A+B, CB). Dann muss noch eine direkte Ordination nur mit der Variablengruppe 2 als Umweltvariablen gerechnet werden (Ergebnis: B+C). Daraus lassen sich die einzelnen Fraktionen bestimmen und das Gewicht der Variablengruppen (z.B. Landnutzung und Standortfaktoren) vergleichen. Die Gesamtvariabilität (D) wird bei jeder Ordination ausgegeben.

Eignung

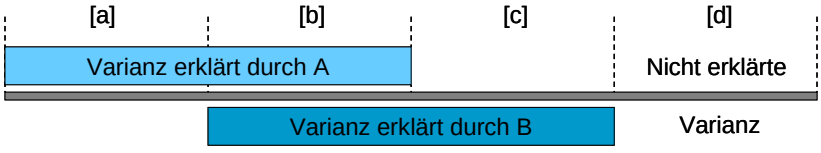
- Ziel: Vergleich des Einflusses von Faktorenkomplexen
- Aufteilen der Variabilität in Raum und Zeit
- Trennen des Einflusses des Standorts und der Landnutzung
- Berücksichtigen der Autokorrelation

Befehle in R

Paket `vegan`: `cca()` oder `rda()` ggf.mit der Angabe einer Formel und den Kovariablen als `Condition`. `cca(Daten.Vegetation~Var1+Var2+Condition(Var3+Var4),Daten.Umwelt)`

Wesentliche Angaben

Datensatz (welche Objekte, welche Eigenschaften, welcher Datentyp, welche Skala?)
Transformation des Datensatzes
Summe der Inertia, Verhältnis *conditional/constrained/unconstrained* Inertia,
Anteil der Inertia, die durch die betrachteten Variablengruppen erklärt wird
Begründung für die Wahl der Variablengruppen
Interpretation der Ordinationsplots



Modell	erklärt	nicht erklärt
-	-	a+b+c+d
A	a+b	c+d
B	b+c	a+d
A+B	a+b+c	d

SCHLÜSSEL ZUR ORDINATION

1. Direkte Gradientenanalyse 2
   2. Wenige Arten 4
      4. Monotone Reaktionen der Arten auf die Gradienten (geringe β-Diversität) **Lineare Regression**
      4. Nichtmonotone Reaktionen der Arten auf die Gradienten (hohe β-Diversität) Verallgemeinerte lineare Modelle **GLM**
   2. Viele Arten 5
      5. Monotone Reaktionen der Arten auf die Gradienten **RDA**
      5. Nichtmonotone Reaktionen der Arten auf die Gradienten 6
         6. Arch-Effekt (soll vermieden werden) **DCCA**
         6. Arch-Effekt spielt keine Rolle **CCA**
1. Indirekte Gradientenanalyse 3
   3. Nur Distanzwerte verfügbar 7
      7. Monotone Reaktionen der Arten auf die Gradienten **PCoA**
      7. Nichtmonotone Reaktionen der Arten auf die Gradienten **NMDS**
   3. Rohdaten verfügbar 8
      8. Monotone Reaktionen der Arten auf die Gradienten 9
         9. Variablen mit gleichem Wertebereich **PCA - Korrelationsmatrix**
         9. Variablen mit unterschiedlichen Wertereichen **PCA - Kovarianzmatrix**
      8. Nichtmonotone Reaktionen der Arten auf die Gradienten 10
         10. Vorab-Festlegung der Dimensionen kein Problem, Problem lokaler Optima stört nicht, Koordinaten der Arten nicht notwendig **NMDS**
         10. Nicht wie oben, aber entweder erscheint Arch-Effekt oder Entzerren und Reskalieren akzeptabel 11
            11. Arch-Effekt wird abgelehnt, Entzerren wird akzeptiert **DCA**
            11. Arch-Effekt stört nicht, oder nur an 1. Achse interessiert **CA**

nach Mike Palmer, The Ordination Web Page,
http://www.okstate.edu/artsci/botany/ordinate, verändert

LITERATURHINWEISE

Lehrbücher

Backhaus, K. 2003. *Multivariate Analysemethoden*. Springer, Berlin, Heidelberg.

ter Braak, C. J. F. 1996. *Unimodal models to relate species to environment*. DLO-Agricultural Mathematics Group, Wageningen.

ter Braak, C. J. F. & Smilauer, P. 1998. *CANOCO reference manual and User s guide to Canoco for Windows*. Microcomputer Power. Ithaca, NY.

Crawley, M. J. 1996. *GLIM for Ecologists*. Blackwell Science, London.

Crawley, M. J. 2002. *Statistical Computing An Introduction to Data Analysis Using S-Plus*. John Wiley & Sons.

Dale, M. R. T. 1999. *Spatial pattern analysis in plant ecology*. Cambridge University Press, Cambridge.

Gauch, H. G. 1982. *Multivariate analysis in community ecology*. Cambridge, Cambridge University Press.

Glavac, V. 1996. *Vegetationsökologie*. G. Fischer, Jena, Stuttgart, Lübeck, Ulm.

Jongman, R. H. G. 1987. *Data analysis in community and landscape ecology*. Pudoc, Wageningen.

Kent, M. & Coker, P. 1992. *Vegetation description and analysis*. CRC Press, Boca Raton.

Legendre, P. & Legendre, L. 1998. *Numerical ecology*. Elsevier, Amsterdam, Oxford.

Leps, J. & Smilauer, P. 2003. *Multivariate analysis of ecological data using CANOCO*. Cambridge University Press, Cambridge.

Leyer, I. & Wesche, K. 2007. *Multivariate Statistik in der Ökologie*. Springer, Berlin, Heidelberg.

