Separating spatial and temporal variation in multi-species community structure using PERMANOVA, a permutational MANOVA

By Kathy Mier
Alaska Fisheries Science Center

Outline

- Explain motivation
- Provide overview of Community Structure Analyses
- Illustrate advantages of PERMANOVA over ANOSIM and BIOENV
- Give example having spatial and temporal effects
- Describe other applications

Common questions asked by fishery biologists:

1) Does abundance of ***a single species*** vary significantly among regions or across years?

2) Does the ***community structure*** (multiple species as they relate to each other in terms of abundance or biomass) vary significantly among regions or across years?

3) Do certain ***environmental variables*** help explain any differences found?

ANOVA vs. MANOVA

univariate

	Species 1
Site 1	25
Site 2	15
Site 3	7
Site 4	8

multivariate

	Species 1	Species 2	Species 3	Species 4
Site 1	25	28	1	20
Site 2	15	18	0	0
Site 3	7	10	0	0
Site 4	8	0	100	7

Community Structure Analyses Overview

•Ecological distance (difference) between samples/variables is defined by an appropriate dissimilarity coefficient.

•Graphical Display of community pattern
- •classification (e.g. cluster analysis)
- •ordination (e.g. NMDS)

•Test for spatial and/or temporal differences between predefined groups, usually using nonparametric methods.

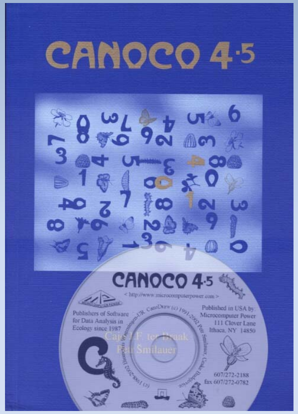
•Determine species most responsible for groupings.

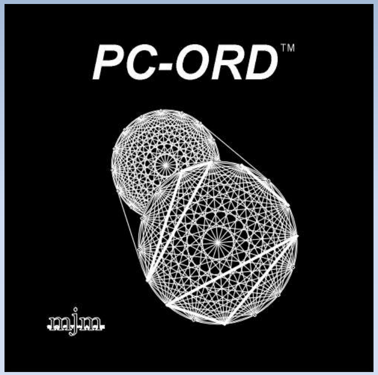
•Linking community patterns to environmental variables.

Specialized Community Structure Software

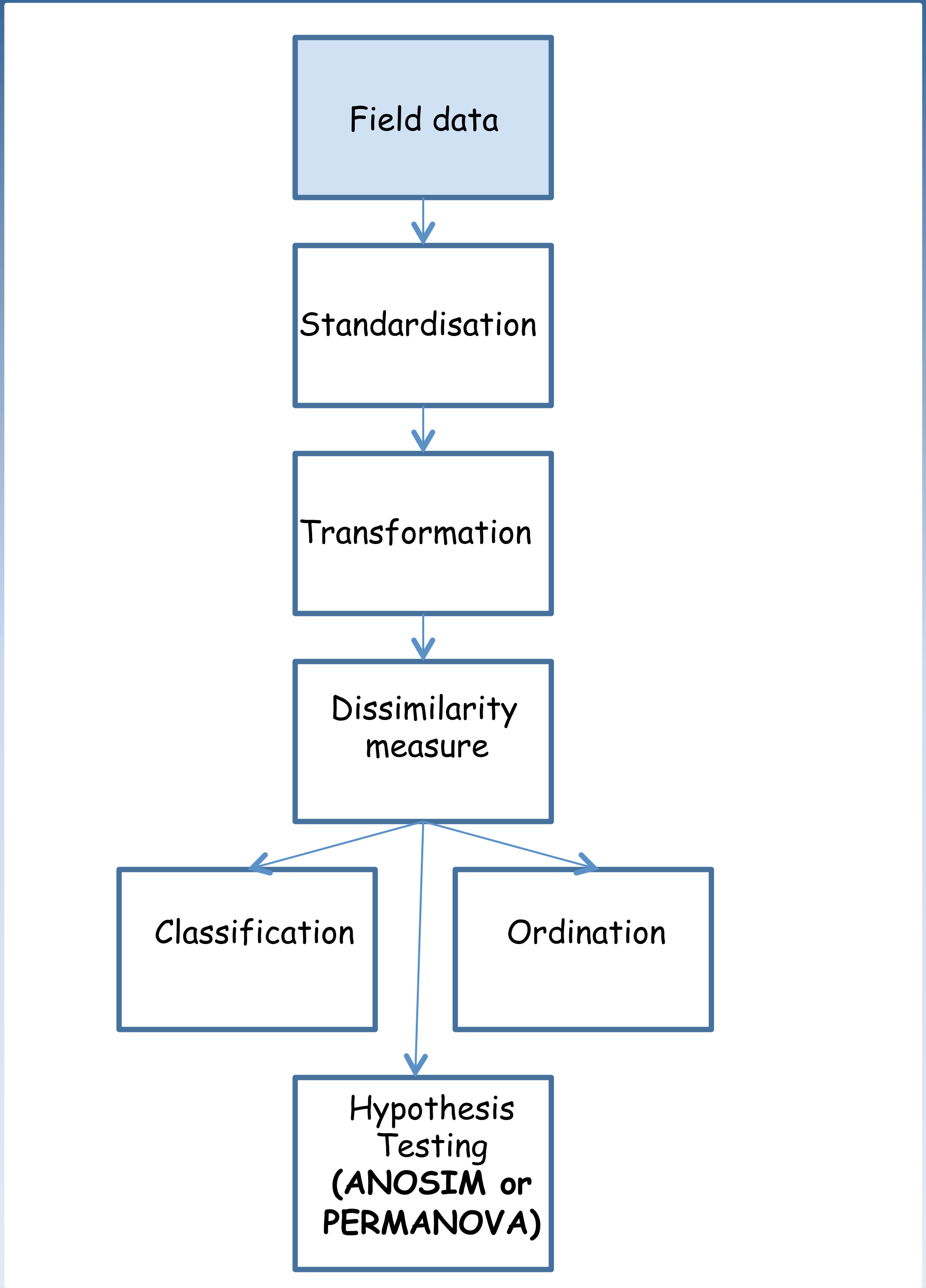
Recurrent Group (Fager)

Decorana/Twinspan





PRIMER-E


Field data
Standardisation
Transformation
Dissimilarity measure
Classification
Ordination
Hypothesis Testing
(ANOSIM or PERMANOVA)

