Evaluating Statistical Methods for Maximizing Classification: An Application using Otolith Tracers from Spotted Seatrout

Stacy Beharry

Where do you belong?

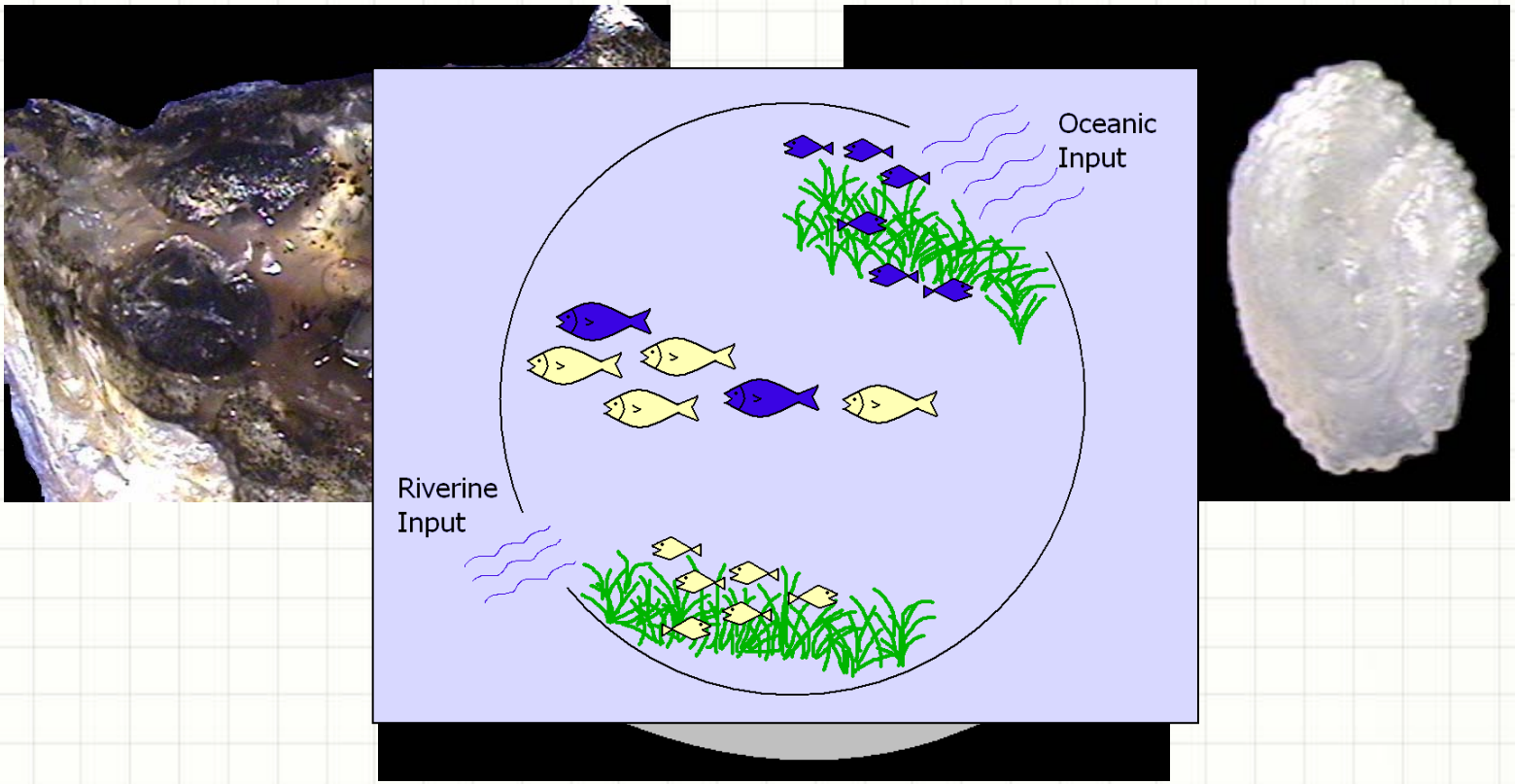
- Important to have methods that accurately separate groups
  - Determining natal sources
  - Identifying stocks
  - Migratory paths
  - Connectivity

- Difficult to separate on observation

Where do you belong?

