An Index of Biotic Integrity for the Eastern Sierra

David Herbst

Sierra Nevada Aquatic Research Laboratory, Mammoth Lakes
University of California, Santa Barbara

CONTRIBUTORS:

Erik Silldorff, Jeff Kane, Tom Suk

Preliminary steps to developing biological standards

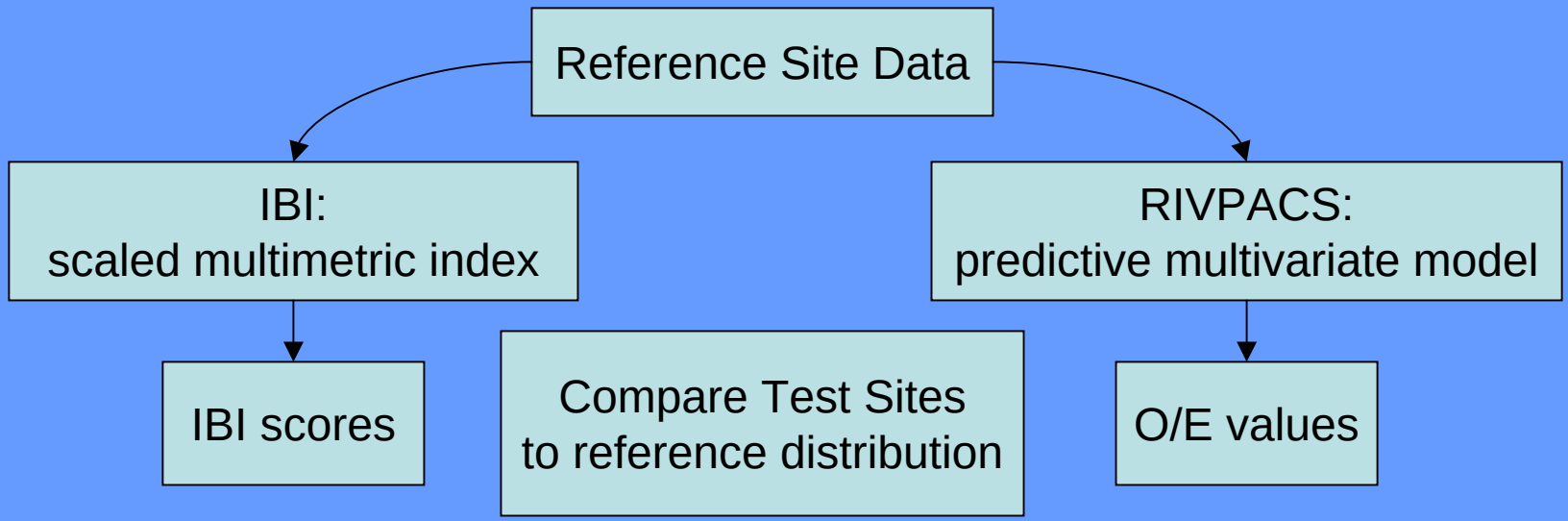
Methods Comparison of CSBP, USFS.R5, UC-SNARL: Field, lab and analysis differences showed compatible results

Adopt Targeted Riffle as standard approach
(USFS 8-sample composite)
[include Multi-Habitat as option]

Conversion of data sets
to uniform standard

Assemble QCed field/lab **data for analysis** and incorporate **inter-annual** repeat sampling and **intra-site** spatial replication variability measures

Overview of analyses and continuing goals



Examine options for impairment thresholds:
below reference range, below R_{mean}-2SD,
below Type I-II balance, sectioning of range…

Eventual Goal: combine with other data sets and regionalize to achieve best sensitivity in assessing biological integrity

