Bio-Objectives Technical Update:
Scoring Tool Development and Testing

RESOURCES AGENCY CALIFORNIA DEPARTMENT FISH & GAME

Aquatic Bioassessment Laboratory

Water Boards

SWAMP Surface Water Ambient Monitoring Program

Technical Update:
Scoring Tool Development and Testing

- **Review of reference work and O/E process**
- **Building the model and early exploration**
- **Performance Tests**
  - **What we measured and why**
  - **Results**: statewide overview and regional comparisons
- **What's next**
- **Questions for Science Panel**

SWAMP
Surface Water
Ambient Monitoring
Program

Technical Team

Peter Ode, *DFG-WPCL/ABL*

Ken Schiff, *SCCWRP*

***Andy Rehn,** *DFG-ABL*

***Raphael Mazor**, *SCCWRP +DFG-ABL*

Larry Brown, *USGS*

Jason May, *USGS*

David Gillett, *SCCWRP*

Eric Stein, *SCCWRP*

Betty Fetscher, *SCCWRP*

Kevin Lunde, *SF Water Board*

David Herbst, *SNARL*

Questions for Science Panel

- What is your opinion of the assessment tool(s) developed so far?
  - How do our tools compare to other state/national efforts?
- What is your opinion of the performance measures we used?
  - What additional performance tests should we apply?
- Which O/E index would you recommend and why?
  - Should we apply evenness correction?
  - Should we regionalize scoring thresholds?
- Are there other scoring tools that we should explore?

SWAMP
Surface Water Ambient Monitoring Program

Objectives:

- Develop scoring tools to objectively assess biological condition of all CA wadeable perennial streams
- Requirement is to balance statewide consistency with regional validity
- Optimize tool based on multiple measures of performance

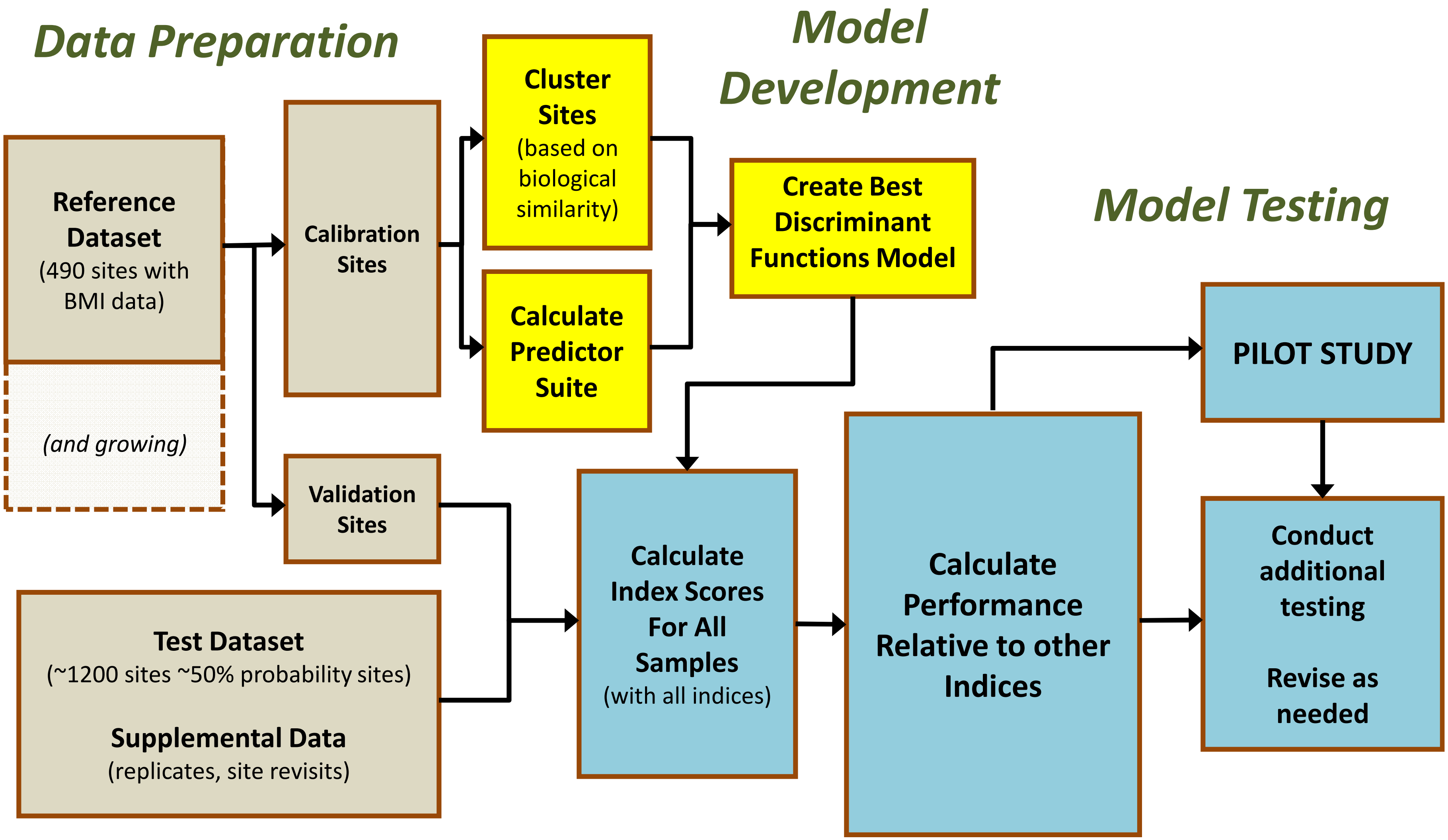
Why Develop A New Tool?

- Existing tools have limitations for statewide application
  - Spatial coverage is limited
  - Reference site definitions not consistent
  - Reference distributions not fully representative

- MMI (IBI) and O/E are both viable approaches; we focused on O/E
  - Designed to predict site-specific expectations, rather than a regional reference average
  - Species loss is a relevant measure of ecological condition
  - Index is amenable to statewide standardization

SWAMP
Surface Water Ambient Monitoring Program

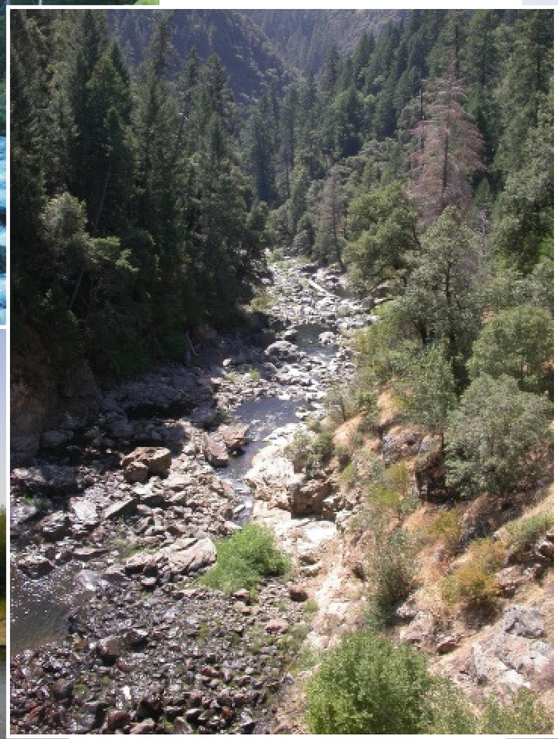
O/E Index Development Process



Scoring Tools Depend on Reference Sites

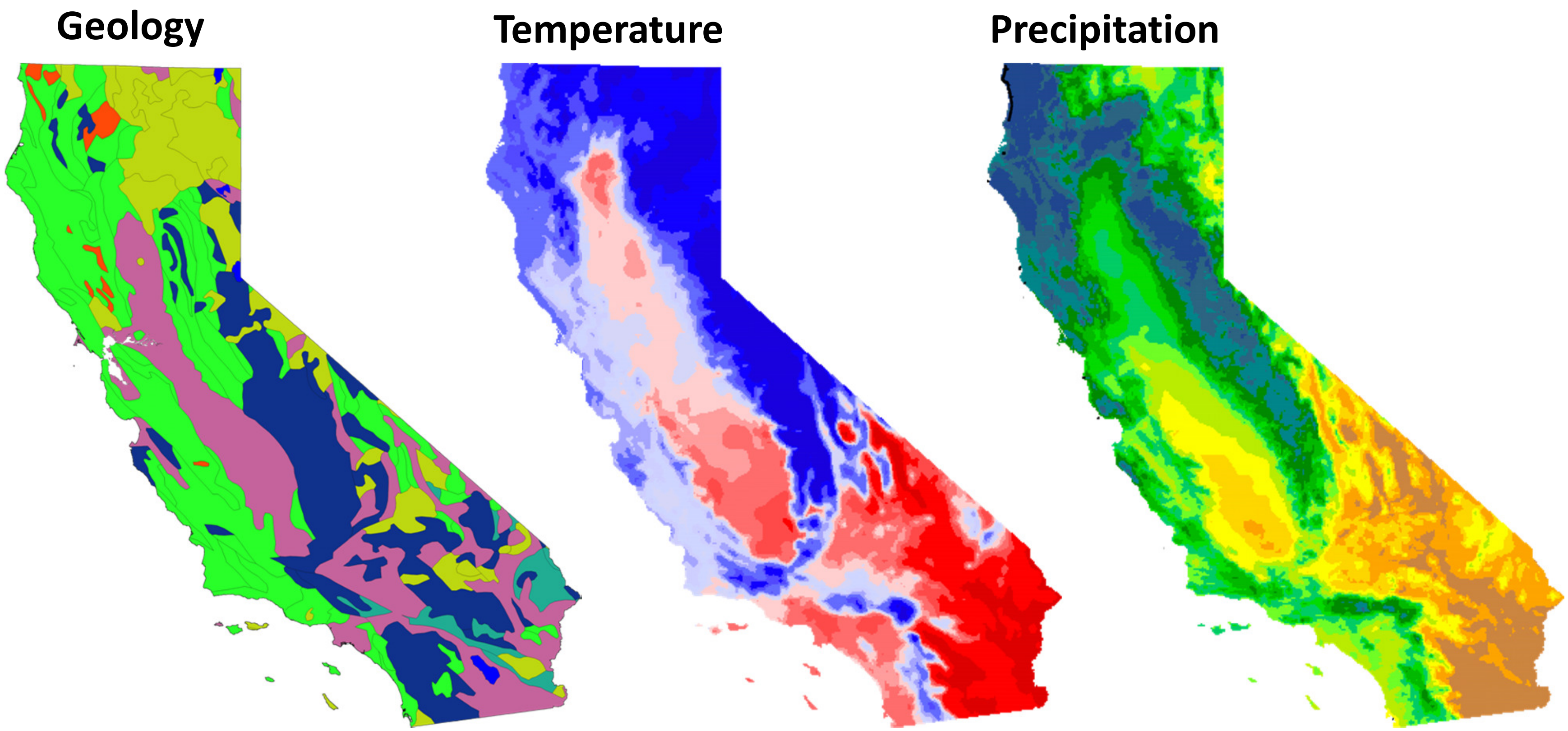
(sites with low levels of disturbance)

"What should the biology look like at a test site?"



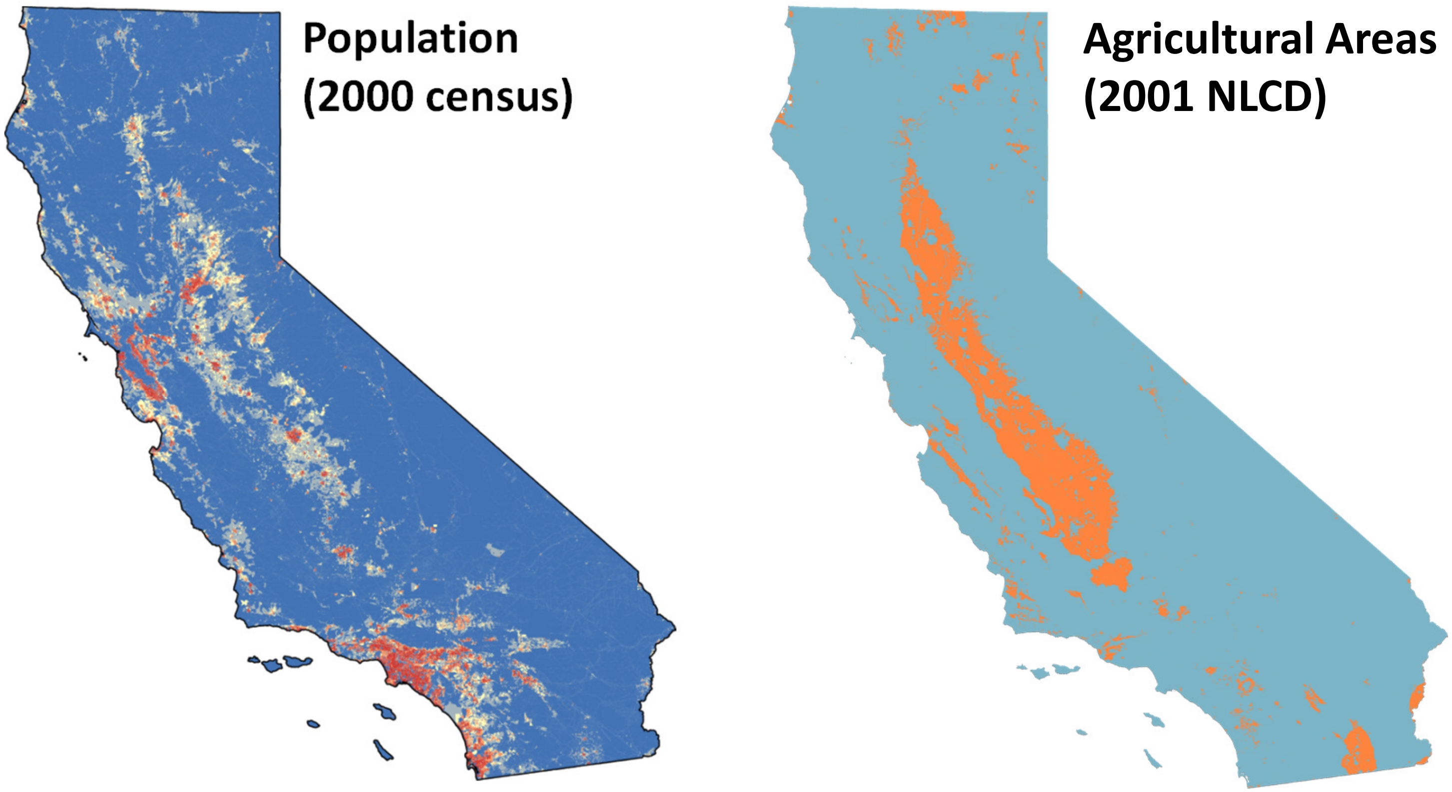
Technical Challenges:

*Strong natural gradients result in a large degree of **natural variation** in biological expectations*



Technical Challenges:

High degree of anthropogenic modification (e.g., impervious surface and intensive agriculture) in some regions



