

Nichtmetrische multidimensionale Skalierung (NMDS)



GEORG-AUGUST-UNIVERSITÄT
GÖTTINGEN

Dr. Heike Culmsee
Vegetationsanalyse & Phytodiversität

Übersicht Ordinationsverfahren

	Linear methods	Weighted averaging	Multidimensional scaling
Unconstrained	Principal Components (PCA)	(Detrended) Correspondence Analysis (DCA, CA)	Metric (PCoA), Non-metric (NMDS)
Constrained	Redundancy Analysis (RDA)	Canonical Correspondence Analysis (CCA)	---

Ziel von Ordinationsverfahren: komplexe Zusammenhänge werden in einem möglichst wenig-dimensionalen Raum abgebildet.

Lösung 1: PCA, DCA, RDA, CCA

- Ersetzen der gemessenen Variablen durch synthetische Variablen (Hauptkomponenten, CA-Achsen)
- Struktur des Ordinationsraums wird durch die Achsen aufgespannt, die wiederum durch die Wahl der Methode festgelegt sind (Euklidische Distanzen, X^2 -Distanzen)

Lösung 2: MDS

- Unähnlichkeits- oder Distanzmatrix zwischen den Objekten
- Die Abstände werden dann direkt in einem Koordinatensystem abgebildet, d.h. die Objekte werden in einen neuen Raum projiziert, so dass ihre wahren Abstände möglichst wenig verzerrt werden.
- Verteilungsunabhängig, d.h. nicht-parametrisch
- Sehr flexible Verfahren: die anfängliche Dreiecksmatrix kann mit verschiedenen Unähnlichkeits- oder Distanzmaßen berechnet werden

Multidimensionale Skalierung (MDS)

⇒ **PCoA** Metrische Multidimensionale Skalierung (= Hauptkoordinatenanalyse)

Distanzen werden weitestgehend bewahrt (wie bei PCA);
Anwendung: räumliche Koordinaten (räumliche Autokorrelation)

⇒ **NMDS** Nicht-metrische Multidimensionale Skalierung

Verwendung von Rängen anstelle von metrischen Werten;
es wird nur auf **Monotonie** geachtet, nicht auf die tatsächlichen Distanzen.

Prinzip und Arbeitsschritte der NMDS

- Wahl des Unähnlichkeits- oder Distanzmaßes

Unähnlichkeits- & Distanzmaße

Qualitative Ähnlichkeitsmaße (presence/absence)

Kontingenzmaße

(basierend auf Vierfeldertafel)

		Art in Aufnahme 1	
		+	-
Art in Aufnahme 2	+	a	b
	-	c	d

$$S_J = a / (a + b + c)$$

Jaccard-Index

$$S_S = 2a / (2a + b + c)$$

Sørensen-Index

a Gesamtzahl der Arten, die in beiden Aufnahmen vorhanden sind,

b Anzahl der Arten, die nur in Aufnahme 1 vorhanden sind,

c Anzahl der Arten, die nur in Aufnahme 2 vorhanden sind

Qualitative Ähnlichkeitsmaße

Bray-Curtis-Koeffizient

$$S_{bc} = \frac{2w}{B + C}$$

- S_{bc} similarity (Ähnlichkeit) zwischen den Samples 1 und 2
- B Σ Abundanzen aller Arten in Sample 1
- C Σ Abundanzen aller Arten in Sample 2
- w Σ der jeweils niedrigsten Abundanzwerte einer Art in den beiden Samples