An iterative process of refinement and integration of data and analytical tools

Next: combine Sierra Nevada data sets with USFS

Lahontan Region

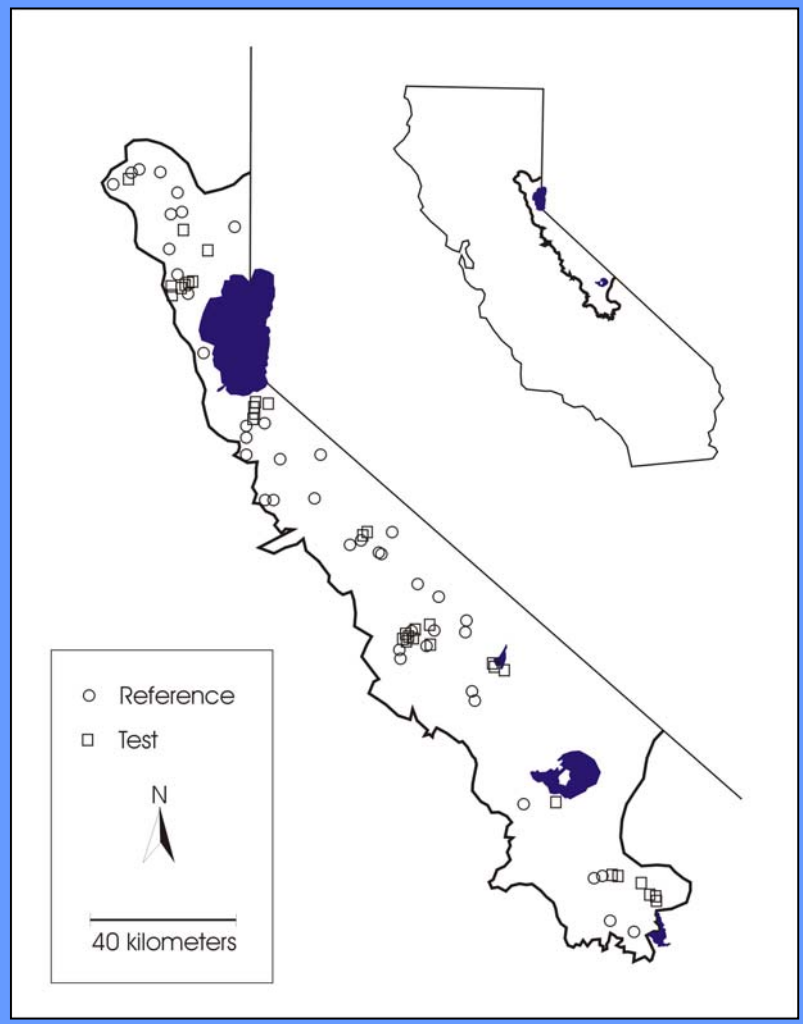
- **Geography: Little Truckee River north to Upper Owens River south**
- **<u>134 site-date surveys</u>:**

  42 reference sites (15 repeated in multiple years =62 total)

  39 test sites (21 repeated =72 total)

<u>Streams represented</u>:

- **1st to 4th order**
- **<1 to 15 m width**
- **most 6000 to 9000 feet elevation**
- **summer index period**
- **gradient <1 to 8%**
- **riffle-pool sequences**



Data Set

Combined criteria for defining reference / test:

- Watershed scale: minimum upstream road xing density
- Reach scale: minimum bank erosion conditions
- Absence of any chronic pollution sources

Sources of variation measured:

- among stream sites forming the reference distribution
- reference and test variation between years of repeat sampling (temporal variation)
- within-site spatial variability in the assessment (riffle replicates within reach)

Standardized format:

- SNARL data converted to 500 fixed-count, CAMLnet taxonomic effort w/ midge and mite resolution to genus (some species or sp.groups)

Metric Selection

- Initial metric screening: 65 reduced to 30
  > No Ref – Test separation indicated by distributions

Further selection optimized according to performance:

1. CV – priority to metrics w/ DQO <20%
2. Signal = Ref/Test mean ratio DQO Ref 50%>Test
3. Signal:Noise = $Ref\text{-}Test_{means}/Ref_{SD}$ DQO Ref 50%>Test
4. Empirical signal:noise = Test sites overlapping into Ref range
   DQO metric rankings minimizing tests >25%tile of reference
5. Reference and Test distributions DQO of near normality, fewest outliers, least overlap of central 25-75% ranges
6. Correlation of metric with human stress gradients DQO R>0.5

