

Development, Evaluation, and Application of a RIVPACS-type Predictive Model for Assessing the Biological Condition of Streams in Region 5 (California) National Forests

Charles Hawkins

Western Center for Monitoring and Assessment of
Freshwater Ecosystems

Utah State University

Joseph Furnish, Ann Carlson, and Ken Roby

USDA Region 5 Forest Service

Lisa Lackey, USFS- Geometronics GIS

Development of a RIVPACS Predictive Model

- Hawkins et al. 2000. Ecological Applications
- 1989 - Are aquatic macroinvertebrates sensitive to timber harvest activities (CWEs)?
- Describes model development & the effectiveness of the predictive model in distinguishing reference vs test sites
- Sierra Nevada & Klamath provinces
- 261 Reference sites & 281 test (impacted) sites

Independent Variables

Variable Means→	Sierra Nevada		Klamath		Ref. vs. Test (Sig.)
	Ref.	Test	Ref.	Test	
Elevation (meters)	1832	1680	818	833	<0.05
Basin Size (acres)	7635	8400	7685	6968	ns
Channel Gradient	4.4	4.6	4.3	5.7	<0.02
Conduct- ivity	86.5	70.7	160.4	183.4	ns

Dependent Variables

Variable Means→	Sierra Nevada		Klamath		Ref. vs. Test (Sig.)
	Ref.	Test	Ref.	Test	
% of Basin Logged	1.3	28.1	1.9	19.0	<0.00001
Stream-road Crossings	20.8	36.9	12.6	26.6	<0.00001
Length of roads (km)	28.8	67.6	27.9	58.3	<0.00001
PO ₄ (ppm)	0.0035	0.0046	0.0036	0.0048	<0.00001

Predictive vs. Multimetric Model

Model Means→	Sierra Nevada		Klamath		Ref. vs. Test (Sig.)
	Ref.	Test	Ref.	Test	
Observed/ Expected (O/E)	1.009	0.900	0.987	0.856	<0.00001
MM-Taxa Richness	30.6	32.7	34.1	34.4	NS
MM-EPT Richness	17.7	19.6	20.0	21.3	<0.002
Shredders (%)	4.8	6.9	7.4	8.0	<0.01

Quick Review of Predictive Models

- Predictive models base assessments on the compositional similarity between observed (O) and expected (E) biota.
- Taxa richness is the unit of measure.

O/E: a Simplified Expression of a Multivariate World

- E is the *number* of native taxa expected to occur at a site in the absence of human-caused stress.
- O is 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 probabilities of capture under natural conditions > than a specified threshold.

O/E has Several Useful Properties as a Measure of Biological Condition

- O/E has an intuitive biological meaning and is interpretable by researchers, managers, stakeholders, and policy makers.
- O/E means the same thing everywhere.
- Its derivation and interpretation are independent of type and knowledge of stressors in a region, i.e., test sites are not used to calibrate the model.
- O/E is quantitative with known error.

O/E is simple to calculate once the probability of capturing each taxon at a site is known.

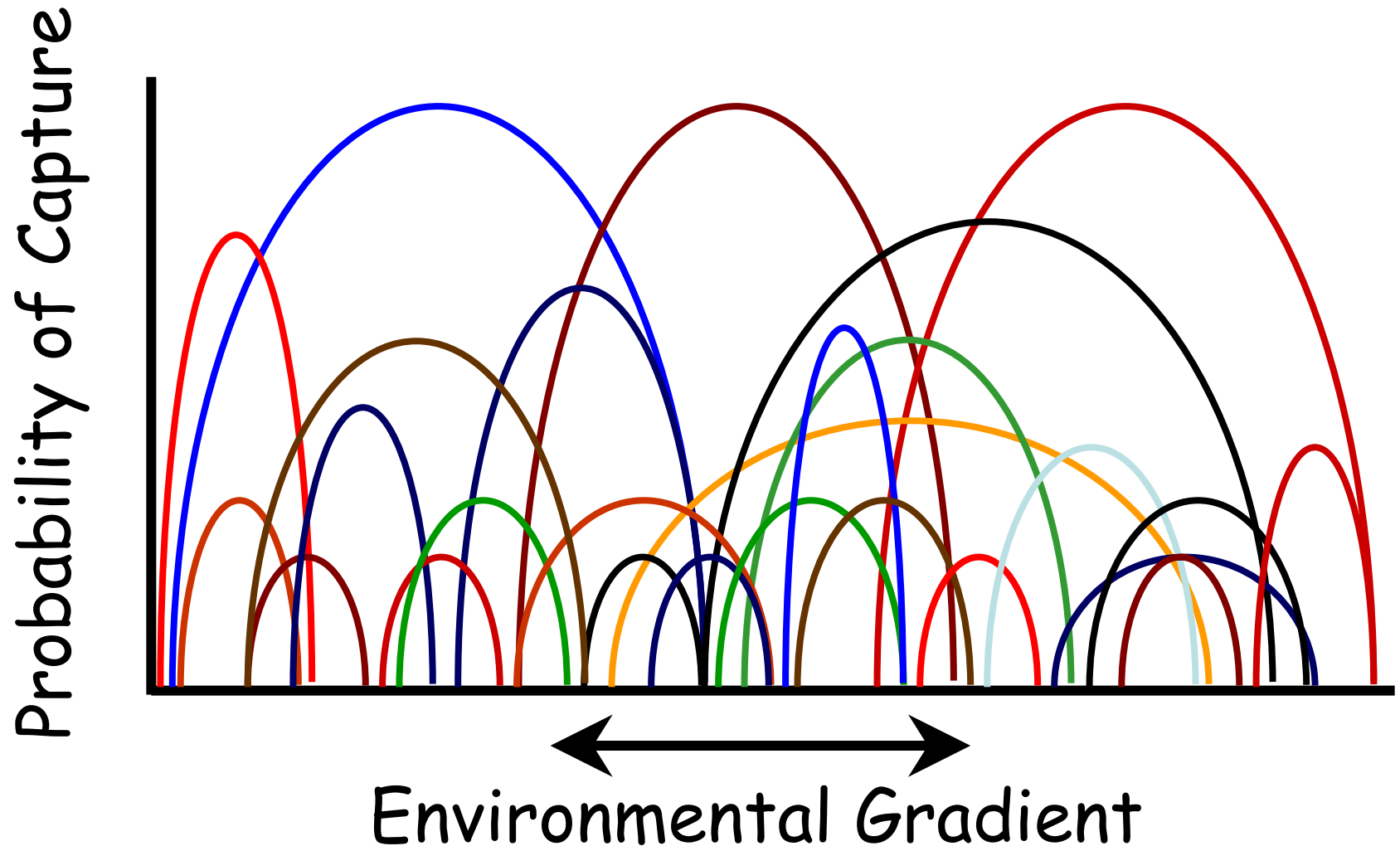
$$E = \sum P_c$$

A threshold P_c can be set to any desired value. Past research has shown that a P_c threshold value of 0.5 results in good precision and sensitivity.

Site 1		
Taxon	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

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

P_c estimates are based on empirical relationships between the distribution of taxa and naturally occurring environmental features (e.g., elevation, stream size, basin geology, etc.) that are not affected by human activity.



Details regarding the construction and application of RIVPACS-type models are available in the literature.

