National Biological Assessment and Criteria Workshop

Advancing State and Tribal Programs

Coeur d'Alene, Idaho

31 March – 4 April, 2003

Index 101

Use of RIVPACS-type Predictive Models in Aquatic Biological Assessment: Theory and Application

Chuck Hawkins, Utah State University;
Rick Hafele, Oregon Dept. of Environmental Quality

The Concept:
O versus E as a Measure of Biological Integrity

$$\frac{O}{E} = \frac{\text{the } \textbf{\textit{set}} \text{ of native taxa expected at a site that are actually observed.}}{\text{the } \textbf{\textit{set}} \text{ of native taxa expected to occur at a site in the absence of human-caused stress.}}$$

The deviation of O from E is a measure of compositional similarity and thus a community-level measure of biological integrity.

O/E has some useful properties as an index of biological condition.

- It has an intuitive biological meaning (taxa are the ecological capital on which all ecosystem processes depend) and is interpretable by researchers, managers, the public, and policy makers.

O/E has some useful properties as an index of biological condition.

- It means the same thing everywhere, which allows direct and meaningful comparisons across regions and states.

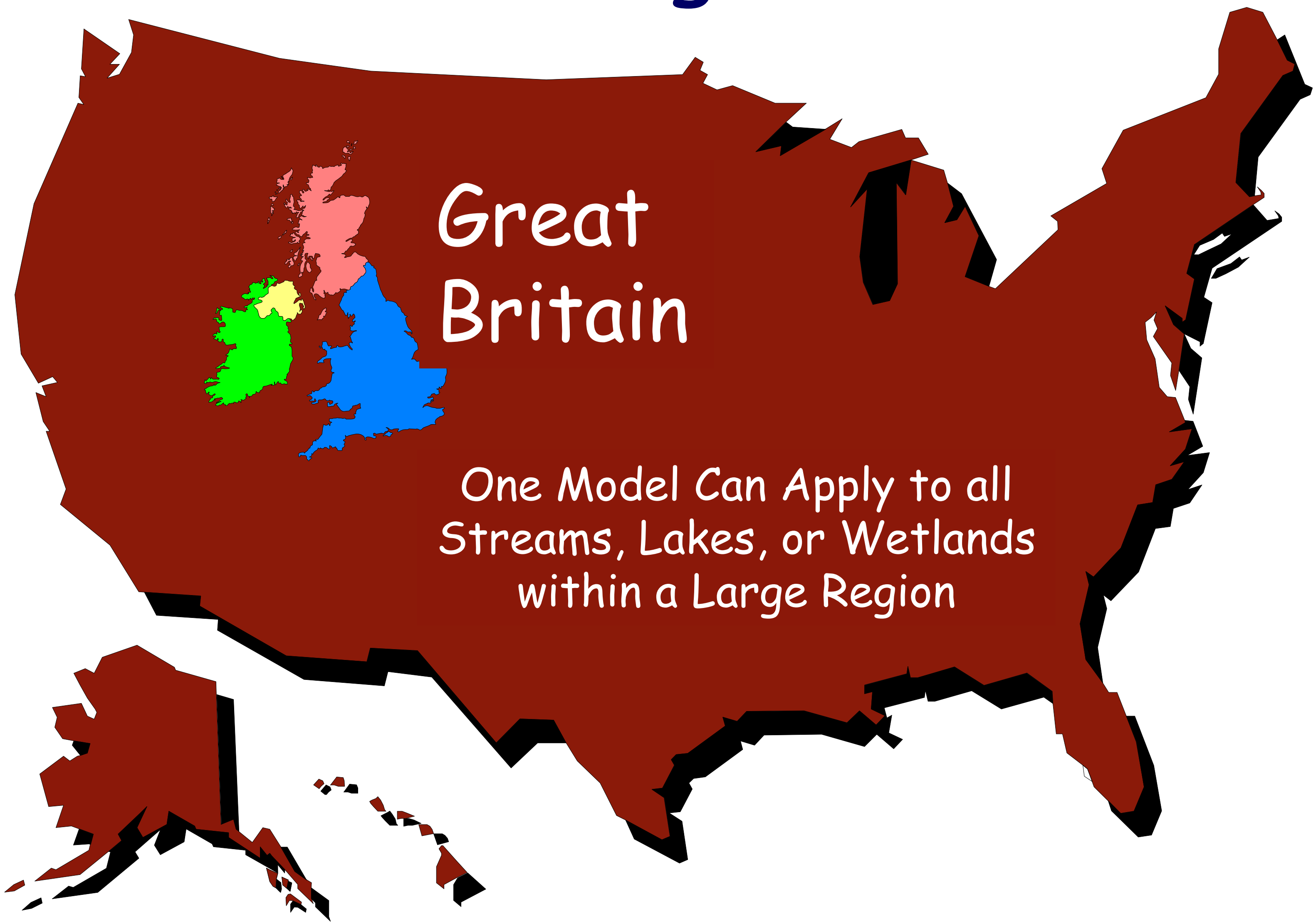
O/E has some useful properties as an index of biological condition.