Spatial/Temporal Effects

Data by site

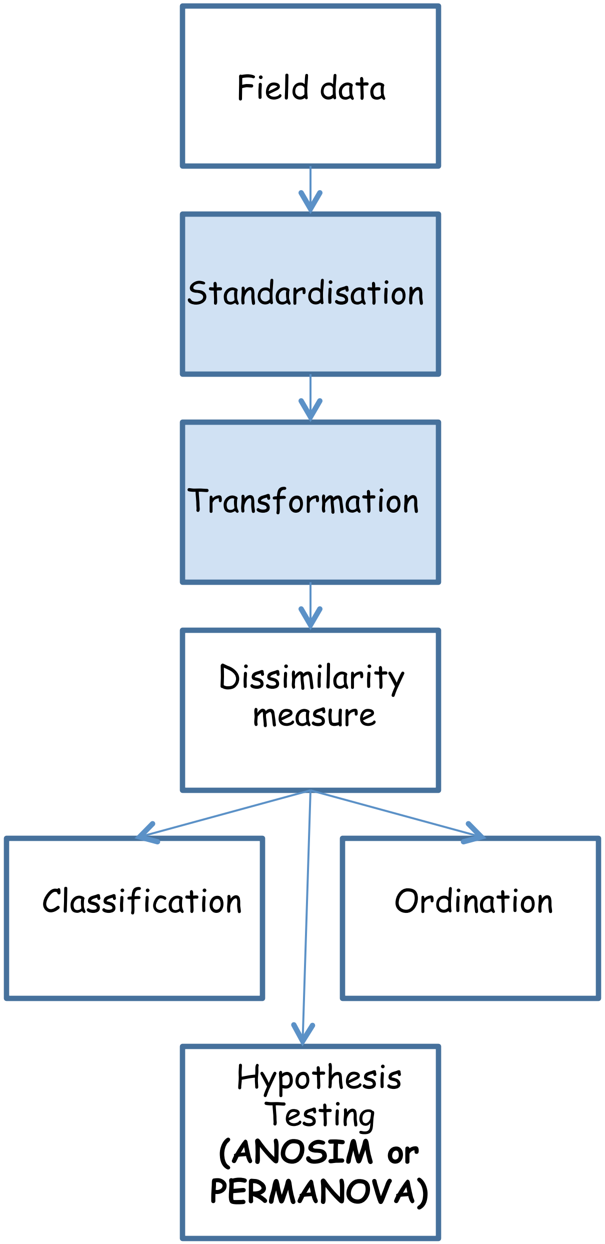
	Species 1	Species 2	Species 3	Species 4
Site 1	25	28	1	20
Site 2	15	18	0	0
Site 3	7	10	0	0
Site 4	8	0	100	7

Test for region effect

Data by year (means/totals)

	Species 1	Species 2	Species 3	Species 4
Year 1	25	28	1	20
Year 2	15	18	0	0
Year 3	7	10	0	0
Year 4	8	0	100	7

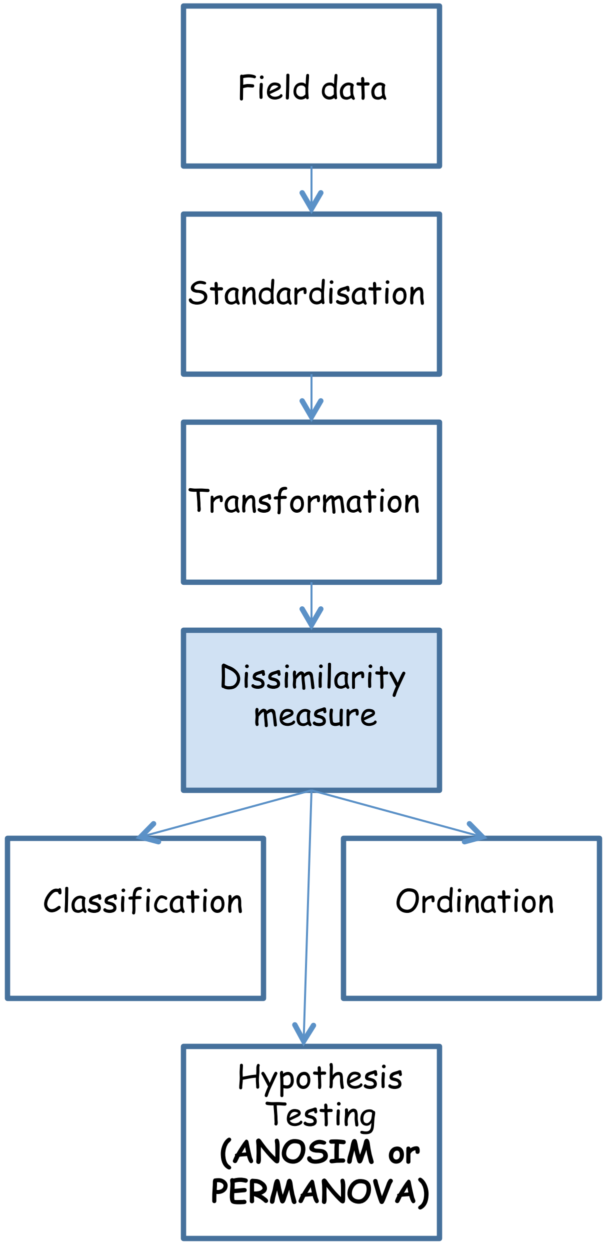
Test for warm/cold effect


Field data
Standardisation
Transformation
Dissimilarity measure
Classification
Ordination
Hypothesis Testing
(ANOSIM or PERMANOVA)

Should data be standardised and /or transformed?

- If comparing ***samples*** (sites), then use 4th root or square root transformation.

- If comparing ***species***, then standardize by species totals or maximum with no transformation.



Field data
Standardisation
Transformation
Dissimilarity measure
Classification
Ordination
Hypothesis Testing
(ANOSIM or PERMANOVA)

Bray-Curtis Dissimilarity Coefficient

$$d^{BCD}(i,j) = \frac{\sum_{k=1}^{n} \left| y_{i,k} - y_{j,k} \right|}{\sum_{k=1}^{n} \left(y_{i,k} + y_{j,k} \right)}$$

for sites i and j, for n species

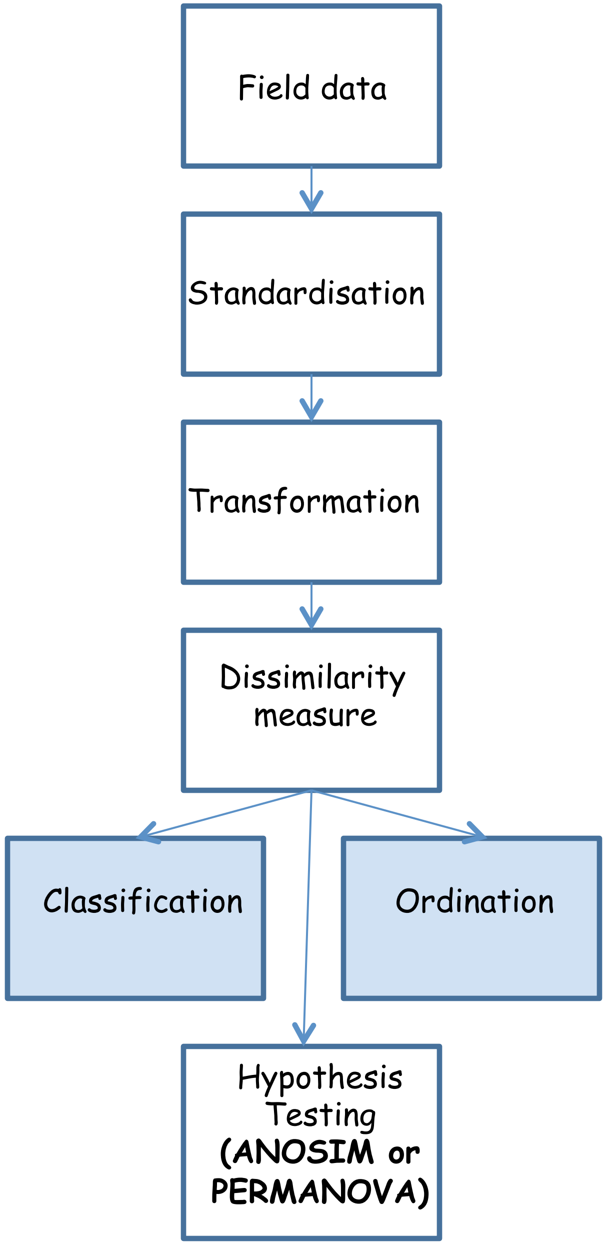
Bray-Curtis Dissimilarity

Raw Data

	Site 1	Site 2	Site 3	Site 4
Species 1	25	28	1	20
Species 2	15	18	0	0
Species 3	7	10	0	0

