National Biological Assessment and Criteria Workshop

Advancing State and Tribal Programs

Coeur d'Alene, Idaho
31 March – 4 April, 2003

RFC 101

Overview of the Role of Classification & Modeling in Establishing Reference Conditions

Presented by

Charles Hawkins, Aquatic, Watershed, and Earth Resources
Utah State University – Logan, UT

The need for 'classification'

- Phil's questions from the previous talk:
  - How does human activity affect aquatic ecosystems and, in particular, aquatic biota?
  - What do we do about our effects on aquatic ecosystems and aquatic biota?
- To answer these questions we need to establish what the biological condition should be given a waterbody's natural potential – I call this the *expected* condition.

However, the world is naturally heterogeneous and 'expected' may not be obvious

- We need to establish the correct match between an assessed site and its expected condition. Ideal is to be both accurate and precise.

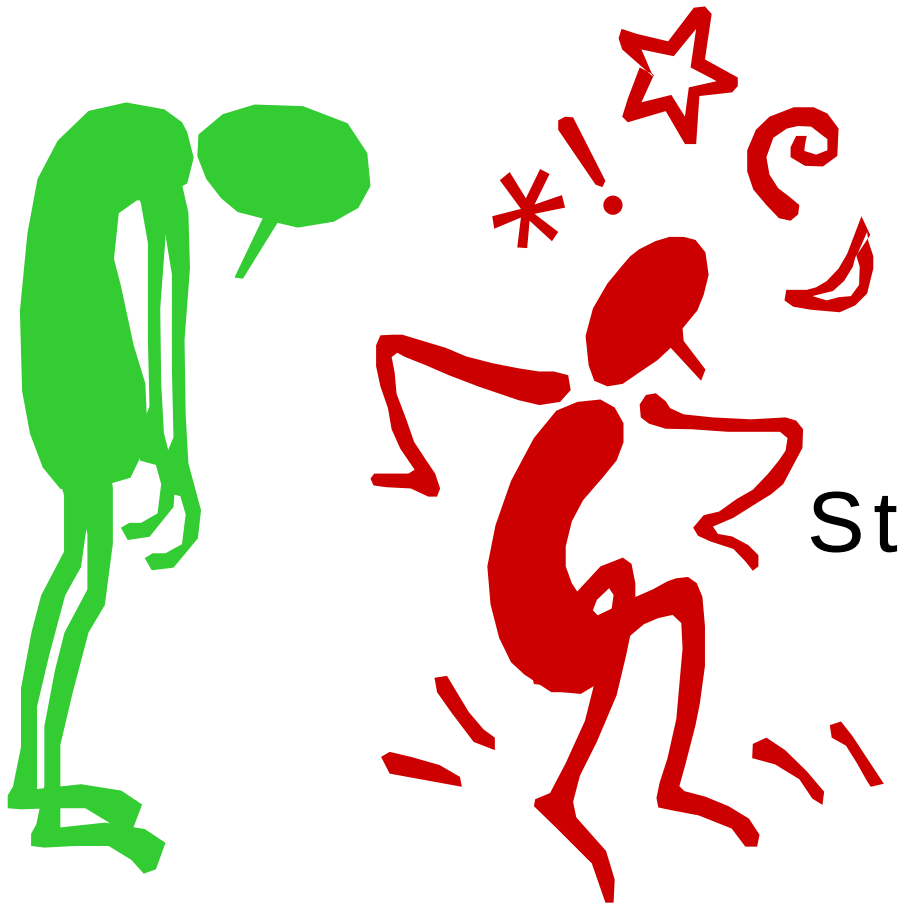
  Two common approaches:

  - ✓ classification, and
  - ✓ modeling

- This step is critical because...

Improper classification leads to bad decisions

Scientist, Manager, or Regulator



Stakeholder

Errors in prediction/classification can result in either of two types of errors of inference regarding the true condition of a site:

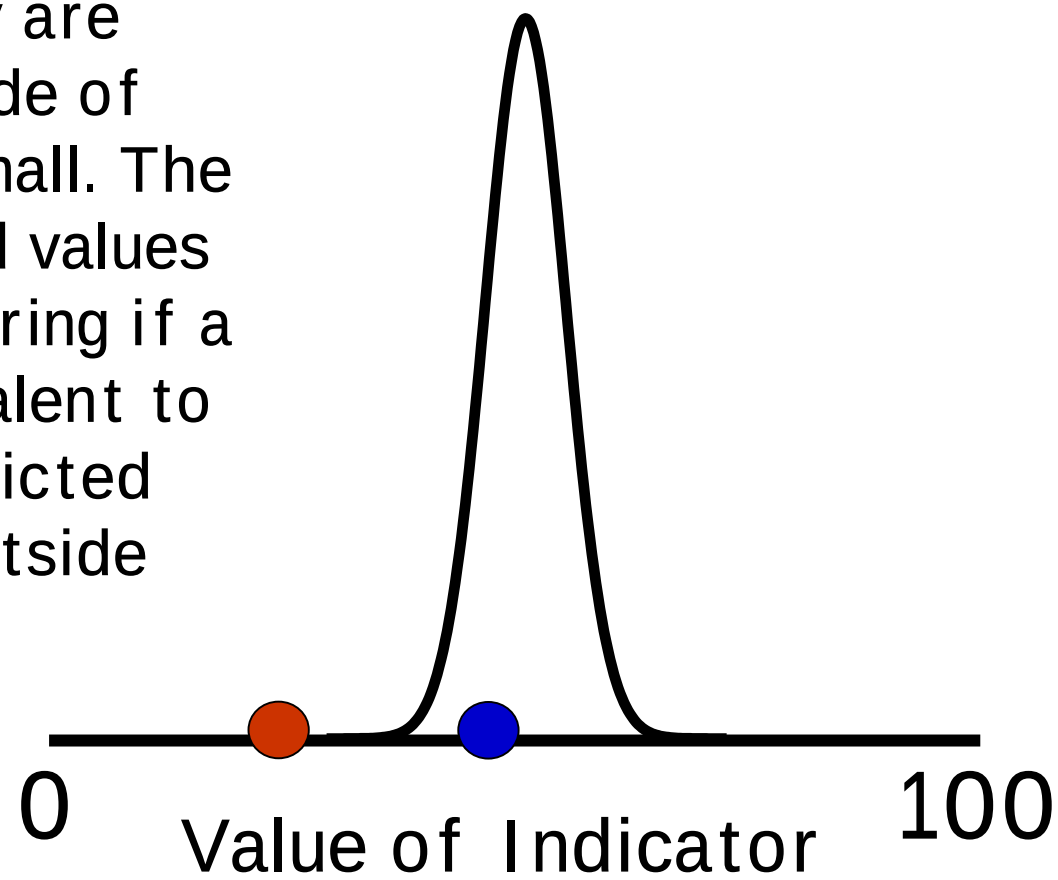
- Type I - false positives, i.e., reject the null hypothesis when it is true.

  costs the regulated community $

- Type II - false negatives, i.e., accept the null hypothesis when it is false.

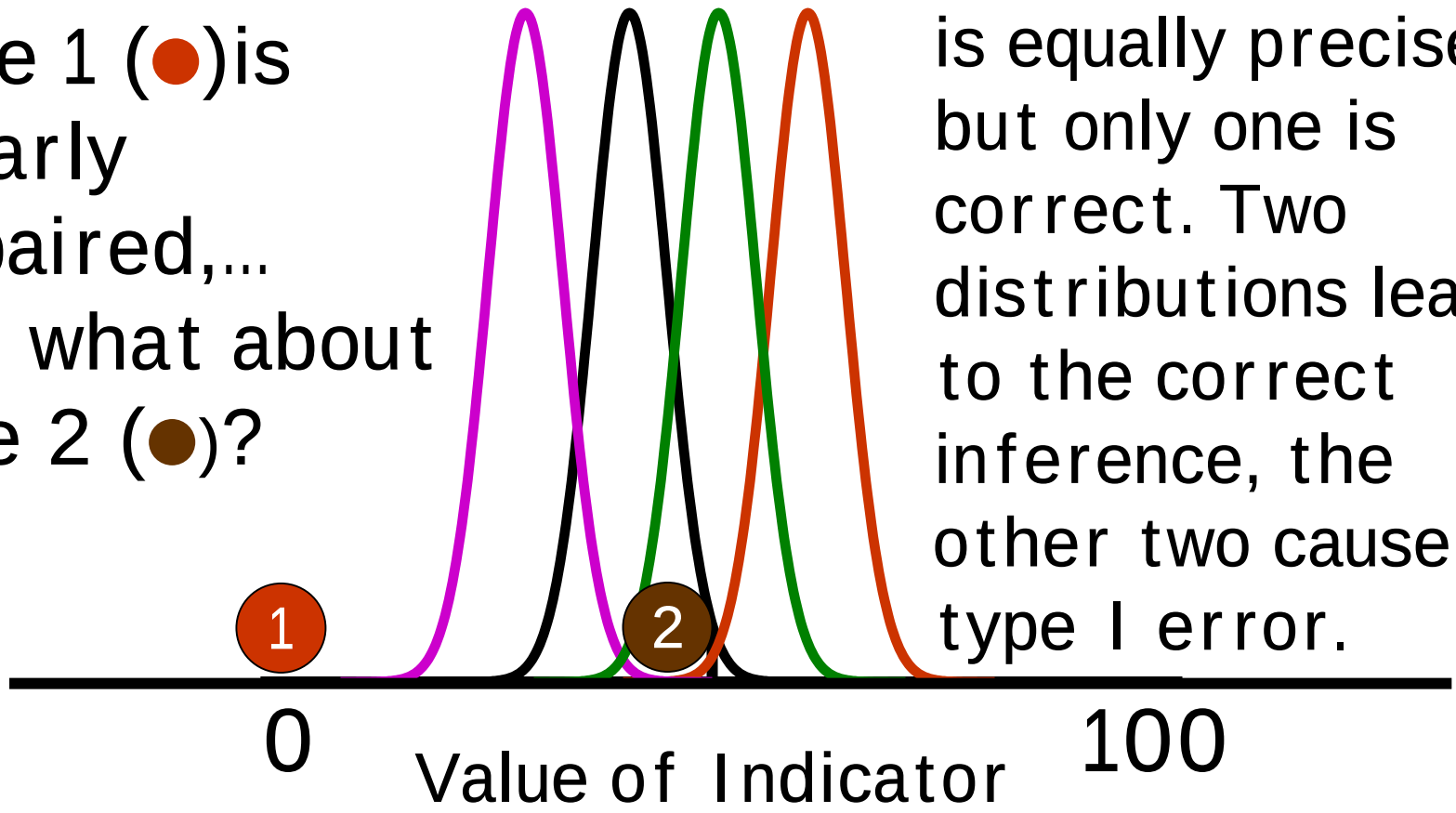
  results in continued degradation of the resource

