PCA – Advanced Examples & Applications

- *Objectives:*

Showcase advanced PCA analysis:

- Addressing the assumptions

- Improving the signal / decreasing the noise

Principal Components (PCA) – Paper II

- Example: Ainley, D.G. et al. (2005).
- Objective: Relate densities of the 12 most abundant species of seabirds to 12 habitat variables:

5 biological, 4 oceanographic, 3 geographic (spatial)

Table 2
The 12 most abundant seabird species recorded during GLOBEC surveys in summer 2000

Species	Raw count	Adjusted count	Composition by number (%)
Sooty Shearwater *Puffinus griseus* (SHSO)	15,138	17,446.9	56.9
Common Murre *Uria aalge* (MUCO)	4982	5370.6	17.5
Cassin's Auklet *Ptychoramphus aleuticus* (AKCA)	2426	2421.8	7.9
Northern Phalarope *Phalaropus lobatus* (PHRN)	1515	1492.5	4.9
Fork-tailed Storm-Petrel *Oceanodroma furcata* (STFT)	892	894.8	2.9
Leach's Storm-Petrel *O. leucorhoa* (STLE)	725	853.1	2.8
Black-footed Albatross *Phoebastria nigripenis* (ALBF)	469	494.7	1.6
Northern Fulmar *Fulmarus glacialis* (FUNO)	507	426.3	1.4
Red Phalarope *Phalaropus fulicarius* (PHRE)	362	353.6	1.2
Pink-footed Shearwater *Puffinus creatopus* (SHPF)	383	345.6	1.1
Western Gull *Larus occidentalis* (GUWE)	231	201.3	0.7
Rhinoceros Auklet *Cerorhinca monocerata* (AKRH)	153	144.8	0.5
Total	27,783	30,446.9	99.2

82.3%

Principal Components (PCA) – Paper II

- Oceanographic variables examined:

sea-surface temperature / salinity, thermocline depth / strength

Table 1
Pearson linear correlation coefficients (r) among physical and biological environmental variables[a]; $n = 554$ survey transects, upwelling period 2000

	JD	Front A	Front B	Front C	SST	SSS	Therm-Depth	Therm-Str	Chl-Max	LL	ML	MH
Front A	0.18											
Front B	0.27	0.85										
Front C	0.60	0.33	0.38									
SST	0.02	0.55	0.54	−0.14								
SSS	0.67	−0.10	−0.08	0.55	−0.44							
Th-Dp	−0.29	0.32	0.27	−0.19	0.42	−0.26						
Th-St	0.08	0.36	0.36	−0.19	0.81	−0.37	0.29					
Chl-Max	−0.31	−0.20	−0.18	−0.34	−0.31	0.39	−0.31	−0.20				
LL-38	−0.59	0.06	−0.03	−0.36	0.23	−0.45	0.29	0.13	−0.24			
ML-120	0.37	0.21	0.28	0.19	0.24	0.11	0.00	0.20	0.06	0.38		
MH-200	0.53	0.20	0.36	0.31	0.13	0.22	−0.06	0.17	0.10	−0.01	0.73	
HH-420	0.50	0.33	0.39	0.33	0.21	0.14	−0.07	0.18	−0.00	−0.05	0.67	0.70

Date **Distance to Fronts** **Chl Max** **Acoustic Biomass**

Principal Components (PCA) – Paper II

- Data Manipulations To Avoid Biases:

- Densities log-transformed to meet normality assumptions
- Nevertheless, residuals generated in the regressions for some species did not meet those assumptions (Skewness / Kurtosis Test for Normality of Residuals, $p < 0.05$)
- Least-squares regression analysis (ANOVA), however, is a very robust procedure with respect to non-normality (Seber, 1977, Kleinbaum et al., 1988)
- Yet, while these analyses yield the best linear unbiased estimator in the absence of normally distributed residuals, p-values near 0.05 must be viewed with caution (Seber, 1977)

Principal Components (PCA) – Paper II

- To avoid double-absences:

- Only 15-min transects in which any given species was recorded were analyzed

- The total sample size for the 12 species was 1209

- Is this an adequate sample size ?

<u>Rule of thumb</u>:

- 5 samples per variable (Tabachnick and Fidell 1989)
- 1209 / 12 ~ 100 samples per variable

Principal Components (PCA) – Paper II

- Analysis Methods:

• Principal components analysis (PCA), in combination with Sidak multiple comparison tests, used to assess differences in habitat selection among 12 seabird species

• To test for significant differences in habitat affinities among seabird species, used two one-way ANOVAs:

In the first, tested for differences among PC1 scores of each species; in the second, compared the PC2 scores

• Differences between two species significant if either one or both PC scores differed significantly

Principal Components (PCA) – Paper II

- Community-Wide Result: First and second PC axes explained 60% of variance in distribution of 12 species

PC	Eigenvalue cumulative proportion
1	0.34
2	0.60
3	0.79
4	0.88
5	0.94
6	0.97
7	0.99
8	1.00

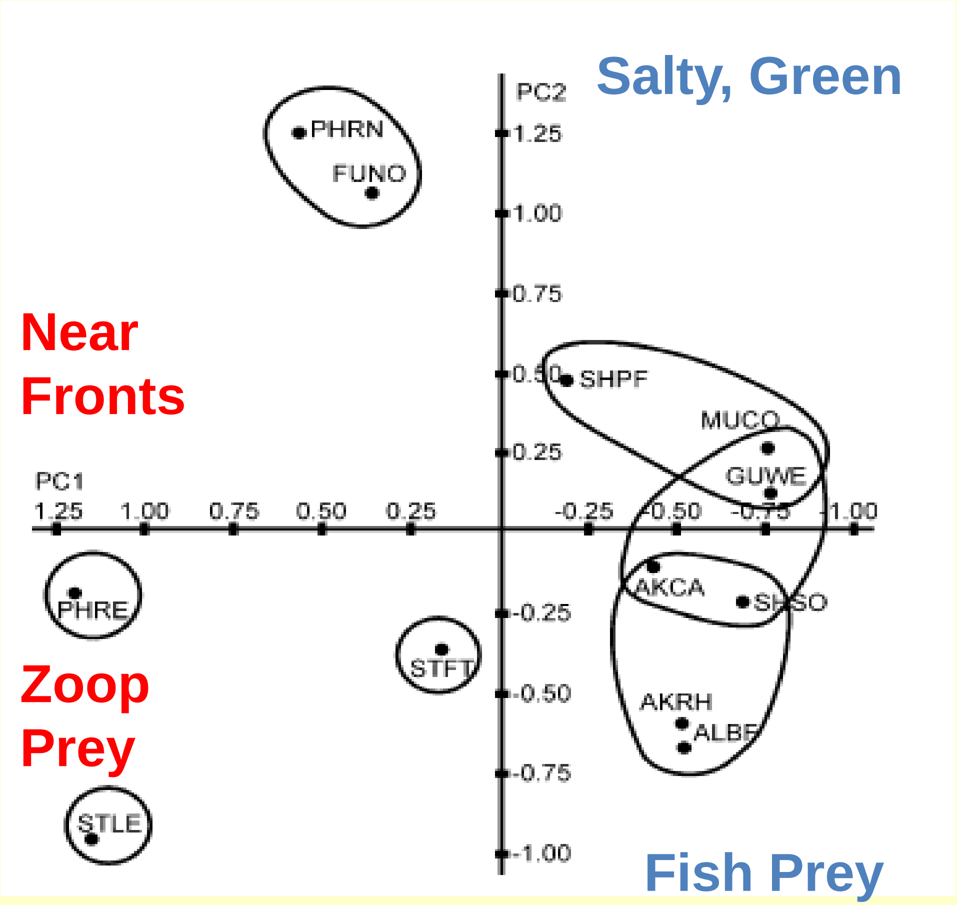
Environmental variable	Eigenvector loading	
	PC1	PC2
Front A	–0.49	–0.29
Front B	–0.55	–0.20
Sea-surface salinity	–0.09	0.58
Chlorophyll maximum	–0.08	0.48
38 kHz backscatter	0.01	–0.52
120 kHz backscatter	0.39	–0.06
200 kHz backscatter	0.43	0.17
420 kHz backscatter	0.41	0.05

Principal Components (PCA) – Paper II

- Species-specific Results:

• Species mapped onto two (independent) dimensions

• Pair-wise associations (tested) denoted by circles



Principal Components (PCA) – Comparisons

- Number of Axes:

- Selected 2 – easy to interpret **(Ainley et al. 2005)**

- Selected 6 – based on eigenvalues > 1 **(Weichler et al. 2004)**

- Display of Results:

