Ordination as a Tool for Unraveling Species-Environment Relationships

Daniel M. Palacios[1,2]

[1]Joint Institute for Marine and Atmospheric Research, University of Hawaii

[2]NOAA, SWFSC, Environmental Research Division, Pacific Grove, CA

Acknowledgments:
• National Oceanographic Partnership Program (ONR/NOPP)

OUTLINE

1. Background
2. Basic concepts
3. Preparatory steps
4. Ordination techniques

Background

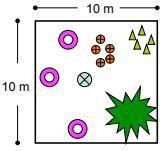
- Concept of dimensionality reduction has been around since Pearson (1901)
- Ordination (PCA) first used in psychology (Hotelling, 1933)
- Biology: developed by vegetation scientists, 50-60's
- Earth sciences: atmospheric, climate scientists, oceanographers, landscape ecologists

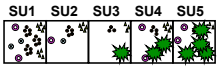
Sources:
McCune & Grace. 2002. Analysis of Ecological Communities. MjM Software Design.
Legendre & Legendre. 1998. Numerical Ecology. 2nd Ed. Elsevier.

Basic concepts

Sampling unit (SU)

- Vegetation plot
- Transect or grid of plots



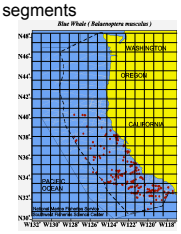


	sp1	sp2	sp3
SU1	3	0	0
SU2	2	1	0
SU3	1	2	1
SU4	0	1	2
SU5	0	0	3

Basic concepts

Marine mammal data

- Point observations
- Need to select a sampling unit based on study goals
  - •Break up transect line into segments
  - •Create a grid



http://swfsc.nmfs.noaa.gov/prd/PROGRAMS/CMMP/distMaps/blue_whale.htm

Basic concepts

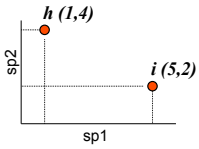
Abundance measures

- Density: #ind/unit area (or length)
- Presence-absence (large-scale studies)
- Size- or age-class data: carries information on demographics. Classes treated as "species"

Basic concepts

$\mathbf{A}_{mxn}$ The data matrix

		sp1	sp2	sp3	sp4	sp5
i	**SU**1	5	2	1	1	1
h	**SU**2	1	4	0		
j	**SU**3	1	1	0		



Sample units in species space

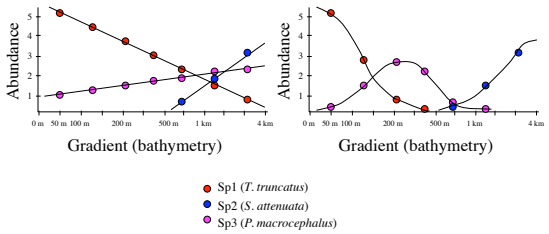
$\mathbf{E}_{mxp}$ The environmental matrix

		sst	chl	z20	zoml	dep
i	**SU**1	25	0.2	65	150	310
h	**SU**2	10.4	70	200	320	
j	**SU**3	19	1.2	80	220	400

Basic concepts

Species on gradients – idealized scenarios

- Linear (underlying assumption in PCA, DFA)
- Gaussian (mean-x, SD, max-y represent species)

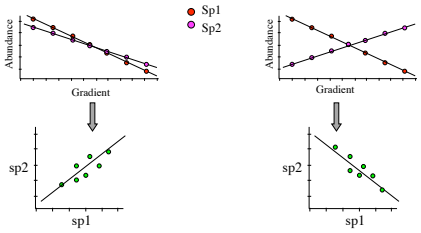


Sp1 (*T. truncatus*)
Sp2 (*S. attenuata*)
Sp3 (*P. macrocephalus*)

Basic concepts

Species associations – positive or negative

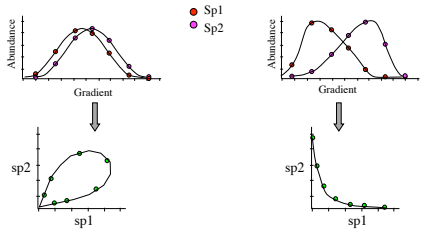
- Linear model



Basic concepts

Species associations – positive or negative

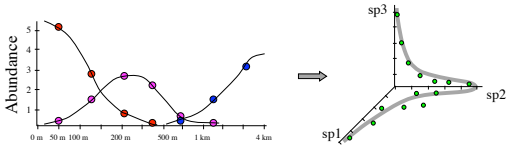
- Gaussian model



Basic concepts

Species associations

- Gaussian model

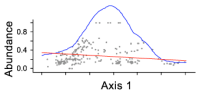


Sp1 (*T. truncatus*)
Sp2 (*S. attenuata*)
Sp3 (*P. macrocephalus*)

Basic concepts

Species on gradients – realistic scenario

- Species responses are "solid"
- Species responses are complex (polymodal, asymmetric, discontinuous)



Preparatory steps

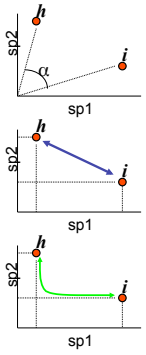
"Raw species data tends to have a few abundant species and many rare species"

- Delete rare species to reduce noise in data set w/o losing much information (<5% of SU's)
- Transformation: statistical and ecological reasons
  - Monotonic: power, log, logit, arcsine, square root
  - Relativization – row-wise or column-wise: by max, by mean, z-score, rank ordering

