

US EPA ARCHIVE DOCUMENT

The IBI Development Process

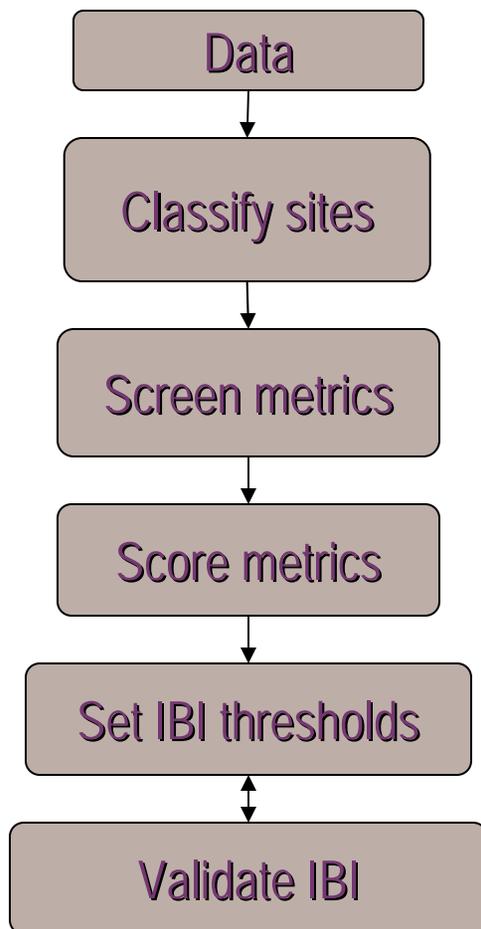
Karen Blocksom

U.S. EPA-ORD-NERL, Cincinnati, OH

October 3-5, 2006

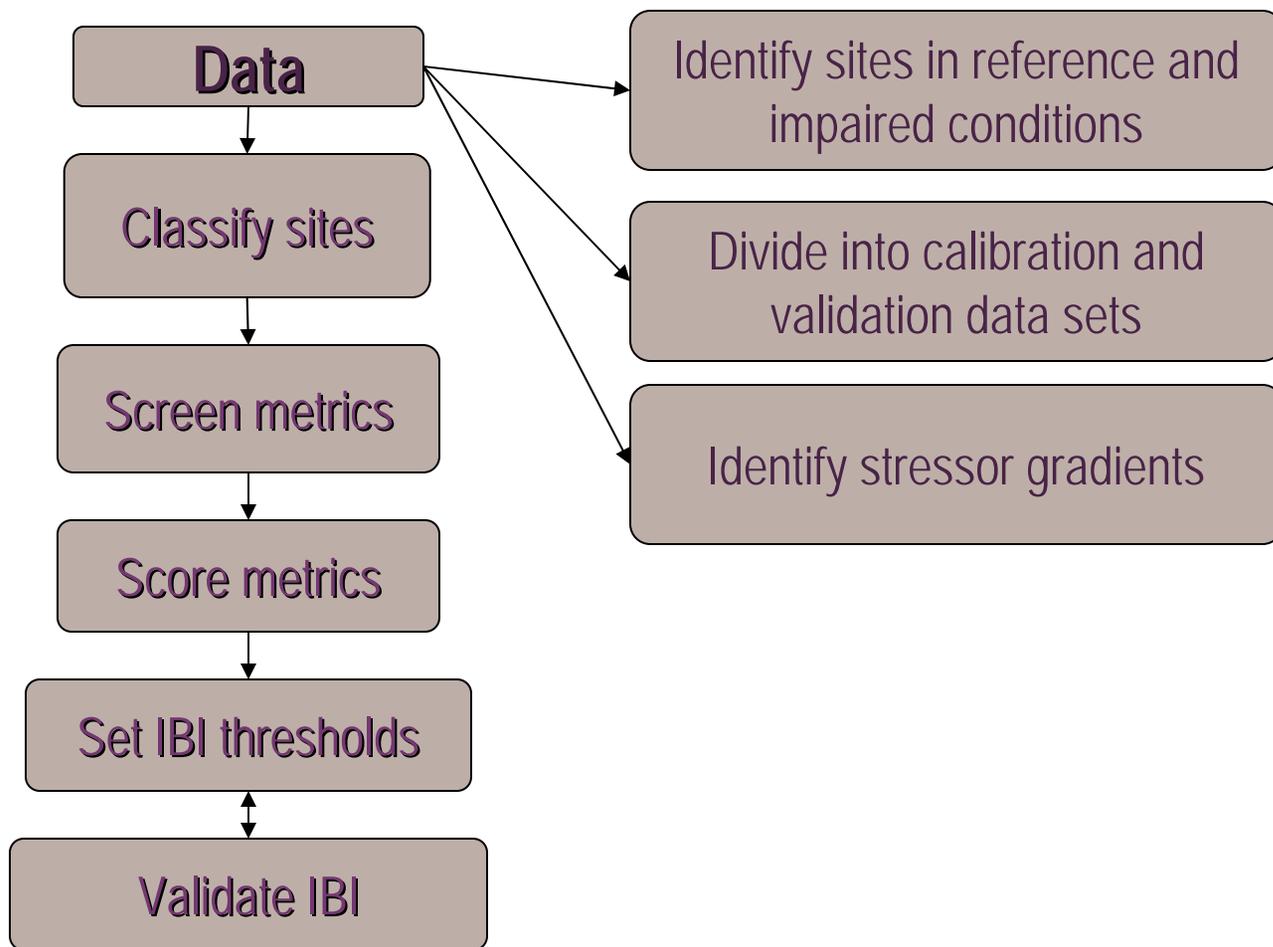
*Although this work was reviewed by EPA and approved for publication, it may not necessarily reflect official Agency policy.

Components of the process



The process is never really completely objective – it is always a combination of statistics and professional judgment.

