Development of RIVPACS models used in the State-Wide Condition Report

Chuck Hawkins
Western Center for Monitoring and Assessment of Freshwater Ecosystems
Aquatic, Watershed, & Earth Resources
Ecology Center
Utah State University

12th Annual Meeting
California Aquatic Bioassessment Workgroup
Davis, California
1 November 2005

Outline

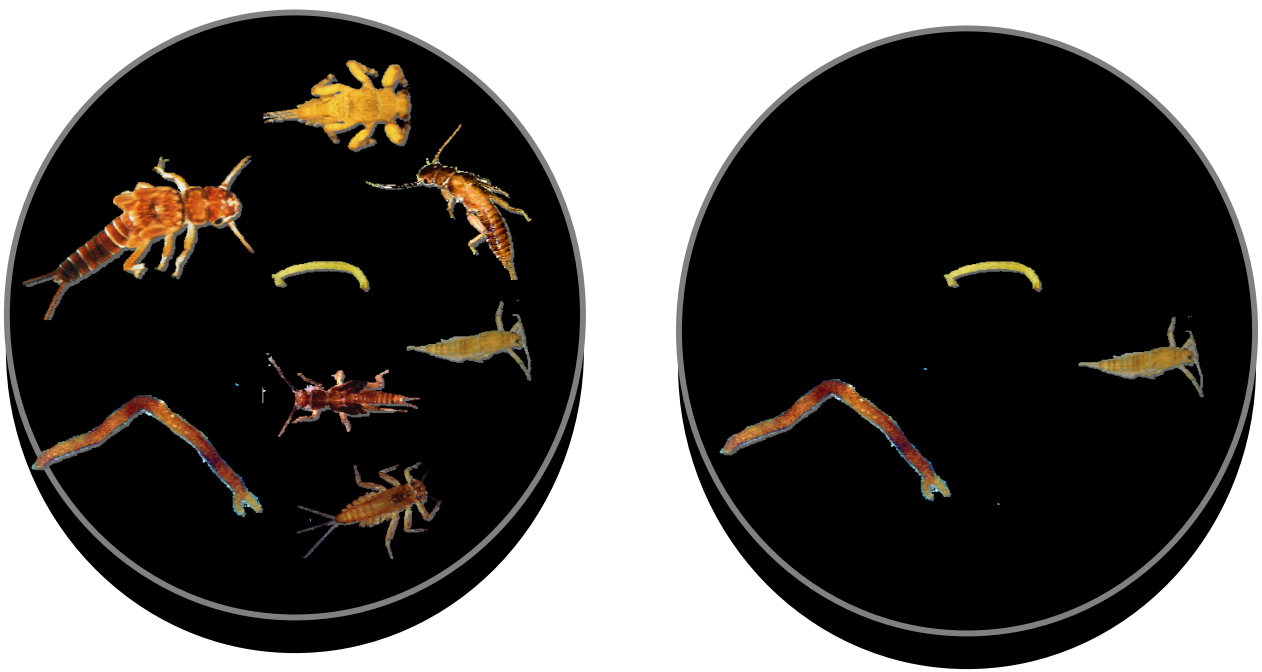
- RIVPACS – O/E as a general measure of biological condition.
- RIVPACS models for CA.

O/E:

Standardized, Site-Specific Assessments

- Recognize that natural ecosystems vary continuously and often markedly in their expected biota.
- Use model to describe the expected biota for individual sites (site-specific).
- Assess biological integrity as the relative degree to which observed (O) biota match that expected (E) for the site (standardized).

O/E is a measure of the taxonomic completeness of the biological community observed at a site



E = 8 taxa

O = 3 taxa

O/E

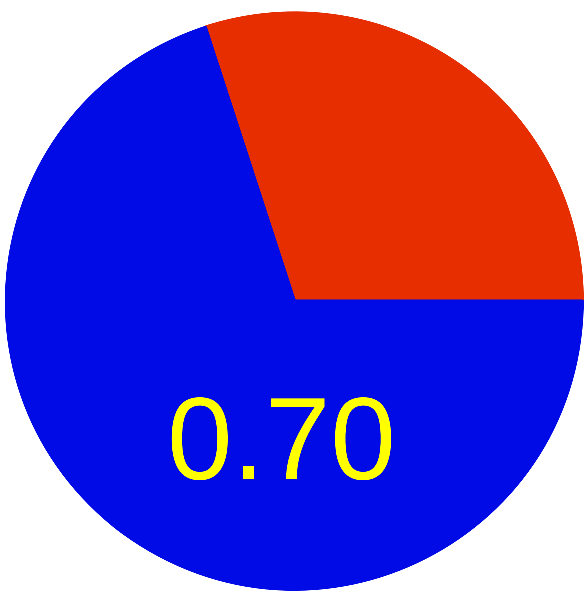
0.38

O/E standardizes assessments across sites that differ naturally in the number of expected taxa

