Comparability of Biological Assessments Derived from Predictive Models and Multimetric Indices of Increasing Geographic Scope

Peter Ode
Aquatic Bioassessment Lab
DFG
Charles Hawkins
Utah State University
Raphael Mazor
SCCWRP

As biomonitoring continues to expand, we're developing many tools for scoring biotic condition: predictive models (O/E) and multimetric indices (IBI/MMI)

…spatial scale of these tools varies greatly and boundaries often overlap

If larger models perform as well as smaller ones then we can save $$$ on indicator development. This potential is very appealing, but larger models have some limitations that may restrict their value for local assessments…

Some new national models give us a good opportunity to test this

QUESTION: what are the consequences of using models derived at different spatial scales…

- – for regional condition assessments?
- – for site-specific assessments?

APPROACH: compare the scores derived from larger models with scores from CA-specific models, using test sets from CA

Review: multimetric (MMI) and predictive (O/E) models convert taxa lists to biological condition scores

MMI (IBIs)

- Convert taxa list to metrics (e.g., # mayfly taxa, % scraper taxa)
- Screen metrics for best model characteristics
  - Responsiveness to human stressors
  - Non-responsiveness to natural gradients
  - Strong signal to noise
  - Non-redundancy
- **Score metrics and assemble into MMI (IBI)**

Observed/Expected (O/E)

- Analyze taxonomic data directly
- Cluster reference sites by taxonomic similarity
- **Identify best predictors of cluster membership:**
  - Environmental gradients
  - Local physical habitat gradients
- **Models predict taxa expected at test sites**

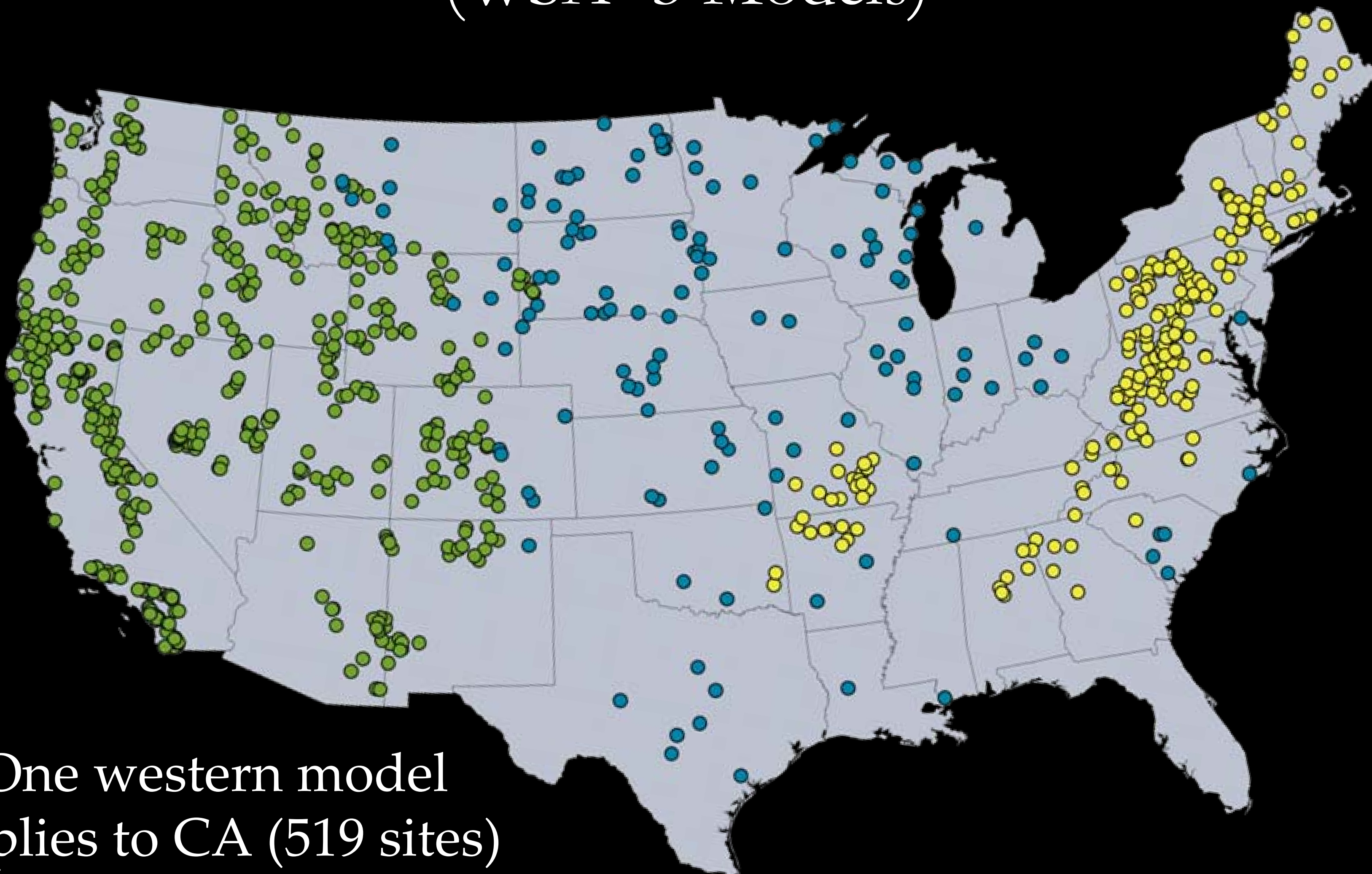
Key Points

- Both MMI and O/E rely on reference sites to establish expected condition

- Geographic range over which models draw reference sites is one of the biggest differences between large-scale and local models (both MMI and O/E)

Following examples are for O/E models, but MMI pattern is similar

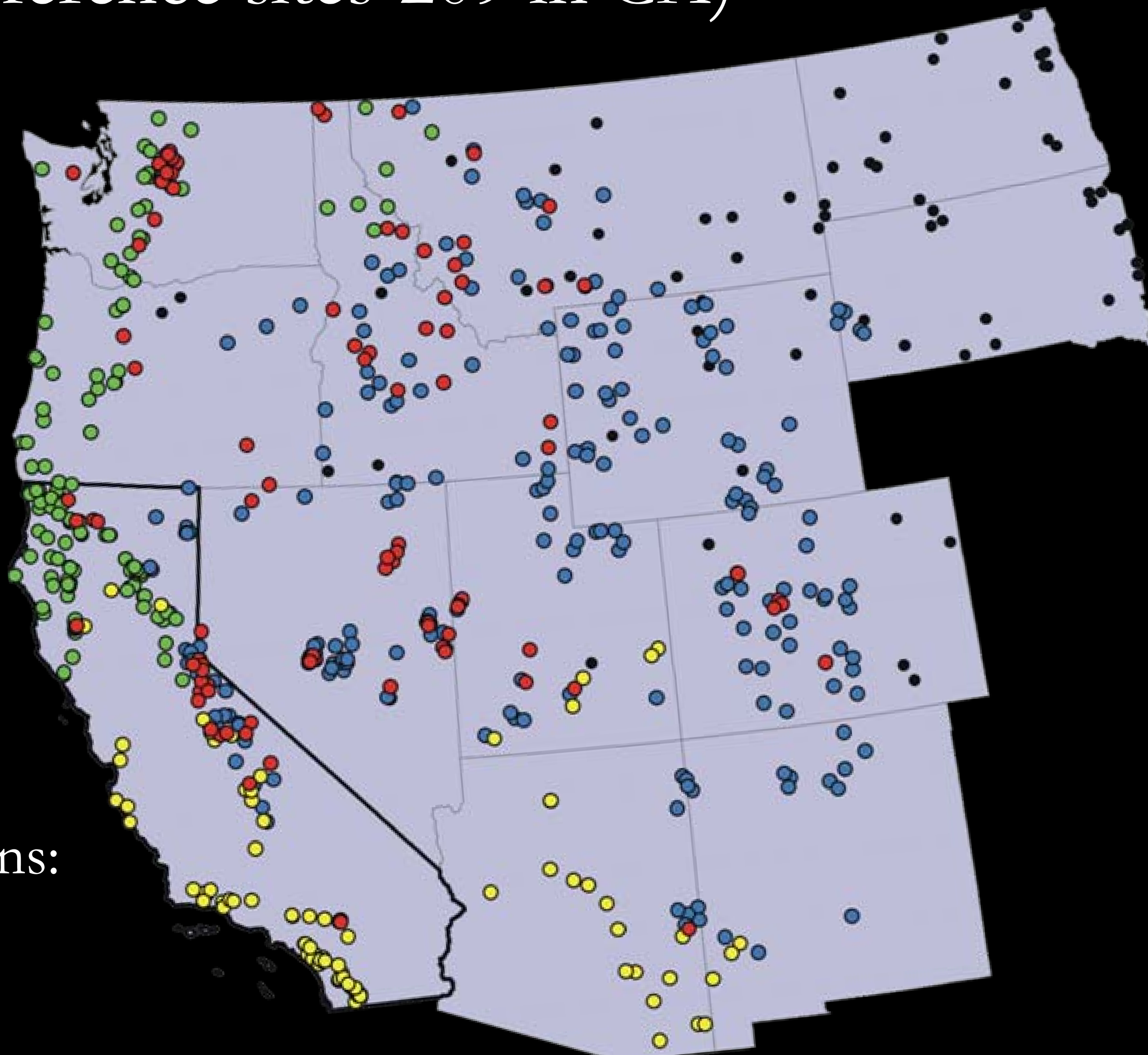
National Wadeable Streams Assessment (WSA -3 Models)