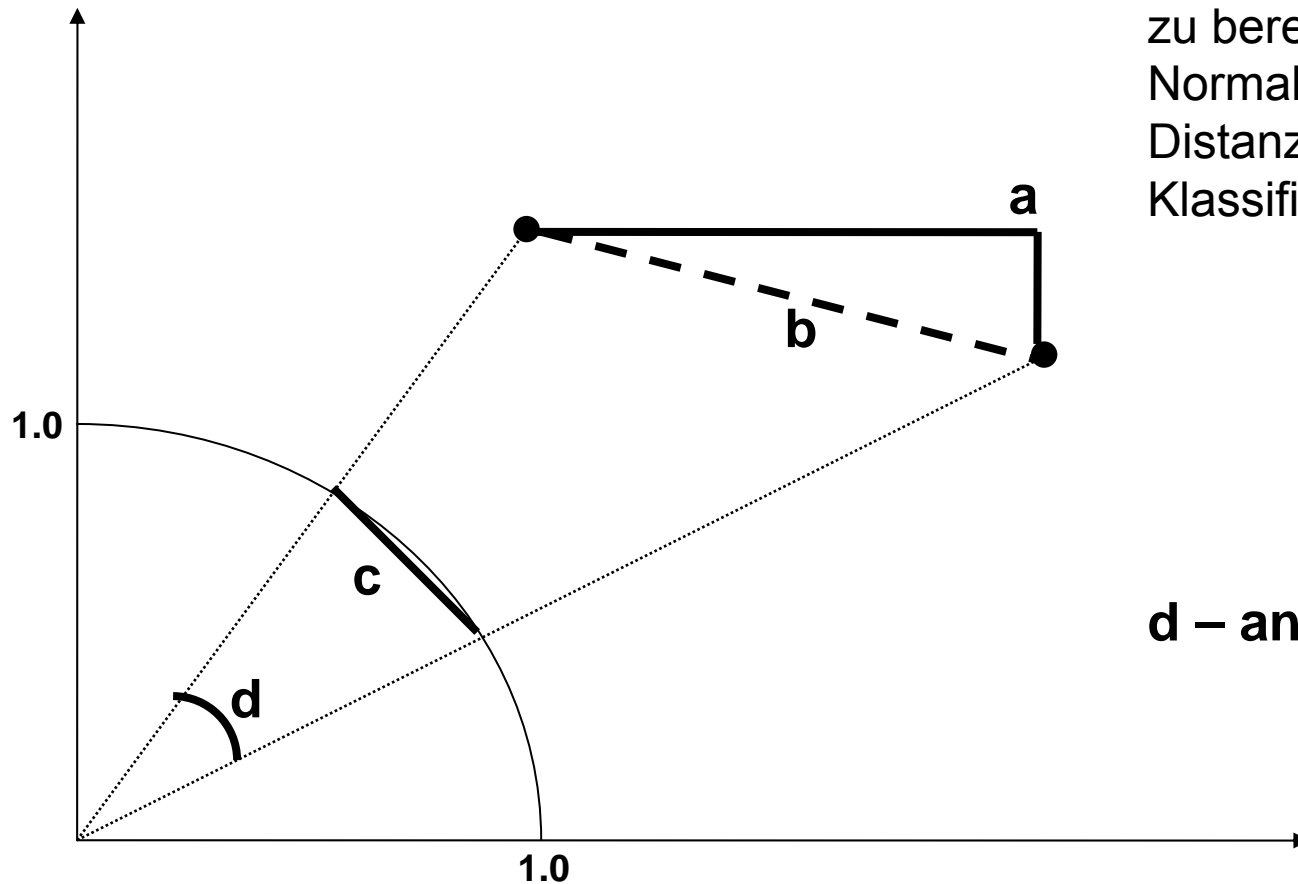
a – City-Block-Distanz, Manhattan distance

b – Euklidische Distanz

bewirkt hohe Bedeutung stark deckender Arten

c – Sehnenstanz (chord distance)

zu berechnen aus der Normalisierung der Euklid. Distanz; wird für Klassifikationen oft verwendet



d – angular separation

Unähnlichkeits- & Distanzmaße

Vergleich verschiedener Unähnlichkeits- und Distanzmaße

	Art A	Art B	Art C	Σ
Objekt 1	0	1	1	2
Objekt 2	1	0	0	1
Objekt 3	0	4	4	8
Σ	1	5	5	11

Tabelle 4.4. Hypothetisches Beispiel einer einfachen Datenmatrix zur Verdeutlichung der Eigenschaften von Abstandsmaßen (in Anlehnung an: Legendre u. Legendre 1998)

Tabelle 4.5. Vergleich qualitativer asymmetrischer Ähnlichkeit (Sörensen), quantitativer asymmetrischer Unähnlichkeit (Bray-Curtis) und quantitativer symmetrischer Distanz (Euklidische Distanz)