- Any marker that remains with the fish until capture
- Selecting markers which capture differences in biotic and abiotic controls
  - Trace element chemistry
  - Stable isotope chemistry
  - Growth
- Discriminant function performed to determine group membership

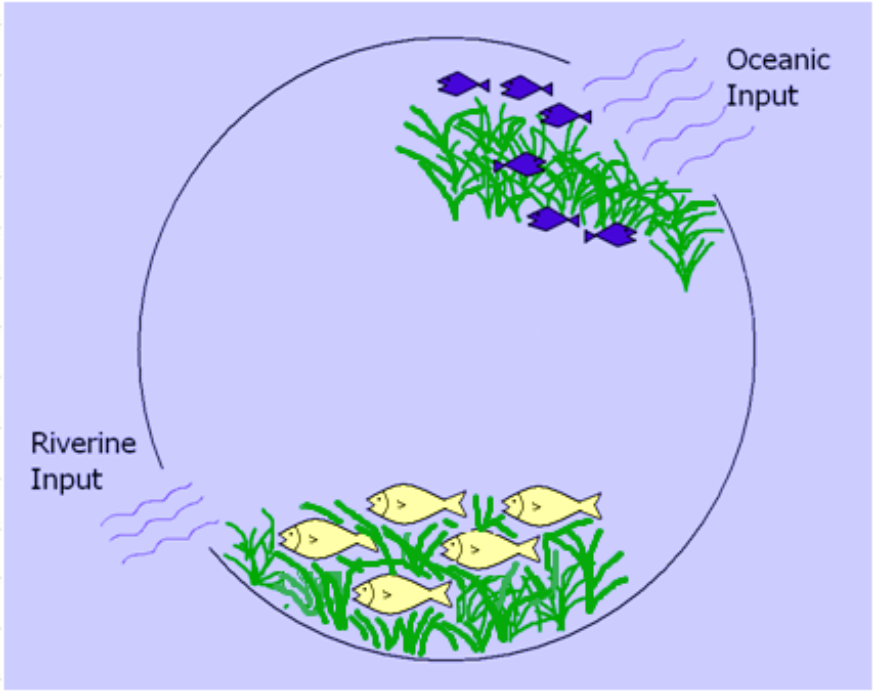
Intro: Natural Tracers



- Otoliths contain chemical signature unique to biogeochemical environment- Trace elements, stable isotope

Intro: Natural Tracers

- Burgeoning use of otolith chemistry to determine movement
- Lots of variables collected
  - ^{7}Li, ^{25}Mg, ^{55}Mn, ^{58}Ni, ^{65}Zn, ^{86}Rb, ^{88}Sr, ^{89}Y, ^{138}Ba, ^{139}La
  - $\delta^{13}C$, $\delta^{18}O$
- Growth