- Plot & table of eigenvalues **(Ainley et al. 2005)**

- Eigenvalues & interpretation (description) **(Weichler et al. 2004)**

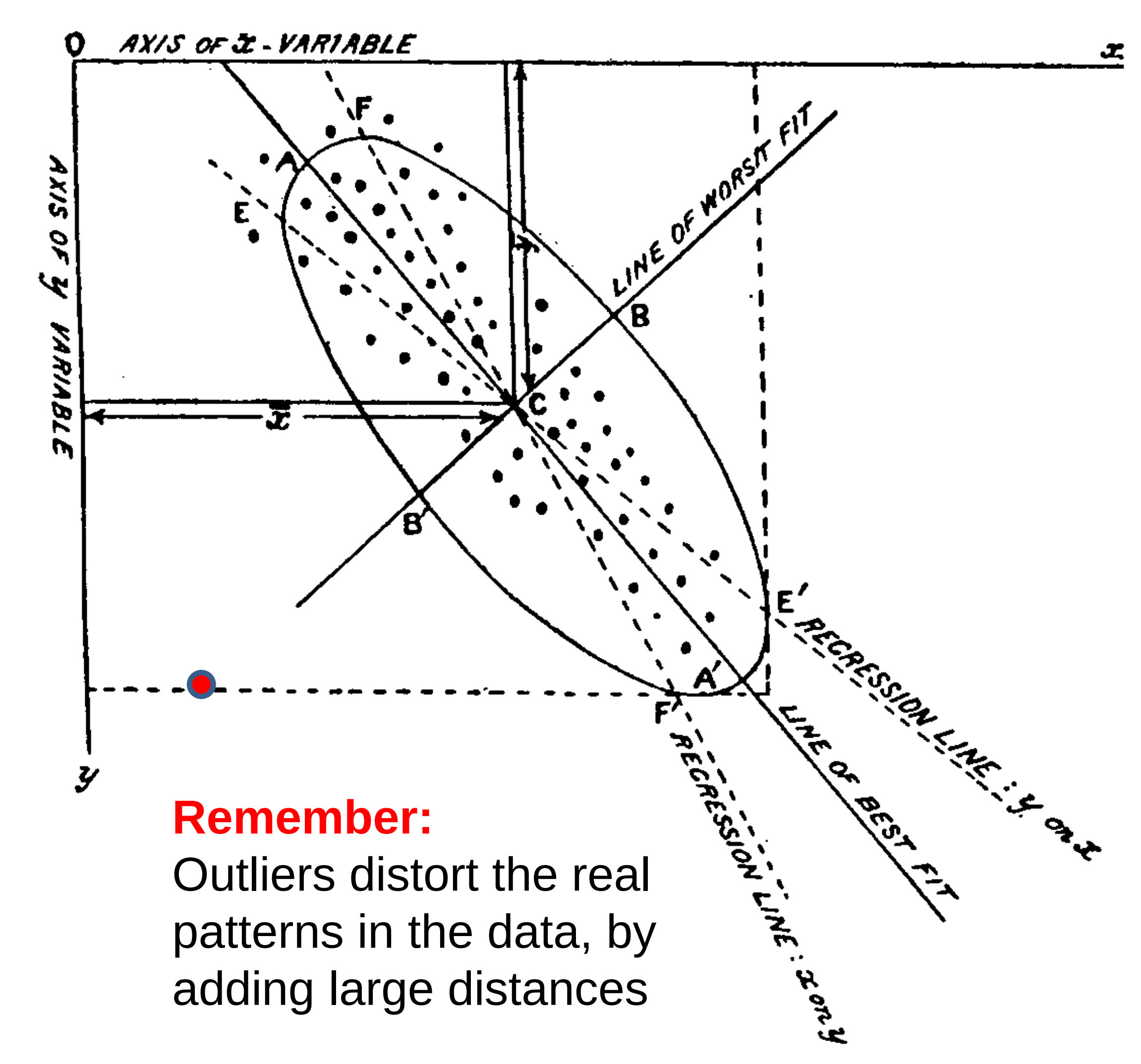
- Significance Tests:

- Pairwise species comparisons (ANOVA) **(Ainley et al. 2005)**

- Correlations with selected variables **(Weichler et al. 2004)**

Influence of Distances in PCA

PCA seeks the strongest patterns, with the largest distances:



Remember:
Outliers distort the real patterns in the data, by adding large distances

Mind the PCA Assumptions

Because it uses linear combinations of response variables to create the axes, PCA is subject to the assumptions of linearity in the relationships among the variables.

- Implicit assumption of linearity in relationships between the responses and gradients represented by ordination axes.
- Ordination axes uncorrelated (by definition, orthogonal).

Assumptions met by relatively homogeneous environmental datasets with few zeroes, and rarely met by species datasets.

Difficulties when PCA used on zero-rich species datasets because they usually violate normality and linearity assumptions(e.g., high skewness, difficult to normalize).

Reducing Noise / Improving Signal

Therefore, critical to determine if a strong influence by dominant species / variables is consistent with the underlying assumptions and with your analysis objectives.

Alternatively, you can think about what steps are needed to reduce their influence and to meet assumptions.

Options for reducing the influence of dominant species involve:

standardizing data to eliminate abundance differences

deleting highly dominant / rare species

creating subsets of variables to analyze separately

Standardizing the Data

- Relativization re-scales all of the data at once, using a common criterion / standard.
- When it is done by columns (e.g., species), variation across plots is retained, but variation across species is standardized.
- When its done by rows (e.g., plots), variation across species is retained, but variation across plots is standardized.

5	Stands			
5	Species			
	Q	Q	Q	Q
	A	B	C	D
s1	1	10	0.1	100
s2	2	20	0.2	200
s3	3	30	0.3	300
s4	4	40	0.4	400
s5	5	50	0.5	500

Sums: 15 150 1.5 1500

Data Relativizations in PC-ORD

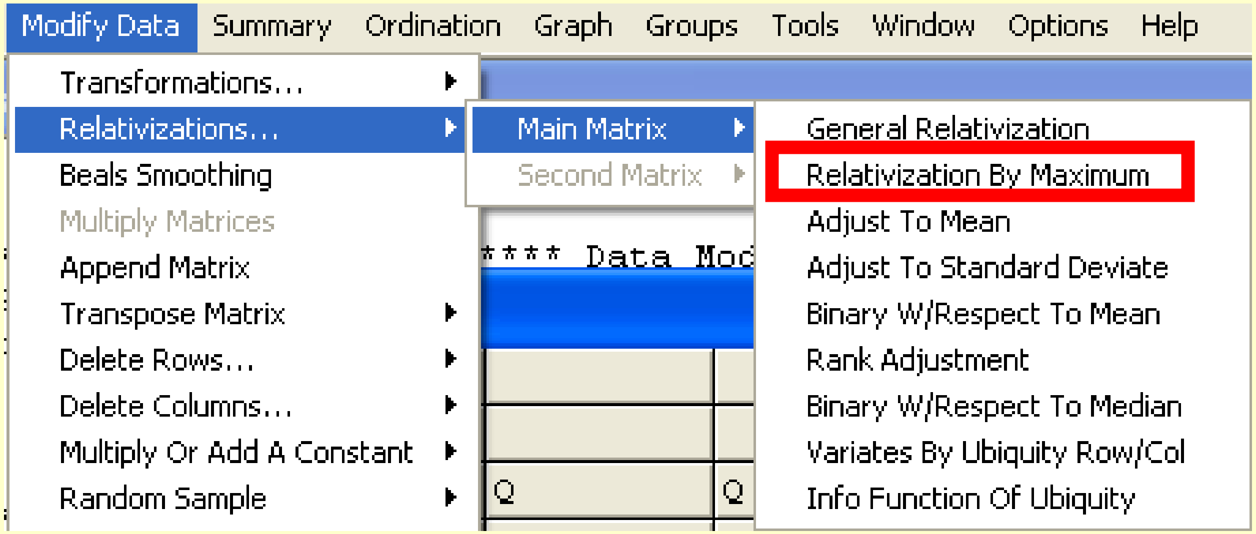
- Relativization by Maximum:

When relativization by maximum is set for columns, each cell in a column is divided by the maximum value in the column, replacing absolute values with proportions of the maximum observed value across all sample units.

This relativization approach is used when the maximum observed value of a given response across all sample units is considered the maximum *potential abundance for that response in* this population-species.

The relativized values for each response represent proportions of their maximum potential. They are applied to species data to equalize the influence of common / rare species and of abundant / nonabundant species.

Data Relativizations in PC-ORD



- Relativization by Maximum: (input: $x \geq 0$; output: from 0 to 1)

Divides each cell's value by the total (for the given row or the given column) – such that values range from 0 to 1.

