United States
Environmental Protection
Agency

Overview: Empirical Approaches for Nutrient Criteria Derivation

Lester L. Yuan, John F. Paul, and Dana Thomas

Office of Research and Development
National Center for Environmental Assessment and National Health and Environmental Effects Research Laboratory

September 7, 2009

United States Environmental Protection Agency

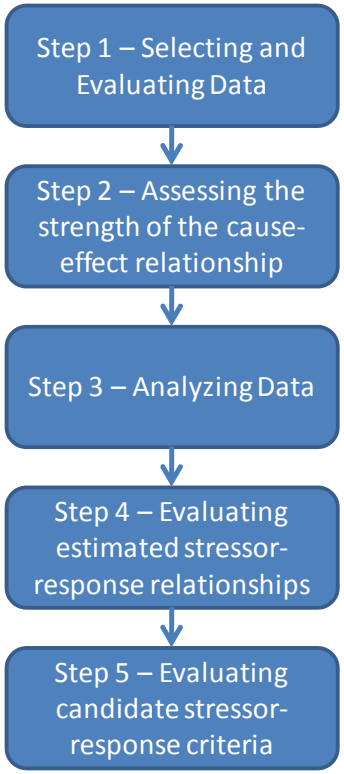
Purpose of the document

- Provide guidance on using known and accepted statistical methods to estimate stressor-response relationships from field data and on deriving nutrient criteria from these relationships to support the advancement of environmental protection and achieve water quality goals.

United States Environmental Protection Agency

Five step framework for using stressor-response relationships to derive nutrient criteria

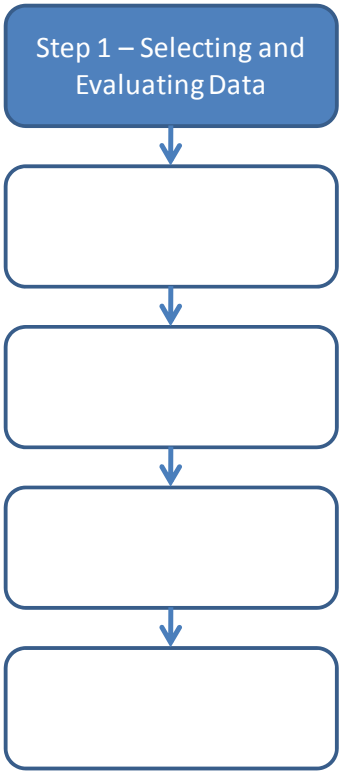
1. Select and evaluate data
2. Assess strength of cause-effect relationship
3. Analyze data
4. Evaluate estimated stressor-response relationships
5. Evaluate candidate stressor-response criteria



United States Environmental Protection Agency

Step 1: Select and evaluate data

- Identify variables that quantify nutrient concentrations and a response variable that relates to a water quality goal.
  - Nutrient concentration variables: (e.g.,TN, TP)
  - Response variables:
    - Relate to water quality goal (e.g., aquatic life protection)
    - Causally relate to changes in nutrient concentrations
- Explore data set to understand relationships between nutrient variables, response variables, and possible confounding variables.



United States Environmental Protection Agency

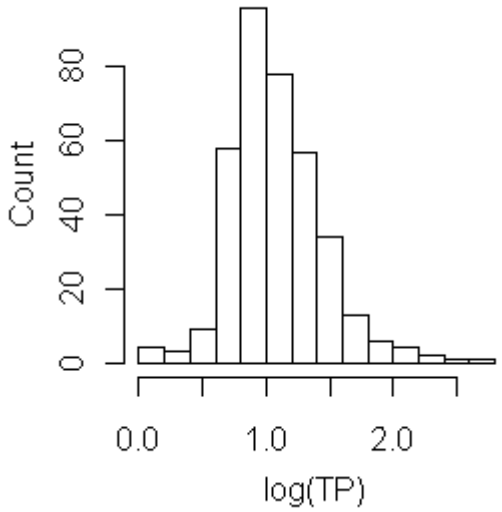
Identify variables

- Example Dataset: Lakes from the Northeast United States, sampled by the Environmental Monitoring and Assessment Program
- Nutrient variable: total phosphorus (TP)
- Response variable: chlorophyll a (chl-a)
  - *Assume for illustrative purposes that chl-a = 15 µg/L is the threshold to support aquatic life protection in Northeast Lakes.*

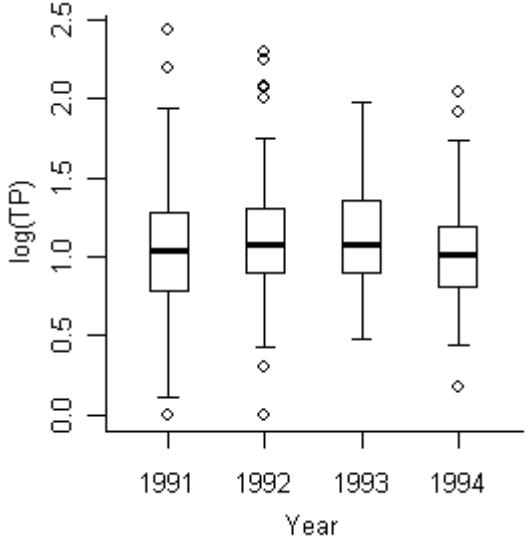
United States Environmental Protection Agency