- Its derivation and interpretation are independent of type and knowledge of stressors in the region.

O/E has some useful properties as an index of biological condition.

- It is quantitative.

O/E has some useful properties as an index of biological condition.



Major Issues for the 101 Course

- Understanding the units of measure.
- Predicting the expected taxa.
- Calculating O/E, the biological condition value.
- Determining if an assessed site is impaired.

Basic Concepts

- Predictive models base assessments on the compositional similarity between observed and expected biota.

The Unit of Measure

- The deviation of O from E is difficult to express in a simple way given the multivariate nature of both terms.
- We need a simple currency that also retains the information content of compositional similarity.
- We also need a way of dealing with the fact that we ***sample*** the biota and thus deal with probabilities not absolutes.

O/E: A Simplified Expression of a Multivariate World

- Define E as the ***number*** of native taxa expected to occur at a site in the absence of human-caused stress.
- Define O as the ***number*** of taxa that are predicted to occur that are actually present.
- The ratio O/E is the ***proportion*** of taxa observed that should have been collected.
- O/E is not based on raw taxa richness; O is constrained to include only those taxa with a probability of capture greater than a stated threshold.

Basic Concepts
(Units of Measure & the Expected Taxa)

	Replicate Sample Number										Freq
Species	1	2	3	4	5	6	7	8	9	10	(P_c)
A	*	*	*	*	*	*	*	*	*	*	1.0
B	*	*		*	*	*		*	*	*	0.8
C	*		*		*	*			*		0.5
D		*	*				*		*	*	0.5
E					*						0.1
Sp Count	3	3	3	2	4	3	2	2	4	3	2.9