5	Stands			
5	Species			
	Q	Q	Q	Q
	A	B	C	D
s1	1	10	0.1	100
s2	2	20	0.2	200
s3	3	30	0.3	300
s4	4	40	0.4	400
s5	5	50	0.5	500
Sums:	15	150	1.5	1500

5	Stands			
5	Species			
	Q	Q	Q	Q
	A	B	C	D
s1	0.200000	0.200000	0.200000	0.200000
s2	0.400000	0.400000	0.400000	0.400000
s3	0.600000	0.600000	0.600000	0.600000
s4	0.800000	0.800000	0.800000	0.800000
s5	1	1	1	1
Sums:	3	3	3	3

Data Relativizations in PC-ORD

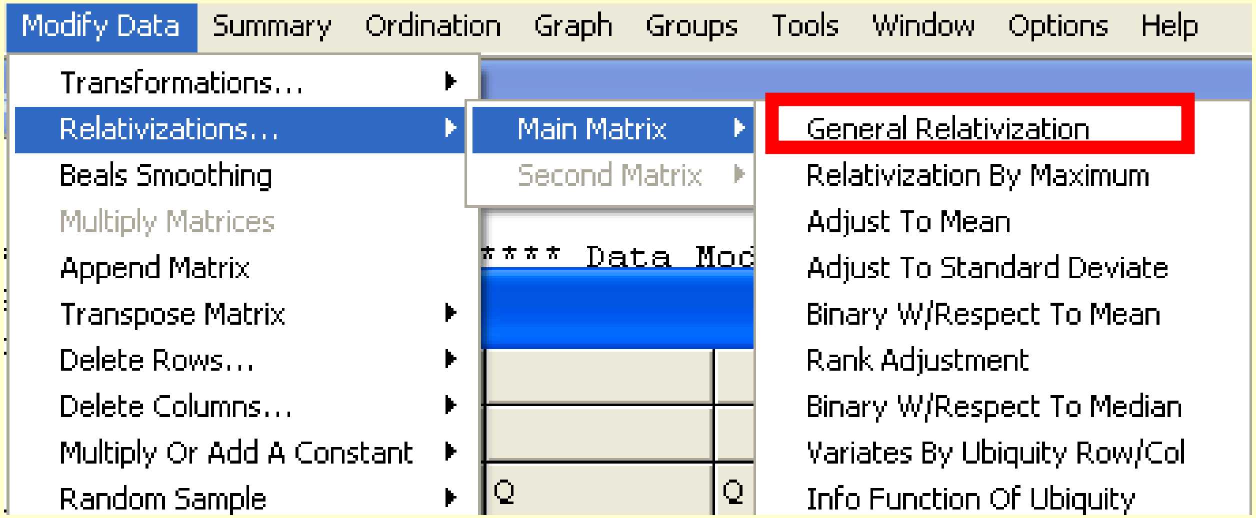
- General Relativization:

General relativization by column totals reduces influence of responses with high total abundance relative to those with low total abundance, because observations are proportional to their intra-response total abundance.

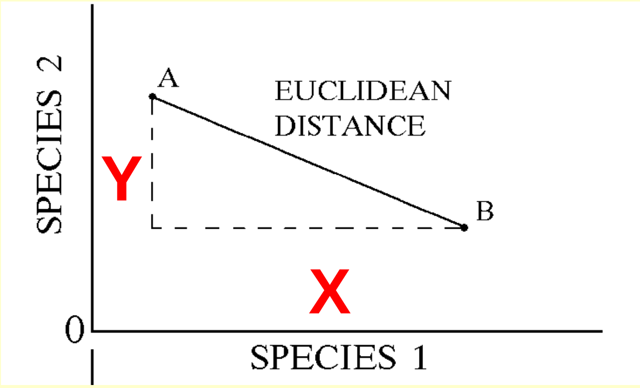
This retains the variation in abundances across sample units, but reduces the influence of very common species and increases the influence of rare species.

NOTE: General relativization by row totals is applicable if your question of interest focuses on giving each sample the same influence. HINT: this is not reasonable when the columns have different variables

Data Relativizations – PC-ORD



$$b = \frac{x_{ij}}{\sqrt[p]{\sum_{j=1}^{n} x_{j}^{p}}}$$



NOTE:
p can take on two values: 1 or 2

Remember:
City Block vs Euclidean

Data Relativizations in PC-ORD

$$b = \frac{x_{ij}}{\sqrt[p]{\sum_{j=1}^{n} x_j^p}}$$

- General Relativization: (input: x $\geq$ 0; output: from 0 to 1)

If p = 1, Relativization is by column (or row) totals. Appropriate for city-block distance measure (e.g., Sorensen).

If p = 2, Relativization is by column (or row) totals. Appropriate for Euclidean distance measure (square root).

General Relativizations

General Relativization: (by totals) makes area under each species distribution response curve = 1

By columns – generalized: (p = 1):

5	Stands			
5	Species			
	Q	Q	Q	Q
	A	B	C	D
s1	1	10	0.1	100
s2	2	20	0.2	200
s3	3	30	0.3	300
s4	4	40	0.4	400
s5	5	50	0.5	500

5	Stands			
5	Species			
	Q	Q	Q	Q
	A	B	C	D
s1	0.06666667	0.06666667	0.06666667	0.06666667
s2	0.1333333	0.1333333	0.1333333	0.1333333
s3	0.2	0.2	0.2	0.2
s4	0.2666667	0.2666667	0.2666667	0.2666667
s5	0.3333333	0.3333333	0.3333333	0.3333333

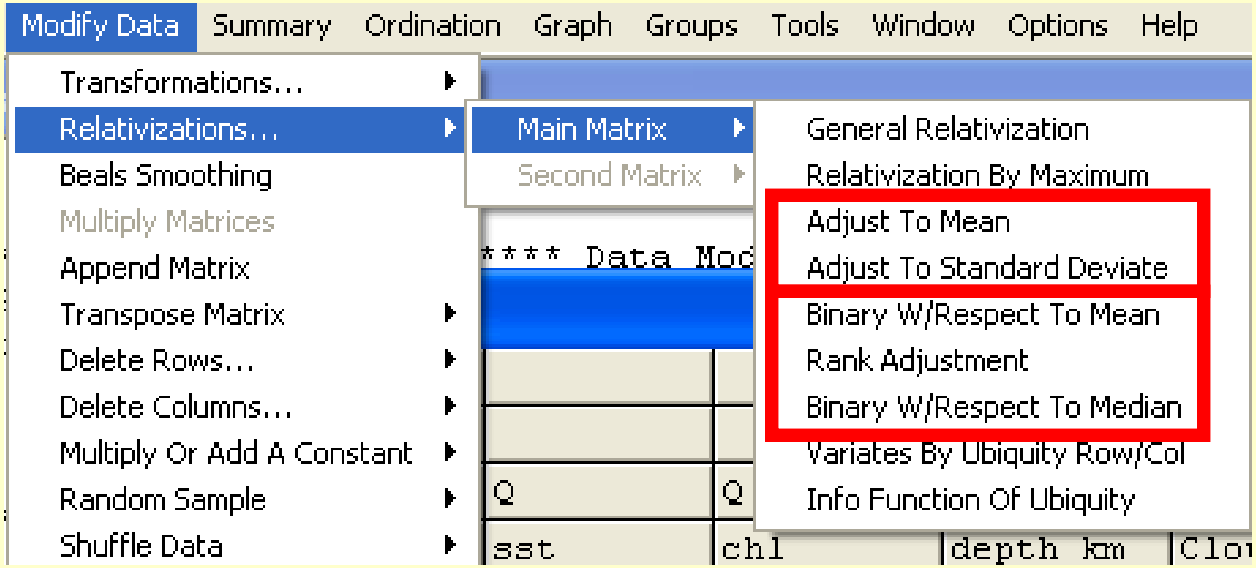
Sums: 1 1 1 1

By columns – generalized: (p = 2):

5	Stands			
5	Species			
	Q	Q	Q	Q
	A	B	C	D
s1	1	10	0.1	100
s2	2	20	0.2	200
s3	3	30	0.3	300
s4	4	40	0.4	400
s5	5	50	0.5	500

5	Stands			
5	Species			
	Q	Q	Q	Q
	A	B	C	D
s1	0.13484	0.13484	0.13484	0.13484
s2	0.26968	0.2696799	0.2696799	0.2696799
s3	0.4045199	0.4045199	0.4045199	0.4045199
s4	0.5393599	0.5393599	0.5393599	0.5393599
s5	0.6741999	0.6741998	0.6741998	0.6741998

Other Data Relativization – PC-ORD



- Deviations: Value – Mean

- Z scores: (Value – Mean) / SD

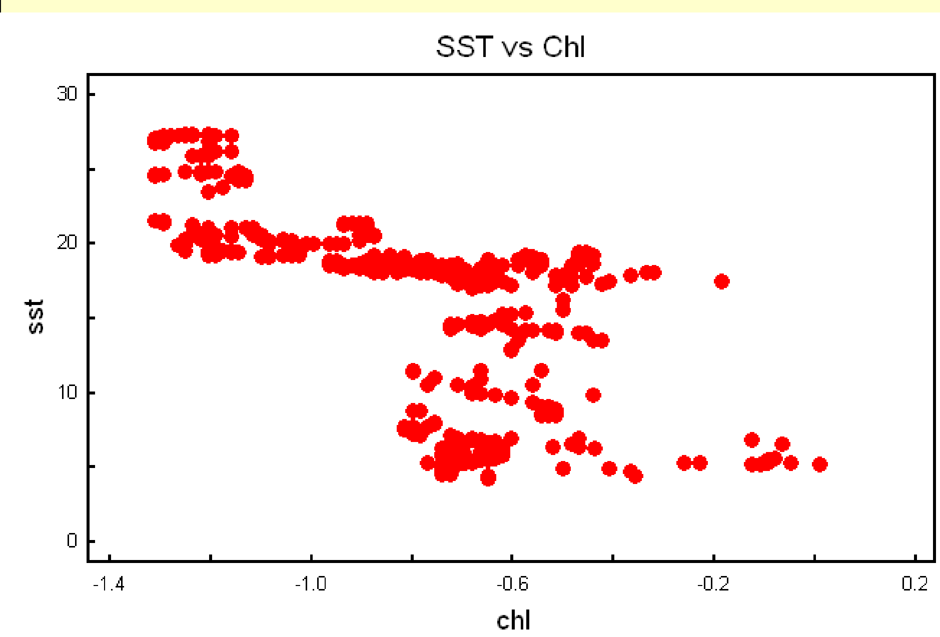
- Binary response: Above (1) / Below (0)