- Extensive modification introduces **gaps in representation** of natural gradients
- Widespread development can make some regions unsuited for standard reference approaches

Reference Criteria for Biological Objectives

Balancing site purity and representativeness

Trade-off: Need to allow limited sources of anthropogenic stress in order to get good representation of all stream types *(this constraint is shared by all bioassessment indices)*

Performance Objectives:

1. Reference pool represents all types of CA streams
2. Biological "quality" is maintained at reference sites

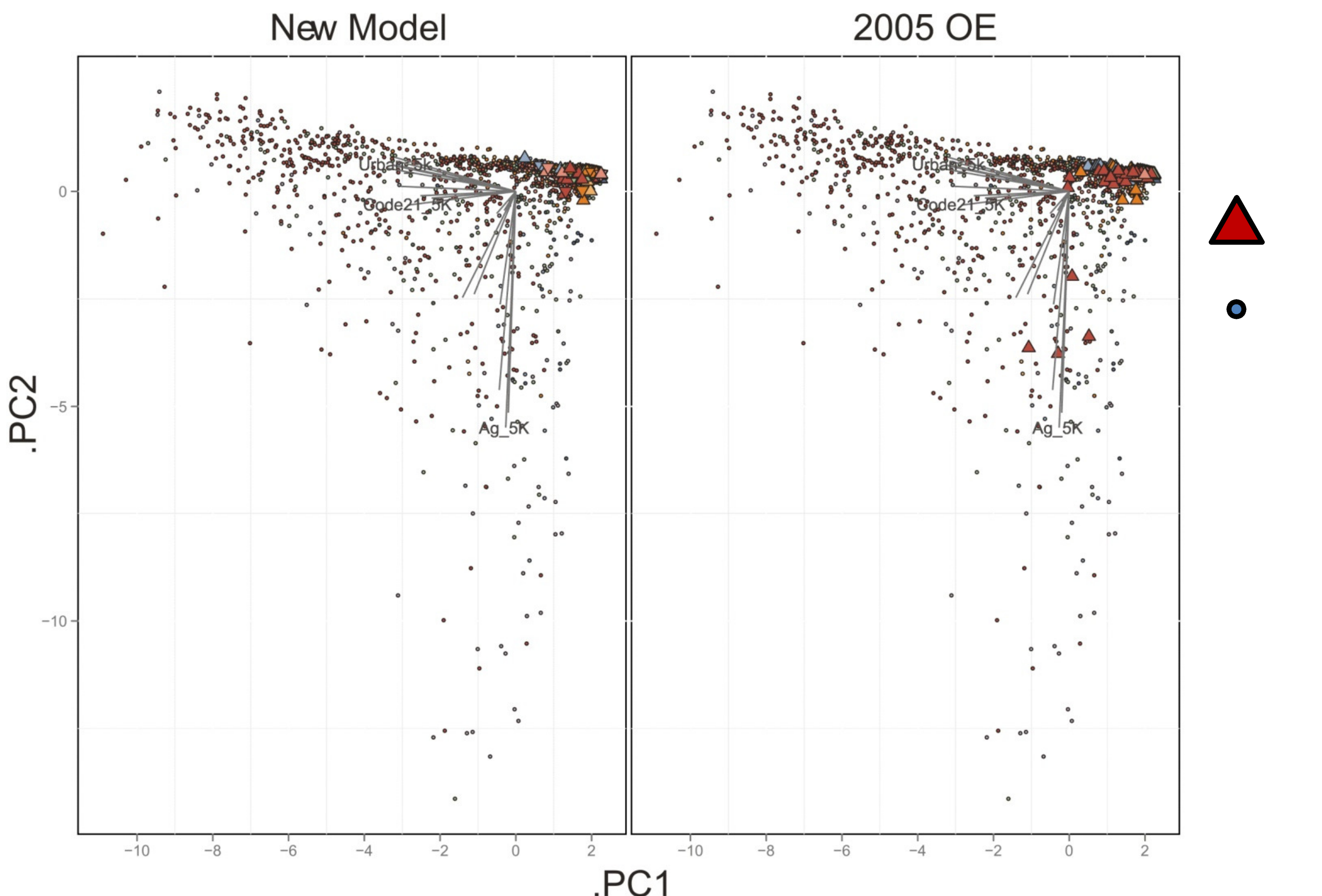
SWAMP
Surface Water Ambient Monitoring Program

Thresholds are comparable or stricter than other CA indices and include many more criteria

Metric	2011 Bio-objectives	South Coast IBI (5k,ws)	North Coast IBI (1k, ws)	Current O/Es (Hawkins 2005)	Restricted Model (1k, 5k, ws)
Local Disturbance (W1_Hall)	**1.5**	-	-	riparian vegetation, erosion, grazing, etc.	**1.0**
% Agricultural	**3,3,10**	**5**	5		**3**
% Urban	**3,3,10**	**3**	3		**3**
% Ag + Urban	**5,5,10**				**5**
% Code 21	**7,7,10**	**in urban**	**in urban**		**5**
Road Dens (km/km^2)	**1.5**	**2.0**	**1.5/ 2.0**		**1.5**
Paved road x-ings (#/ws)	**5/10/50**				**5/10/50**
TN, TP (mg/L)	**3.0/ 0.5**	-	-		**3.0/ 0.5**
Nearest Dams	**>10 km**	-	-		**>10 km**
Active Producing Mines	**0** (5k)	-	-		**0** (5k)
% Canals & Pipelines	**10**	-	-		**10**
Gravel Mine Density	**0.1** (r5k)				**0.1** (r5k)
Conductivity	**<2000 uS, + <99%, >1%**				**<2000 uS, + <99%, >1%**
BPJ Screen	**X**	**X**	**X**	**X**	**X**

Reference sites confined to streams with few sources of anthropogenic stress

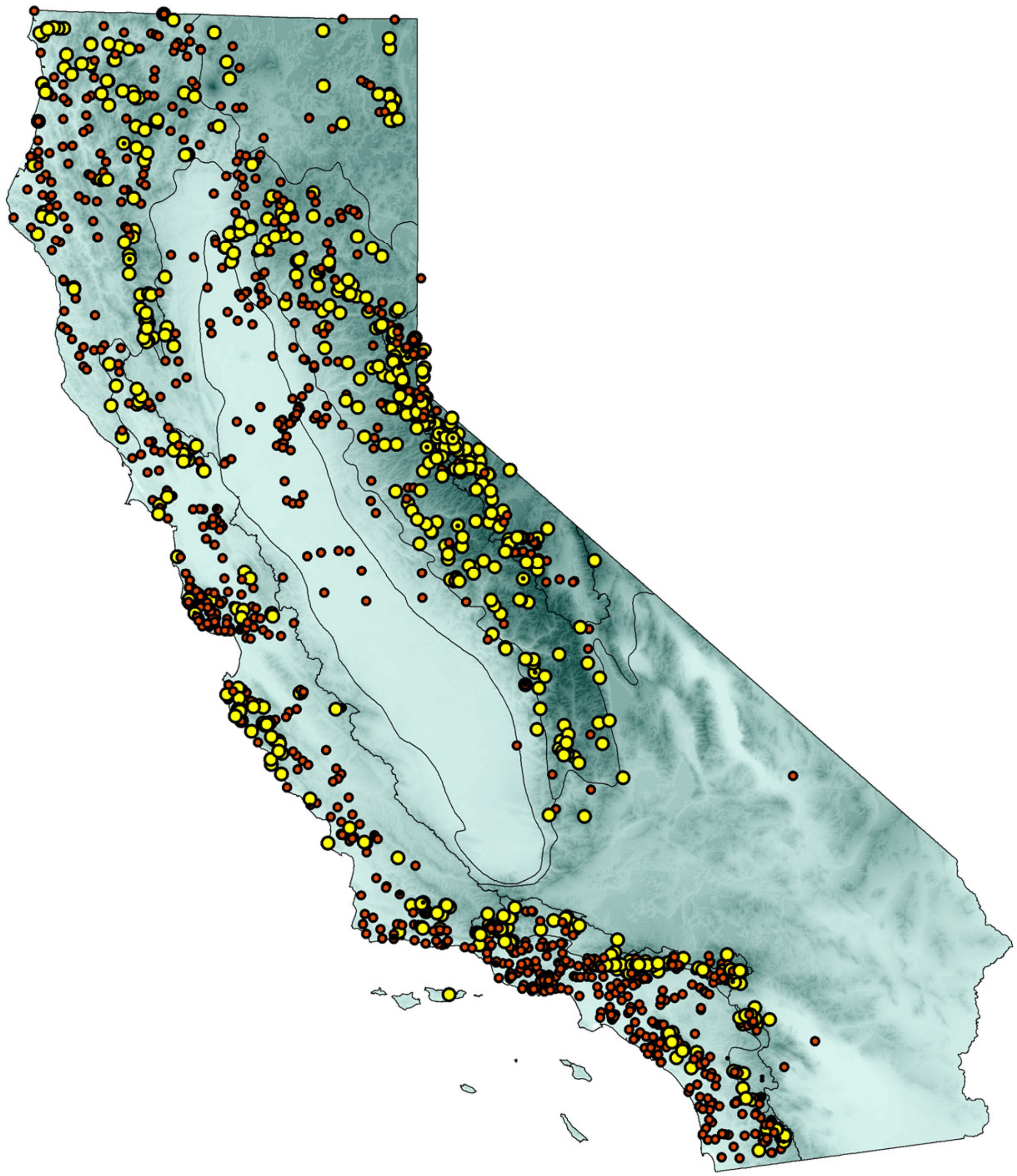
PCA axes are composite stressor gradients



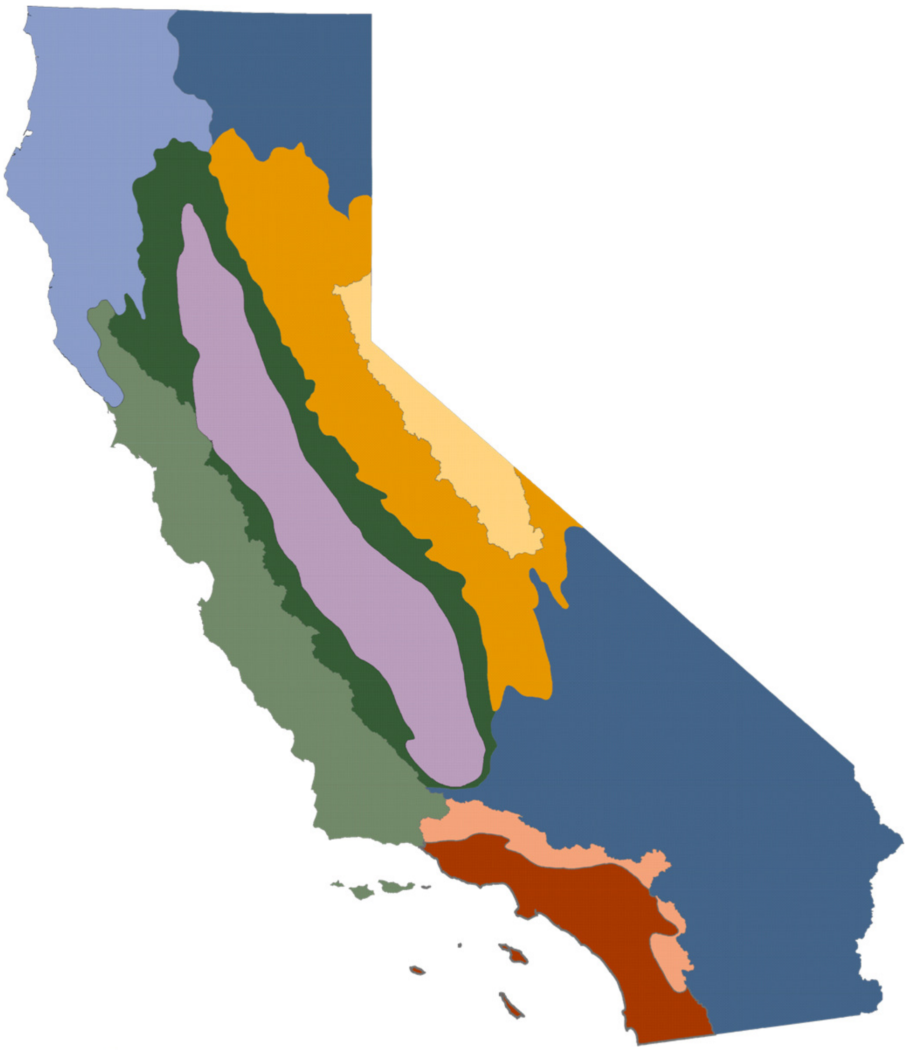
SWAMP
Surface Water Ambient Monitoring Program

Reference Sites

REGION	n
North Coast	79
Central Valley	1
Coastal Chaparral	87
Interior Chaparral	30
South Coast Mountains	96
South Coast Xeric	22
Western Sierra	131
Central Lahontan	142
Deserts + Modoc	27
TOTAL	**615**



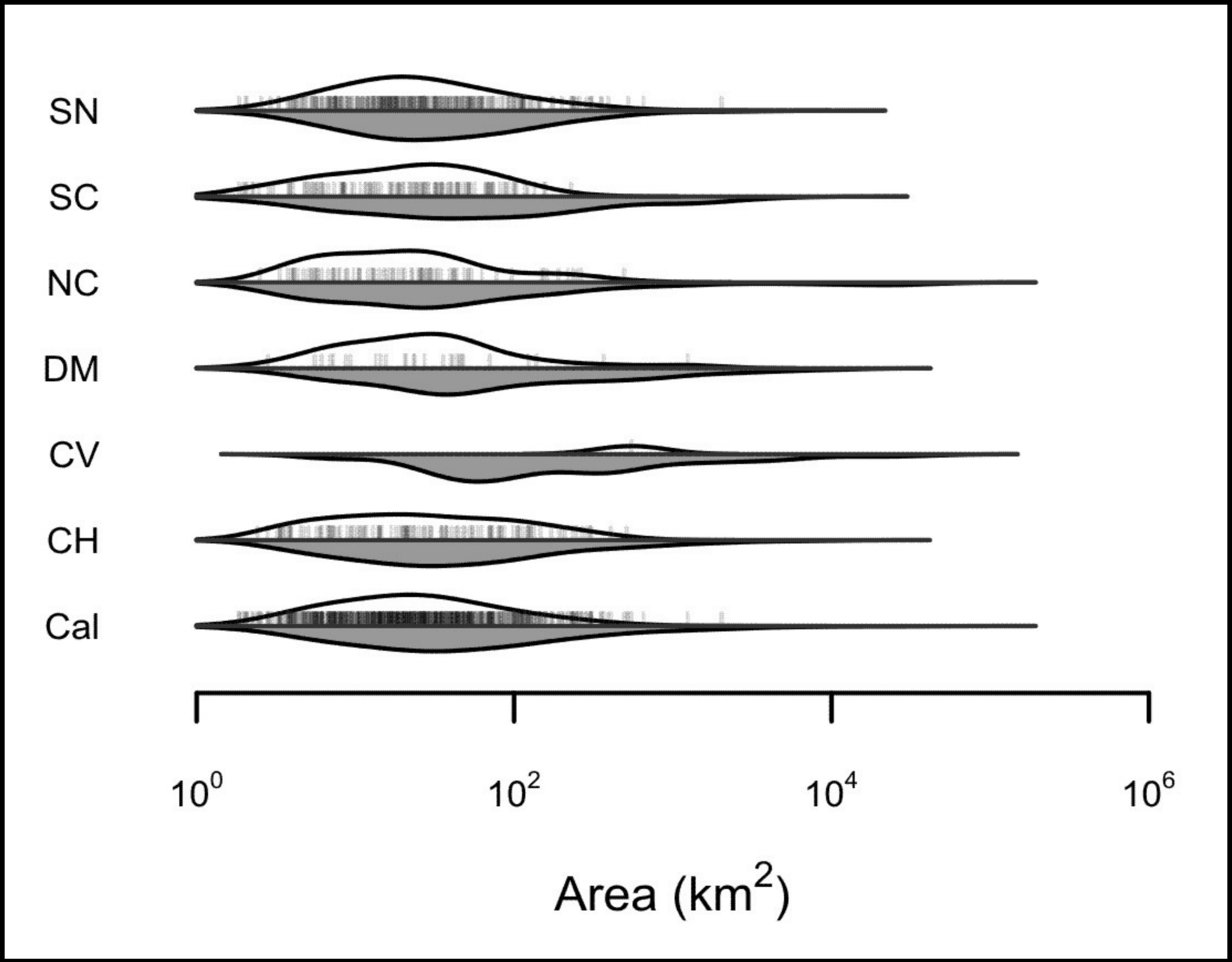
PSA Regions and Codes



Code	Region
CH-co	Chaparral (coastal)
CH-in	Chaparral (interior)
CV	Central Valley
DM	Deserts + Modoc
NC	North Coast
SC-m	South Coast (mountains)
SC-x	South Coast (xeric)
SN-cl	Sierra Nevada (central Lahontan)
SN-ws	Sierra Nevada (west slope)