Examine data distributions

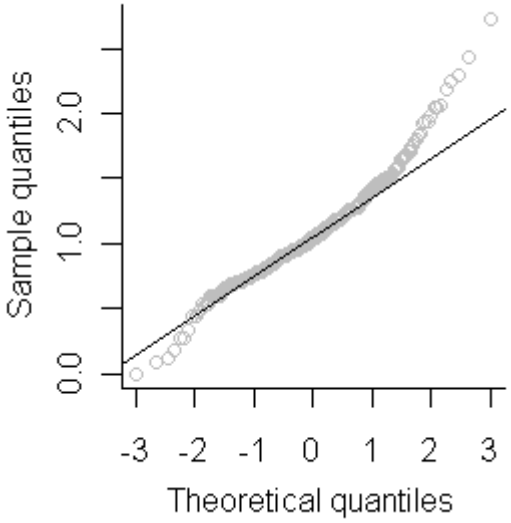
Histograms



Boxplots

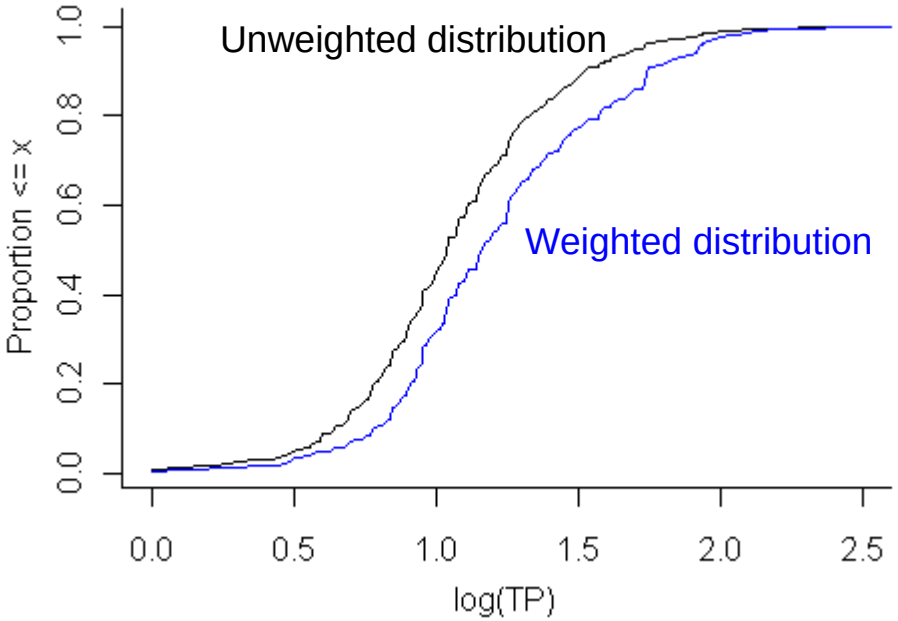


Q-Q plots



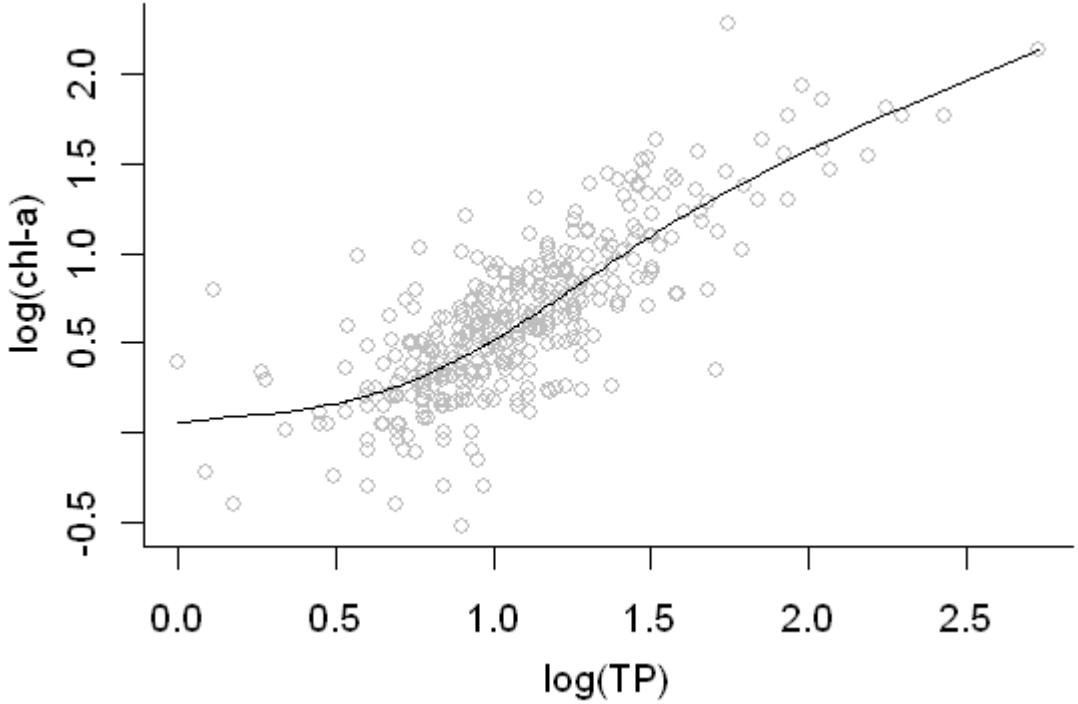
Cumulative distribution functions

- View distribution of data and allows for weighting of data from probability samples.