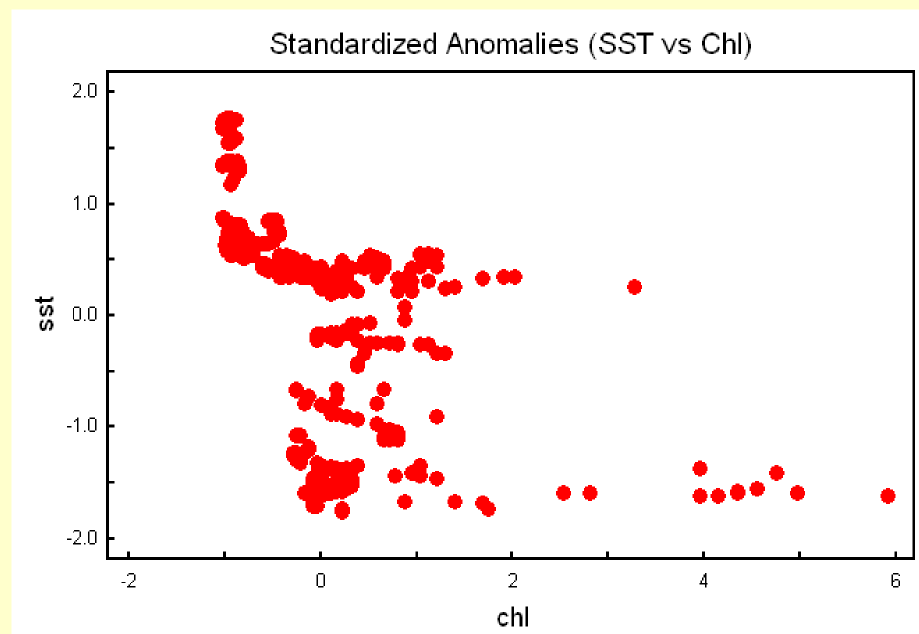
- Ranks: Assigns ranks
(e.g., 0, 0, 6, 9 would receive the ranks 1.5, 1.5, 3, 4)

Data Relativization – Remember …

Note:

Need to accept TEMP file





Relativizations – Recommendations

Do not use relativization by maximum when any data < 0

Do not use general relativization when any data < 0

Cannot use standard deviates with “empty” data groups (rows / columns) - Why not ?

NOTE: Fine to use with negative data

PCA Example – Upwell

Where do we start ? Data Exploration + Summarization

```
              Summary of:     5 vars          N =   240 samples
---------------------------------------------------------------------
----------------------
 Num.  Name        Mean       Stand.Dev.       Sum     Minimum     Maximum
---------------------------------------------------------------------
    1 time        1996.000       5.786    479040.0000  1986.042    2005.958
    2 MEI          -0.4816       2.817      -115.6      -11.75        8.678
    3 PDO          -0.3167       1.731       -76.01      -5.429       3.489
    4 upwell36      5.371       49.488     1289.0000   -131.000     200.000
    5 upwell39     21.150       64.645     5076.0000   -296.000     240.000
---------------------------------------------------------------------
AVERAGES:     404.3       24.89        0.9704E+05  308.4      491.6       239.2
---------------------------------------------------------------------
```

What do we look for ?

Value Ranges – Typos, Possible Transformations

Unequal Sums – Different Weights

PCA Example – Upwell

What do we look for ?

```
                    Skewness  Kurtosis
----------------------------------
    1 time             0.000    -1.162
    2 MEI             -0.460     2.669
    3 PDO             -0.140    -0.162
    4 upwell36         0.846     1.802
    5 upwell39        -0.021     2.934
----------------------------------
   Averages:           0.045     1.216
```

```
        1200 cells in main matrix
Percent of cells empty =     0.333
```

-1 < Skewness < +1

Few Vacant Cells

PCA Example – Upwell

How can we solve unequal sums (weights) of variables ?

Relativize by Maximum (Columns)

Summary of: 5 vars N = 240 samples

Num.	Name	Mean	Stand.Dev.	Sum	Minimum	Maximum
1	time	0.995	0.003	238.8085	0.990	1.000
2	MEI	-0.5550E-01	0.3246	-13.32	-1.354	1.000
3	PDO	-0.9078E-01	0.4961	-21.79	-1.556	1.000
4	upwell36	0.027	0.247	6.4450	-0.655	1.000
5	upwell39	0.088	0.269	21.1500	-1.233	1.000
	AVERAGES:	0.1927	0.2681	46.26	-0.7616	1.000

PCA Example – Upwell

How can we solve unequal sums (weights) of variables ?

Standard Deviates by Columns

Summary of: 5 vars N = 240 samples

Num.	Name	Mean	Stand.Dev.	Sum	Minimum	Maximum
1	time	-0.1490E-07	1.000	-0.3576E-05	-1.721	1.721
2	MEI	-0.3204E-07	1.000	-0.7689E-05	-3.999	3.251
3	PDO	0.4470E-07	1.000	0.1073E-04	-2.954	2.199
4	upwell36	-0.1490E-07	1.000	-0.3576E-05	-2.756	3.933
5	upwell39	0.3104E-08	1.000	0.7451E-06	-4.906	3.385
AVERAGES:		-0.2806E-08	1.000	-0.6735E-06	-3.267	2.898

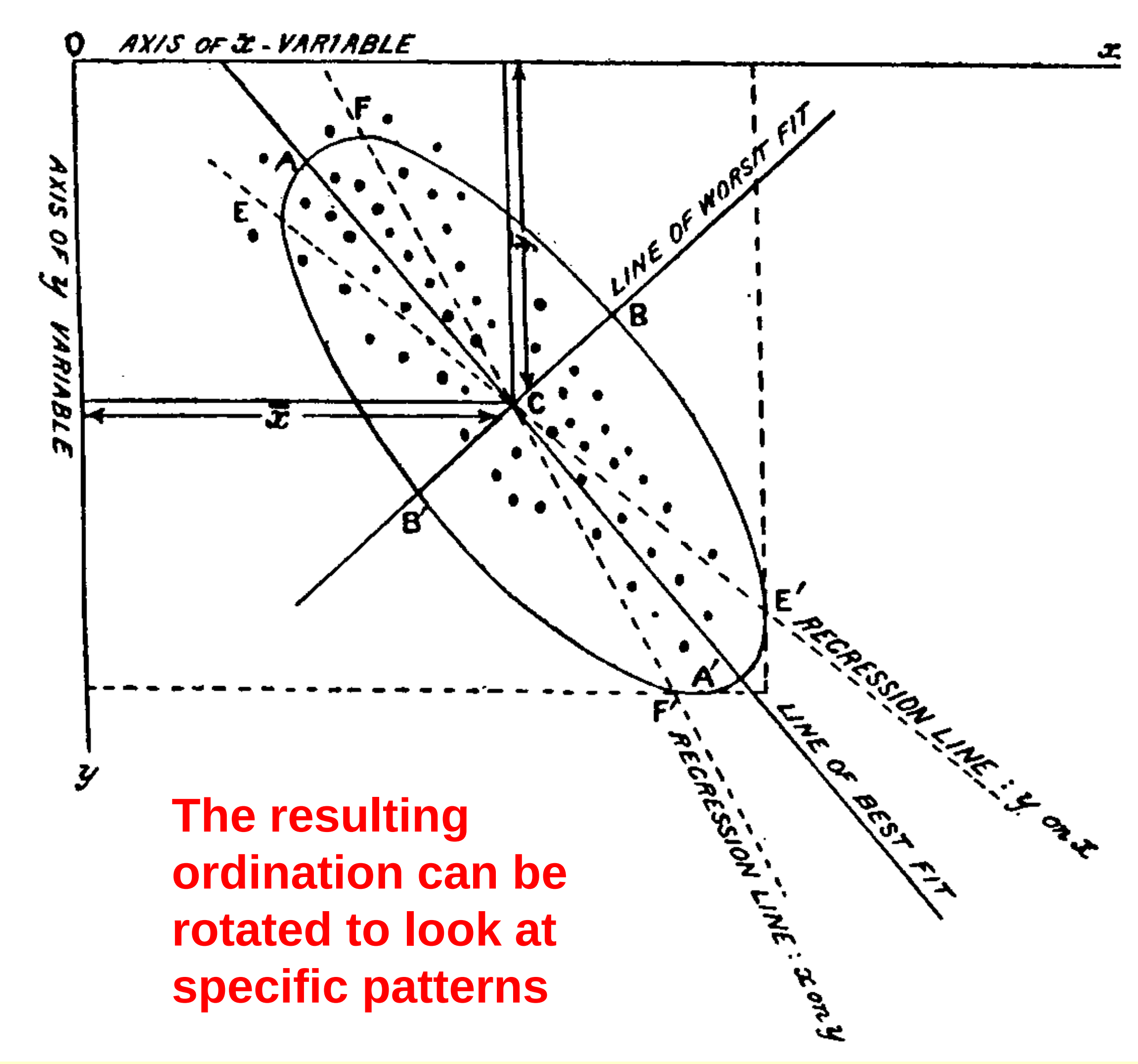
Mind your Relativizations

Not all datasets amenable to all relativizations: some are mathematically incompatible, others fail to relativize the samples / species.