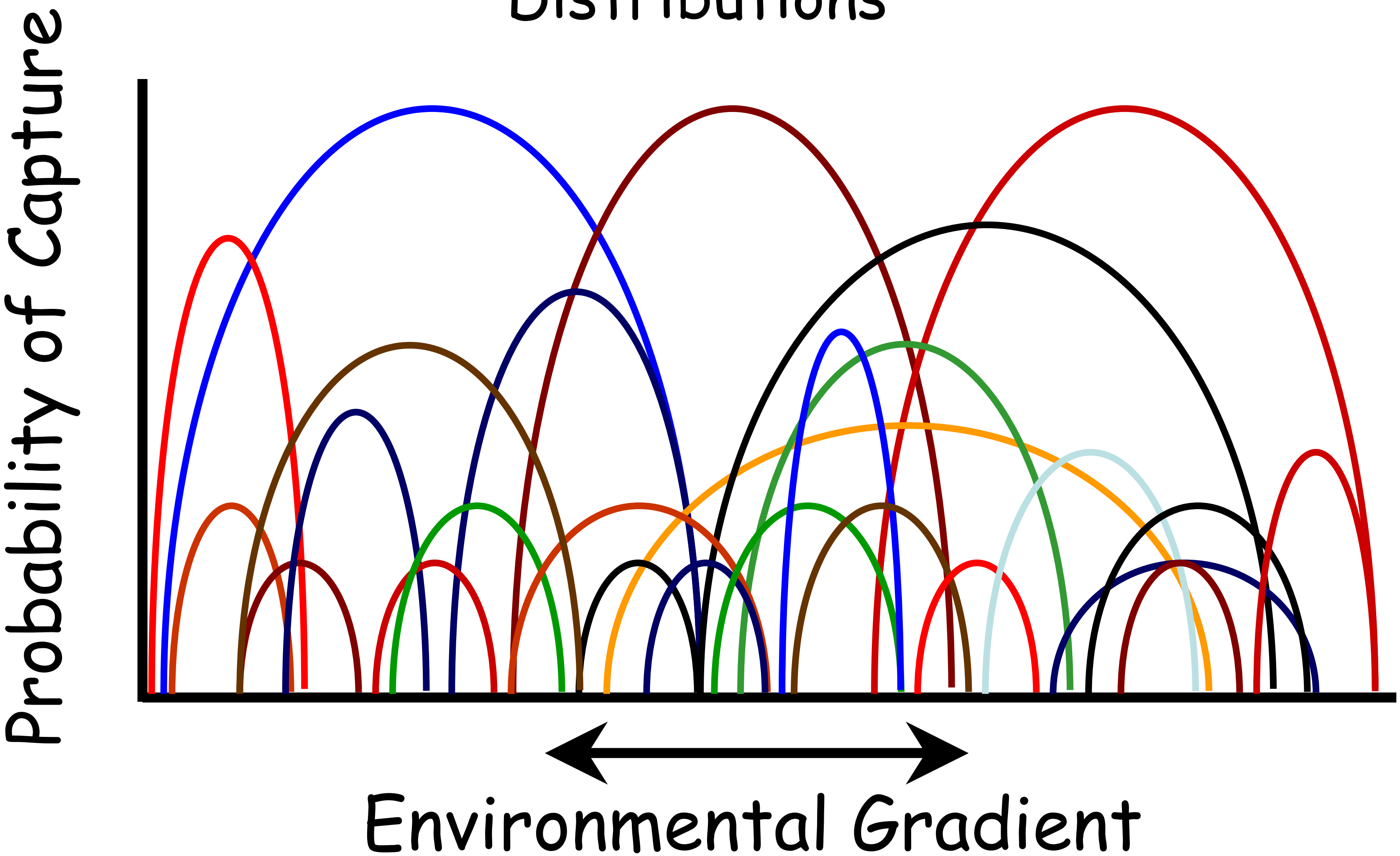
Species Richness is the Currency.

$E = \sum P_c = 0$ number of species / sample = 2.9.

O/E as a Measure of Impairment

Expected Biota		Observed Biota			
Species	Pc	O_1	O_2	O_3	O_4
A	1.0	*	*	*	*
B	0.8	*		*	
C	0.5		*		
D	0.5	*			
E	0.1				
F	0				*
Expected Sp Count	2.9	3	2	2	1
	O/E	1.03	0.69	0.69	0.34

This is the Challenge:
Estimating the Probabilities of Capture of Many Different Taxa that Exhibit Individualistic Distributions



The basic approach to modeling pc's and estimating E was worked out by Moss et al.*

River InVertebrate Prediction and Classification System (RIVPACS)

*Moss, D., M. T. Furse, J. F. Wright, and P. D. Armitage. 1987. The prediction of the macro-invertebrate fauna of unpolluted running-water sites in Great Britain using environmental data. Freshwater Biology 17:41-52.