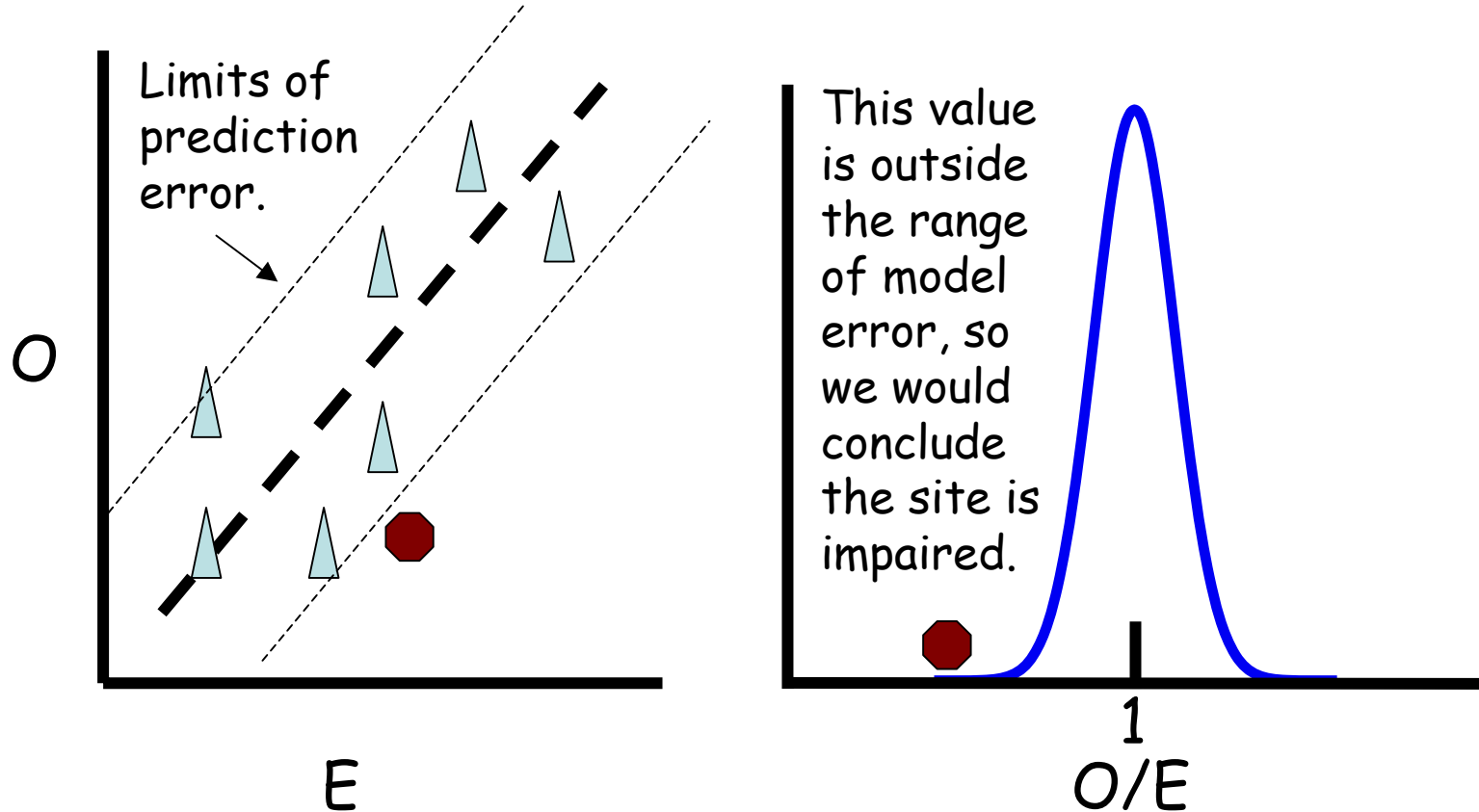
E-mail chuck.hawkins@usu.edu for a list of literature.

Quantifying prediction error is an important aspect of RIVPACS assessments.

For reference sites:

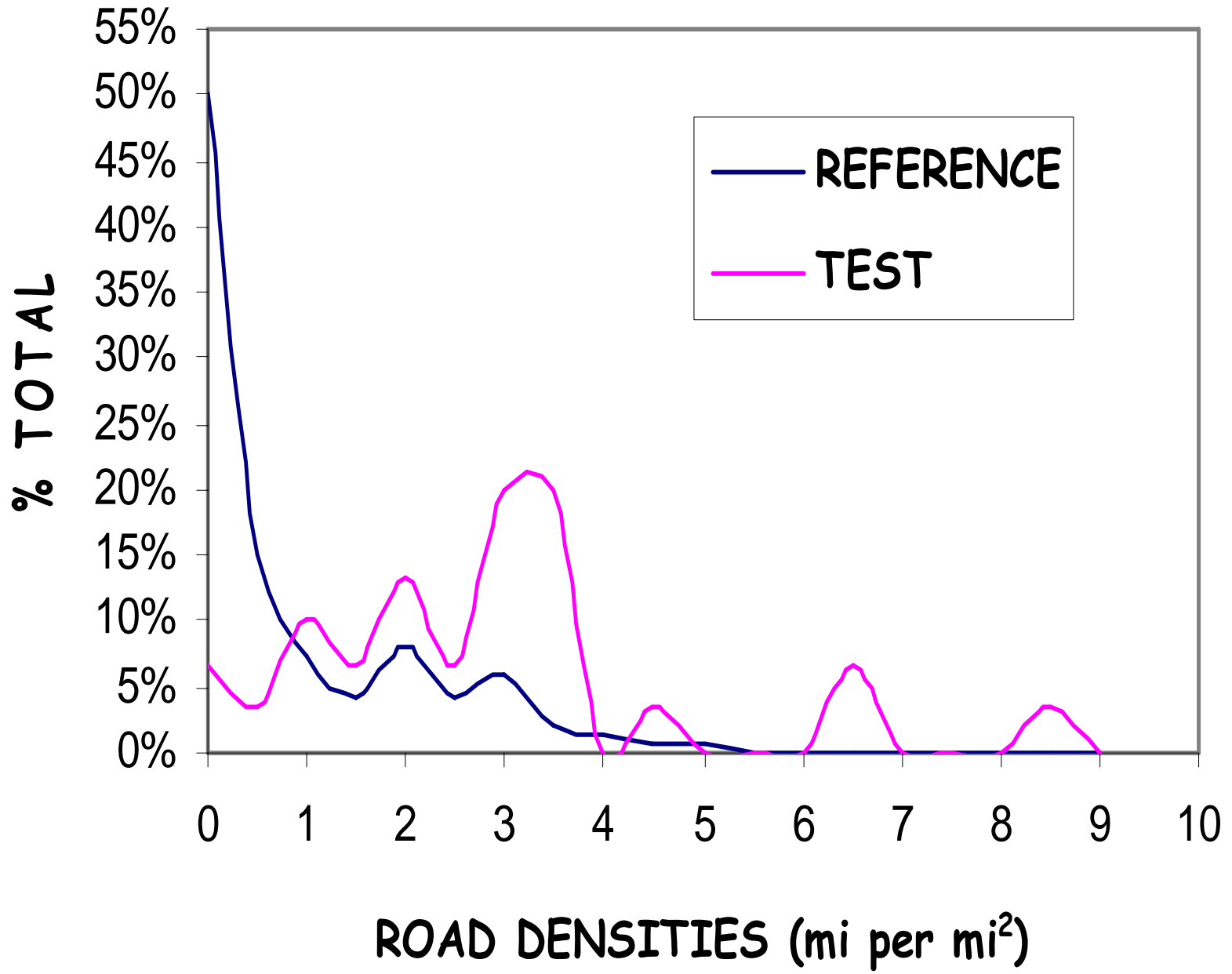
- O should be strongly correlated with E and the relationship should have a slope of 1 (accurate).
- O/E values derived from reference sites should have a small standard deviation (precise).

Model error is used to statistically determine if a new observation is either equivalent to reference or impaired.