Check Data Ranges / Sums BEFORE
Check Data Ranges / Sums AFTER

Rotating the Ordination

PCA seeks the strongest patterns, with the largest distances:



The resulting ordination can be rotated to look at specific patterns

PCA Tools – Rotation

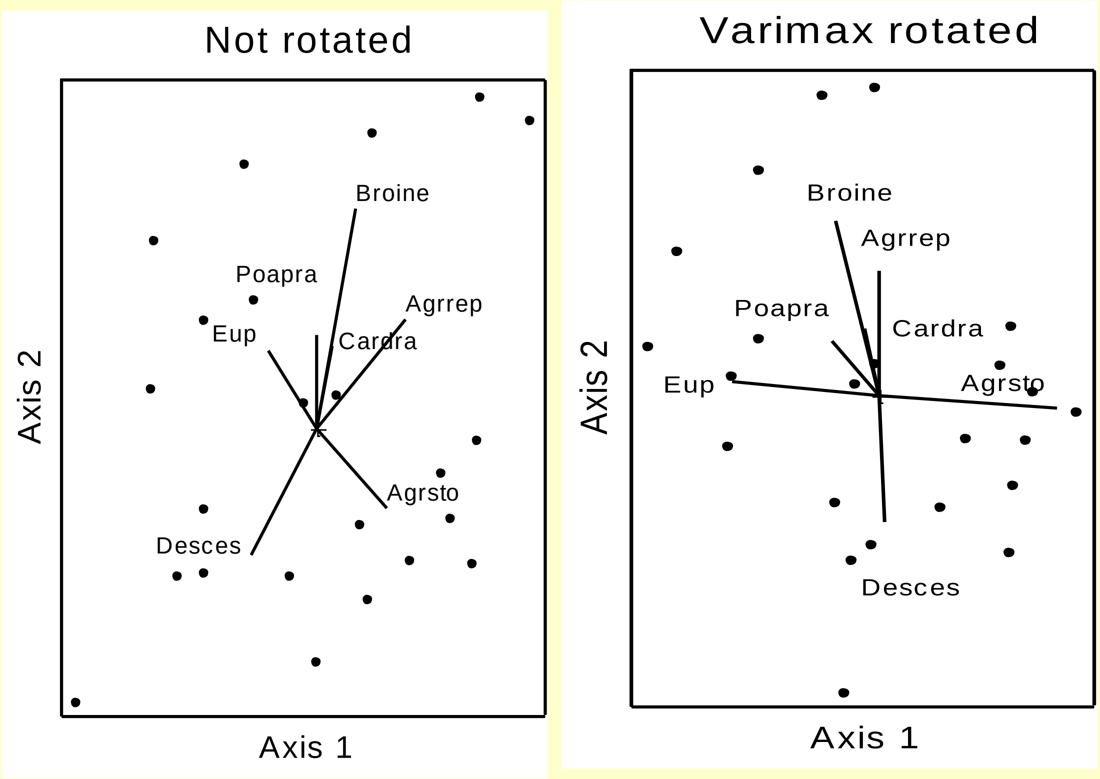
- Results: Rotation (VARIMAX)

Samples:

points

Species:

vectors



Comparison of ordination of sample units in species space before and after varimax rotation. Note the improved alignment of the species vectors with the ordination axes in the rotated ordination.

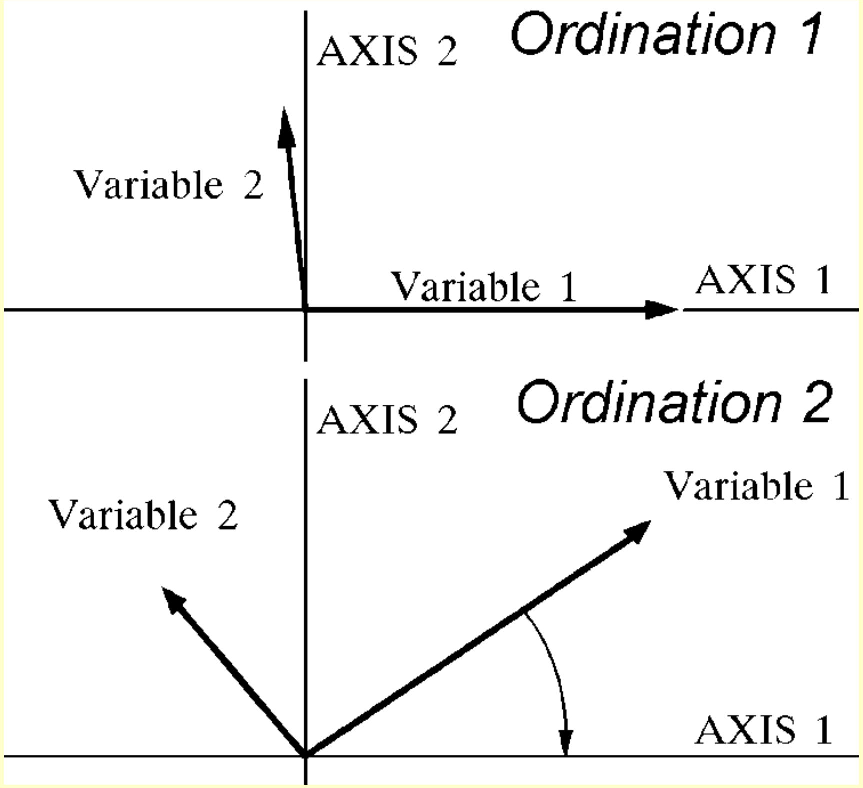
PCA Tools - Rotation

Rotation to align patterns from separate ordinations facilitates comparisons across studies:

In Ordination 1, the point cloud has been rotated to maximize loading of Variable 1 onto Axis 1.

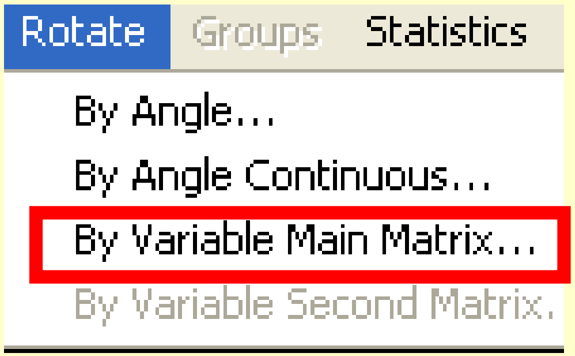
In Ordination 2, the same dominant trends were found but at an angle to those found in Ordination 1.

Therefore, Ordination 2 can be rotated through an angle (shown by arrow) so that it aligns Variable 1 with Axis 1.