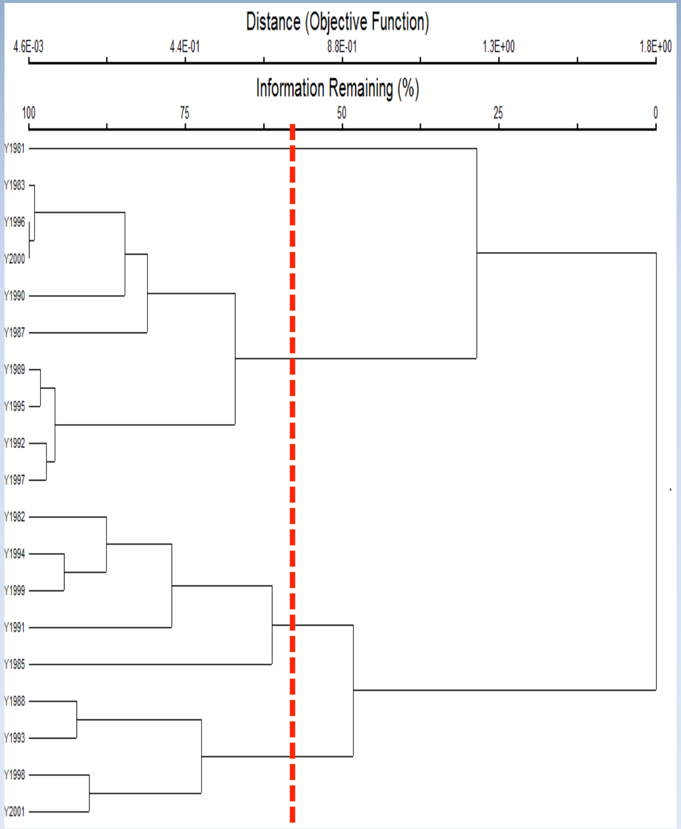
Bray-Curtis of sites

Site	1	2	3	4
1	0			
2	0.087	0		
3	0.958	0.965	0	
4	0.403	0.474	0.905	0

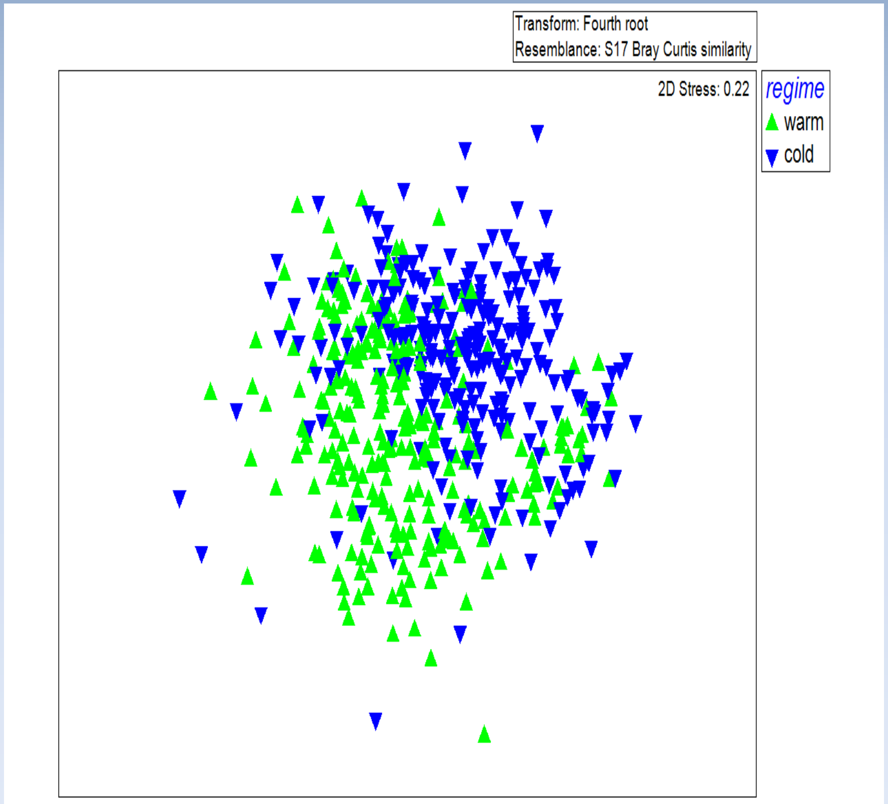

Field data
Standardisation
Transformation
Dissimilarity measure
Classification
Ordination
Hypothesis Testing
(ANOSIM or PERMANOVA)

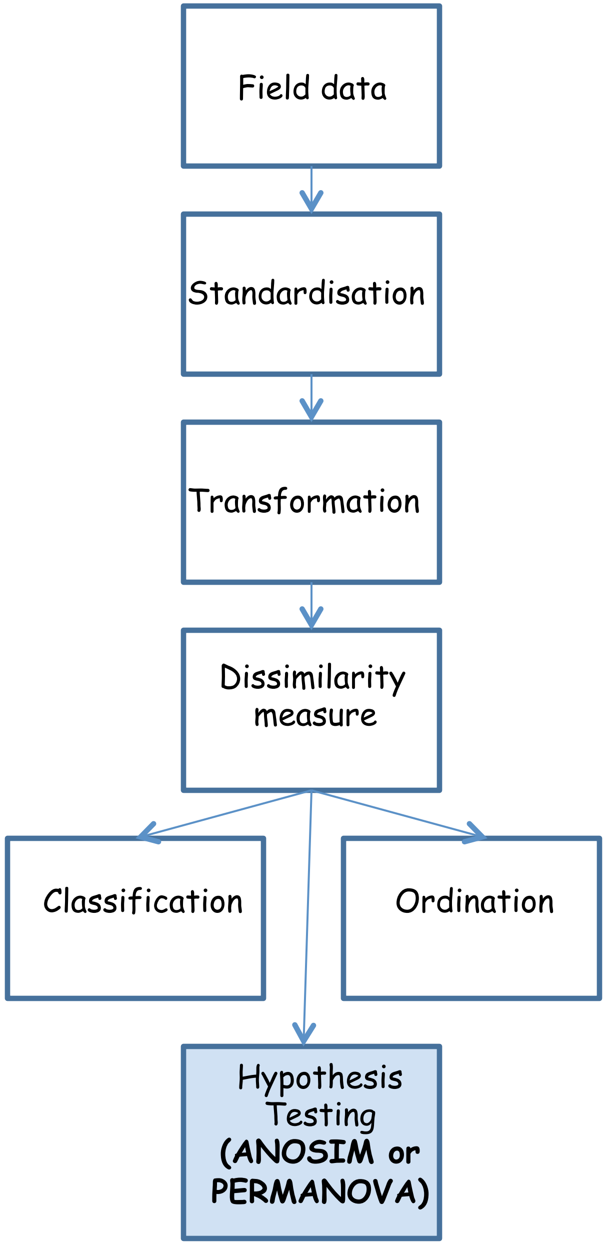
Descriptive Analyses

Cluster Analysis

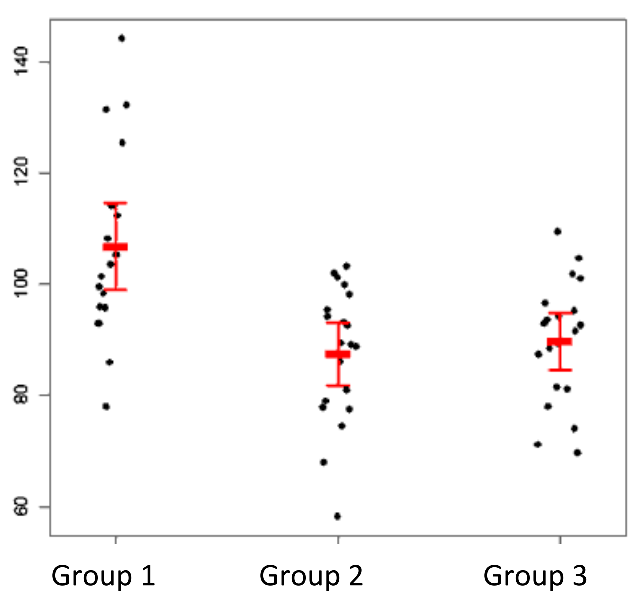


NMDS




Field data
Standardisation
Transformation
Dissimilarity measure
Classification
Ordination
Hypothesis Testing
(ANOSIM or PERMANOVA)