Classifications or predictions are always made with some error or uncertainty. Predictions are accurate if, on average, they give the right result. They are precise if the magnitude of error or uncertainty is small. The distribution of predicted values forms the basis for inferring if a new observation is equivalent to reference (within predicted values) or impaired (outside predicted values).

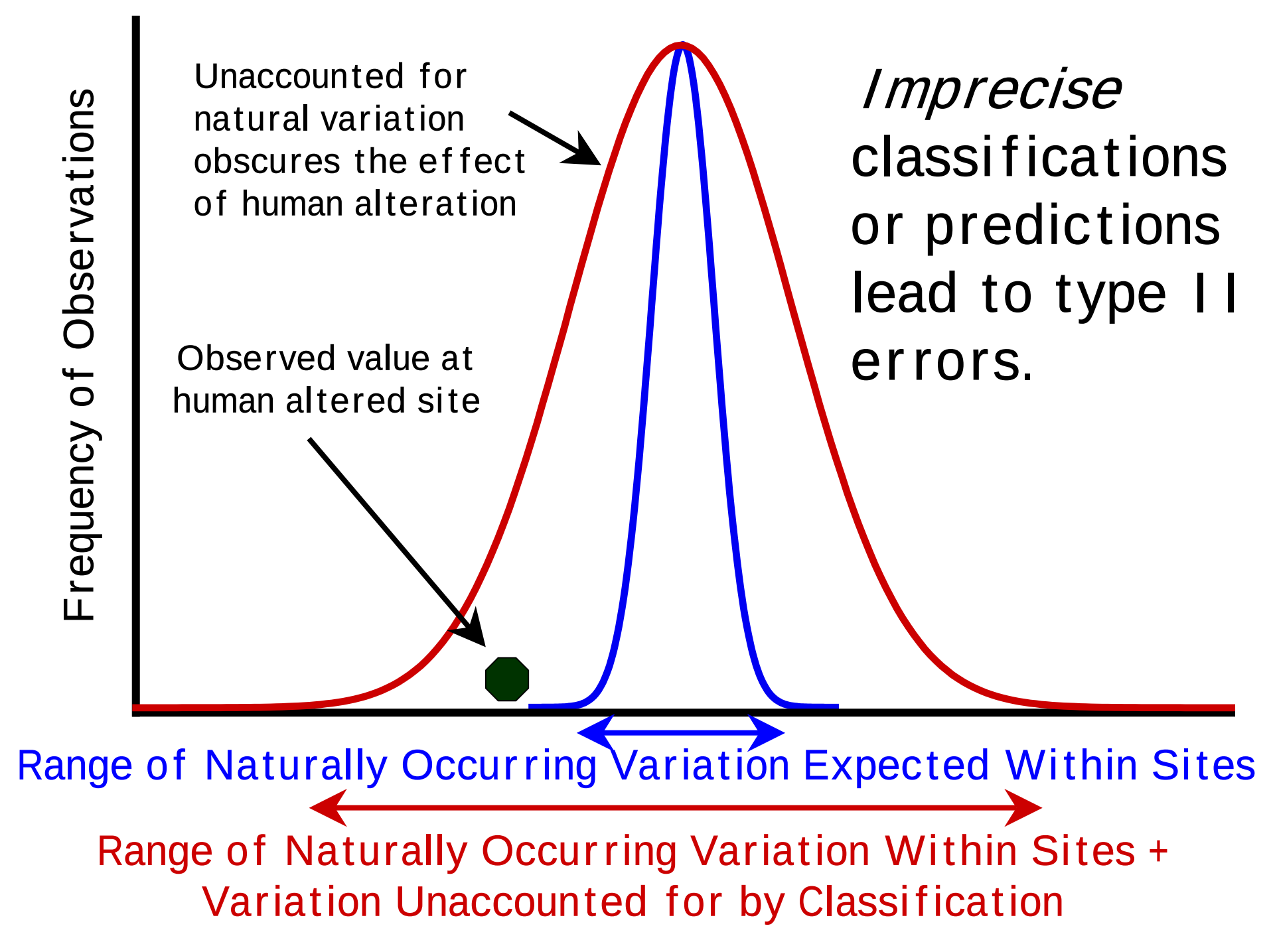


Classifications that produce *inaccurate* estimates of expected condition produce Type I errors

Site 1 (●)is clearly impaired,… but what about site 2 (●)?



Each distribution is equally precise, but only one is correct. Two distributions lead to the correct inference, the other two cause a type I error.


Frequency of Observations
Unaccounted for natural variation obscures the effect of human alteration
Observed value at human altered site
Imprecise classifications or predictions lead to type II errors.
Range of Naturally Occurring Variation Expected Within Sites
Range of Naturally Occurring Variation Within Sites + Variation Unaccounted for by Classification

Classification and modeling are 2 related approaches to estimating the reference (expected) condition

- Both approaches are used to make predictions regarding the conditions that should occur at sites lacking historical information.