Site 1

O = 7

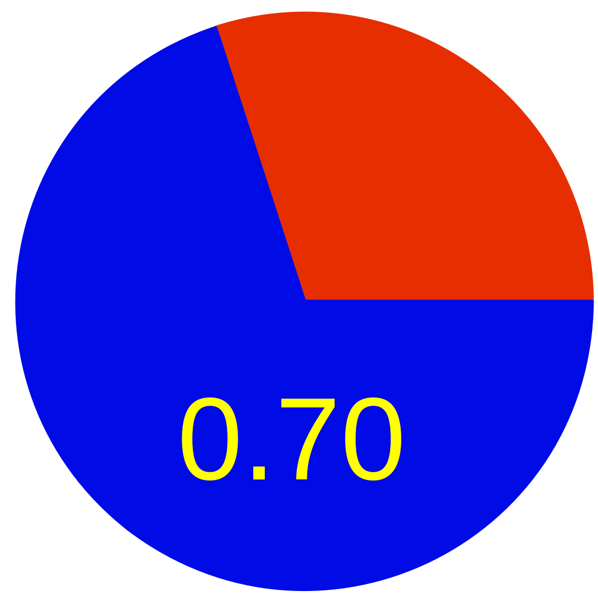
E = 10



Site 2

O = 21

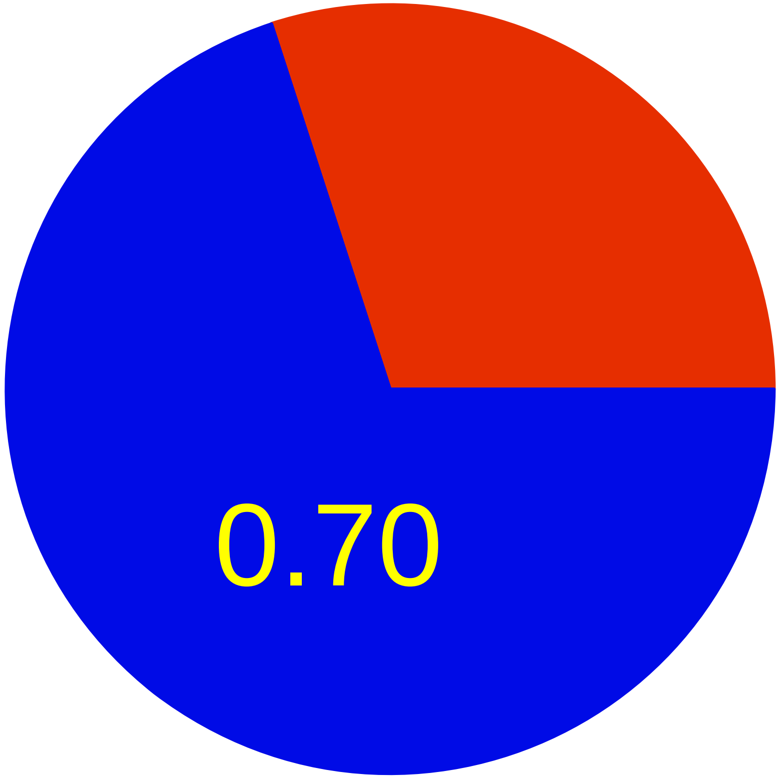
E = 30



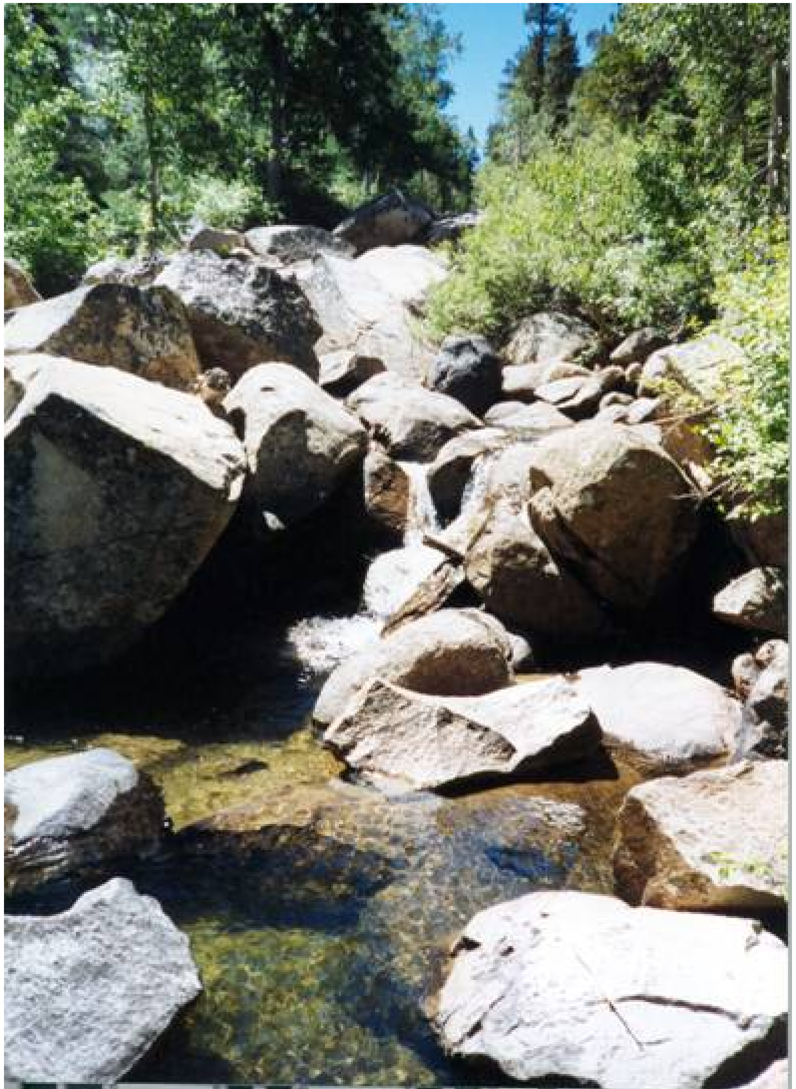
O/E is a measure of '**ecological capital**', a fundamental component of biological integrity[1].

[1]Ecological Indicators for the Nation. 2000. National Academy Press.

Site or regional assessment



O/E Allows Comparison of "Apples" and "Oranges"



The Technical Challenge:

Accurately and precisely describing the biota expected in different waterbodies in a State.

O/E Modeling and Assessments

1. Develop statistical models that predict the probabilities of capturing (*pc*) any taxon in the region of interest at any assessed site.
2. Compute O/E from sample data (O) and predictions of (E) derived from estimates of *pc*.
3. Assess site condition in the context of model error.