One way ANOVA



Source of Variation	d.f.	SS	MS	F_0
Factor A (between groups)	a-1	$SSA = \sum_{i=1}^{a} n_i \left(\bar{y}_{i.} - \bar{y}_{..}\right)^2$	$MSA = \frac{SSA}{(a-1)}$	$\frac{MSA}{MSE}$
Error (within groups)	N-a	$SSE = SST - SSA$	$MSE = \frac{SSE}{(N-a)}$	
Total	N-1	$SST = \sum_{i=1}^{a} \sum_{j=1}^{n} \left(y_{ij} - \bar{y}_{..}\right)^2$		

Why not use classical MANOVA??

•MANOVA, unlike ANOVA, is ***NOT robust to violations of parametric assumptions***. Most ecological data is overdispersed, nonnormal (right-skewed), with lots of zeros.

•MANOVA requires the number of variables to be less than number of samples

What do we do?

- Use analyses based on ***dissimilarity coefficients***

- Use ***permutation techniques*** that do not require parametric assumptions, since no distribution is being assumed

    - ANOSIM (Analysis of Similarity)
    - PERMANOVA (Permutational MANOVA)

What is ANOSIM?
(*Analysis of similarity*)

•A nonparametric multivariate ANOVA using ***ranked distance or dissimilarity*** between pairs of samples or variables (Bray-Curtis)

•Uses ***R test statistic***: $$R = \frac{\bar{r}_B - \bar{r}_W}{\frac{1}{2}n(n-1)}$$

•Uses **permutation techniques** to compute p-value, therefore does not assume a particular distribution

What is "PERMANOVA"? (*Permutational MANOVA*)

•An extension to ANOSIM, using ***distance or dissimilarity*** between pairs of samples or variables (Bray-Curtis)

•Uses ***Pseudo-F test statistic***: $$F = \frac{SS_A / a - 1}{SS_W / N - a}$$

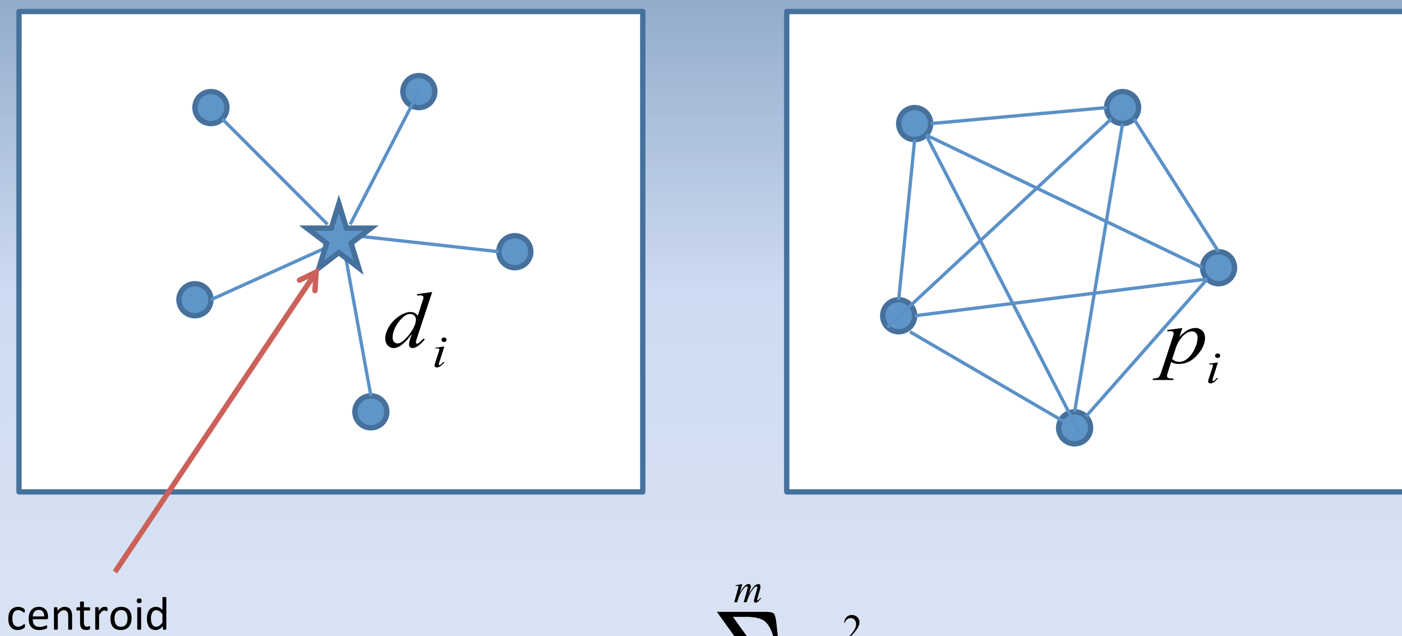
•Uses **permutation techniques** to compute p-value, therefore does not assume a particular distribution

Why use PERMANOVA over ANOSIM?

- PERMANOVA , uses ***actual Bray-Curtis*** coefficients where ANOSIM uses only ***ranks of Bray-Curtis,*** therefore preserving more information

- PERMANOVA allows for ***partitioning of variability***, similar to ANOVA, therefore allowing for more complex designs (multiple factors, nested factors, interactions, covariates)

Huygen's theorem



$$\sum_{i=1}^{n} d_i^2 = \frac{\sum_{i=1}^{m} p_i^2}{n}$$

PERMANOVA partitioning of SS

Site	1	2	3	4
1	0			
2	0.087	0		
3	0.958	0.965	0	
4	0.403	0.474	0.905	0

SS_T = SS of all / N

		Group 1		Group 2	
	Site	1	2	3	4
Group 1	1	0			
	2	0.087	0		
Group 2	3	0.958	0.965	0	
	4	0.403	0.474	0.905	0