Metrics not meeting DQOs in any category: 30 reduced to 22

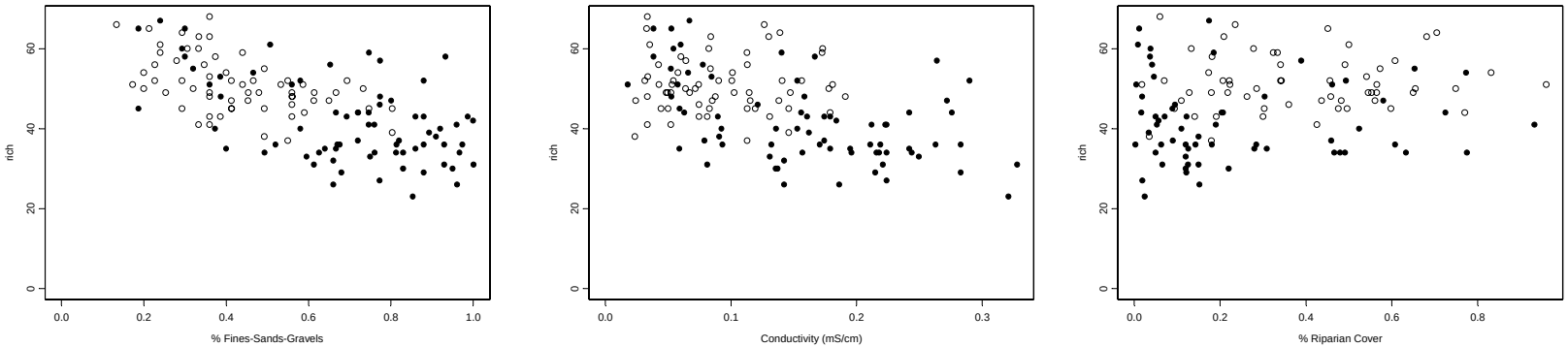
Scale to 0-10 as 10^{th} percentile test to median of reference range

Inter-correlation or conceptually redundant: 22 reduced to 13

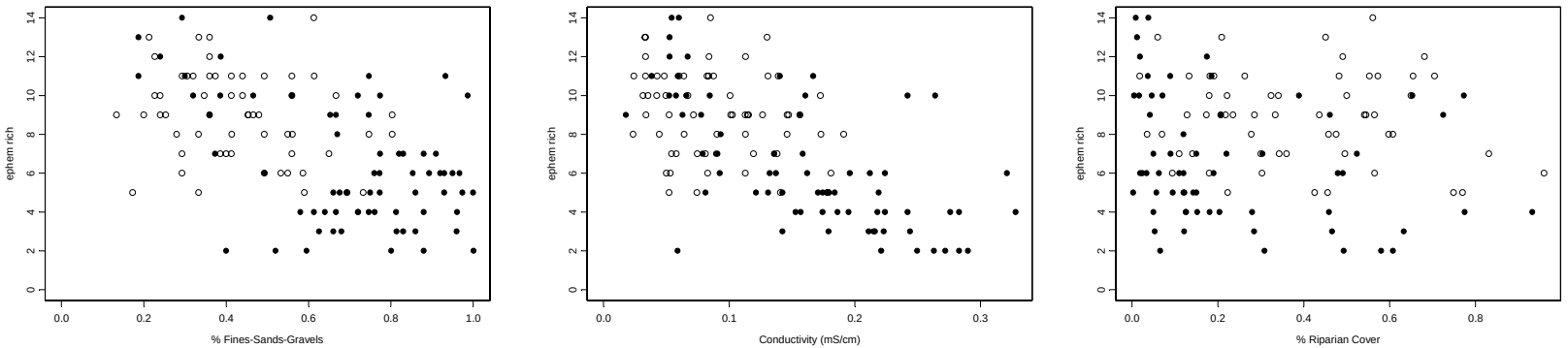
Examples of some richness metric responses to disturbance gradients:

Sedimentation	Conductivity	Riparian Cover

rich



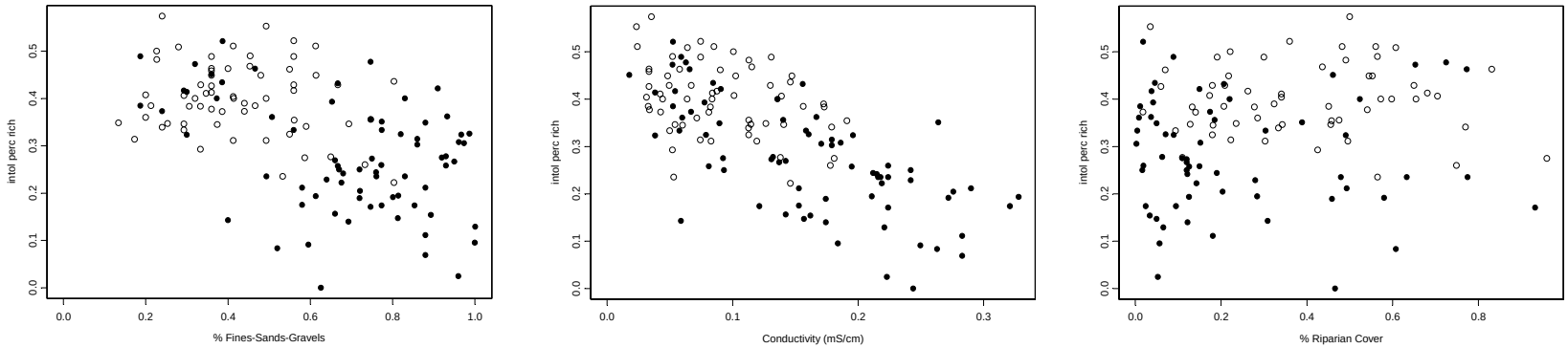
ephem rich



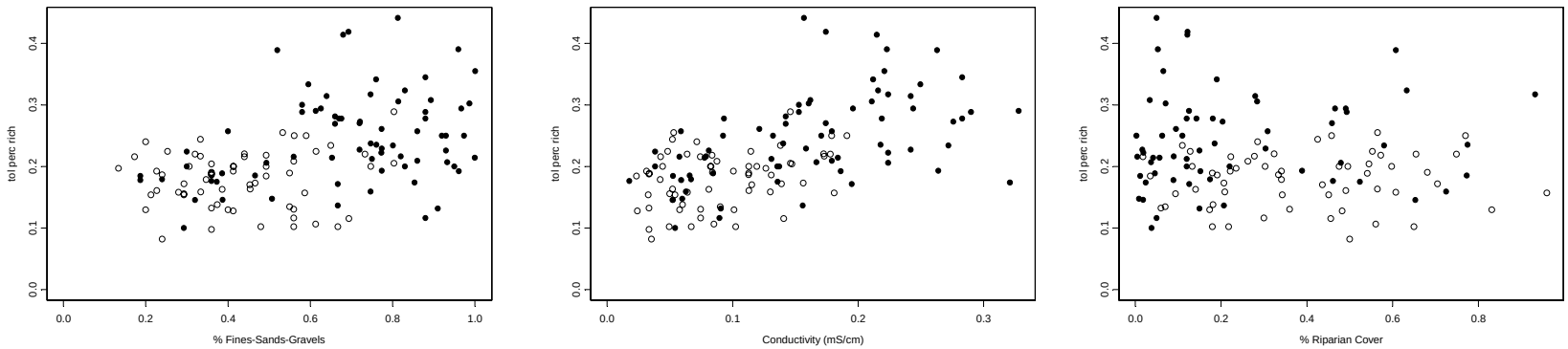
Examples of some tolerance metric responses to disturbance gradients:

Sedimentation | Conductivity | Riparian Cover

intol perc rich



tol perc rich



IBI construction

- Combine 13 metrics in different groups (20) balancing low inter-correlations with mix of types (richness, composition, tolerance, and function)
- Rank alternatives according to best performance (as with metrics) > 3 options selected

12-Metric IBI	10-metric IBI	8-metric IBI
rich	rich	rich
ephem rich	ephem rich	
plecop rich	plecop rich	
trichop rich	trichop rich	trichop rich
acari rich	acari rich	acari rich
chiro.perc rich	chiro.perc rich	chiro.perc rich
		intol perc rich
tol perc rich	tol perc rich	
pred rich		
ept abund		
shredder	shredder	shredder
dominance 3	dominance 3	dominance 3
bi	bi	bi

	12-Metric	10-Metric	8-Metric
Noise	0.13	0.13	0.14
Signal	1.55	1.53	1.53
Signal:Noise Ratio	2.74	2.65	2.45
Overlap @ 25th	0.21	0.18	0.19

Very similar but use 12-metric IBI as provisional recommendation to maintain stability & flexibility

Regulatory application: options for Lahontan to consider

Thresholds: methods comparison showed minimum Type II error at Type I =15-20% (also corresponds to natural break in distribution)

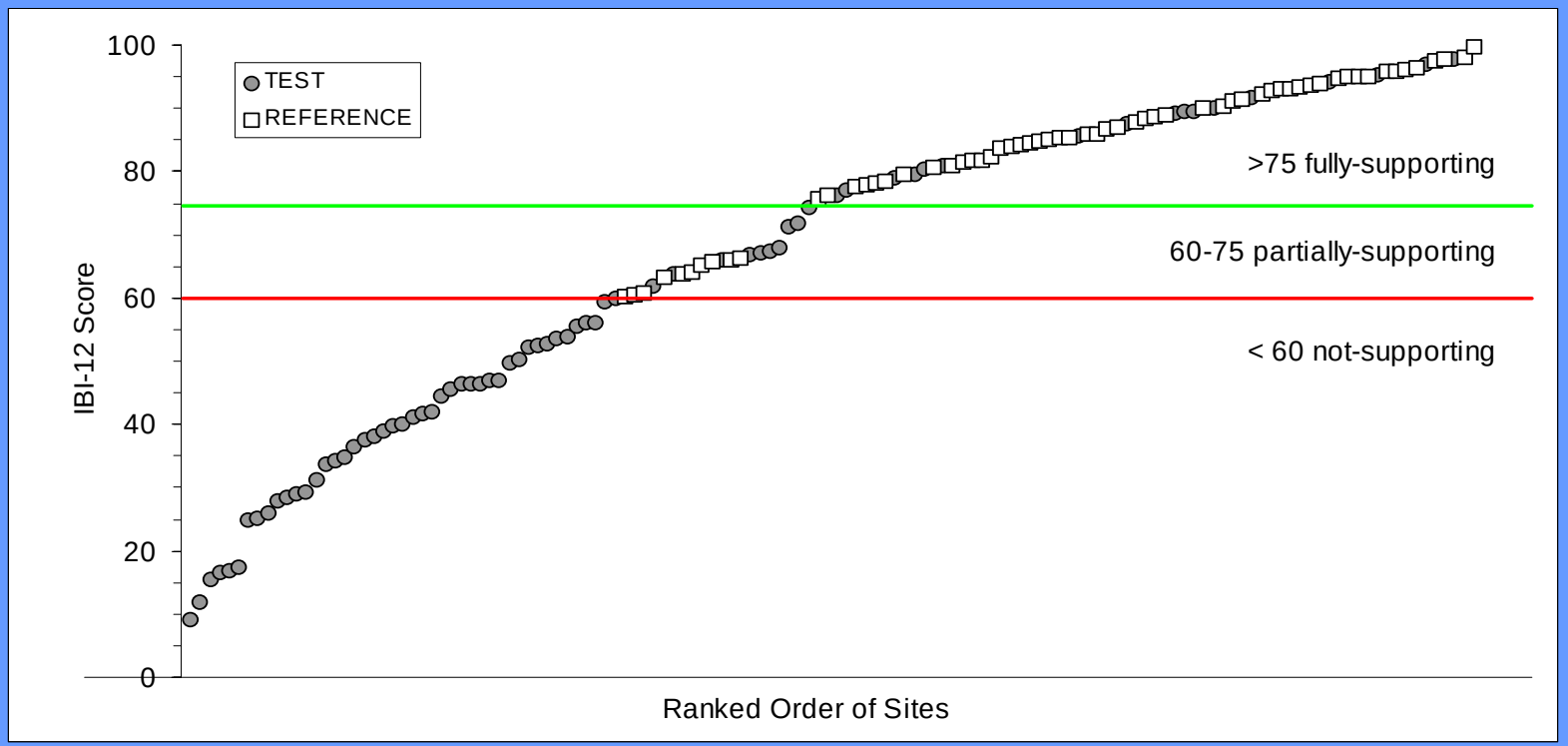
Supporting = above Type I error rate threshold (or >25% reference?)