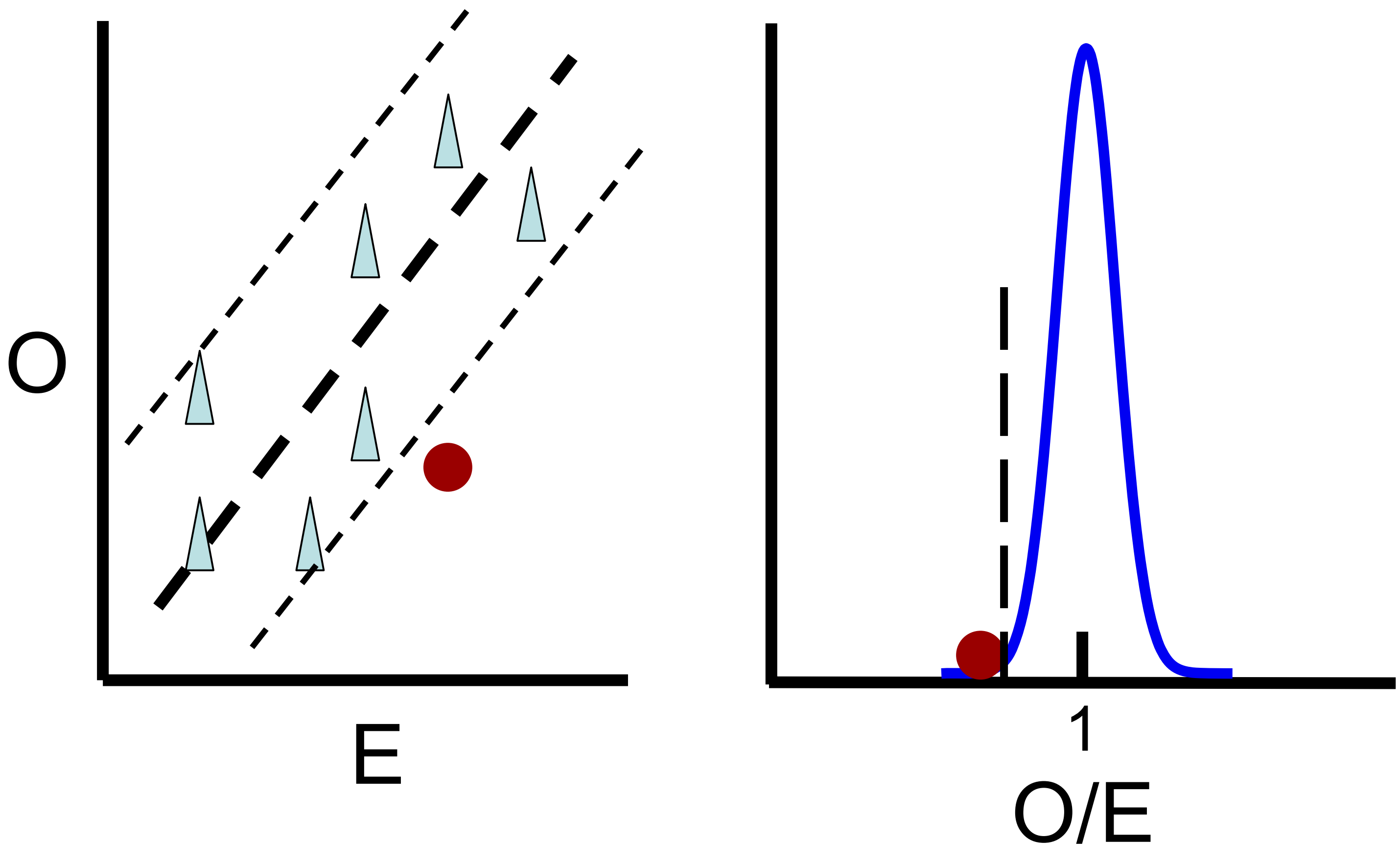
How O/E is <u>Calculated:</u>

Sum of taxa pc's estimates the number of taxa (E) that should be observed at the site given standard sampling.

Taxa	pc	O	O_2	O_3
Atherix	0.92	*	*	
Baetis	0.86	*		*
Caenis	0.70		*	*
Drunella	0.63		*	*
Epeorus	0.51	*		
Farula	0.32			
Gyrinus	0.07			
Hyalella	0.00			
E	4.01	3	3	3

$$O/E = 3 / 4.01 = 0.75$$

Need to Estimate Prediction Error for Site Assessments



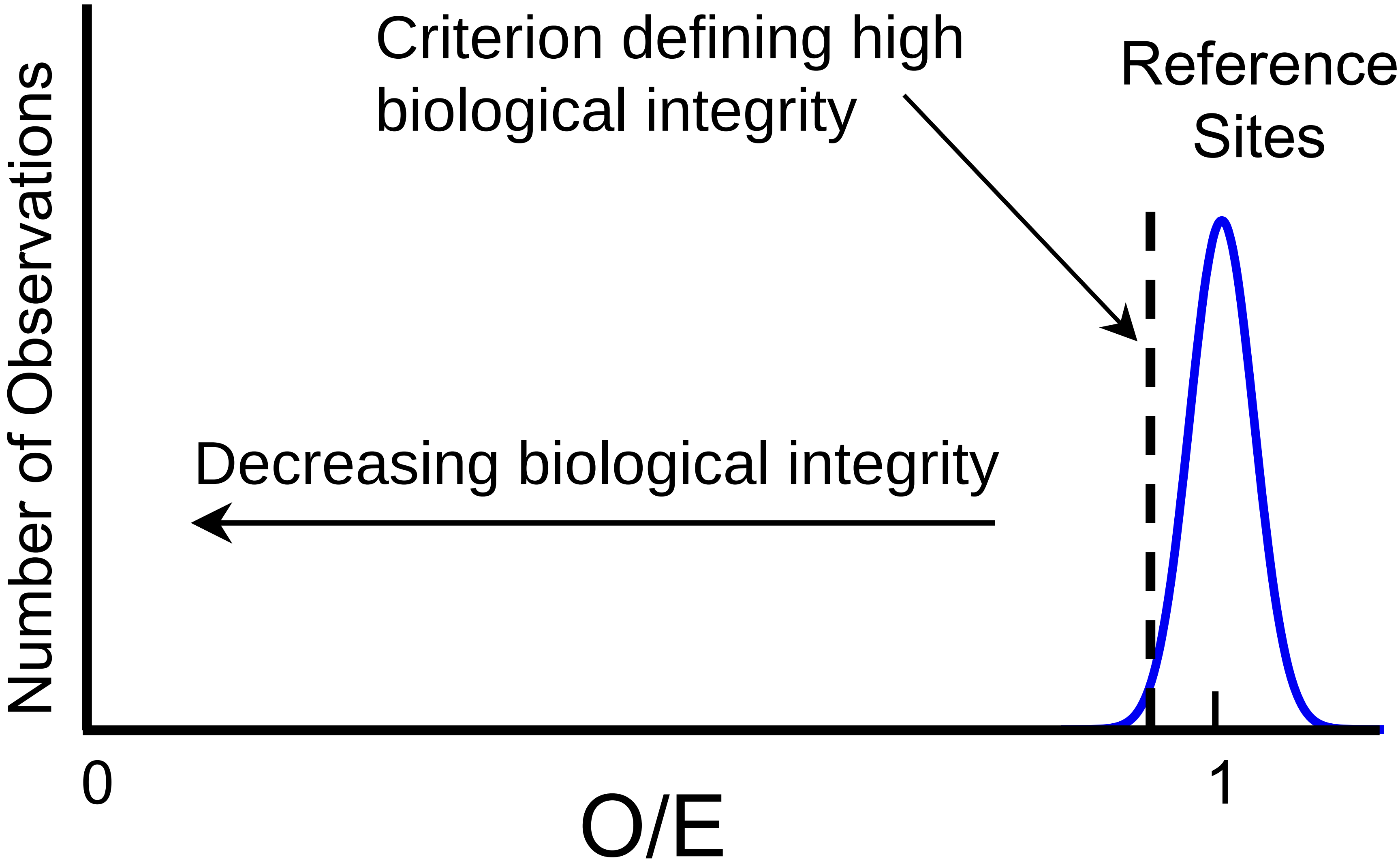
How Good Are the Models?