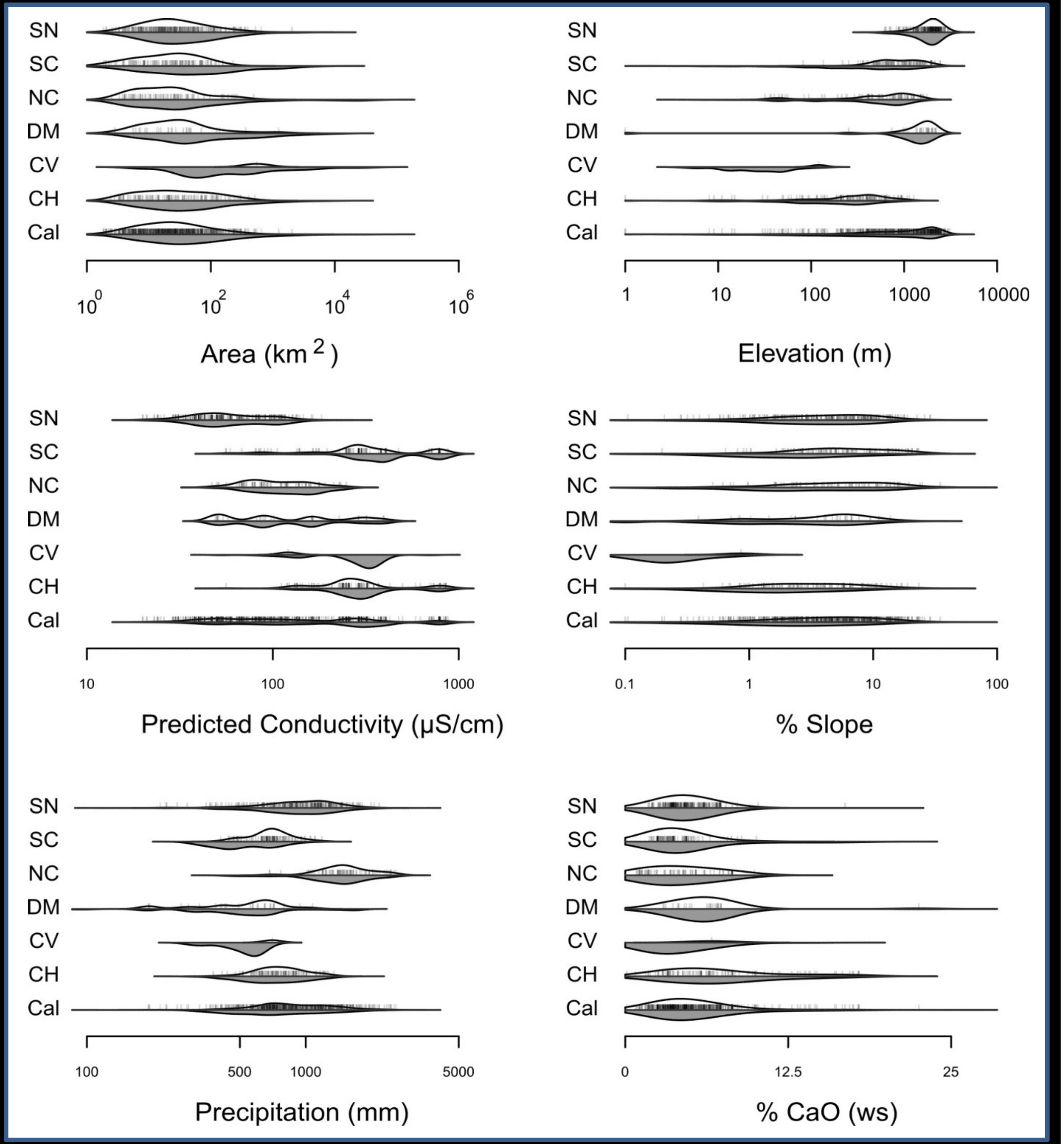
Environmental Representativeness:

"Beanplots" used to compare match between reference and overall distributions

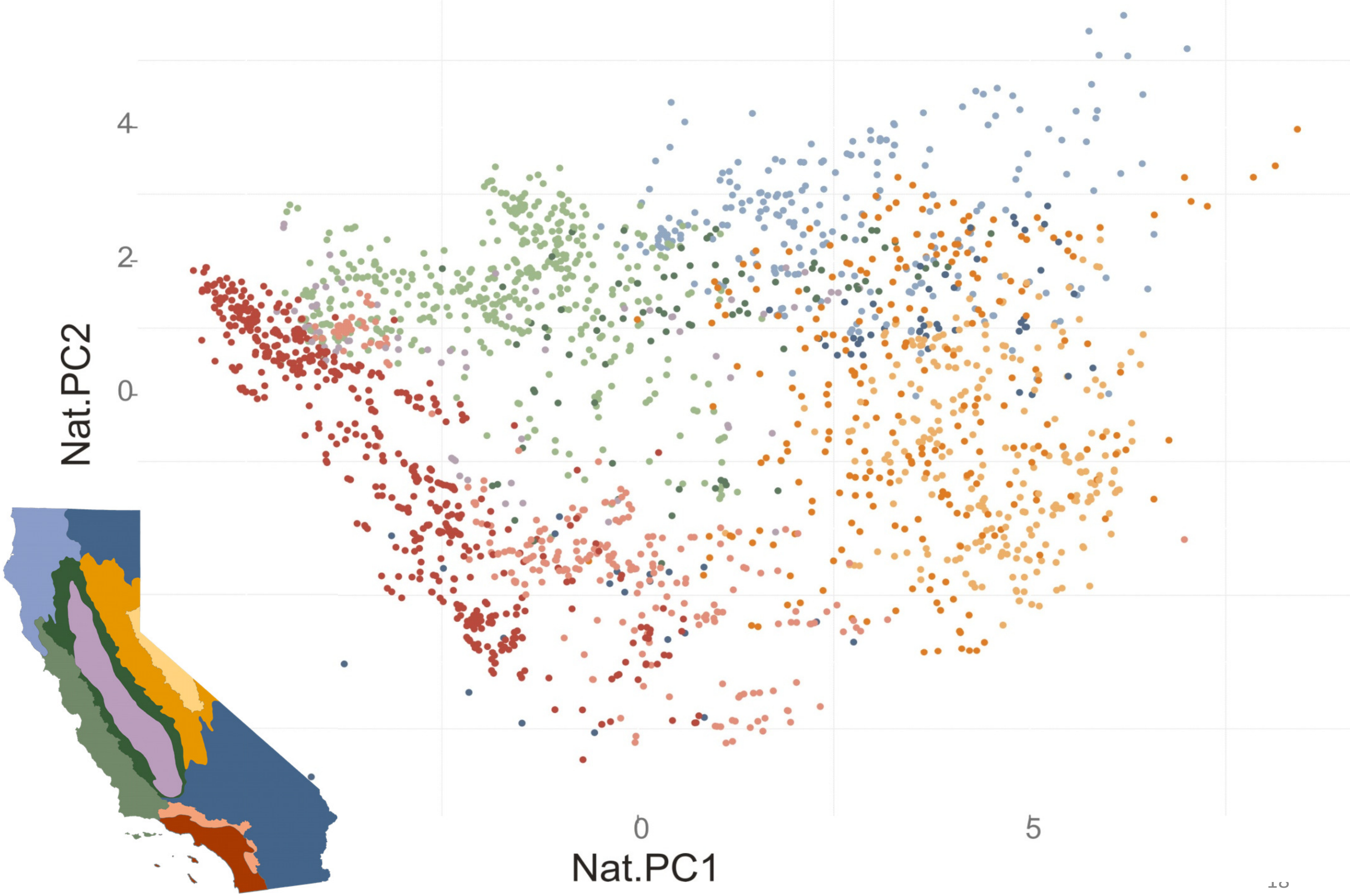


Univariate Gradient Representation

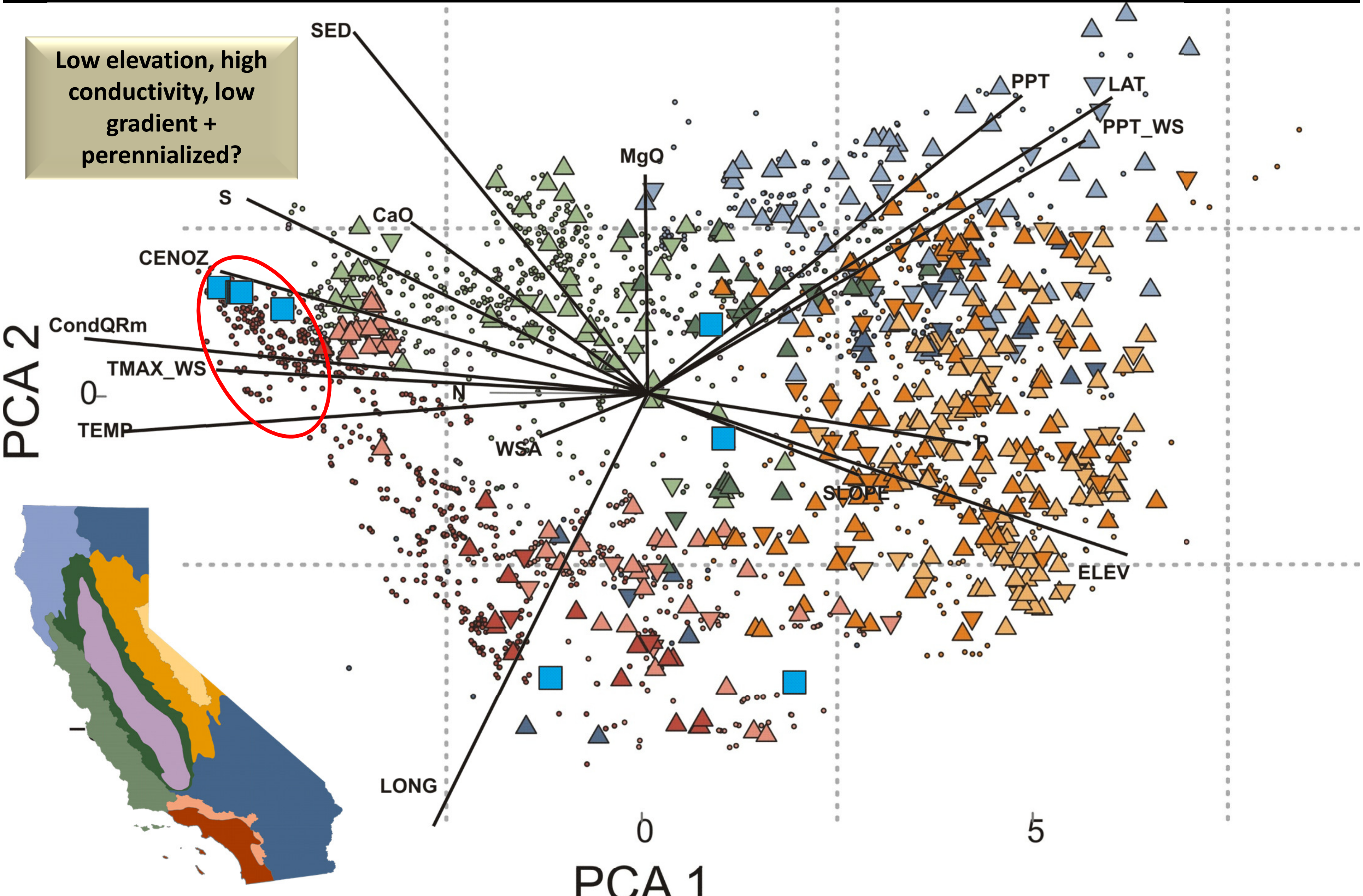
- Overall excellent representation in most regions
- Central Valley and South Coast (xeric only) very under-represented
- Very low gradient , large watershed, low elevation settings slightly under-represented in Chaparral/ S. Coast