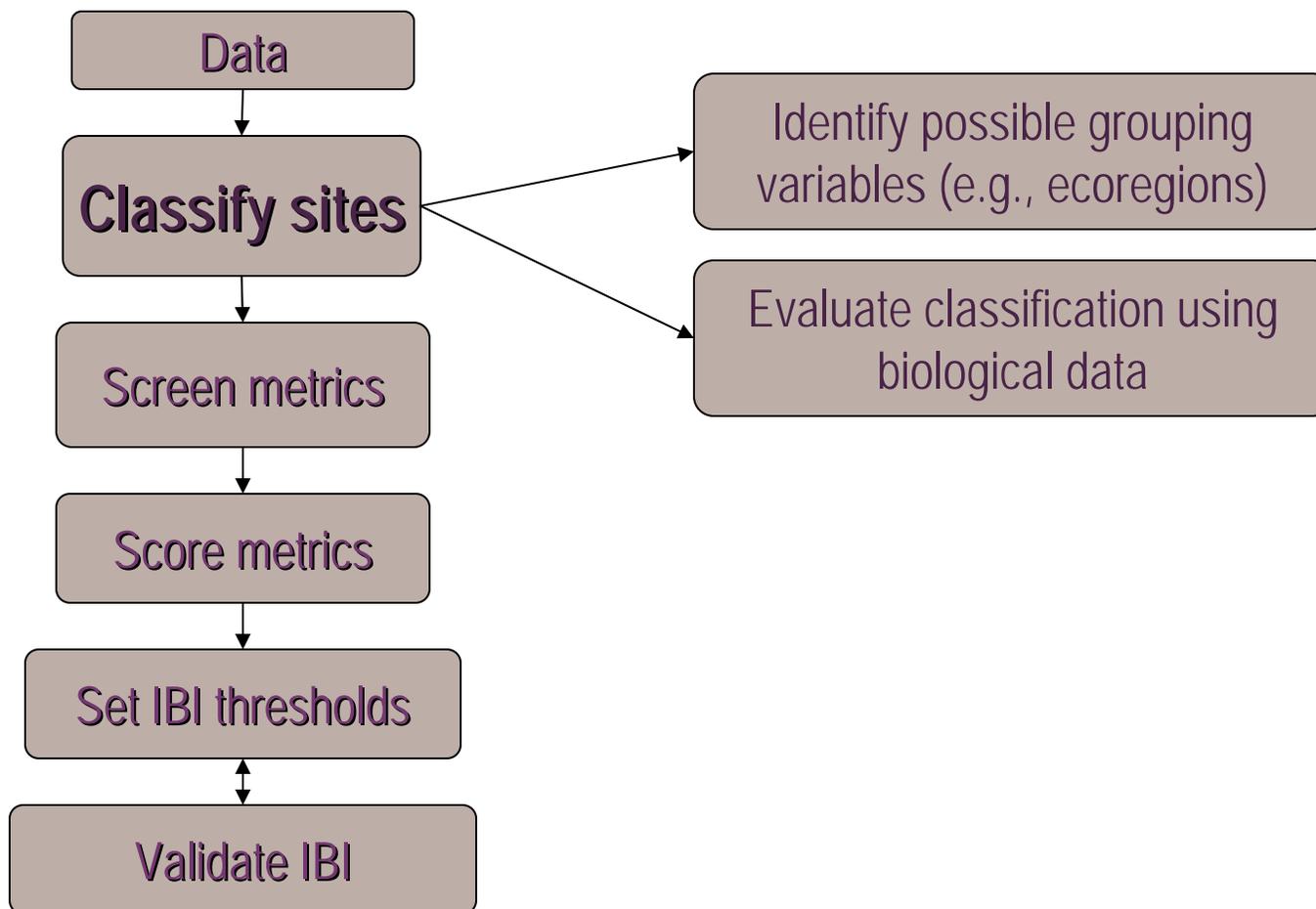
Components of the process



Data

- Reference (least disturbed/minimally disturbed) and impaired (“trashed”) conditions:
 - Determine *a priori* using non-biological data if possible
 - Can affect many aspects of IBI development, so spend time getting this right
- Divide data into calibration and validation data sets
 - Randomly but with some stratification to ensure representation across rivers and other major areas
 - Approximately 20-30% of samples in validation data set
 - Depending on size of data set, may postpone this step until assembling index
- Identify stressor gradients
 - These will be relevant for assessing metric and IBI responsiveness

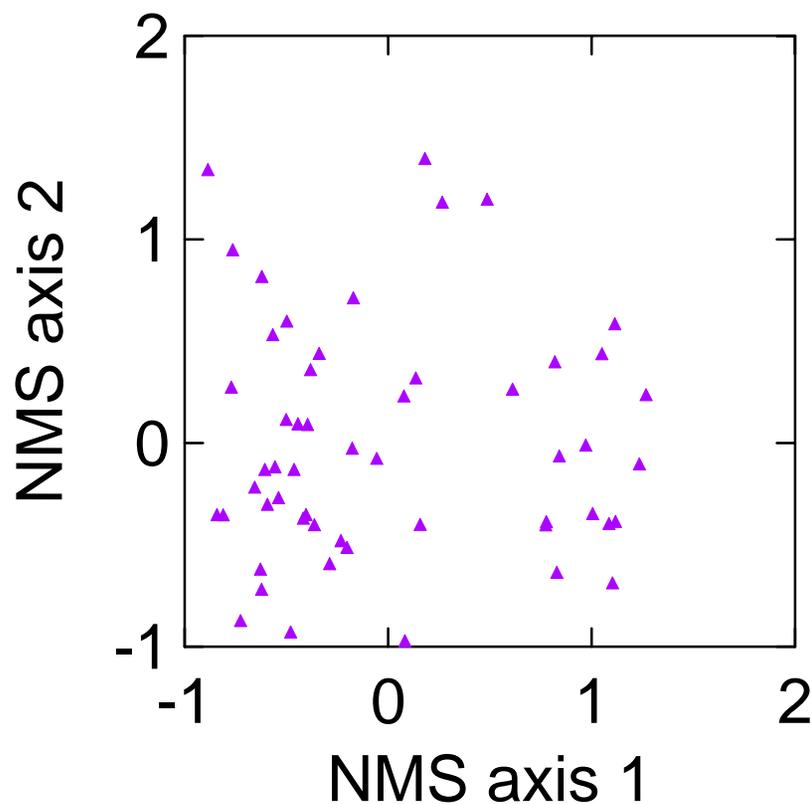
Components of the process



Classification

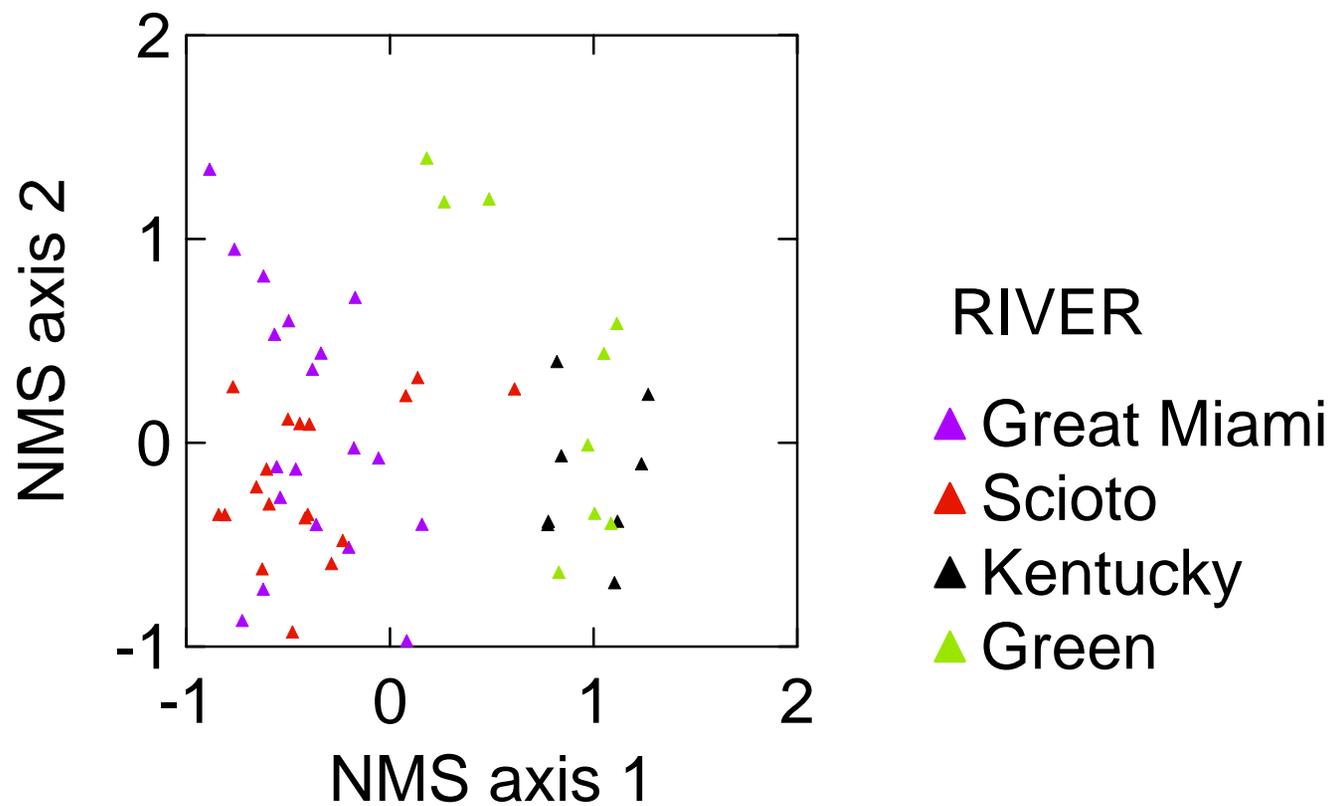
- Typically reference sites only, to avoid introducing stressor gradients into process – segregate by river (i.e., Missouri, upper Miss, Ohio)?
- Non-metric multidimensional scaling (NMDS) on raw abundance data
 - Smaller stress value is better, <20 or better
 - May drop rare taxa (e.g., present at fewer than 5% of sites, or <5% of abundance at <10% of sites)
 - Bray-Curtis distance measure commonly used
- Identify alternative (pre-existing) classification schemes:
 - Omernik ecoregions
 - Basins
 - Other classifications based on hydrology, geomorphology, etc. (e.g., The Nature Conservancy, Forest Service)
- Look for separation or clustering in NMDS based on different classification schemes

