Scoring Biological Integrity

the California Stream Condition Index (CSCI)

Resources Agency California Department Fish & Game

Aquatic Bioassessment Laboratory

SCCWRP Established 1969

California Water Boards

SWAMP Surface Water Ambient Monitoring Program

- **Scoring Tool Enhancements**
  - Update to O/E component
  - Integrating predictive MMI techniques
  - Our recommendations
- **Setting Thresholds**
- **Statewide and Regional Extent Estimates**
- **Questions for the panel**

SWAMP
Surface Water Ambient Monitoring Program

Technical Team

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

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

Larry Brown, *USGS*

Jason May, *USGS*

David Herbst, *SNARL*

Peter Ode, *DFG-WPCL/ABL*

Ken Schiff, *SCCWRP*

David Gillett, *SCCWRP*

Eric Stein, *SCCWRP*

Betty Fetscher, *SCCWRP*

Kevin Lunde, *SF Water Board*

****** Scientific Review Panel***


How do we convert a list of species into a condition score?
NABS (www.benthos.org)

Qualities of a good scoring tool

Technical Qualities

- precise
- accurate
- responsive

Regulatory Qualities

- universally applicable
- easy to relate to ecological condition
- easy to compare to a standard

Two common approaches for quantifying biotic condition

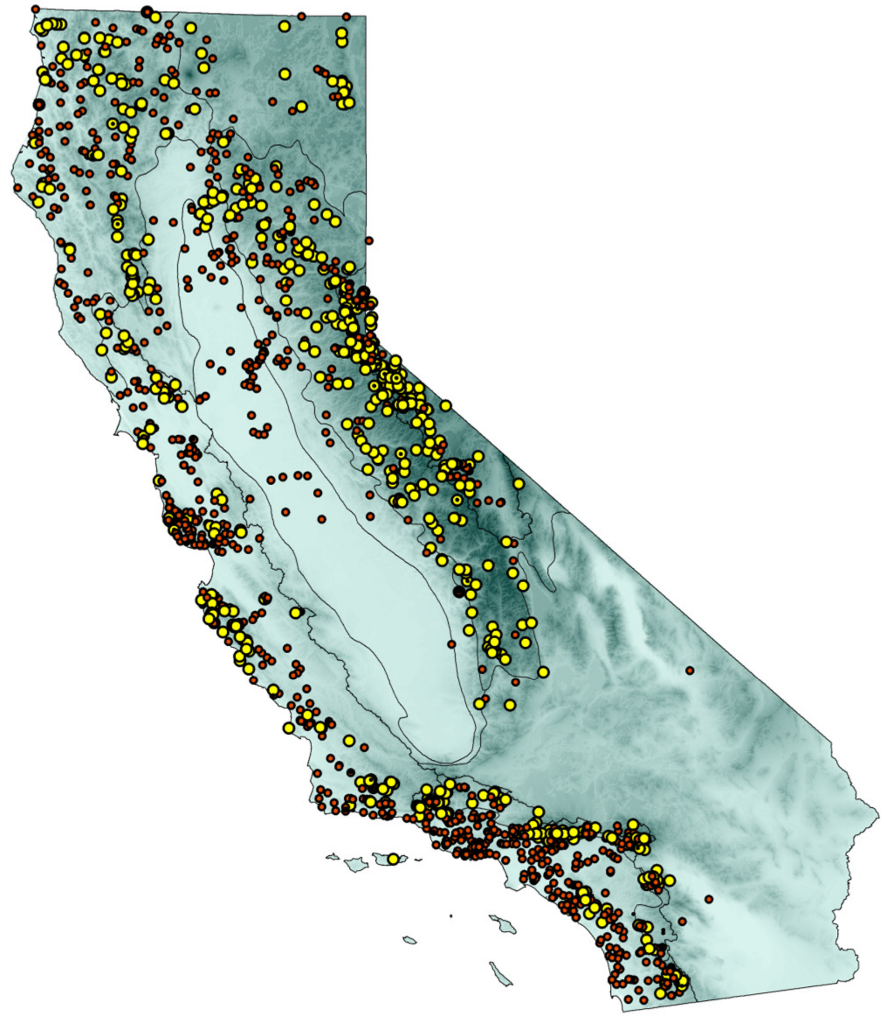
Species loss indices (e.g., O/E indices)

Ecological structure indices (e.g., multi-metric indices including IBIs)

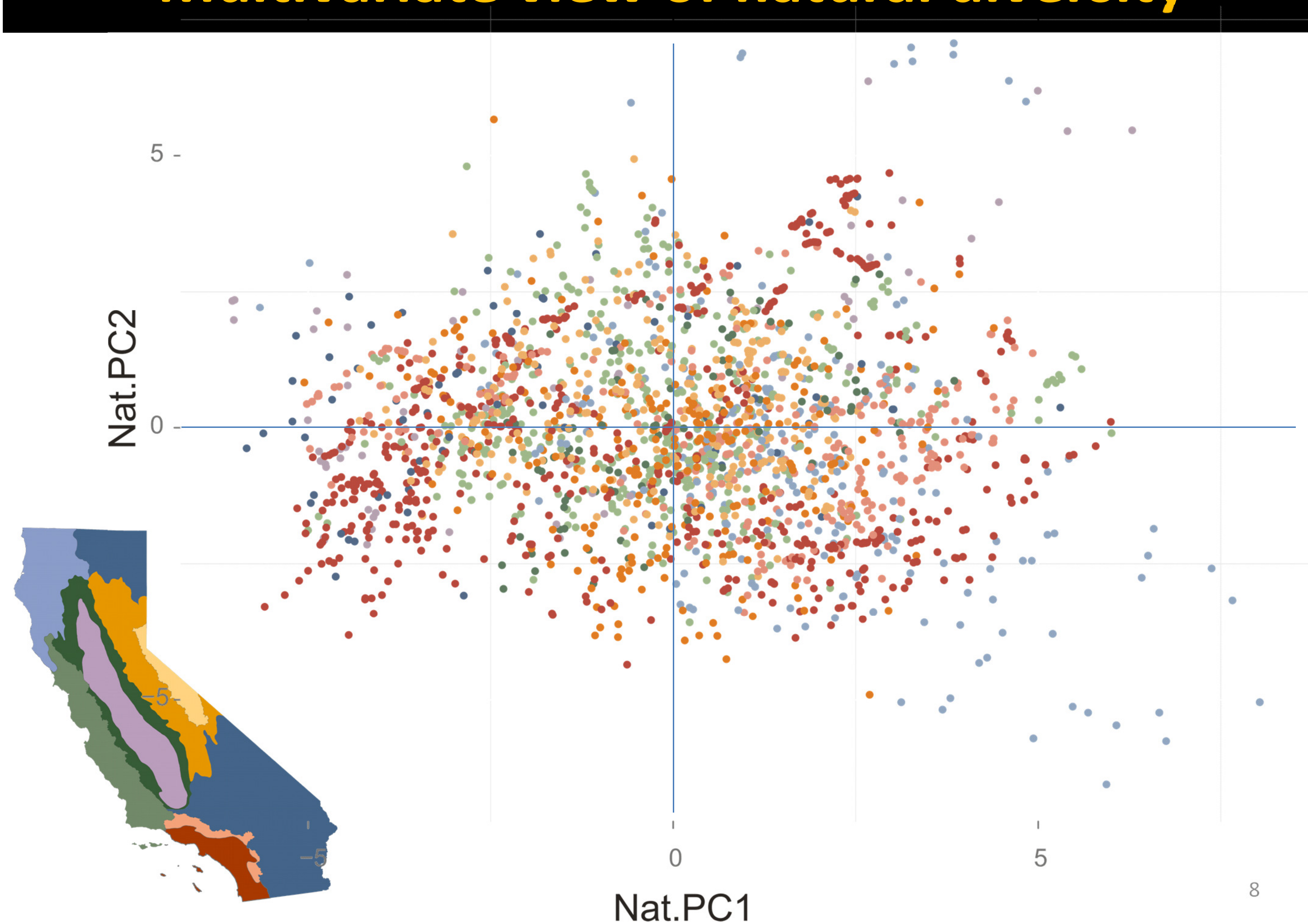
Scoring tools rely on reference sites to establish expected conditions

