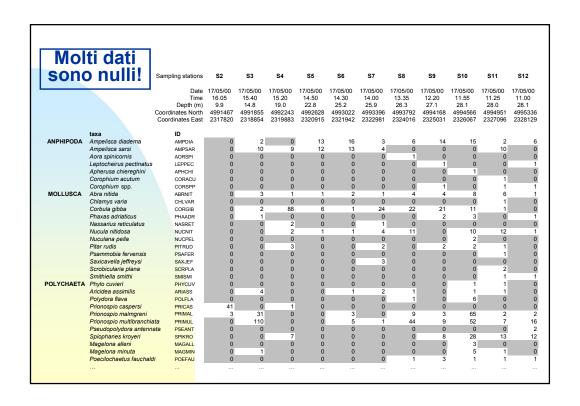
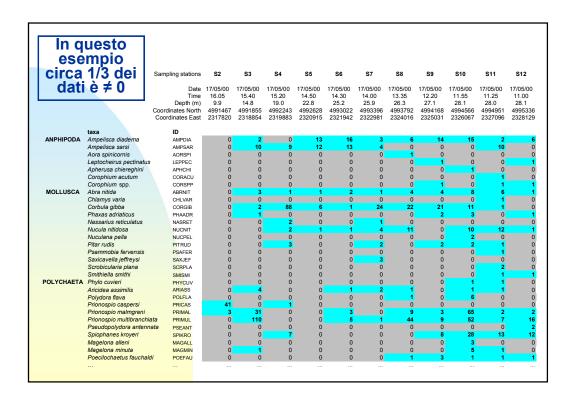
Analisi dei dati ecologici

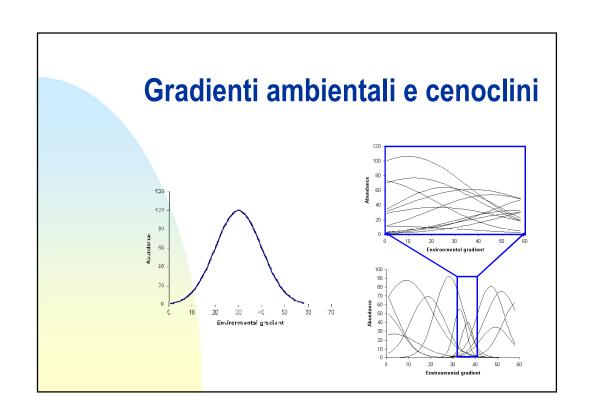
I da	ati												
		Sampling stations	S2	S3	S4	S5	S6	S7	S8	S9	S10	S11	S12
		Date Time Depth (m) Coordinates North Coordinates East	17/05/00 16.05 9.9 4991467 2317820	17/05/00 15.40 14.8 4991855 2318854	17/05/00 15.20 19.0 4992243 2319883	17/05/00 14.50 22.8 4992628 2320915	17/05/00 14.30 25.2 4993022 2321942	17/05/00 14.00 25.9 4993396 2322981	17/05/00 13.35 26.3 4993792 2324016	17/05/00 12.20 27.1 4994168 2325031	17/05/00 11.55 28.1 4994566 2326067	17/05/00 11.25 28.0 4994951 2327096	17/05/00 11.00 28.1 4995336 2328129
	taxa	ID											
ANPHIPODA	Ampelisca diadema	AMPDIA	0	2	0	13	16	3	6	14	15	2	6
	Ampelisca sarsi	AMPSAR	0	10	9	12	13	4	0	0	0	10	0
	Aora spinicornis	AORSPI	0	0	0	0	0	0	0	0	0	0	0
	Leptocheirus pectinatus		0	0	0	0	0	0	0	1	0	0	1
	Apherusa chiereghini	APHCHI	0	0	0	0	0	0	0	0	0	0	0
	Corophium acutum	CORACU	0	0	0	0	0	0	0	0	0	0	0
	Corophium spp.	CORSPP	0	0	0	0	0	0	0	1	0	1	1
MOLLUSCA	Abra nitida	ABRNIT	0	3	1	1	2	1	4	4	8	6	1
	Chlamys varia	CHLVAR	0	0	0	0	0	0	0	0	0	0	0
	Corbula gibba	CORGIB	0	2	88	6	1	24	22	21	11	1	0
	Phaxas adriaticus Nassarius reticulatus	PHAADR	0	1	0 2	0	0	0	0	2	3	0	1
	Nassarius reticulatus Nucula nitidosa	NASRET NUCNIT	0	0	2	1	1	4	11	0	10	12	1
POLYCHAETA	Nucula milioosa Nuculana pella	NUCPEL	0	0	0	0	0	0	0	0	2	12	0
	Pitar rudis	PITRUD	0	0	3	0	0	2	0	2	2	1	0
	Pitar rugis Psammobia fervensis	PSAFER	0	0	0	0	0	0	0	0	0	1	0
	Saxicavella jeffreysi	SAXJEF	0	0	0	0	0	3	0	0	0	0	0
	Scrobicularia plana	SCRPLA	0	0	0	0	0	0	0	0	0	2	0
	Smithiella smithi	SMISMI	0	0	0	0	0	0	0	0	0	1	1
		PHYCUV	0	0	0	0	0	0	0	0	1	1	0
	Aricidea assimilis	ARIASS	0	4	0	0	1	2	1	0	1	1	0
	Polydora flava	POLFLA	0	0	0	0	0	0	1	o o	6	0	0
	Prionospio caspersi	PRICAS	41	0	1	0	0	0	0	0	0	0	0
	Prionospio malmgreni	PRIMAL	3	31	0	0	3	0	9	3	65	2	2
	Prionospio multibranchia	ata PRIMUL	0	110	0	0	5	1	44	9	52	7	16
	Pseudopolydora antenn		0	0	0	0	0	0	0	0	0	0	2
	Spiophanes kroyeri	SPIKRO	0	0	7	0	0	0	0	8	28	13	12
	Magelona alleni	MAGALL	0	0	0	0	0	0	0	0	3	0	0
	Magelona minuta	MAGMIN	0	1	0	0	0	0	0	0	5	1	0
	Poecilochaetus fauchalo	di POEFAU	0	0	0	0	0	0	1	3	1	1	1