Sörensen-Ähnlichkeit			Bray-Curtis-Unähnlichkeit			Euklidische Distanz					
	O1	O2	O3		O1	O2	O3		O1	O2	O3
O1	1			O1	0			O1	0		
O2	0	1		O2	1	0		O2	1.7	0	
O3	1	0	1	O3	0.6	1	0	O3	4.2	5.7	0

Prinzip und Arbeitsschritte der NMDS

- Wahl des Unähnlichkeits- oder Distanzmaßes;
- Durchführung einer ersten NMDS mit möglichst vielen Dimensionen (normalerweise 6), Shepard-Diagramm;
- Ausprobieren möglichst vieler Startkonfigurationen (ca. 50-100); Einschätzung der Schwankungsbreite des Stresses (lokales Minimum);

Nicht-metrische Multidimensionale Skalierung (NMDS)

Stress = ein Maß für die Güte einer Ordination

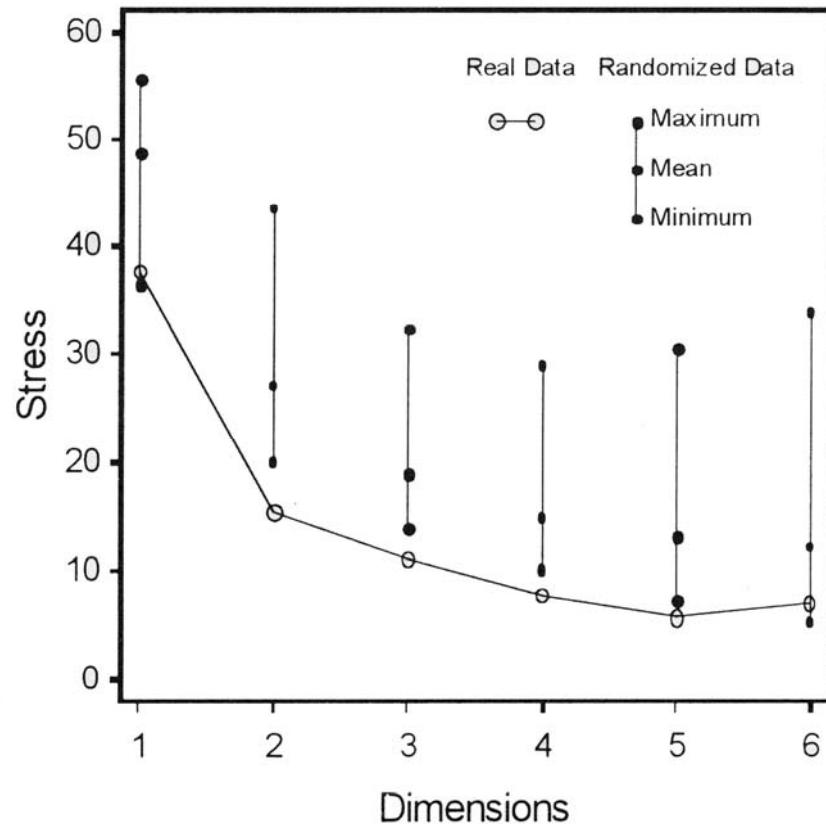


Figure 16.3. A scree plot shows stress as a function of dimensionality of the gradient model. "Stress" is an inverse measure of fit to the data. The "randomized" data are analyzed as a null model for comparison. See also Table 16.1.

Der Stress-Wert gibt an, wie stark die Lagebeziehungen im Ordinationsraum hinsichtlich ihrer Monotonie von denen in der ursprünglich berechneten Distanzmatrix abweichen

1. Schritt: Zufällige Anordnung der Objekte im Ordinationsraum => Stress am Anfang hoch.
2. Iterativ (schrittweise) wird versucht, die Objekte besser anzuordnen, also den Stress zu verringern. => Abweichungen von der monoton steigenden Beziehung werden verringert.