  - Classification predicts the expected biotic condition of a waterbody from previously observed associations between biotic attributes and categorical descriptors of a waterbody's environmental setting. *Classification results in a finite number of predicted outcomes*.
  - Modeling predicts the expected biotic condition by mathematically describing how biota vary along environmental gradients. *Modeling results in an infinite number of predicted outcomes*.

In both approaches, environmental descriptors are used for prediction that are usually *surrogates* for the proximal environmental factors that actually influence biota.

Surrogate	Causal Factor	Response
Elevation →	Temperature →	Biota

The factors used for prediction can also vary in terms of the spatial and temporal scale at which they occur.

Classification/modeling systems can be:

- Single Factor or Multi-factor
- Single Scale or Multi-scale
- Hierarchical or Non-hierarchical

A few of the causal factors associated with naturally occurring biotic variation among sites:

Temperature

Sediment Size

Water Chemistry

Hydrology

Food type

The use of reference sites in classification and modeling

- The use of reference sites is an *empirical* approach to estimating reference condition.
- Accurate and precise predictions from reference site data depend on:
  - Agreed upon and acceptable criteria for defining reference site quality,
  - Acceptable means of extrapolating/interpolating.

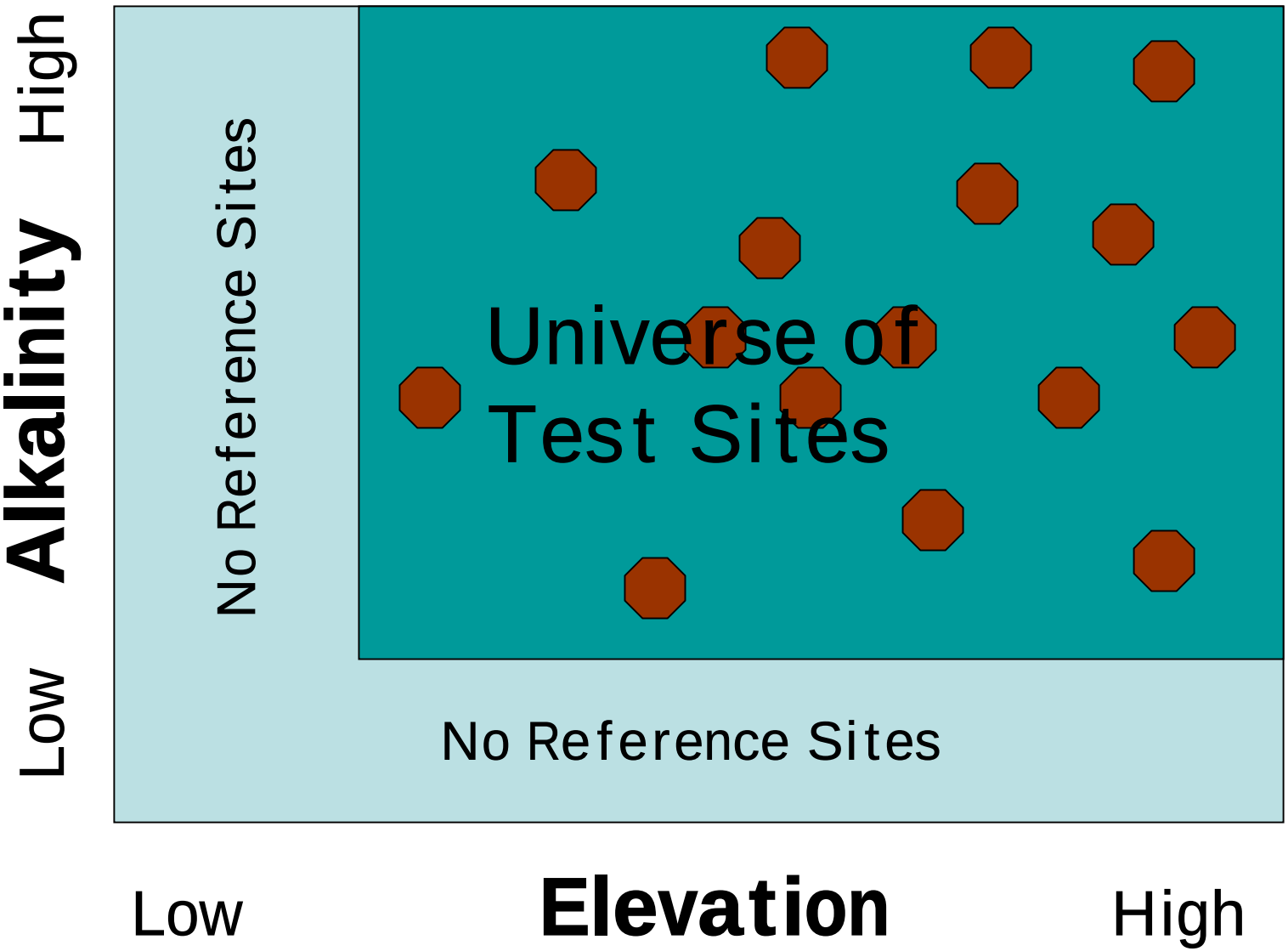
I will treat 2 issues that influence predictive accuracy and precision:

- Are reference sites representative of the resource of interest?
- How much natural variability do we need to account for?

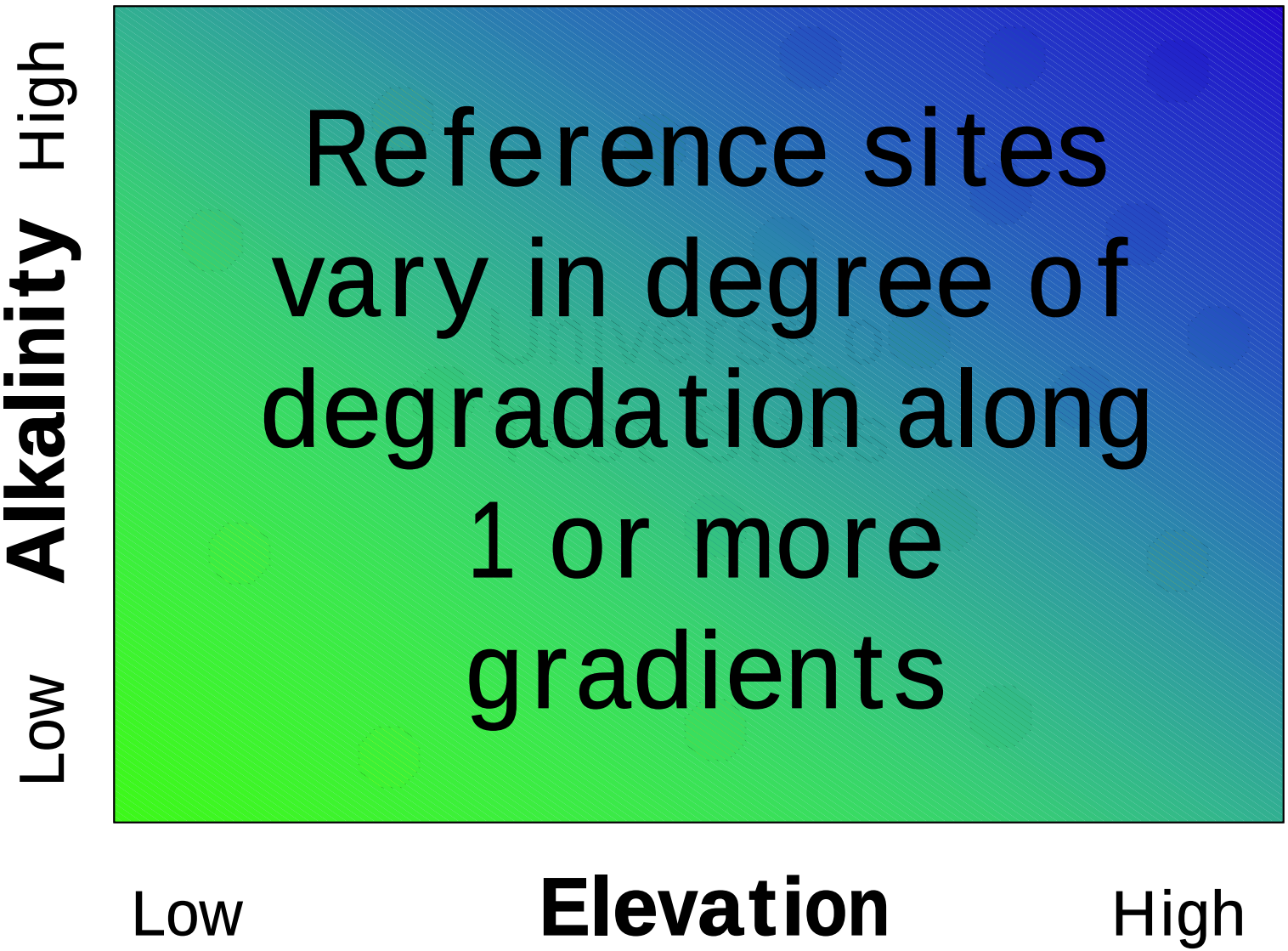
Are reference sites representative of the resource of interest?