SS_W = SS of each group /n

$$SS_A = SS_T - SS_W$$

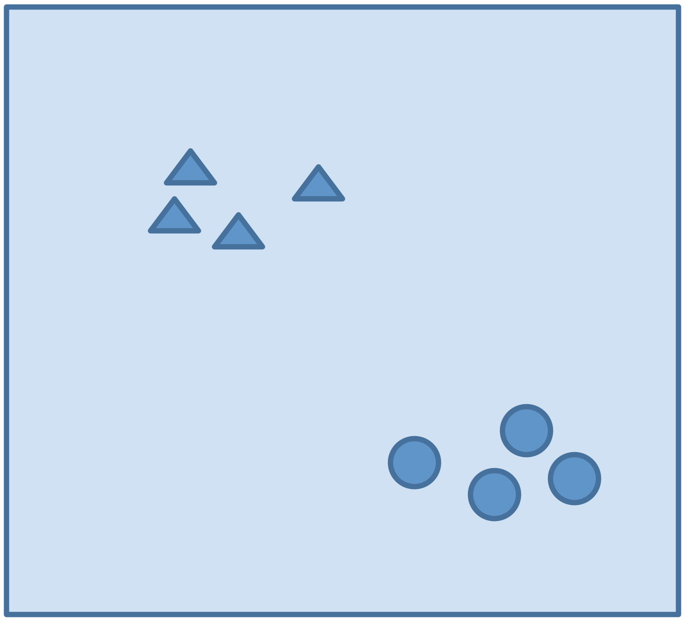
Classic F statistic

$$F = \frac{SS_{Among} / a - 1}{SS_{Within} / N - a}$$

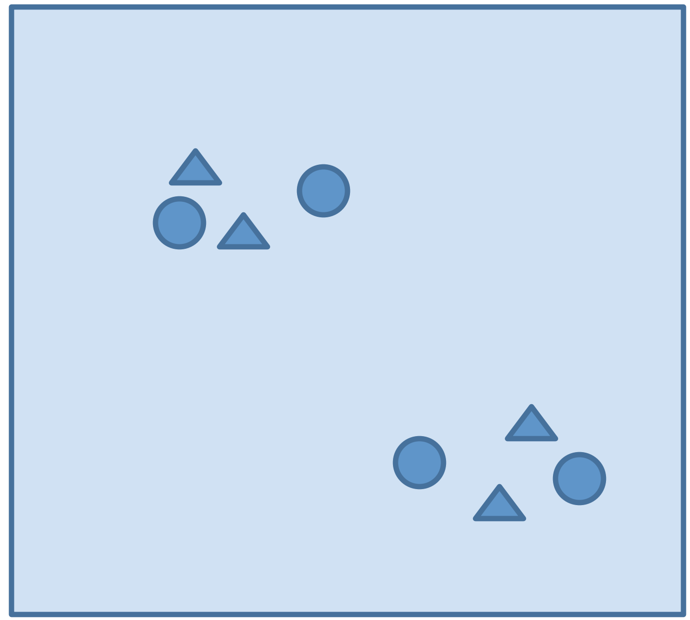
For PERMANOVA, we call this a Pseudo-F

We cannot use a traditional F table to compute the p-value, but must create our own distribution by randomly shuffling the group labels onto sample units

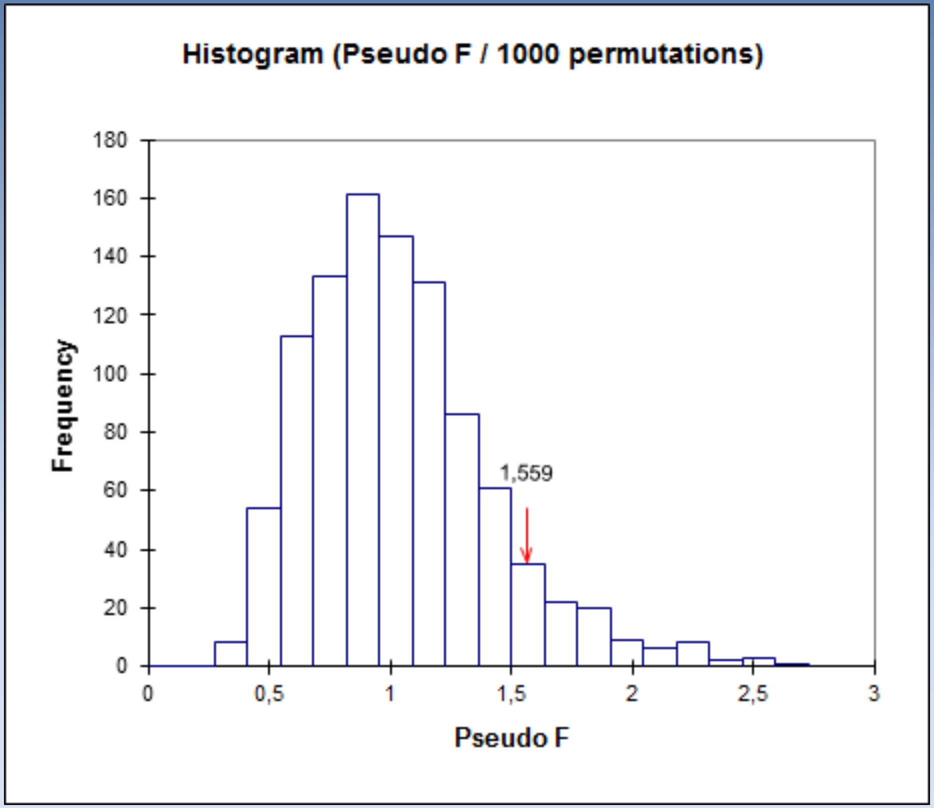
Test by permutation



Pseudo-F
(observed)



Pseudo-F
(after permutation)



$$P = \frac{(No.\,of\ \ F' \geq F) + 1}{(Total\ no.\,of\,F') + 1}$$

Zooplankton Example

1) Was the community structure of Zooplankton ***different between warm (2003-2005) and cold years (2006-2009)*** in the Bering Sea.

2) If so, what ***environmental variables*** contributed to the difference?

From: Eisner et al. *"Climate-mediated changes in zooplankton community structure for the eastern Bering Sea"(in prep)*

Spatial/Temporal Effects

2003

	Species 1	Species 2	Species 3	Species 4
Site 1	25			
Site 2	15			
Site 3	7			
Site 4	8			