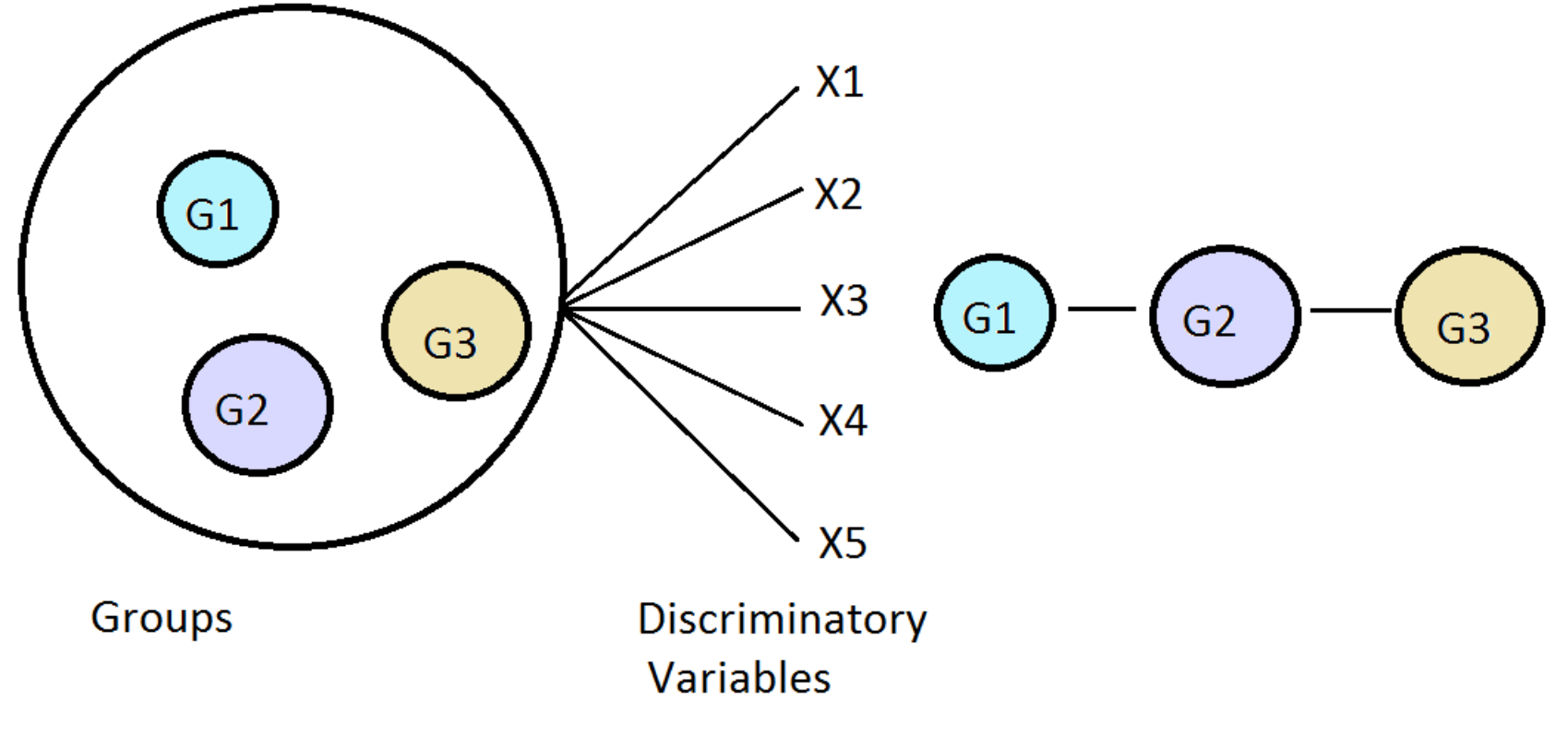


Variable importance and selection

- Linear Discriminant Function Analysis used to delineate groups
  - Parametric technique
  - Equal var–covariance matrices between groups
  - Groups drawn from multivariate normal distributions
- Used in a variety of fields
  - Classification of Venezuelan spirituous beverages by means of discriminant analysis Hernández-Caraballo et al 2003

How is natal source determined

- Using variables collected from the otolith



Adapted from Discriminant Function Analysis by W. Klecka

Variable importance and selection

- Are all variables needed to produce the maximum classification?
- Do all variables convey useful information?
- What about variables that convey similar information

Variable importance and selection

- In building model variables may be used
  - Everything collected
  - Commonly used (faith based approaches)
  - New techniques allow for variables to be collected
  - Means are significantly different among groups (ANOVA)

- How should the model be constructed?