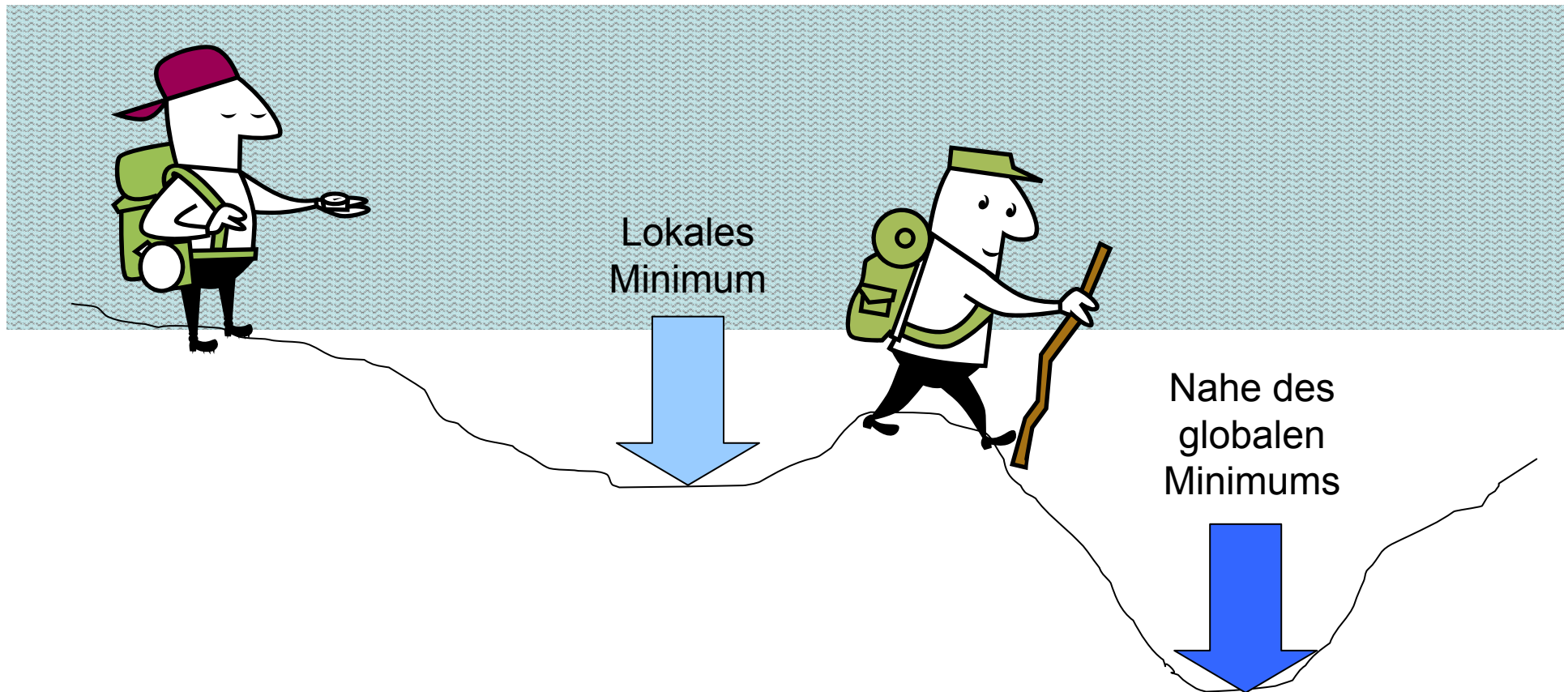
Interpretation der Stress-Ergebnisse

Clarke's rules of thumb

- | | |
|-------|--|
| < 5 | An excellent representation with no prospect of misinterpretation. This is, however, rarely achieved. |
| 5-10 | A good ordination with no real risk of drawing false inferences |
| 10-20 | Can still correspond to a usable picture, although values at the upper end suggest a potential to mislead. Too much reliance should not be placed on the details of the plot. |
| > 20 | Likely to yield a plot that is relatively dangerous to interpret. By the time stress is 35-40 the samples are placed essentially at random, with little relation to the original ranked distances. |
-