- Null Model O/E SD = variation in reference sample O/E values estimated by creating a model from 1 class (i.e., all streams are assumed to be alike).
- Random Sampling Error (RSE) = variation associated with only sampling error, i.e., the ‘perfect model’.
- Model O/E SD = variation in reference O/E values after accounting for effects of model predictor variables on E.

RSE Good Model Poor Model Null

0.10 0.30

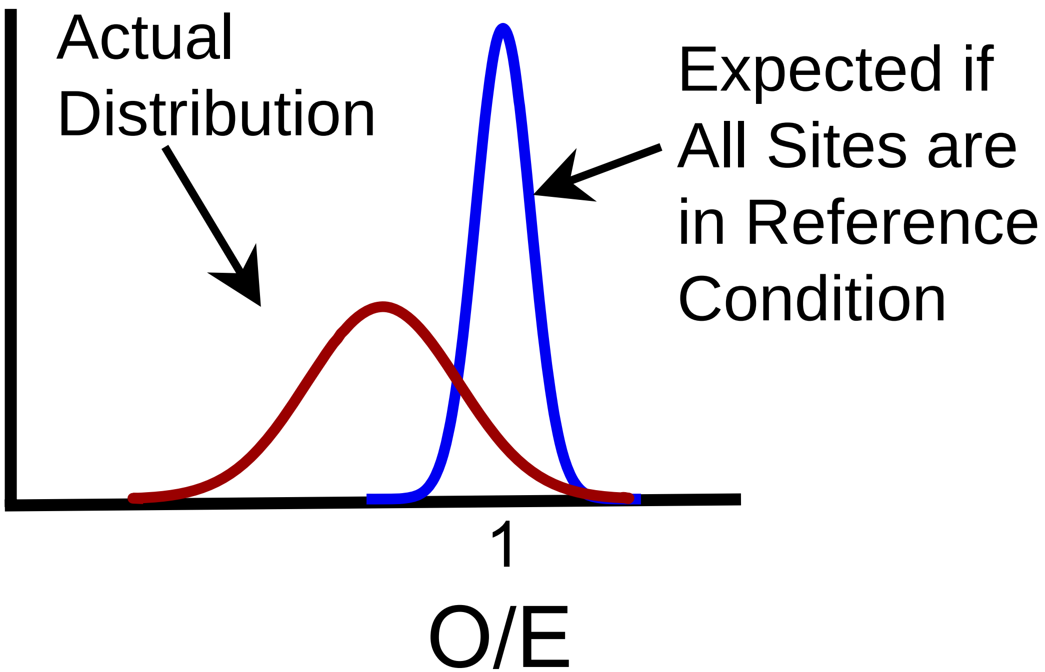
Assessing Biotic Condition

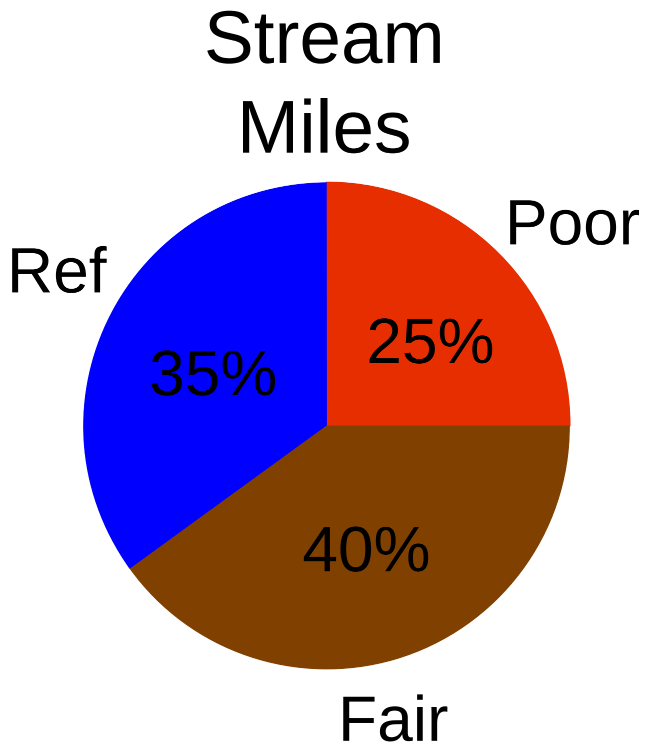


Standardized units of O/E allow assessments to be directly compared across diverse types of streams and easily aggregated for regional assessments

- General Accounting Office (2000)
- Heinz Center Report (2002)
- EPA Draft Report on the Environment (2003)

For Regional Assessments, We Want to Compare the Distribution of Observed O/E Values with the Expected Distribution