The R5 Model

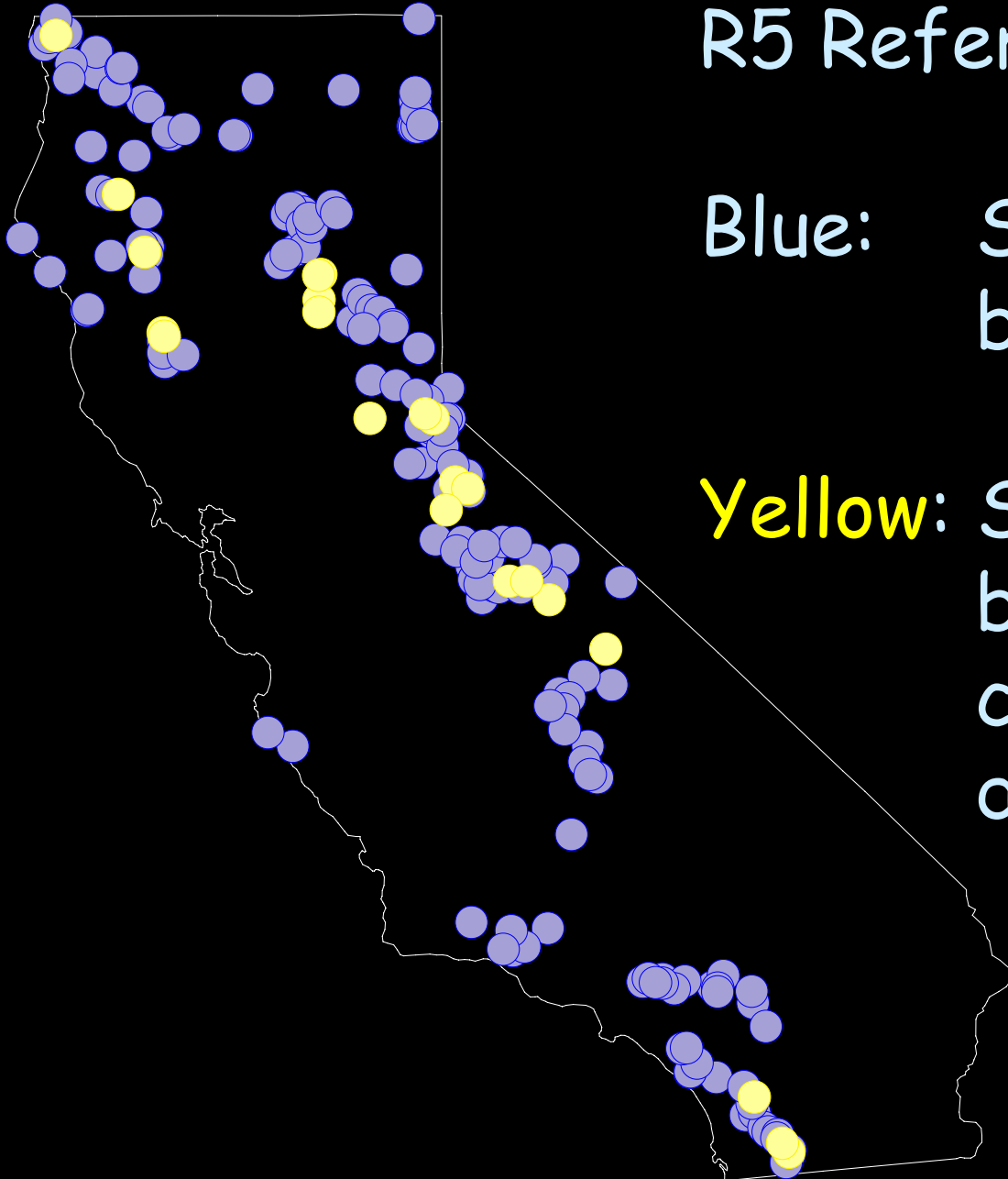
- 198 Reference Sites:
 - 176 sites used to build the model.
 - 24 sites were excluded because of low counts (<300 individuals) or missing data.
- 180 taxa used in the model.
- The model was applied to 63 test sites.

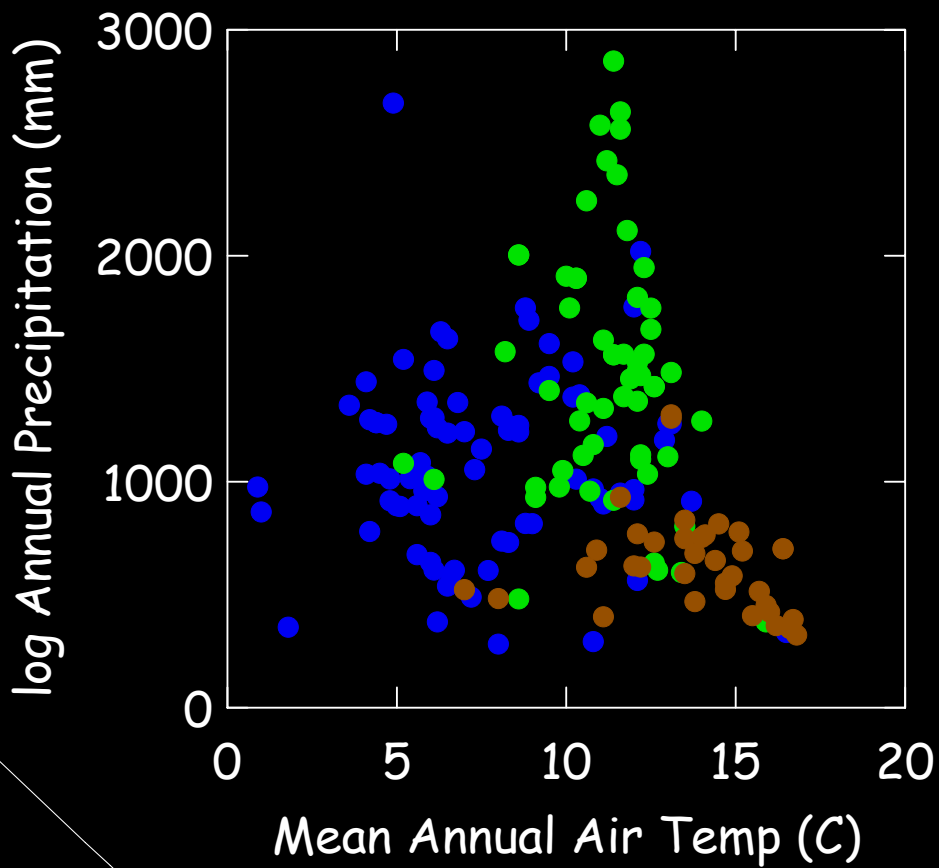
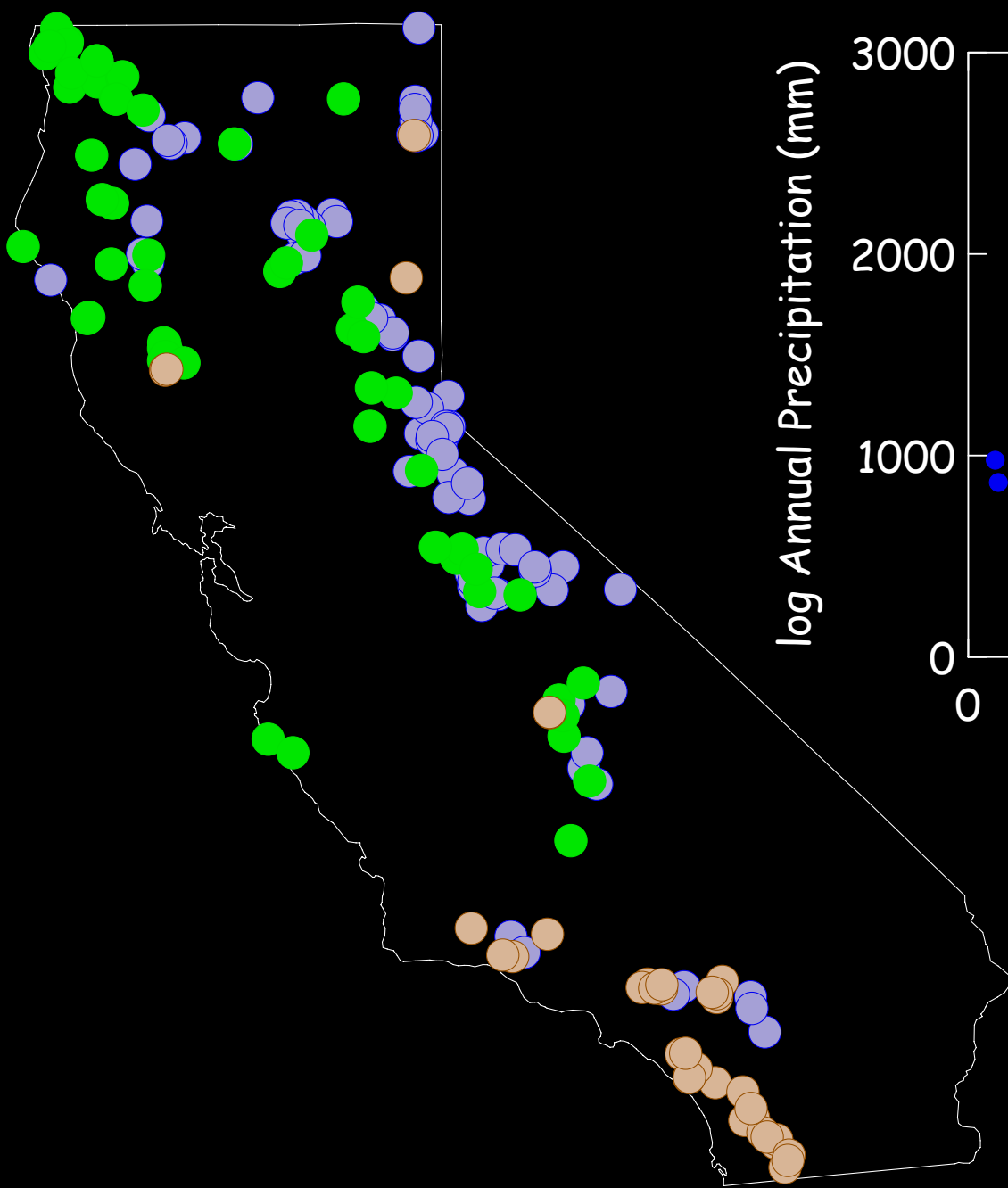