One western model applies to CA (519 sites)

Western EMAP Models
(WEMAP, 629 reference sites-209 in CA)

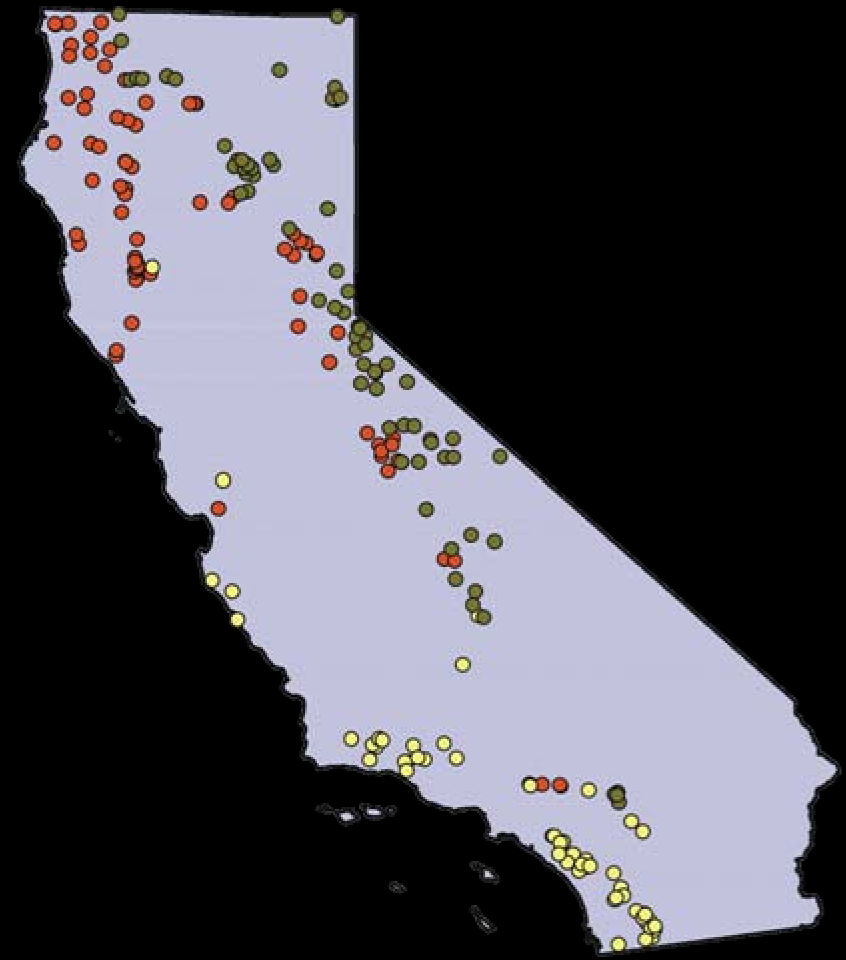
5 separate sub-models
(represented by colors)
… 4 apply to CA



Two WEMAP versions:

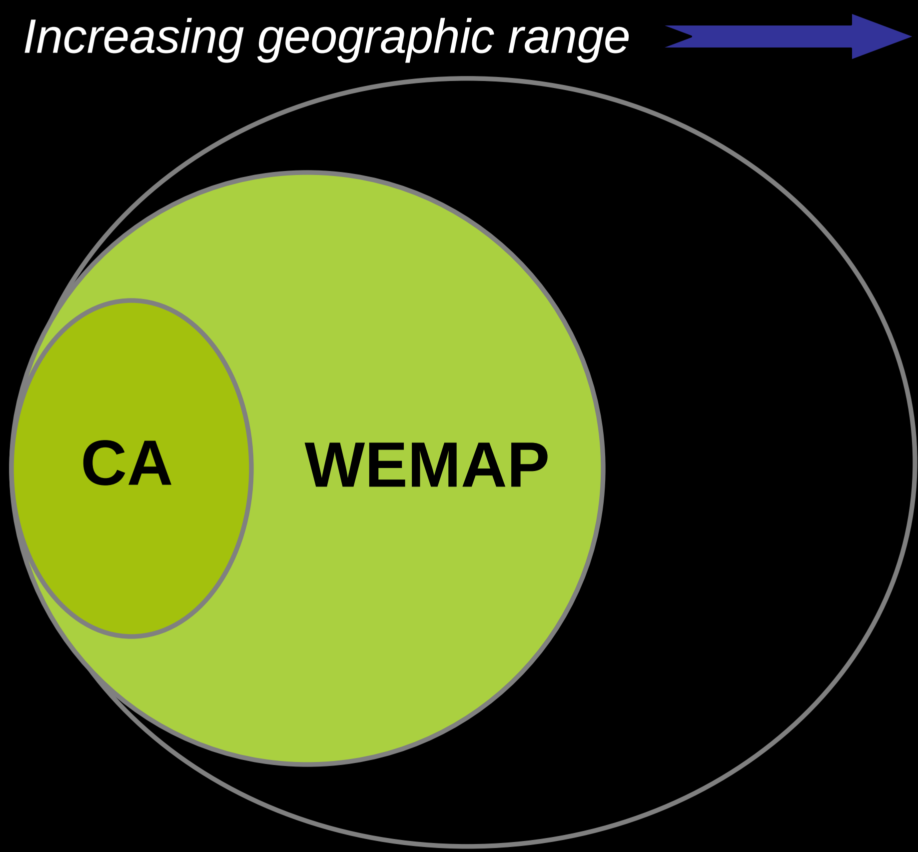
1. WEMAP (null)
2. WEMAP (full)

California Models (206 sites)



3 CA sub-models had better performance than a single model

Spatial Relationships of Model Reference Sites

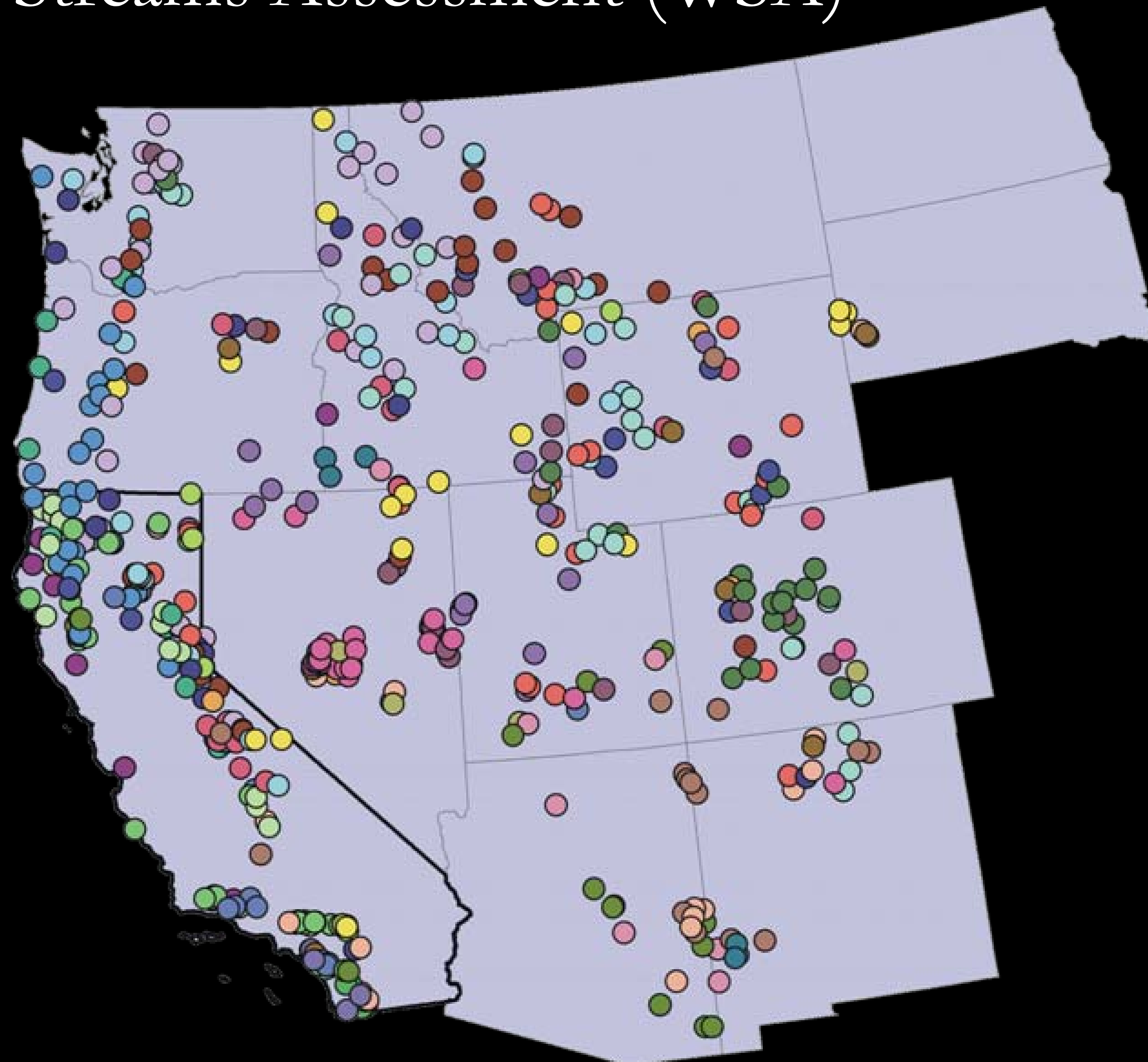


As we expand the spatial scale of the models, we're expanding the geographic area from which we combine reference sites into the clusters used to predict "E".