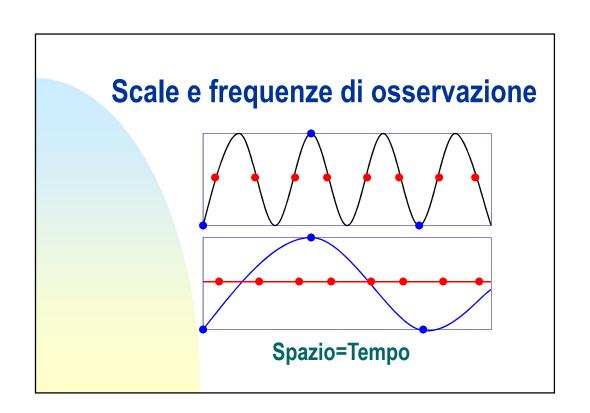
Caratteristiche dei dati ecologici

- I dati sono "sparsi", cioè hanno molti valori nulli (a volte la maggioranza!)
- La gran parte delle specie presenti è rara.
- I fattori ambientali che influenzano la distribuzione delle specie sono molteplici e combinati fra loro,...
- ...ma quelli veramente importanti sono pochi (bassa dimensionalità intrinseca).
- I dati contengono molto "rumore" sia per eventi stocastici e contingenti, sia per l'errore di osservazione (anche in condizioni ideali le repliche sono diverse!)
- L'informazione è spesso ridondante (la specie A è associata alla specie B, ma questa può essere associata alla specie C, etc.): questo è un problema, ma è anche ciò che rende possibile interpretare i dati ecologici.



Piani di campionamento

- Non esistono informazioni preliminari, le scale spazio-temporali non sono note:
 - ♦ piano randomizzato
- Esistono informazioni preliminari, le scale spazio-temporali sono note:
 - ◆ piano regolare
- Esistono sufficienti dati di riferimento per descrittori accessori (covarianti):
 - ◆ piano stratificato



La cassetta degli attrezzi.

- Ordinamento (PCA, MDS, NMDS, CA, DCA, CCA, etc.)
- Classificazione (algoritmi gerarchici, kmeans, reti neuronali, etc.)
- Analisi spaziale (correlogrammi, variogrammi, kriging, co-kriging, etc.)
- Analisi di serie (periodogrammi, runs tests, cross-correlation, cross-association, etc.)
- Confronti fra dati multivariati (MRPP, test di Mantel, INDVAL, etc.)
- Reti neurali
- ...