Classification - example

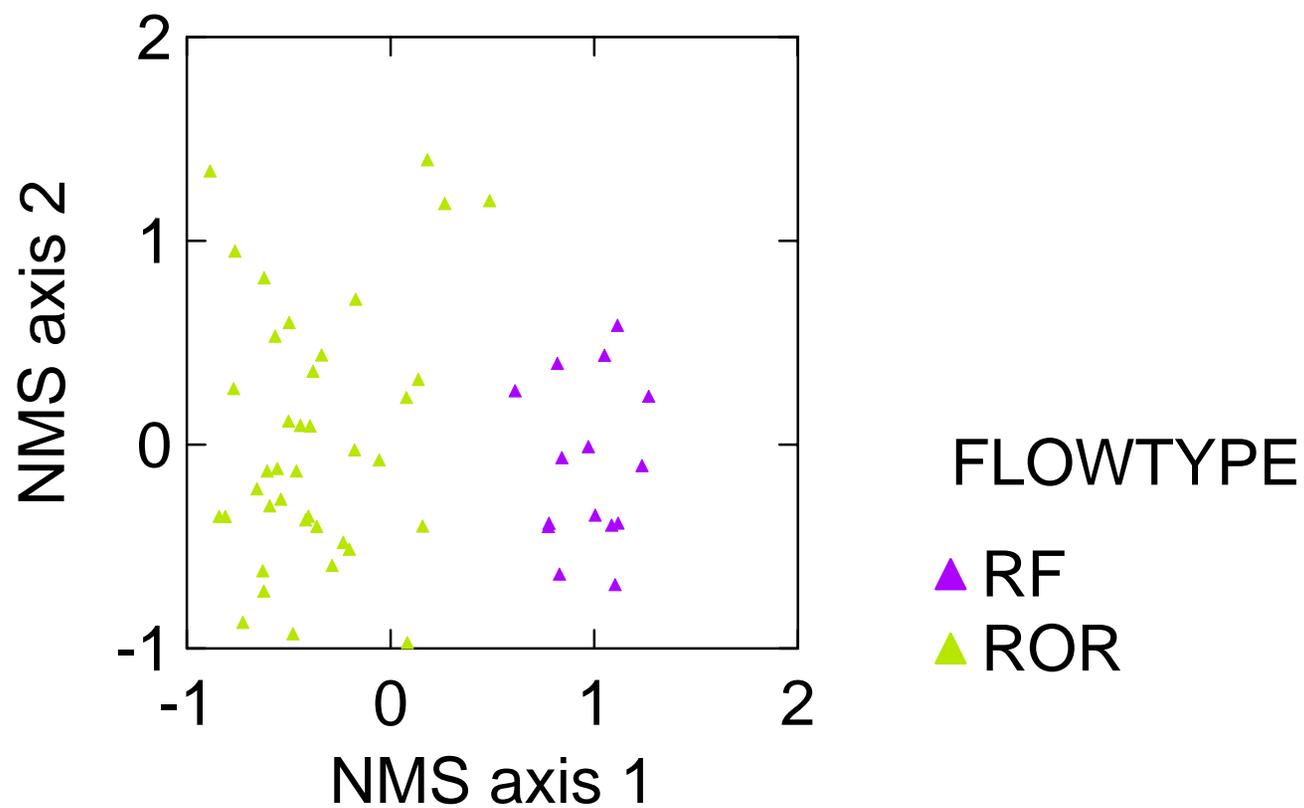


- Large river study
- 54 sites
- 4 rivers
- Macroinvertebrates
- Genus-level data