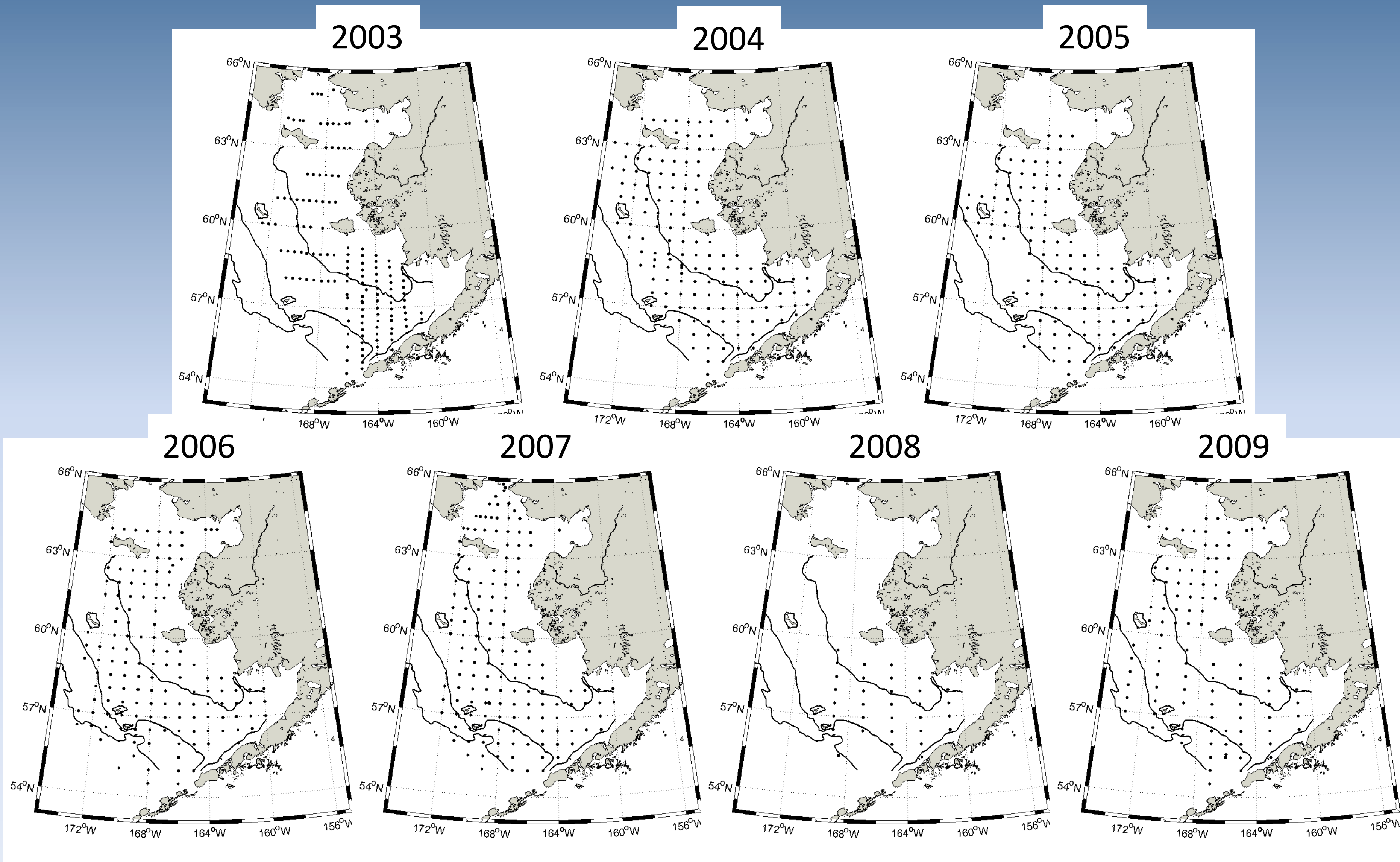
2004

	Species 1	Species 2	Species 3	Species 4
Site 1	25			
Site 2	15			
Site 3	7			
Site 4	8			

2005…2009

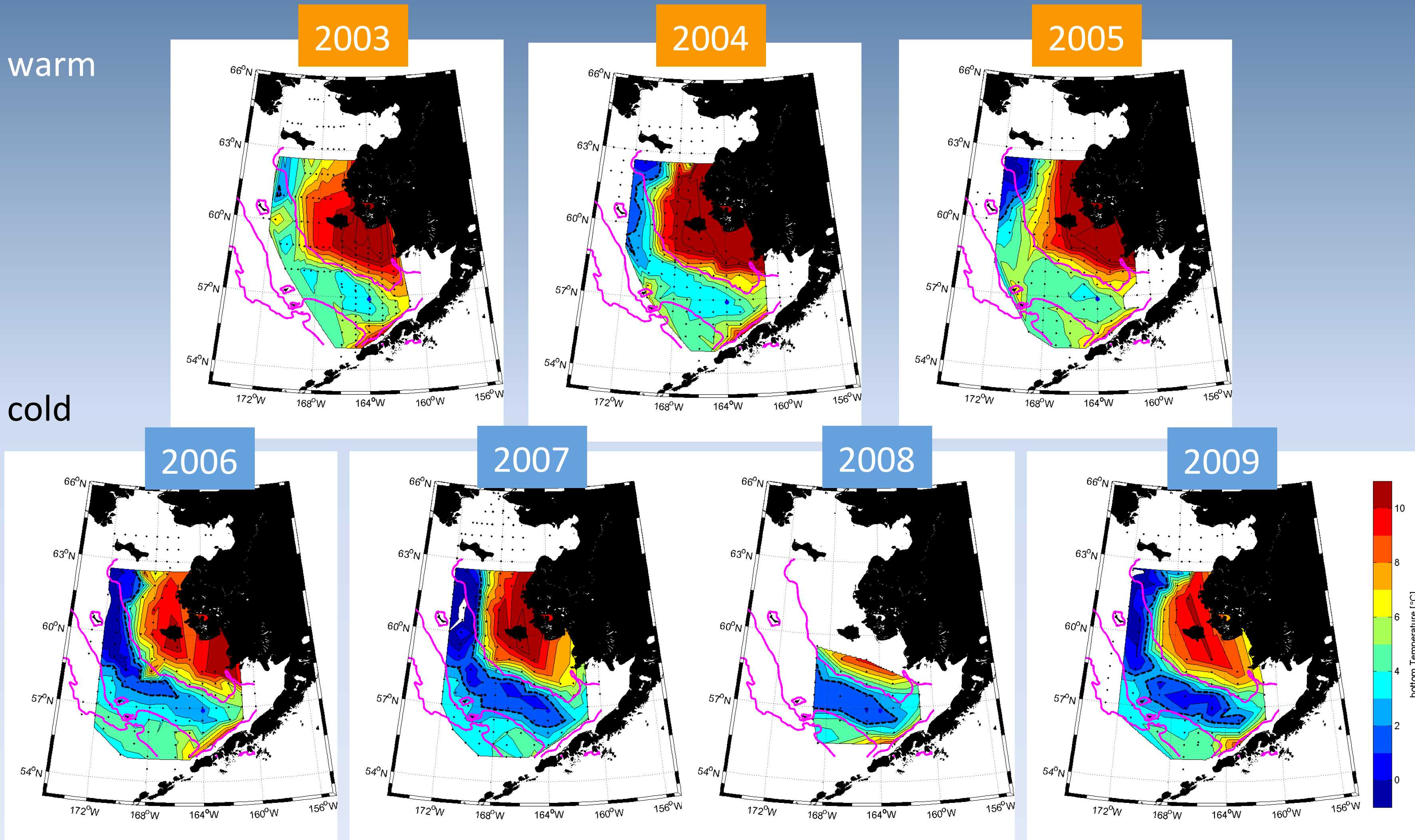
	Species 1	Species 2	Species 3	Species 4
Site 1	25	28	1	20
Site 2	15	18	0	0
Site 3	7	10	0	0
Site 4	8	0	100	7