PCA Tools – Rotation

- Rotation aligns ordination to highlight certain patterns



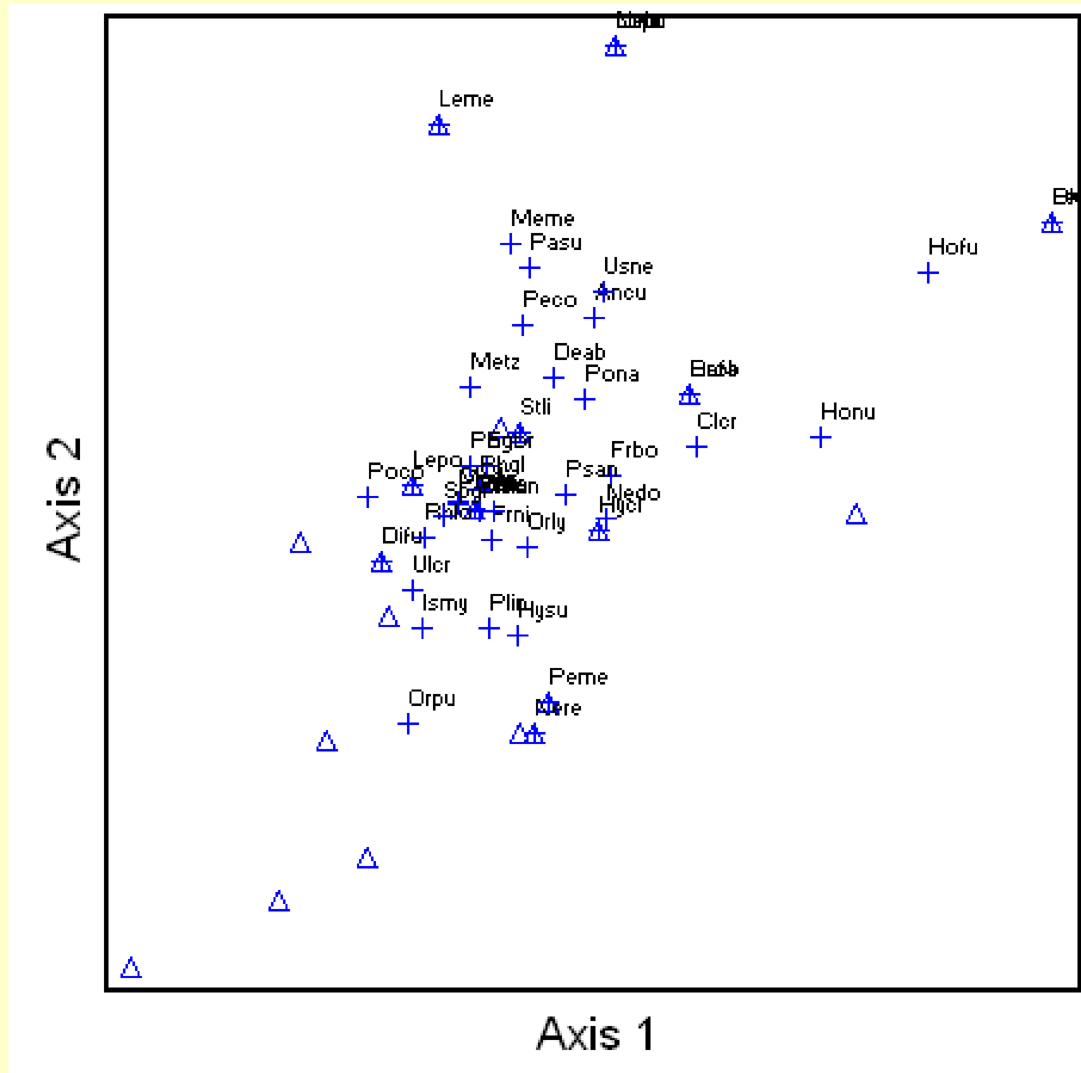
NEDO Axes Loadings

Axis 1: + 0.51

Axis 2: - 0.74

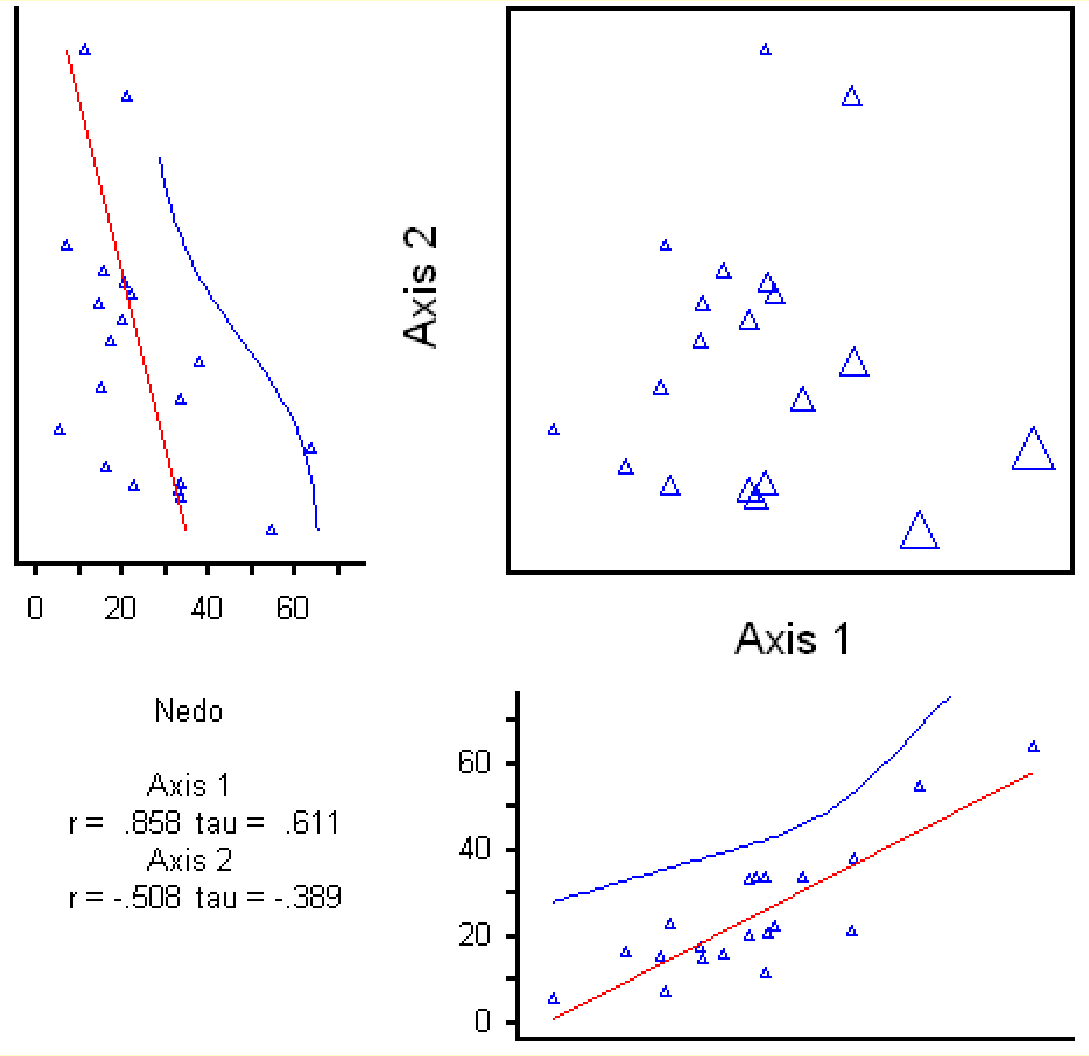
- Rotation by NEDO

Stretch plot along direction of most variation for species



PCA Tools – Rotation

- Looking at a Specific Species Response



Correlations

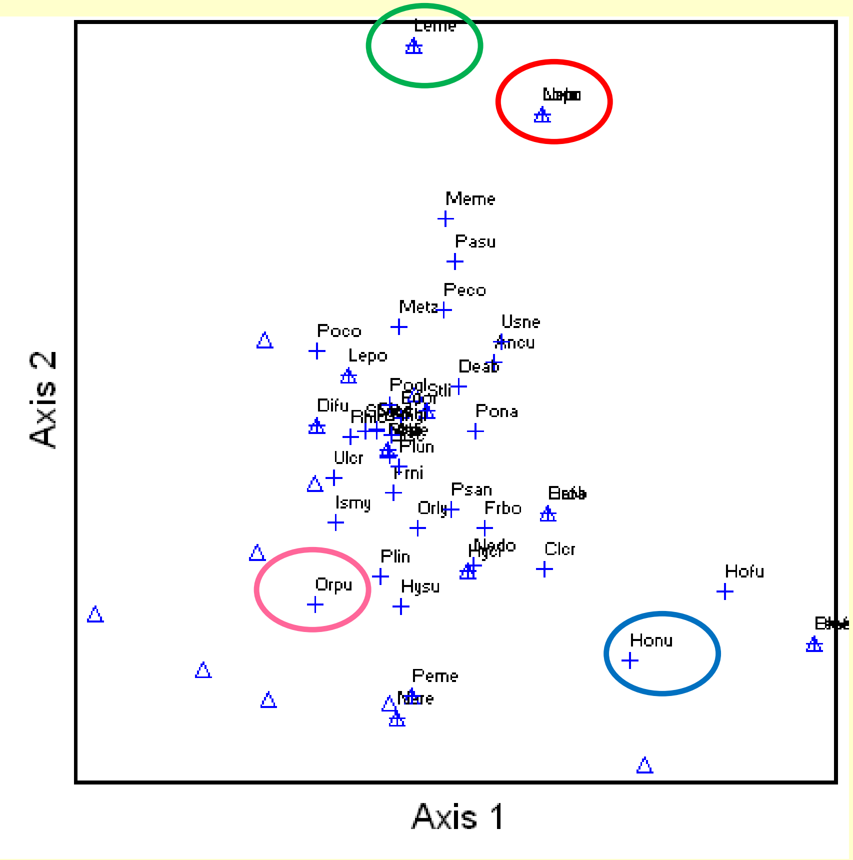
NEDO

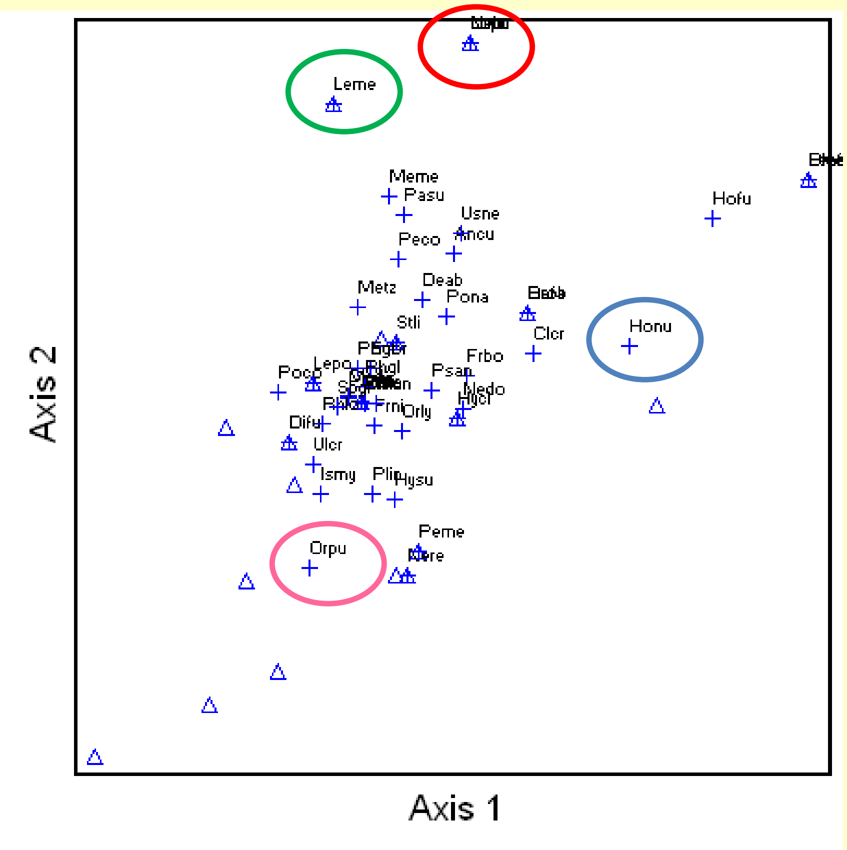
Axis 1: + 0.86

Axis 2: - 0.51

PCA Tools - Rotation

- Rotation aligns ordination to highlight certain patterns





NOTE: loadings of the species on the axes and the correlations of the species with the axes will change after rotation is implemented

Mind your Rotations

Report all rotations in results.

Check Axis Correlations / % Variance BEFORE
Check AxisCorrelations / % Variance AFTER

PCA Next Steps – Example 1

- Use PCA to synthesize cross-correlated environmental variables into independent (orthogonal) patterns

- Use new synthetic variables to compare categorical variables (groups) using ANOVA / GLMs

Marine Pollution Bulletin 64 (2012) 2588–2598

Contents lists available at SciVerse ScienceDirect

Marine Pollution Bulletin

ELSEVIER

journal homepage: www.elsevier.com/locate/marpolbul

Baseline

Persistent organic pollutants in the endangered Hawaiian monk seal (*Monachus schauinslandi*) from the main Hawaiian Islands

Jessica Lopez [a,b,*], Daryle Boyd [c], Gina M. Ylitalo [c], Charles Littnan [d], Ronald Pearce [c]

PCA – Example of Next Steps

- Principal Component Analysis (PCA) used to assess patterns of shared variation in 71 POP analytes.

  6 DDTs,47 PCB congeners, 8 chlordane isomers,
  3 hexachlorohexanes dieldrin, mirex, aldrin,
  hexachlorobenzene , and 10 PBDE congeners.

- Considered three categorical variables:

  - Three age / sex groups compared in the analysis: juveniles, adult males, and adult females.

  - Two sample origins: necropsy (dead) / biopsy (alive).

  - Two tissues sampled: serum (blood) and fat.

PCA – Example of Next Steps

- Sample Outliers:

- Data log transformed and examined to determine the existence of outliers (> 3 S.D. deviations from mean).

- Two adult male outliers (one high and one low) were removed for statistical analysis following these criteria.

- Empty Variables:

- POP analytes that were below LOQ in > 75% of samples removed to reach recommended 5:1 sample / variable ratio

PCA – Example of Next Steps

- Significant PCA axes selected using alpha = 0.05, using 999 randomizations.
- One significant PCA axis accounted for 74.89% of variance.

Table 4
Loading values on one significant PCA axis and Pearson correlations (*r* values) of POPs with PCA Axis 1.