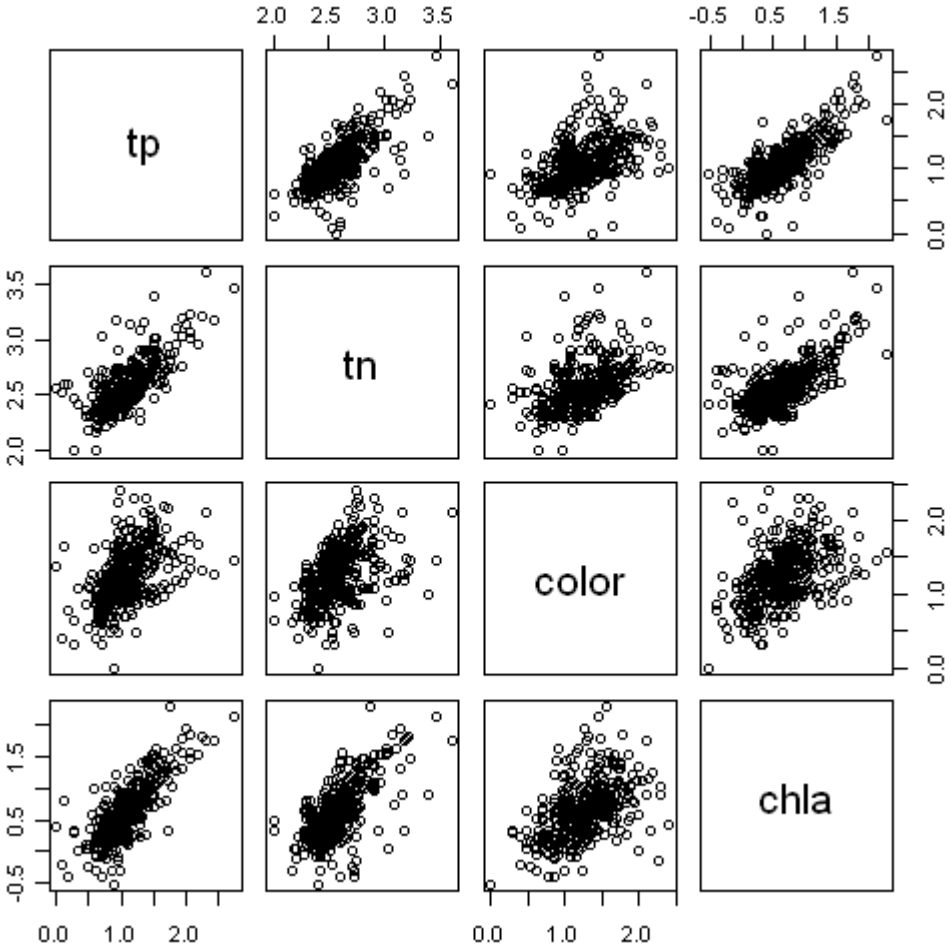


Fit nonparametric curves to data

- Helps visualize relationships between variables

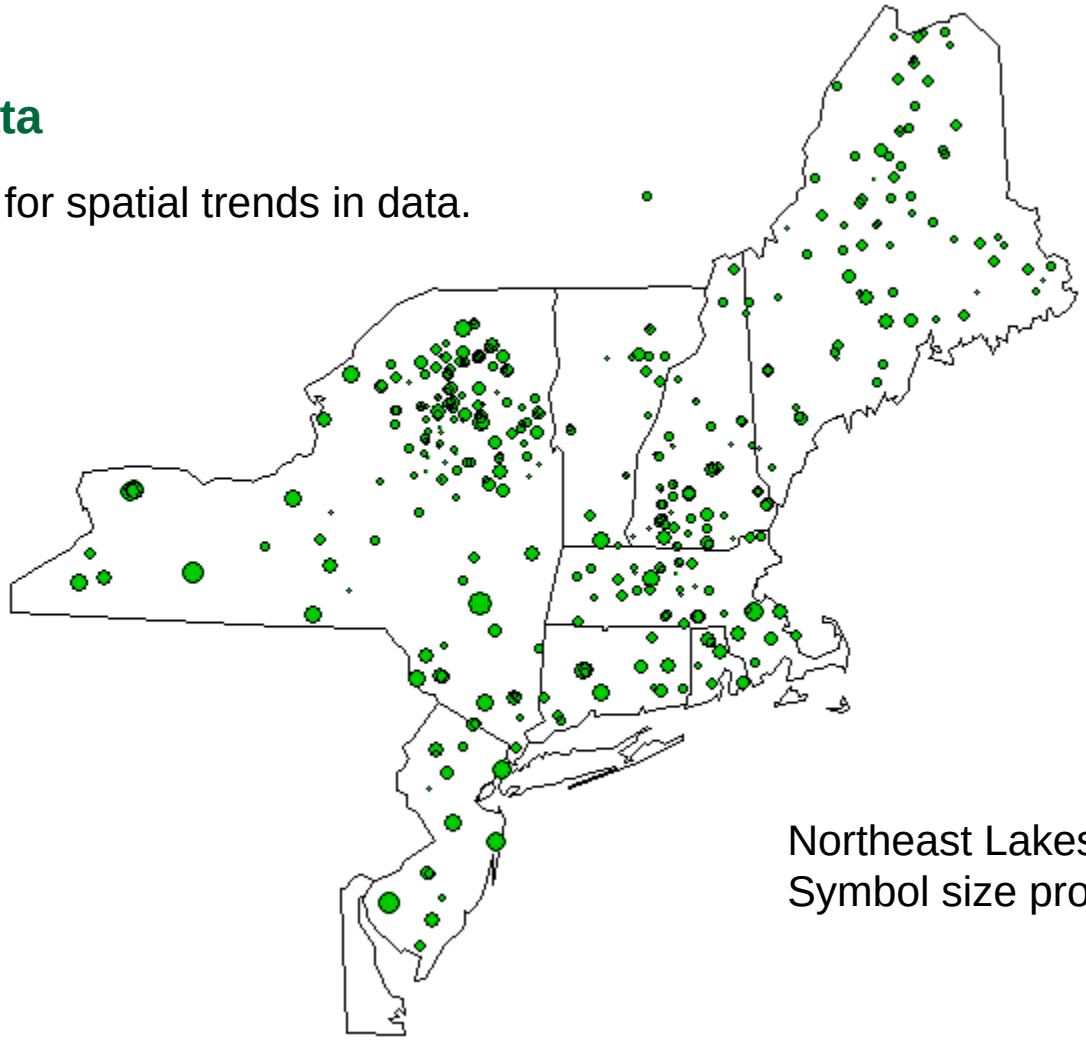


Examine covariance among several different variables



Map data

Look for spatial trends in data.

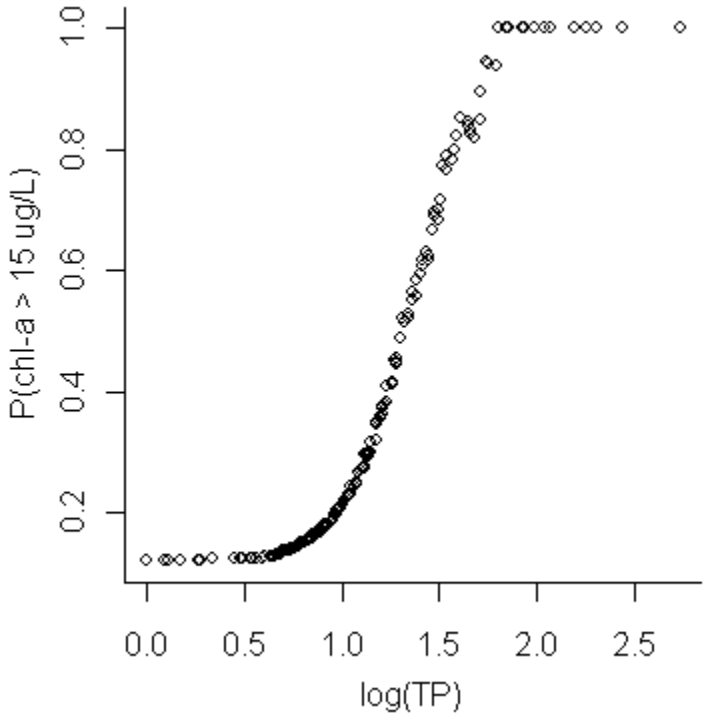


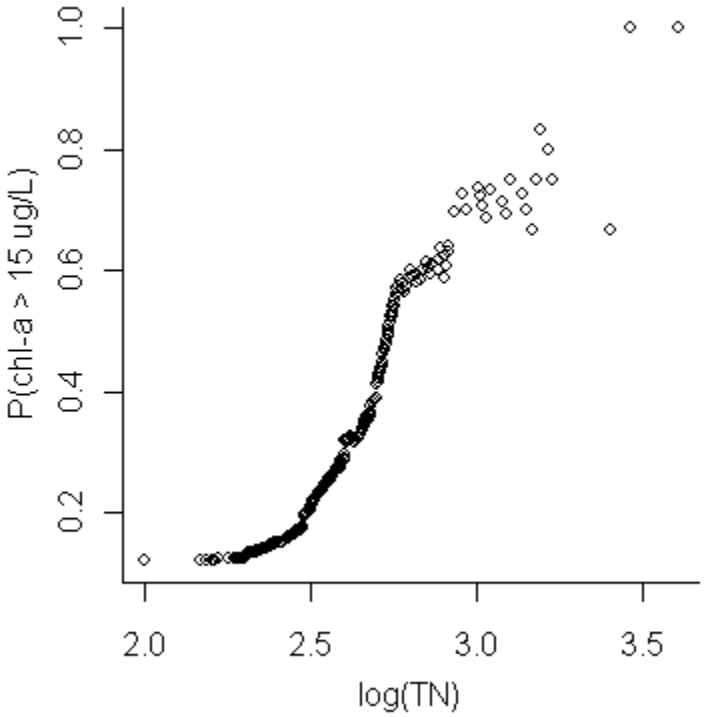
Northeast Lakes
Symbol size proportional to log(chl-a)

United States Environmental Protection Agency

Conditional probability analysis (CPA)

CPA quantifies the probability of not supporting water quality goal, given different nutrient concentrations.