- 485 reference sites used to develop scoring models

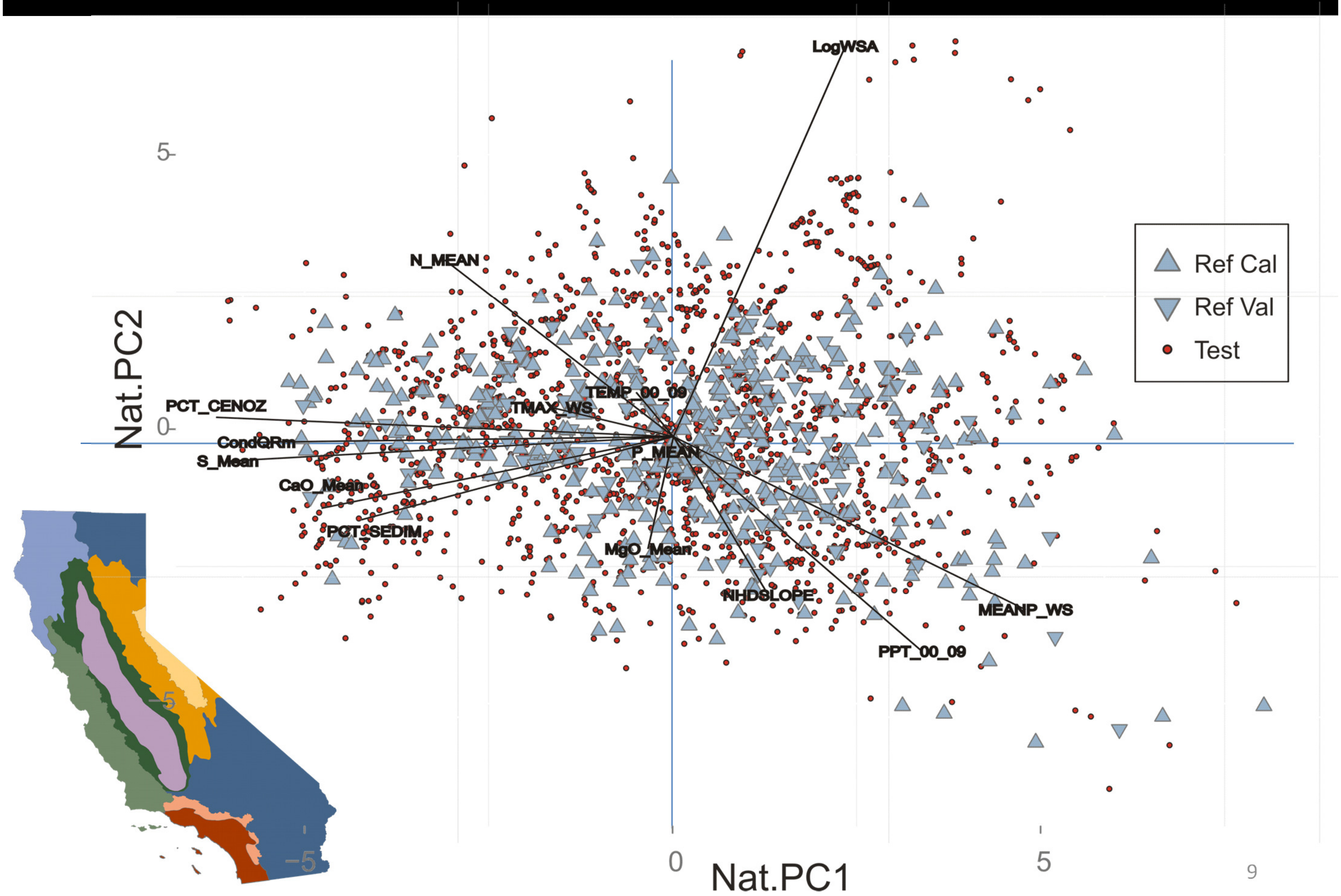
- Excellent coverage of CA's natural stream diversity



Multivariate view of natural diversity



Multivariate view of natural diversity



Species Loss Index (O/E)

Compare number of **observed** ("O") taxa to number of **expected** ("E") taxa

Step 1. Cluster reference sites based on biological similarity

Step 2. Identify natural gradients that best explain clusters (=predictors)

Step 3. Use predictor values at test sites to predict species expected to be observed

Index score is an estimate of taxonomic loss

O/E Update

- April index performed well
- **Reference pool adjustments:**
  - added sites to target under-represented gradients
  - dropped sites based on stakeholder feedback
- New discriminant functions model was not as precise as the April model
- Experimented with climatic sub-models, random forest techniques, predictor selection

O/E Update

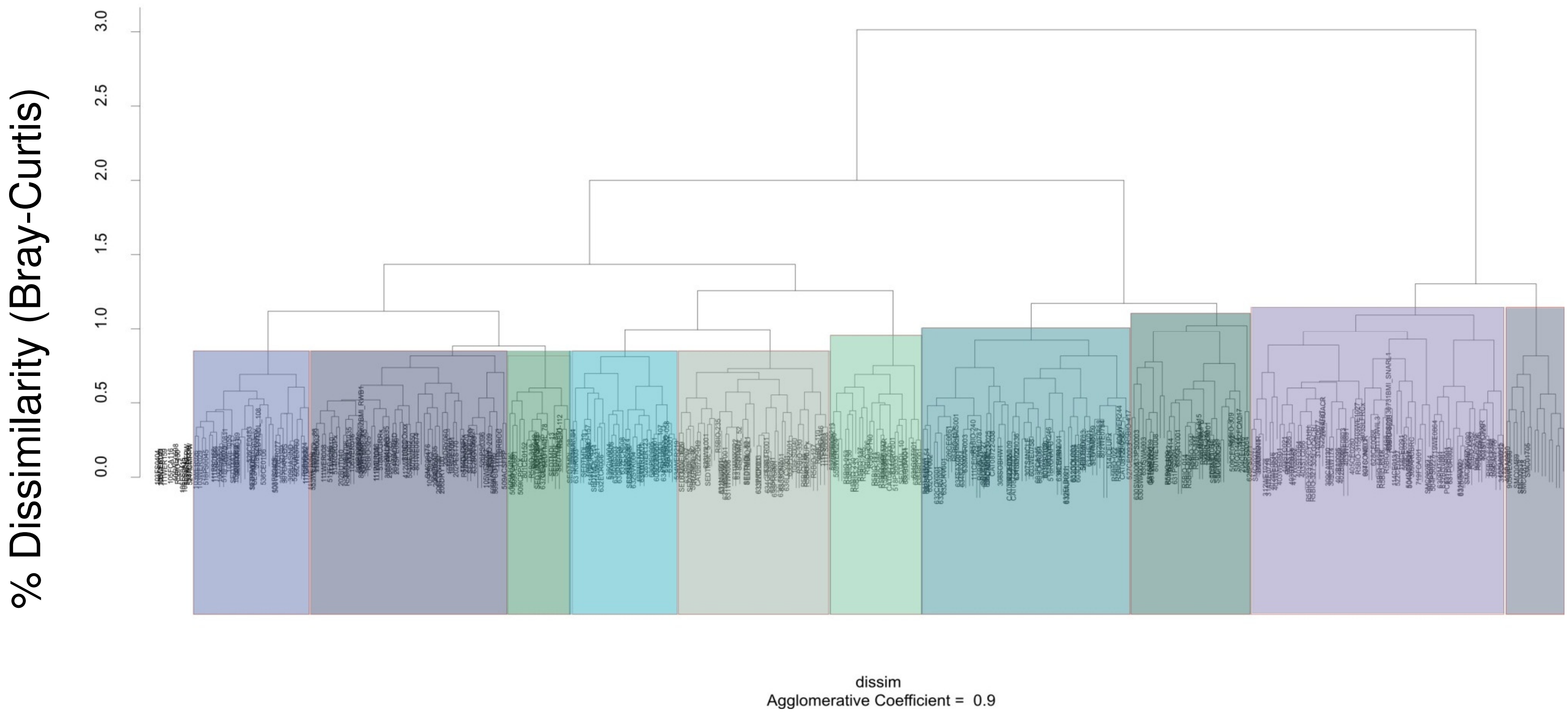
Final Model (Random Forests, 10 clusters, 4 predictors):

- Average Monthly Temperature (2000-2009)
- Average Monthly Precipitation (2000-2009)
- Log Watershed Area
- Site Elevation

Performance was very similar to our April O/E index

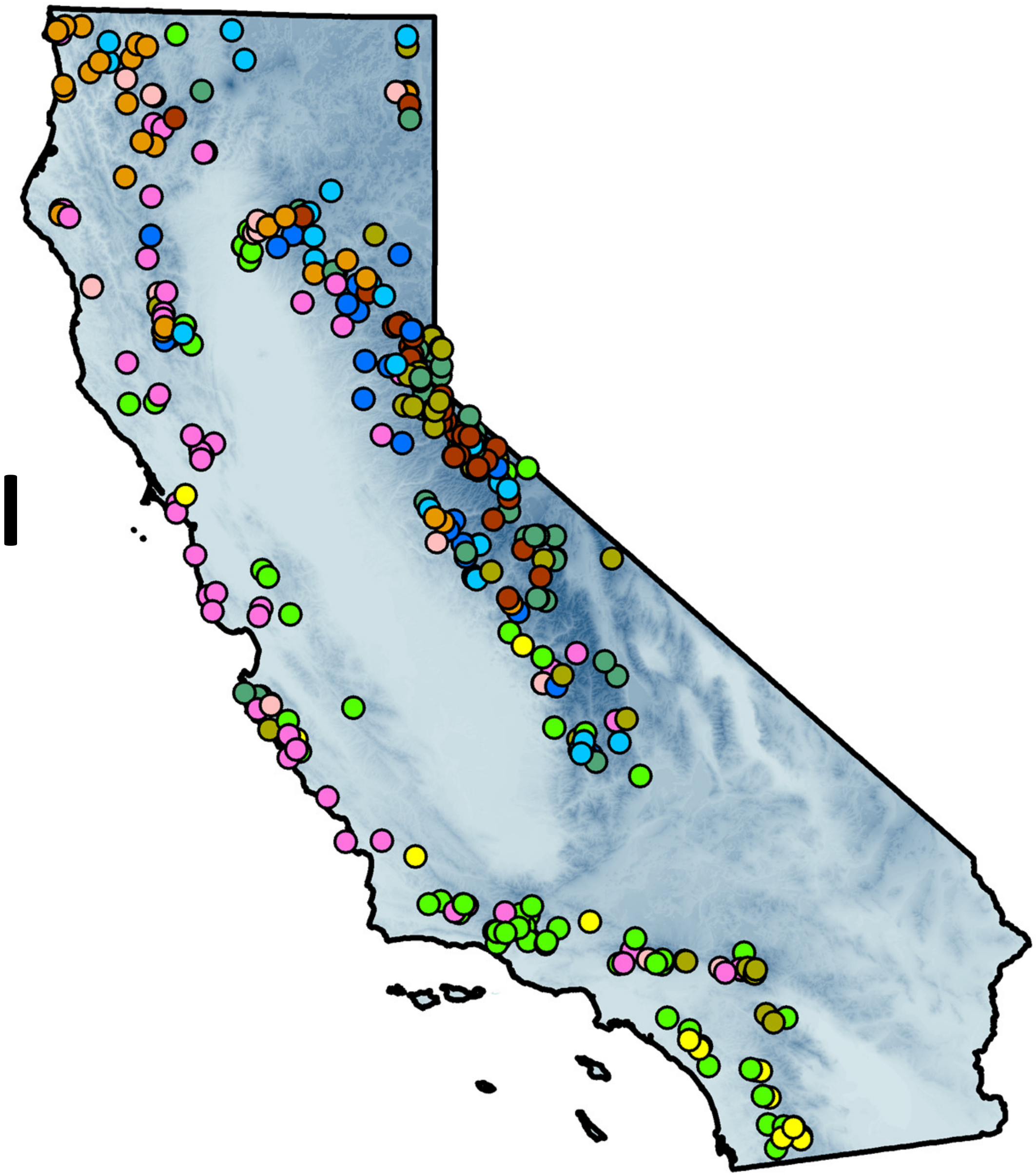
Cluster biological similarity

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



13

10 biological clusters



14

Multi-metric Indices (MMIs)