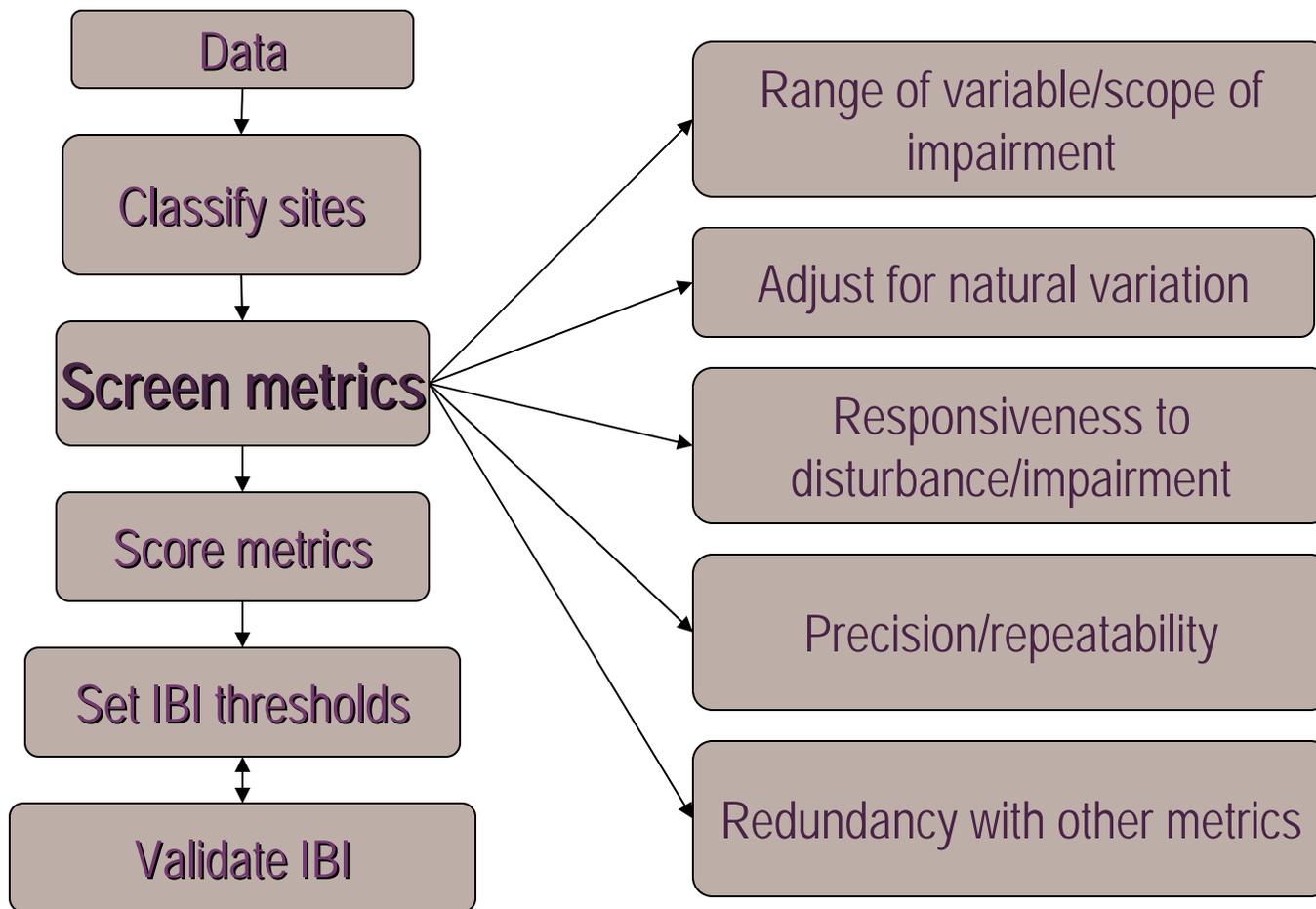
Classification - example



Classification - example

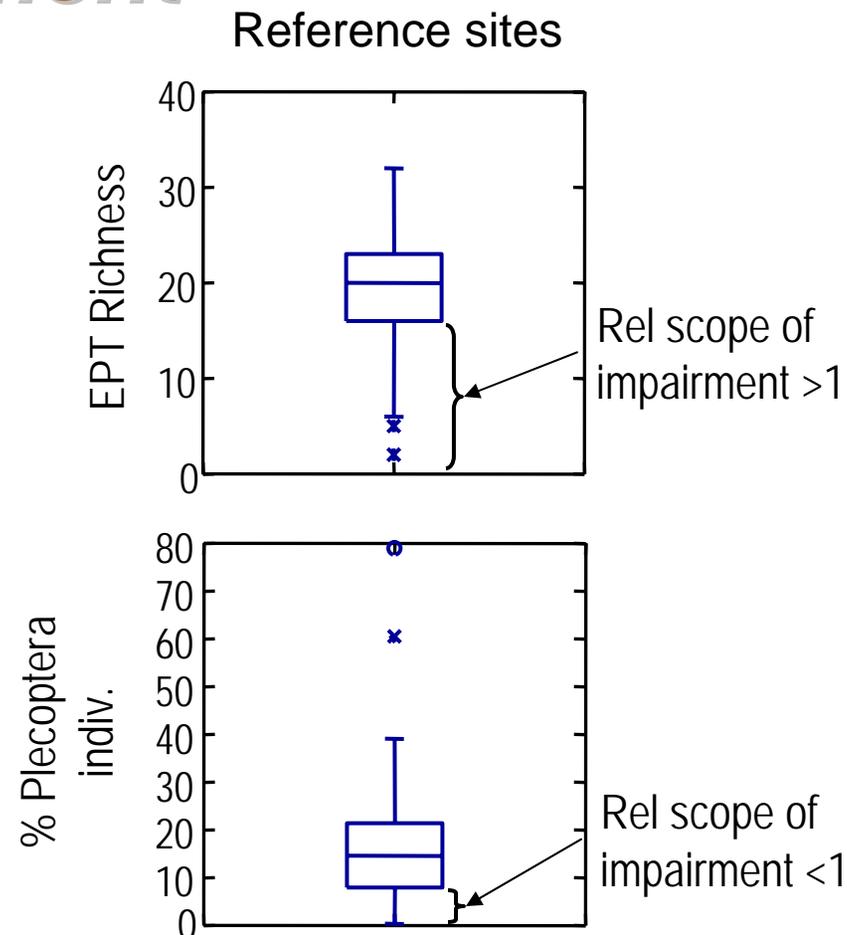


Components of the process



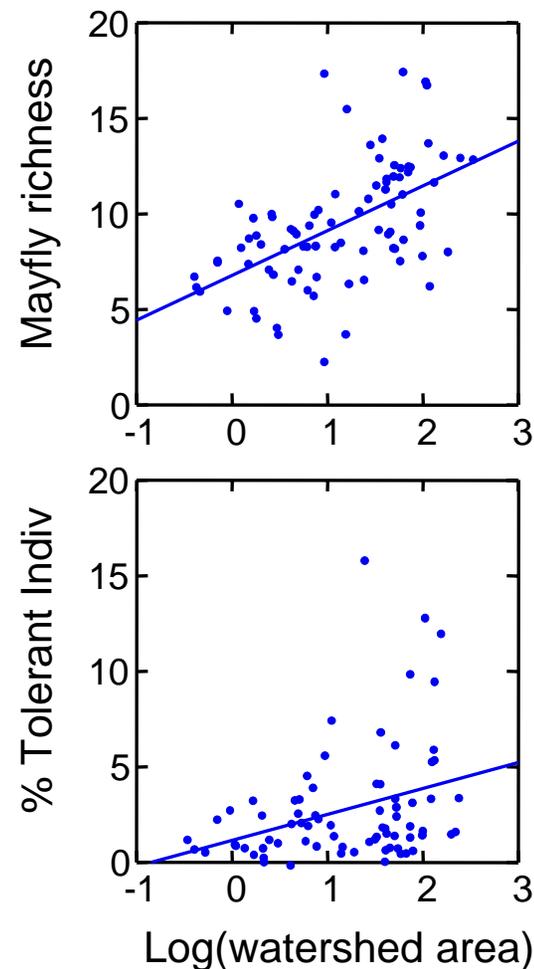
Screening metrics: Range/scope of impairment

- Set criteria for different types of metrics
- Examples:
 - Percentage metrics-range <10
 - Richness metrics-range <5
- Scope of impairment
 - Variable range below reference 25th percentile
 - Compare to interquartile range of reference sites (relative scope of impairment)
 - >1 acceptable, <1 unacceptable



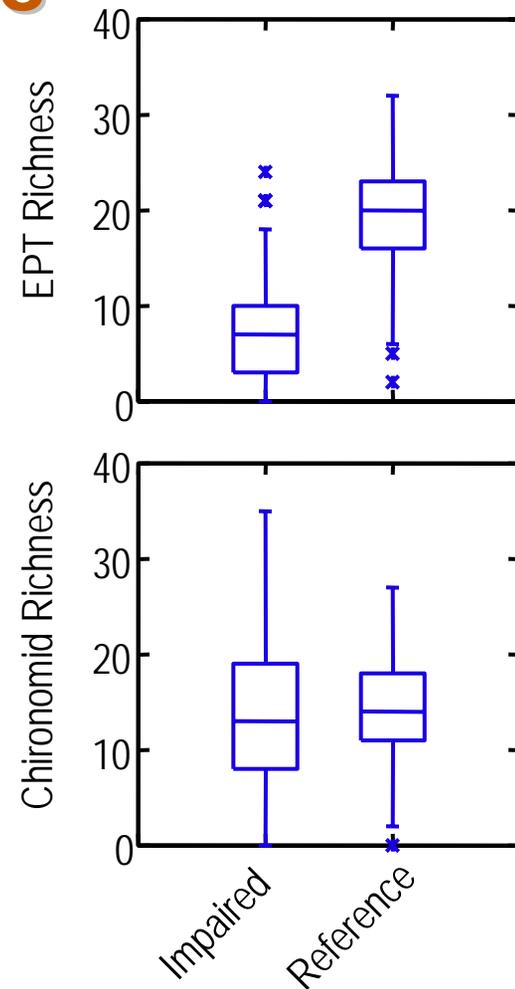
Screening metrics: Adjust for natural variation

- Examine plots of metrics vs. continuous natural factors (e.g., Julian day, river mile, flow)
- Use only reference sites in plots
- Calculate residuals for all sites from regression
 - Type of regression may depend on type of trend detected (i.e., linear or wedge-shaped)
 - Quantile regression available in R
- Avoid automation using simple correlations, as many important relationships may be missed
- Make sure relationships make ecological sense to avoid spurious correlations
- Use caution with variables which might be strongly related to stressors of interest



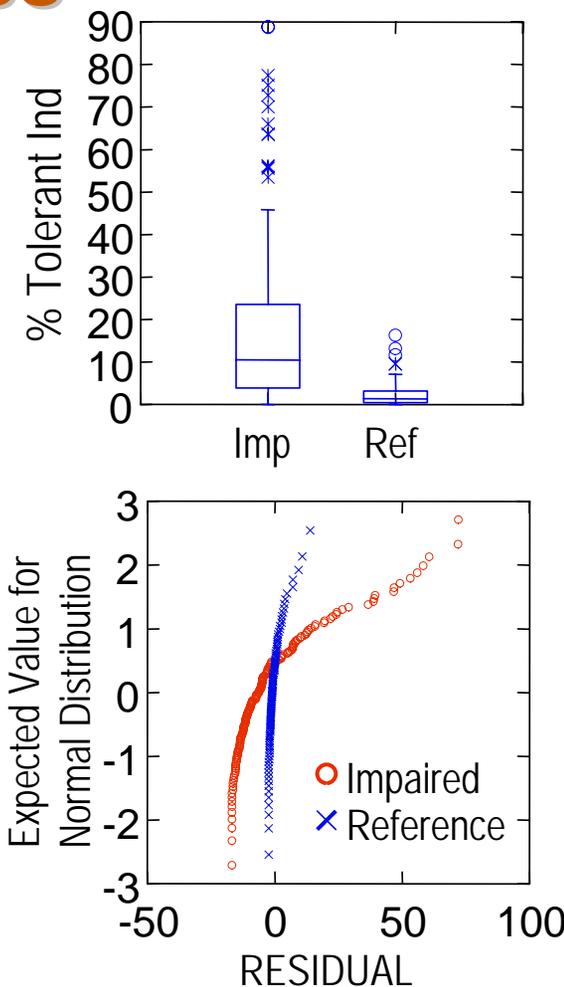
Screening metrics: Responsiveness

- Box plots of reference vs. impaired sites
 - Assess degree of overlap of interquartile ranges (boxes) and medians
 - Accept only metrics with some minimum level of separation
 - May also calculate discrimination efficiency (DE) as proportion of impaired sites with values below 25th percentile of reference (or similar)
- Formal tests
 - May not be appropriate to apply simple F- or t-tests to all metrics without checking assumptions (i.e., normality, homogeneous variance)
 - Nonparametric tests do not consider degree of difference if significant