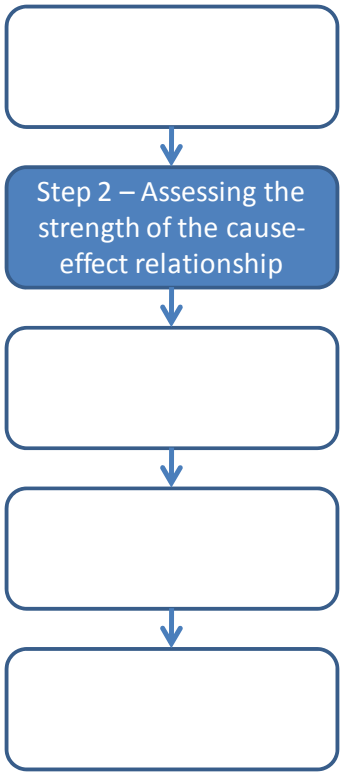


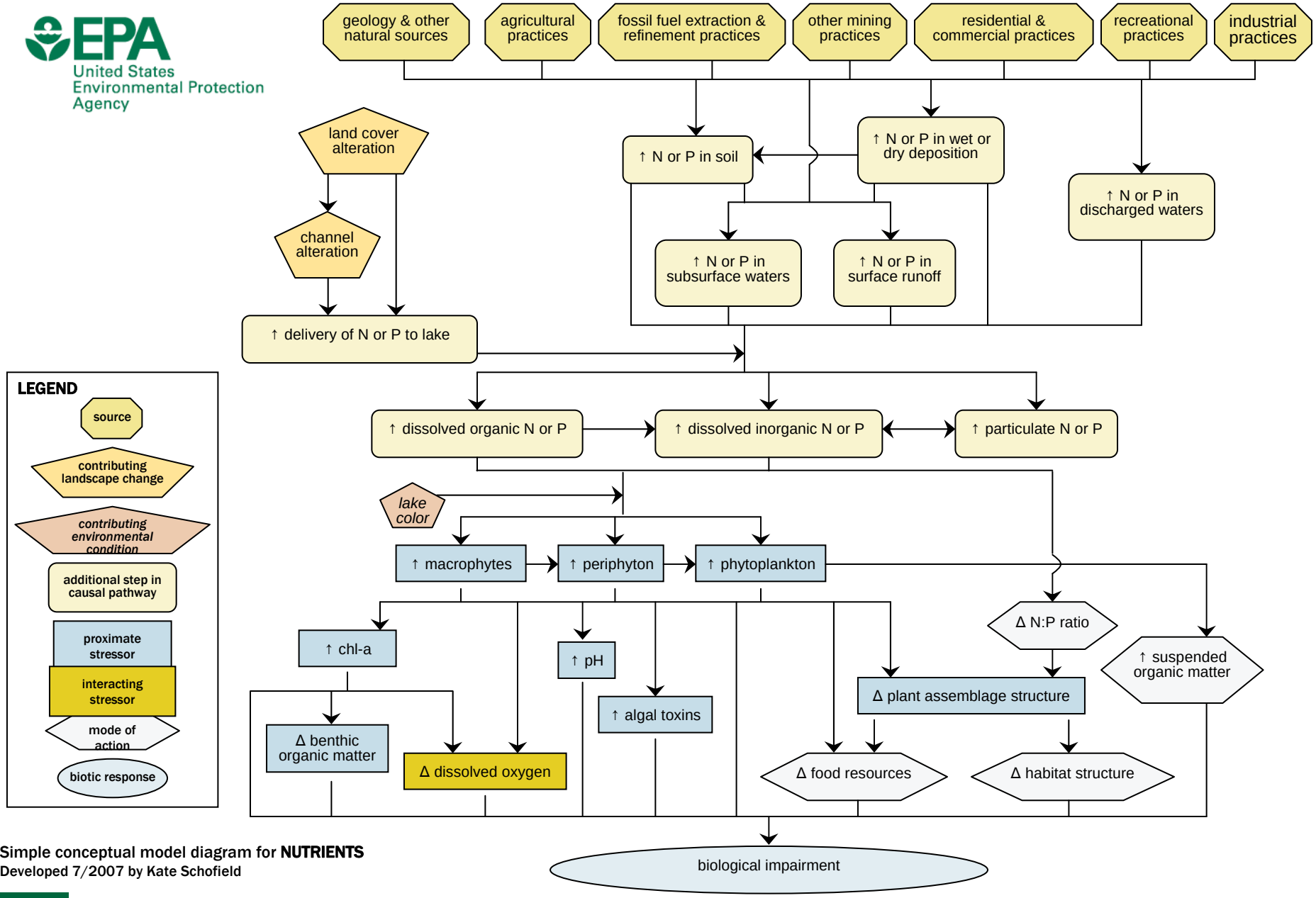


United States Environmental Protection Agency

Step 2: Assess strength of the cause-effect relationship

- Conceptual model: Describe sequence of relationships by which nutrients affect response variable
- Existing literature: Document cause-effect relationships observed in other studies.
- Additional analyses:
  - Structural Equation Modeling
  - Propensity Score Analysis





Simple conceptual model diagram for NUTRIENTS
Developed 7/2007 by Kate Schofield

United States Environmental Protection Agency

Cause-effect relationships between TP and chl-a in literature

For example, increased total phosphorus increases chlorophyll a concentration

Supporting evidence includes:

- Whole lake fertilization experiments (Schindler 1974, 1977; Stockner and Shortreed, 1985; Carpenter et al. 1996)
- Consistent relationships estimated from observational data collected from many different locations.

United States Environmental Protection Agency

Step 3: Analyze data

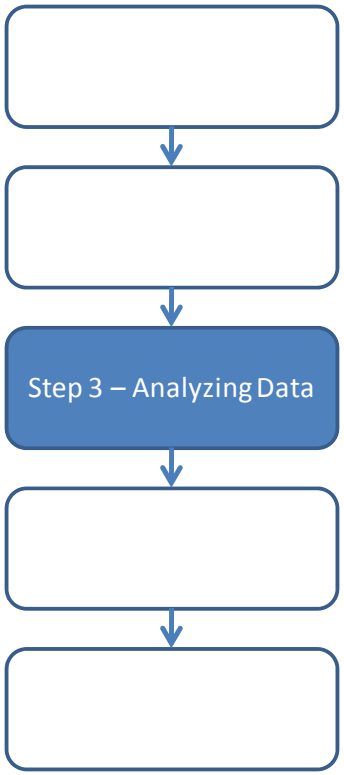
Two distinct approaches for informing criterion derivation from a stressor-response relationship.

1. Threshold is known for response variable

   Stressor-response relationship is used to "translate" response threshold to a protective nutrient concentration.

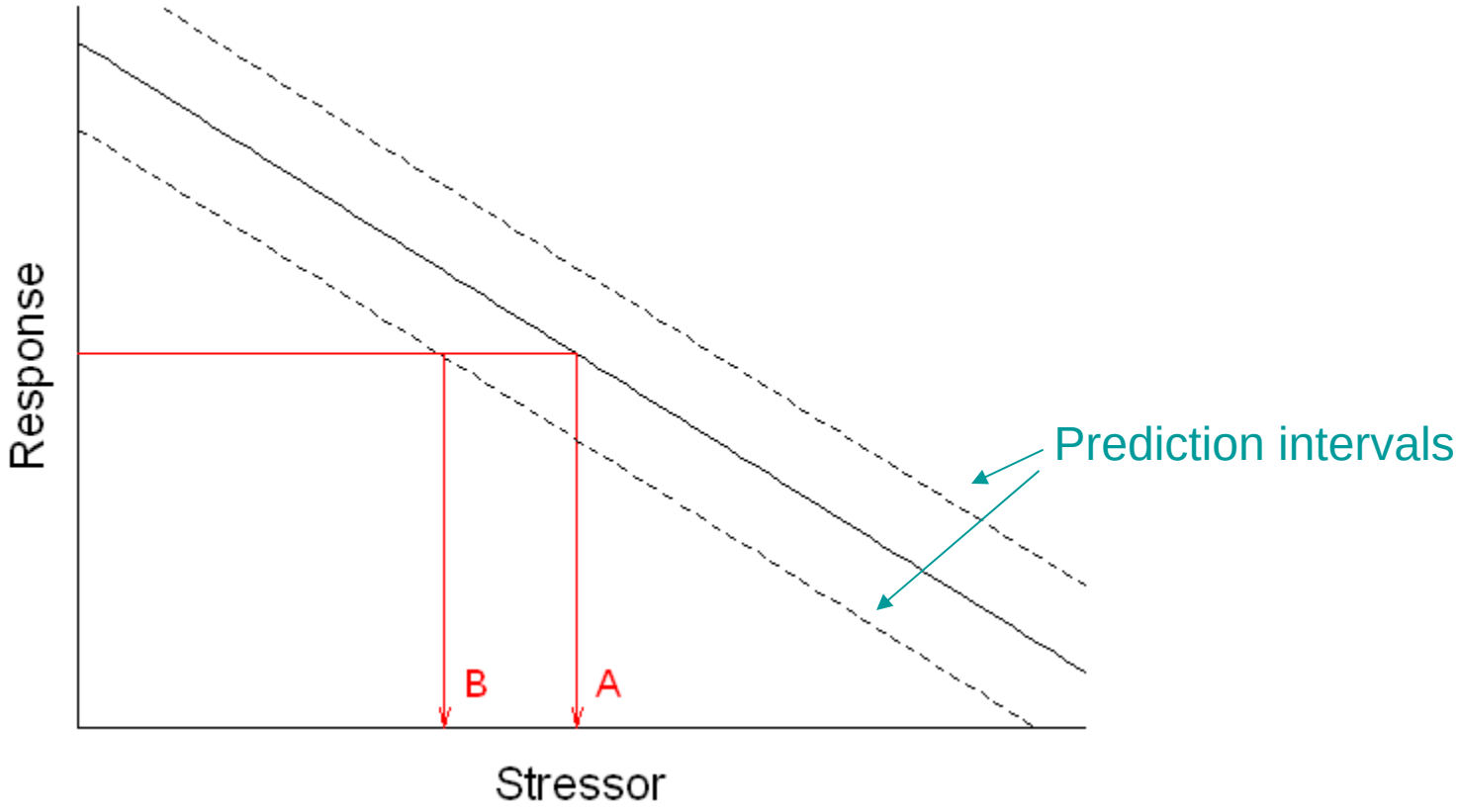
2. No threshold available for response variable

   Use characteristics of the stressor-response relationship to inform the derivation of a criterion.