- This problem really boils down to whether the range of environmental and biological conditions in the population of reference sites is equivalent to the range that would occur in the population of all other sites of interest.
- Reference site 'quality' will almost always vary across classes of sites, so we must be careful about what we mean by "reference".

Reference Representativeness



Reference Representativeness

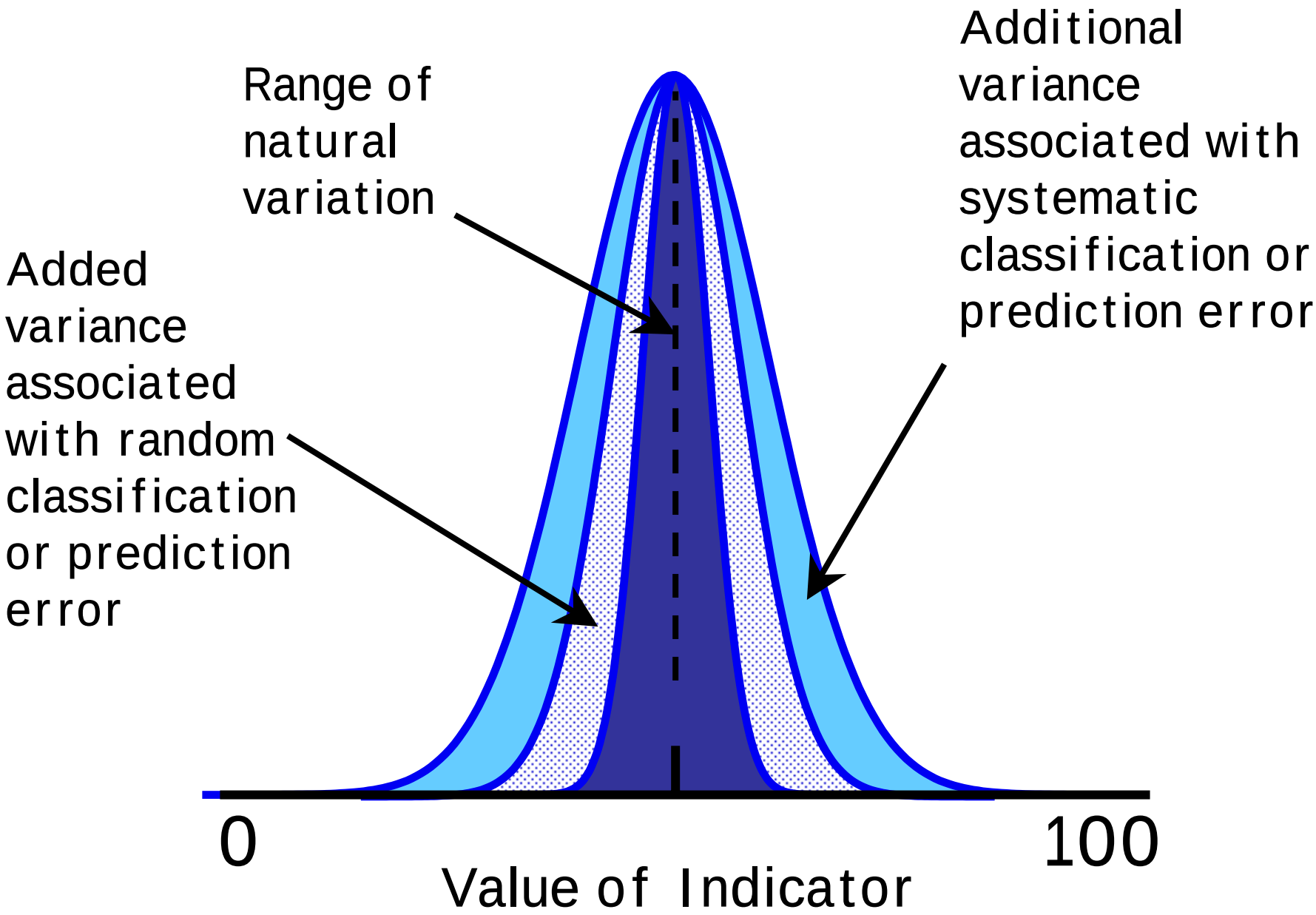


Accounting for natural variability - how much is enough?

