Sediment Quality Objectives

For California Enclosed Bays and Estuaries

Development of Chemistry Indicators

February 28, 2006

Presentation Overview

- Summary of previous results
- SSC recommendations
- Indicator development and evaluation results
- Results interpretation
- Data integration

Chemistry Indicator Development

- **Develop indicator(s) that reflect biological effects caused by contaminant exposure**
- **Identify important geographic, geochemical, or other factors that affect relationship between chemistry and biological effects**
- **Develop thresholds and guidance for use in MLOE framework**

Approach

- Investigate multiple approaches
  - Existing methods used by other programs
  - Existing methods calibrated to California
  - New approaches
- Evaluate SQG performance
  - Use CA data
  - Use quantitative and consistent approach
  - Select methods with best performance

July 2005 SSC Meeting

- Interim progress report
  - Presented preliminary results
- Candidate SQG selection
- Data set preparation and characteristics
  - Potential for geographic strata
- Preliminary evaluation
  - Refine SQG candidate list

Candidate SQGs

- **National SQGs**
  - **Published approaches with readily available values**
  - **Used in other assessment programs**
- **Calibrated SQGs**
  - **National approaches refined to reflect California conditions**
  - **Normalization to sediment characteristics**
- **New SQGs**
  - **Novel methods**

Empirical SQGs

SQG	Metric	Source
ERM Effects Range Median Analysis of diverse studies and effects values	Mean Quotient for Chemical Mixture	Long et al.
Consensus MEC Mid-range effect concentration Geometric mean of similar guidelines	Mean Quotient for Chemical Mixture	MacDonald et al, Swartz, SCCWRP
SQGQ-1 Mid-range effect concentration Subset of chemical guidelines from various sources	Mean Quotient for Chemical Mixture	Fairey et al.
Logistic Regression Regression model for each chemical	Probability of Toxicity (Pmax) for Chemical Mixture	Field et al.

Mechanistic SQGs

SQG	Metric	Source
EqP Organics Acute and chronic effects Organic Carbon Normalized	Sum of Toxic Units (TU)	EPA + CA Toxics Rule

Candidate SQGs

- National versions
  - Mean ERM quotient
  - Logistic regression (National Pmax)
  - Mean SQGQ1 quotient
  - Mean Consensus quotient
- Calibrated SQGs (regional)
  - Logistic regression (National Pmax)
  - Logistic regression (California Pmax)
  - ERM quotient (CA ERM)
- New SQGs
  - Kappa

Calibration and Validation Datasets

- CA data for bays and estuaries
  - Multiple studies and locations
  - Screened for quality and completeness
- Calibration/development dataset
  - Calibration of SQGs
  - Development of new SQGs
- Validation dataset
  - Approximately 30% of data, not used for calibration
  - Representative of contamination gradient and geographic regions

Correlation With Toxicity

North

SQG	Spearman Correlation
Kappa	0.54
N. CA ERM	0.37
NOAA ERM	0.37
N. CA Pmax	0.35
Consensus	0.29
SQGQ1	0.28
National Pmax	0.27
EqP organics	-0.08

South

SQG	Spearman Correlation
Kappa	0.46
S. CA Pmax	0.32
NOAA ERM	0.29
S. CA ERM	0.28
SQGQ1	0.25
Consensus	0.22
National Pmax	0.22
EqP organics	-0.08

Preliminary Results

- Differences in contaminant mixtures in North and South regions
  - Evidence that empirical relationships between chemistry and toxicity also vary by region
- Normalization to sediment characteristics not effective
  - TOC and iron normalization did not improve relationships with toxicity
- Potential for improved predictive ability using new or calibrated SQGs
  - Stronger correlation with amphipod mortality
  - Greater sensitivity and specificity

SSC Recommendations

- Drop EqP SQGs as a candidate chemical indicator
  - Valuable supplemental information, however
- Evaluate statewide and regional versions of SQGs
  - Base final recommendations on performance
- Investigate kappa statistic
  - Suitability for use in SQG development
  - Effects of sample size and data distribution
- Continue to develop Kappa SQG
  - Use expanded chemical list
- Evaluate SQGs for relationship to benthic community effects
  - Consider use of multiple SQGs with different approaches

Presentation Overview

- Summary of previous results
- SSC recommendations
- Indicator development and evaluation results
- Results interpretation
- Data integration

Recent Activities

- Investigated Kappa statistic and refined analysis methods
- Completed SQG development
  - Calibration of ERM and Pmax
  - New approaches
- Repeated SQG comparison with toxicity
  - Revised data set
  - Refined SQGs
  - New methodology
- Compared SQGs with Benthos

Kappa Statistic

- Developed in 1960-70's
  - Peer-reviewed literature describes derivation and interpretation
- Used in medicine, epidemiology, & psychology to evaluate observer agreement/reliability
  - Result reflects magnitude of agreement between different measures
  - Accommodates multiple categories of classification
  - Multiple thresholds can be adjusted by user
  - Similar application to SQG development and assessment
  - Sediment quality assessment is a new application