Species list is converted into metrics representing diversity, ecosystem function, and sensitivity to stress

Taxon	Count
Mayfly species 1	43
Mayfly species 2	12
Mayfly species 3	2
Beetle species 1	1
Beetle species 2	1
Midge genus 1	65
Midge species 1	3
Midge species 2	10
Midge genus 2	3
Dragonfly species 1	2
Stonefly species 1	1
Stonefly species 2	14
Worm species 1	9
Worm species 2	2

- # mayfly taxa
- # predator taxa
- % sediment tolerant taxa
- % herbivore taxa
- % mayfly individuals

15

Why develop an MMI?

- Science panel recommended exploring MMI
- MMIs have useful qualities
  - Measure ecological attributes other than species loss
  - Very responsive to stress
  - May work well where species-specific predictions are tricky

- **New techniques available** (*see Hawkins and Vander Laan presentation at 2011 CABW*)
  - Adds site-specific adjustments to traditional MMIs

Building a predictive MMI (pMMI)

follows methods of Hawkins and Vander Laan

Step 1. Calculate lots of metrics at reference and stressed sites

***Step 2.** Create models that adjust metric values to account for major natural sources of metric variation

Step 3. Select metrics based on ability to discriminate reference from stressed sites

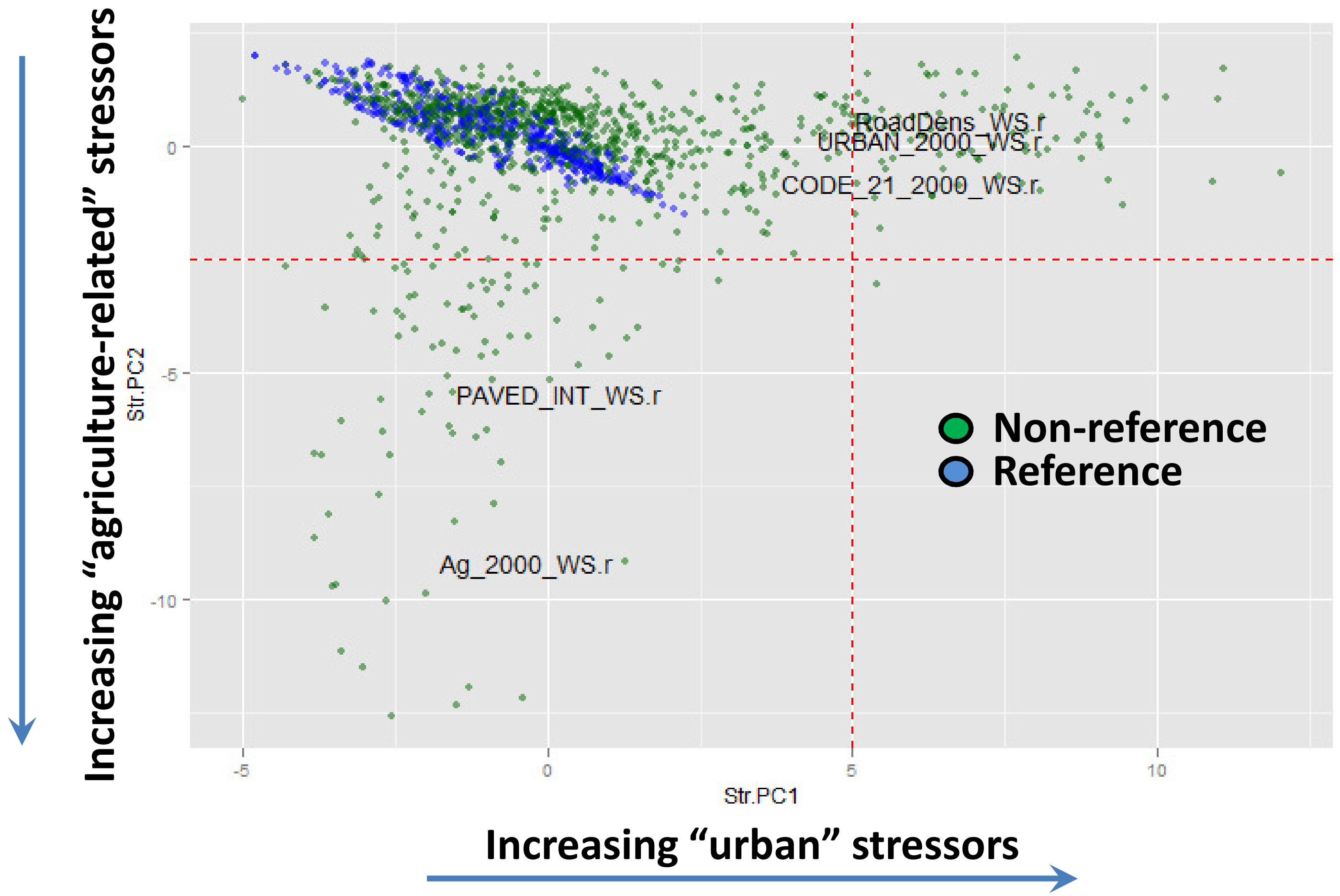
Step 4. Score metrics (after Cao et al. 2007) and assemble into composite pMMI

Step 1. Calculate metrics at reference sites and stressed sites

- **Sample Information:**
  – 1520 sites had "adequate" samples (i.e., >450 bugs) = 2813 samples
  – 514 are reference (same definition as O/E)
  – 175 are highly stressed (84 Ag, 91 Urb)
  – The rest are "test"
- **Calculate Metrics**
  – Used SWAMP's new bioassessment reporting module
  – Subsample to 500 organisms, calculate at SAFIT Level 1 (midges to family)
  – Reject samples <450 specimens
- **Use 80% for model development, 20% to validate**

Identifying stressed sites

PCA with all GIS stressor variables (after removing effects latitude, longitude, and elevation) – stress cutoffs arbitrary



Metrics: the usual suspects

Class	Abundance-based	# Taxa	% Taxa
Taxonomic	% EPT	EPT taxa	% EPT taxa
	[not considered]	Coleoptera taxa	% Coleoptera taxa
	[not considered]	Diptera taxa	% Diptera taxa
	% Chironomidae	[NA]	[NA]
	[not considered]	Non-insect taxa	% Non-insect taxa
	Shannon Diversity	Taxonomic richness	
FFG	% Collectors	Collector taxa	% Collector taxa
	% Predators	Predator taxa	% Predator taxa
	% Scrapers	Scraper taxa	% Scraper taxa
	% Shredders	Shredder taxa	% Shredder taxa
Tolerance	% Intolerant	Intolerant taxa	% Intolerant taxa
	% Tolerant	Tolerant taxa	% Tolerant taxa
	Weighted tolerance value		

Step 2. Adjust metric values to account for influence of natural gradients

- Random forests models (1000 trees) allow us to predict site-specific reference expectation for each metric
- Most influential gradients (all GIS-based):

• Latitude • Longitude • Elevation Range • Site Elevation • Precipitation • Temperature • log Watershed Area	• Soil Erodability • Soil Bulk Density • Soil Permeability • Hydraulic Conductivity	• MgO_Mean • Surfur_Mean • SumAve_Phos • CaO_Mean • Mean Phosphorus • Mean Nitrogen

- If Rsq $\geq$ 10%, use metric residuals (observed – predicted). Otherwise, use raw value

Step 3. Select most responsive metrics

- Select metrics with the best ability to discriminate reference from stressed (i.e., highest t-values – all > t=10)
- Avoid selecting redundant metrics
  - If R^2 with any previously selected metric > 0.5, do not select
  - Avoid "philosophical redundancy"(e.g., EPT taxa and % EPT)

Step 4. Score metrics and assemble into composite pMMI *(follows Cao et al. 2007)*

- **Score metrics**
  - Decreasing metrics:(Obs – Min)/(Max– Min)
  - Increasing metrics:(Obs – Max)/(Min– Max)
    - Max = 95th percentile of reference
    - Min = 5th percentile of stressed

- **Sum** 10 metrics and **adjust** scale to be equivalent to O/E (divide score by mean of reference)