The larger models contain a smaller proportion of CA reference sites and thus, are increasingly influenced by the characteristics of sites outside the state.

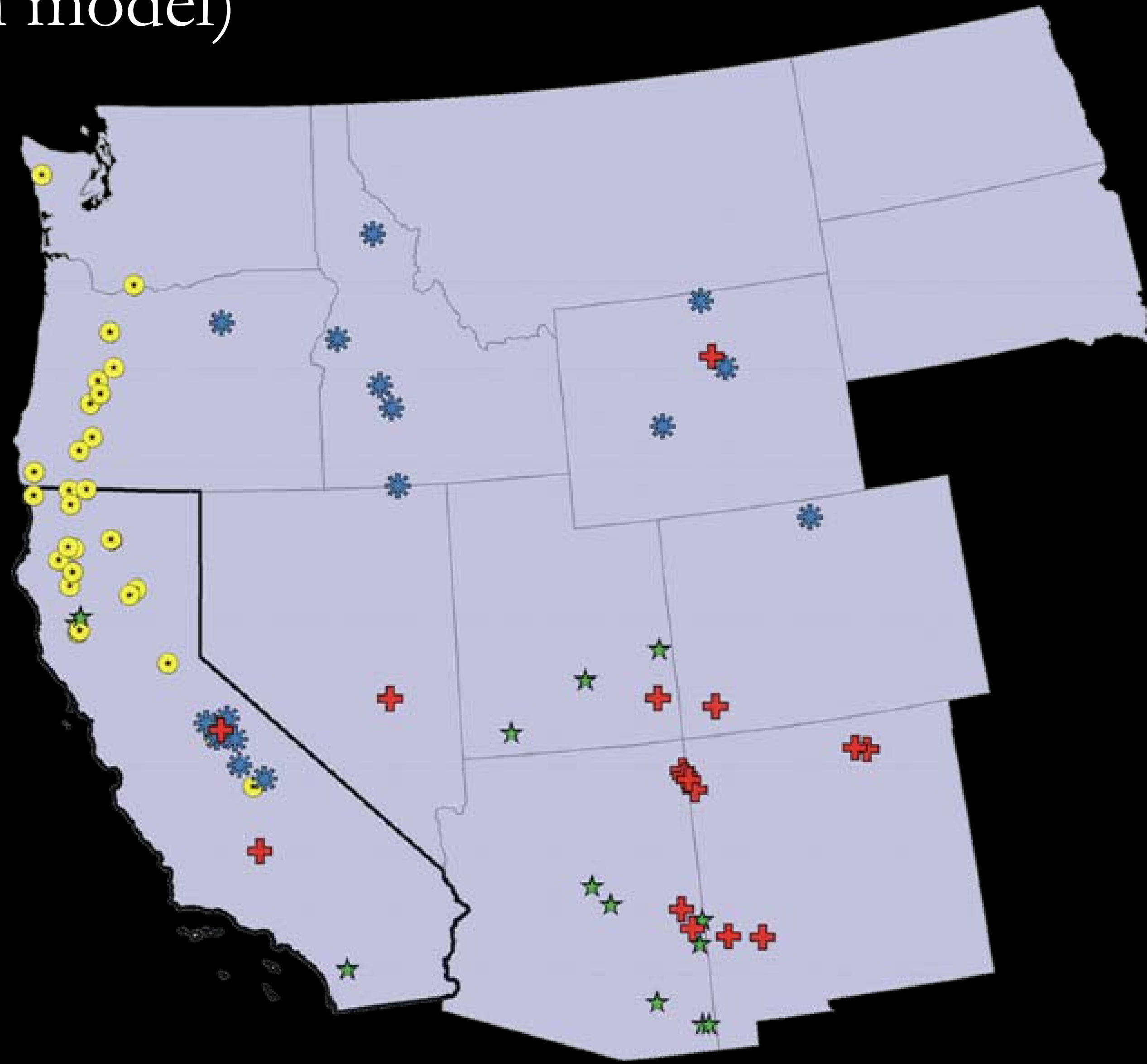
Scale Example (O/E): Wadeable Streams Assessment (WSA)

Western Model = 30 clusters (24 in CA)



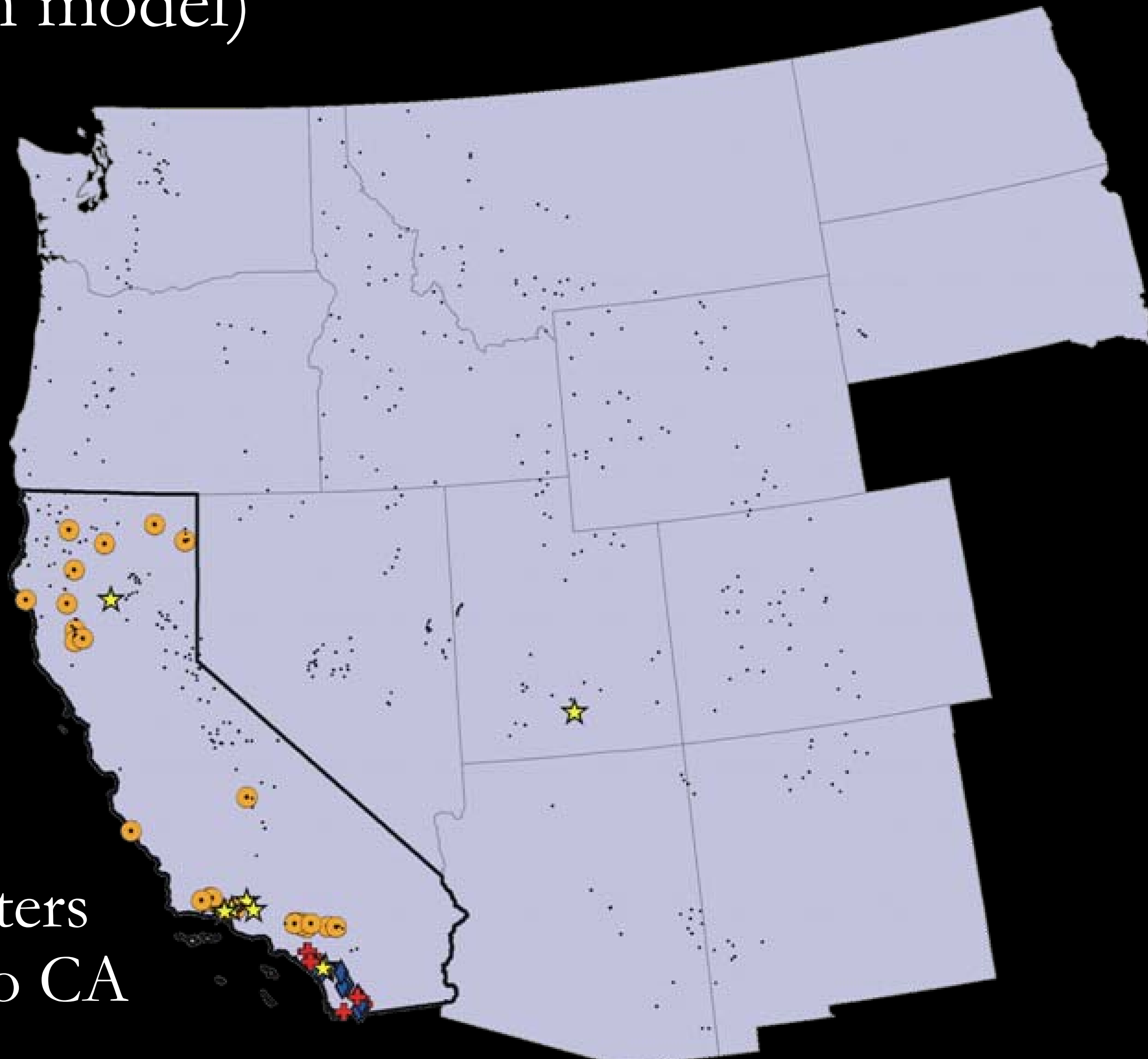
Great variability in geographic range of clusters….