Caveats

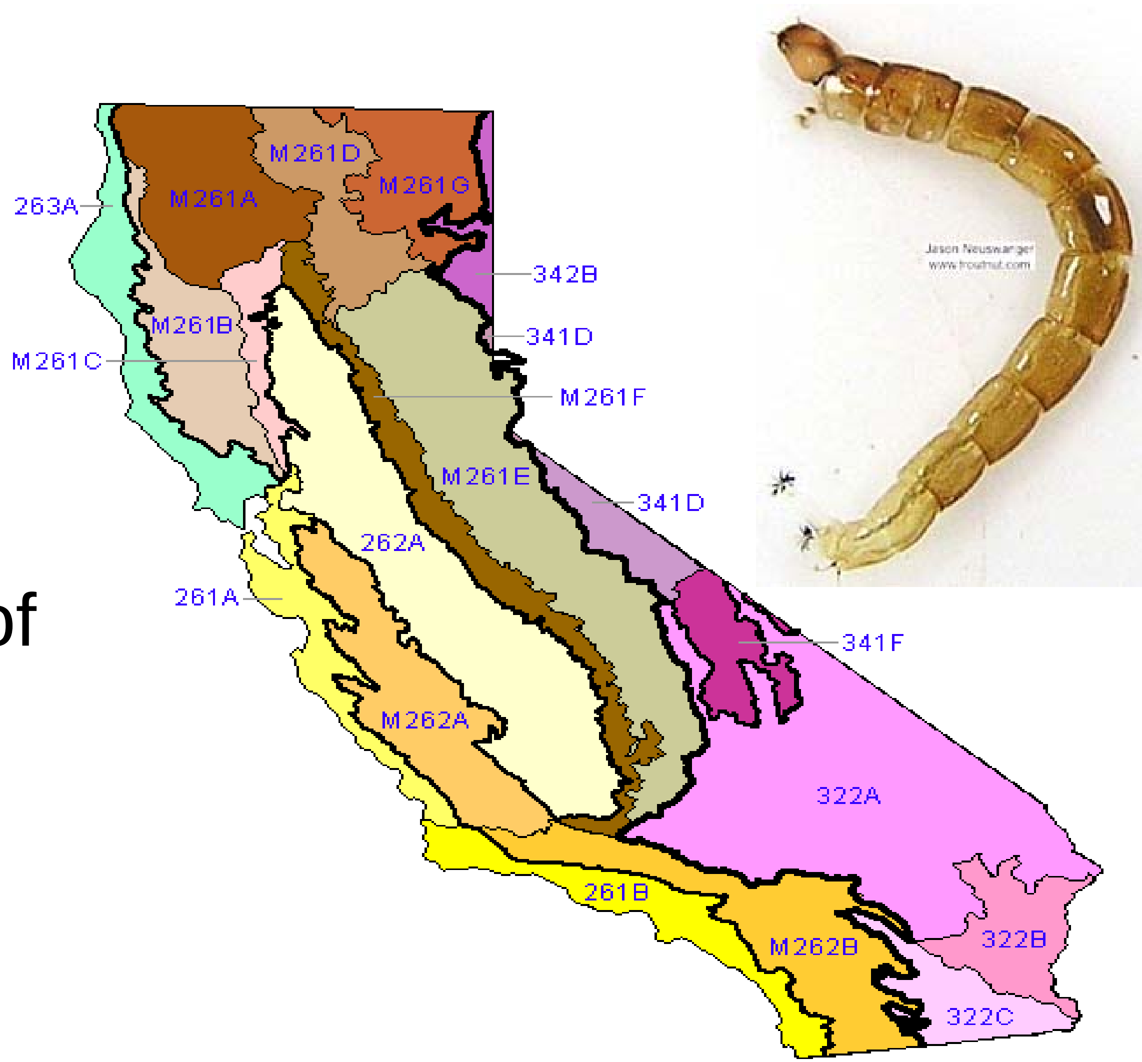
- Taxonomy must be consistent across streams and regions.
- Sampling methods must be similar.
- Reference site quality must be similar.

History of the California Models

- 1998 - Proof-of-concept model built with data collected by USU from USFS lands (Ecological Applications, 2000, 10:1456–1477)
- 2002 - Small contract with USFS to build a robust model for region-wide application.
- 2003 - Single model based on 1 yr of samples looked promising.
- 2004 – USFS filled in reference site data gaps.
- 2005 – Western EMAP data (plus USU STAR data) become available. Single model based on combined data was imprecise.
- 2005 – Separated sites into 3 hydro-climatic regions. Models perform well.