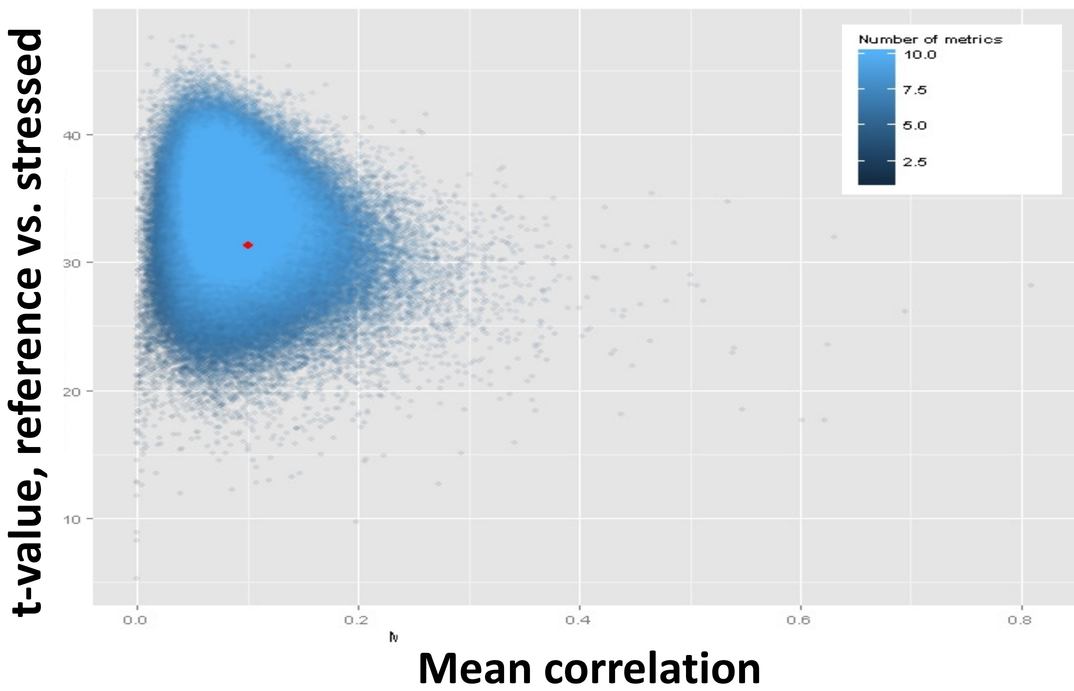
Final Metrics

Metric	Mod v Null	% explained by RF model	|t|	Response
Collector taxa	Modeled	11	13.2	Decrease
Coleoptera taxa	Modeled	40	17.6	Decrease
Diptera taxa	Null	7	13.5	Decrease
Intolerant taxa	Modeled	53	32.2	Decrease
Predator taxa	Modeled	11	13.6	Decrease
Scraper taxa	Modeled	38	20.0	Decrease
Shredder taxa	Modeled	42	19.1	Decrease
% Non-Insect Taxa	Modeled	15	18.1	Increase
Shannon diversity	Modeled	16	10.7	Decrease
Tolerance value	Modeled	32	12.4	Increase

Evaluated multiple MMIs

All subsets of 30 metrics

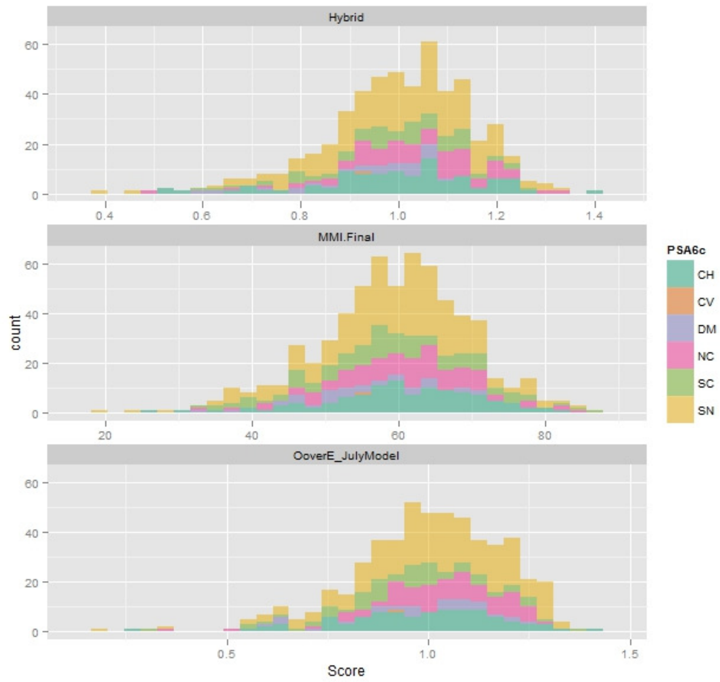
(~100,000 MMIs; 10 metrics max, no redundancy)



- Nearly all MMIs discriminate (reflects pre-screening of metrics?)
- More metrics = convergence to central tendency, better validation
- Thousands of other MMIs are probably just as good as ours

Comparing Performance of 3 Scoring Tools

1. Species Loss Index (O/E)
2. Ecological Structure Index (pMMI)
3. Combined Index ("hybrid")



Created a common validation set for performance measures so we're comparing apples to apples

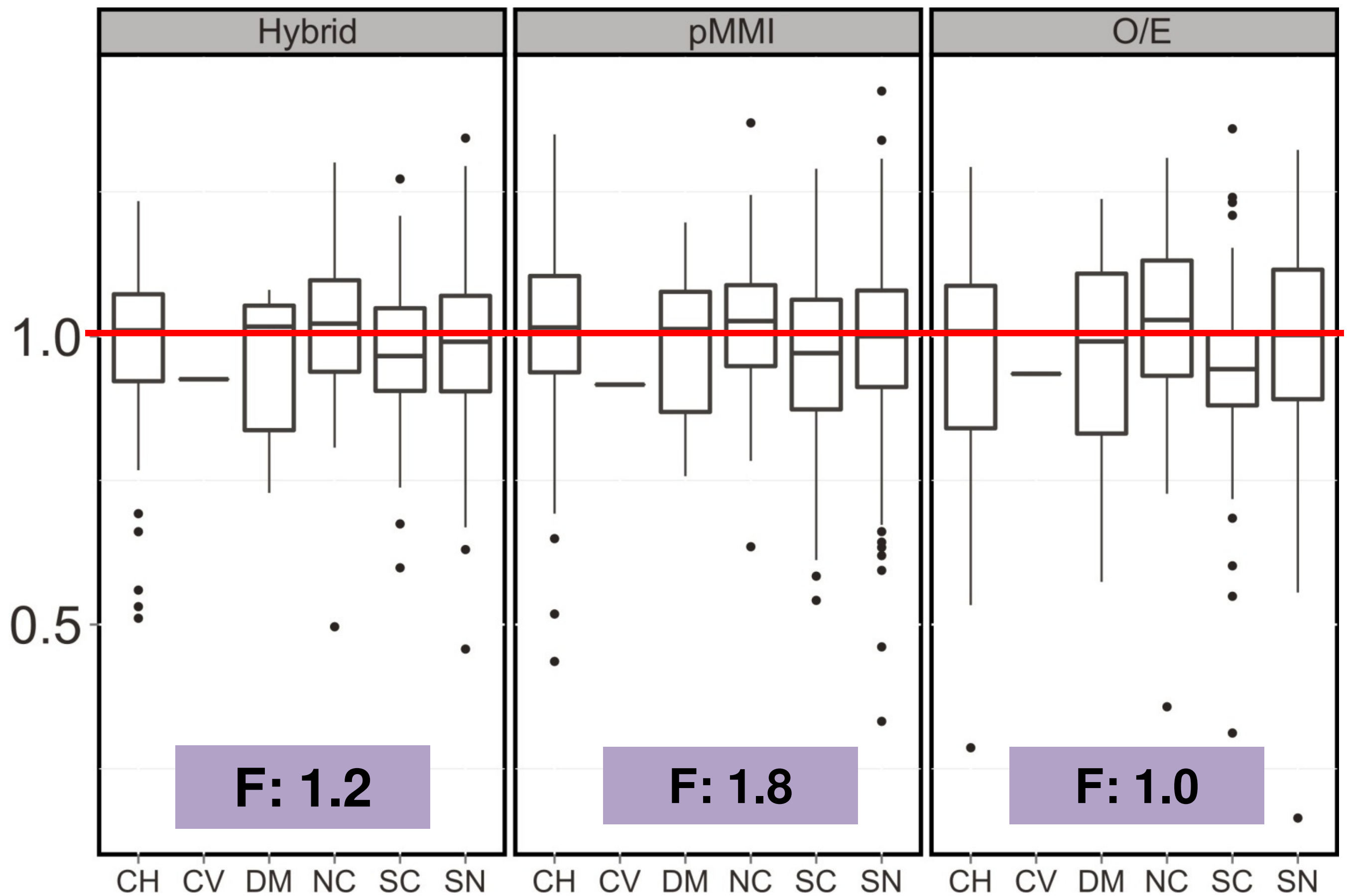
Measuring Performance

All evaluations used a common dataset

Class	Property	Measure	O/E	pMMI	Hybrid
Precision	Variance of reference sites	SD	**0.19**	**0.15**	**0.14**
Sensitivity/ Responsiveness	Discrimination	t-value	**9.5**	**17.6**	**15.3**
	Variance explained by stress	Random forest model	**25%**	**56%**	**49%**
Accuracy/ Bias	Variance explained by natural gradients (ref sites)	Random forest model	**-7%**	**-9%**	**-8%**
	Difference among PSA regions (ref sites)	ANOVA	**1.0 (ns)**	**1.8 (ns)**	**1.2 (ns)**
Replicability	Within-site variability	Mean within-site SD	**0.10**	**0.10**	**0.08**