Partially supporting = between minimum reference and threshold

>transitional / uncertain condition: in multiple years of testing few sites scored consistently within this class (marginal references & recovering tests?)

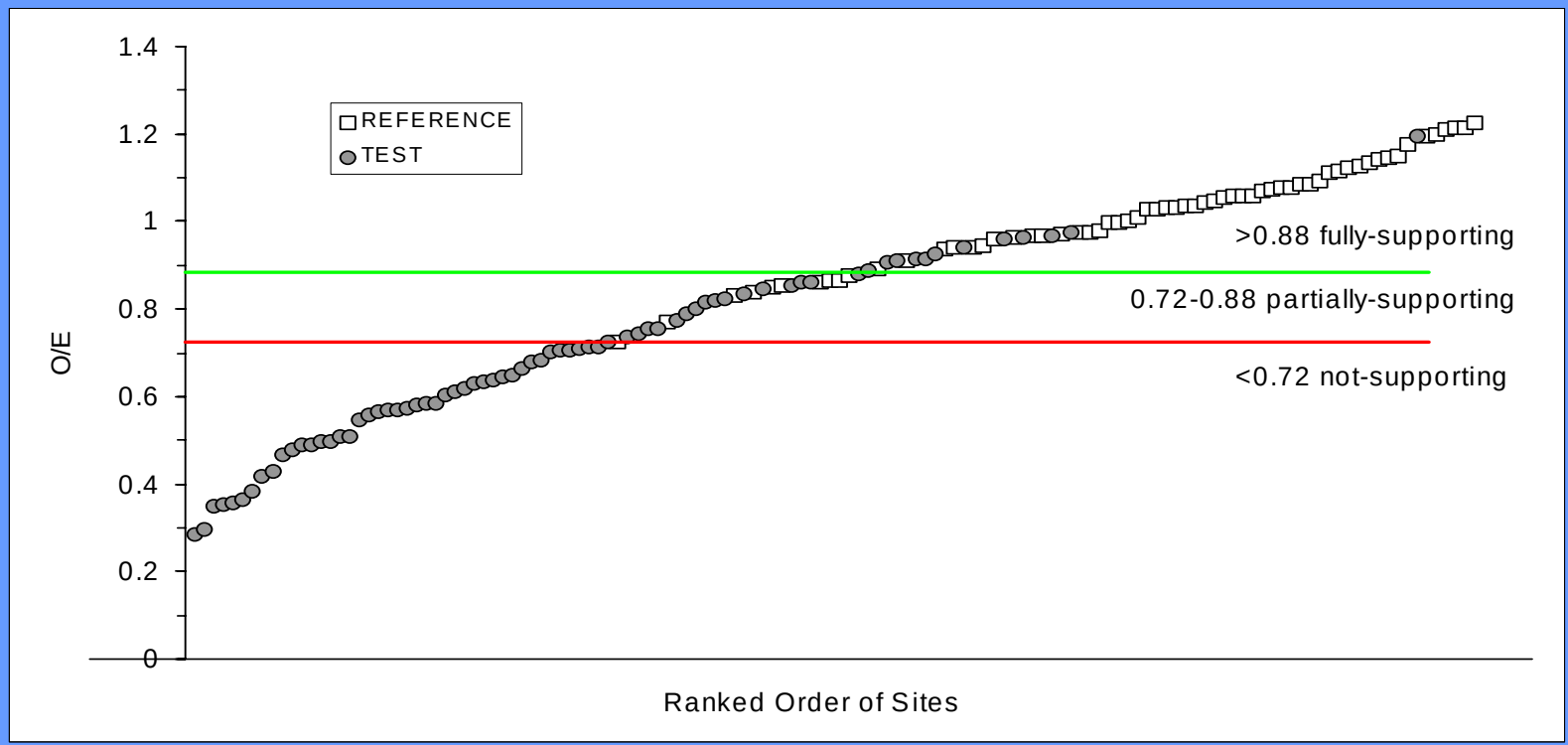
Not supporting = below minimum reference (or Ref_{mean}-2SD?)