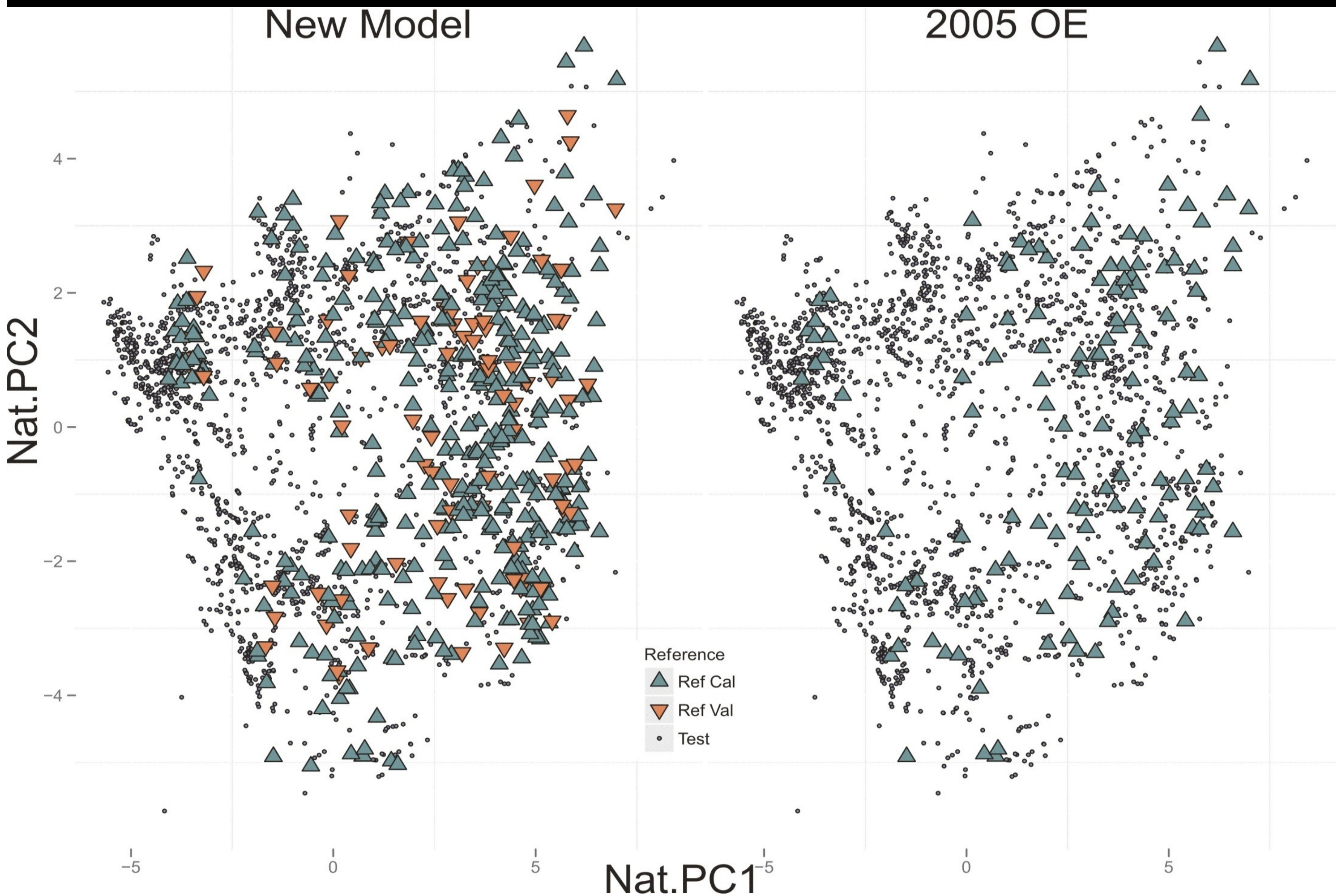
Multivariate view of natural diversity



Multivariate evaluation of representativeness



New reference pool fills in many gaps



Reference Conditions: Performance Summary

Stream Type Representation

- Overall excellent representation in most regions (absent in Central Valley, fewer in SoCal xeric region)
- Some under-representation of very low gradient, high conductivity, low elevation settings in Chaparral and South Coast

Biological Integrity

- Greatly reduced anthropogenic sources of variation in biological assemblages in reference pool

SWAMP
Surface Water Ambient Monitoring Program

Observed/ Expected Indices

Developed in UK (Wright and others 1970s-1980s, RIvPACS) – now widely used worldwide

Species-based approach: Compare number of **observed** ("O") taxa to number of **expected** ("E") taxa

"Expected" taxa at a test site are modeled using predictive modeling techniques

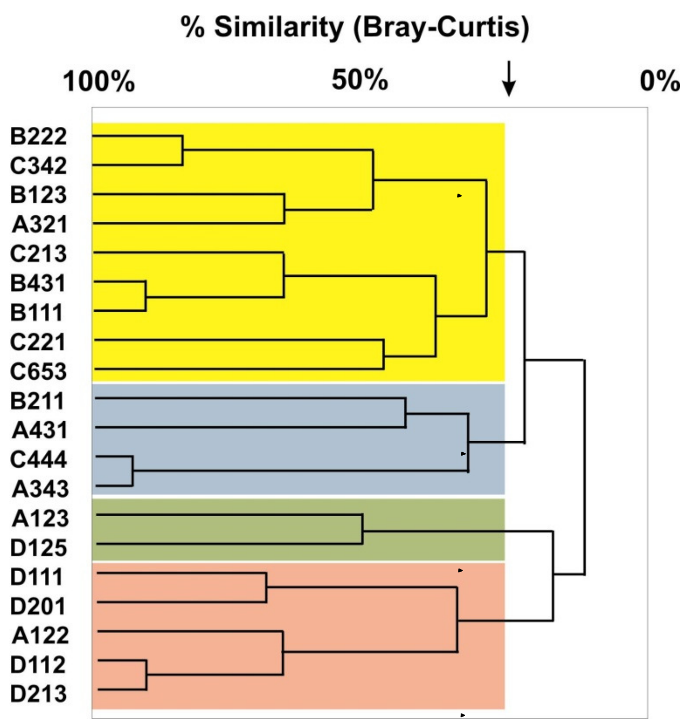
Compare test site to subsets of the reference sites that are physically similar to the test site *(geology, climate, elevation, latitude, etc.)*

Index score is a direct measure of taxonomic loss

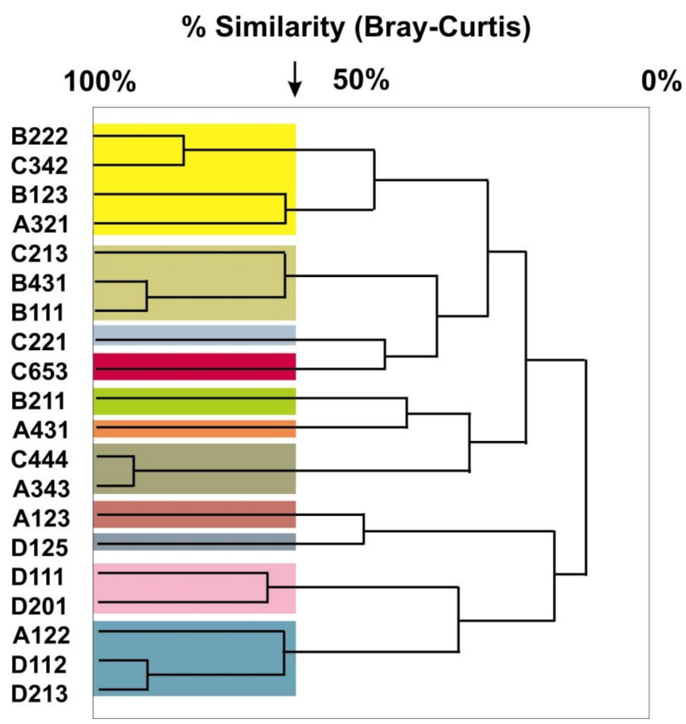
Estimating “E”: Step 1

Group reference sites based on biological similarity

Clustering techniques used to identify groups of reference sites with similar species composition



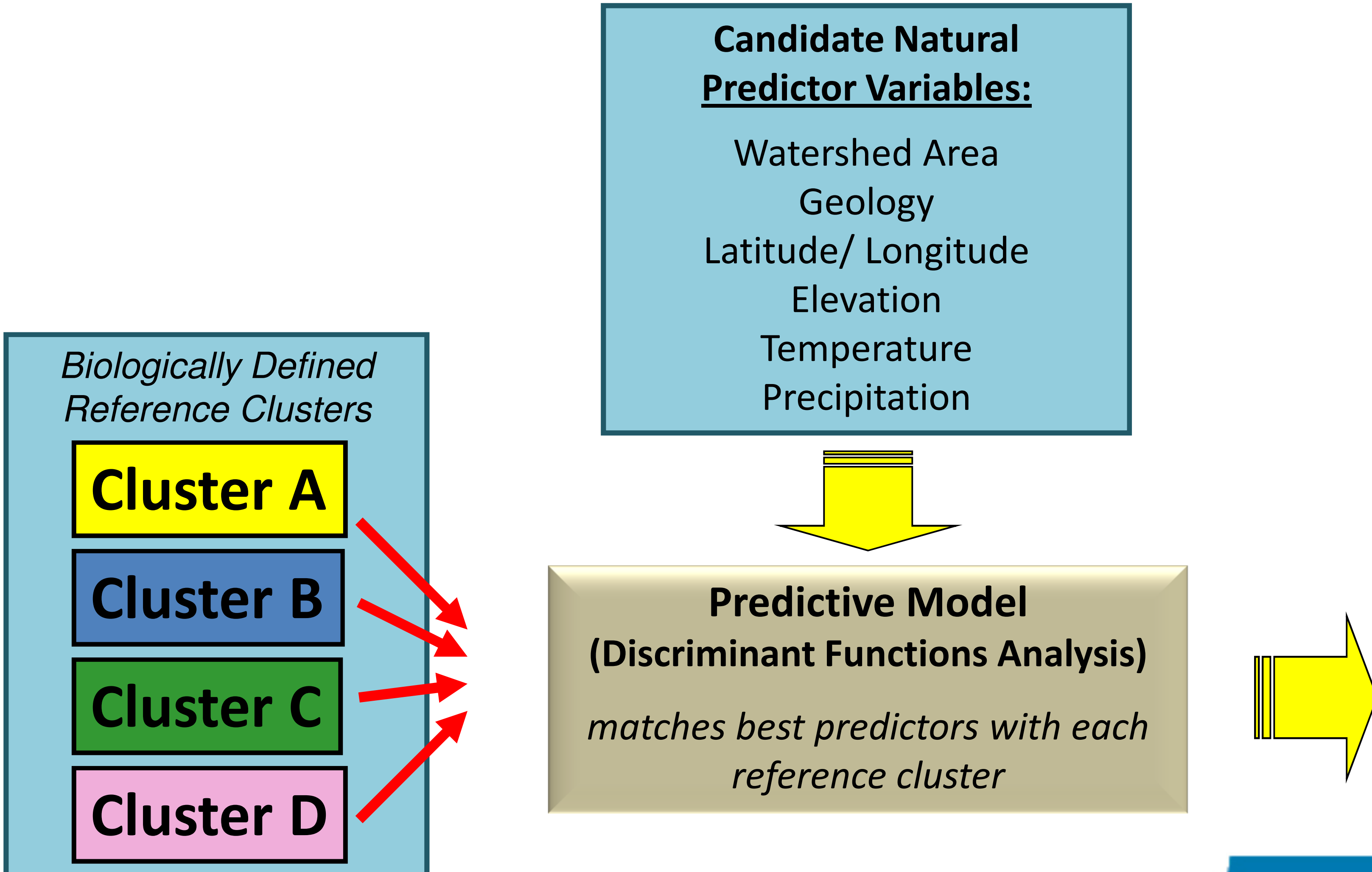
4 classes



11 classes

Estimating "E": Step 2

Develop model that will predict cluster membership for new sites



Estimating "E": Step 3

Estimate capture probabilities

Use discriminant model output + frequencies of occurrence within each class to estimate **probabilities of capture (PC)** for each taxon at a given site

Predictor Values at Test Site

Predictive Model
(matches predictors with each reference class)

Cluster	Site's probability of cluster membership	Frequency of species X (*Kogotus sp.*) in cluster	Expected contribution to PC
A	0.5	0.6	0.30
B	0.4	0.2	0.08
C	0.1	0.0	0.00
D	0.0	0.0	0.00
Probability of ***Kogotus sp.*** being in sample if site is in reference condition			**0.38**

Estimating "E": Step 4

Sum taxon occurrence probabilities estimate the number of native taxa (E) that should be observed (O)

Taxon	pc	O
Atherix	0.70	*
Baetis	0.92	*
Caenis	0.86	
Drunella	0.63	
Epeorus	0.51	*
Kogotus	**0.38**	
Gyrinus	0.07	
Hyalella	0.00	*
Count	4.07	3