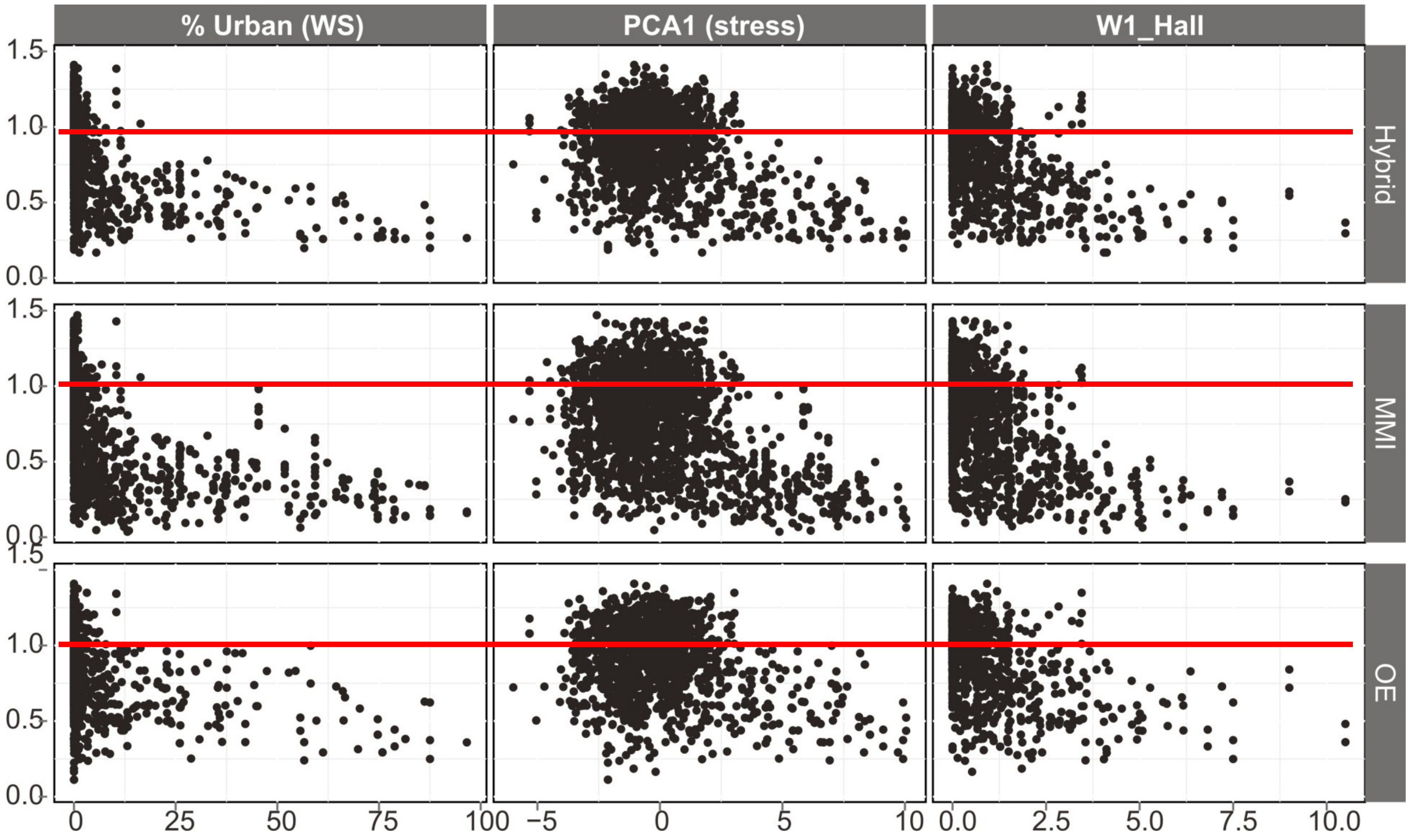
Statewide Consistency

Distribution of reference scores by PSA region

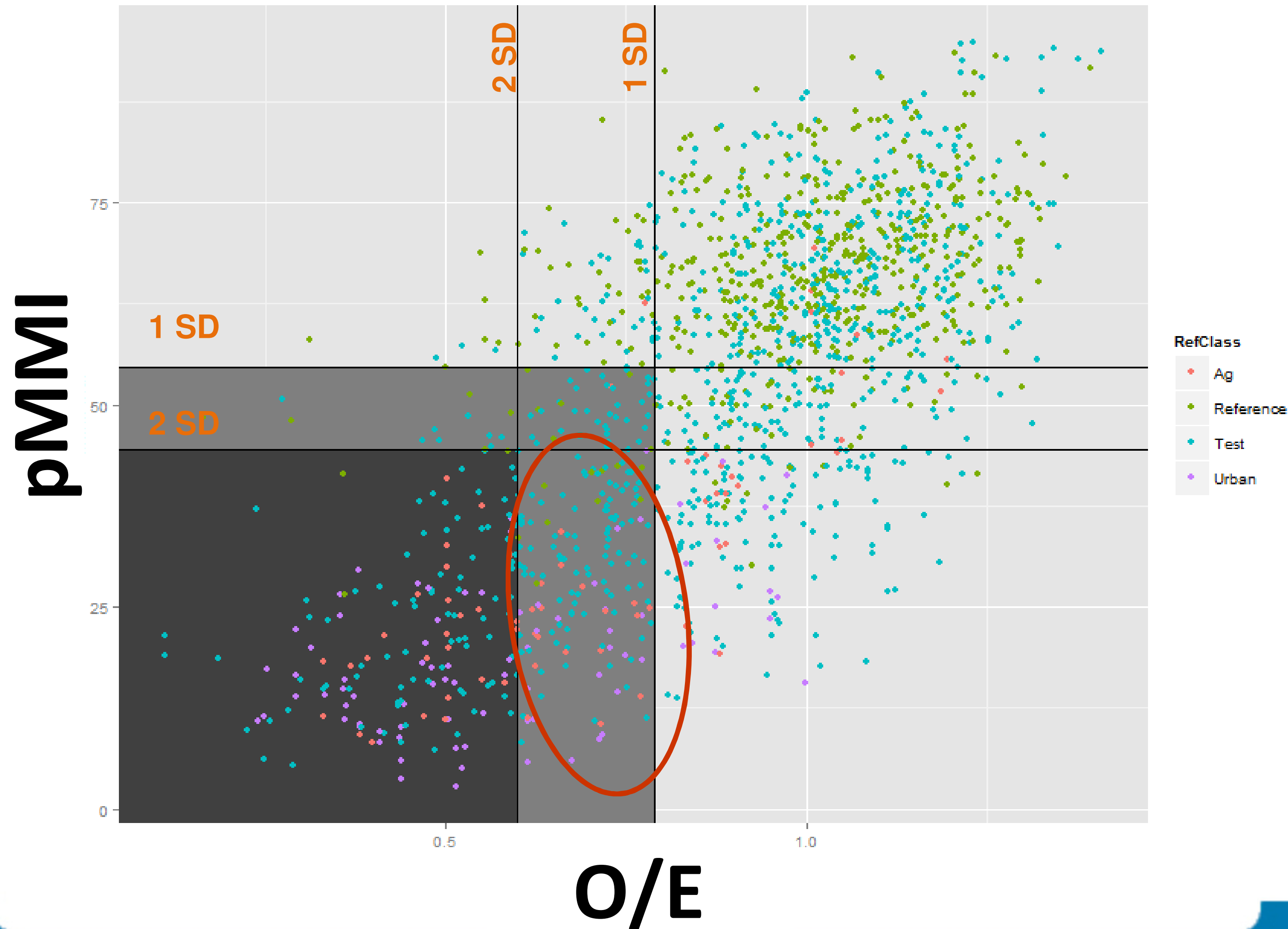


SWAMP
Surface Water Ambient Monitoring Program

Responsiveness to stress

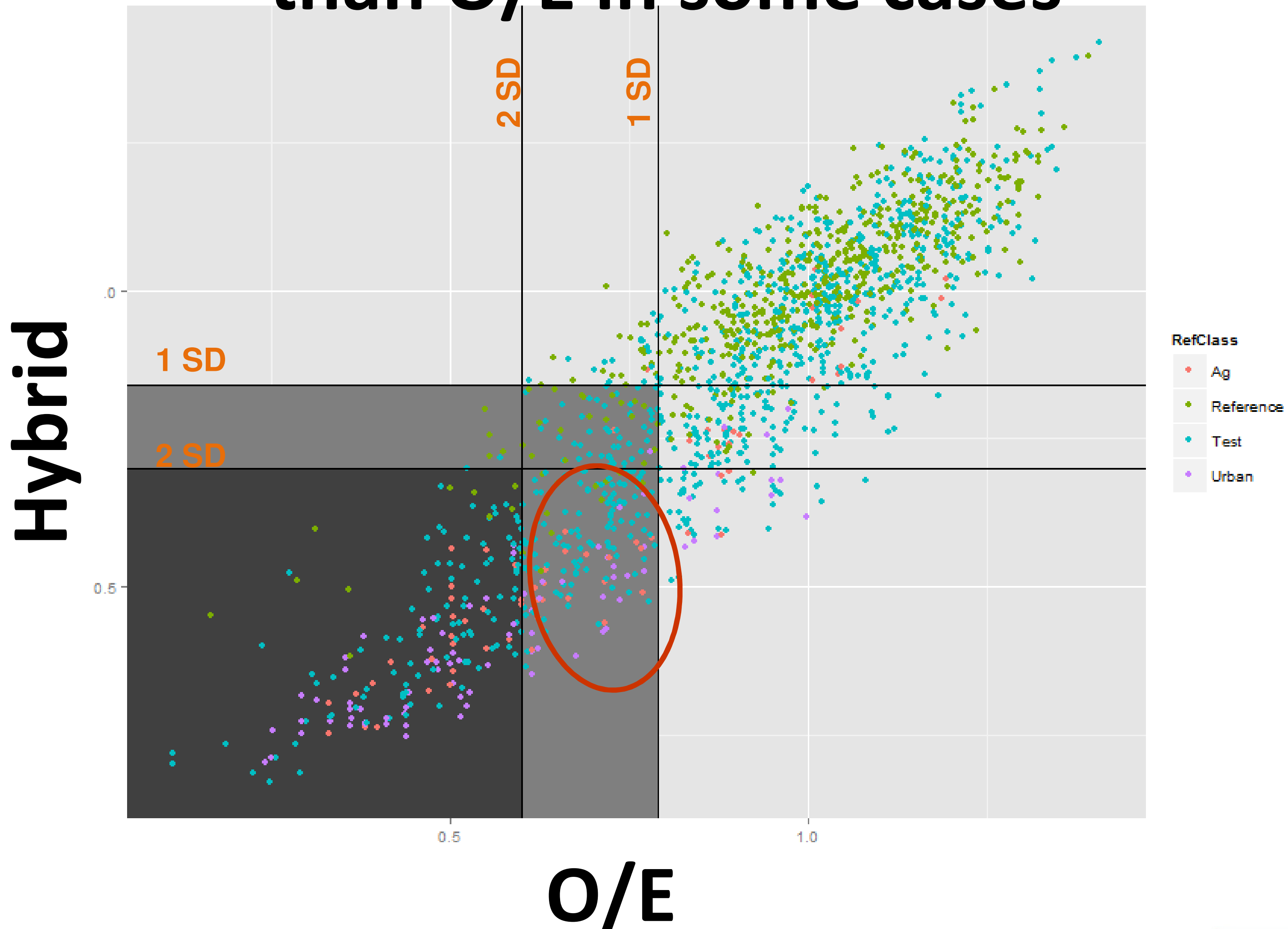


pMMI and O/E have general agreement, but tell us somewhat different things

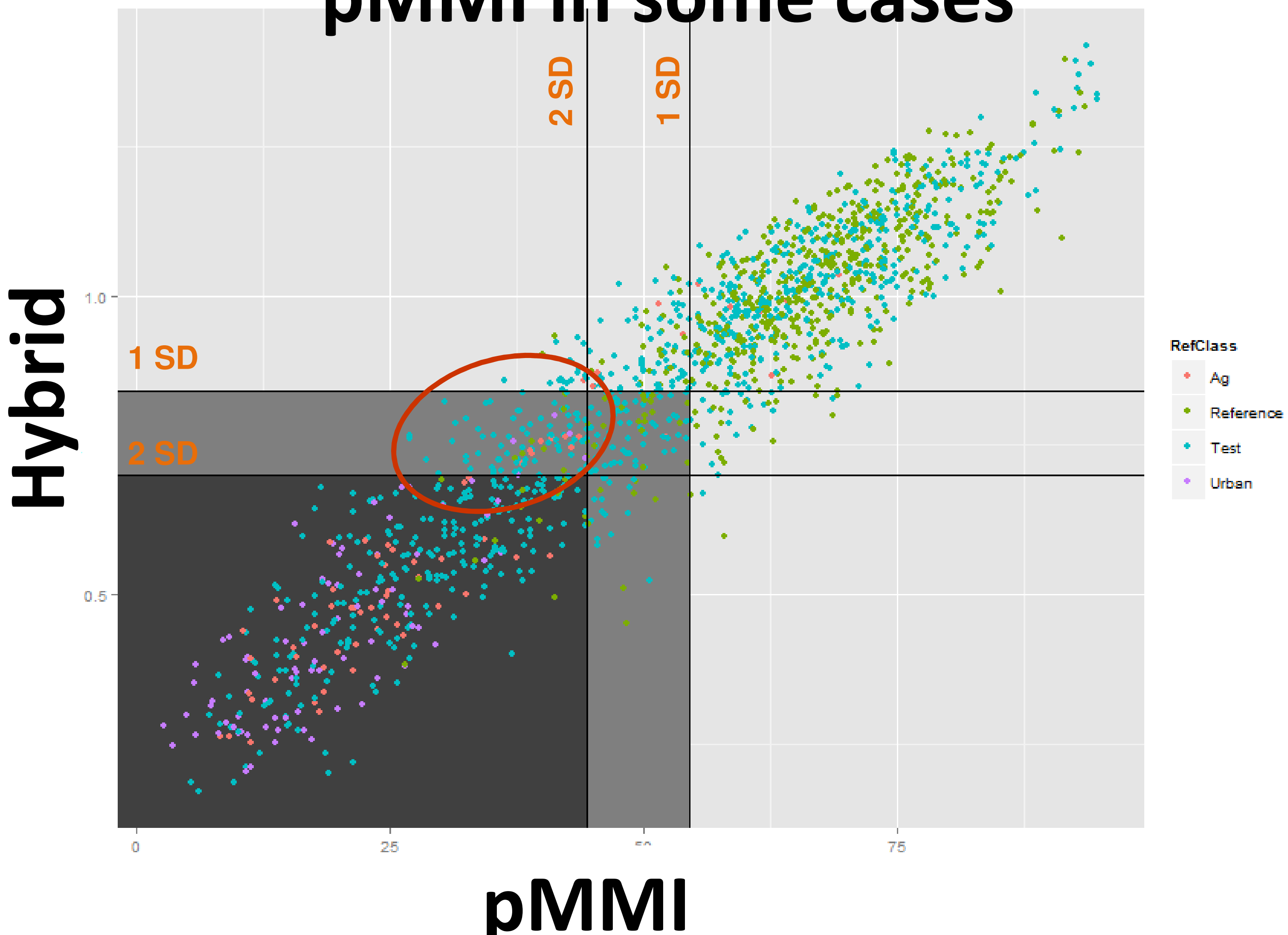


ace Water
bient Monitoring
gram

Hybrid more likely to find impairment than O/E in some cases



Hybrid less likely to find impairment than pMMI in some cases



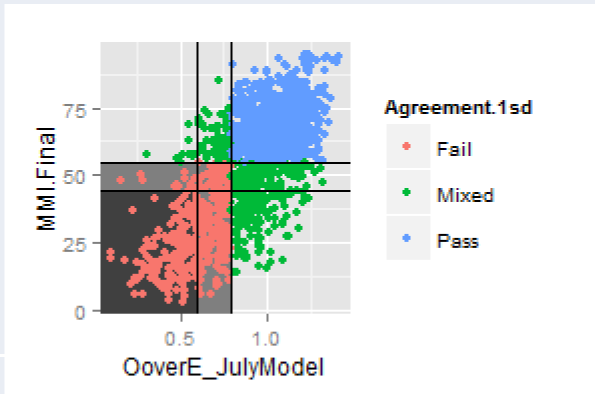
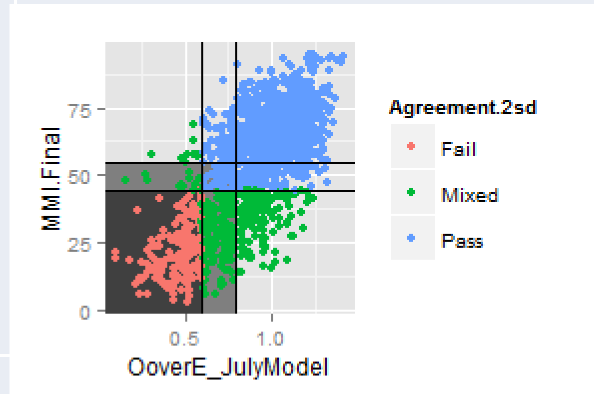
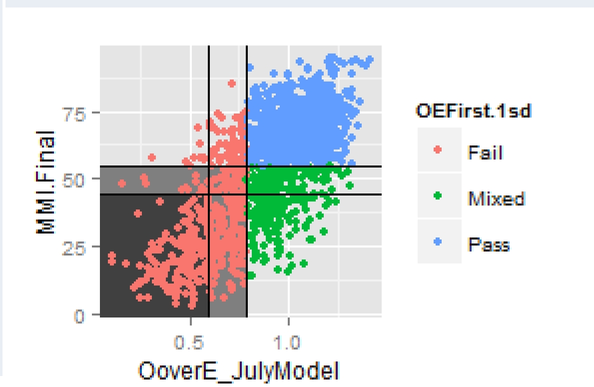