R5 Reference Sites:

Blue: Sites used to build the model.

Yellow: Sites excluded because of low counts or missing data.





Biotic composition among reference sites was strongly related to climatic variation. Sites classified into 3 biologically-defined groups clustered by temperature and precipitation.

Model Predictors

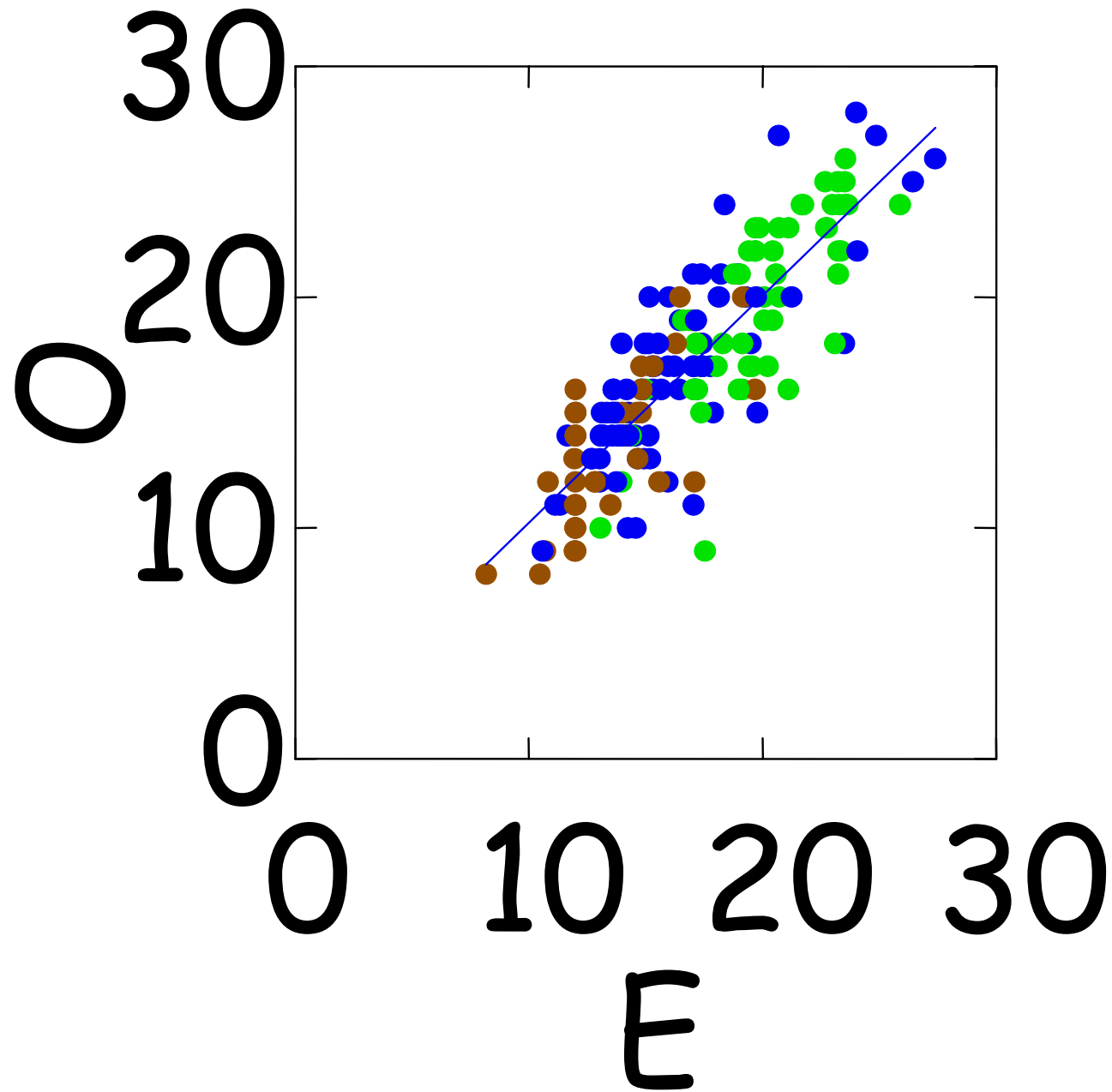
(in order of relative importance)

<u>Variable</u>	<u>Units</u>
Elevation	meters
Wet Days*	number
Latitude	decimal degrees
Sampling Date	days past 1 Jan
Watershed Area	log m ²
Annual Precipitation*	log millimeters
Longitude	decimal degrees
Alkalinity	sqrt ppm

* *Obtained from PRISM climate database*

The $P_c > 0.5$ model accounted for 72% of the variation in taxa richness among reference sites. Slope = 0.98.

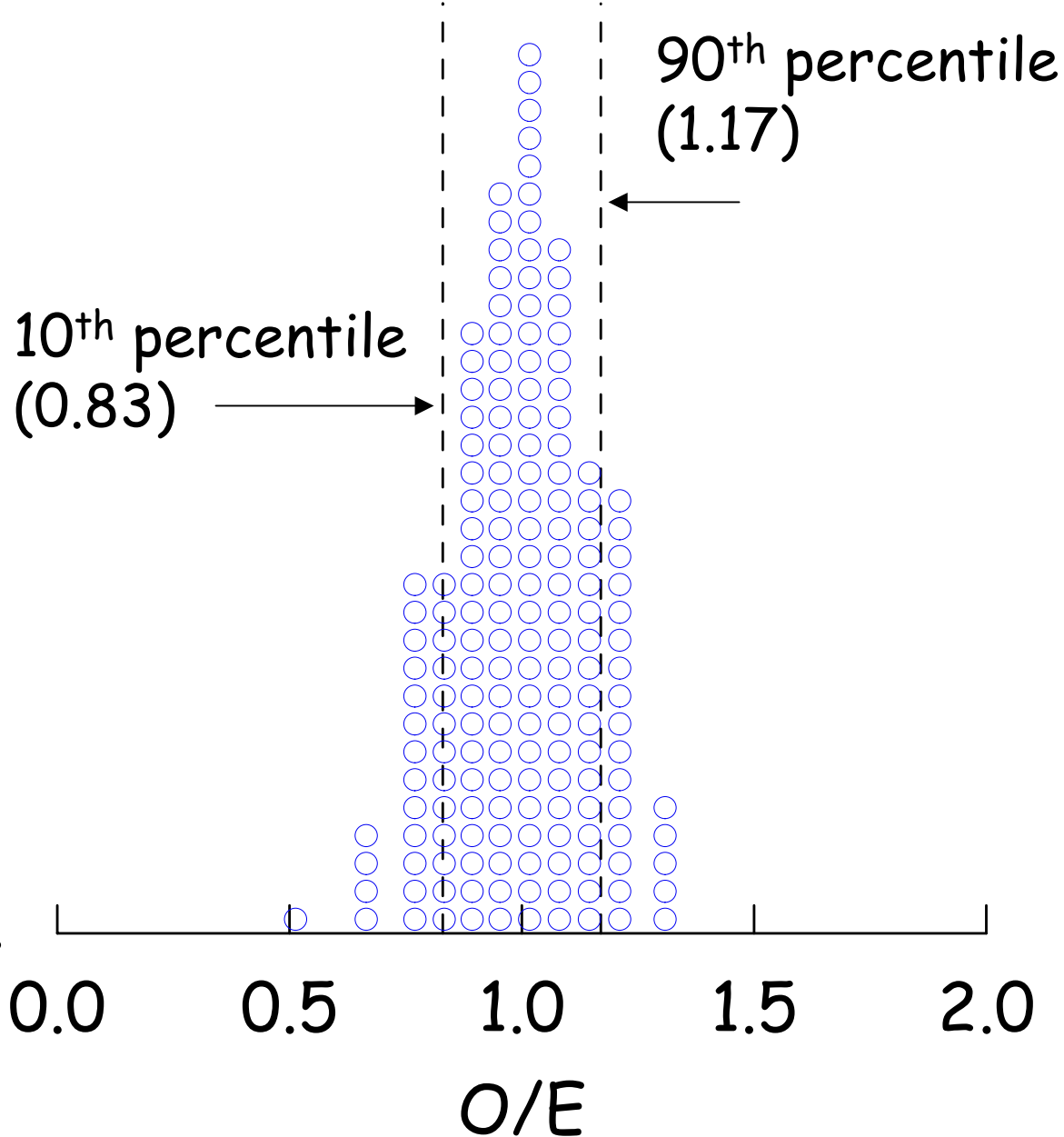
Note that the model accounted for considerable variation in richness within classes of sites.