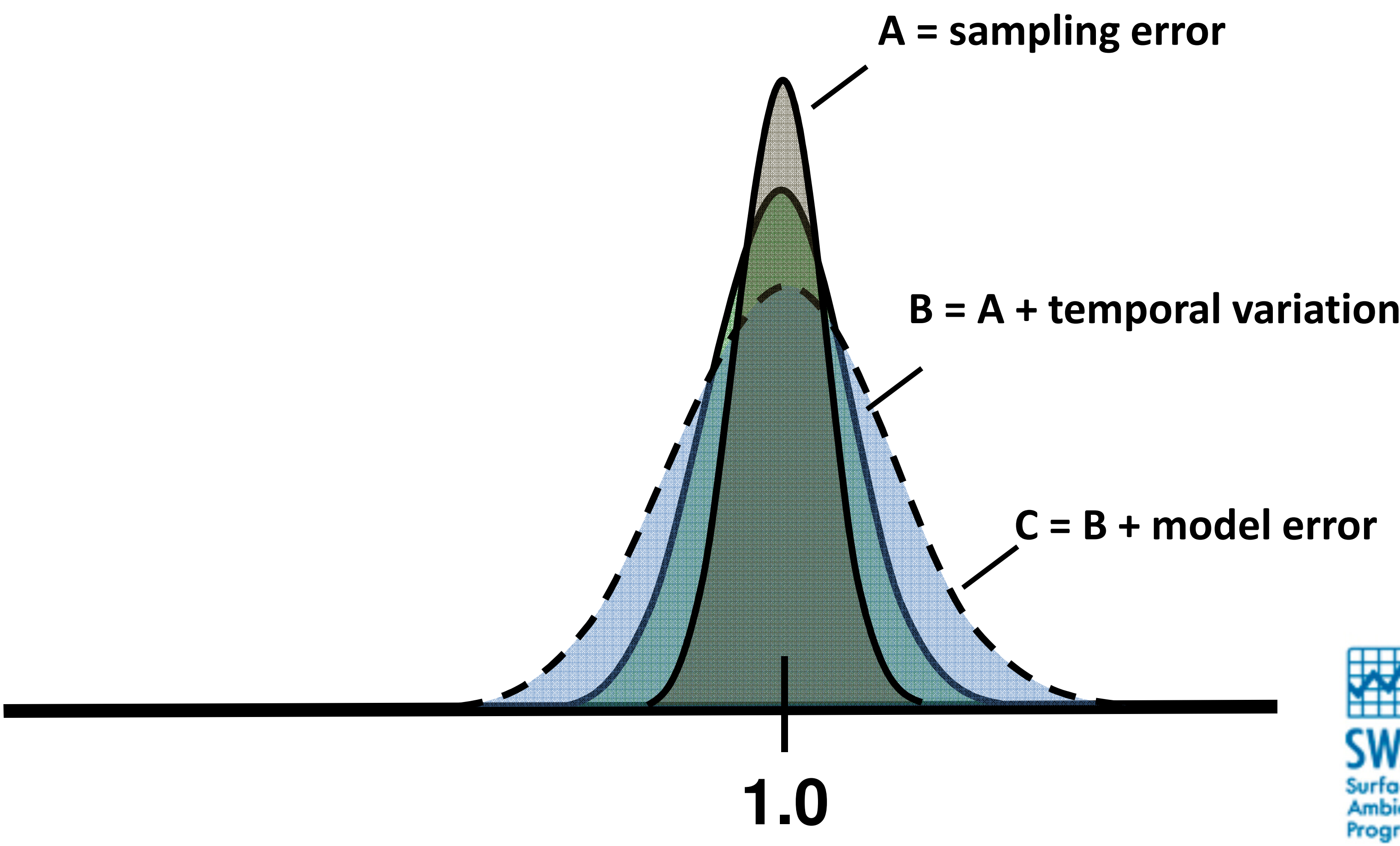
O/E = 3 / 4.07

O/E = 0.74

O/E Score
Indicates proportion of *native* assemblage present at test site

SWAMP
Surface Water Ambient Monitoring Program

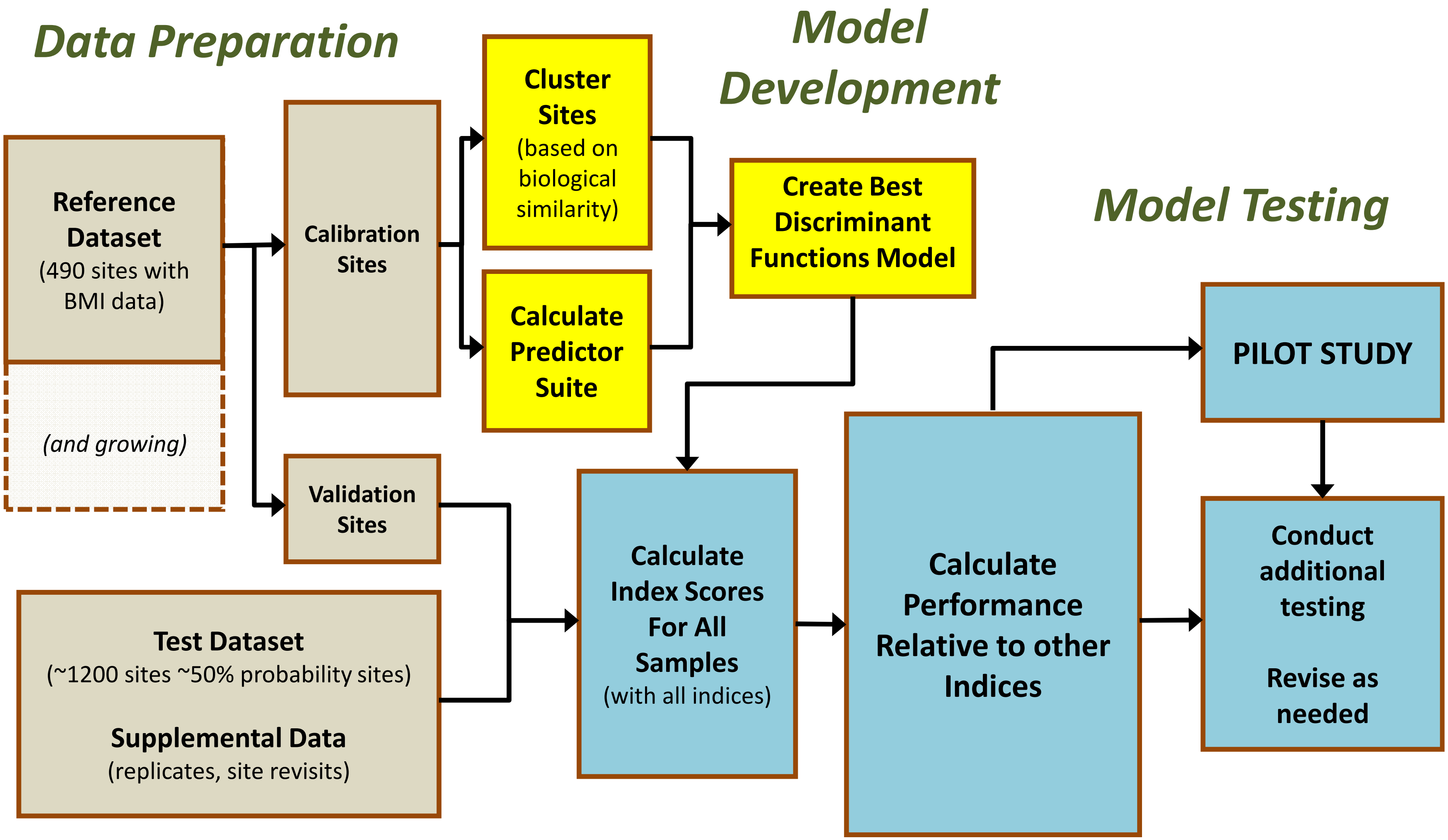
Sources of variation in O/E scores



(after Hawkins et al. 2010)

Surface Water Ambient Monitoring Program

O/E Index Development Process

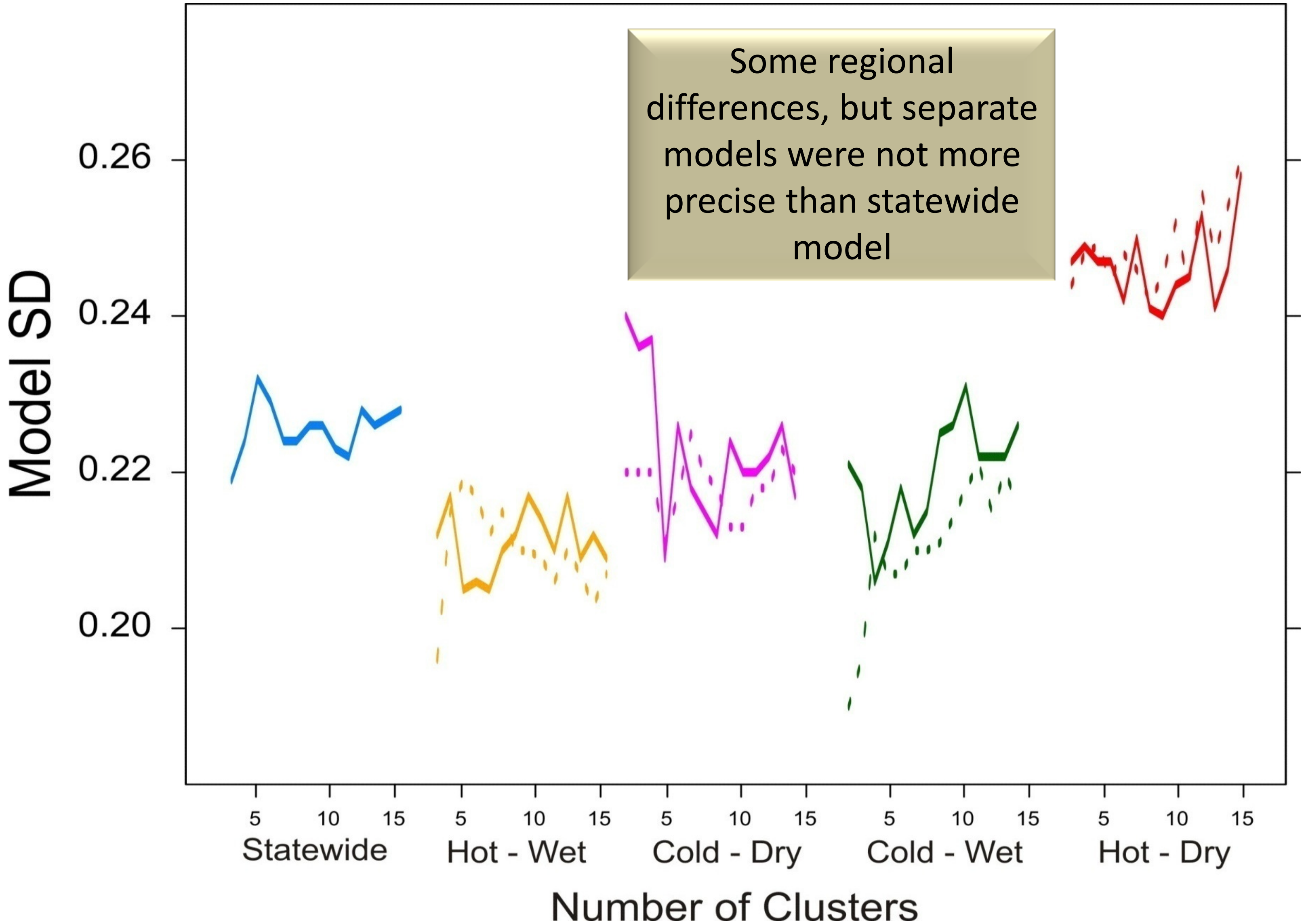


Subjects of Preliminary Exploration

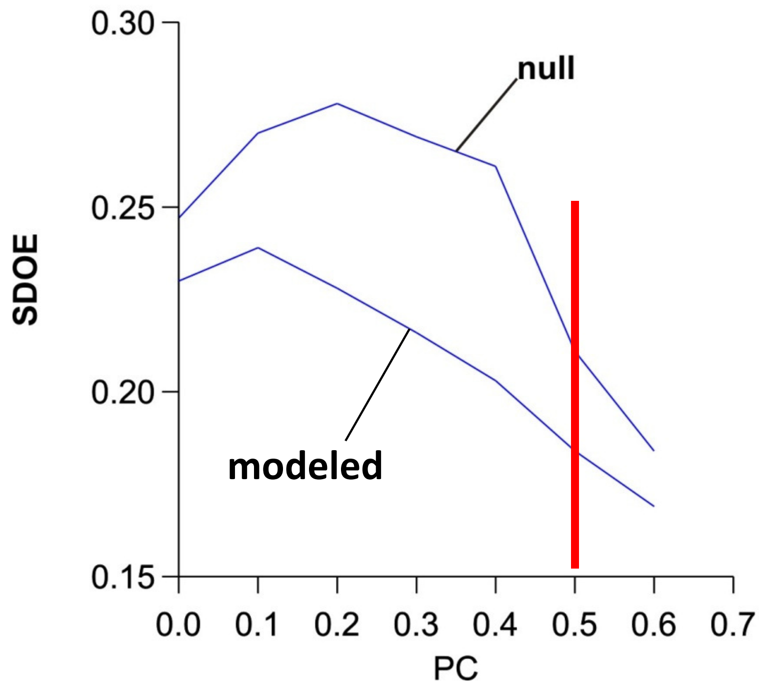
- **Climatic sub-models** – subdivisions of CA helped in 2005 model, but were no better than state model this time
- **Cluster size** – little to no impact on model performance
- **Probabilities of capture** – degree of inclusion of rare taxa: settled on a common threshold of $p > 0.5$ to reduce noise
- **Recent climate predictors** – last year's Temp and PPT had little predictive value in models
- **Bray-Curtis Index** – alternate to O/E using the B-C measure of dissimilarity: good responsiveness, but low precision
- **Null models** – no clusters, test site compared to all others
- **Evenness correction –** reduces confounding effect on richness

Each climate group modeled separately

Sub-group summaries using statewide model



Sensitivity Analysis: probability of capture



Fine Tuning the Scoring Tools

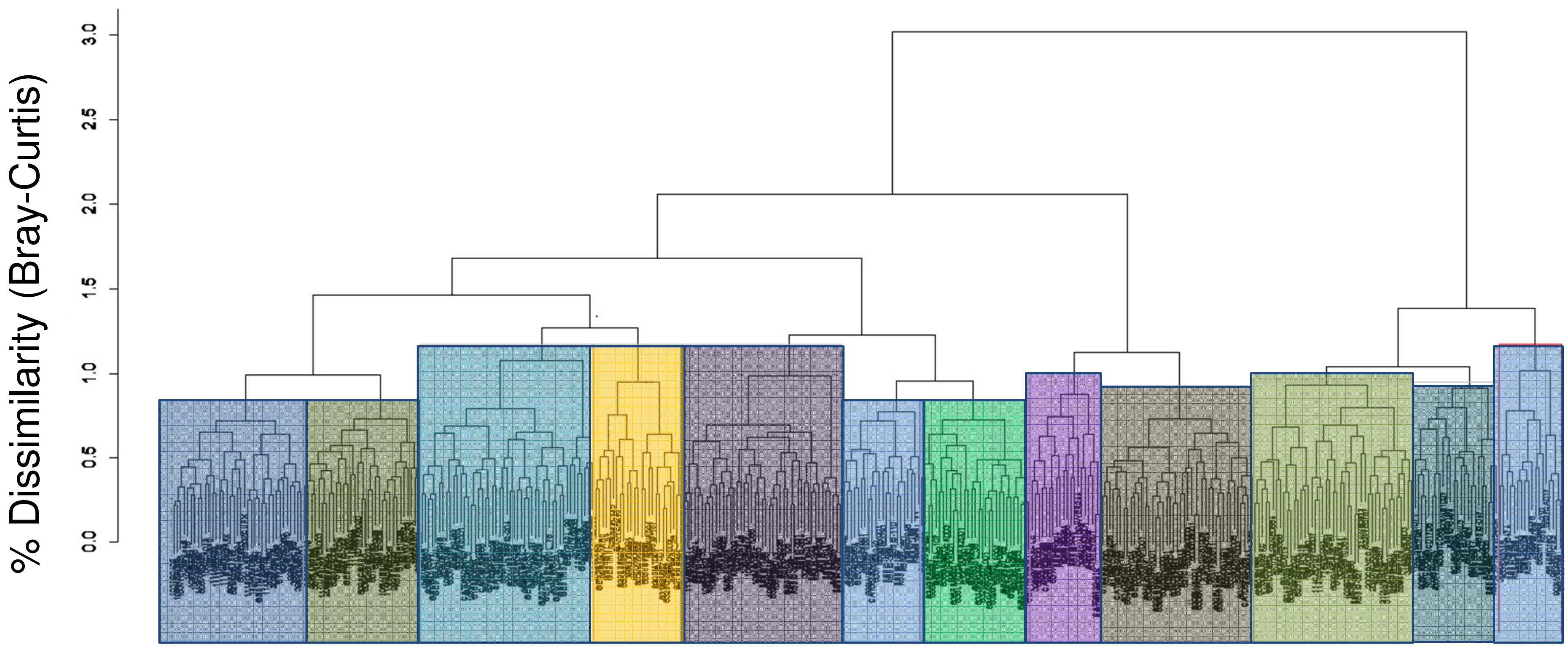
- We are now exploring variations on a set of related O/E indices
- All are likely to perform well
- Some variation in performance measures and implications for setting impairment thresholds

Data Preparation & Initial Decisions

- 615 reference sites identified in reference task
- Taxonomic effort standardized to SAFIT I (a): *mostly genus level IDs, with Diptera: Chironomidae to subfamily*
- Subsample to standard 400 individual count
- 490 sites were suitable for modeling *(i.e., had sufficient BMI counts after removing ambiguous taxa)*
- Prepare 34 **natural** predictor variables
- Split dataset into **calibration** and **validation** sets (80:20, 392 sites in calibration set)