- How much we need to account for is a function of how small of a response we want/need to detect, which needs to be decided by stakeholders up front!!!

The expected condition of a site will always be somewhat "fuzzy" because:

- Unimpaired sites are not static - they are in dynamic equilibria.
- Measurement error associated with estimating the value of an indicator.
- *Variance associated with the effects of unmeasured naturally occurring factors.*


Range of natural variation
Additional variance associated with systematic classification or prediction error
Added variance associated with random classification or prediction error
0
100
Value of Indicator

The goal of classification and modeling is to partition, and thus control for, the effects of natural factors.

What is the appropriate scale?

Variance in invertebrate assemblage composition among 694 Swedish streams

(Sandin & Johnson 2001)

Spatial Scale	% Variance
Regional	23
Catchment	32
Local	45

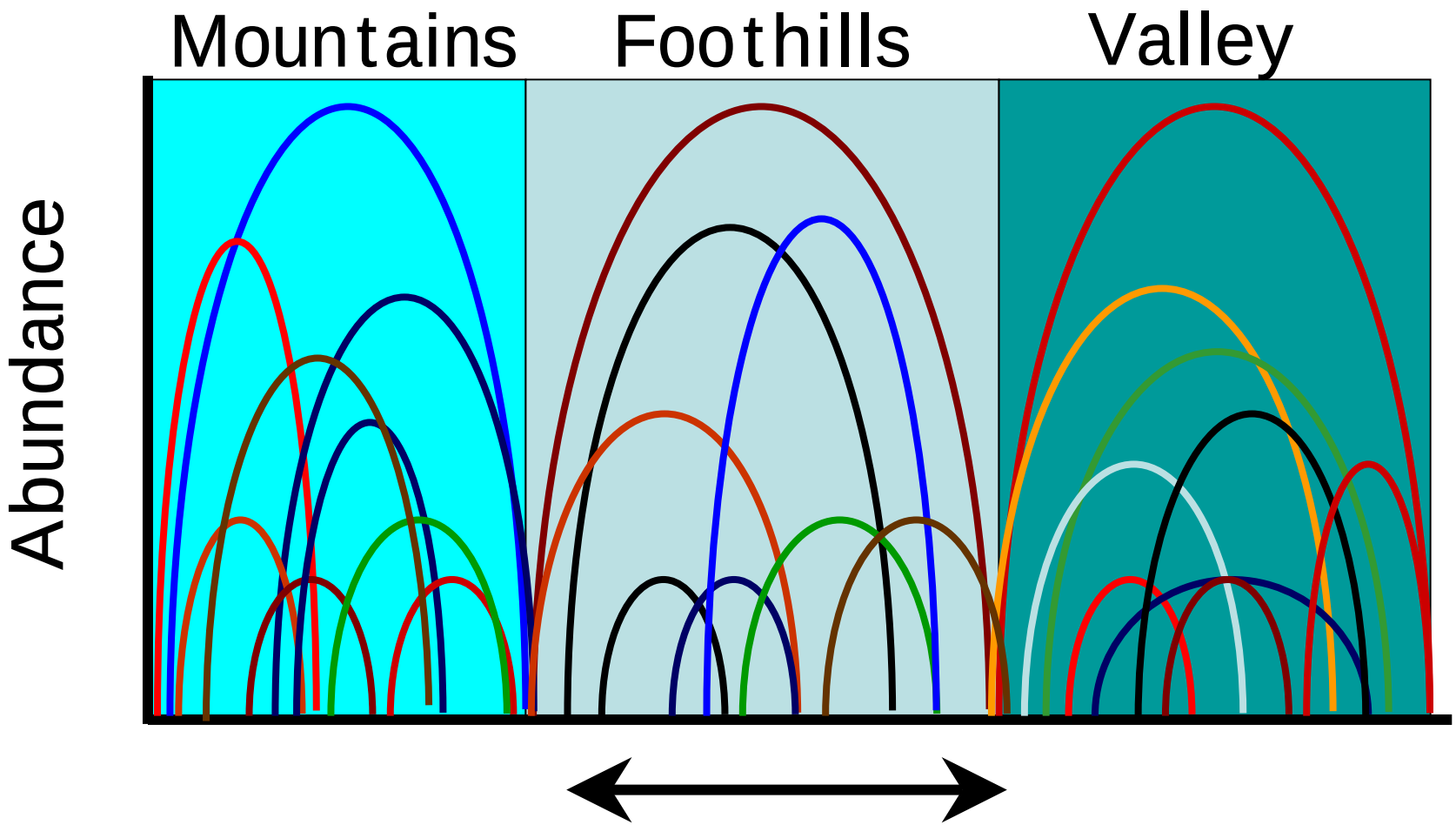
Regional = latitude, longitude, ecoregion
Catchment = geology, landuse
Local = velocity, depth, elevation, substrate, water chemistry

The data of Sandin and Johnson imply that we should consider factors that operate at several spatial scales if we are to effectively partition naturally occurring variation.