Misure di distanza

$$D_{jk} = \sqrt{\sum_{i=1}^{p} (x_{ij} - x_{ik})^2}$$

$$D_{jk} = \sum_{i=1}^{p} \left| x_{ij} - x_{ik} \right|$$

$$D_{ij} = \sum_{i=1}^{p} \frac{|x_{ij} - x_{ik}|}{(x_{ij} + x_{ik})}$$

$$D_{jk} = r \sqrt{\left(\sum_{i=1}^{p} \left| x_{ij} - x_{ik} \right|^{r}\right)}$$

Misure di similarità Per i coefficienti binari (dati di presenza/assenza) Numero di taxa presenti in entrambi i Osservazione j campioni Numero di taxa **Osservazione** *k* presenti solo nel b 🖛 campione k Numero di taxa d _ presenti solo nel campione j Numero di taxa p = a + b + c + d assenti in entrambi i campioni

Misure di similarità

Dati binari (qualitativi)

Indice di concordanza

$$S_{jk} = \frac{a+d}{a+b+c+d}$$
 simmetrico

Indice di Jaccard

$$S_{jk} = \frac{a}{a+b+c}$$
 asimmetrico

Indice di Sørensen

$$S_{jk} = \frac{2a}{2a+b+c}$$
 asimmetrico

Dati nanfitativi

Coeff. di Bray-Curtis

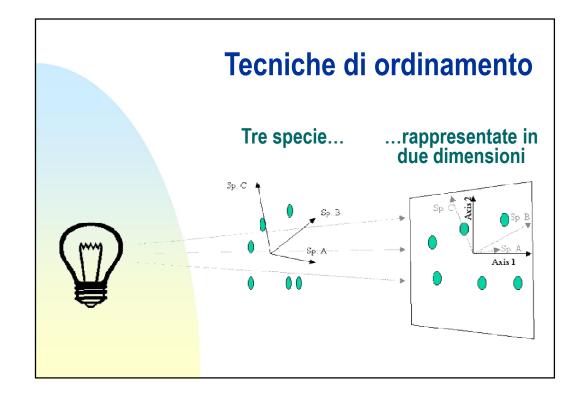
$$S_{ij} = \frac{\sum_{i=1}^{s} |x_{ij} - x_{ik}|}{\sum_{i=1}^{s} (x_{ij} + x_{ik})}$$

Similarità e dissimilarità

Dissimilarità = 1 - Similarità

Un coefficiente di dissimilarità è di tipo metrico se:

- 1. $D_{ij} = 0 \text{ se } j = k$
- 2. $D_{jk} > 0$ se $j \neq k$ Semi-metrica
- 3. $D_{jk}=D_{kj}$
- 4. $D_{jk}+D_{kh}\geq D_{jh}$ (assioma della diseguaglianza triangolare)



Perchè l'ordinamento?

"Ordination primarily endeavors to represent sample and species relationships as faithfully as possible in a low-dimensional space."

Gauch (1982)

Perchè...

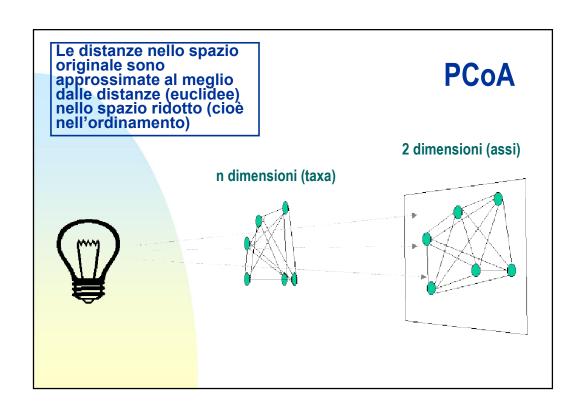
- E' impossibile visualizzare efficacemente insiemi di dati multidimensionali complessi
- Un'analisi multivariata è più economica e più efficiente di n analisi univariate
- Gli assi dello spazio ridotto di solito rappresentano gradienti ambientali interpretabili
- Se si effettuano anche test statistici, si evitano i problemi legati alle comparazioni multiple
- Concentrando l'attenzione solo su alcuni assi si evita di considerare il "rumore"