The model was accurate (mean = 1.01) and more precise (SD = 0.14) than many models.

10th and 90th percentiles are often used as thresholds to infer impairment.

Frequency Distribution of Reference Site O/E values

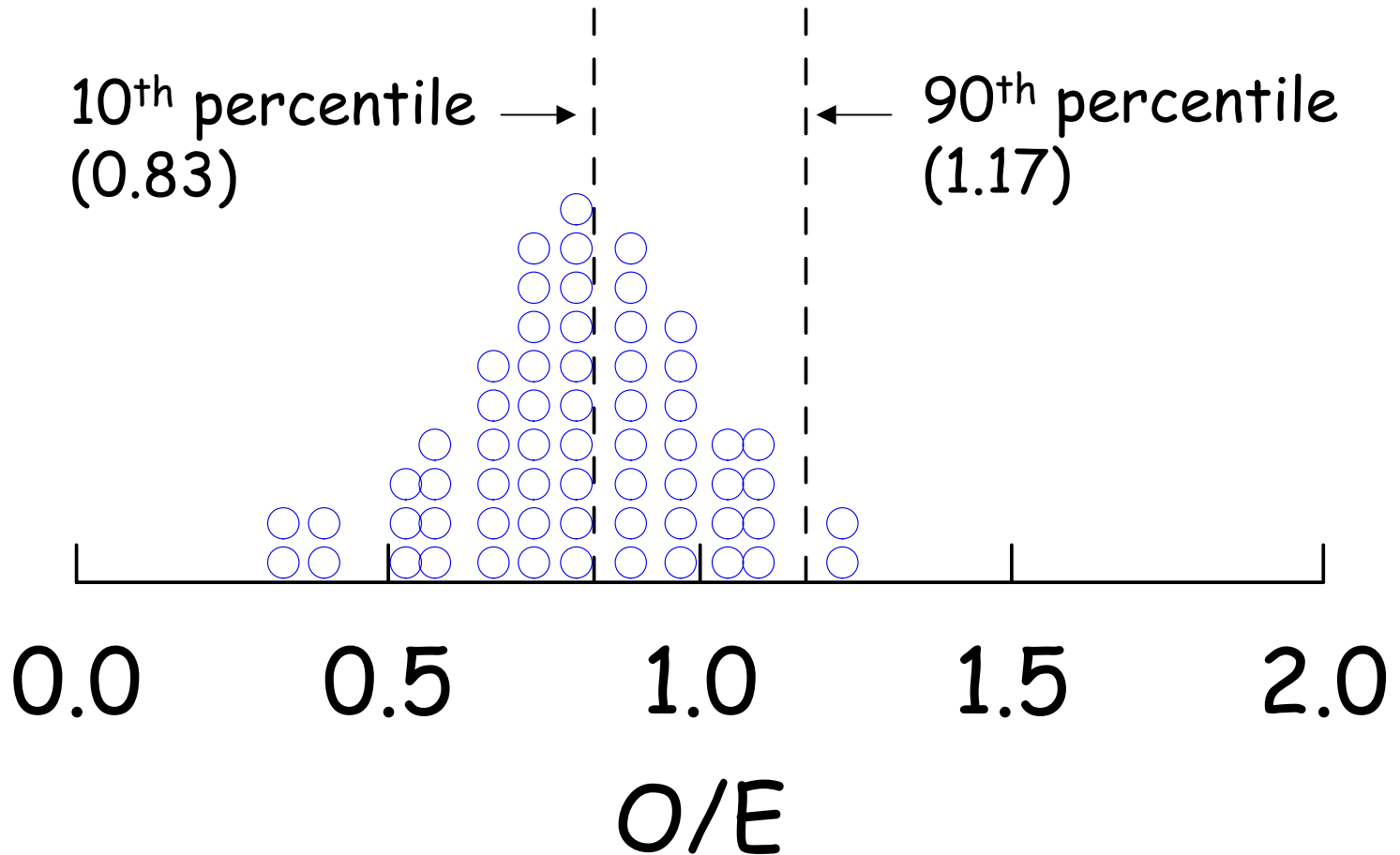


Frequency Distribution of Test Site O/E Values

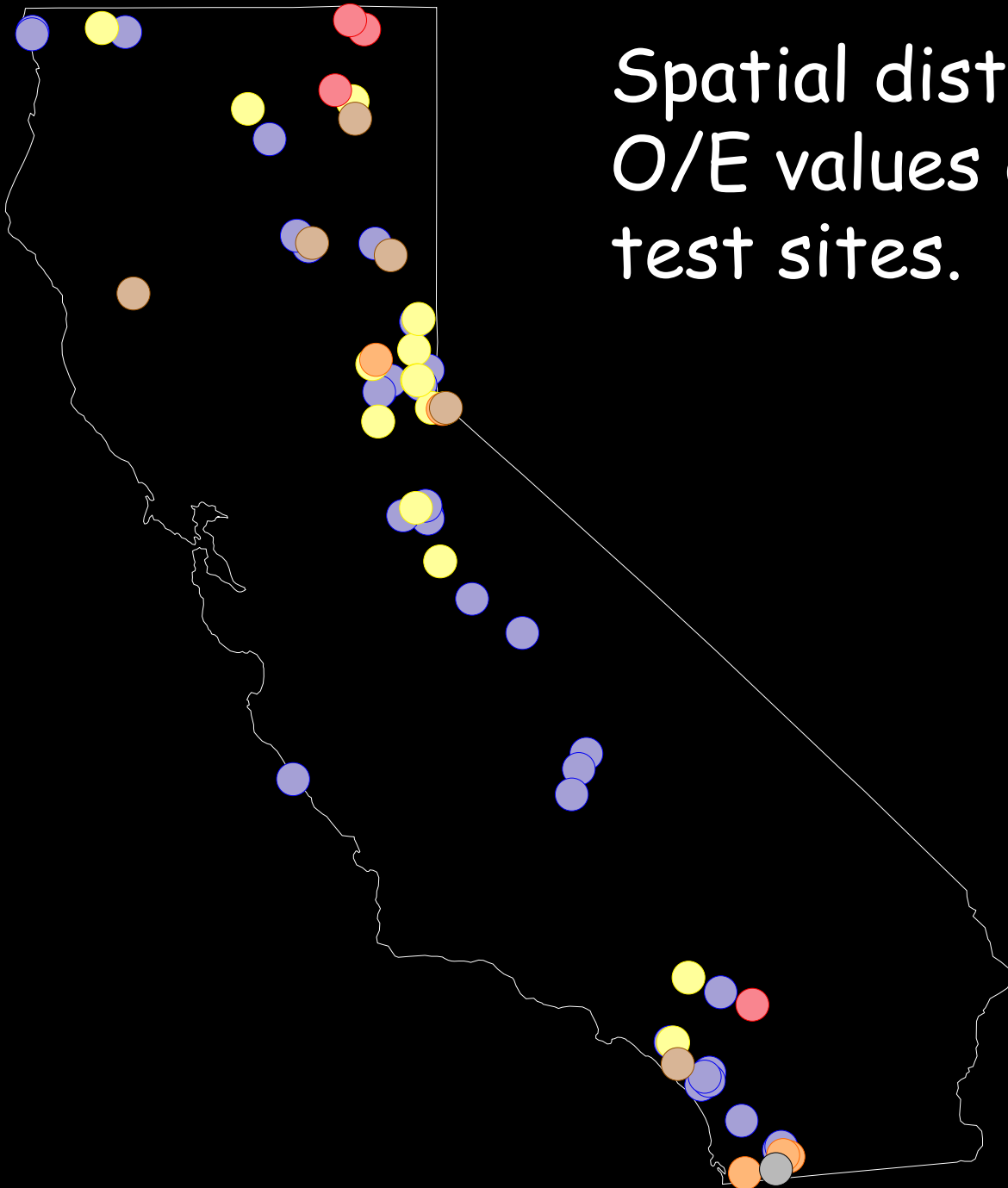
Mean = 0.81.