RIVPACS-type Models: 8 Basic Steps

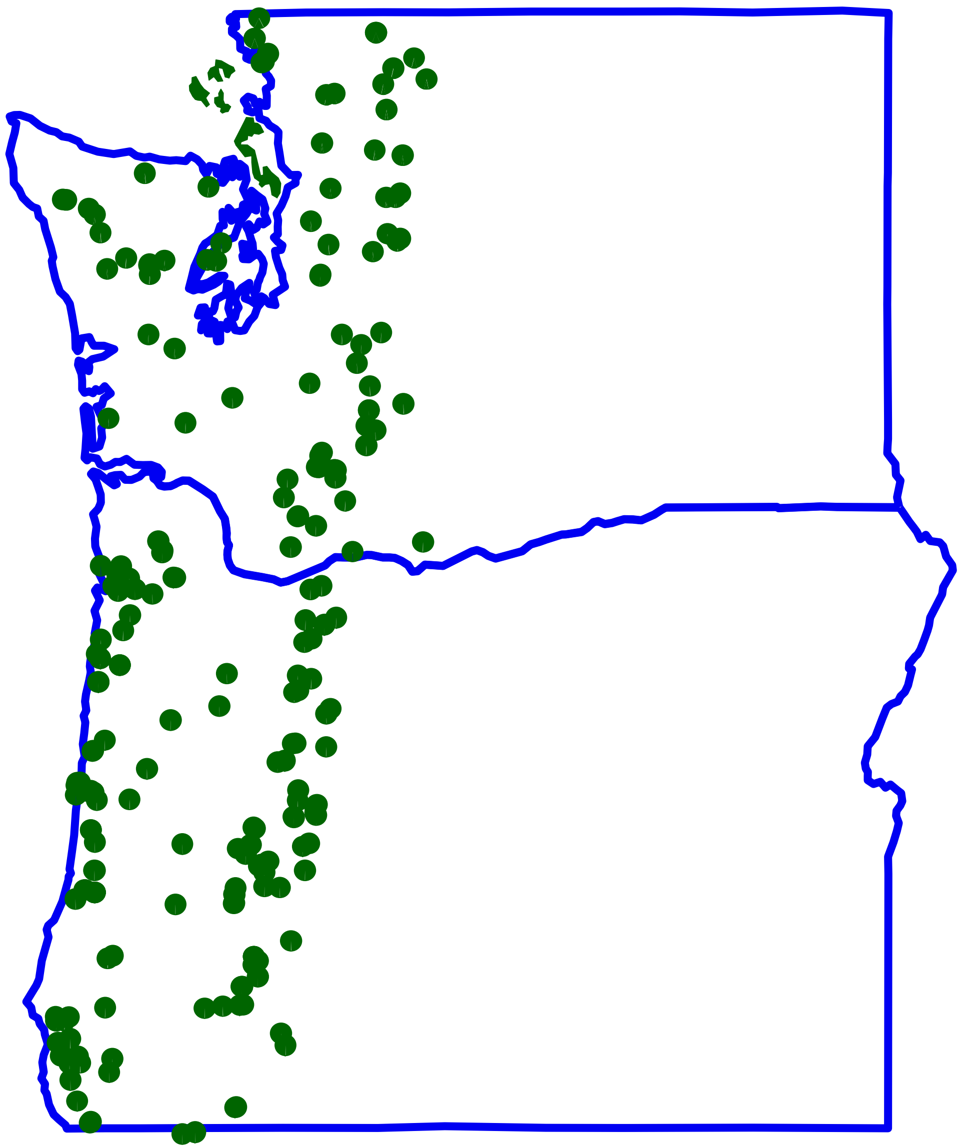
1. Establish a network of reference sites.
2. Establish standard sampling protocols.
3. Classify sites based on their biological similarity.
4. Estimate individual probabilities of capture by relating environmental setting to the biological classification (multivariate statistics).

For each assessed site:

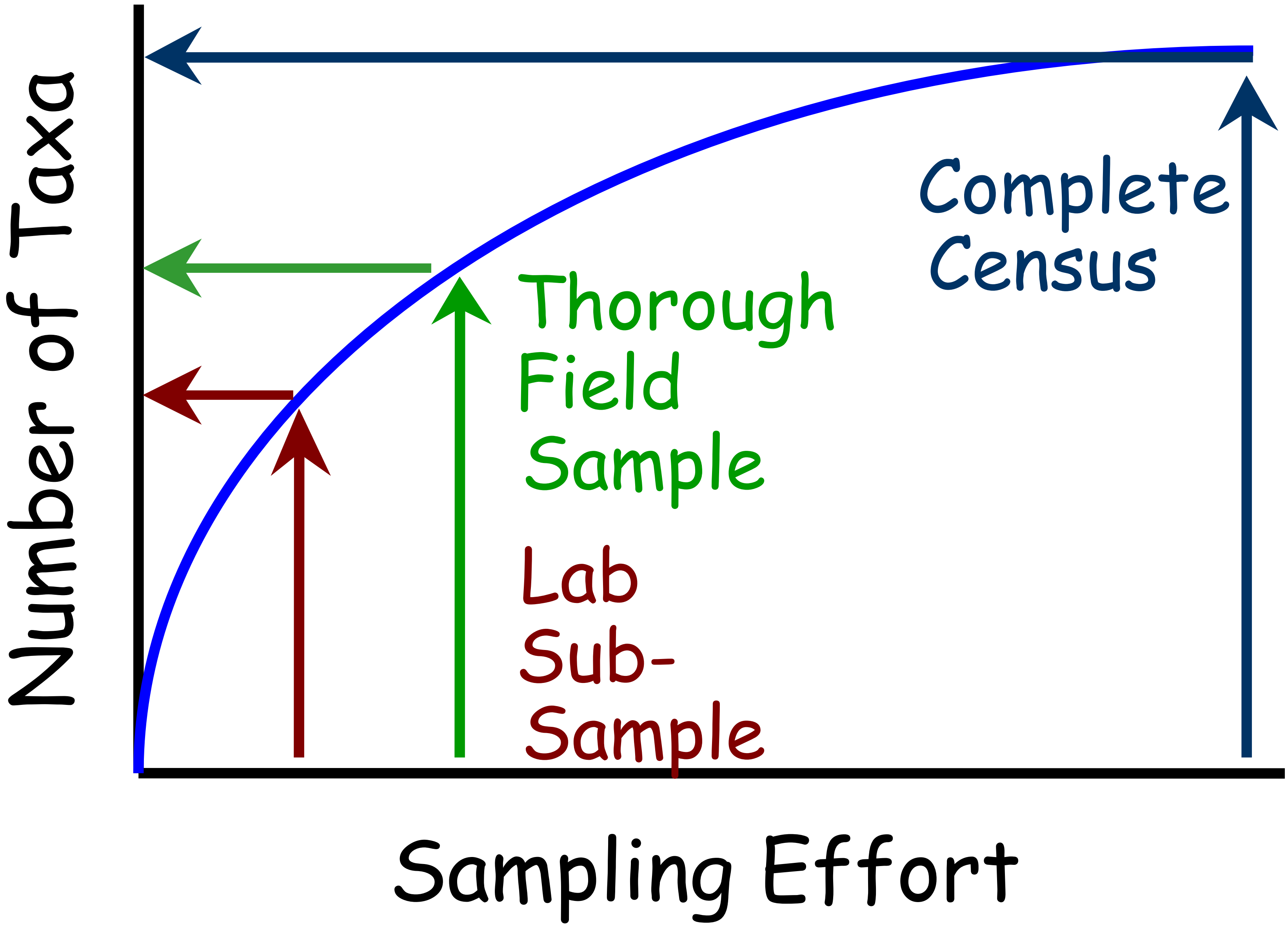
5. Sum p_c's to estimate E.
6. Count O
7. Calculate O/E.
8. Determine if observed O/E is different from reference?