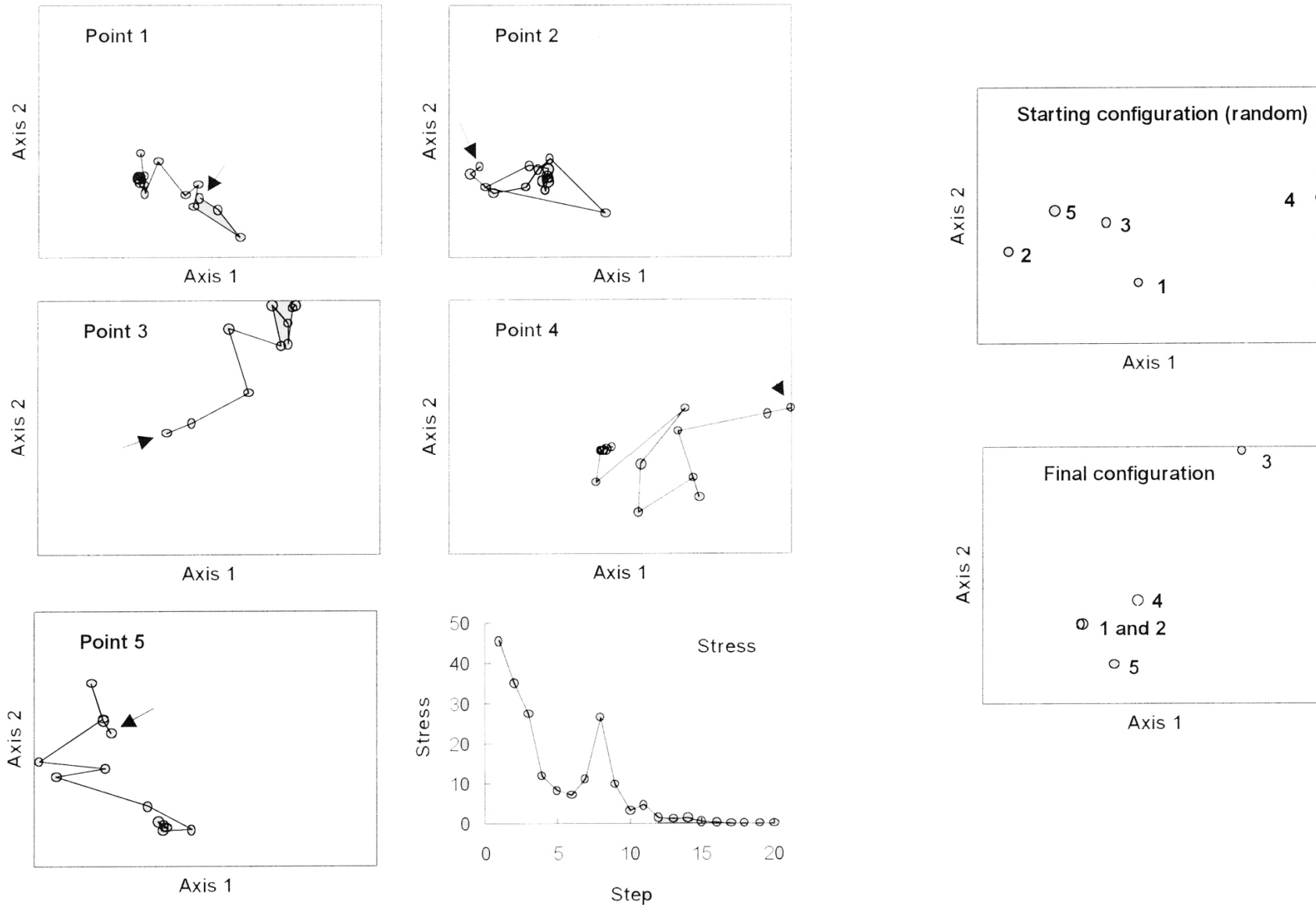
Nicht-metrische Multidimensionale Skalierung (NMDS)

Vom lokalen zum globalen Minimum



Nicht-metrische Multidimensionale Skalierung (NMDS)

Using random perturbations to find best final configuration



Prinzip und Arbeitsschritte der NMDS

- Wahl des Unähnlichkeits- oder Distanzmaßes;
- Durchführung einer ersten NMDS mit möglichst vielen Dimensionen (normalerweise 6), Shepard-Diagramm;
- Ausprobieren möglichst vieler Startkonfigurationen (ca. 50-100); Einschätzung der Schwankungsbreite des Stresses (lokales Minimum);
- Festlegen der Dimensionen (Achsen) und abschließende NMDS