National Wadeable Streams Assessment (WSA- western model)



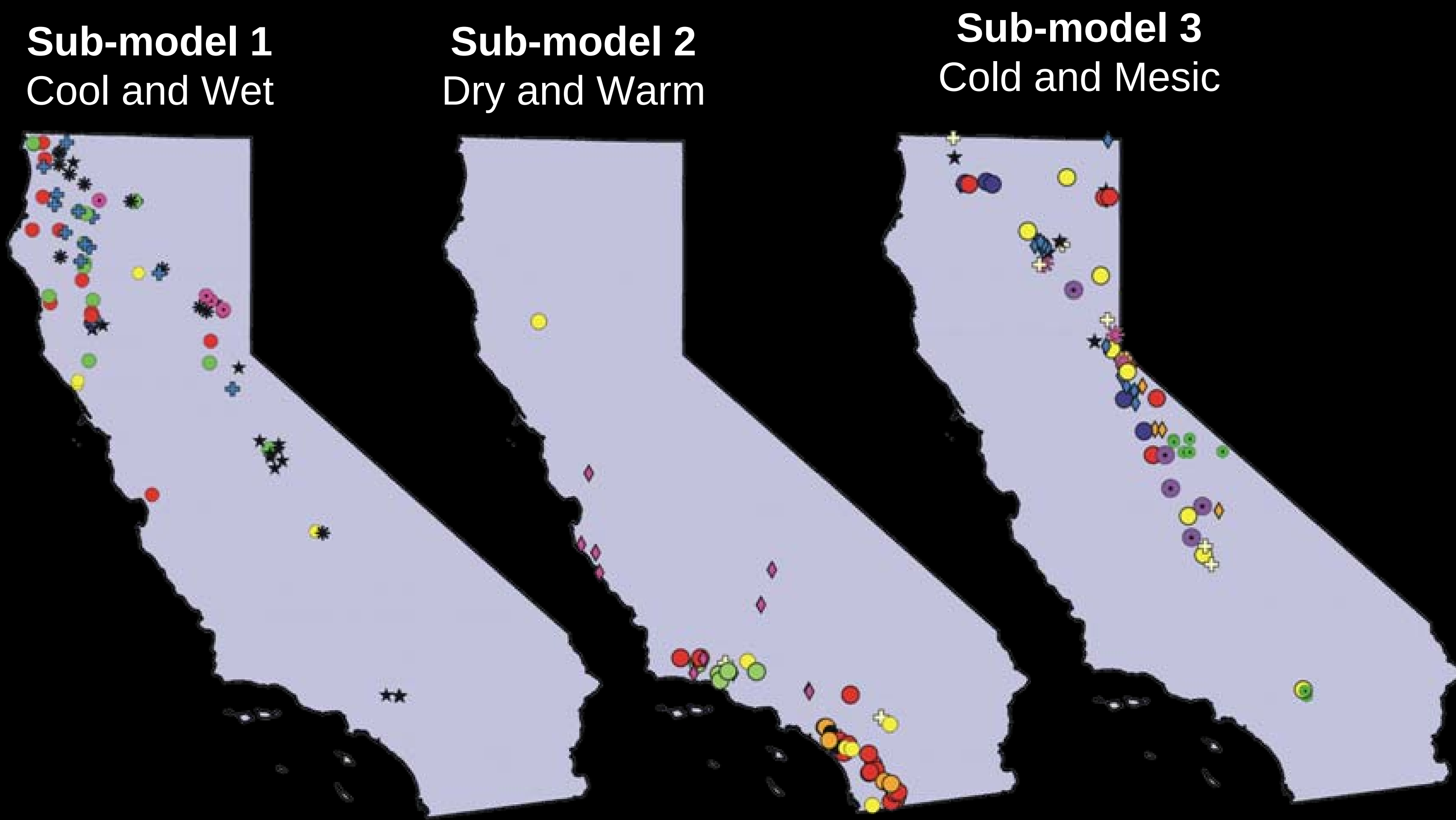
Most clusters are widespread…

National Wadeable Streams Assessment (WSA –western model)



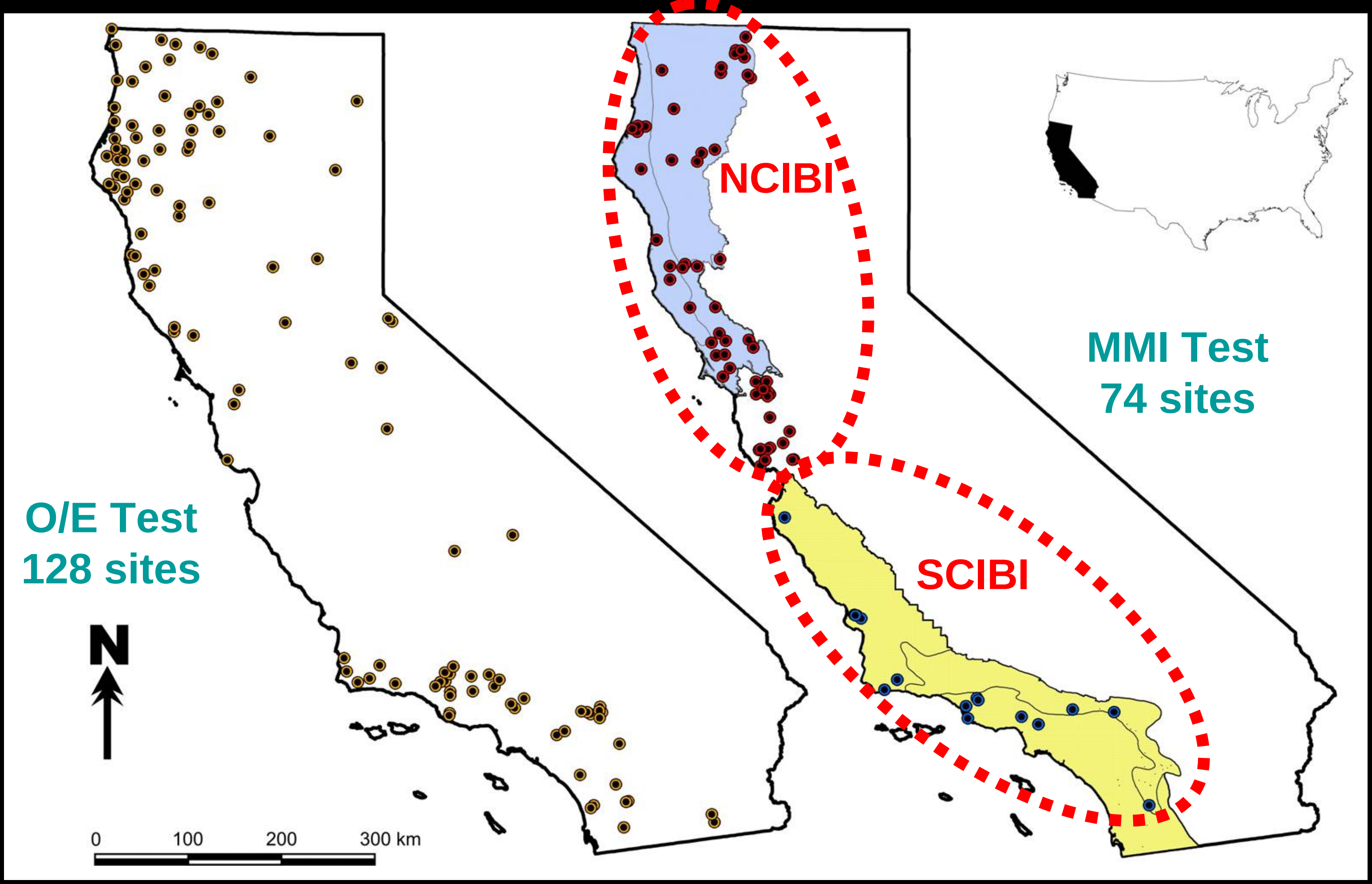
… but a few clusters
are restricted to CA