Creating RIVPACS Models

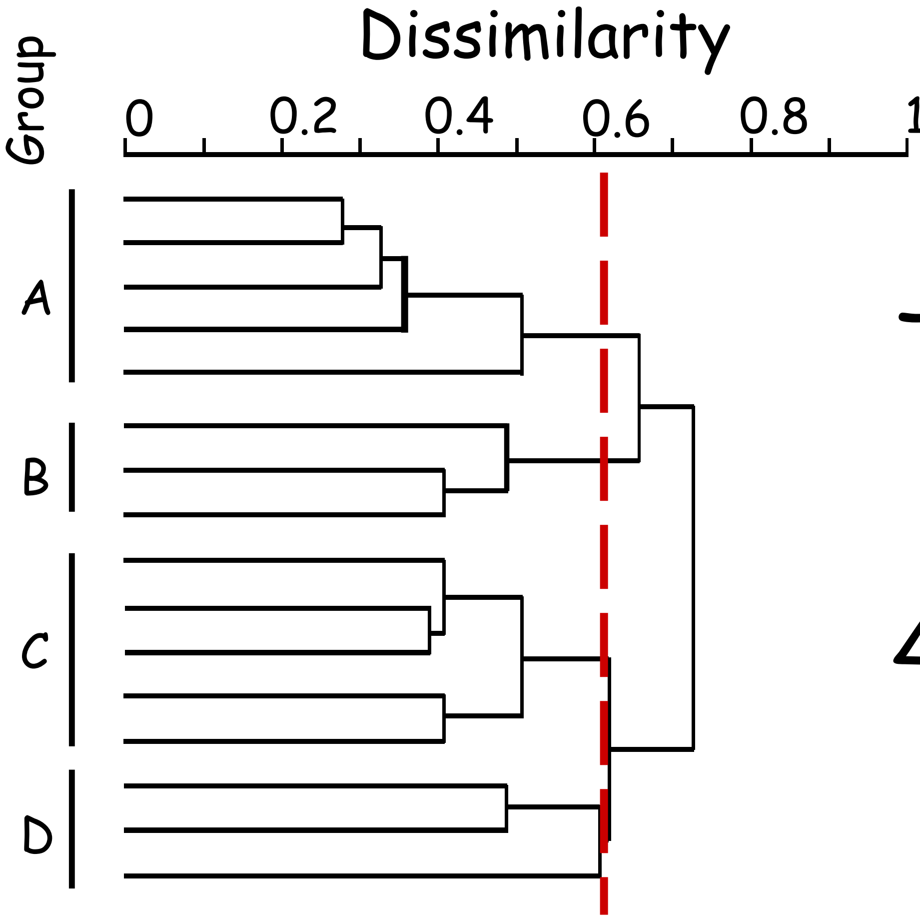
1. Establish a network of reference sites that span the range of environmental conditions in the region of interest.



2. Use standard protocols to sample biota and habitat features.


Number of Taxa
Complete
Census
Thorough
Field
Sample
Lab
Sub-
Sample
Sampling Effort