Rao's Test of Additional Info

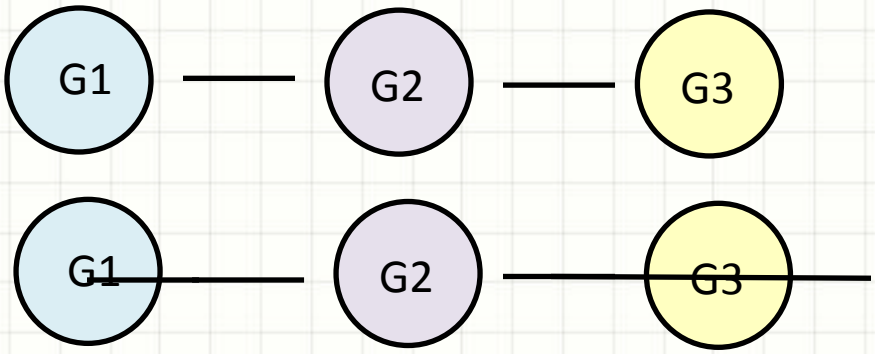
M1= A+B

M2= A+B+**C+D**

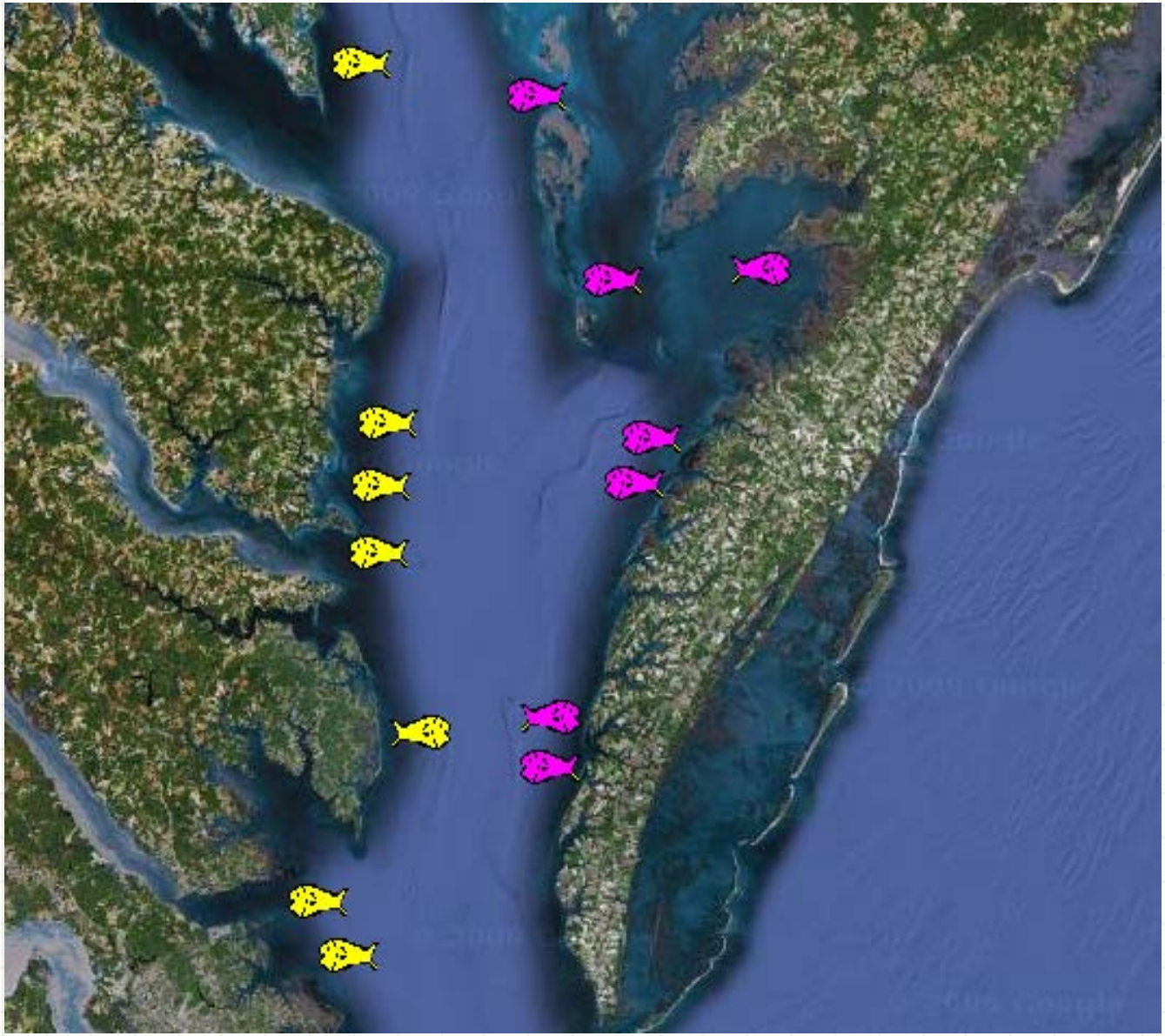
- Determine if **C** and **D** convey any additional information beyond A and B
  - i.e. does the separation distance increase with the addition of each variable