California Model Clusters



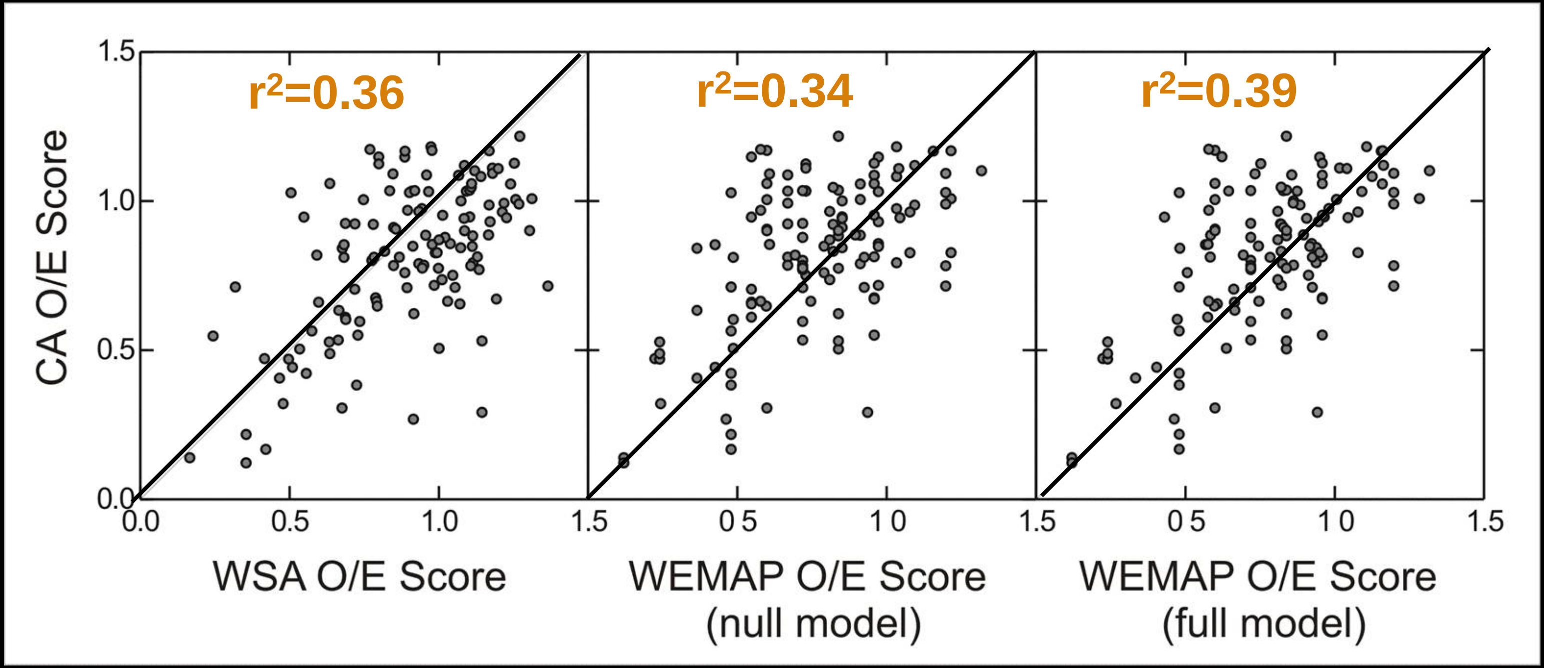
Clusters occur at much smaller spatial scales than in the larger models…
…allows local environmental gradients to become more prominent

Test datasets for O/E and MMI comparisons



- Use tests set to compare larger models to CA models
- Assume CA models represent the “truth”
- Score sites with CA models, WEMAP models and WSA models
- Compare precision, sensitivity, accuracy and bias

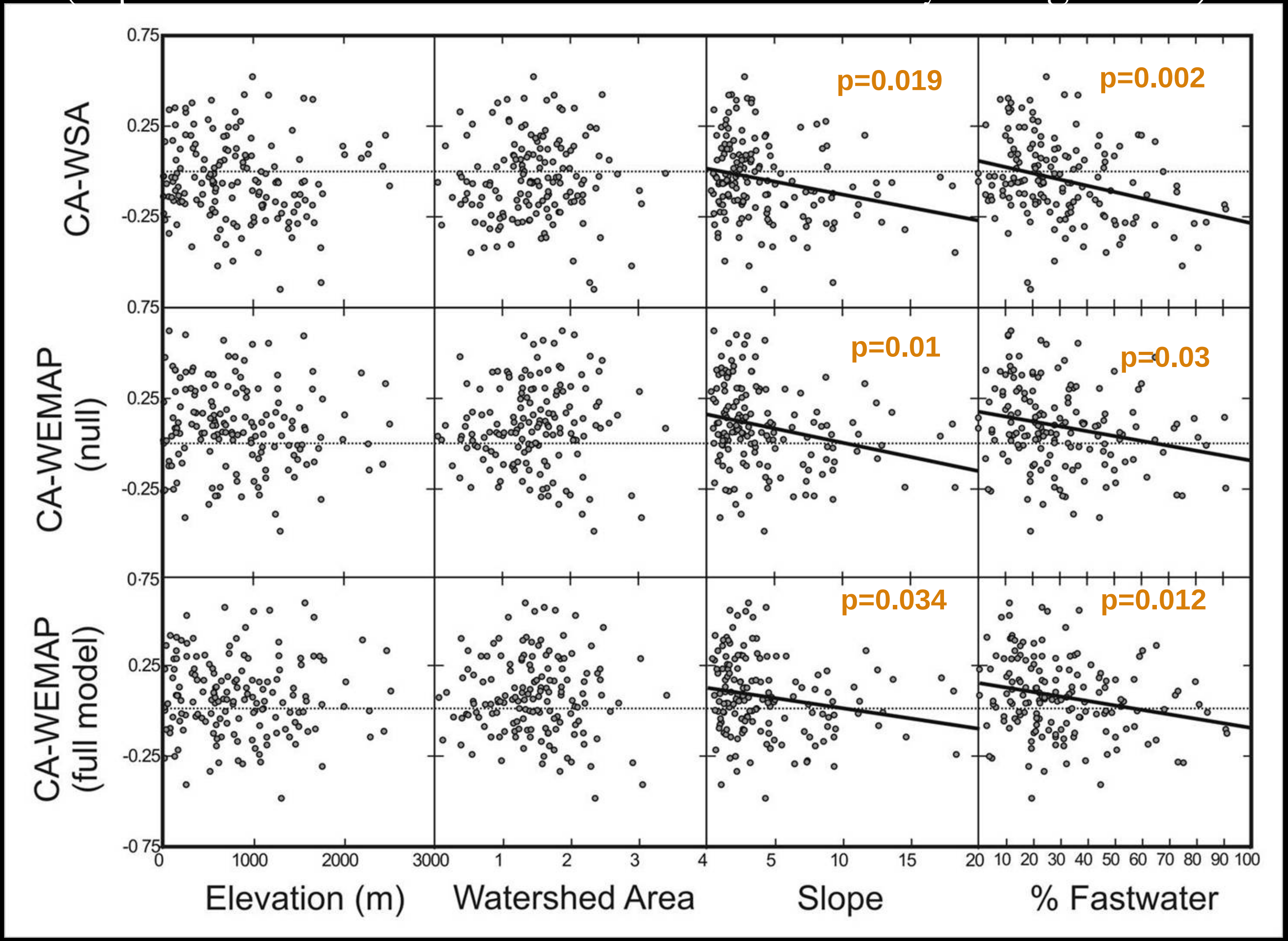
“do the O/E models give sites the same score?” (correlations with CA Models)



Weak agreement between national and CA models

“do differences in O/E model scores vary with key environmental gradients”

(expect a flat line if score differences don’t vary with gradient)



Difference between larger models and CA models tends to vary with increasing % slope and % fast water

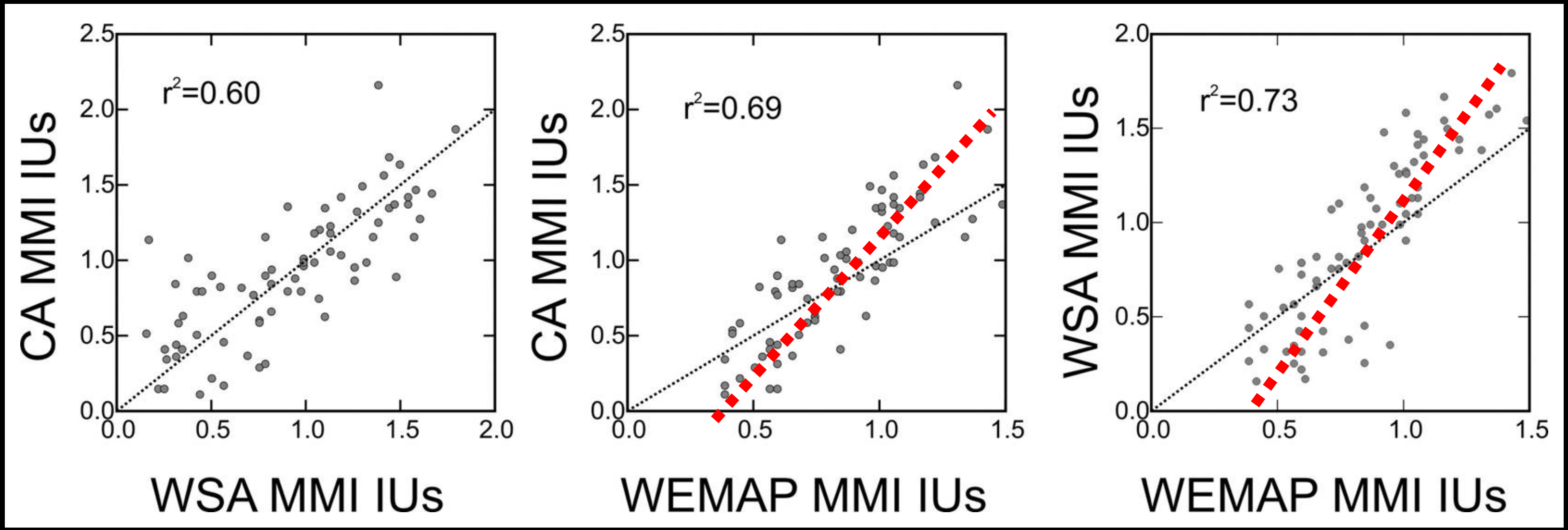
O/E Impairment Decisions: Impaired (I) or Not Impaired (NI)