Range = 0.33 - 1.25.

60% of test sites were not reference quality based on 10th and 90th percentile thresholds.



Spatial distribution of
O/E values observed at
test sites.



● > 0.81

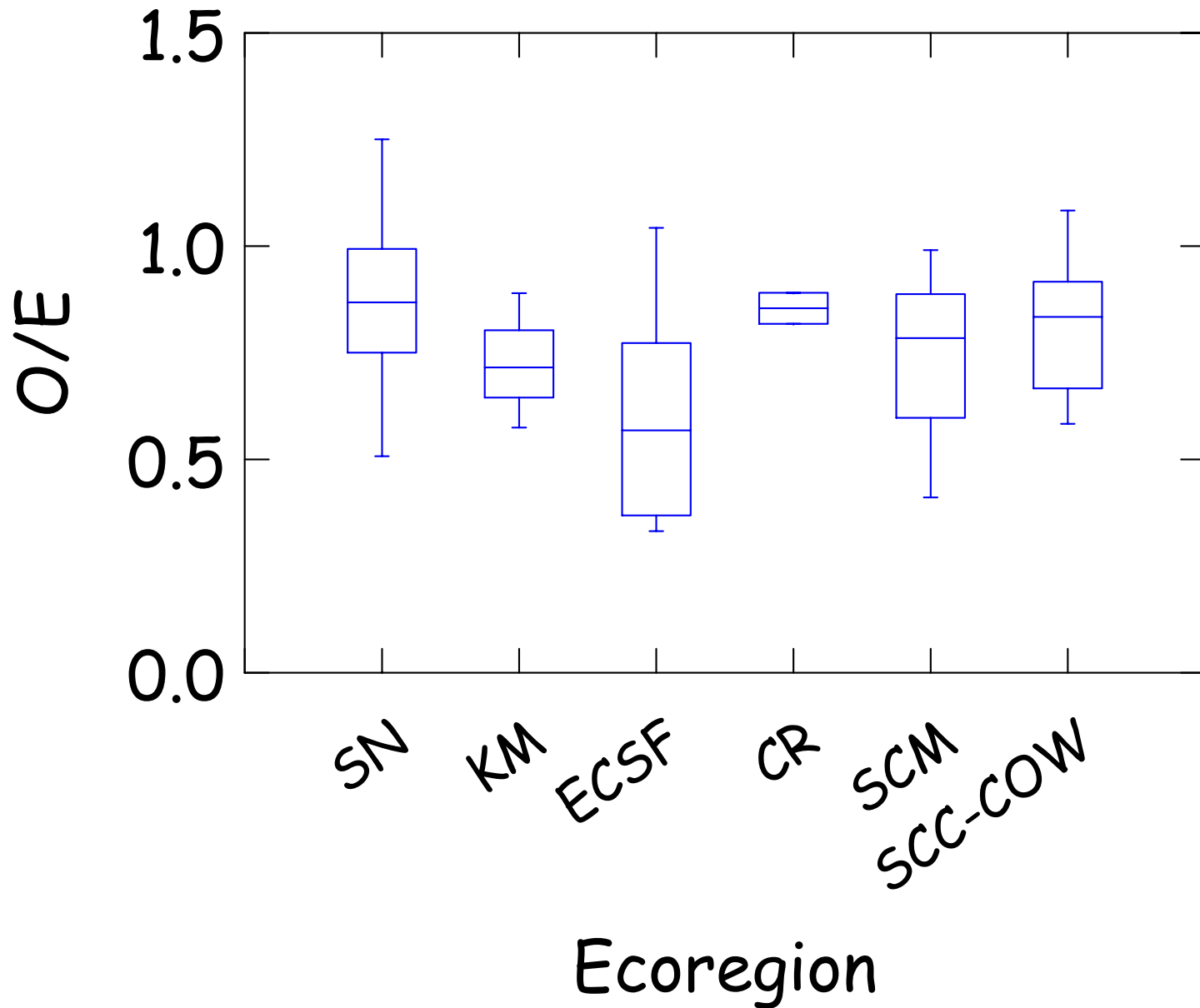
● $0.7 - 0.81$

● $0.6 - 0.7$

● $0.5 - 0.6$

● < 0.5

Distribution of O/E values at test sites by ecoregion



Model Output Can Also be Used to Identify Taxa that Decrease or Increase Relative to Expectations

Example based on the 63 R5 Test Sites

Decreasing Taxa (ranked)

Oroperla (Perlodidae)
Arctopsyche (Hydropsychidae)
Dolophilodes (Philopotamidae)
Serratella (Ephemerellidae)
Ordobrevia (Elmidae)
Rhithrogena (Heptageniidae)
Ampumixis (Elmidae)
Peltoperlidae
Doroneuria (Perlidae)
Perlinodes
Drunella doddsi
Ameletus

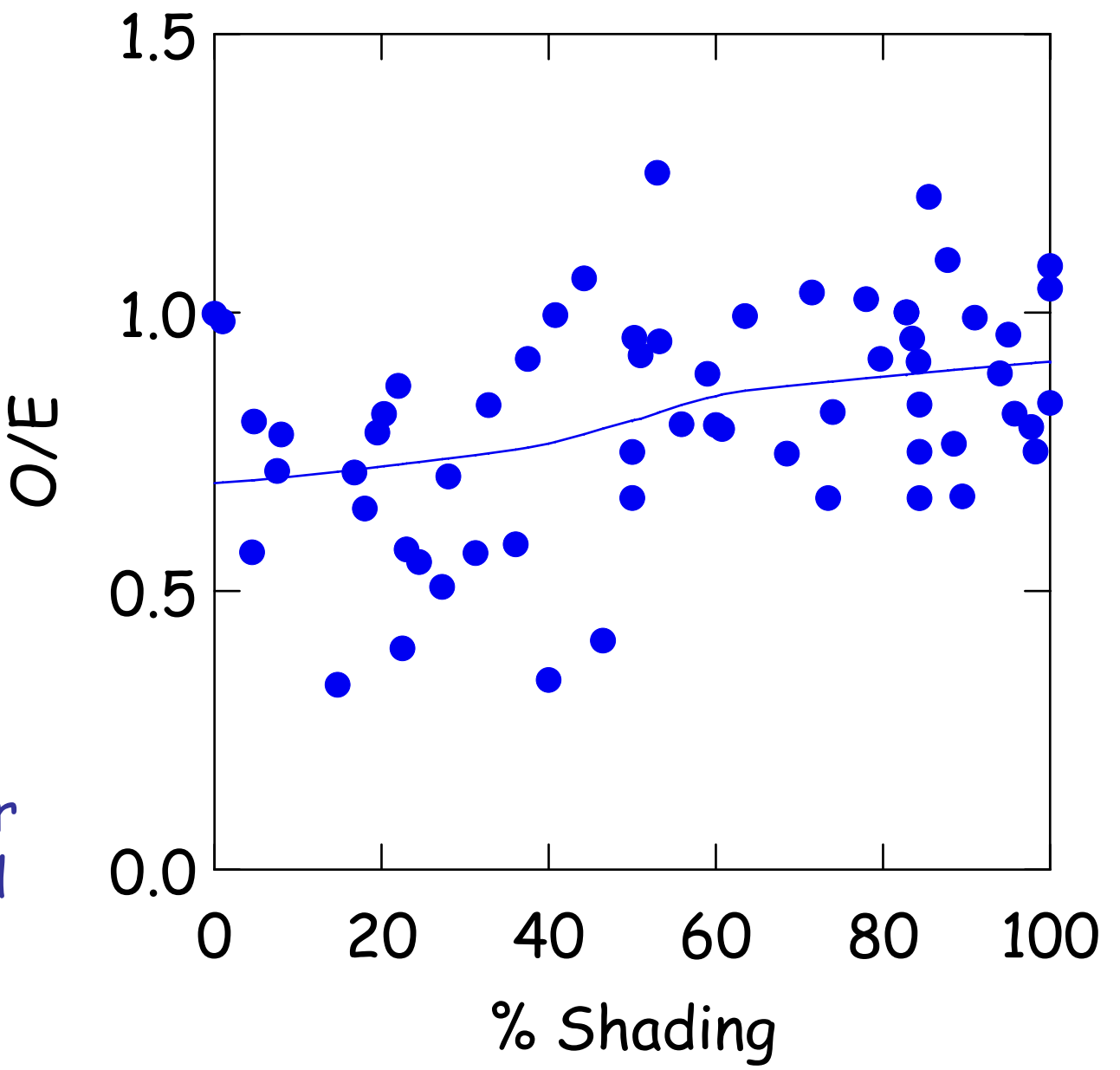
Increasing Taxa (ranked)