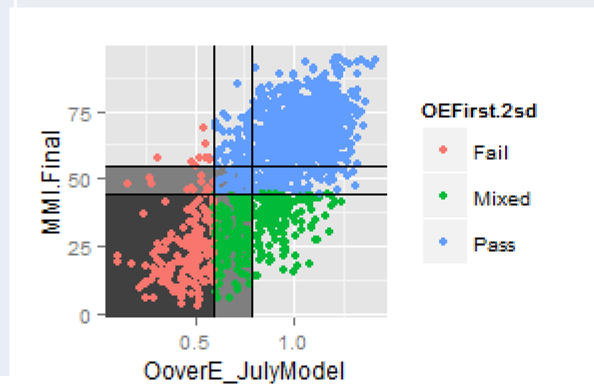
Both pMMI and O/E have desirable qualities

- pMMI is precise and very responsive to stress (but it was designed to be)
- % species loss is an intuitive, meaningful measure of condition
- Both are accurate and applicable throughout state
- Potential for complementarity is great -- we explored a few options (see Science Panel)

Options for using 2 indices

- Hybrid
  - Equal
  - Unequal weight
- Agreement/ Disagreement
- Use one to verify the other

Multi-index Approaches

	1sd	2sd
Agreement (gray area when disagree in both directions)		
O/E First (only MMI disagreement used to set gray area)		