Kappa

Evaluates agreement between 2 methods of classification

- Chemical SQG result
- Biological effects result (toxicity or benthos)
- Partial credit based on magnitude of differences

Toxicity Result	SQG Result (potential for effect)			
	Unlikely	Marginal	Moderate	High
Nontoxic				
Low				
Moderate				
High				

T3 T2 T1

Kappa Analysis Output

- **Kappa (*k*)**
  - **Similar to correlation coefficient**
  - **Confidence intervals**
- **Thresholds**
  - **Optimized for correspondence to observed effects**
  - **Can be applied to chemistry data to predict effect category (*cat*)**
    - **E.g., Category 1, 2, 3, or 4**

Kappa Concerns

- Results are sensitive to data distribution
  - Misleading results possible with nonsymmetric data sets
  - Data should have even distribution of responses
  - Use the same data when making comparisons between approaches
- No definitive guidance regarding how to interpret results
  - No absolute criteria for “good” or “bad” agreement
  - Comparisons of relative agreement are recommended
- We revised our analysis method to minimize these factors

Analysis Strategy

- Calibration and validation data sets
  - Calibration: indicator development and calibration
  - Validation: independent data (30%) to confirm results
- Subsample data set
  - Equal proportion of samples in each toxicity category
  - Reduces bias in results due to preponderance of nontoxic samples
- Analyze multiple data subsamples
  - Use mean/median to get robust estimate of correlation or performance metrics
  - Estimate of variability

Calibration of SQGs

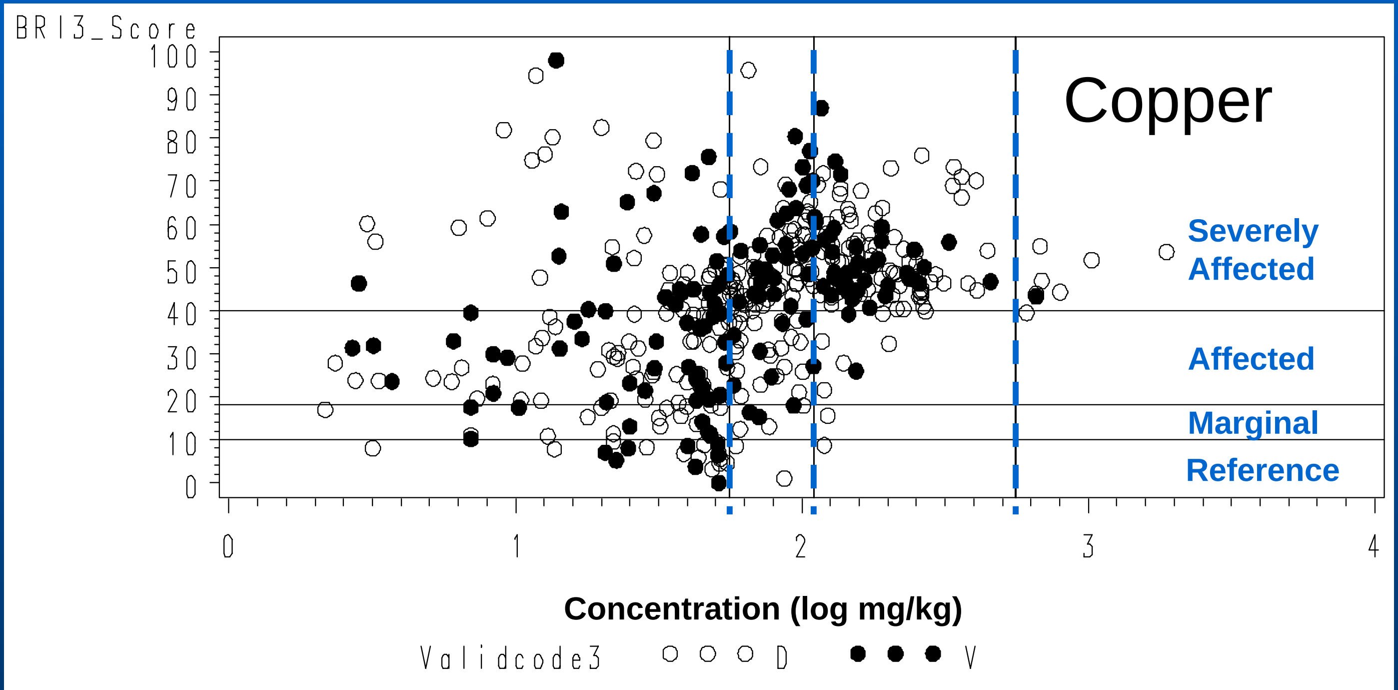
- Adjustment of national approaches based on California data
- ERM
  - Derived CA-specific values using modified method of Ingersoll et al.
  - Sample-based analysis
  - Statewide, North, and South versions
- Logistic Regression Model (Pmax)
  - Developed chemical modes using CA data
  - Compared fit to data against national models
  - Selected best combination of models
  - Statewide, North, and South versions

