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**

<table>
<thead>
<tr>
<th></th>
<th>Species 1</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Site 1</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 2</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 3</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 4</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Species 1</th>
<th>Species 2</th>
<th>Species 3</th>
<th>Species 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site 1</td>
<td>25</td>
<td>28</td>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td>Site 2</td>
<td>15</td>
<td>18</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Site 3</td>
<td>7</td>
<td>10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Site 4</td>
<td>8</td>
<td>0</td>
<td>100</td>
<td>7</td>
</tr>
</tbody>
</table>

univariate

multivariate
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
Field data

Standardisation

Transformation

Dissimilarity measure

Classification

Ordination

Hypothesis Testing (ANOSIM or PERMANOVA)
Spatial/Temporal Effects

<table>
<thead>
<tr>
<th></th>
<th>Species 1</th>
<th>Species 2</th>
<th>Species 3</th>
<th>Species 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site 1</td>
<td>25</td>
<td>28</td>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td>Site 2</td>
<td>15</td>
<td>18</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Site 3</td>
<td>7</td>
<td>10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Site 4</td>
<td>8</td>
<td>0</td>
<td>100</td>
<td>7</td>
</tr>
</tbody>
</table>

Data by site

<table>
<thead>
<tr>
<th></th>
<th>Species 1</th>
<th>Species 2</th>
<th>Species 3</th>
<th>Species 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Year 1</td>
<td>25</td>
<td>28</td>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td>Year 2</td>
<td>15</td>
<td>18</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Year 3</td>
<td>7</td>
<td>10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Year 4</td>
<td>8</td>
<td>0</td>
<td>100</td>
<td>7</td>
</tr>
</tbody>
</table>

Data by year (means/totals)

Test for region effect
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 4\textsuperscript{th} 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

\[ d_{BCD}(i, j) = \frac{\sum_{k=1}^{n} |y_{i,k} - y_{j,k}|}{\sum_{k=1}^{n} (y_{i,k} + y_{j,k})} \]

for sites \( i \) and \( j \), for \( n \) species
Bray-Curtis Dissimilarity

### Raw Data

<table>
<thead>
<tr>
<th>Species</th>
<th>Site 1</th>
<th>Site 2</th>
<th>Site 3</th>
<th>Site 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Species 1</td>
<td>25</td>
<td>28</td>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td>Species 2</td>
<td>15</td>
<td>18</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Species 3</td>
<td>7</td>
<td>10</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

### Bray-Curtis of sites

<table>
<thead>
<tr>
<th>Site</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.087</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.958</td>
<td>0.965</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.403</td>
<td>0.474</td>
<td>0.905</td>
<td>0</td>
</tr>
</tbody>
</table>
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)** | \( a-1 \) | \( \sum_{i=1}^{a} n_i (\bar{y}_i - \bar{y}_\cdot)^2 \) | \( \frac{MSA}{MSE} \)  
Error **(within groups)** | \( N-a \) | \( SSE = SST - SSA \) | \( MSE = \frac{SSE}{(N-a)} \)  
Total | \( N-1 \) | \( SST = \sum_{i=1}^{a} \sum_{j=1}^{n} (y_{ij} - \bar{y}_\cdot)^2 \) |  

Where:
- \( y_{ij} \) is the observation in the \( i \)th group and \( j \)th observation.
- \( \bar{y}_i \) is the mean of the \( i \)th group.
- \( \bar{y}_\cdot \) is the overall mean.
- \( a \) is the number of groups.
- \( n_i \) is the number of observations in the \( i \)th group.
- \( N \) is the total number of observations.

The ANOVA table helps in determining whether the differences between the group means are statistically significant.
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

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

centroid
PERMANOVA partitioning of SS

\[ SS_T = \frac{SS \text{ of all}}{N} \]

\[ SS_W = \frac{SS \text{ of each group}}{n} \]

\[ SS_A = SS_T - SS_W \]
Classic F statistic

\[ F = \frac{SS_{Among}}{SS_{Within}} \times \frac{a - 1}{N - a} \]

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)
\[ P = \frac{(\text{No. of } F' \geq F) + 1}{(\text{Total no. of } F') + 1} \]
Zooplankton Example


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**

<table>
<thead>
<tr>
<th>Site</th>
<th>Species 1</th>
<th>Species 2</th>
<th>Species 3</th>
<th>Species 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site 1</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 2</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 3</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 4</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**2004**

<table>
<thead>
<tr>
<th>Site</th>
<th>Species 1</th>
<th>Species 2</th>
<th>Species 3</th>
<th>Species 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site 1</td>
<td>25</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 2</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 3</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Site 4</td>
<td>8</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**2005...2009**

<table>
<thead>
<tr>
<th>Site</th>
<th>Species 1</th>
<th>Species 2</th>
<th>Species 3</th>
<th>Species 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Site 1</td>
<td>25</td>
<td>28</td>
<td>1</td>
<td>20</td>
</tr>
<tr>
<td>Site 2</td>
<td>15</td>
<td>18</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Site 3</td>
<td>7</td>
<td>10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Site 4</td>
<td>8</td>
<td>0</td>
<td>100</td>
<td>7</td>
</tr>
</tbody>
</table>
BASIS Station Locations 2003–2009
Cold pool (< 2°C) indicated by dashed black line.
NMDS plot of all stations and years
NMDS of all stations and years

Transform: Fourth root
Resemblance: S17 Bray Curtis similarity

DepthDomain
- Inner
- Middle
- Outer

2D Stress: 0.22
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

RegionDept
- N Inner
- N Middle
- S Inner
- S Middle
- S Outer
Does community structure differ between warm and cold years?

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

** 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

Latitude 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?

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

** Term has one or more empty cells

Significant Regime effect over and above the variation of this effect among blocks
## Best-fit PERMANOVA with Environmental Variables

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