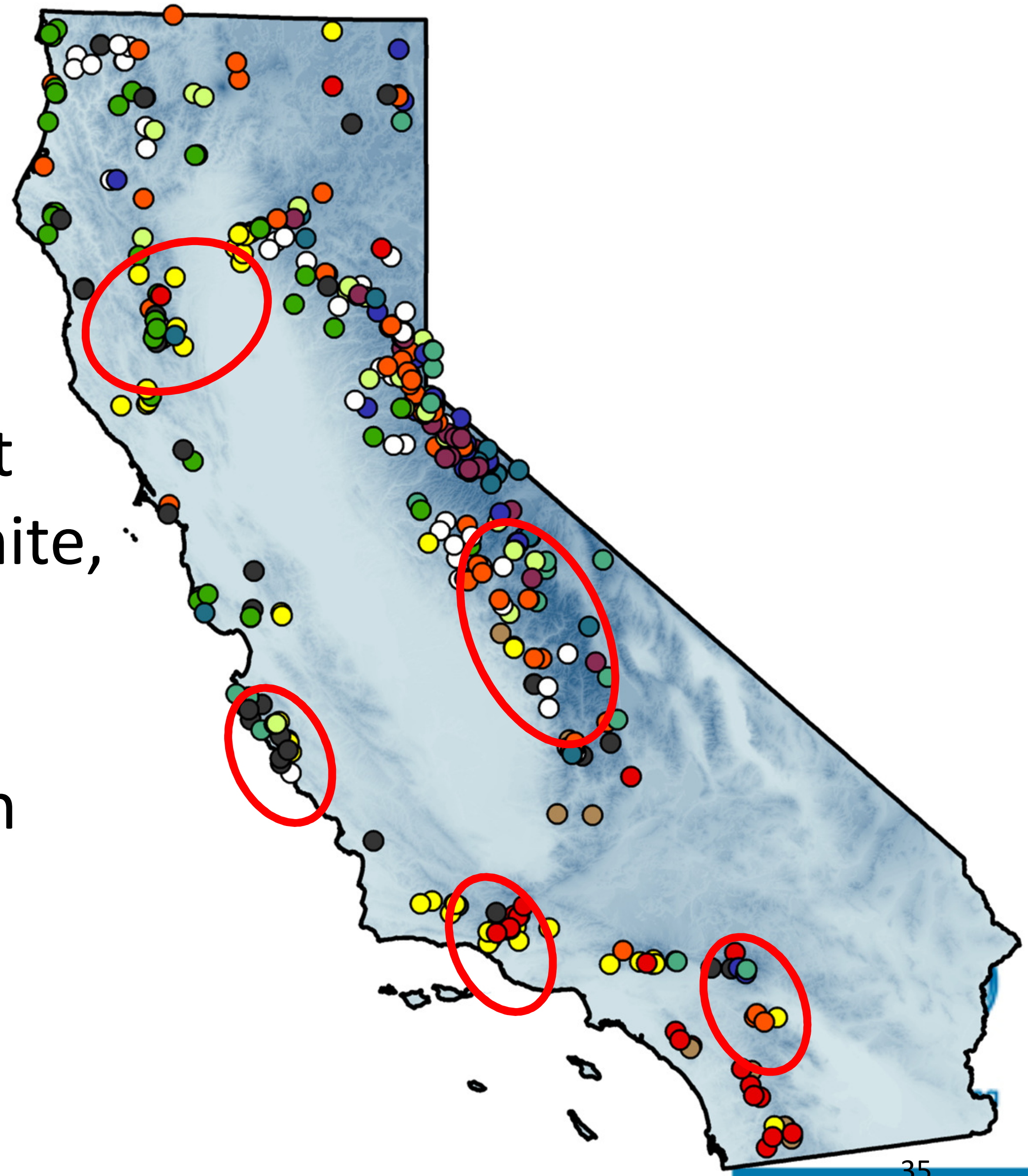
Cluster biological similarity

(Bray-Curtis dissimilarity, flexible-β = -0.25, rare taxa removed if < 5% of sites)



12 biological clusters

- Several large, geographically coherent clusters (e.g., green, white, purple)
- Several pockets of high variability



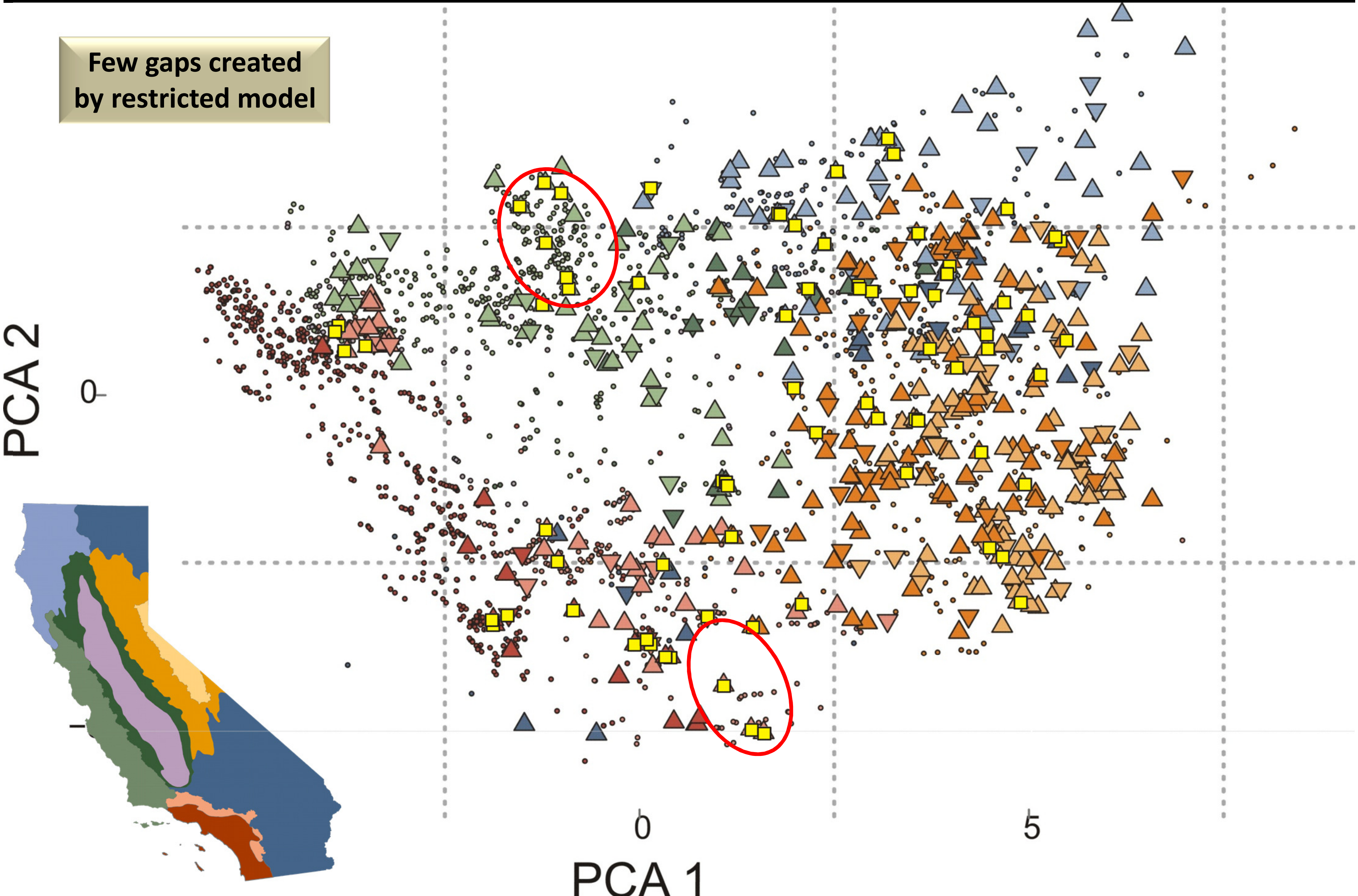
Discriminant Functions Models

- Examined all possible subsets of DFA models using 10 predictors (winnowed from 34)
- Best model had 5 predictors. More predictors did not improve model performance
- Added a second model with tighter reference criteria to evaluate sensitivity

“Restricted” tightened a few sensitive variables

Metric	2011 Bio-objectives	South Coast IBI (5k,ws)	North Coast IBI (1k, ws)	Current O/Es (Hawkins 2005)	Restricted Model (1k, 5k, ws)
Local Disturbance (W1_Hall)	**1.5**	-	-	riparian vegetation, erosion, grazing, etc.	**1.0**
% Agricultural	**3,3,10**	**5**	5		**3**
% Urban	**3,3,10**	**3**	3		**3**
% Ag + Urban	**5,5,10**				**5**
% Code 21	**7,7,10**	**in urban**	**in urban**		**5**
Road Dens (km/km^2)	**1.5**	**2.0**	**1.5/ 2.0**		**1.5**
Paved road x-ings (#/ws)	**5/10/50**				**5/10/50**
TN, TP (mg/L)	**3.0/ 0.5**	-	-		**3.0/ 0.5**
Nearest Dams	**>10 km**	-	-		**>10 km**
Active Producing Mines	**0** (5k)	-	-		**0** (5k)
% Canals & Pipelines	**10**	-	-		**10**
Gravel Mine Density	**0.1** (r5k)				**0.1** (r5k)
Conductivity	**<2000 uS, + <99%, >1%**				**<2000 uS, + <99%, >1%**
BPJ Screen	**X**	**X**	**X**	**X**	**X**

82 fewer sites in “Restricted” set



2 new O/E indices

Evaluate sensitivity to somewhat stricter reference criteria

Feature	"New" model	"Restricted" model
Clusters	12	9
Sites	490 (392 Cal/ 98 Val)	408 (325 Cal /83 Val)
Reference Screens		tighter standards for roads, Ag, Urban, riparian disturbance
Model Predictors	**Elevation**	**Elevation**
	log Watershed Area	**log Watershed Area**
	log Predicted Conductivity	**log Predicted Conductivity**
	average Temp (2000-2009)	**average Temp (2000-2009)**
	average PPT (2000-2009)	

- All predictors are GIS based
- Climate data from PRISM Climate Center
- Conductivity predictions from Utah State (John Olson and Chuck Hawkins)

Scoring Tool Performance Measures

1. Applicability – the extent of the stream population that can be scored accurately with the index
2. Precision – variability of scores for sites considered to be in similar condition (e.g., reference sites)
3. Accuracy – proximity of score to “true” condition
4. Responsiveness – ability to discriminate impaired sites and sensitivity to gradients of stress
5. Repeatability – similarity of scores for repeated measurements

SWAMP
Surface Water Ambient Monitoring Program

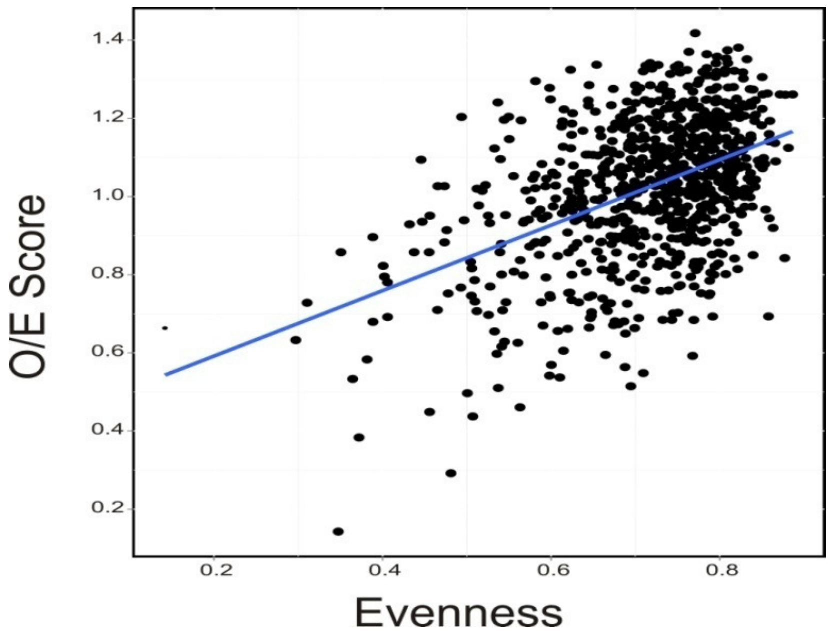
Indices used in comparisons

Name	Description
O/E	O/E index (modeled with 5 predictors)
***O/E_ec**	**O/E index with evenness correction**
O/E_null	O/E index with no predictors (null model)
O/E_null_ec	O/E null model with evenness correction
r_O/E	restricted O/E index (modeled with 4 predictors)
r_O/E_ec	restricted O/E index with evenness correction
O/E (2005)	**2005 O/E index (Chuck Hawkins, 3 submodels)**
NCIBI	**North Coast IBI**
SCIBI	**South Coast IBI**

SWAMP Surface Water Ambient Monitoring Program

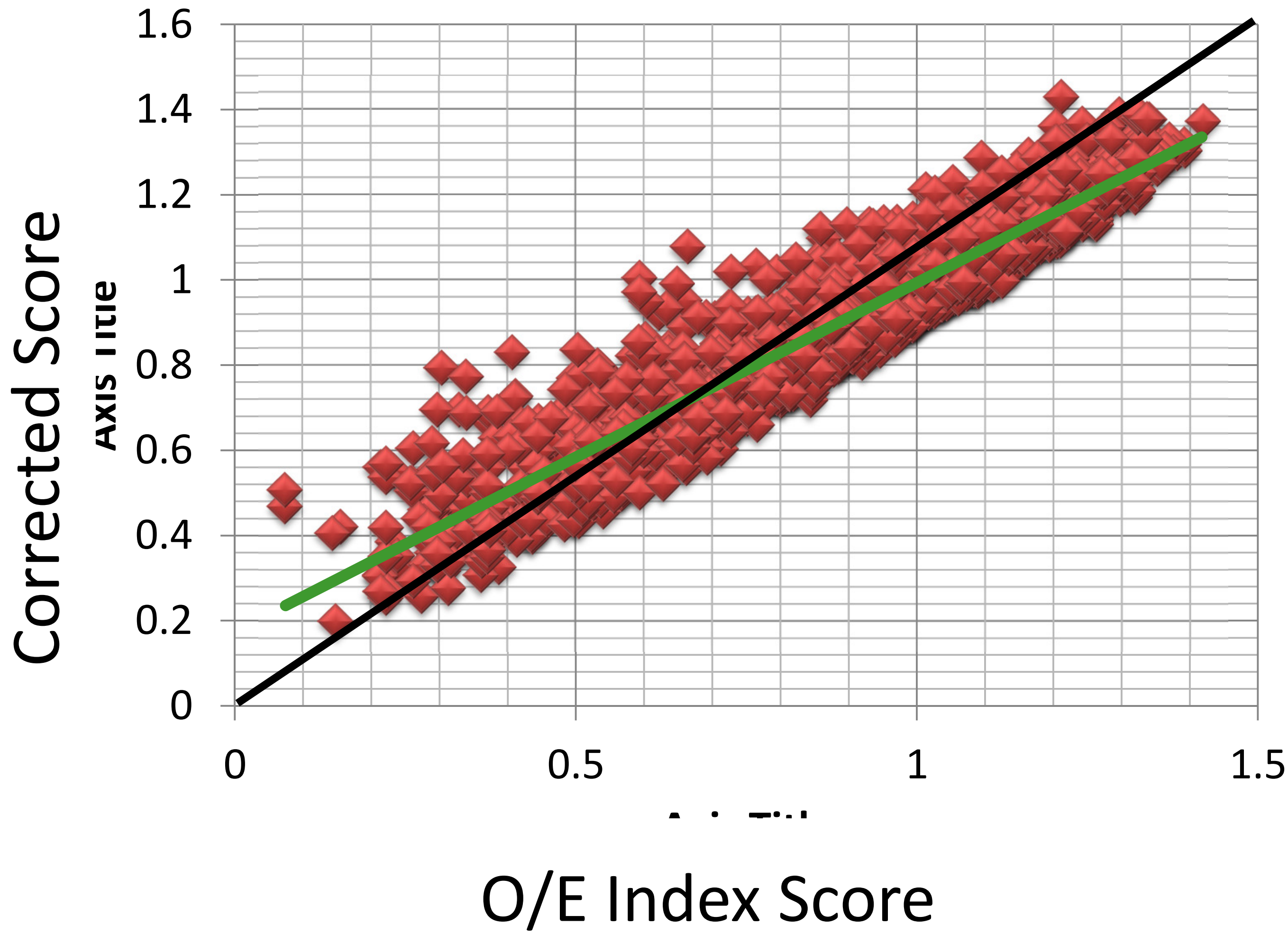
Why an evenness correction?