We recommend an equal-weight combined index

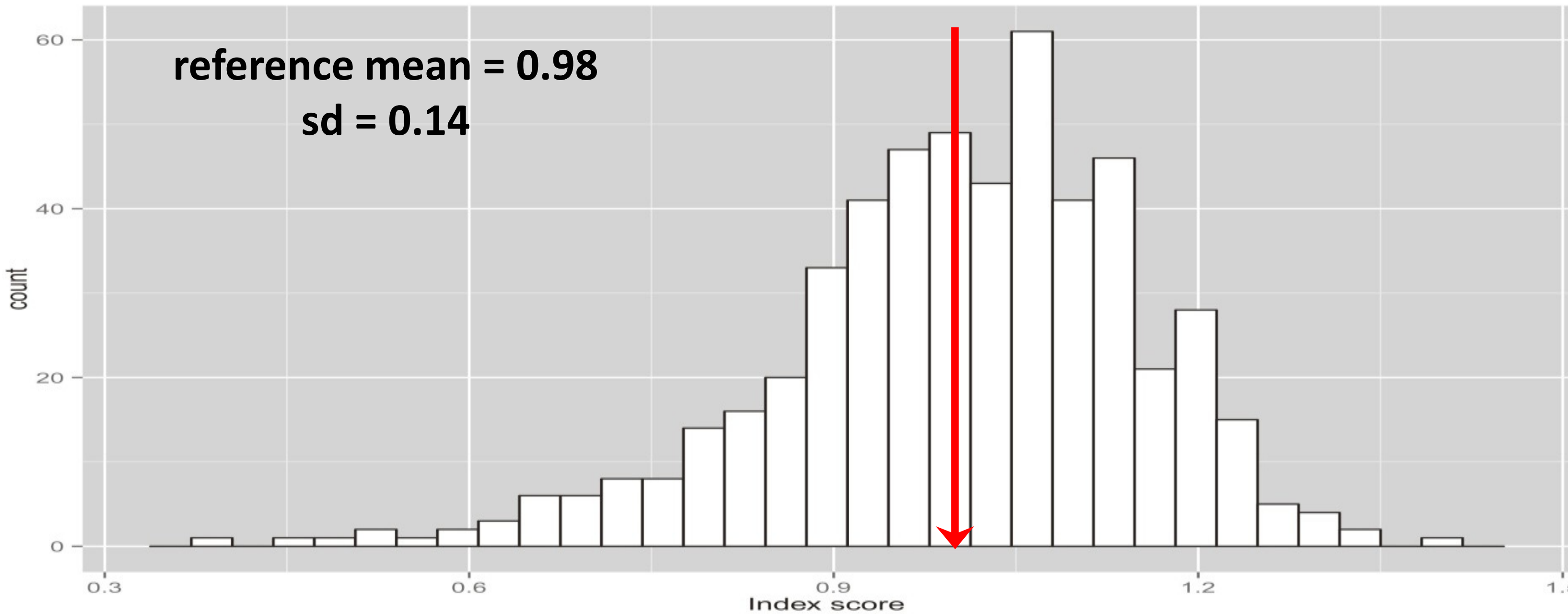
- Retains some of the better qualities of both indices, tempers weaknesses
- Retains the precision and high sensitivity of the MMI and the independence of the species loss data
- Can be disaggregated into component MMI and O/E
  - Don't lose information by combing
  - Reference expectations for all components are available
- No objective *a priori* reason to weight
- Implementation is easier with a single score

California Stream Condition Index (CSCI)

Part A: Ecological Structure Component (pMMI)

Part B: Taxonomic Loss Component (O/E)

CSCI is a simple average of the two scores



Options for setting thresholds

- **Statistical criteria**
  - Standard deviation
  - %-ile of reference distribution

- **Ecological criteria**
  - Acceptable species loss or change in community structure

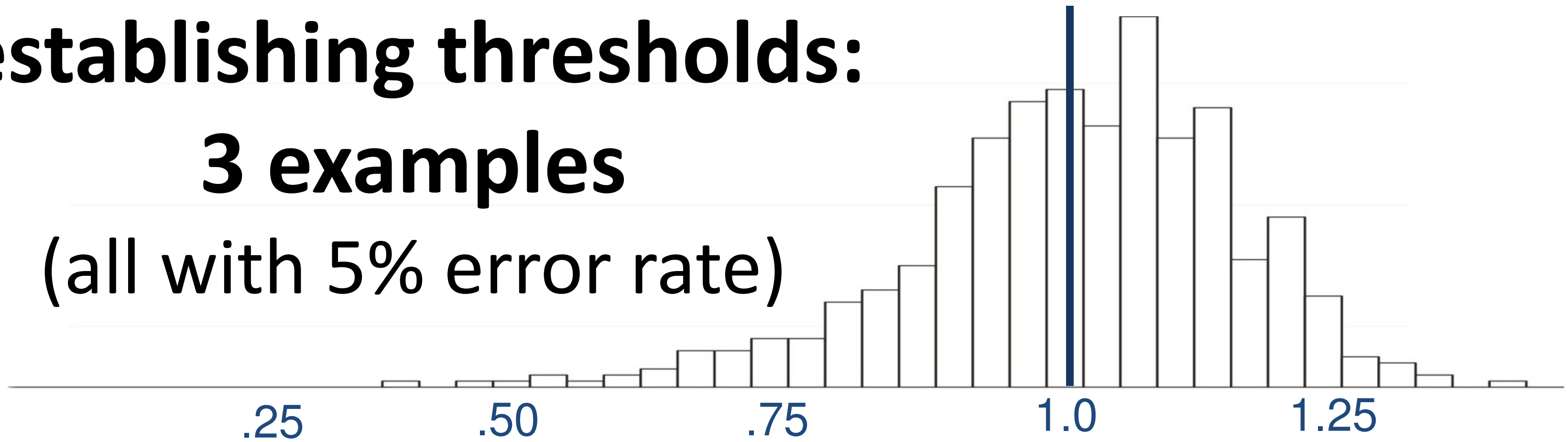
SWAMP
Surface Water Ambient Monitoring Program

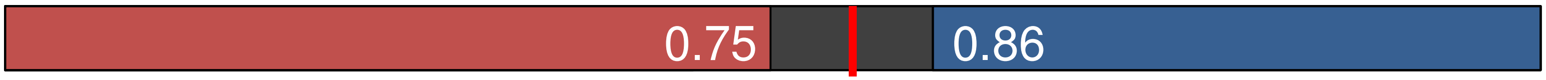
We recommend statistically defined thresholds with a gray area

- Widely accepted practice with broad acceptance
- Ecological benchmarks are appealing biologically, but we have limited basis for setting these objectively
- Gray area is helpful way to express uncertainty in whether a sample reflects site condition

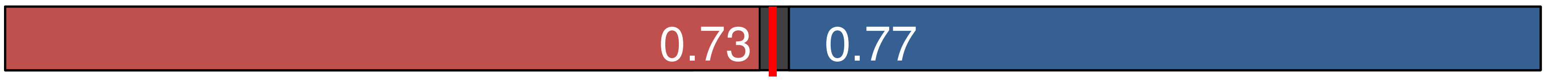
Statistical approaches to establishing thresholds: 3 examples

(all with 5% error rate)





95% and 85% confidence that site is not equivalent to reference



95% confidence that the 95% threshold is where we think it is

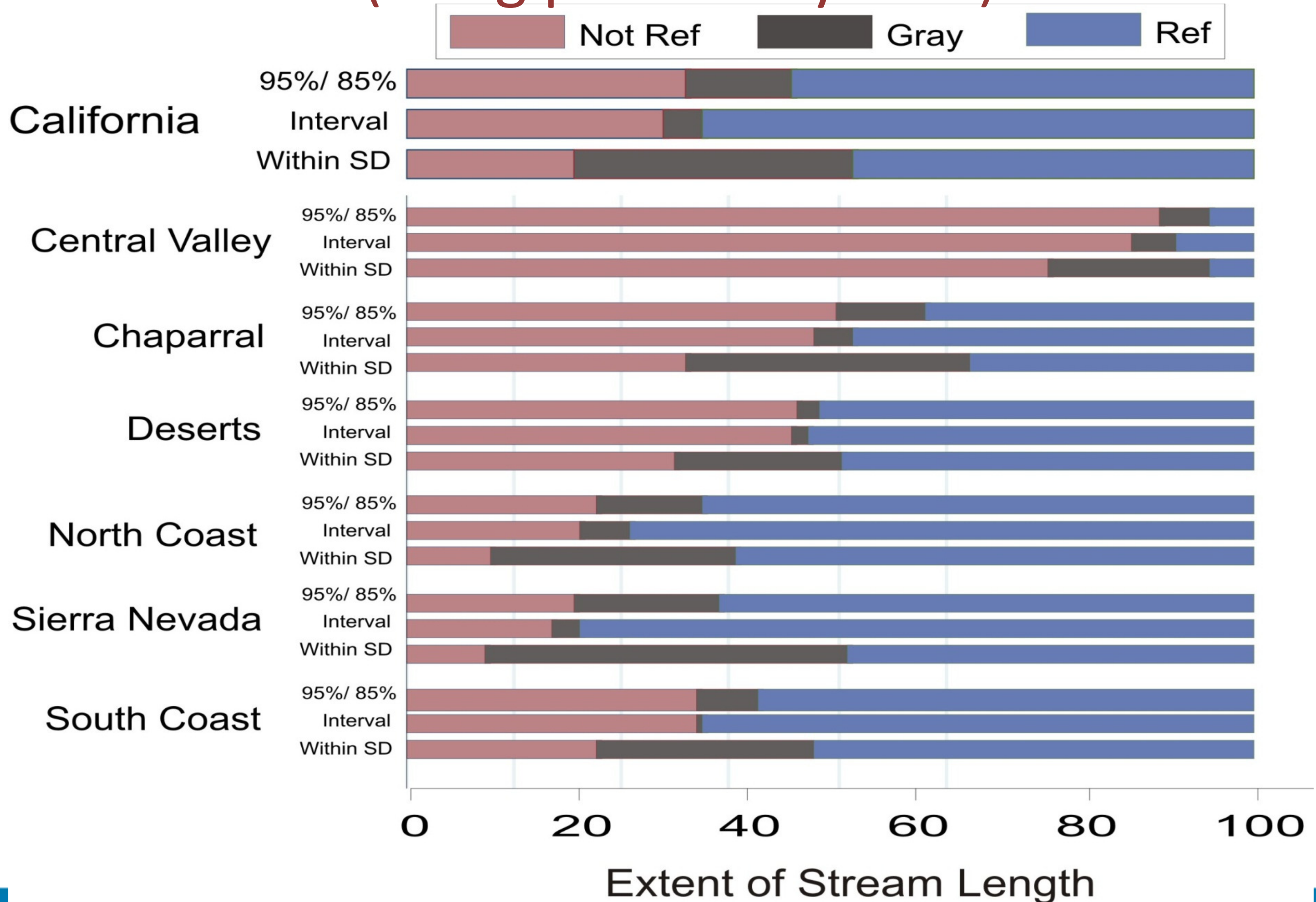
Use within-site error rate to establish uncertainty around threshold

Different approaches for multiple samples (i.e., increasing certainty about site condition)

- Formal t-test vs. threshold
  - **Pass** if site mean > threshold; **Fail** if site mean < threshold
  - **Gray**: mean ~ threshold
    - Different responses given power of the test
      - Low power: More sampling
      - High power: Apply strict threshold comparison (no gray zone)
- What about Type II error?
  - Compare test distribution to reference distribution?
  - Set alpha at 0.10 or higher?
- Other ideas?