IBI-12



RIVPACS model: cross-validation assessment

- **Clustering = 5 groups of 5-20 each**
- **Discriminant analysis = latitude, stream width and annual precipitation provide best predictors**
- **Reference mean = 1.02 (0.72 – 1.22), CV = 0.114**
- **Use same threshold criteria to define three condition classes**



How do assessments compare?

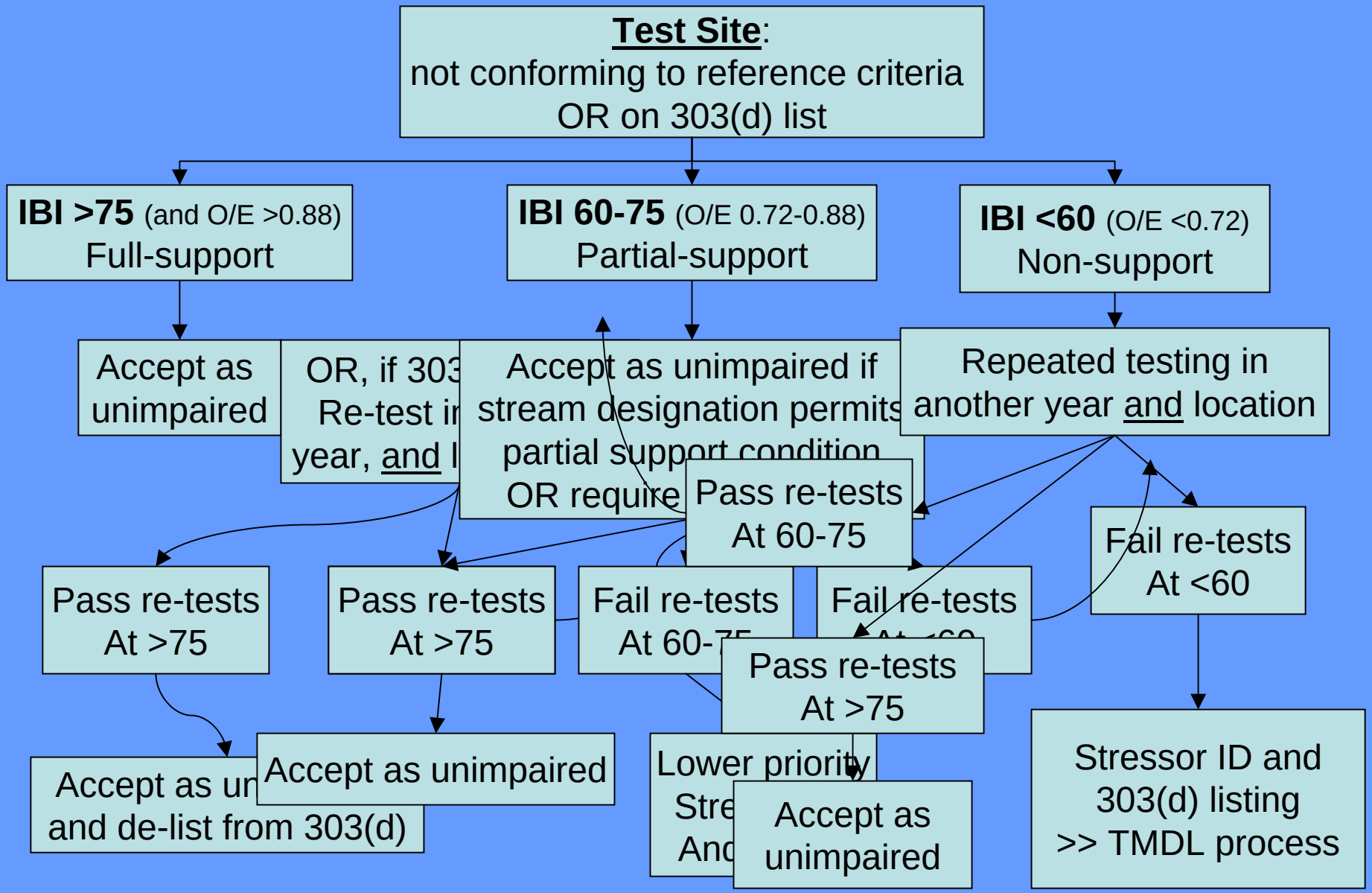
CROSS-VALIDATION CONDITION CLASS ASSESSMENTS						
		REFERENCE		TEST		
		FULL	PARTIAL	FULL	PARTIAL	NOT
IBI	12	52	10	17	10	45
	10	52	10	16	17	39
	8	52	10	15	33	24
RIVPACS	O/E	52	10	13	15	44