3. Classify sites in terms of their compositional similarity.

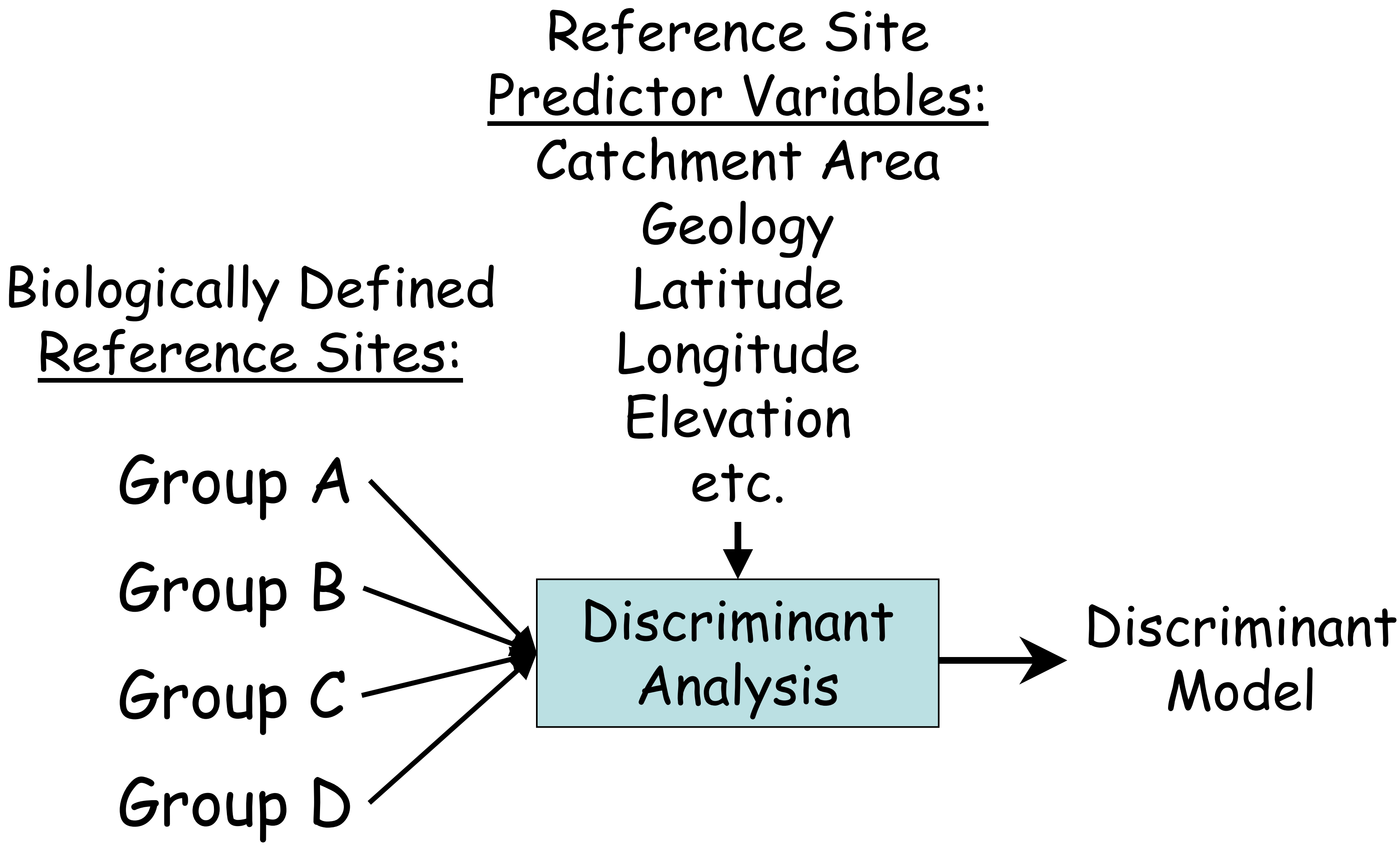


This cluster analysis shows 4 'groups' of sites

4. Derive a multivariate model to predict from environmental features the probabilities of sites belonging to biologically-defined groups and the probabilities of capturing each taxon.

P_c = f(elevation, watershed area, geology)

The Discriminant Model



Combining the Discriminant Model + Frequencies of Occurrence Provides Estimates of Probabilities of Capture

Test Site

Predictor Variables

↓

Discriminant Model

Group Probabilities		Taxon 1 Reference Site Frequencies	
50% Group A	*	60% Group A	= 0.30
40% Group B	*	20% Group B	= 0.08
10% Group C	*	0% Group C	= 0.00
0% Group D	*	0% Group D	= 0.00
			P_c = 0.38

5. Sum p_c's to estimate the number of taxa (E) that should be observed at the site based on standard sampling.

Species	P_c
1	0.70
2	0.92
3	0.86
4	0.63
5	0.51
6	0.32
7	0.07
8	0.00
E	4.01

6. Determine O, the number of predicted taxa that were collected (O).

7. Calculate O/E.

Species	P_c	O
1	0.70	*
2	0.92	*
3	0.86	
4	0.63	
5	0.51	*
6	0.32	
7	0.07	
8	0.00	
E	4.01	3