Use an impairment threshold (mean – 2sd for all models) to compare number of **false negatives** and **false positives** relative to the expectation of the CA models

		CA Model 1 (n=59)		CA Model 2 (n=44)		CA Model 3 (n=25)		TOTALS (n=128)		ALL
		I	NI	I	NI	I	NI	I	NI	
CA ("truth")	I	16	-	17	-	6	-	39	-	39 I
	NI	-	43	-	27	-	19	-	89	89 NI
WSA	I	7	1	7	2	0	1	14	4	18 I
	NI	9	42	10	25	6	18	25	85	110 NI
WEMAP (null)	I	10	4	10	3	4	2	25	9	34 I
	NI	5	39	7	24	2	17	14	80	94 NI
WEMAP (full)	I	12	6	10	3	4	3	26	12	38 I
	NI	4	37	7	24	2	16	13	77	90 NI

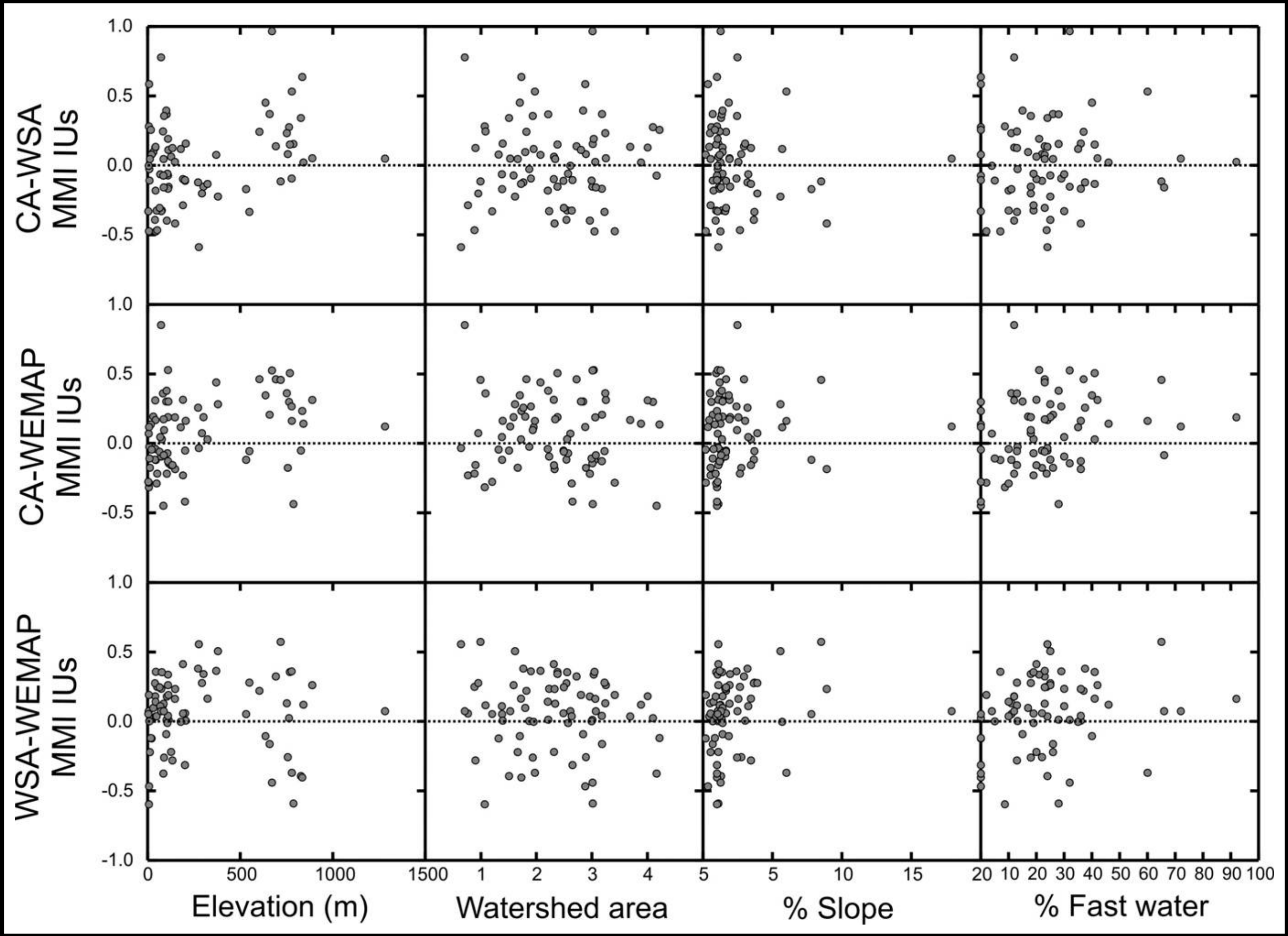
- WSA generally very forgiving (misses 2/3 of impaired sites, few false +)
- WEMAP misses 1/3 of impaired sites, but ~ same number of false + as false -

Do the MMI models give sites the same score?
(correlations with CA Models)



- Much greater agreement than for O/E comparisons
- But inconsistent scoring along condition gradient for WEMAP

MMI Model Bias vs. Environmental Gradients



No bias for these gradients

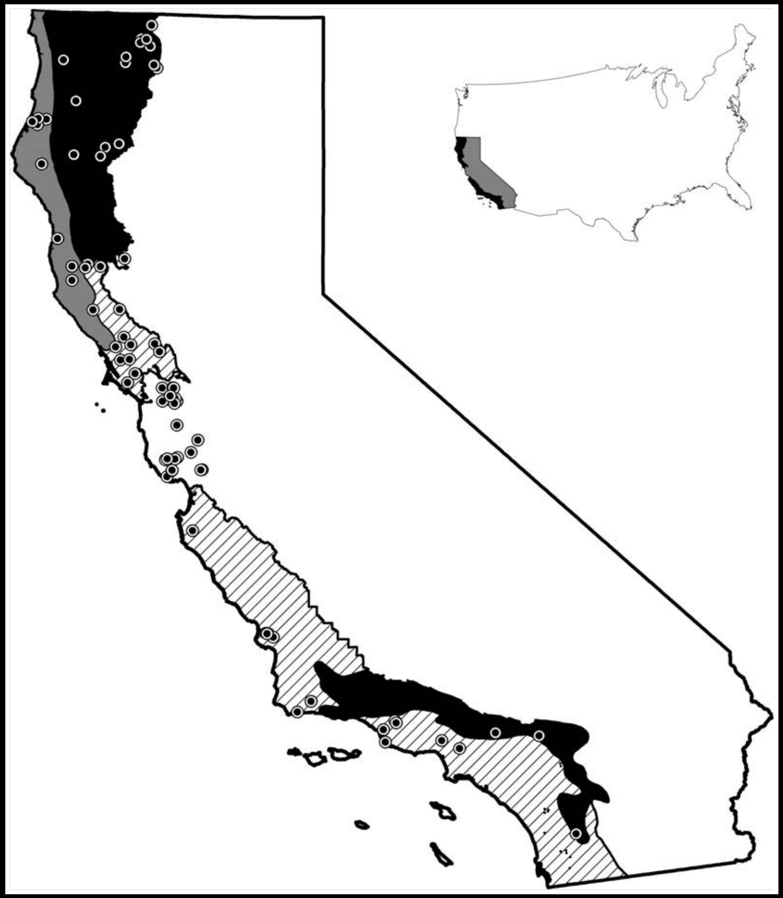
MMI Impairment Decisions:
Impaired (I) or Not Impaired (NI)

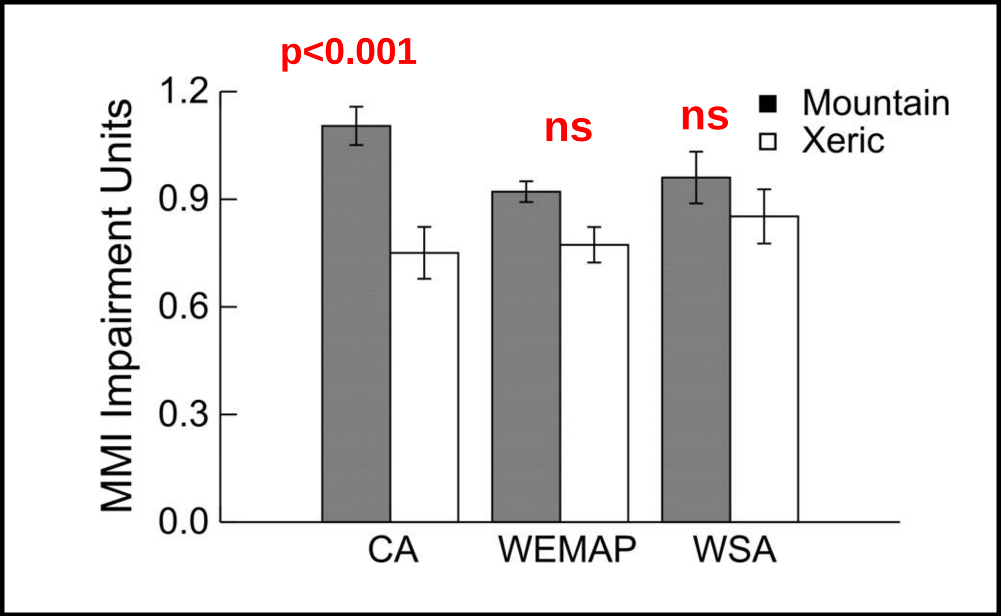
number of **false negatives** and **false positives** relative to CA models