- **test sites passing as fully supporting remain about the same among IBIs with slight decrease for O/E**
- **test sites move from not supporting to partial supporting as IBI metric set is reduced (downgrading of reference minimum)**

- **O/E and IBI-12 are in best agreement above and below not-supporting limit**
- **85% agreement in reference sites graded as full-supporting**
- **90% agreement in test sites graded as not-supporting**
- **RIVPACS grades more test sites as partial than full-supporting than IBI-12**

Along with separate models based on multi-habitat sampling, confidence in accurate site evaluations can be achieved through an integrated assessment analysis approach

Regulatory Decision Tree: derive certainty in assessments from repeated tests between years, and consistent results within and between analytical approaches



Uncertainty in site assessment: spatial and temporal variation & partial-support condition

- Between-year variation: 2-4 yrs of repeat sampling at 15 Ref & 21 Test
- Within-site variation: 5 replicates between adjacent riffles - all surveys

Some measures to consider in defining condition classes:

- Both spatial and temporal variation average SD=8-9 units
- About 84% of multi-year repeats are in agreement for site condition assessments (others change by one class only)
- Lower range of reference distribution: many scoring systems call this the “fair” range (sometimes the 25th percentile of the reference), here our standard, based on Type I-II trade-off is ~16th percentile, and many sites falling into this zone were references that in multiple years of testing otherwise scored above the threshold: partial-support

  >shows that references sometimes score in this zone of uncertain condition due to natural variability
- 2 SD below reference mean = criterion for fair to poor limit used in other studies = 62 (IBI-12 minimum =60) - 3 classes above & 2 below?
- MDD based on riffle replicates (maximizes within-reach variability) averages 7.7 for 5 replicates, 16.3 for 3 replicates (80% at p=0.05)
- EPA and human disturbance gradient: 6 classes – another option for dividing the aquatic life use attainment categories