Step 3: Analyze data

Translate a known response threshold



United States Environmental Protection Agency

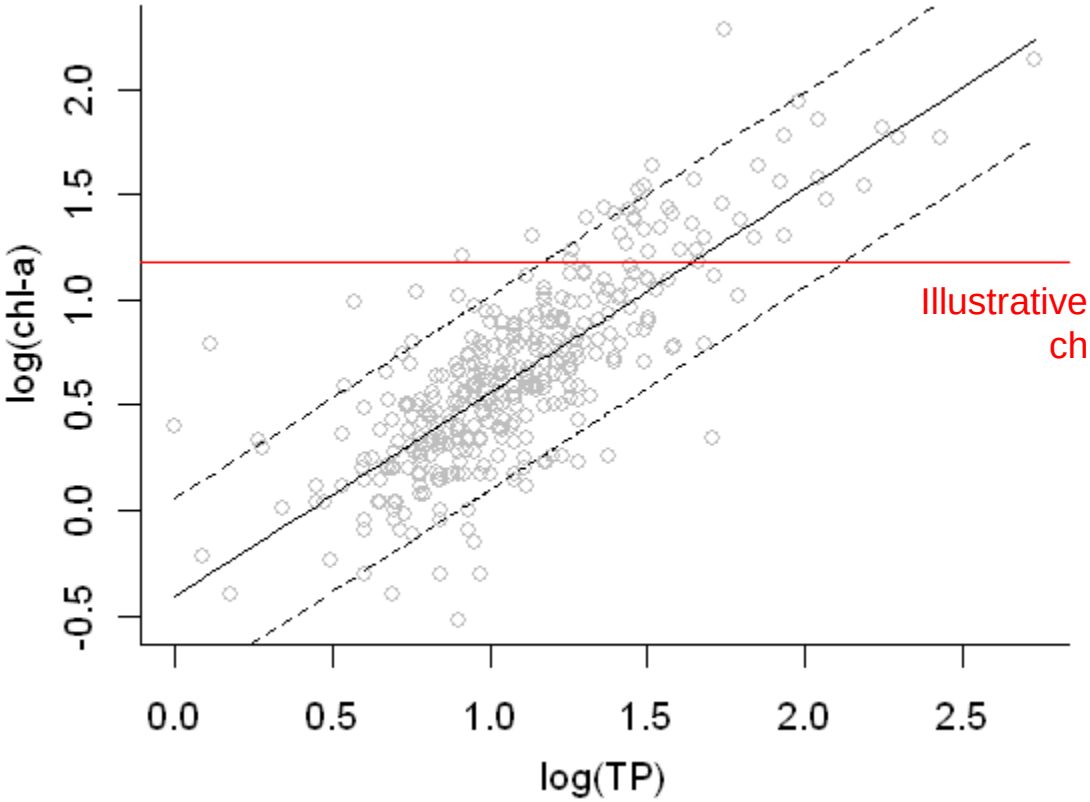
Translate a known response threshold:

Methods for estimating the stressor-response relationship

- Simple linear regression
- Quantile regression
- Logistic regression

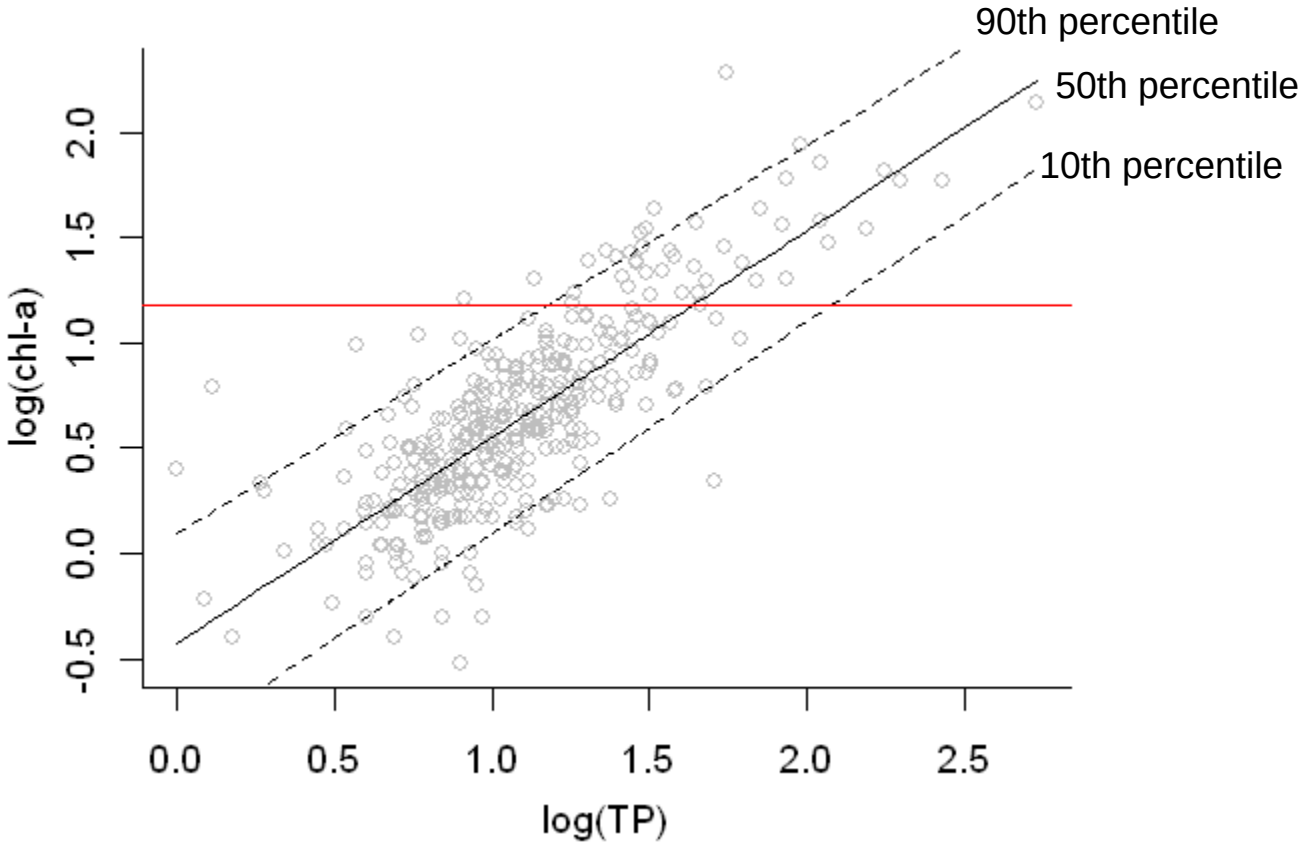
- Multiple linear regression

Simple linear regression



Illustrative response threshold:
chl-a = 15 µg/L

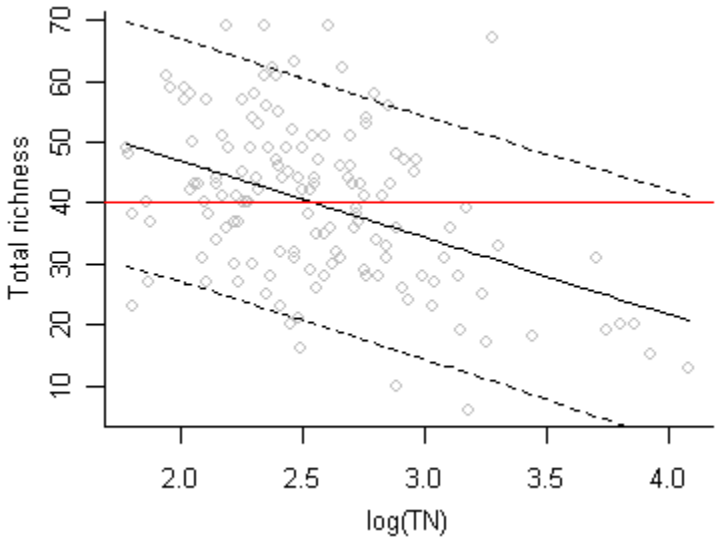
Quantile regression



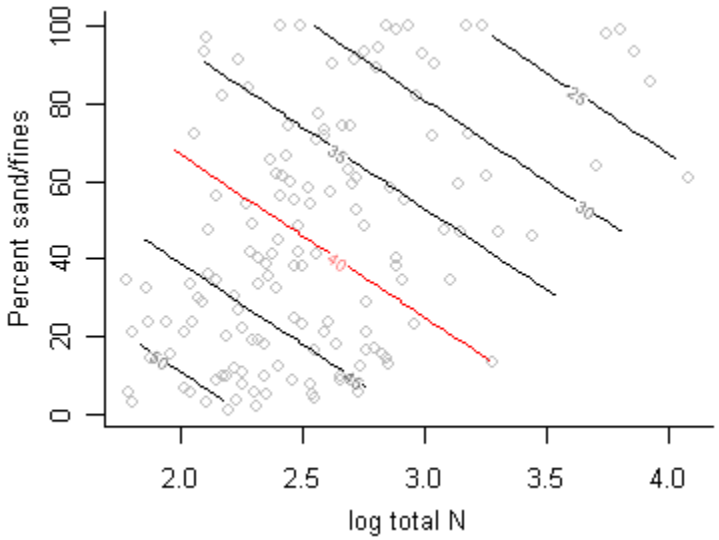
Multiple linear regression

Multiple linear regression models can be useful when nutrients are one of many factors influencing biota.

Many factors, in addition to nutrients, influence macroinvertebrate richness:



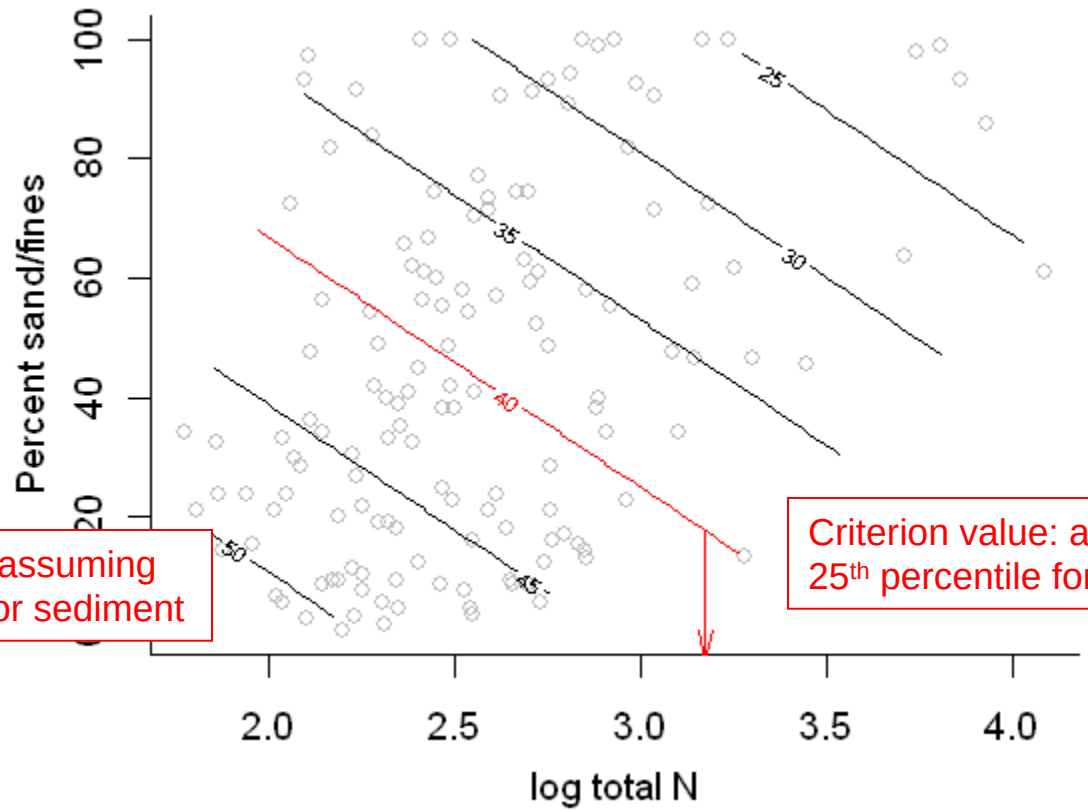
$R^2 = 0.19$



$R^2 = 0.32$

Data from EMAP-West

Use results from multiple linear regression to translate an existing response threshold



Contour lines indicate modeled mean macroinvertebrate richness.

Criterion value: assuming 75th percentile for sediment

Criterion value: assuming 25th percentile for sediment

United States Environmental Protection Agency

Step 3: Analyze data

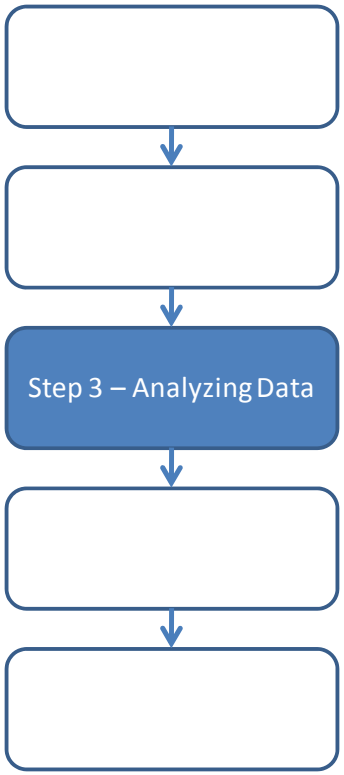
Two distinct approaches for informing criterion derivation from a stressor-response relationship.

1. Threshold is known for response variable

   Stressor-response relationship is used to "translate" response threshold to a protective nutrient concentration.

2. No threshold available for response variable

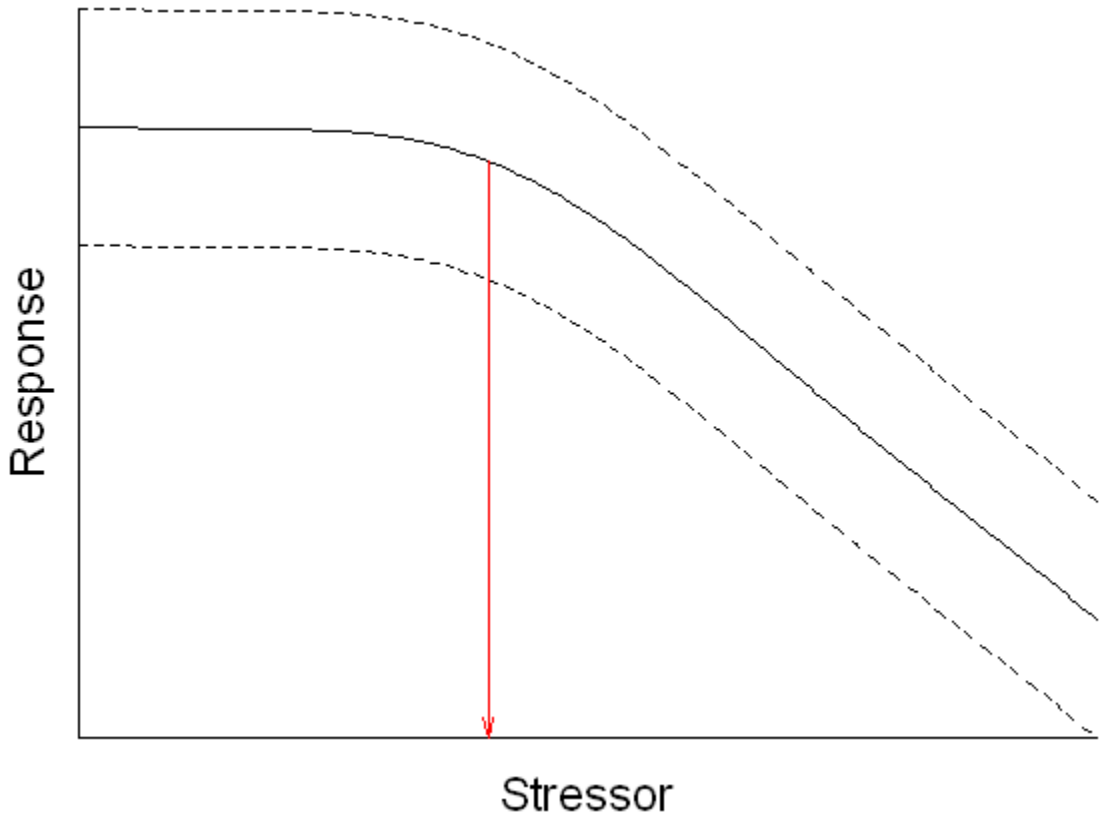
   Use characteristics of the stressor-response relationship to inform the derivation of a criterion.