POP	Axis 1	
	loading value	*r* value
Trans-nonachlor	−0.2106	−0.630
pcb 17	0.3206	0.983
pcb 18	0.3247	0.990
pcb 28	0.2897	0.979
pcb 31	0.3295	0.985
pcb 33	0.3433	0.977
pcb 66	0.0398	0.298
pcb 70	0.2288	0.869
pcb 99	−0.1320	−0.852
pcb 105	−0.1059	−0.718
pcb 110	0.2052	0.939
pcb 118	−0.1205	−0.799
pcb 128	−0.1679	−0.924
pcb 138	−0.1713	−0.983
pcb 149	0.0227	0.160
pcb 153	−0.1892	−0.930
pcb 170	−0.1864	−0.825
pcb 180	−0.1974	−0.800
pcb 183	−0.1881	−0.860
pcb 187	−0.2052	−0.814
ppDDE	−0.2299	−0.625

PCA – Example of Next Steps

- Sample loading values compared using ANOVA to assess whether common patterns of POP levels associated with different age/sex groups (juvenile, adult ♀, adult ♂), origin (live biopsy vs. necropsy), or tissue (blubber vs. serum).

  - No significant differences among age/sex groups in PCA loading values, indicating that shared variation of POPs did not differ between age/sex groups.

  - Significant difference between the 2 sample origins ($p = 0.02$), suggesting a difference in POPs between necropsy and live animal samples.

  - Significant difference between blubber samples and serum samples ($p < 0.001$).

PCA Next Steps – Example 2

- Use PCA to synthesize cross-correlated environmental variables into independent (orthogonal) patterns

- Use new synthetic variables to explain other response variables (like species counts) using other statistical methods (GLMs, GAMs)

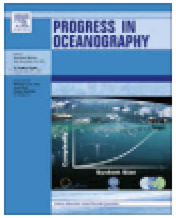
Progress in Oceanography 84 (2010) 242–254

Contents lists available at ScienceDirect

Progress in Oceanography

journal homepage: www.elsevier.com/locate/pocean

ELSEVIER



Top-down and bottom-up factors affecting seabird population trends in the California current system (1985–2006)

David G. Ainley [a,*], K. David Hyrenbach [b,c]

PCA – Next Steps

- Published Example: Ainley & Hyrenbach (2010).
- Objective: Relate seabird densities to five cross-correlated environmental variables: MEI, PDO, upwelling 39, upwelling 36, SST

Table 2
Pearson linear correlation coefficients (r) among environmental variables; $n = 276$ monthly values during 23 yr (October 1984–September 2007).

Correlation coefficient	MEI	PDO	UP_S	UP_N
p-Value				
MEI	–	0.492	−0.150	−0.139
PDO	<0.001	–	−0.266	−0.234
UP_S	0.002–0.005	<0.001	–	0.593
UP_N	0.02–0.05	<0.001	<0.001	–

PCA – Next Steps

- Objective: Also considered lagged environmental data: winter, early spring, late spring

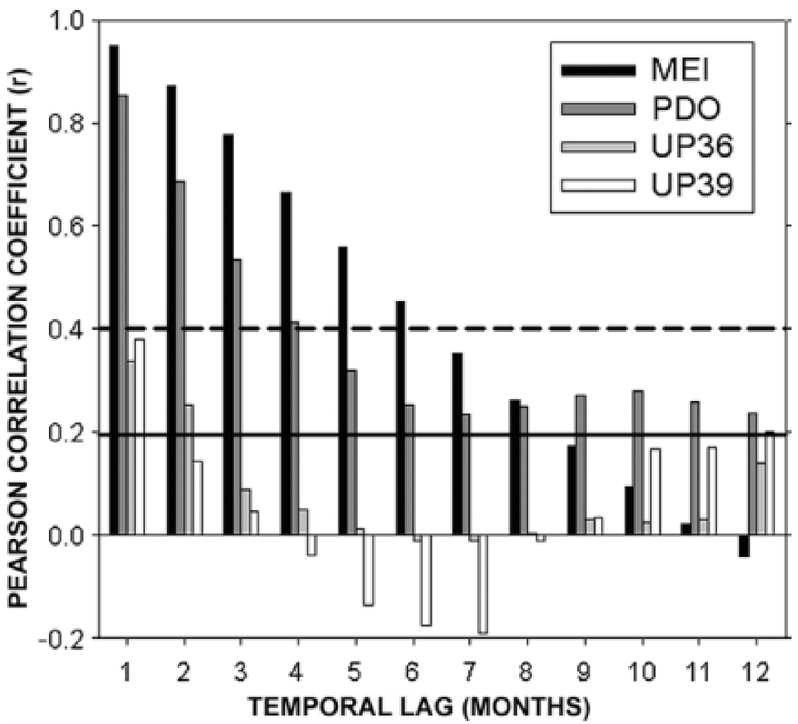


Fig. 3. Autocorrelations of four monthly environmental variables over a range of temporal lags from 1 to 12 mo. The horizontal lines indicate two levels of statistical significance defined by alpha = 0.001 (solid line) and 0.0001 (dashed line).