Extent of stream length by region

(using probability data)



Questions about thresholds

Are there other options we should consider for guarding against Type I error (false positives)?

Can you suggest objective ways to protect against Type II error (false negatives)? Is there a way to incorporate a “safety factor”?

Do you favor one of the approaches for bounding a gray area?

Can you recommend strategies for dealing with multiple data points?

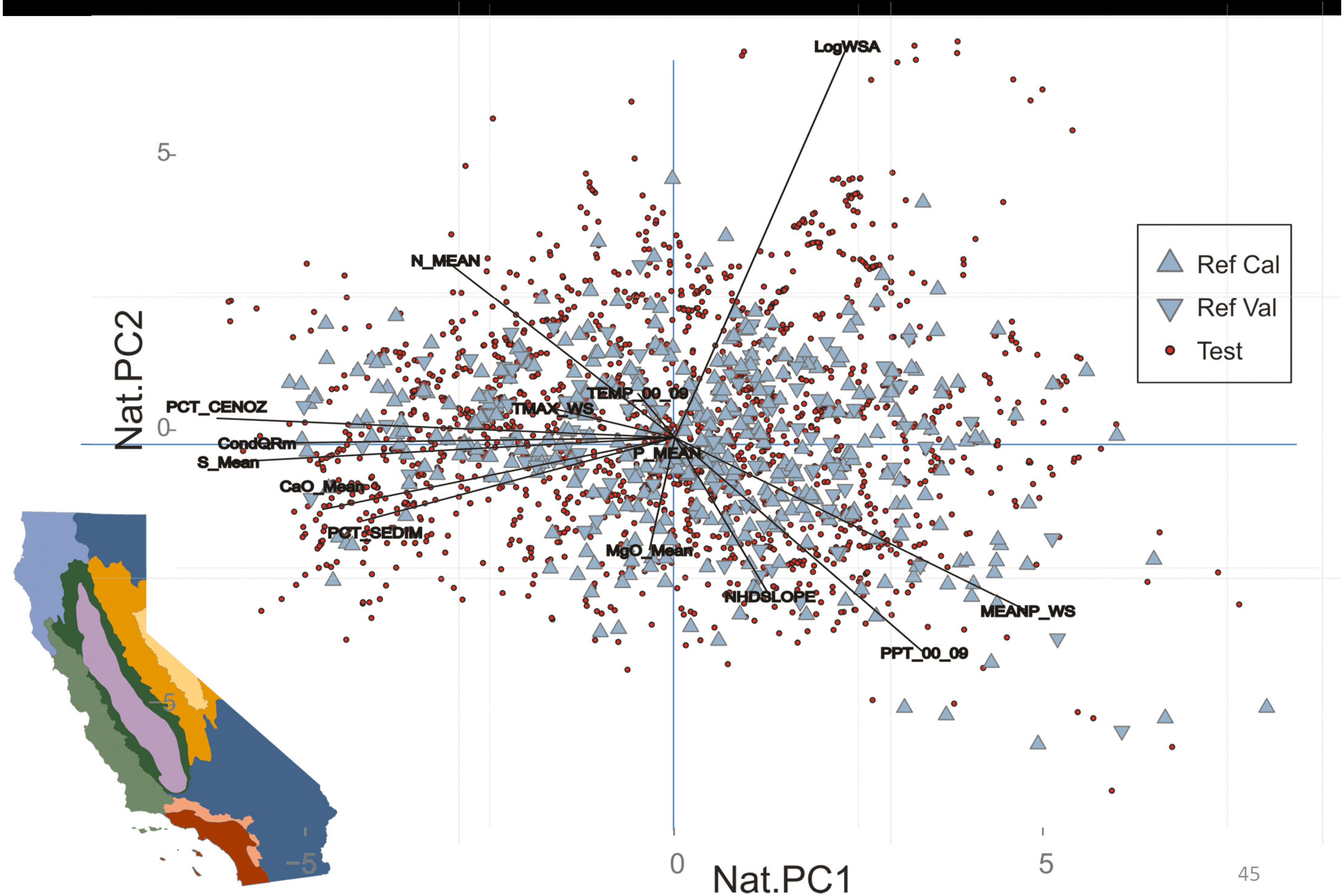
What's next? (Part I):
Quantify applicability of tool

Goal: develop an objective means for determining whether a test site can be appropriately scored(i.e., "is a test site within the "experience" of the model)

- Develop a multivariate applicability test (e.g., Mahalanobis distance)?
- Univariate tests?
- Other ideas?

- How do we define a criterion of acceptance?
- Could be a good way to establish exceptions for truly unique environmental settings.

Multivariate view of natural diversity



What's next? (Part II)
Automation and Documentation

Automate calculations

- Package GIS layers
- Make standard calculation and reporting tools available

Document, document, document

- Journal articles
- Website 101 and FAQ
- Website appendices

SWAMP
Surface Water Ambient Monitoring Program

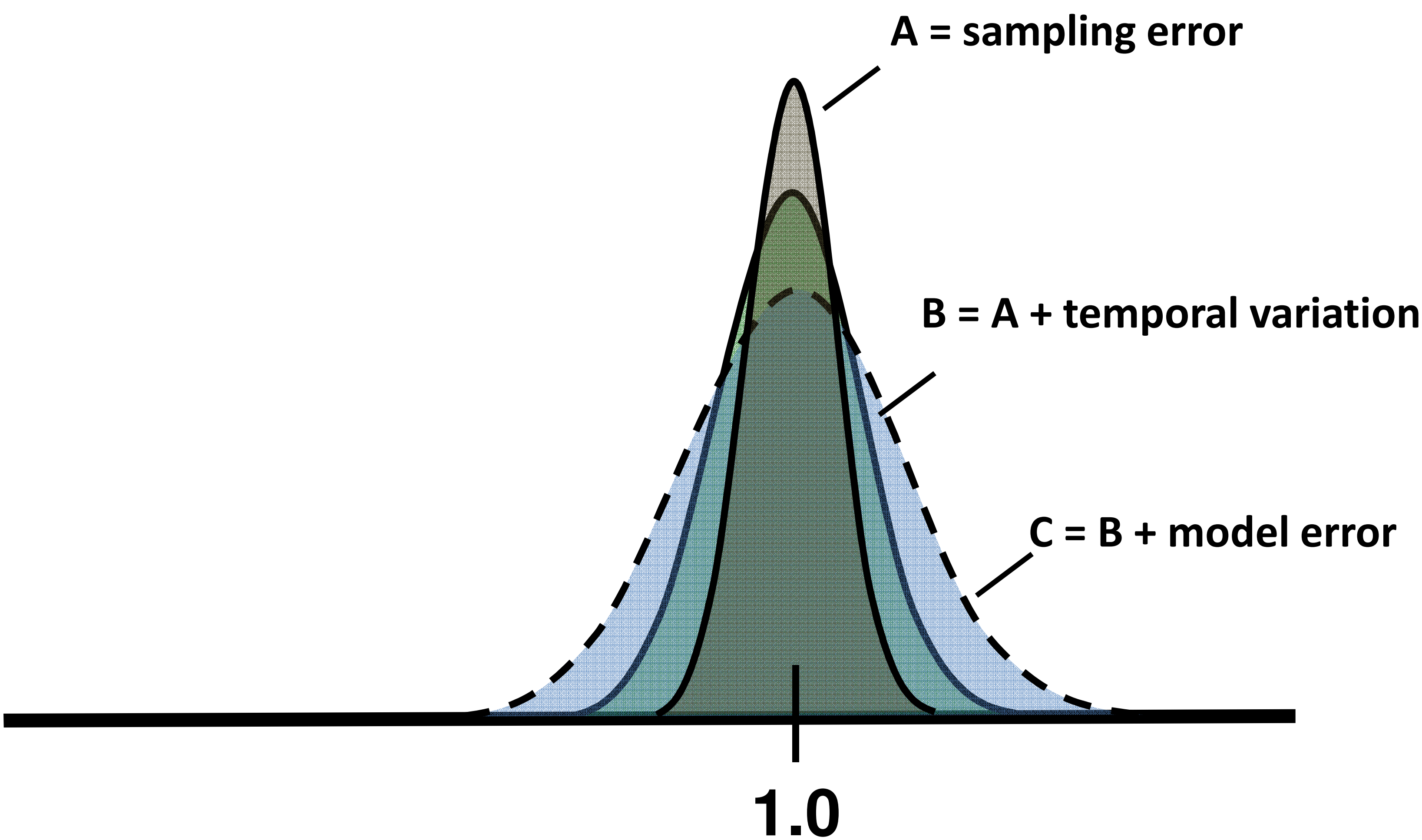
Questions for the panel (Part I)

- Are our scoring tools ready to support implementation?
- Are there other factors we should consider before finalizing our scoring tool recommendations?
    - Combination index versus other options
    - Inclusion of a grey area or not
    - Balancing Type I and II errors
- Gray area options
    - Should we explicitly deal with multiple data points in our gray area approach?

Questions for the panel (Part II)

- Recommendations for exploring and quantifying limits of tool?
- Recommendations for automation?
- Recommendations for documentation?

Sources of variation in index scores



(after Hawkins et al. 2010)