United States Environmental Protection Agency

Step 3: Analyze data

Use characteristics of the stressor-response relationship

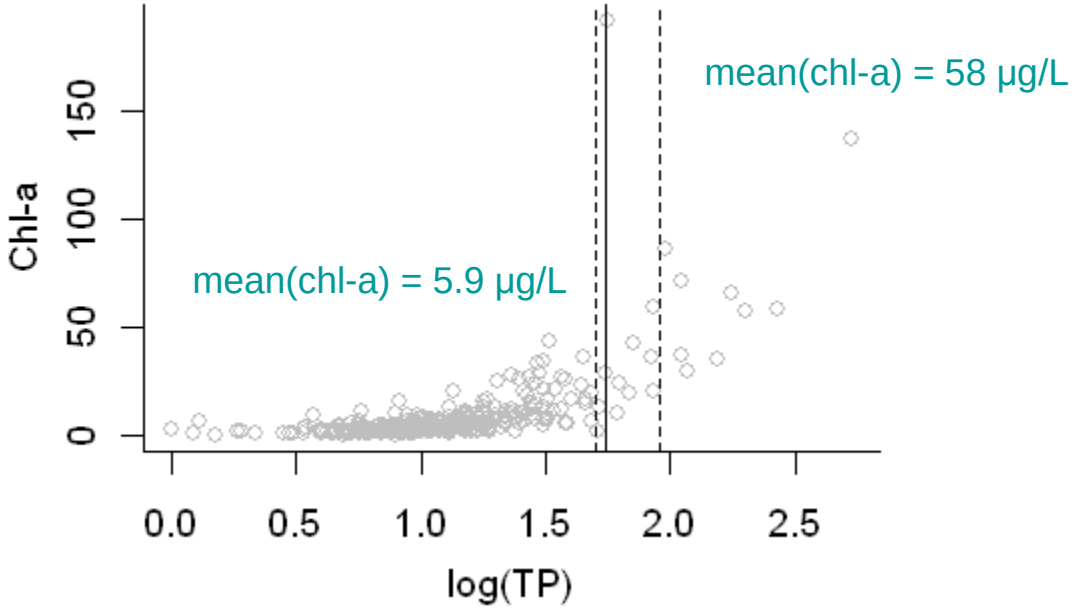


United States Environmental Protection Agency

Methods for identifying characteristics of the stressor-response curve

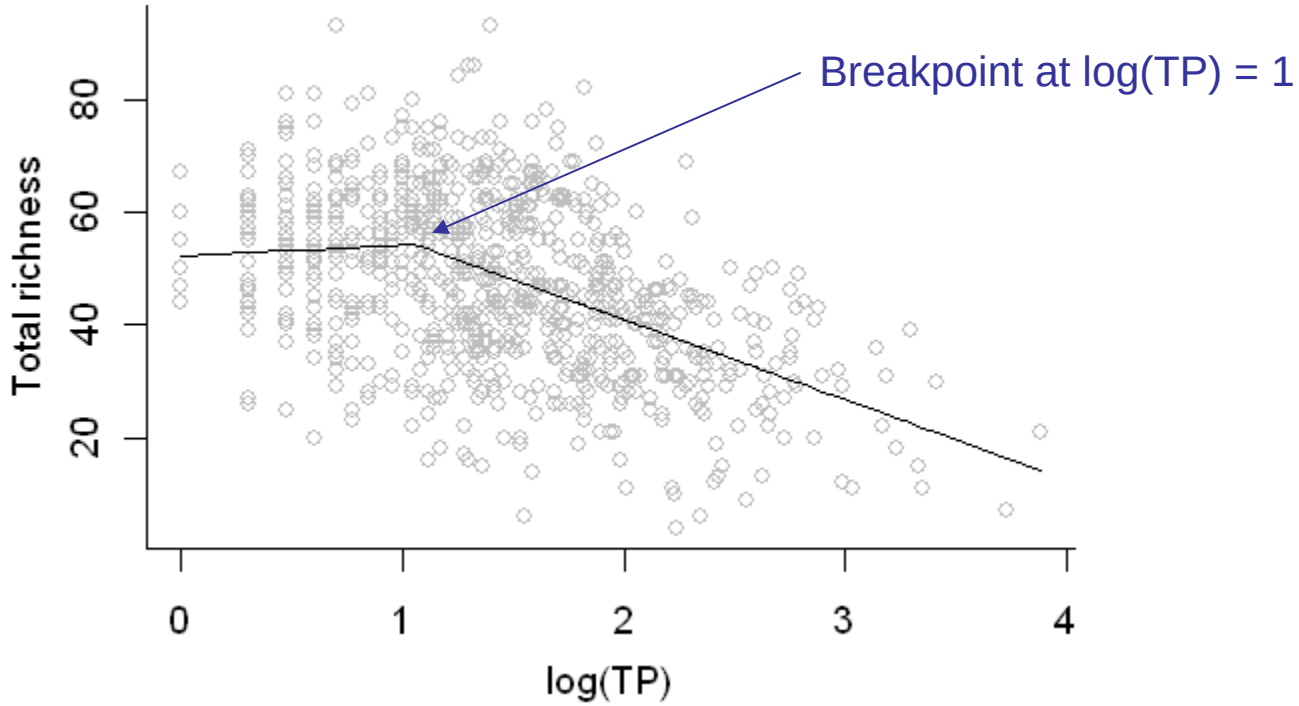
- Nonparametric changepoint analysis
- Discontinuous regression models

Nonparametric changepoint analysis



Data from Northeast Lakes

Discontinuous linear regression

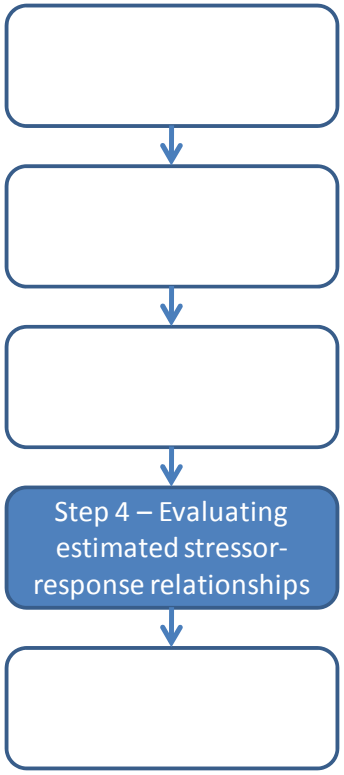


Data from EMAP-West

United States Environmental Protection Agency

Step 4: Evaluate estimated stressor-response relationships

- Validation:
  - Apply models to independent data to quantify predictive accuracy.
- Assess uncertainty in model
  - Model assumptions (e.g., space-for-time substitution, regression model assumptions)
  - Model structure: included variables and functional form
- Select stressor-response model