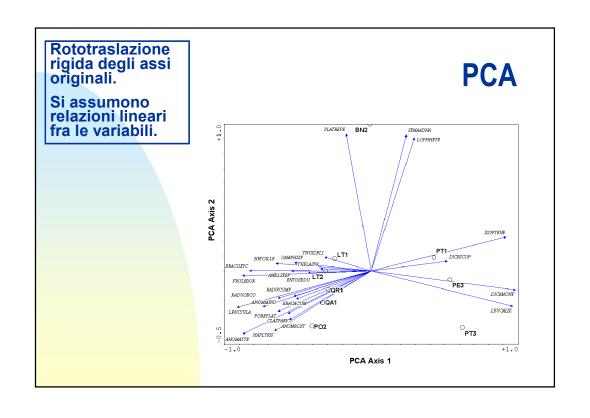
E inoltre...

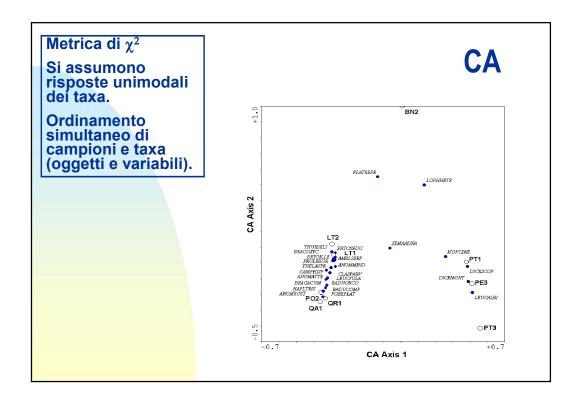
- Fino a non molto tempo fa l'obiettivo dei metodi di ordinamento era di tipo prettamente descrittivo (più un'arte che una scienza, quindi!).
- Con la CCA, i test statistici d'ipotesi sono stati accoppiati alle tecniche di ordinamento, superando l'approccio descrittivo (cfr. ter Braak 1985)

Analisi indiretta di gradiente

- Metodi basati su distanze
 - ◆ Ordinamento polare (Bray-Curtis)
 - ◆ Analisi delle Coordinate Principali (PCoA)
 - Multidimensional Scaling Nonmetrico (NMDS)
- Metodi basati su autovalori/autovettori
 - ◆ Modello lineare
 - Analisi delle Componenti Principali (PCA)
 - ◆ Modello unimodale
 - Analisi delle Corrispondenze (CA)
 - Analisi delle Corrispondenze Detrendizzata (DCA)







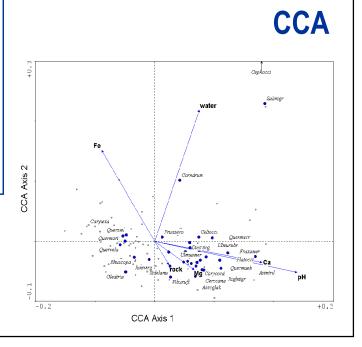
Analisi diretta di gradiente

- Modello lineare
 - ♦ Analisi di Ridondanza (RDA)
- Modello unimodale
 - ◆ Analisi Canonica delle Corrispondenze (CCA)
 - ◆ Analisi Canonica delle Corrispondenze Detrendizzata (DCCA)

E' concettualmente affine alla CA, ma l'ordinamento di campioni e taxa è vincolato a combinazioni lineari di variabili ambientali.

