Conditional Probability

N-STEPS Objectives

- Provide regions, states, and tribes with support related to nutrient criteria development
- Provide access to expert assistance with issues related to nutrient criteria development and implementation
- Improve communication nationwide.

What Is It?

A conditional probability is the probability of an event occurring when some other event has occurred and conditional probability analysis is an approach that allows a user to estimate the likelihood of exceeding some response threshold for a given nutrient concentration. This approach is based on the assumption that as nutrient concentrations increase, the likelihood of an impact on some negative response increases. This approach has great promise in identifying appropriate nutrient criteria to other measures where criteria thresholds have already been developed – for example, dissolved oxygen, pH, or biocriteria.

Example Question: What is the probability of exceeding our dissolved oxygen criterion over a range of different nutrient concentrations and where is the threshold?

How is it Applied to Nutrient Criteria Development?

Nutrient criteria development involves three main processes: identifying relationships between biological responses and nutrient stressors, examining these relationships, and establishing nutrient and/or biological thresholds or criteria.

Conditional probability analysis (CPA) is statistical tool to evaluate the relative risk of impairment of biological attributes along a nutrient gradient. It calculates the probability of exceeding a given threshold (e.g., biocriterion) given a set nutrient concentration; it can be used, therefore, to identify which nutrient conditions have a significant probability of being associated with adverse biological conditions. By combining CPA with change-point analysis, one can identify the nutrient threshold indicating a high risk of exceeding some other criterion.

How Does It Work?

Conditional probability calculates the probability of an event occurring (e.g., DO<y) when it is known that some other event has occurred (e.g., TP>x). The nutrient concentrations are treated as discrete random variables and probability functions are calculated. Functionally, a two-step procedure is used. One identifies the subset of samples were nutrients exceed some threshold and from those sites, one determines the samples which have exceeded the response threshold. This is the subset of samples for which nutrients exceed some value (x), which are also impacted for the response variable (y). Calculating these values iteratively over the range of nutrient concentrations generated an empirical conditional probability curve. Confidence intervals can be generated for this curve using resampling techniques, like bootstrapping (Paul and McDonald 2005). Thresholds along this curve are identified using variations of change-point analysis (See other fact sheet on change-point analysis).

Data Requirements

Independently collected numeric data in the form of paired observations are required. These are preferably continuous data, although discrete numeric variables (e.g., taxa richness) could also work. The greater the range of environmental conditions encompassed the better. One way to assure a large range is to use a gradient design and select sites along as large a gradient as possible.

What Should You Look For & Report?

Examine the conditional probability plots. They should show a clear trend in response with changes in nutrient concentration. Steep relationships and clear changes in the response make threshold identification easier. Confidence intervals can be used to help identify thresholds (see Paul and McDonald 2005). Confidence intervals for the change-points can also be reported (see fact sheet on Change-Point Analysis), depending on the approach used.

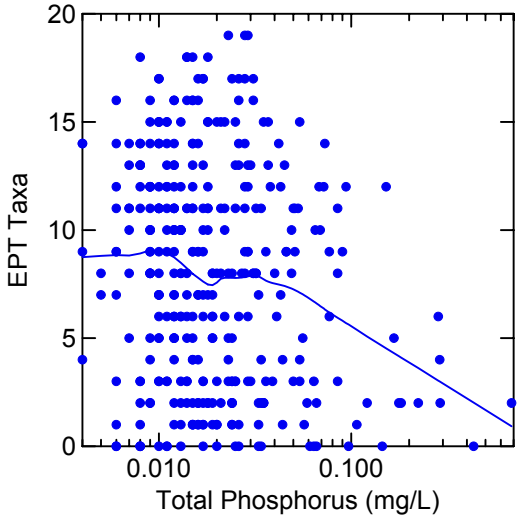
Pros	Cons
• Effective way to identify nutrient thresholds	• Requires substantial data, probabilistic data is best
• Nice approach for linking criteria together	• Lack of significance does not mean lack of association
• Quantitative measure of thresholds with error estimate	• Requires knowledge of an appropriate threshold for response variable
	• Lack of significance does not mean lack of association

Alternatives

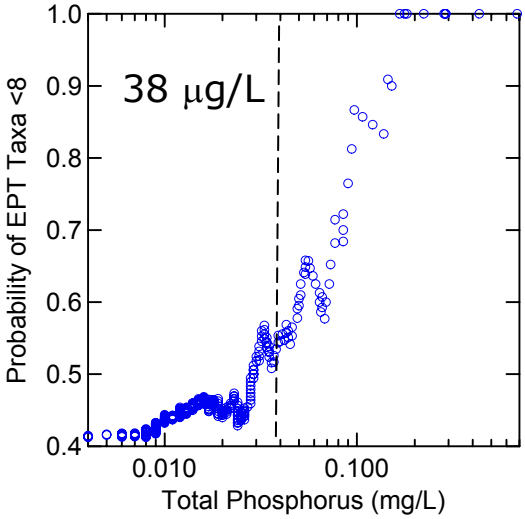
- Change-point analysis alone

Citations

Paul, J.F. and M.E. McDonald. 2005. Development of empirical, geographically specific water quality criteria: a conditional probability analysis approach. Journal of the American Water Resources Association 41:1211-1223



Scatterplot with Loess Line



Conditional Probability Plot