- Samples with low evenness can impair our ability to accurately predict richness (a big deal for O/E models)
- Correction minimizes confounding effects of evenness

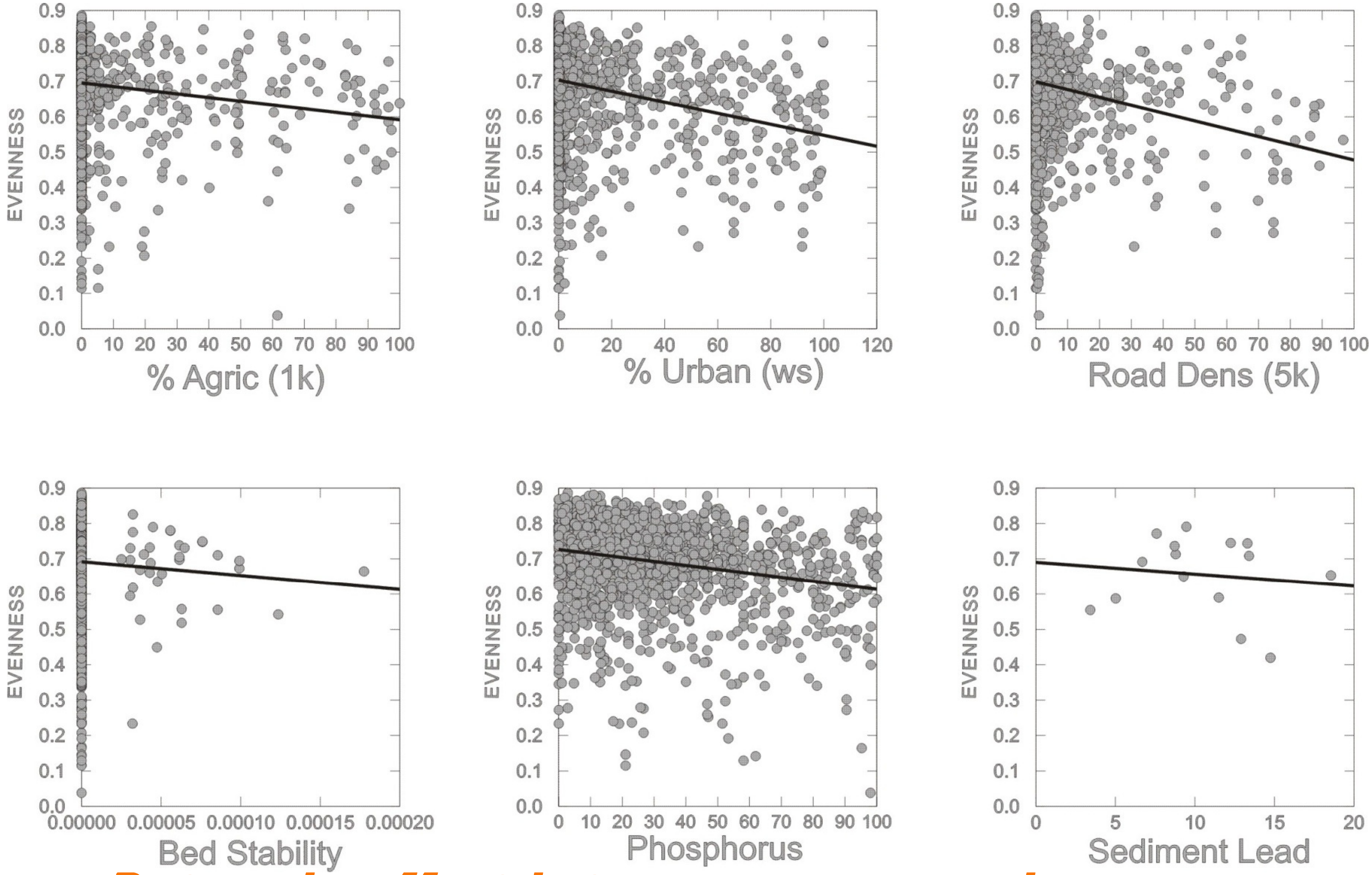


Taxon	Sample 1	Sample 2
Atherix	10	3
Baetis	11	90
Caenis	12	2
Drunella	9	1
Epeorus	15	1
Kogotus	13	1
Gyrinus	21	1
Hyalella	9	1
Count	100	100
Richness	9	9

Evenness correction is biased at low O/E scores

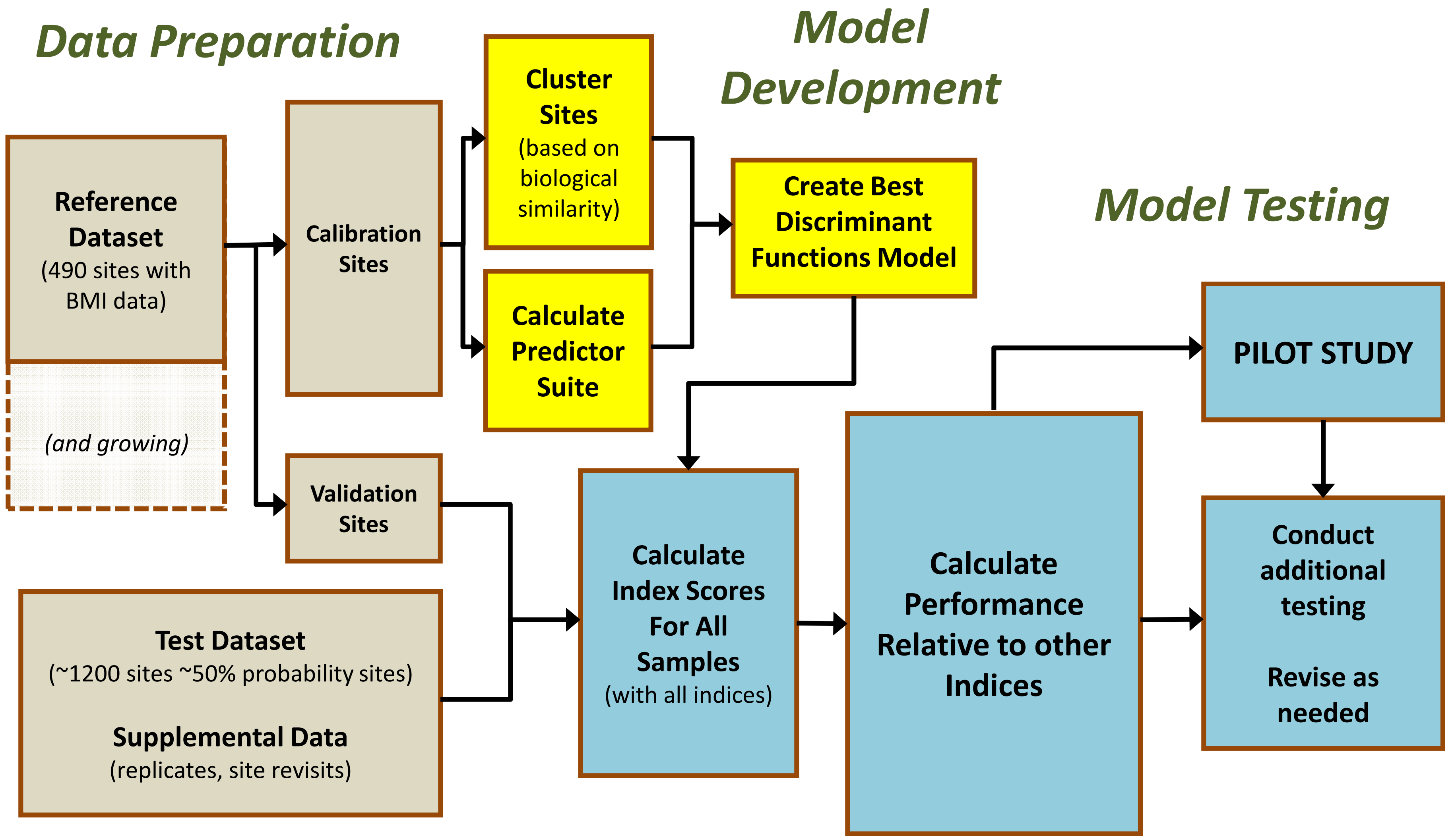


Evenness was weakly related to impairment



But main effect is to compress scoring range

O/E Index Development Process



Performance Highlights

- Compare variants of new scoring tools
  - "new" O/E vs. "restricted" O/E
  - Clustering vs. no clustering
  - Evenness correction vs. no correction
- Compare 2 new indices with existing scoring tools
  - "Current" O/Es (Hawkins 2005, 3 submodels)
  - SoCal IBI, NorCal IBI

SWAMP
Surface Water Ambient Monitoring Program

Applicability

The extent of the stream population that can be scored accurately with an index

Why do we care? Objective test of whether the environmental setting of a given site meets the conditions for scoring with an index

How do we measure it?

- Range test: are test sites within range of reference predictors? (e.g., elevation, watershed area, etc.)
- Distance (in multi-dimensional space) of a test site to the nearest reference cluster

SWAMP Surface Water Ambient Monitoring Program

Precision

variability of scores for sites considered to be in similar condition (e.g., reference sites)

Why do we care?

- Used to establish impairment thresholds (smaller SD means easier to detect deviation from reference)
- Determines how large a difference the index can detect

How do we measure it?

- Standard deviation of reference sites
- Replicate scoring consistency

SWAMP Surface Water Ambient Monitoring Program

Precision

standard deviation of reference sites (validation set)

- Modeled indices are more precise than null indices
- Evenness-corrected indices are a little more precise than uncorrected indices
- 2 new indices have comparable precision

Model	SD
O/E	0.18
***O/E_ec**	0.16
O/E_null	0.21
O/E_null_ec	0.19
r_O/E	0.17
***r_O/E_ec**	0.17

SWAMP
Surface Water Ambient Monitoring Program

Responsiveness/ Sensitivity

ability to discriminate impaired sites and sensitivity to gradients of stress

Why do we care?

- Assures that index can detect difference from expected conditions and is responsive across a gradient of stress

How do we measure it?

- Relative strength of discrimination between reference and test sites
- Strength of relationship between index score and gradients of stress

SWAMP
Surface Water Ambient Monitoring Program

Responsiveness:

discrimination between reference and test sites

- Modeled indices are more responsive than null indices
- 2 new O/E indices are equivalent
- Evenness corrected variants are equivalent to uncorrected indices

Model	t-value
O/E	17.6
***O/E_ec**	17.5
O/E_null	12.8
O/E_null_ec	12.1
r_O/E	17.0
***r_O/E_ec**	16.9

SWAMP Surface Water Ambient Monitoring Program

Responsiveness/Sensitivity

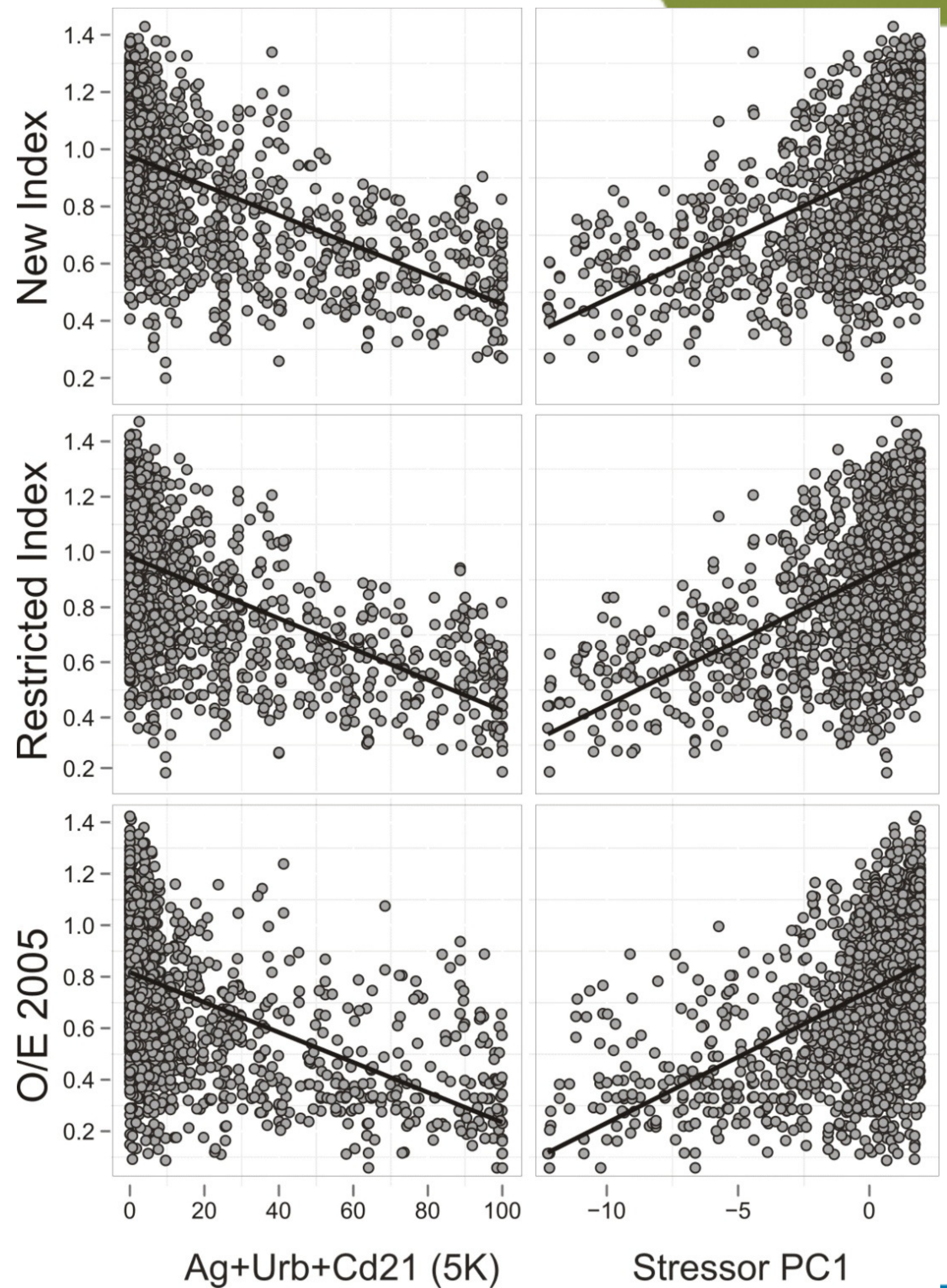
sensitivity to stressor gradients

- **Scores vs. stressor Gradients**
  - Look for “wedge-relationships” (absence of high scores at stressed sites)