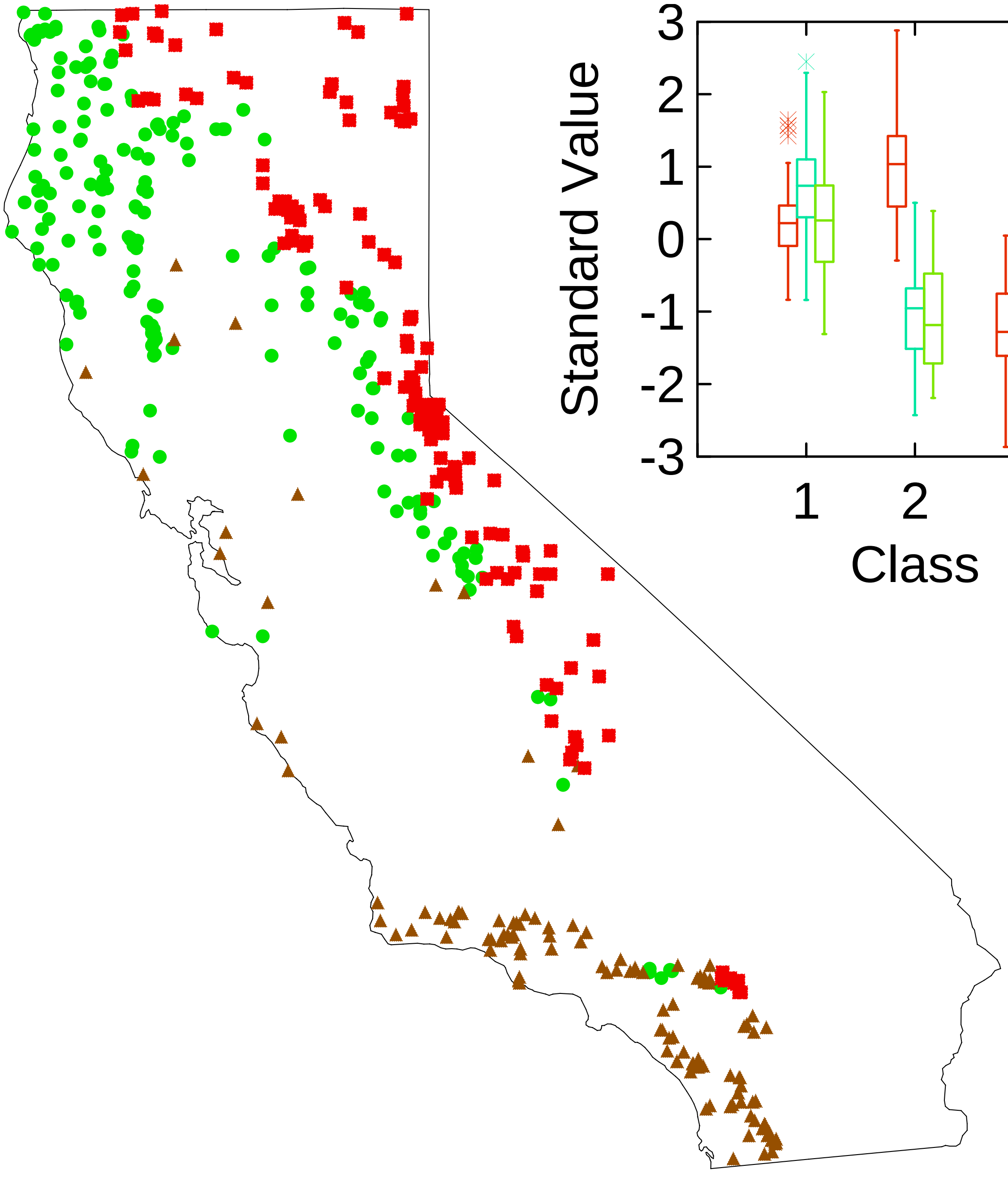
The Current California Models

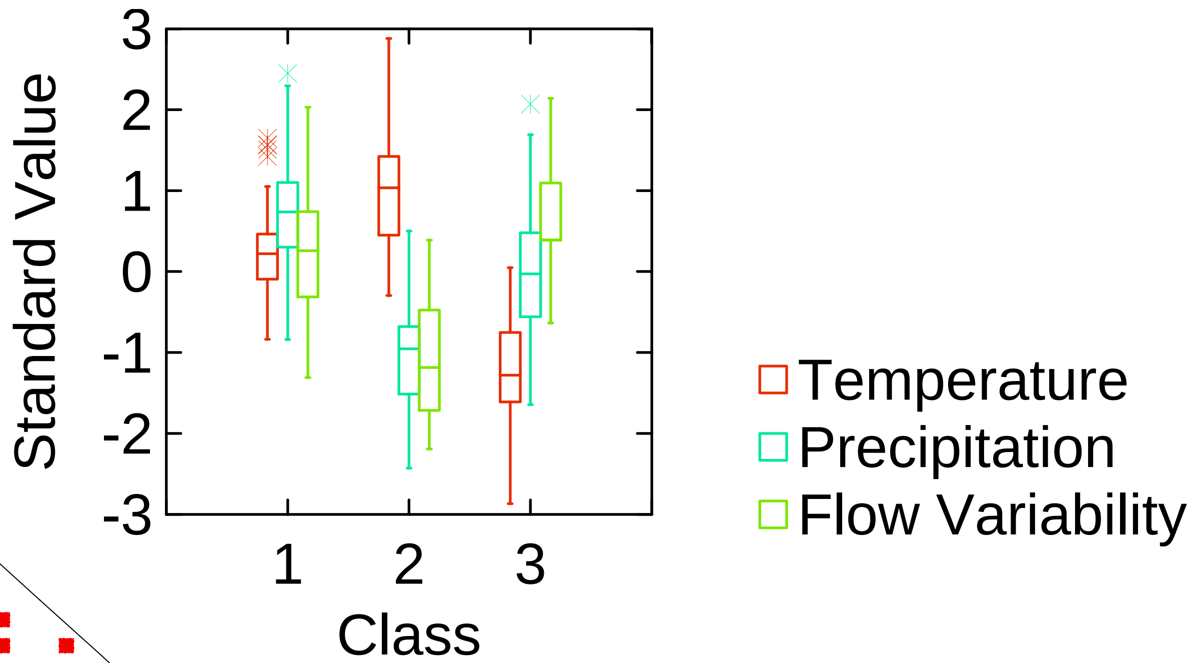
- Data Used
- How Many Models?
  - One would be nice, but……
  - Midges = 1/3 of taxa, but….
- General Performance



Data Used

- 446 Candidate Samples
  - 260 R5 USFS
  - 181 Western EMAP
  - 5 USU (portions of CA assumed to be OR!)
- Reference Sites
  - 240 total after dropping samples with problems
  - 206 with >= 300 individuals
- Non-Reference Sites
  - R5 = targeted sites
  - EMAP = random selection