Tricorythodes
Tipula
Coenagrionidae
Helicopsyche
Argia
Hydroptila
Cheumatopsyche
(Hydropsychidae)
Pisidium (Pea clam)
Pericoma (Psychodidae)
Gomphidae
Physella (FW pulmonate snail)
Dixa

An example of the response in O/E to riparian condition.

O/E at test sites generally increased as riparian shading increased.

Other stressors also influenced O/E, but stressor data were limited in this data set.



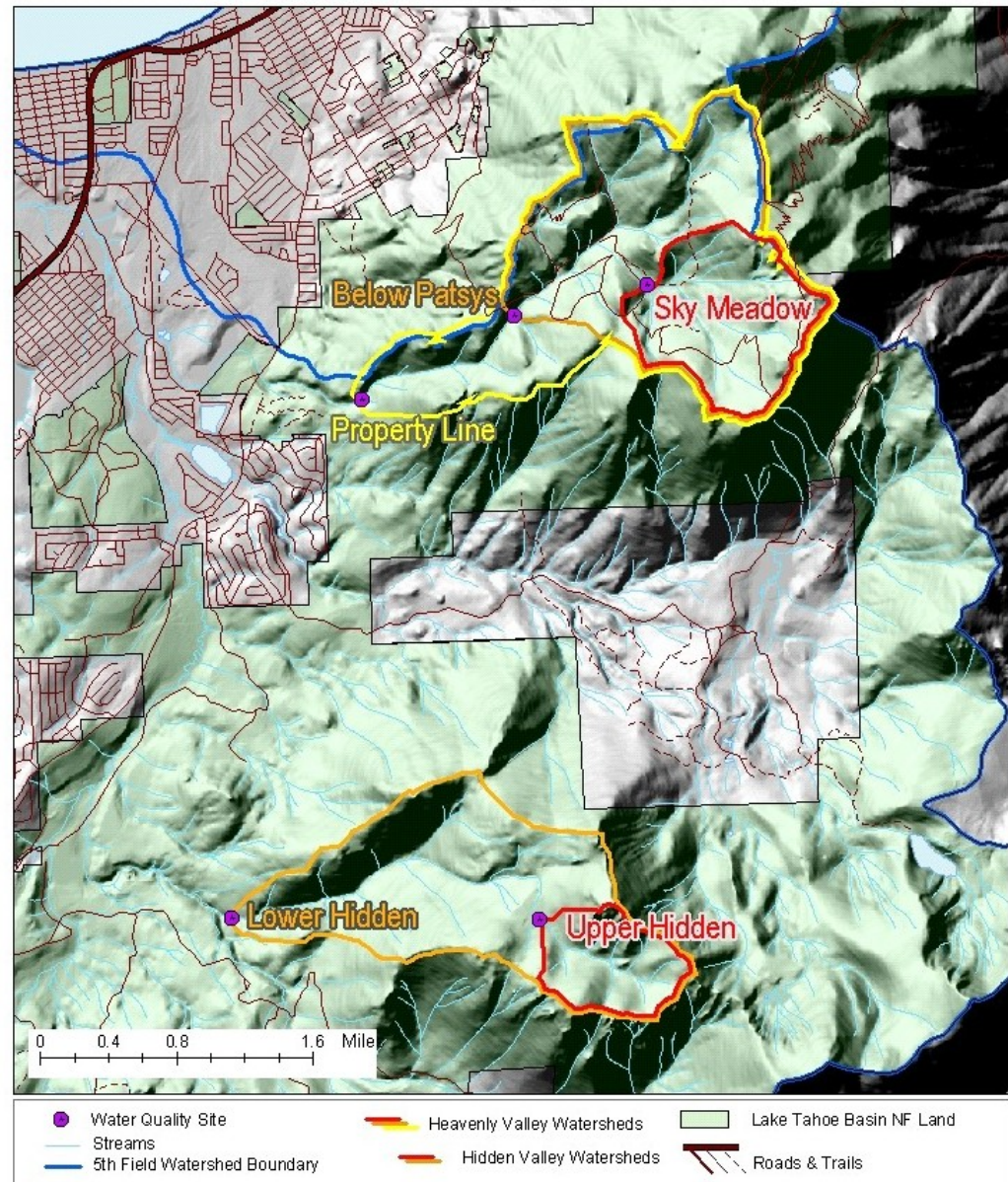
Using the Model

- The model is undergoing final evaluation at Utah State University.
- Use will require input of 2 ASCII data files:
 - Taxa by Sample file
 - Predictors by Sample file
- Model output includes:
 - Site check (Is the site within the model's range?).
 - O/E values based on P_c 's of 0.0 and 0.5.
 - P_c 's for all taxa at all sites.
 - Expected and observed number of sites for all taxa for the set of sites submitted to the model.