		CA Mountain (n=33)		CA Xeric (n=41)		TOTALS (n=74)		ALL
		I	NI	I	NI	I	NI	
CA ("truth")	I	12		32		44		44 I
	NI		21		9		30	33 NI
WSA	I	11	4	26	1	38	5	42 I
	NI	1	17	6	8	7	25	32 NI
WEMAP	I	11	8	30	1	42	9	50 I
	NI	1	13	2	8	3	21	24 NI

- Both large models had a significant number of disagreements with CA
- WSA similar overall assessment, but WEMAP overestimated impairment

MMI test sites were ~equally divided between mountain (shaded) and xeric (hatched) ecoregions





CA MMIs detected a difference in condition between xeric and mountain sites that neither WEMAP nor WSA picked up

Results Summary

O/E Results

Accuracy:

- Larger models had relatively low agreement with CA models (r^2 0.34-0.39)

Gradients:

- WSA and WEMAP tend to underscore sites relative to CA modles as % slope and % fast-water habitats increased

Impairment Decisions:

- WSA model strongly underestimates impairment, WEMAP (null) slightly underestimates impairment
- WSA tends to overestimate site quality relative to all others (sometimes by a lot)

MMI Results

Accuracy:

- Much greater agreement than for O/E comparisons (r^2 0.60, 0.69)
- But inconsistent scoring along condition gradient for WEMAP

Gradients:

- No bias for the gradients we compared

Impairment Decisions:

- Both large models had a significant number of disagreements with CA
- WSA similar overall assessment, but WEMAP overestimated impairment
- CA MMIs found a difference in condition between Xeric and Mountain sites in the test set that neither WEMAP nor WSA detected

Does it matter?

" it depends on what question you're asking"

Overall Condition Assessments (e.g., 305b reporting)- Accuracy is more important than precision… (we can make up for low precision by looking at large numbers of samples)

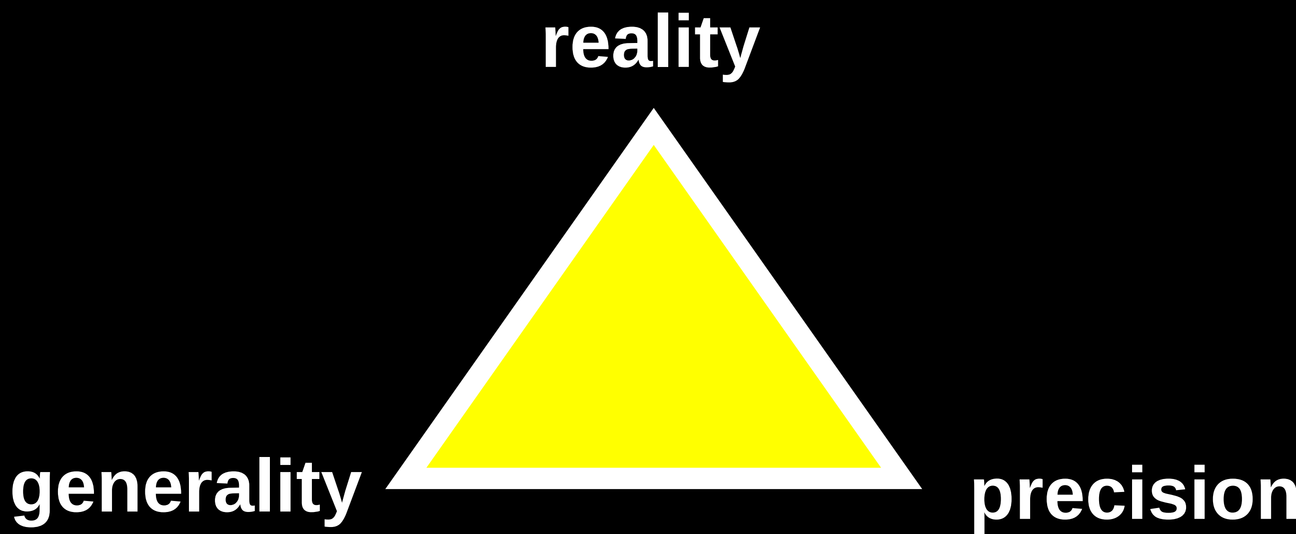
WEMAP O/E model and WSA MMI model might be OK for large assessments, but WSA O/E underestimates impairment and WEMAP MMI overestimates impairment too often

Site Specific Monitoring (i.e., most bioassessment applications) Where both accuracy and precision are important, this is pretty strong evidence that we still need local models …

WSA and WEMAP models are not appropriate because they get it wrong too often

Limits to model optimization

Levins (1966) postulated a triangular relationship among three desirable model traits…



..we can optimize any two of these, but not all three

Thus, as we expand the geographic scope of the models (i.e., increase generality), we have to sacrifice either reality, precision or some of both

Acknowledgements

Field Crews:
- EMAP-West
- USFS/ Utah State University
- CA Aquatic Bioassessment Laboratory (ABL, Shawn McBride, Jennifer York)

Invertebrate Taxonomists:
- EcoAnalysts
- Utah State Bug Lab
- ABL (Doug Post, Dan Pickard, Joe Slusark, Brady Richards)

Funding and Data Sources
- USFS, Joseph Furnish - CA models
- EPA-Office of Research and Development, WEMAP and WSA

Cut?

Predictor variables have a lot of overlap, but each sub-model has unique combinations

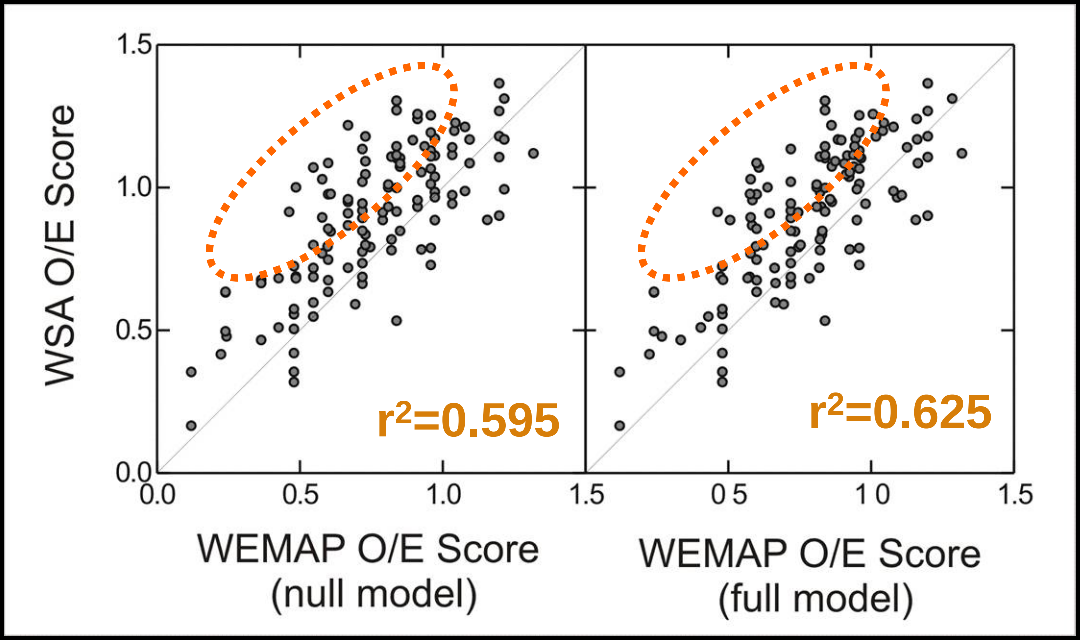
California Predictors	
Group 1	Watershed Area
	Longitude
	Latitude
	Temperature
Group 2	Longitude
	Precipitation
Group 3	Watershed Area
	Temperature

WEMAP Predictors	
Group 2	Watershed Area
	Longitude
	Elevation
	Precipitation
All other CA groups use null models (no predictors)	

WSA Predictors
Watershed Area
Longitude
Day of Year
Min. Temperature
Elevation
Precipitation
Slope

Cut?