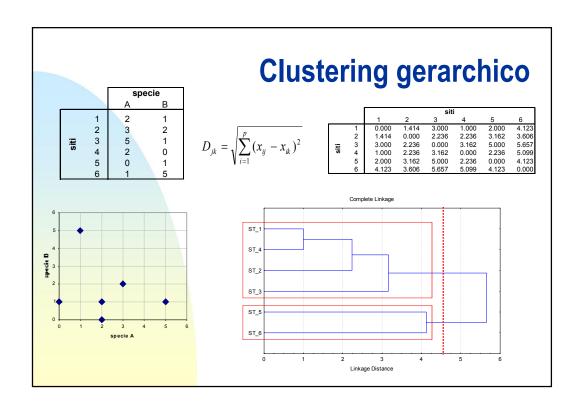
Rappresentazione

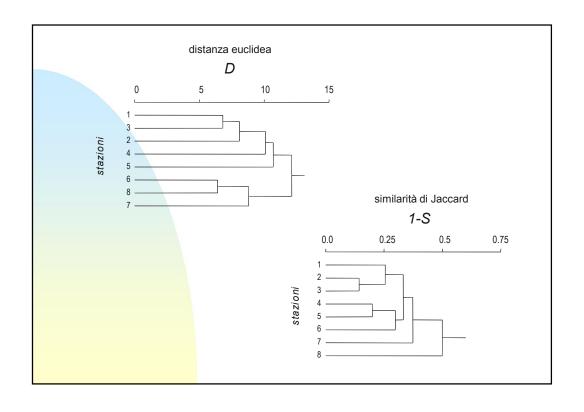
Rappresentazione simultanea di campioni, taxa e gradienti ambientali.

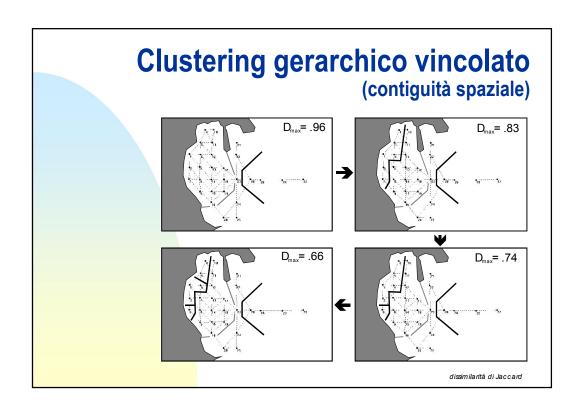


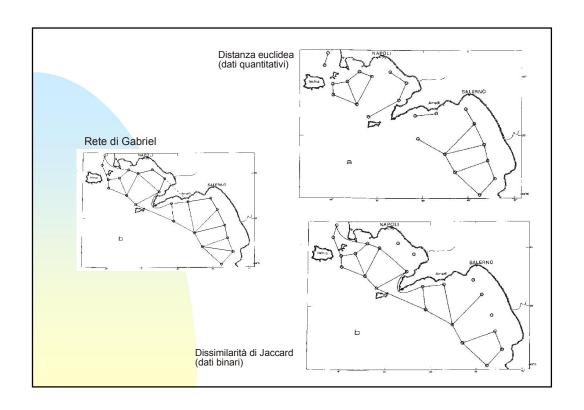
Clustering (classificazione)

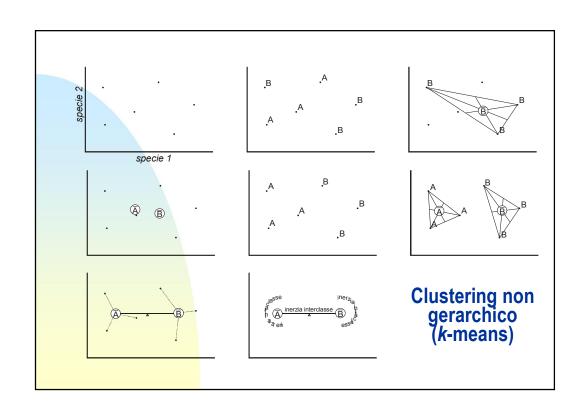
- Obiettivi:
 - ◆ Formare gruppi omogenei di entità (osservazioni, campioni, siti, specie, etc.)
 - ◆ Identificare discontinuità (nello spazio, nel tempo)
- Algoritmi:
 - ◆ Gerarchici
 - ♦ Non gerarchici
- Vincolati
- •Non vincolati