Should we use classification or modeling?

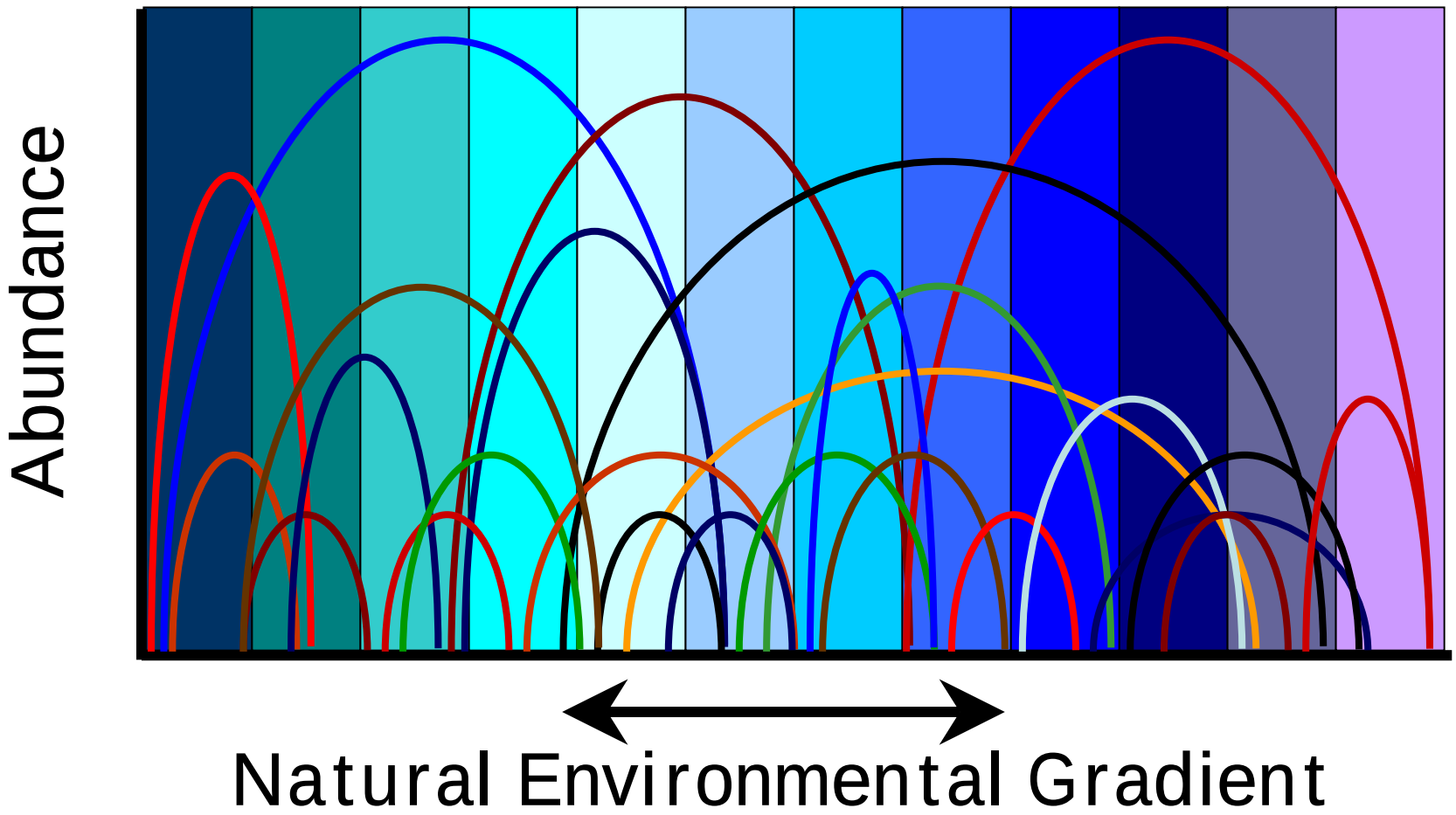
The answer depends on how good our assumptions are regarding how the natural world is structured and organized.

Aggregated Distributions of Species



Individualistic distributions of species

How many classes?



Much work has been, and continues to be, directed toward determining the most effective methods for classifying waterbodys (e.g., ecoregions, bioregions, stream size).

Specific examples will be given by others following this talk.

The most frequently used approach to predict the taxa expected at a site from models is based on the RIVPACS method (*River InVertebrate Prediction and Classification System)*

Treatment of RIVPACS will be covered in Index Development 101 and 201.

Questions