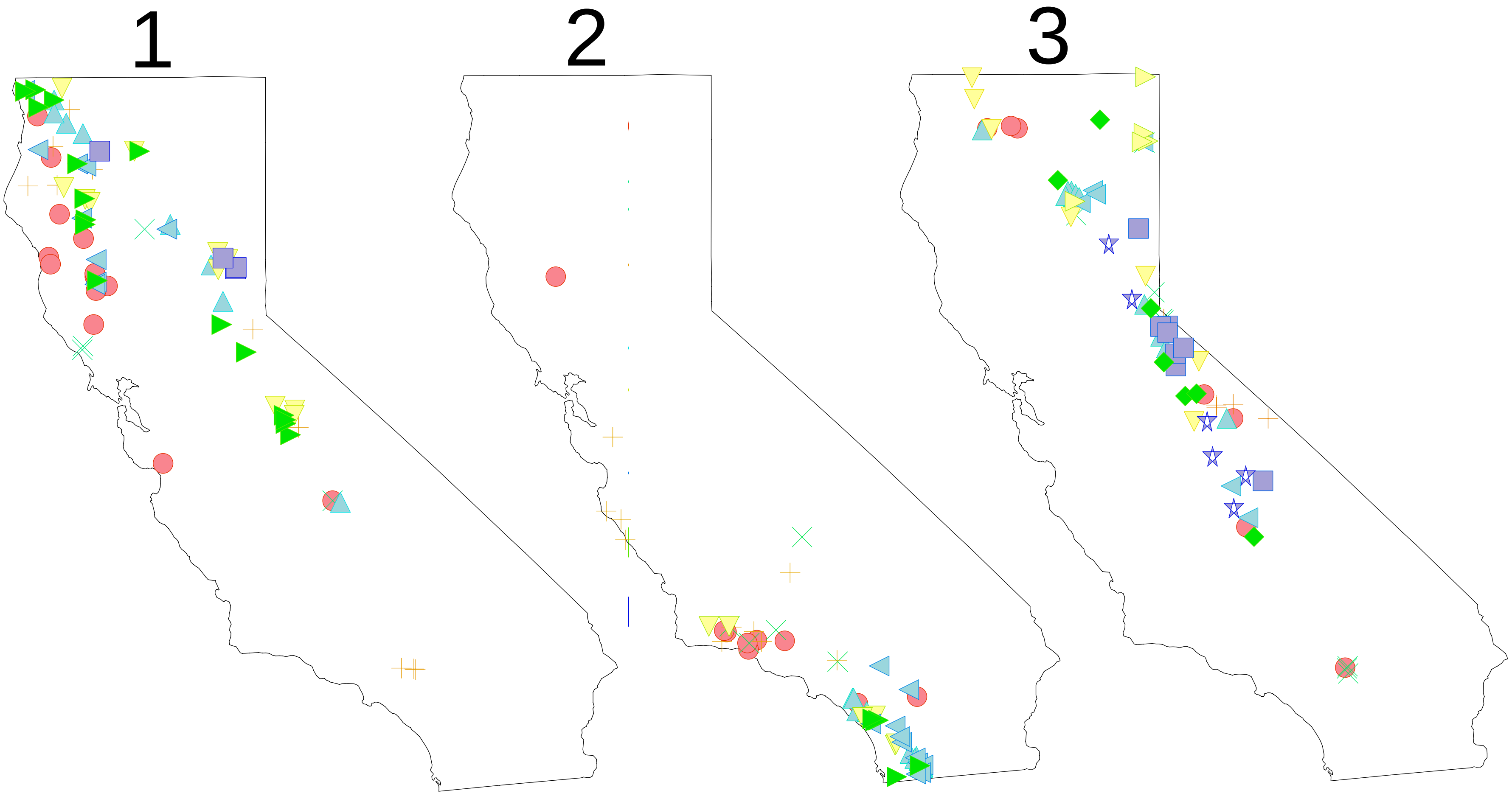


Tried one model. Performance was marginal, so...

Created a 3 class hydro-climatic classification based on data from all sites.

Distribution of Reference Sites for Each of the 3 Hydro-climatic Types

(biotic classes used in modeling are color coded)



Potential Predictor Variables

Map

- Latitude
- Longitude
- Elevation

GIS-Derived

- Mean annual precipitation
- Mean annual air temperature
- Basin area
- % Basin geology (7 classes)

Field

- Channel slope (%)
- % Substrate (64-250mm)
- Log alkalinity
- Sampling date

Used John Van Sickle's All Possible Subsets 'R' program to select the 'best' model for each hydro-climatic stream type from >32,000 possible models.

Best = optimal combination of:

1. precision (small reference O/E SD)
2. use of map/GIS variables
3. avoiding over-fit models

Models and Predictor Variables (midges to genus)

Model 1 (8 classes)		Model 2 (11 classes)		Model 3 (10 classes)	
Variable	F	Variable	F	Variable	F
WSA	9.99	Long	7.66	WSA	6.51
Long	7.62	Precip	4.42	Temp	3.60
Lat	6.90				
Temp	2.81				

Models and Predictor Variables (midges to subfamily)

Model 1		Model 2		Model 3	
Variable	F	Variable	F	Variable	F
WSA	8.84	Long	5.52	WSA	7.13
Temp	8.46	% Sed	2.67	Temp	4.25
Lat	8.14	Precip	1.80		

Hydro-Climatic Class 1
(Wet and Cool)

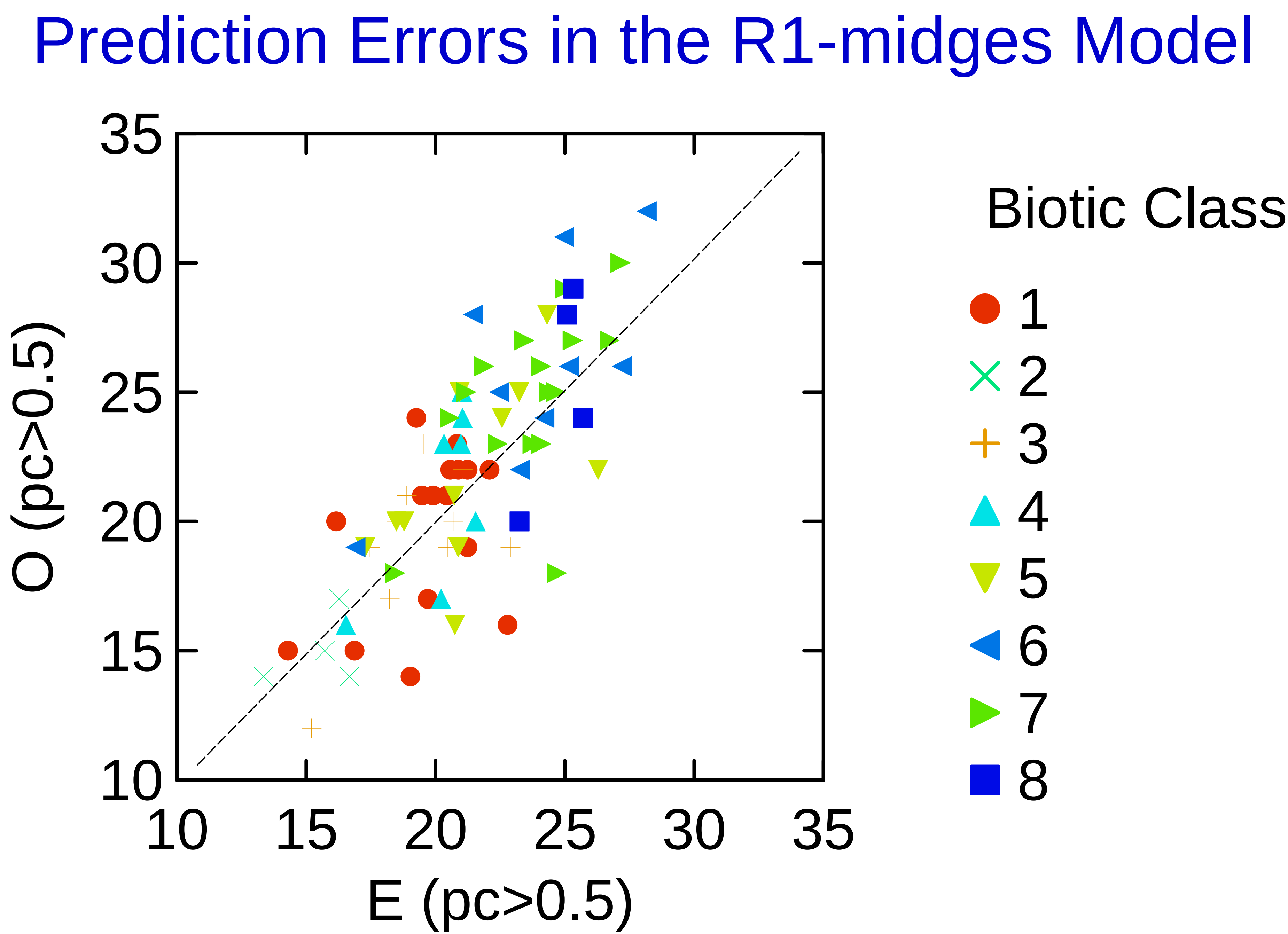
Parameter	Midges to Genera		Midges to Subfamilies	
	R	T	R	T
Mean	1.03	0.84	1.03	0.84
Model SD	0.13		0.13	
Null SD	0.17		0.15	
RSE	0.11		0.11	