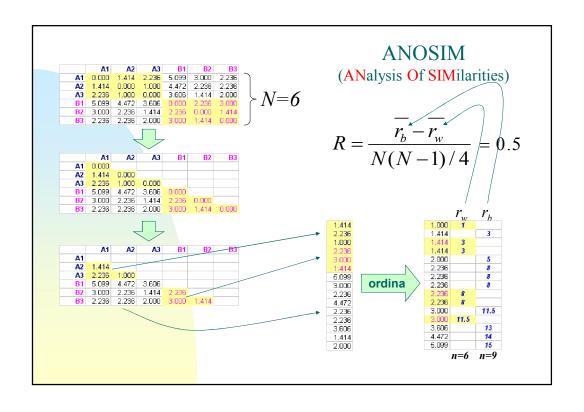
Test basati su permutazioni

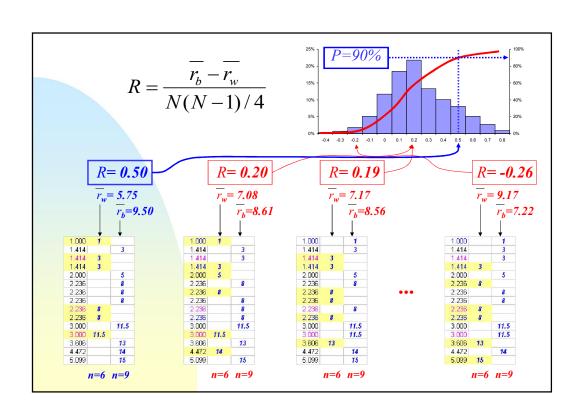
- Sono un caso speciale dei test di randomizzazione, che utilizzano serie di numeri casuali formulare delle inferenze statistiche.
- La potenza di calcolo dei moderni PC ha reso possibile la loro applicazione diffusa.
- Questi metodi non richiedono che siano soddisfatte particolari assunzioni circa la distribuzione dei dati.
- Quindi, questi metodi sono molto più adatti dei tradizionali test statistici (es. *t*-tests, ANOVA, etc.) in applicazioni ecologiche.

Test basati su permutazioni



- Si definisce una statistica il cui valore sia proporzionale all'intensità del processo o della relazione studiati
- Si definisce un'ipotesi nulla H₀
- Si crea un set di dati basati sul "rimescolamento" di quelli realmente osservati (la modalità di "rimescolamento" viene definita in funzione dell'ipotesi nulla)
- Si ricalcola la statistica di riferimento e si compara il valore con quello osservato
- Si ripetono gli ultimi due punti molte volte (es. 1000 volte)
- Se la statistica osservata è maggiore del limite ottenuto nel 95% dei casi basati su "rimescolamento", si rigetta H₀





Indicator Species Analysis

L'abbondanza relativa RA
$$_{kj}$$
 della specie j nel gruppo di campioni k è $\longrightarrow RA_{kj} = \frac{x_{kj}}{\displaystyle\sum_{k=1}^g x_{kj}}$

La frequenza media
$$\mathsf{RF}_\mathsf{kj}$$
 della presenza di una specie j nel gruppo di campioni k è $\longrightarrow RF_{kj} = \sum_{i=1}^{n_k} \frac{b_{ijk}}{n_k}$

Combinando abbondanze relative (RA) e frequenze medie (RF) si ottiene quindi il valore indicatore (IV) $\longrightarrow IV_{kj} = RA_{kj} \cdot RF_{kj} \cdot 100$

Indicator Species Analysis

Test di Mantel

	Δ	0.0	12	26	1.8	3.2
Matrice X		1.2				
distanze geografiche	С	2.6	3.1	0.0	1.1	4.2
anotanio googramono	D	1.8	0.5	1.1	0.0	3.4
	=	3 2	27	12	2 /	0.0

Matrice Y dissimilarità cenotica
 A
 B
 C
 D
 E

 A
 0.00
 0.29
 0.56
 0.45
 0.49

 B
 0.01
 0.00
 0.48
 0.06
 0.12

 C
 0.21
 0.17
 0.00
 0.27
 0.59

 D
 0.07
 0.04
 0.16
 0.00
 0.02

 E
 0.45
 0.34
 0.78
 0.21
 0.00

Statistiche di Mantel

assoluta

$$z = \sum_{i=1}^{n} \sum_{j=1}^{n} x_{ij} y_{ij}$$

standardizzata

$$r = \frac{1}{(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{(x_{ij} - \bar{x})}{s_x} \cdot \frac{(y_{ij} - \bar{y})}{s_y}$$

La distribuzione di riferimento si genera ricalcolando la statistica dopo permutazioni aleatorie di una delle due matrici o (per matrici molto grandi) approssimando una distribuzione t di Student.

Problema: valutare gli effetti della protezione sulla struttura delle comunità.

- La struttura delle comunità varia in funzione delle risposte delle specie a gradienti ambientali complessi.
- L'impatto antropico altera o modula questi gradienti.
- La protezione dell'ambiente riduce l'impatto antropico o lo trasferisce in altre aree.