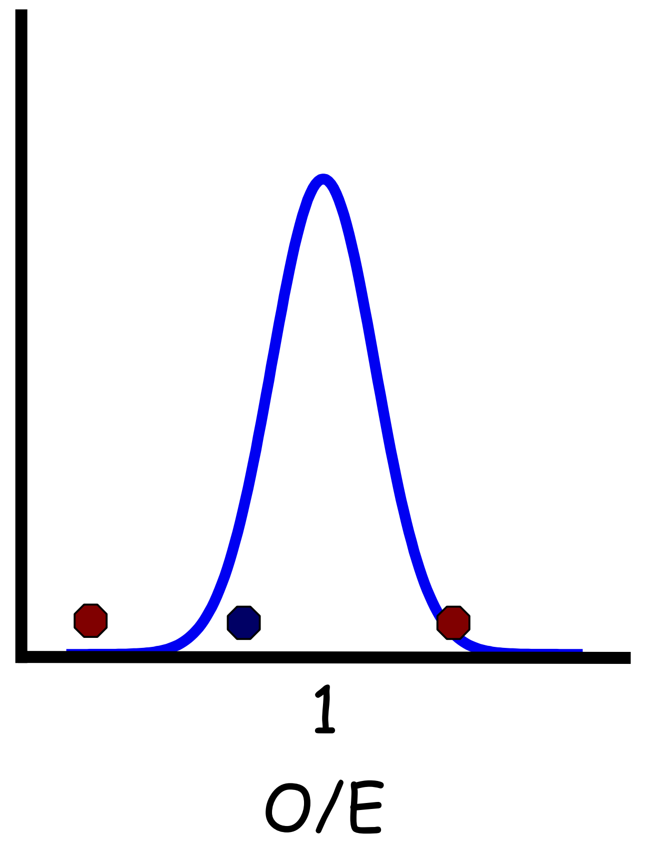
O/E = 3 / 4.01 = 0.75

8. Determine if the O/E value is significantly different from the reference condition by comparing against model predictions and error.

Statistical Issues Regarding Inferences of Impairment

Single Sites/Samples

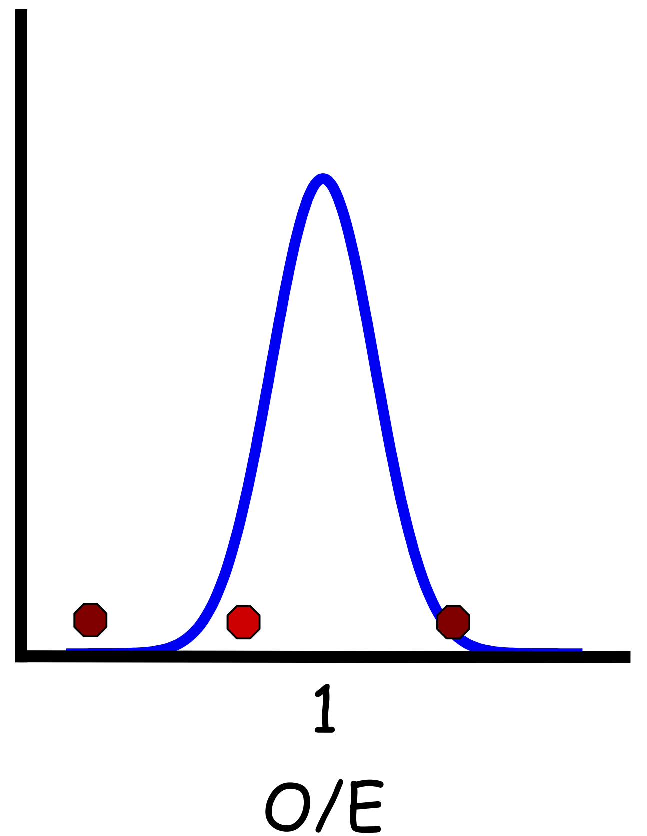
Hypothesis: the observed O/E value is from the same distribution of values estimated for reference sites, i.e., the site is equivalent to reference.



Statistical Issues Regarding Inferences of Impairment

Multiple Sites or Replicated Samples at a Site

Hypothesis: the observed mean is different from 1 (the reference mean).



RIVPACS-type Models: 8 Basic Steps

1. Establish a network of reference sites.
2. Establish standard sampling protocols.
3. Classify sites based on their biological similarity.
4. Estimate individual probabilities of capture by relating environmental setting to the biological classification (multivariate statistics).

For each assessed site:

5. Sum p_c's to estimate E.
6. Count O
7. Calculate O/E.
8. Determine if observed O/E is different from reference?

RIVPACS Outputs Can Also Be Used to Identify Sensitive and Tolerant Taxa

Sensitivity Index:

$$\frac{\textbf{\# sites taxon was observed}}{\textbf{\# sites taxon was expected}}$$