- **Different types of gradients examined**
  - Proximate, mechanistic (metals, pyrethroids, ions)
  - Proximate, non-mechanistic (habitat, nutrients)
  - Ultimate (land cover)
  - Synthetic (PCA axes)

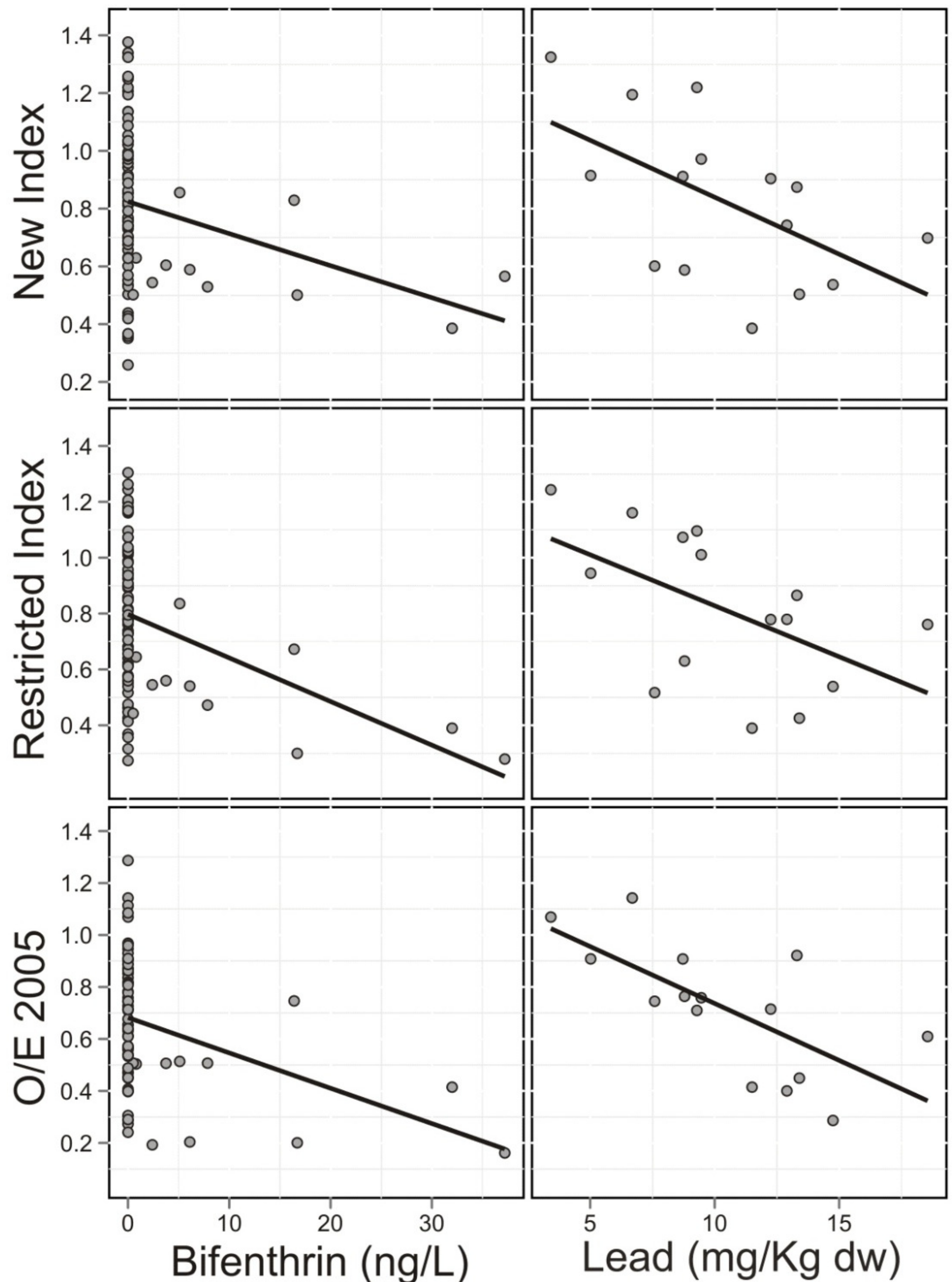
SWAMP
Surface Water Ambient Monitoring Program

Responsiveness of new and old indices to landuse development and a composite index of stress

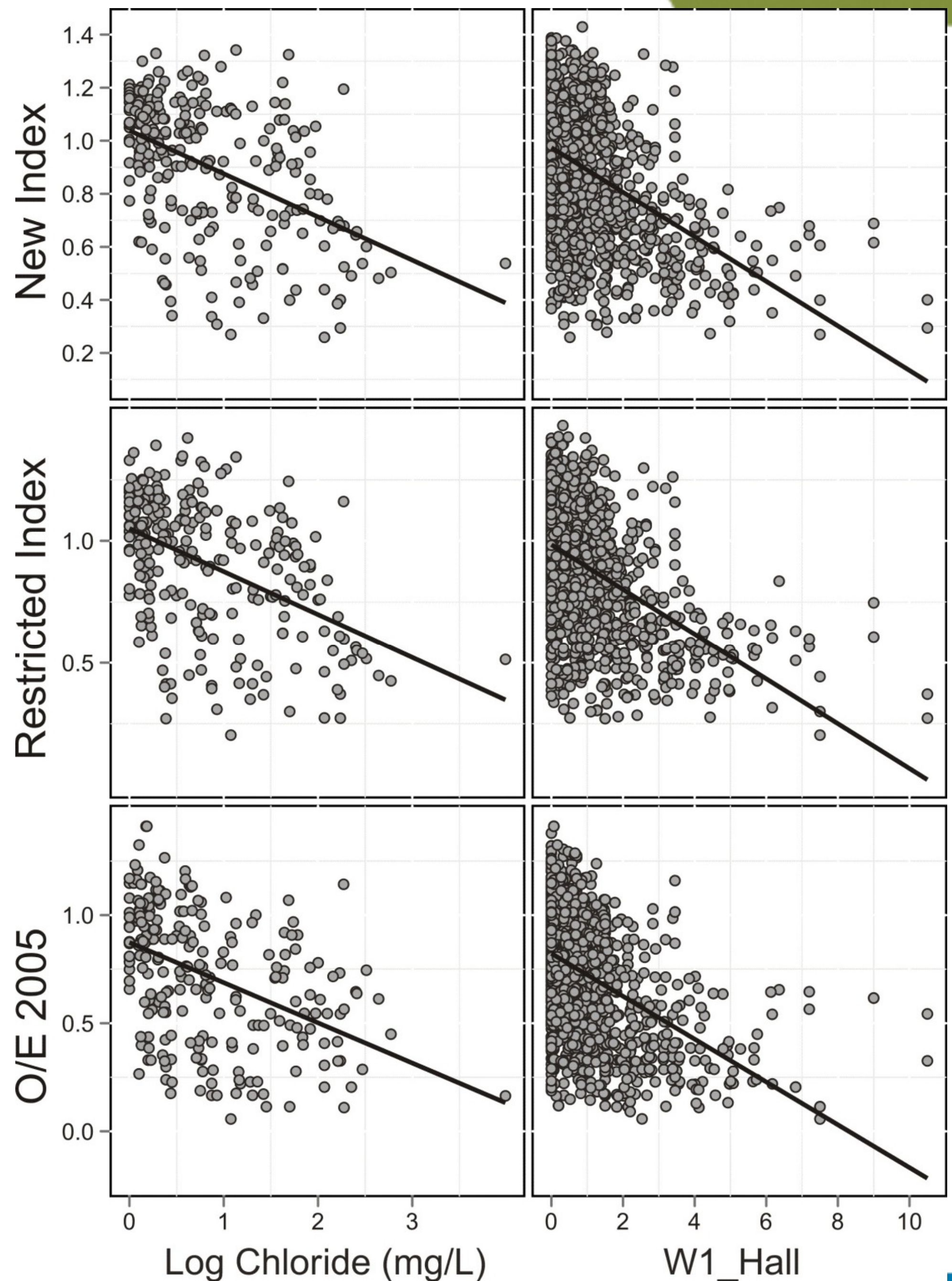


Responsiveness of new and old indices to sediment chemistry

(data only available for some sites in SoCal)



Responsiveness of new and old indices to chloride and riparian disturbance
(W1_Hall)



Accuracy

proximity of score to "true" condition

Why do we care?

Accurate indices give accurate condition assessments, but direct measures of "truth" are elusive

How do we measure it? *(indirectly, by looking for bias)*

- Compare scores at ref sites by region
- Compare scores at ref sites vs. natural gradients
- Estimate residual natural variance not explained by scoring tool

SWAMP
Surface Water Ambient Monitoring Program

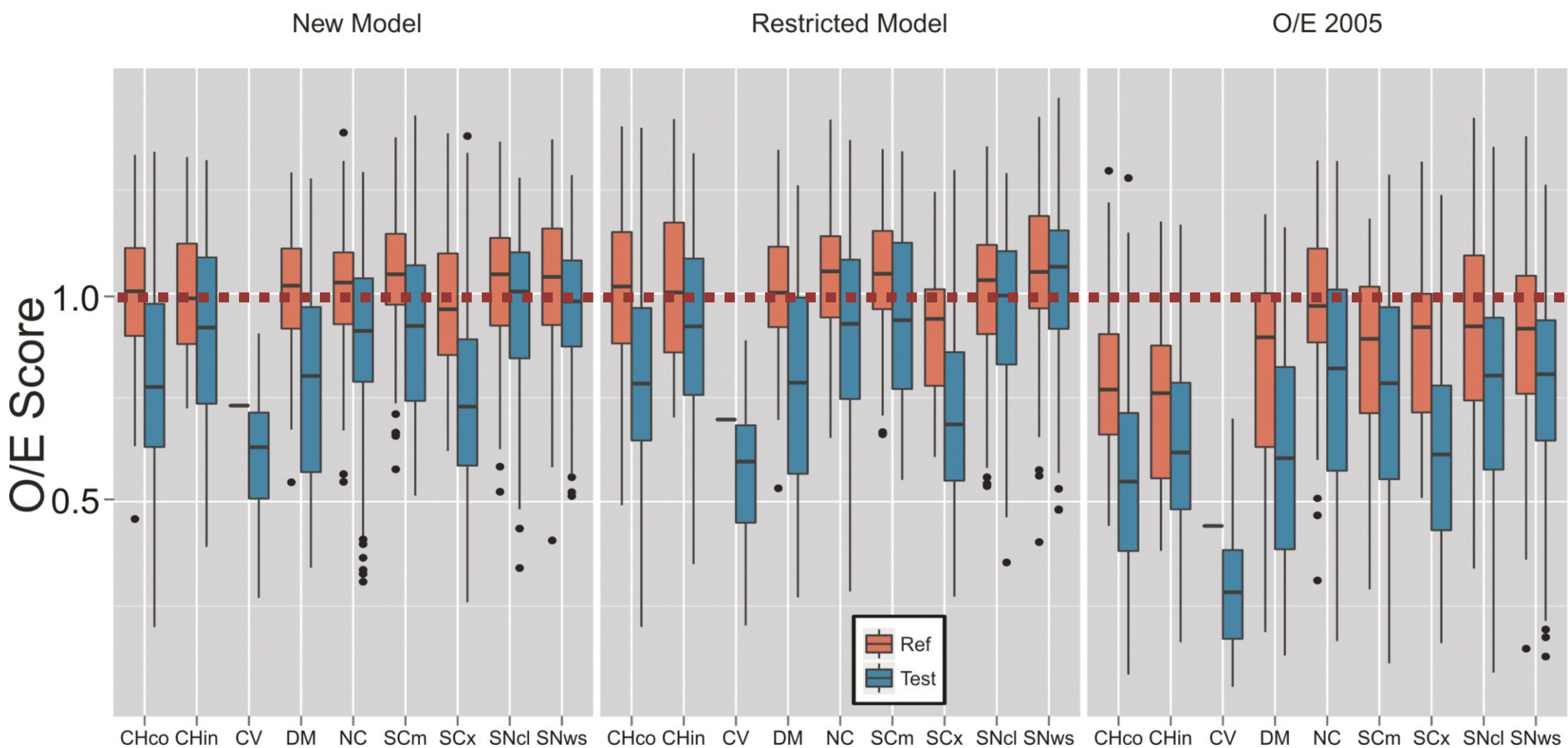
Regional consistency from a statewide index

- Restricted O/E index has more regional bias
- Evenness correction makes slight improvements

Comparisons with Current Tools

INDEX	Precision (sd or CV)			Accuracy (%)		Responsiveness (t-value)
	Reference Calibration	Reference Validation		Residual Natural Variance		Reference v. Test
O/E	0.18	0.18		25		17.6
O/E_ec	0.17	0.16		20		17.5
r_O/E	0.19	0.17		11		17.0
r_O/E_ec	0.18	0.17		9		16.9
O/E_2005	0.23	0.20		53		14.3
SoCal IBI	0.26	0.16		14		10.5
NorCal IBI	0.17	0.14		31		4.4

Old vs. New O/E Comparisons



- New models have little regional bias and are more precise
- Reference test discrimination is similar, but strong overall bias

Performance Summary

New indices:

- New indices are as good or better than earlier indices and generally comparable performance
  - Better precision
  - Better accuracy
  - Better discrimination of test – reference
  - New O/E scores higher than old O/E and IBIs
- Evenness correction and restricted versions have a mix of pros and cons

SWAMP
Surface Water Ambient Monitoring Program

What's Next

Optimization of scoring tool and exploring implications for different applications (e.g., influence of temporal variability, recent climate, effectiveness in different regions of the state)

Precision (consistency tests)

- Consistency of assessment at true replicates
- Long-term (inter- and intra-annual) consistency

Accuracy (bias)

- Explore sources and implications of differences between old and new scoring tools, including separation of natural and anthropogenic sources
- Explore effects of recent climate and temporal variability

Applicability

Questions for Science Panel

- What is your opinion of the assessment tool(s) developed so far?
  - How do our tools compare to other state/national efforts?
- What is your opinion of the performance measures we used?
  - What additional performance tests should we apply?
- Which O/E index would you recommend and why?
  - Should we apply evenness correction?
  - Should we regionalize scoring thresholds?
- Are there other assessment tools that we should explore (e.g., modeled MMIs)?

SWAMP Surface Water Ambient Monitoring Program

O/E Index Development Process