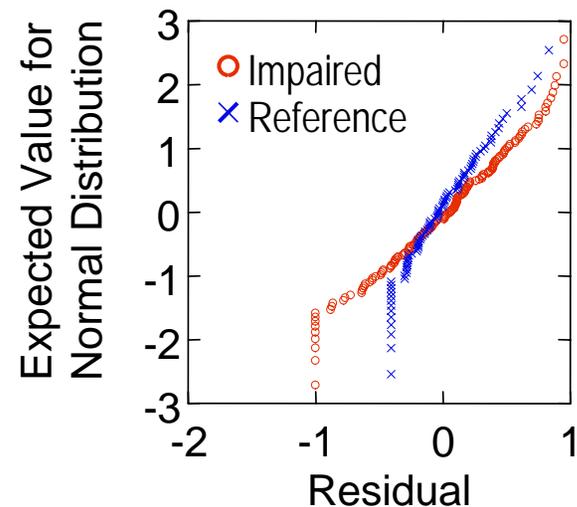
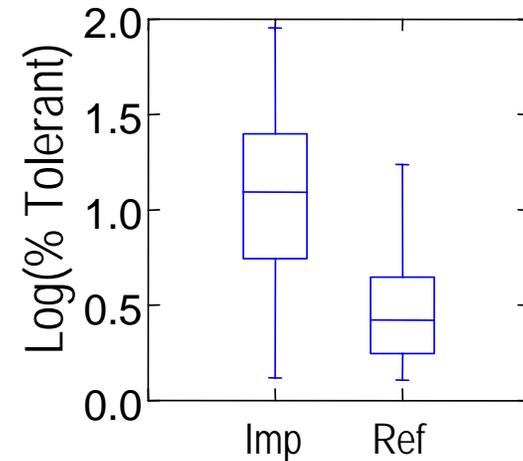
Screening metrics: Responsiveness

- Example: % Tolerant Individuals
- Untransformed, difficult to assess separation of impaired/reference
- From ANOVA, $F=49$
- BUT, very skewed distributions apparent in residuals



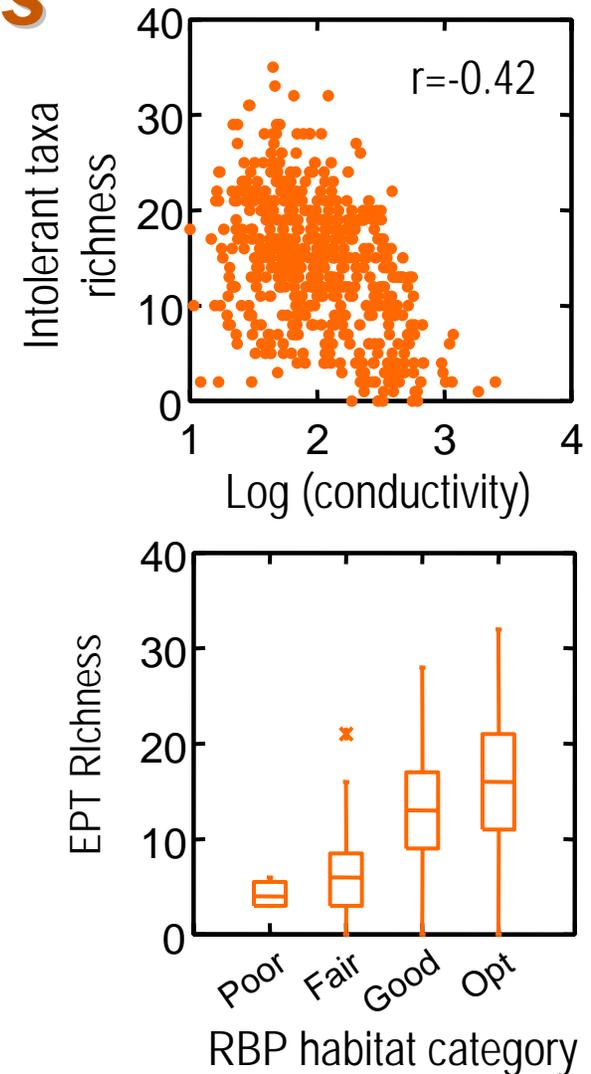
Screening metrics: Responsiveness

- When transformed as $\log(x+1)$, more normally distributed
- More clear separation of reference/impaired – $F \sim 106$
- Residuals no longer as strongly skewed



Screening metrics: Responsiveness

- Plots of individual or composite measures of disturbance vs. metrics
- Categorizing variables into several levels can show patterns more clearly
- Correlations do not provide as much information on relationships if large sample sizes



Screening metrics: Precision/Repeatability

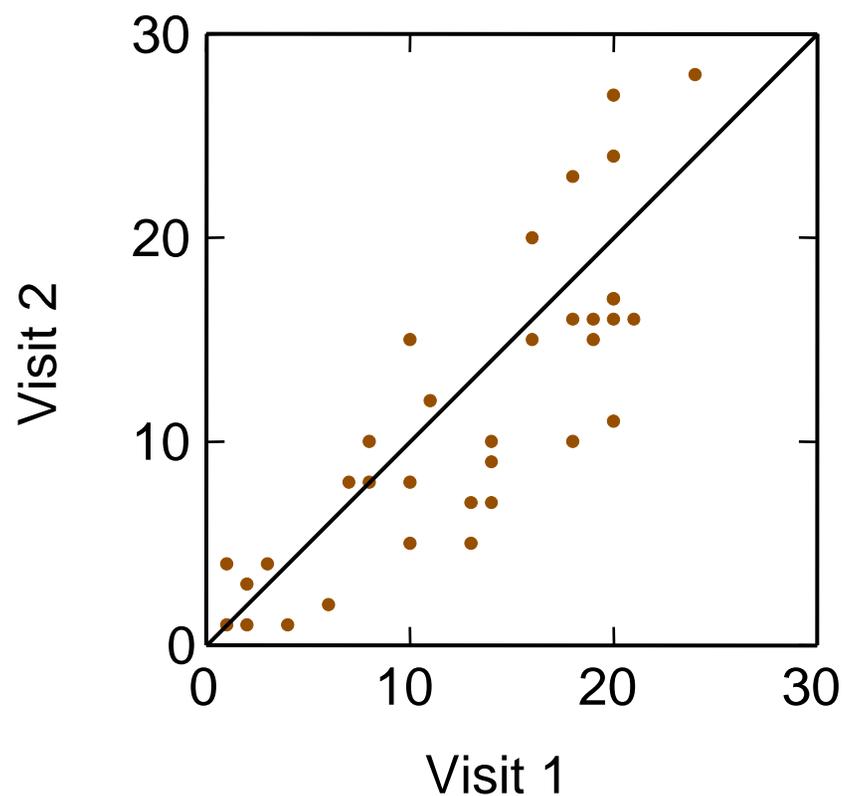
- Calculate Signal:Noise ratio (S/N) based on Kaufmann et al. (1999):
 - Signal/Noise = (Variance among sites)/(Variance within sites)
 - ANOVA with site ID(year) as random factor, year as fixed effect
 - Sites with revisits, include all sites to “enhance” signal
 - $S/N = (F_{\text{site}} - 1) / c_1$, where c_1 is:
 - number of reps if all the same (e.g., 2 per site)
 - otherwise, coefficient of site ID component in Type III expected mean squares (SAS)
 - Meaning of values:
 - Under 2 – poor repeatability/precision
 - 2-6 – moderate
 - >6 – good

Screening metrics: Precision/Repeatability

Example of S/N:

EPT Richness

$$\begin{aligned} S/N &= (9.44 - 1)/2 \\ &= 4.22 \end{aligned}$$



Screening metrics: Precision/Repeatability

- Alternative approaches

- Relative Percent Difference (RPD):

$$RPD = \frac{|Y_1 - Y_2|}{\bar{Y}} \times 100$$

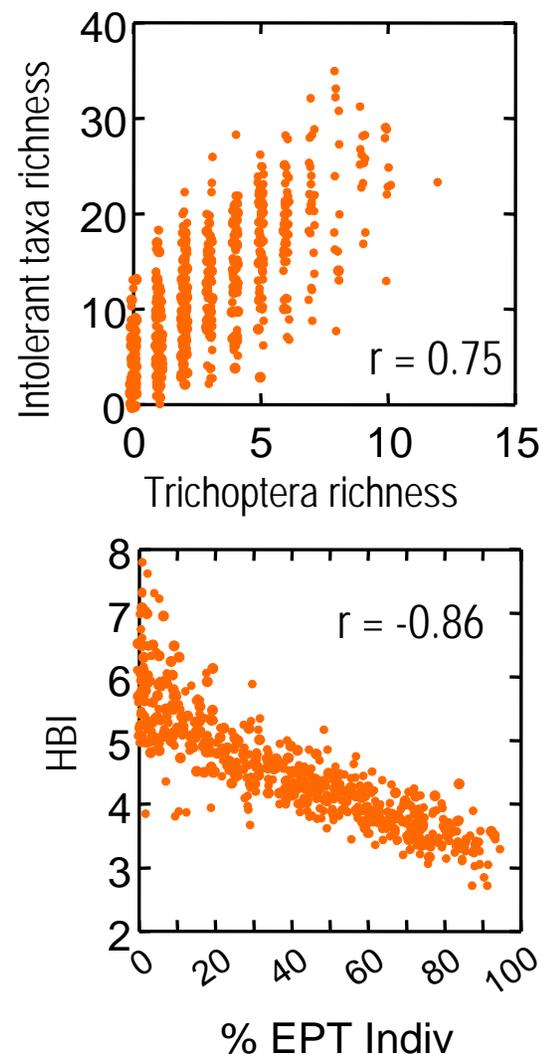
- Coefficient of Variation (CV):

$$CV = \frac{RMSE}{\bar{Y}} \times 100$$

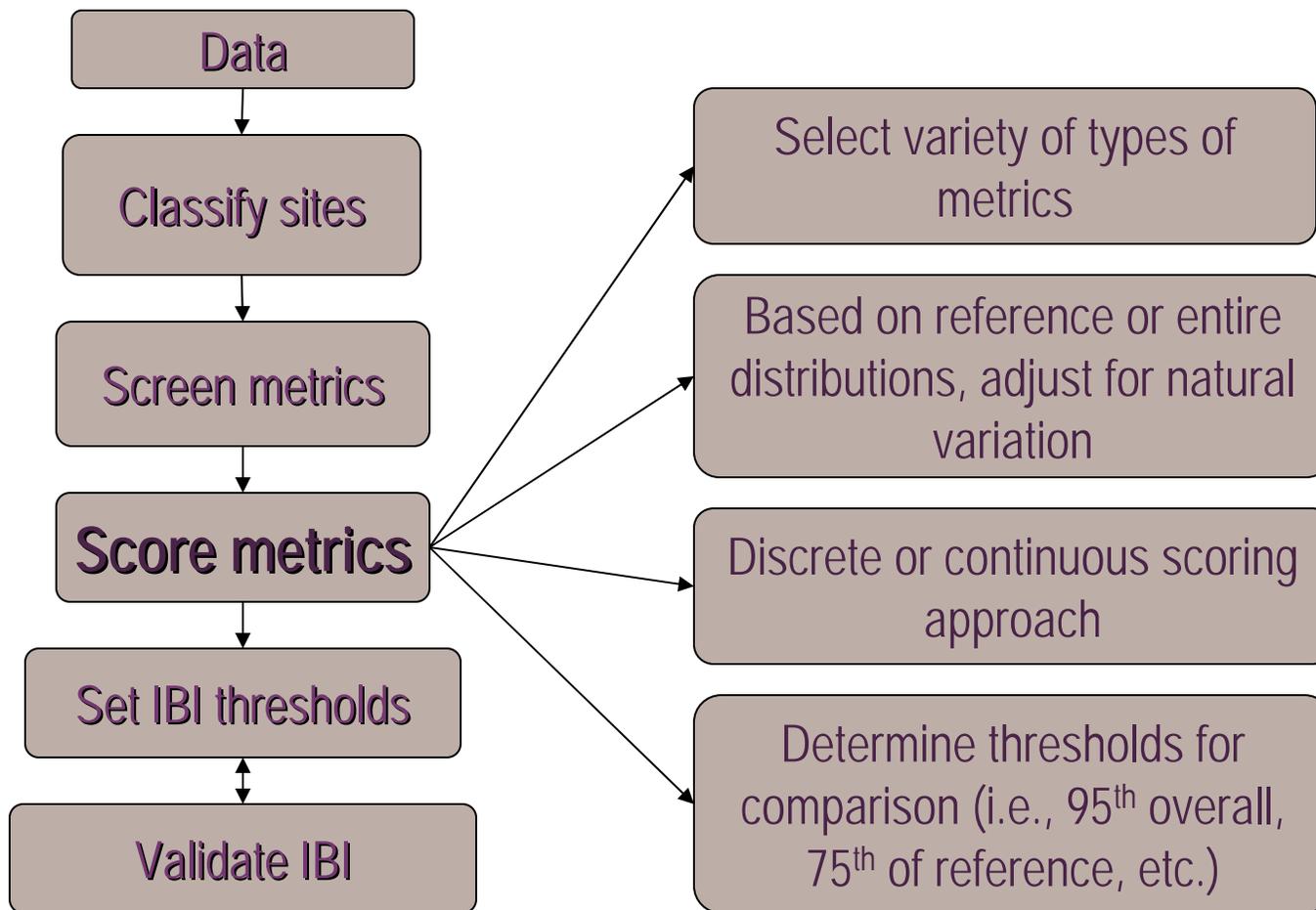
- Both approaches require setting some criteria for judging adequate metric precision/repeatability

Screening metrics: Redundancy

- Pearson correlations of among all (remaining) metrics
- For $r > 0.8$ (or other predetermined value), choose one of pair, based on:
 - Other measured characteristics of metric
 - Ease of calculation
 - Ecological meaning
- May also look at bivariate plots to verify redundancy



Components of the process

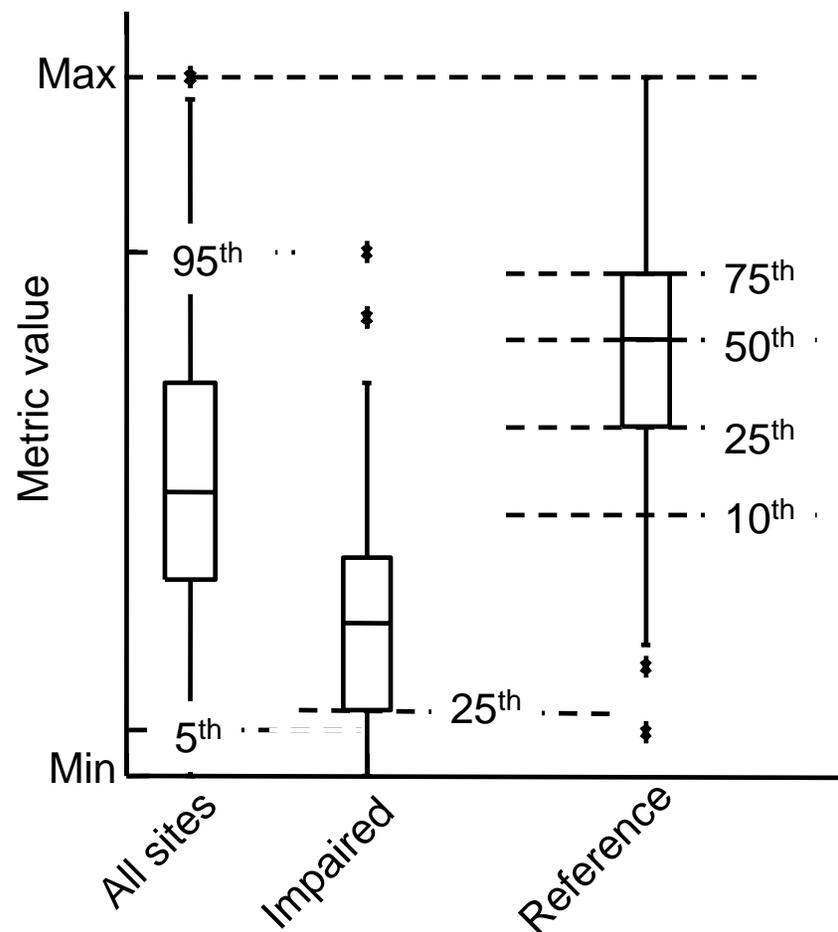


Scoring metrics: Selecting metrics

- Use metric evaluation information to select suites of candidate metrics and test as index alternatives
- Incorporate metrics representing various assemblage characteristics:
 - Taxa richness
 - Taxonomic composition
 - Tolerance
 - Trophic level
 - Reproductive guilds
 - Behavioral habits (e.g., clingers)
- Identify separate suites of metrics for each class?
- Try to include more metrics rather than fewer
- Use calibration data only for this step