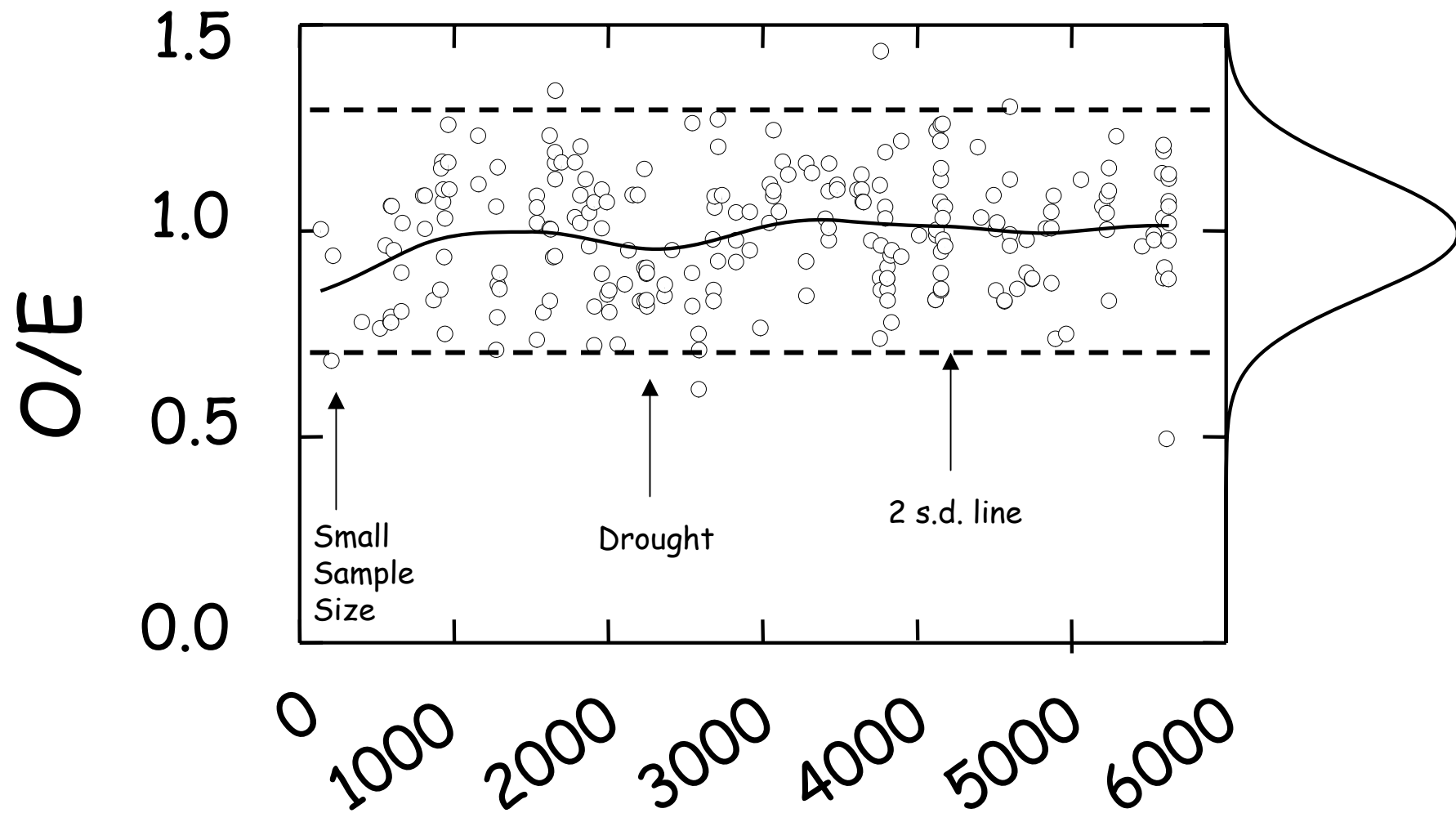
Results of RIVPACS Analysis for Heavenly & Hidden sites

- | <u>Reference Site</u>
<u>Scores</u> | <u>Heavenly Valley</u>
<u>TMDL Site</u>
<u>Scores</u> |
|---|---|
| • Lower Hidden = 0.97 | • Property Line = 0.56 |
| • No paired site. | • Below Patsy's = 0.57 |
| • Upper Hidden
(Intermittent) = 0.45 | • Sky Meadows = 0.41 |

Heavenly Valley TMDL Bioassessment Monitoring: 3 sites in Heavenly Valley and a reference watershed in Hidden Valley, with similar size, aspect and elevations.



Invertebrate Data from North Carolina Reference Sites



Days Since 1 January 1983
From Hawkins 5/2001

Data Files

(details regarding file construction will be available on our web site soon)

○ Predictor variable data matrix:

- Samples as rows and variables as columns.
- Variable names must be spelled exactly as used in the model.
- Units of measure must be those used in the model.

○ Taxa data matrix:

- Samples as rows and taxa as columns.
- Taxa must be collected with R5 or compatible field methods.
- Taxonomic resolution must be consistent with that used in the model.

The R5 Model currently runs on the Western Assessment Center's server.
Access requires a username and password.

Welcome!

Please enter your username and password to log in.

Username:

Password:

LOGIN

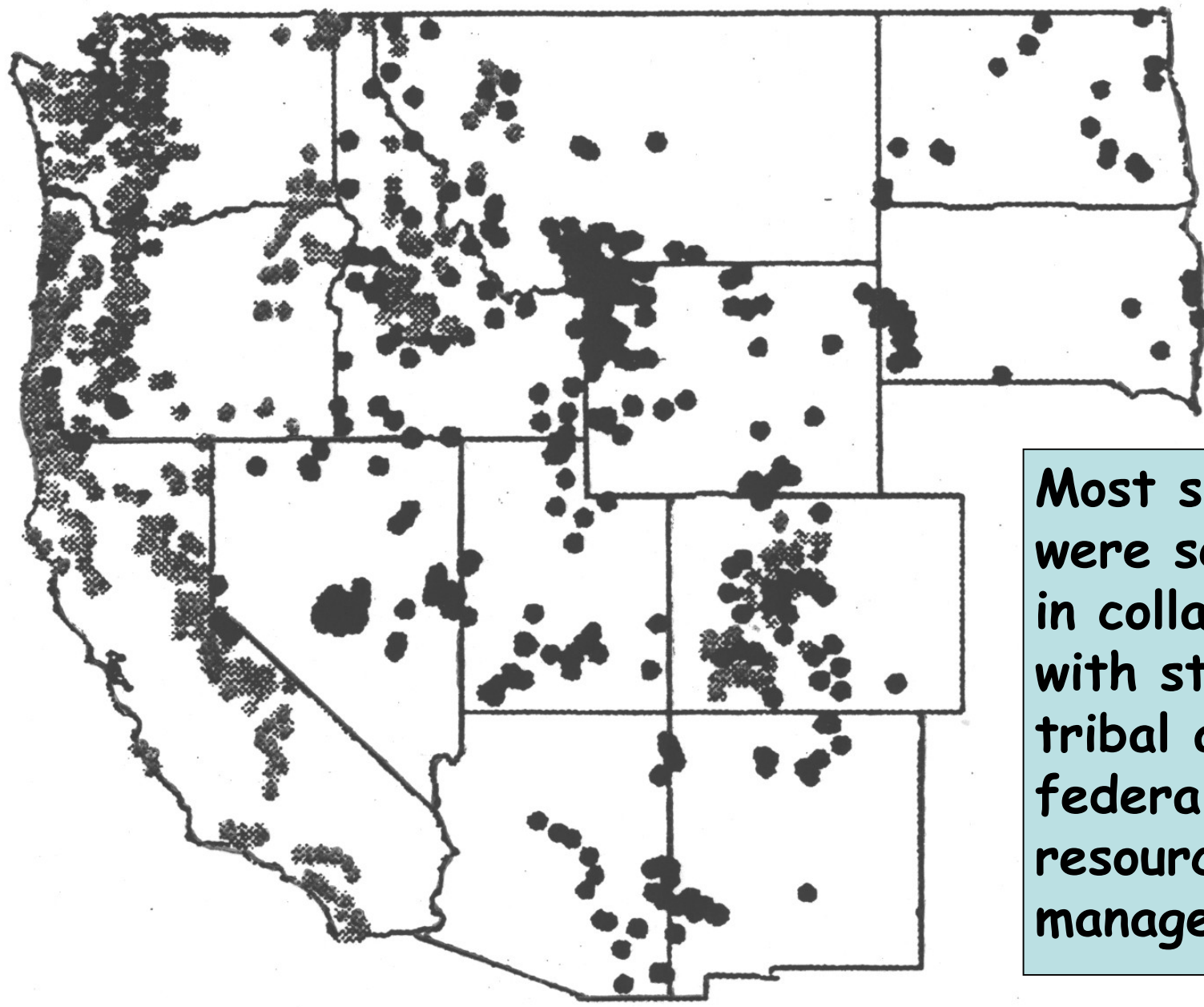
Please logout and close the browser window when you finish,
to prevent others from using any of your submitted information.

Conclusions

- The current R5 model is among the best performing models constructed to date.
- It is applicable to many stream conditions throughout California:
 - 32.7 to 42.0 latitude.
 - -124.3 to -116.4 longitude.
 - 20 to 149 wet days per year.
 - 40 to 7800 mm of precipitation per year.
 - 37 to 3000 m in elevation.
 - 17.5 to 5,160 km² watershed area.
 - 134 to 288 day of the year.
 - 1 to 300 ppm alkalinity.
- It can be expanded to be applicable to streams and sampling dates outside of this range with the addition of new reference sites sampled in new regions and on earlier or later dates.

The locations of 1064 "reference" sites.

From Chuck Hawkins, Utah State University, 11/2003



Most sites were selected in collaboration with state, tribal and federal water resource managers.