Preparatory steps

Distance matrix

- Correlation (r = cos α)
- Euclidean
- City block



Ordination techniques

- Ordination ≡ Arrangement of items along a scale to show similarities
- Most common use is to describe the strongest patterns of species composition
- Direct gradient analysis (constrained ordination): items are positioned according to environmental factors (CCA)
- Indirect gradient analysis: items are arranged according to covariation and association among species (reflects community response to environment)

Ordination techniques

Mechanics:

Reduce distance matrix to a small set of structures

$\mathbf{A}_{mxn}$ ( ) $\mathbf{D}_{mxm}$ $\mathbf{S}_{mxn}$ ( )

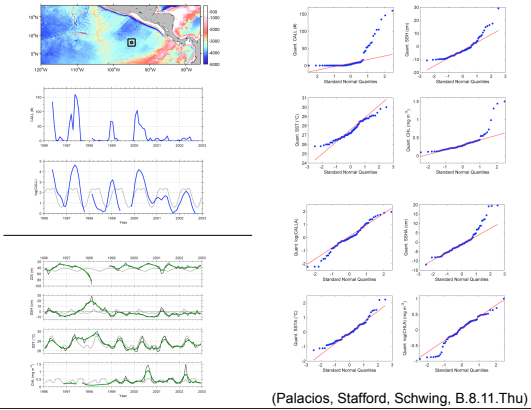
Assess effectiveness:

- Proportion of variation represented by each structure
- Correlation between ordination axis and species abundance
- Strength of relationship to a second matrix ($\mathbf{E}_{mxn}$)
- Comparison with null model (randomization)

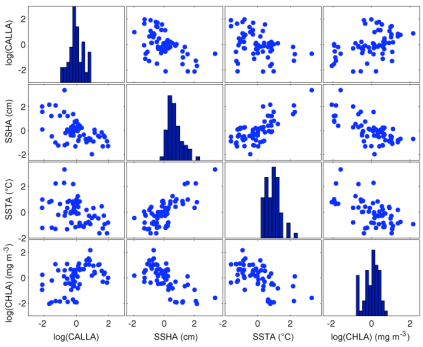
Ordination techniques

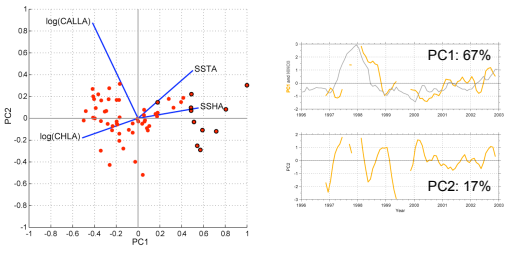
Principal Component Analysis (PCA)

- Requires multivariate normality, linearity
- Seeks strongest linear correlation structure among variables
- Decompose the correlation matrix into scores (the value of each SU on each axis) and loadings (correlation between each variable and each component)
- Strength: ideal for data w/ linear relationships among variables (e.g., environmental data)
- Weakness: linear model is poor fit to community data, leads to horse-shoe effect in ordination space



(Palacios, Stafford, Schwing, B.8.11.Thu)





Ordination techniques

Nonmetric Multidimensional Scaling (NMS)

- Developed by Kruskal (1964). No underlying model.
- Can use any type of distance and relativization scheme
- Based on iterative optimization methods (computer intensive)
- Iterates search for best SU positions in reduced space that minimizes stress.
- Strength: the use of ranked distances linearizes the relationships between distances in species space and in ordination space
- Weakness: finding unique solution because of local minima

Cetacean community around the Galapagos Is:

Abundance for 9 cetaceans (spotted, spinner, common, striped, bottlenose and Risso's dolphin; pilot, sperm, Bryde's whale).

Abundance matrix = $\mathbf{A}_{309x9}$

Environmental variables:

PC1 - scores from 1st principal component

PC2 - scores from 2nd principal component

PC3 - scores from 3rd principal component

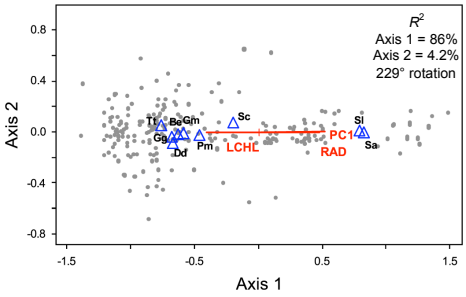
LCHL - log-transformed chl

RAD - distance from the center of the archipelago

Environmental matrix = $\mathbf{E}_{309x5}$

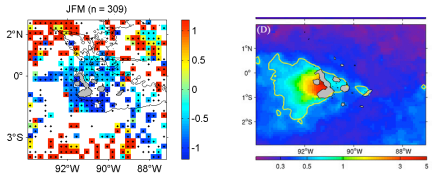
(Palacios, 2003)

NMS Ordination

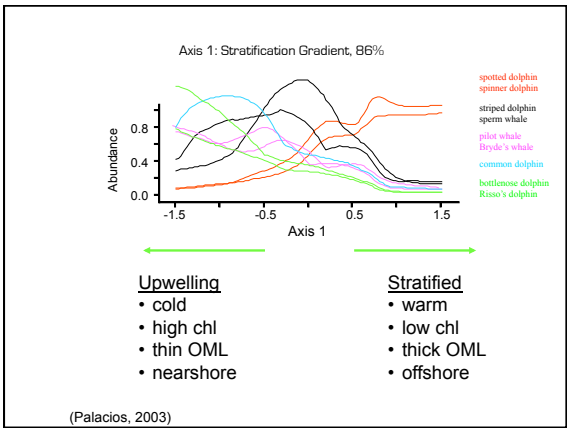


(Palacios, 2003)

Spatial representation of scores on Axis 1



(Palacios, 2003)



(Palacios, 2003)