A + B

A + B + C + D

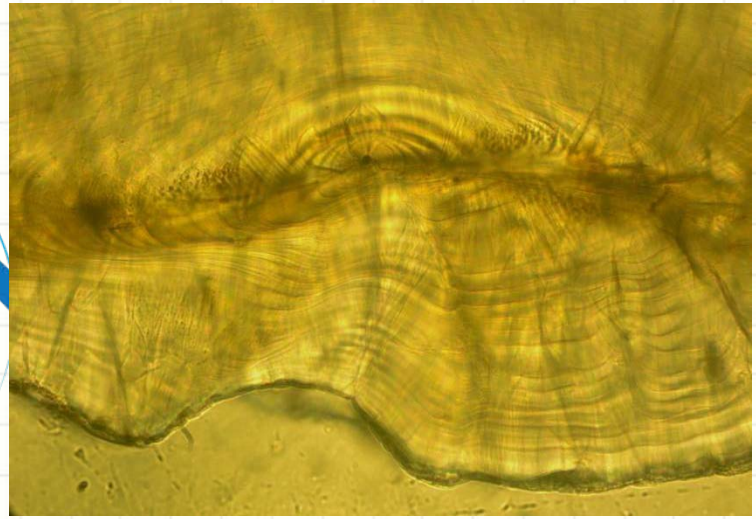


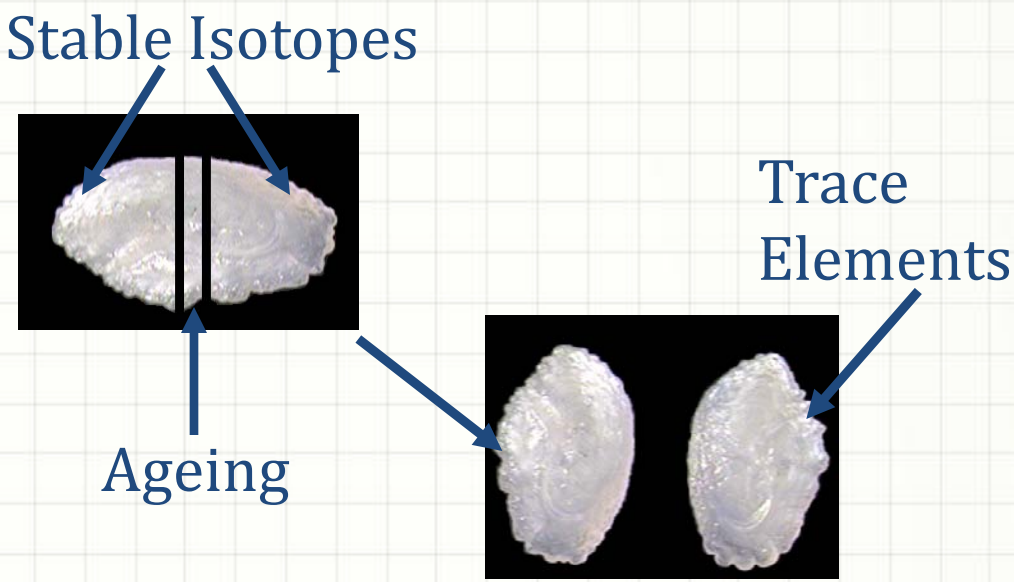
Methods



Methods

- Solution based ICP-MS - ratio to calcium
- Stable isotope
- Daily ages $Y_t = \beta_0 \, e^{\left(-e^{\left(-\beta_1 (t-\beta_2)\right)}\right)}$