Scoring metrics: Select scoring approach

- Many types of scoring approaches
- Set thresholds based on:
 - reference/impaired sites
 - all sites
- Scoring can be:
 - Discrete (e.g., 1-3-5)
 - Continuous (e.g., 0-100)



Scoring metrics: Setting thresholds

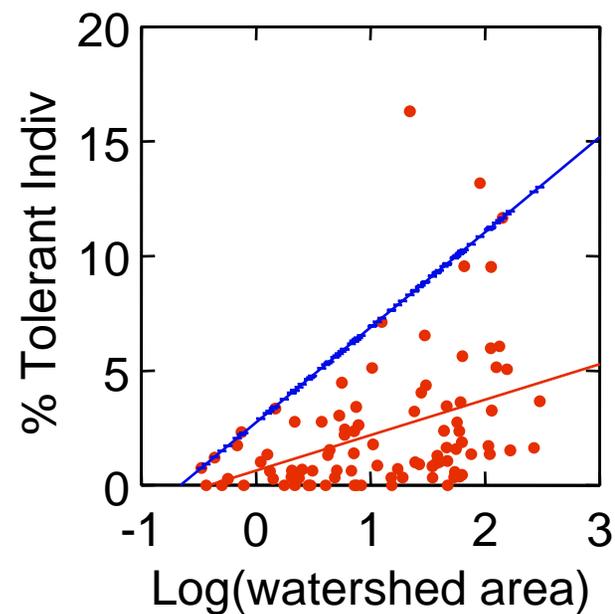
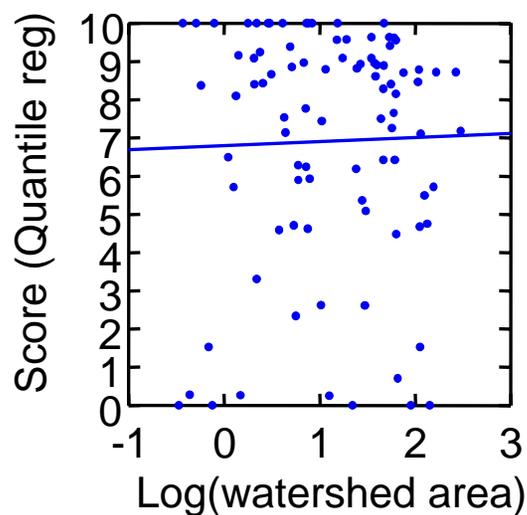
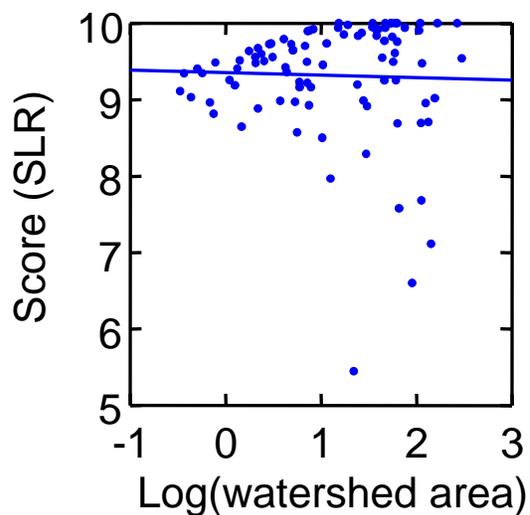
- Reference/Impaired Approach:
 - Placement of thresholds depends on confidence in quality or accuracy of reference/impaired designations
 - Larger percentile of reference corresponds to lower confidence in quality (i.e., using 10th percentile implies higher quality reference relative to using 25th percentile)
 - Avoid extremes to exclude outliers
- All Sites Approach
 - Typically use 95th percentile/5th percentile, depending on direction of response to stress
- Use consistent approach across metrics

Scoring metrics: Setting thresholds

- When continuous scoring is performed, thresholds are used as the “ceiling” and “floor”, and scores are the position of the observed value in that interval
- Scoring is applied differently for metrics that increase and those that decrease with disturbance
- Typically, we apply thresholds with the assumption that the metric decreases with increasing disturbance (e.g., EPT taxa) – for example:
 - $\text{Score} = (\text{Metric value} - \text{Floor}) / (\text{Ceiling} - \text{Floor}) * 10$
- However, metrics like HBI, % Tolerant taxa, etc. increase with disturbance, so the floor is the more desirable state – for example:
 - $\text{Score} = (\text{Ceiling} - \text{Metric value}) / (\text{Ceiling} - \text{Floor}) * 10$

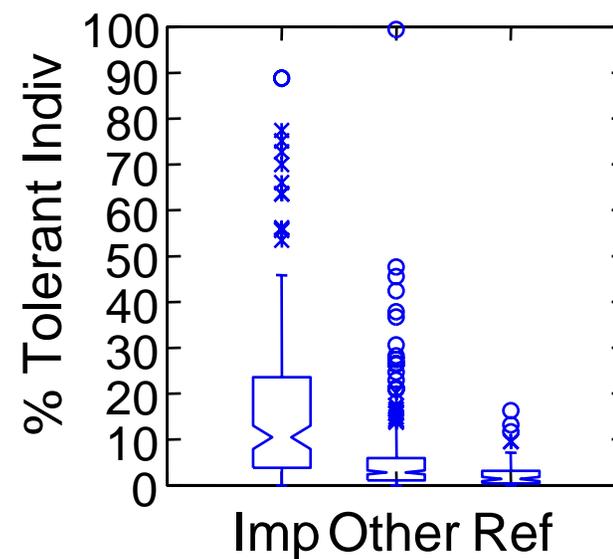
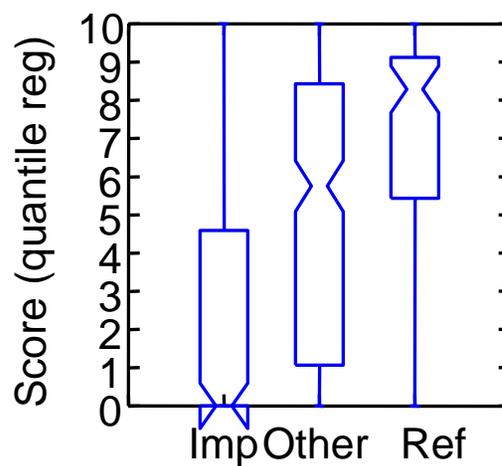
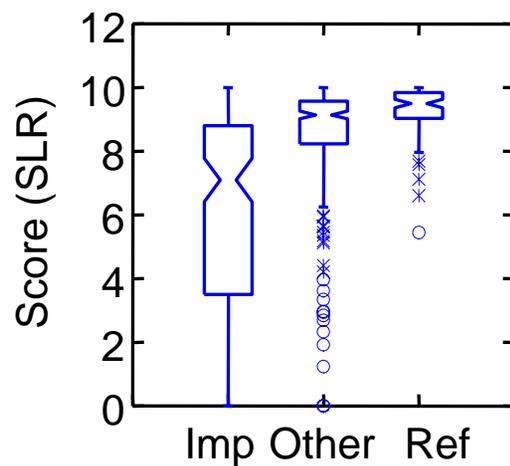
Scoring metrics: Adjusting for natural variation

- If relationship to natural factors (i.e., watershed area) is wedge-shaped, use quantile regression to establish thresholds
- Simple linear regression in these cases affects scores at ends of natural gradients



Scoring metrics: Adjusting for natural variation

- If relationship to natural factors (i.e., watershed area) is wedge-shaped, use quantile regression to establish thresholds
- Simple linear regression in these cases affects scores at ends of natural gradients