BASIS Station Locations 2003-2009

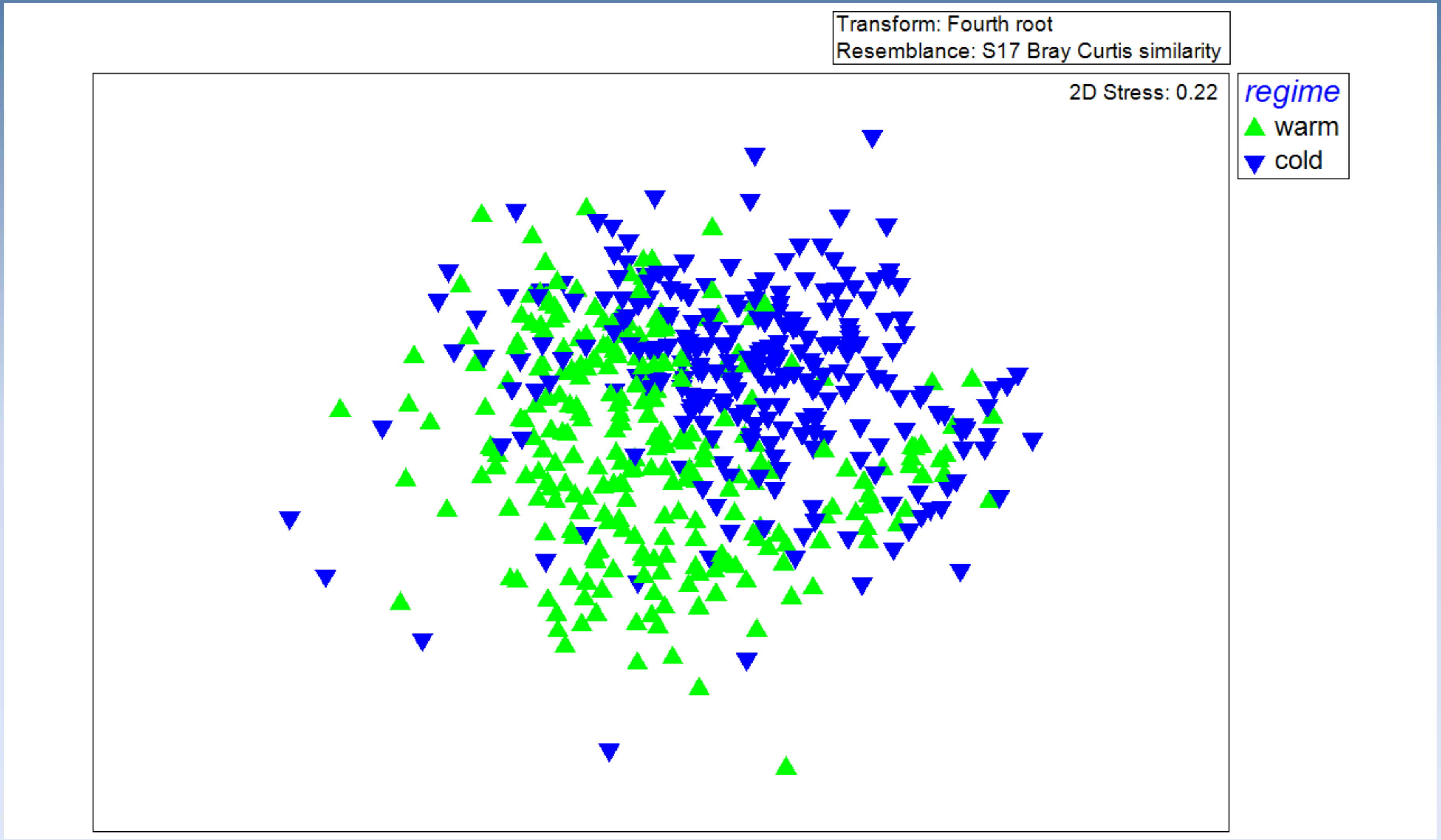


Bottom temperature (bottom 10m)

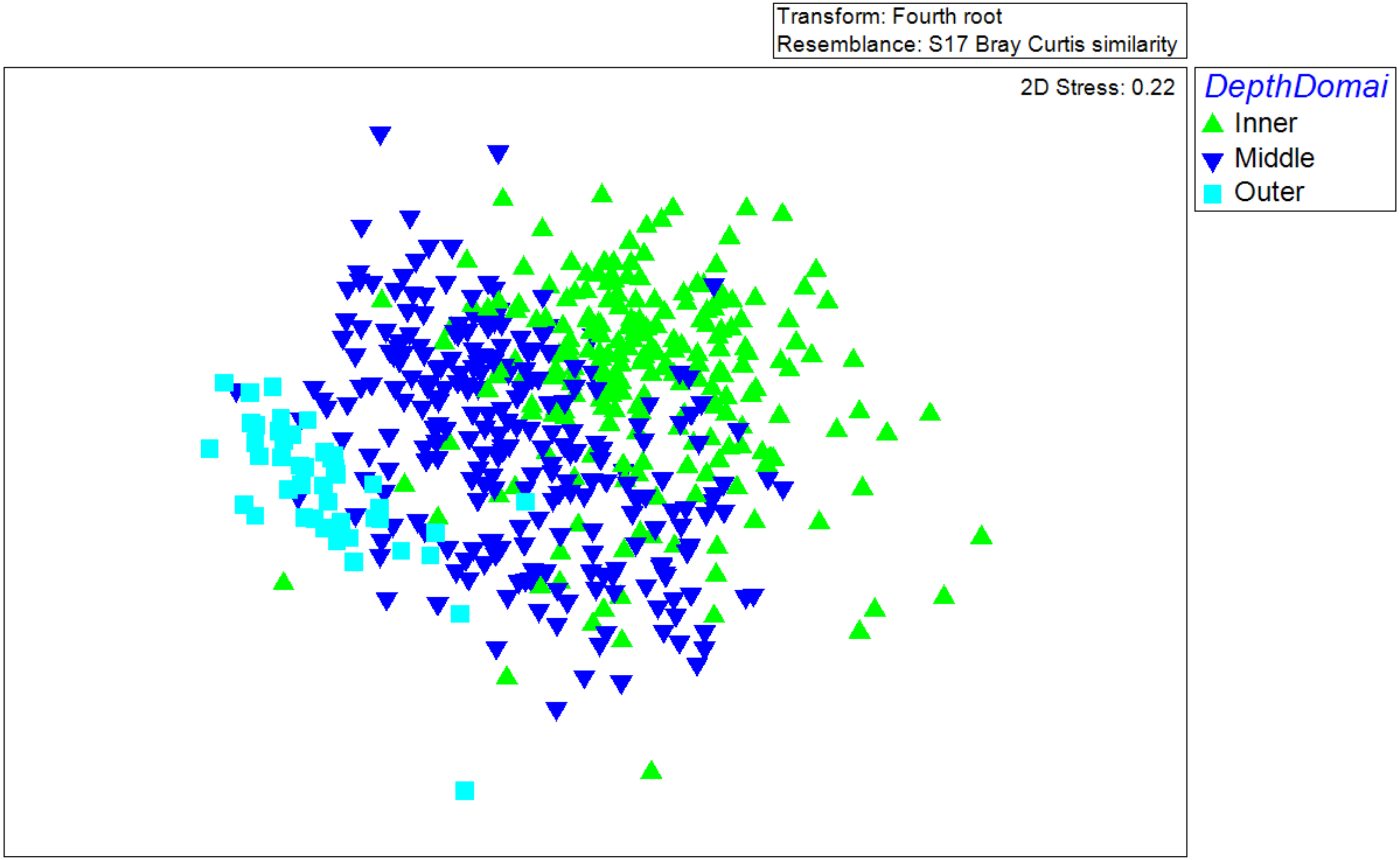
Cold pool (< 2°C) indicated by dashed black line.



NMDS plot of all stations and years



NMDS of all stations and years

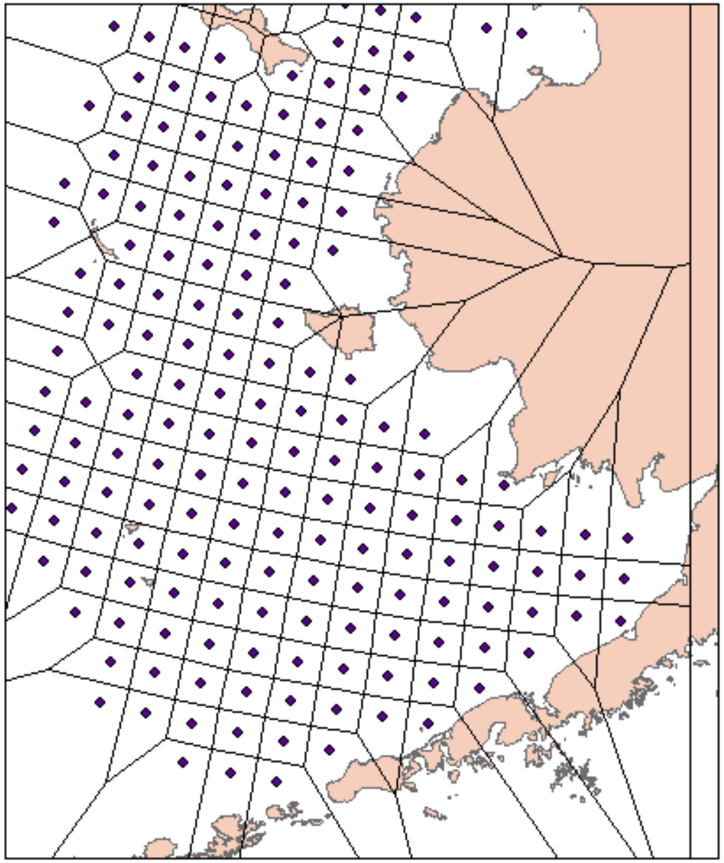


How do we account for uneven sampling across years?