Methods

- LDFA to classify juvenile spotted seatrout (*Cynoscion nebulosus*) to nursery areas using each variable
  - ^{7}Li, ^{25}Mg, ^{55}Mn, ^{86}Rb, ^{88}Sr, ^{89}Y, ^{138}Ba (7)
  - $\delta^{13}C$, $\delta^{18}O$ (2)
  - β_0, β_1, β_2 (3)

- Rao's test of additional information to determine conveyed by each additional variable

Methods

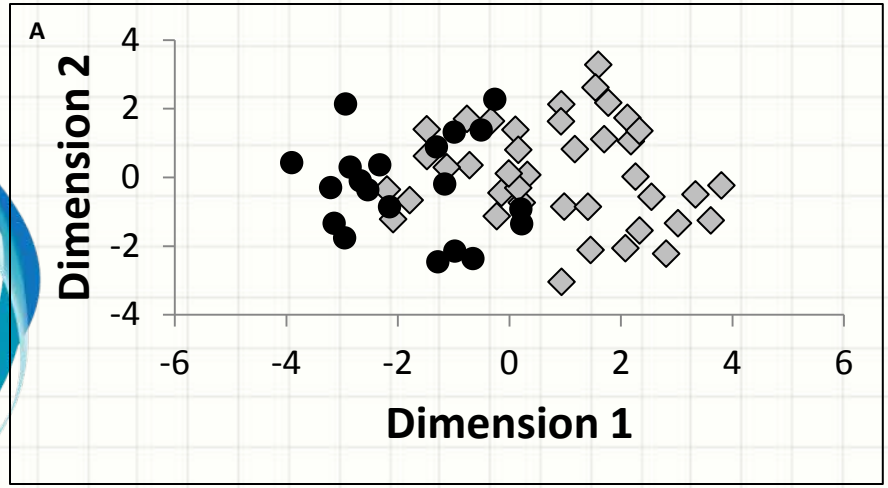
- Relies on the Mahalanobis distance between groups
- Uses a F-distribution
- Easily programmed in SAS or Excel

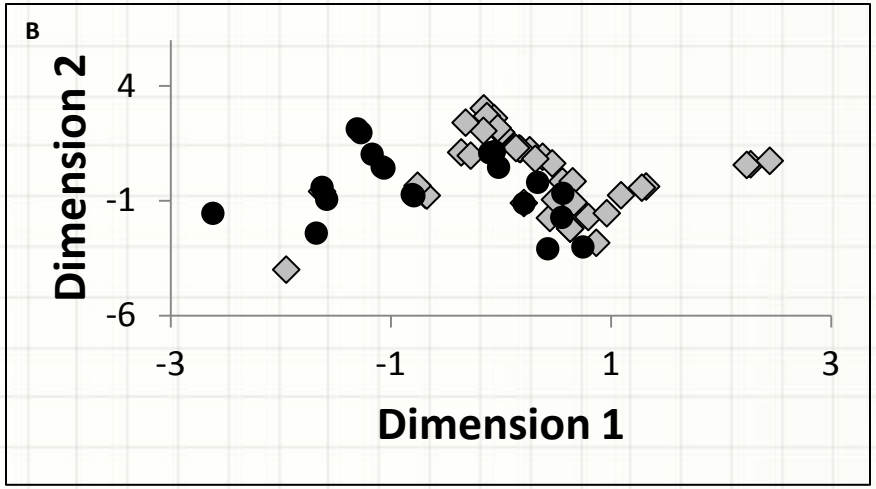
$$U_{q,p} = \frac{1 + \frac{N_1 N_2}{(N_1 + N_2)(N_1 + N_2 - 2)} D_{p+q}^2}{1 + \frac{N_1 N_2}{(N_1 + N_2)(N_1 + N_2 - 2)} D_p^2} - 1$$