PCA – Next Steps

- Results: Four PC axes described 83 % of variability

Table 3
Results of principal component analysis showing the loadings of the different components of the four dominant multi-variate environmental factors and the proportion of the observed variance explained by each factor. Those variables with large loadings (>0.5) are in bold font.

Process – time period	Variable	PC1	PC2	PC3	PC4
ENSO – winter	MEI_1	**+0.790**	+0.348	−0.145	+0.018
ENSO – early spring	MEI_2	**+0.840**	+0.398	−0.004	-0.122
ENSO – late spring	MEI_3	**+0.708**	+0.371	+0.400	-0.259
PDO – winter	PDO_1	+0.497	+0.092	−0.081	**+0.822**
PDO – early spring	PDO_2	**+0.757**	+0.172	−0.105	**+0.554**
PDO – late spring	PDO_3	**+0.762**	-0.207	+0.434	+0.232
Upwelling North – winter	$UP\text{-}39_1$	−0.396	-0.444	**+0.619**	+0.153
Upwelling North – early spring	$UP\text{-}39_2$	**−0.609**	**+0.572**	+0.164	+0.216
Upwelling North – late spring	$UP\text{-}39_3$	**−0.704**	**+0.513**	−0.078	+0.059
Upwelling South – winter	$UP\text{-}36_1$	−0.471	−0.166	**+0.643**	+0.288
Upwelling South – early spring	$UP\text{-}36_2$	-0.466	**+0.651**	+0.425	+0.011
Upwelling South – late spring	$UP\text{-}36_3$	**−0.785**	+0.373	−0.113	+0.146
Sea surface temp. – late spring	SST	**+0.685**	+0.236	+0.476	−0.356
Eigenvalues		5.78	1.91	1.64	1.42
Cumulative variance explained		0.44	0.59	0.72	0.83

Assessed temporal trends in PC factors using Spearman rank correlations (df = 19, rs critical = 0.433). Tests indicated no trends in spring-time environmental conditions sampled during the study period:

PC1 (rs = 0.195, 0.50 > p > 0.20), PC2 (rs = +0.238, 0.50 > p > 0.20), PC3 (rs = 0.005, p > 0.50), and PC4 (rs = +0.018, p > 0.50).

PCA – Next Steps

- Results:

Related seabird densities to 4 PC factors and "time" using GLM tests:

- R squared
- P value
- # of variables

Table 6

Results of multiple regression models of the relationship between seabird density and environmental variables. The total number of significant variables and the percent of variance explained (adjusted r^2) by the best-fit model are shown, alongside the associated *p*-values for each of the five independent variables considered. The sign (±) of the coefficients are shown for marginally significant ($p < 0.10$). Significant ($p < 0.05$) variables are highlighted in bold font. Highly significant variables, adjusted for multiple comparisons (*alpha* = 0.05/16 = 0.003), are underlined. Diving and surface-foraging species are listed separately, in decreasing order of abundance. See Table 4 for definitions of the species acronyms.

	Variables	r^2	PC1	PC2	PC3	PC4	Year
SHSO	2	22.5	0.375	−0.052	0.612	+0.095	0.971
MUCO	2	29.8	+0.020	−0.090	0.248	0.197	0.580
AKCA	**1**	**62.1**	**0.895**	**0.675**	**0.575**	**0.224**	**−0.001**
COBR	1	23.1	+0.018	0.161	0.970	0.703	0.490
AKRH	**1**	**20.4**	**0.800**	**0.492**	**0.411**	**0.520**	**−0.026**
GUPI	**2**	**58.8**	**0.005**	**0.437**	**0.939**	**0.949**	**−0.030**
GUWE	**3**	**58.2**	**0.450**	**−0.040**	**0.735**	**+0.091**	**−0.001**
PHRE	1	25.7	−0.013	0.329	0.729	0.954	0.120
GUCA	**0**	**0**	**0.218**	**0.867**	**0.798**	**0.344**	0.621
ALBF	**2**	**37.3**	**0.224**	**0.224**	**+0.056**	**0.264**	**−0.003**
SHPF	**0**	**0**	**0.174**	**0.792**	**0.558**	**0.529**	0.223
STAS	**1**	**19.0**	**0.335**	**0.106**	**0.843**	**0.428**	**−0.031**
FUNO	2	21.5	−0.091	+0.062	0.577	0.377	0.349
PELB	1	18.1	+0.035	0.552	0.628	0.295	0.641
GUSA	1	18.8	0.537	0.326	+0.032	0.675	0.877
STLE	1	21.9	+0.020	0.310	0.786	0.100	0.362

PCA – Next Steps

- Results: Species with significant responses to PC1

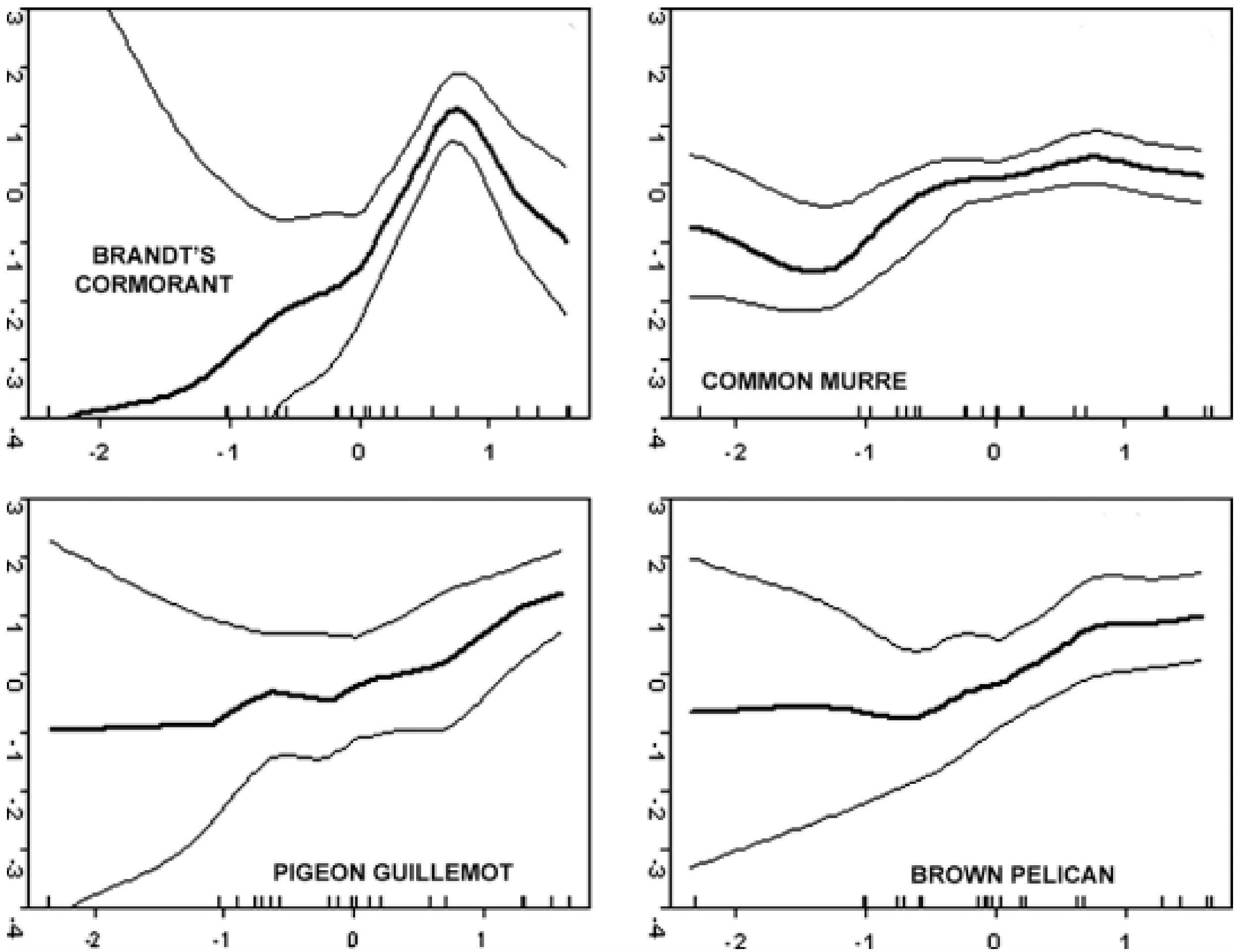


Fig. 6. Generalized additive models (GAM) relating seabird density (number km^{-2}) to PC1 scores. Only those species with significant responses ($p < 0.05$) which occurred in >75% of annual cruises are shown: brandt's cormorant, pigeon guillemot, common murre, brown pelican.

Summary – Next Steps 1

PCA synthesized complex patterns into orthogonal axes

Other statistical tests performed with resulting PC loadings

This allows performing categorical comparisons (i.e., ANOVA)

Summary – Next Steps 2

PCA synthesized complex patterns into orthogonal axes

Other statistical tests performed with resulting PC loadings

This allows relating species abundances (non-normal data) to the PCA factors using other statistics (i.e., GLMs, GAMs)