McCune, B., Grace, J. B. & Urban, D. L. 2002. *Analysis of ecological communities*. MjM Software Design, Gleneden Bearch, Or.

Pielou, E. C. 1977. *Mathematical ecology*. Wiley, New York.

Pielou, E. C. 1984. *The interpretation of ecological data*. Wiley, New York.

Podani, J. 1994. *Multivariate data analysis in ecology and systematics*. SPB Academic Publ., Den Haag.

Podani, J. 2000. *Introduction to the exploration of multivarate biological data*. Backhuys, Leiden.

Sachs, L. 2004. *Angewandte Statistik*. Springer, Berlin, Heidelberg.

Upton, G. J. G. & Fingleton, B. 1990. Spatial data analysis by example. Wiley, Chichester.

Venables, W. N. & Ripley, B. D. 1999. *Modern applied statistics with S-PLUS*. Springer, New York, Berlin, Heidelberg.

Whittaker, R. H. 1978. *Ordination of plant communities*. Junk, Den Haag.

Whittaker, R. H. 1978. *Classification of plant communities*. Junk, Den Haag.

Wildi, O. 1986. *Analyse vegetationskundlicher Daten*. Veröffentlichungen des Geobotanischen Instituts der Eidgenössischen Technischen Hochschule, Stiftung Rübel, in Zürich: 90.

Wildi, O. & Orlóci, L. 1996. *Numerical exploration of community patterns*. SPB Acad. Publ., Den Haag.

Folia Geobotanica Heft 42/2 (2007): Sonderheft zu Fragen der statistischen Auswertung von vegetationskundlichen Datensätzen.

Internetadressen

The ordination web page (Michael Palmer)	http://www.okstate.edu/artsci/botany/ordinate/
Vegan: R functions for vegetation ecologist	http://cc.oulu.fi/~jarioksa/softhelp/vegan.html
The Comprehensive R Archive Network	http://cran.r-project.org/
PC-ORD	http://www.ptinet.net/~mjm/pcordwin.htm
JUICE	http://www.sci.muni.cz/botany/juice/
Skript von Jari Oksanen	http://cc.oulu.fi/~jarioksa/opetus/metodi/
Splus and R for Ecologists	http://labdsv.nr.usu.edu/splus_R/splus.html
Interessante Links zum Thema	http://www.homepages.ucl.ac.uk/~ucfagls/ncourse/links.htm
Statistikskript von Dormann & Kühn	http://www.ufz.de/data/Dormann2004Statsskript16256348.pdf

GLOSSAR

Ähnlichkeit	Maß für die Gemeinsamkeiten in der Ausprägung der Variablen zweier Objekte (Arten und ihre Abundanzen in zwei Aufnahmen)
Arch-Effekt	Krümmung der Punktwolke im Ergebnis der CA auf den ersten beiden Achsen bei starkem Einfluss eines Gradienten
Biplot	Plot, der gleichzeitig Objekte und Variablen (Aufnahmen und Arten) darstellt
CA	Korrespondenzanalyse (Ordinationsverfahren)
CCA	Kanonische Korrespondenzanalyse (Ordinationsverfahren)
Cluster	Klasse – Ergebnis eines Klassifikations- oder Clusterverfahrens
Code Replacement	Ersetzen der Abundanzklassen der Vegetationstabelle für die numerische Verarbeitung (Eingabe in ein Rechenprogramm)
DCA	Entzerrte Korrespondenzanalyse (Ordinationsverfahren), basiert auf CA mit anschließendem Entfernen des Arch-Effekts
Dendrogramm	Baumförmige Grafik, die das hierarchische Ergebnis eines Klassifikationsverfahrens abbildet.
Distanz	Entfernung zweier Aufnahmen (Objekte) im Raum, der durch die Variablen aufgespannt wird (Artenraum)
Doppelnull-Problem	Verzerrungen des Gradienten im Artenraum, die darauf beruhen, dass ein vegetationskundlicher Datensatz viele Nullstellen hat und sich aus gemeinsamen Nullstellen in verschiedenen Aufnahmen vermeintliche Ähnlichkeiten ergeben.
Hufeisen-Effekt	Krümmung der Punktwolke im Ergebnis der PCA auf den ersten beiden Achsen bei starkem Einfluss eines Gradienten, die Enden des Gradienten werden nach innen gebogen und bilden damit nicht mehr die äußersten Punkte entlang der ersten Achse
Klassifikation	Einteilen der Objekte in Klassen
Multiple Verfahren	Mehrere unabhängige und eine abhängige Variable
Multivariate Statistik	Mehrere unabhängige und mehrere abhängige Variablen
NMDS	Nichtmetrisches mehrdimensionales Skalieren (Ordinationsverfahren) (auch NMS)
Ordination	Anordnung der Objekte nach Ähnlichkeit der Ausprägung der Eigenschaften
PCA	Hauptkomponentenanalyse (Ordinationsverfahren)
Plot	Grafische Darstellung des Ergebnisses eines Ordinationsverfahrens
RDA	Redundanzanalyse (Ordinationsverfahren)
Transformation	Umwandeln der Abundanzwerte (Werte in der Tabelle) entweder monoton (alle Werte unabhängig voneinander nach dem gleichen Verfahren) oder als Vektortransformation (abhängig von den anderen Werten in der Tabelle). Ziel: Erfüllen von Anforderungen statistischer Verfahren, Angleichen der Wertebereiche, Herabgewichten großer Abundanzen,…
Triplot	Plot der zugleich Objekte, Variablen und Umweltvariablen darstellt