Methods

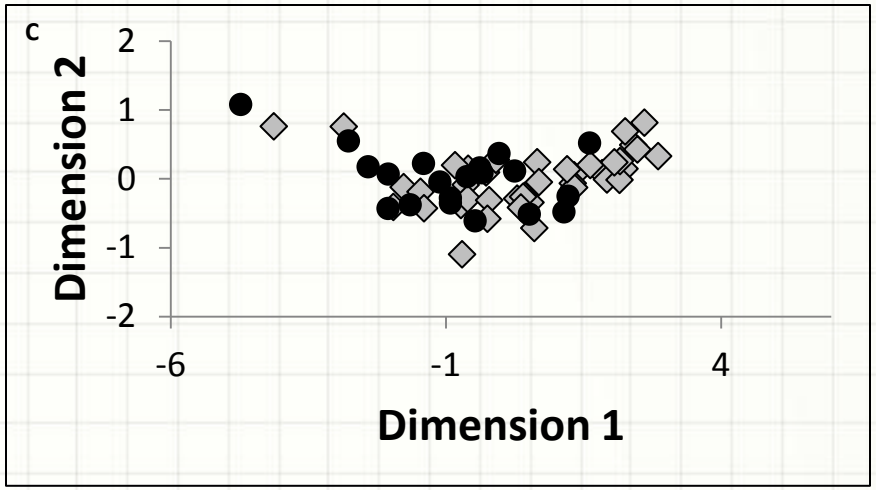
- Compared the overall error of the model selected by
  - Rao's test
  - Variables that differed between groups
  - Commonly used variables
  - All variables collected

Results





	♦ ES	● WS
TE	90	60
SI	92.5	55
G	87.5	25



Results

- All variables did not have the same power to separate groups

		Classification	
Variable	D^2	ES	WS
Ba	3.55	87.5	70.0
Sr	2.39	87.5	70.0
$\delta^{13}C$	2.20	92.5	70.0
Li	0.96	85.0	50.0
Mg	0.94	87.5	35.0
Mn	0.76	85.0	40.0
β_0	0.63	95.0	20.0
β_1	0.62	85.0	30.0
β_2	0.61	85.0	25.0
Y	0.30	90.0	25.0

Results

Ba

Sr

$\delta^{13}C$

Li

Mg

Mn

β_0

β_1

β_2

Y

	Classification %			Rao's	
Model	ES	WS	% Error	F	p
Ba Sr	90	70	20.0	1.67	0.20
Ba C*	87.5	85	13.8	8.97	0.00
Ba C Li	87.5	85	13.8	0.01	0.92

Results

	Variable	Classification %		Overall
		ES	WS	Error %
Rao's	Ba C	87.5	85	13.8
Common	Mg Mn Sr Ba C O	85	80	17.5
ANOVA	Li Mg Mn Sr Y Ba C β_0 β_1 β_2	87.5	75	18.8
All collected	Li Mg Mn Rb Sr Y Ba C O β_0 β_1 β_2	90	50	30.0

Discussion

- Variables convey different amounts of information including all can decrease classification
- All variables do not convey useful information
- If variables are collinear one may be sufficient

Ecological impacts

- Ecologically important areas can be identified
- Long term productivity can be determined
  - Production of adult fish form each natal source can be determined
  - Estimation of cohort mortality can be established
  - Directly asses survivorship
- Populations can be effectively managed
  - Prioritize management decisions

Acknowledgements

- Dr Cynthia Jones
- Dr Dayanand Naik
- Dr Jason Schaffler
- Center for Quantitative Fisheries Ecology
- National Science Foundation

Department of Ocean, Earth & Atmospheric Sciences
Old Dominion University

CQFE