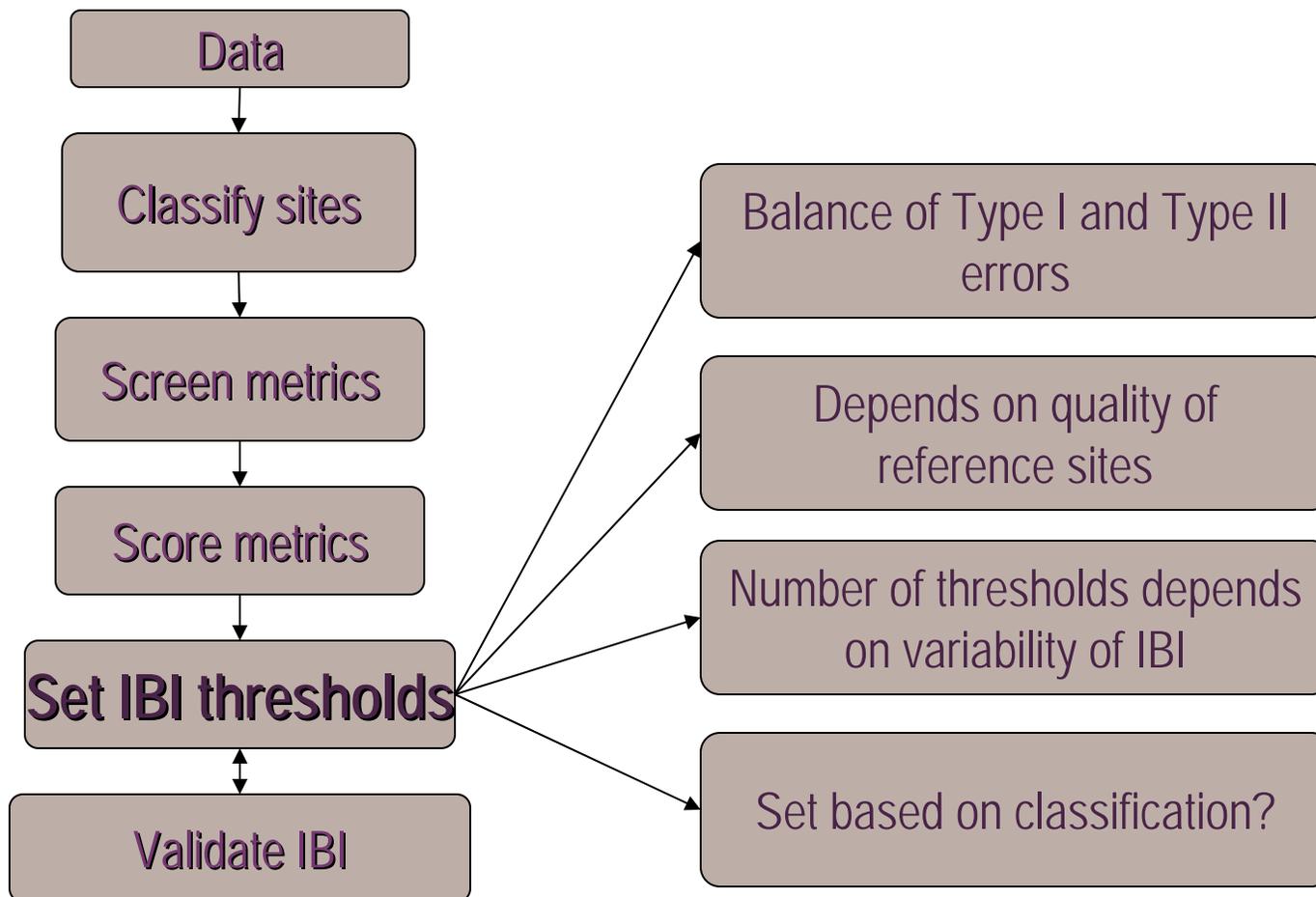
Scoring metrics: Aggregate into index

- Ensure that all metrics scores are in the same direction (i.e., a higher score = better condition)
- Combine metrics into index using either average or sum
- Perform this step for all suites of metrics being considered

Next steps:

- Validation of IBI
- Setting biocriteria or condition thresholds
- Not necessarily a linear process

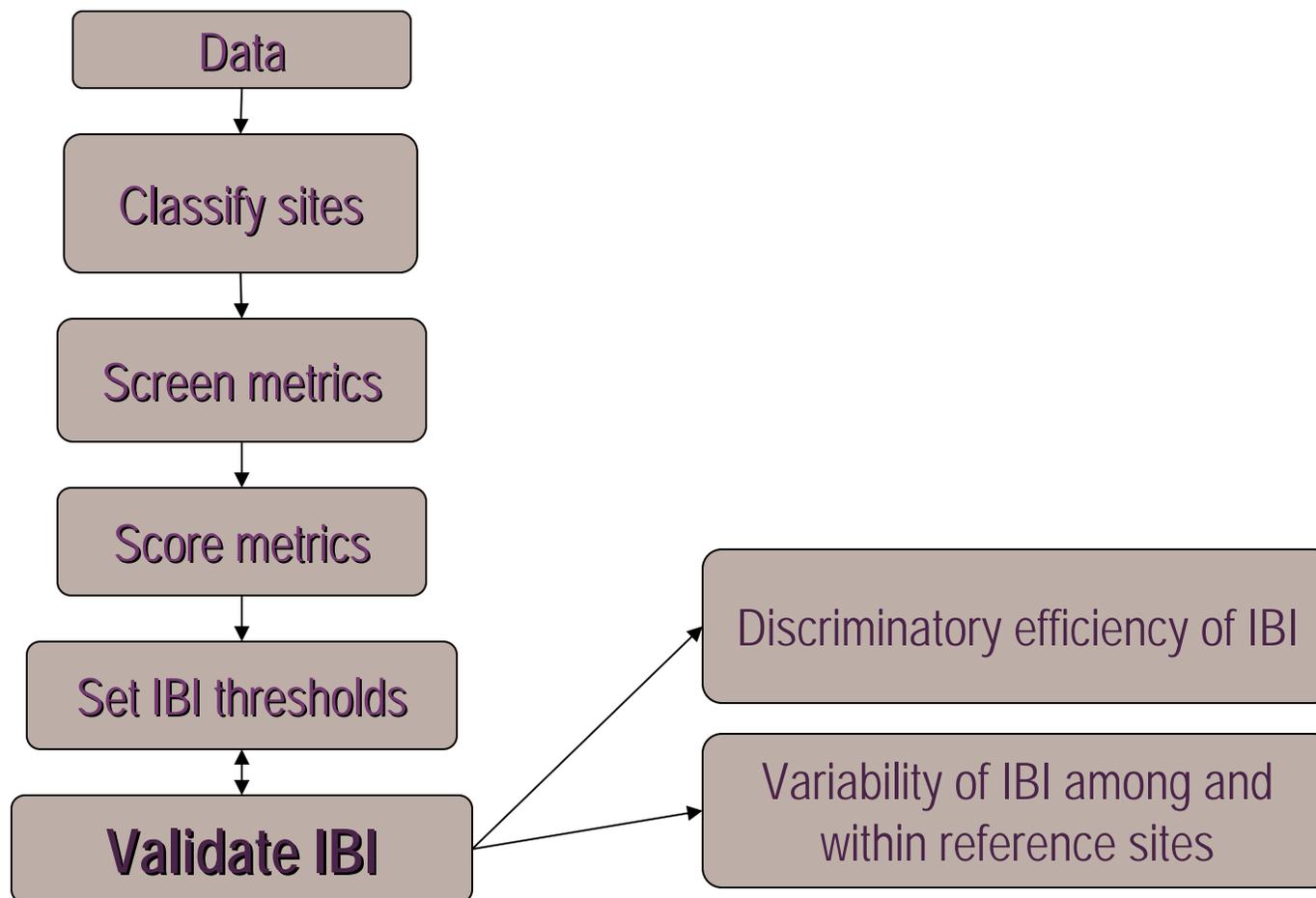
Components of the process



Set IBI thresholds

- Typically a percentile of reference condition is identified as threshold(s) for condition classes
- Set for different classes if reference site index scores do not strongly overlap
- Step may occur before or after validation of IBI – may use condition classes in validation
- Number and placement of thresholds could be based on:
 - Balance of Type I and II errors
 - Response to disturbance gradient
 - Overlap of reference and impaired sites
 - Distribution and variation of reference

Components of the process



RESEARCH & DEVELOPMENT

Building a scientific foundation for sound environmental decisions

Validate IBI

- Calculate precision and responsiveness measures for IBIs for calibration and validation data sets
- Compare outcomes between data sets
- Poor consistency in results may indicate a less desirable option
- Also may mean that disturbance or other gradient is not consistent between the two data sets
 - If this is the case, redivide data but also recalibrate metric scores using new calibration data
- Use characteristics of IBI to determine the most effective and robust choice

Final Thoughts

- There is no single “right” way to develop an IBI- multiple alternatives are available at each step
- Reference sites should be re-evaluated periodically
- Thresholds should be revisited occasionally or with specific regularity
- Should not be a one-time process, particularly if overall condition improves over time