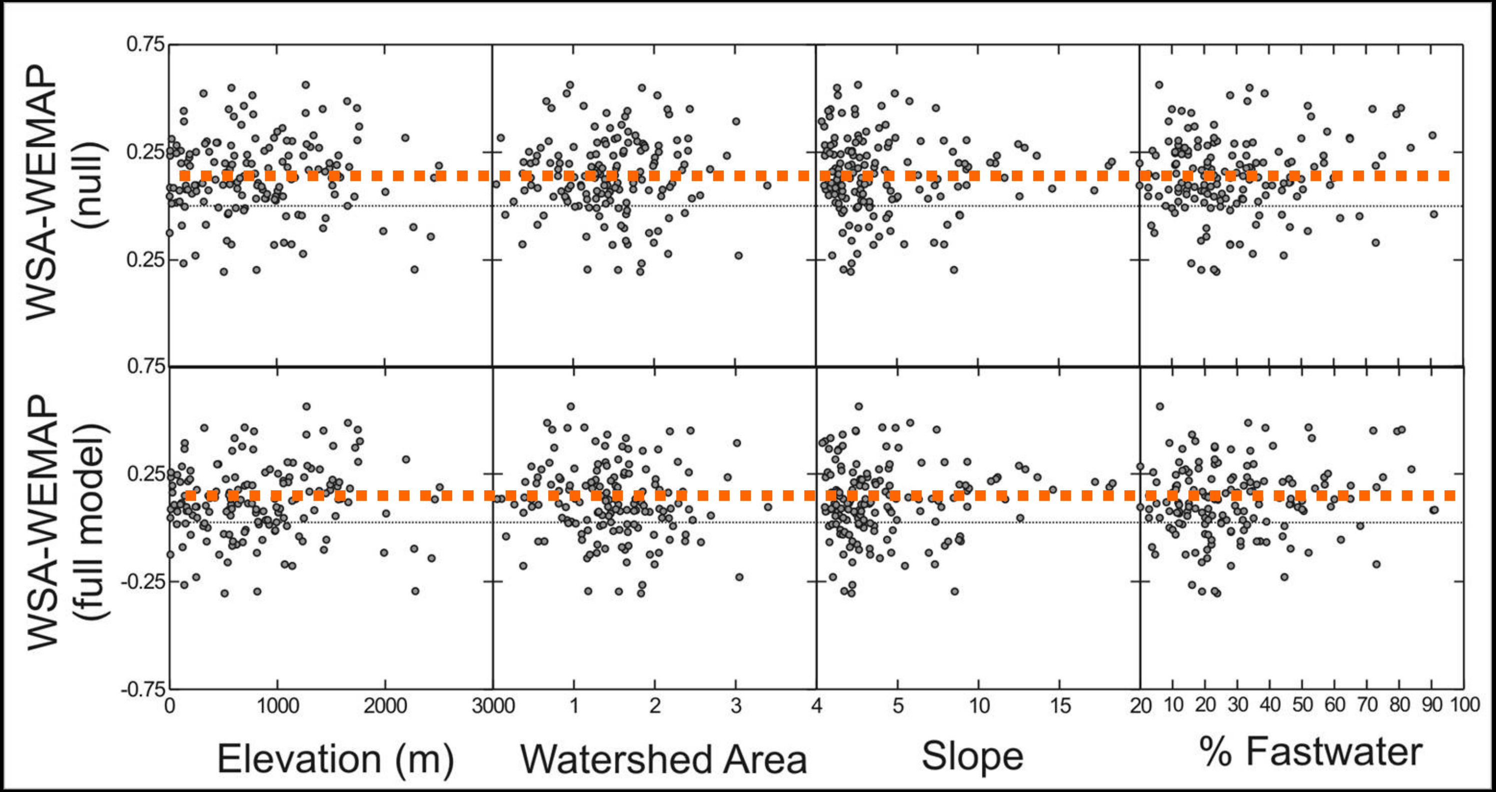
Correlations between the two larger O/E models



Stronger correlations, but prominent bias in WSA model

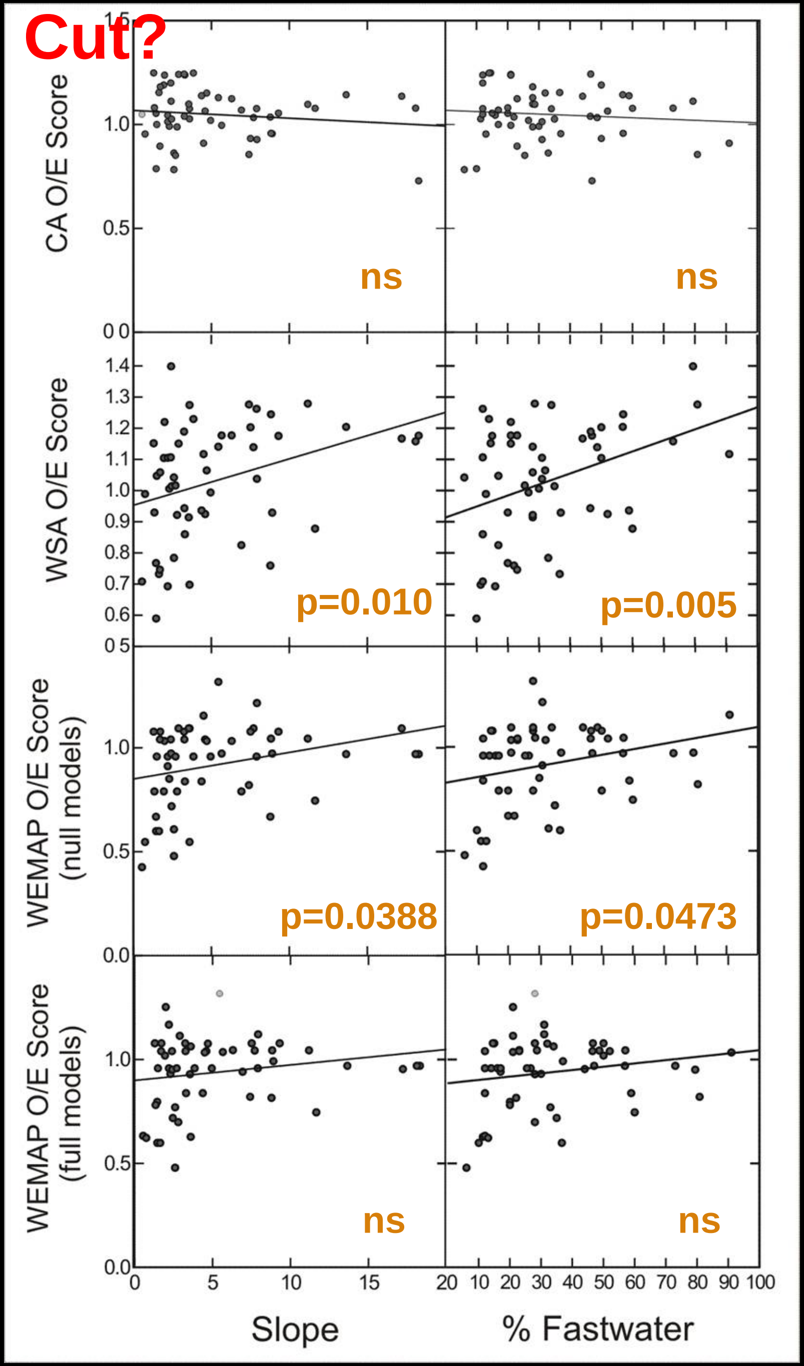
Cut?

Model Bias vs. Environmental Gradients (large models)



WSA model tends to score sites ~.15 units higher than WEMAP models, but no bias with gradients…

… are both large models failing to account for these gradients or is CA getting it wrong?



O/E scores for **reference sites** versus environmental gradients under the 4 models

(expect a flat line if models account for gradients)

- CA models clearly not affected by these gradients

- Both WSA and WEMAP (null, but not full) had significant relationships with % slope and % fast water that were not accounted for by the models

Cut?

O/E Results: “how sensitive are the models?” (standard deviations)

California		WEMAP (null)		WEMAP (full)		WSA	
Group 1	**0.13**	Group 1	0.388	Group 1	0.243	**Western**	**0.198**
Group 2	**0.17**	**Group 2**	**0.20**	**Group 2**	**0.15**		
Group 3	**0.16**	**Group 3**	**0.20**	**Group 3***	**0.20**		
		Group 4	**0.22**	**Group 4***	**0.22**		
		Group 5	**0.17**	**Group 5***	**0.17**		

- California models are more precise than either WEMAP or WSA
- WEMAP (full) slightly more precise than WEMAP (null) or WSA

Predictors (the major environmental gradients) are a key to understanding model differences

1. Models vary both in the specific predictors and in the geographic range of the predictor gradients….

2. Predictors that work for larger geographic areas may miss or under-represent regionally important gradients…

3. Variation in predictor association within a taxonomic group (e.g., species within a genus) that occurs across the geographic range of a model can also influence model accuracy and precision...

… do these factors affect performance?

Conclusions:

- Site-specific monitoring (i.e., most bioassessment applications) will require models that account for locally important gradients

- More work needed to determine optimal scale for model development and which environmental factors influence this scale