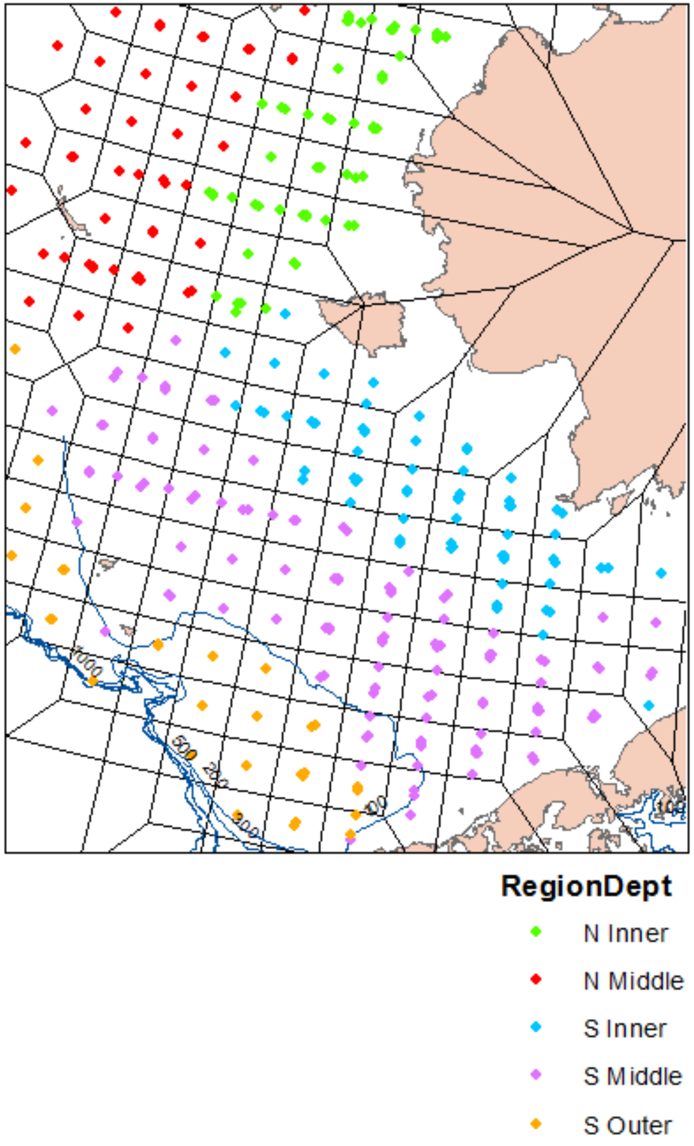
And how do we tease out the spatial variability from the temporal variability?

Creating spatial "Blocks"

BASIS Target Stations



BASIS 2003-2009 Sampled Stations



Does community structure differ between warm and cold years?

S Inner

PERMANOVA table of results

Source	df	SS	MS	Pseudo-F	P(perm)
Regime	1	8515.8	8515.8	2.8989	0.002
Block	21	23486	1118.4	1.7107	0.001
Year(Regime)	5	12584	2516.8	3.85	0.001
RegimexBlock**	20	20055	1002.7	1.5339	0.002
Res	78	50991	653.73		
Total	125	1.22E+05			

Significant Regime effect over and above the variation of this effect among blocks

** Term has one or more empty cells

What environmental variables help explain the difference between warm and cold years?

16 environmental variables

From Niskin bottle samples:

- Nitrate and ammonium (at surface and below pycnocline)
- Chlorophyll a from surface
- Percentage of large phytoplankton (>10 micrometers/total chl a)

From CTD:

- Temperature and salinity averaged over mixed layer depth
- Temperature and salinity averaged over water below the pycnocline
- Chlorophyll a determined by fluorescence (fluorometer on CTD),
- Stability over the top 70m,
- Day of sea ice retreat
- % time in upwelling

Lattitude and Longitude

Why not use BIO-ENV?

What is BIOENV?

BIOENV (or ***BEST***) is an iterative procedure in PRIMER that links environmental variables to community structure by seeking the ***best*** subset of environmental variables that explains the community structure.

PERMANOVA vs. BIOENV

We did not use BIOENV in this case because we wanted to ***remove the spatial effects***, so that we could tease out what was driving the temporal effects only (specifically, ***warm vs. cold years***).

Does community structure differ between warm and cold years?

S Inner						
PERMANOVA table of results						
Source	df	SS	MS	Pseudo-F	P(perm)	
Regime	1	8515.8	8515.8	2.8989	0.002	Significant Regime effect over and above the variation of this effect among blocks
Block	21	23486	1118.4	1.7107	0.001	
Year(Regime)	5	12584	2516.8	3.85	0.001	
RegimexBlock**	20	20055	1002.7	1.5339	0.002	
Res	78	50991	653.73			
Total	125	1.22E+05				

** Term has one or more empty cells

Best-fit PERMANOVA with Environmental Variables

S Inner							
Source	df	SS	MS	Pseudo-F	P(perm)		
Block	21	25510	1215	1.7541	0.001		
IntChlaTop50m	1	4857	4857	3.706	0.012		
% Time upwelling	1	10027	10027	4.2999	0.014		
T above MLD	1	2569	2569	2.8856	0.030		
Regime	1	4449	4449	2.0256	0.079		
Year(Regime)	4	7842	1960	2.9718	0.001		
RegimexGRID	20	19086	954	1.4661	0.006		
Year(Regime)xBlock	68	43242	636	1.2798	0.171		
Residual	8	3975	497				
Total	12 5	121560					

Summary of Example

•NMDS plots suggested a possible difference in community structure ***across regions and between warm and cold years***

•PERMANOVA showed that there was a ***significant difference*** in the community structure of zooplankton between warm and cold years ***over and above*** the variation of this effect among blocks.

•Environmental factors that help to explain this significant Regime effect after removing spatial variability, include ***IntChlatop50m, %Time in Upwelling, and T from MLD***.

Data applications of PERMANOVA

- Abundance
- Biomass
- Diet
- Lengths
- Otolith measurements
- Impact studies

Recent studies using PERMANOVA or ANOSIM

Busby et al. “Spatial and temporal patterns in summer ichthyoplankton assemblages on the eastern Bering Sea shelf 1996-2007” ***(compares multi-species abundance and lengths across years)***

Wilson et al. “Food-related benefits of rearing in an outlying region: walleye pollock off Kodiak Island, Gulf of Alaska” ***(analyzes pollock diet and distribution of zooplankton)***

Jump et al. “Feeding Ecology and Niche Separation of Juvenile Northern Rock Sole and Yellowfin Sole in the Eastern Bering Sea” ***(compares diets where both species co-occur and occur alone)***

Recap

•Community structure analysis is a very useful approach for multi-species data because we can more precisely define dissimilarity/similarity with Bray-Curtis coefficient

•Pretreatment of data is essential, since this drives the analysis (elimination of very rare species, transformation of samples, standardisation of species).

•Descriptive methods (classification, ordination) allow for graphic representations which may help to see whats going on.

•Hypothesis testing such as ANOSIM or PERMANOVA, with PERMANOVA being more versatile, can be applied, even if data is nonnormal.

•PERMANOVA, because it can partition the variance as in an ANOVA, is useful for examining significant enviromental effects.

Thanks to…

- PRIMER developers, including Bob Clarke, Ray Gorley, and Marti Anderson for their correspondence

- Lisa Eisner and EMA group for the use of their data