Hydro-Climatic Class 2
(Dry, Warm, Flashy)

Parameter	Midges to Genera		Midges to Subfamilies	
	R	T	R	T
Mean	1.04	0.76	1.02	0.73
Model SD	0.17		0.16	
Null SD	0.19		0.19	
RSE	0.15		0.14	

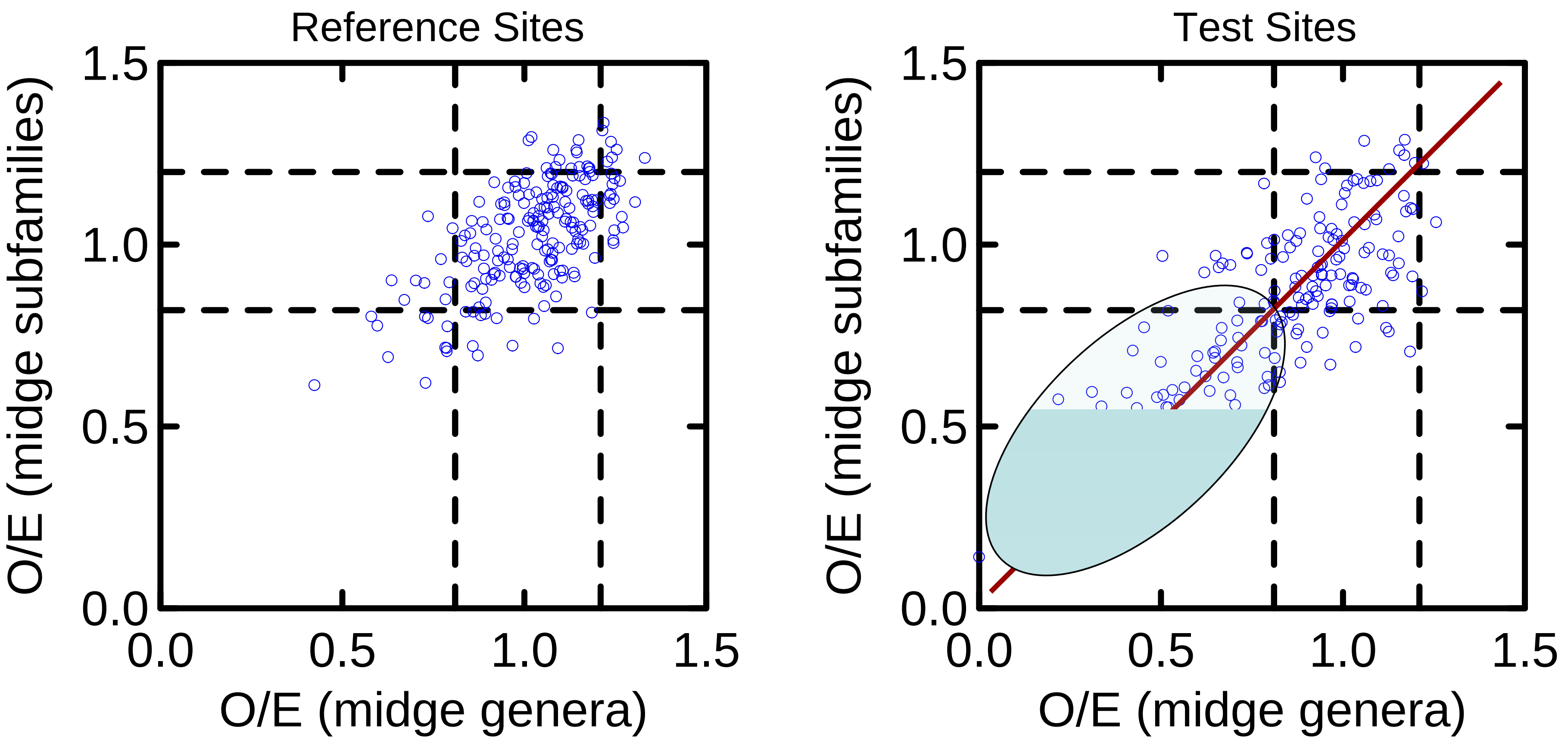
Hydro-Climatic Class 3
(Mesic and Cold)

Parameter	Midges to Genera		Midges to Subfamilies	
	R	T	R	T
Mean	1.01	0.80	1.03	0.81
Model SD	0.16		0.15	
Null SD	0.18		0.18	
RSE	0.14		0.14	



Prediction Errors in the R1-midges Model
O (pc>0.5)
E (pc>0.5)
10
15
20
25
30
35
Biotic Class
1
2
3
4
5
6
7
8

Relationship Between O/E Values Based on Models With and Without Midge Genera (outputs from all 3 models combined)



Summary of Model Performance

1. All 3 models are substantially better than null models in precision and similar to good to excellent models developed elsewhere.
   - Should be excellent for site-specific assessments.
2. Precision was similar among the 3 models.
   - Aids in regional comparisons and state-wide integration.

Summary of Model Performance

3. Models based on midges identified to subfamily were nearly identical, on average, in their assessments as models based on midges identified to genus.
   - USFS and CA/EMAP assessments can be compared/combined.
4. These ‘subfamily’ models did systematically underestimate the degree of impairment at impaired sites and hence represent a conservative assessment.

Information

- chuck.hawkins@usu.edu
- www.cnr.usu.edu/wmc
- People
  - Joseph Furnish (R5 USFS)
  - Pete Ode and Andy Rehn (CA ABL)