United States Environmental Protection Agency

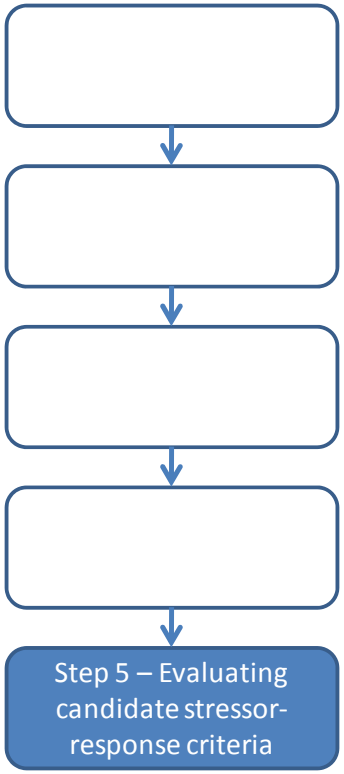
Evaluating models for Northeast Lakes

- Quantile regression and simple linear regression provide similar results.
  - *Regression assumptions are reasonably supported*
- Multiple regression model does not substantially increase explanatory power of the model.
  - *Simple linear regression provides representative model*.
- Select simple linear regression model of log(chl-a) vs. log(TP)

United States Environmental Protection Agency

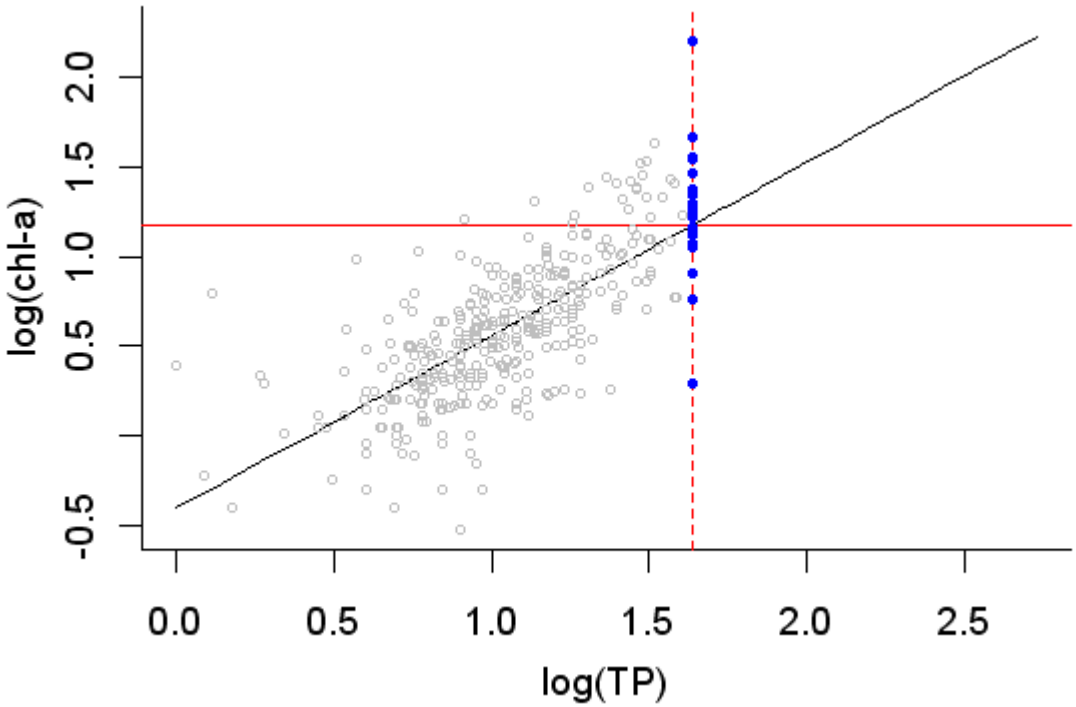
Step 5: Evaluate candidate stressor-response criteria

- Estimate the predicted effects of different candidate criteria to achieve water quality goals.



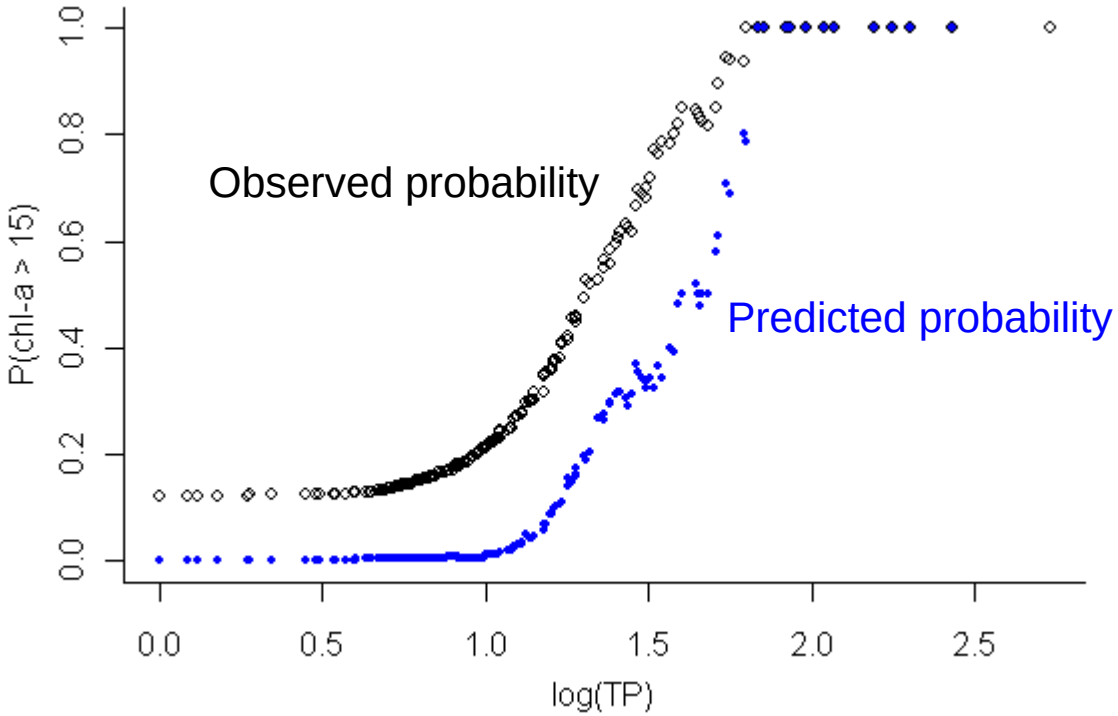
United States Environmental Protection Agency

Predicting the effects of a criterion value on achieving water quality goals

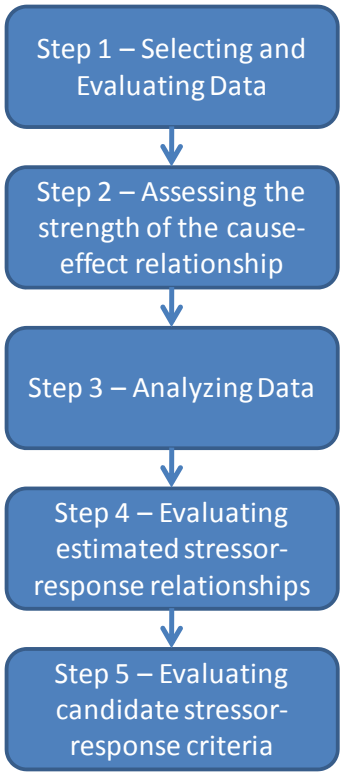


Assess effects of different candidate criteria

Compare *observed* probabilities of not achieving water quality goals with *predicted* probabilities, assuming that nutrient concentrations are reduced to a specified criterion value.



United States Environmental Protection Agency

Conclusions

- Statistical methods presented within a five-step framework for deriving nutrient criteria.
- Resulting analyses provide information useful for decisions regarding final nutrient criteria.