New SQG Characteristics

- Compatible with multiple line of evidence assessment framework
  - Categorical classification and multiple thresholds
- Capability to include new contaminants of concern
  - Not dependent on values from other approaches
- Adaptable to different application objectives
  - Thresholds can be adjusted
- Able to use toxicity and benthic community impact data in development
  - Accept nominal and ordinal data
- Result reflects uncertainty of empirical relationship
  - Some type of weighting based on strength of relationship

Kappa SQG Development

- **Derived Kappa and thresholds for individual target chemicals using biological effects data**



Kappa SQG Application

- **Calculate Kappa score for each chemical in sample**
  - ***Score* = *k* x *cat***
    - **k = kappa (reflects strength of empirical relationship)**
    - **cat = predicted effect category (1-4, based on thresholds)**
- **Calculate mean weighted Kappa score**
  - **Average of *k* x *cat* for all chemicals**
  - **Each constituent contributes to final classification in a manner proportional to reliability of relationship**
  - **Chemical mixture model**
- **Mean kappa score used in manner analogous to mean SQG quotient or Pmax**

Candidate SQGs

- National versions (statewide)
  - Mean ERM quotient
  - Logistic regression (National Pmax)
  - Mean SQGQ1 quotient
  - Mean Consensus quotient
- Calibrated SQGs (regional)
  - Logistic regression (California Pmax)
  - ERM quotient (CA ERM)
- New SQGs
  - Kappa_Tox
  - Kappa _Benthos

SQG Selection Process

Three principal factors considered

- **Performance using CA data**
- **Feasibility of application**
  - **Simple approach favored over complicated**
- **Capability for revision**
  - **New chemicals**
  - **New data**

SQG Performance Evaluation

- Conduct separate evaluations for toxicity and benthos effects
  - Potentially different relationships with chemistry
  - Less data available for benthos
- Examine SQG correlations with effect
  - Select short list of SQGs for further evaluation
- Compare predictive ability over a range of thresholds
  - Confirm and refine correlation results
- Compare statewide and north/south SQG versions

Correlation with Toxicity
Validation Data

	SQG	Statewide	North	South
Regional	Kappa		0.37	0.31
	CA Pmax		0.19	0.40
	CA ERMq		0.24	0.18
Statewide	Kappa	0.27	0.38	0.31
	CA Pmax	0.33	0.37	0.42
	Nat. Pmax	0.22	0.17	0.32
	CA ERMq	0.19	0.24	0.17
	NOAA ERMq	0.26	0.30	0.27
	Consensus	0.25	0.24	0.30
	SQGQ1	0.16	0.27	0.26

Short list of SQGs for further evaluation:

Kappa, CA Pmax, National Pmax, NOAA ERM

Predictive Ability

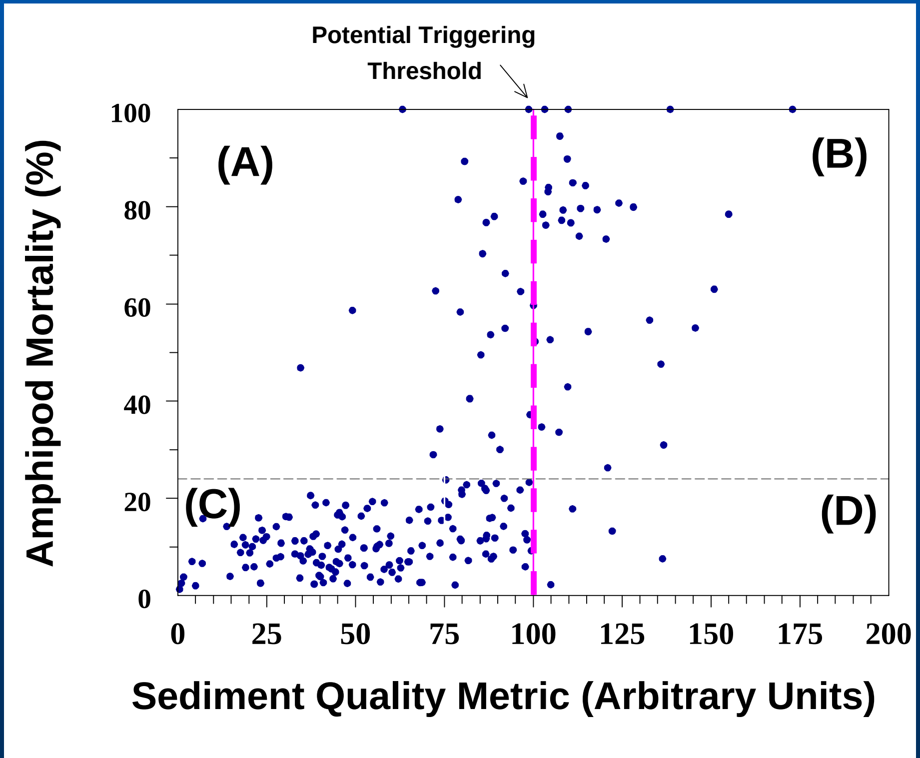
Multiple measures are needed to describe important characteristics

Negative Predictive Value =C/(C+A) x 100
(percent of no hits that are nontoxic)
=**Nontoxic Efficiency**

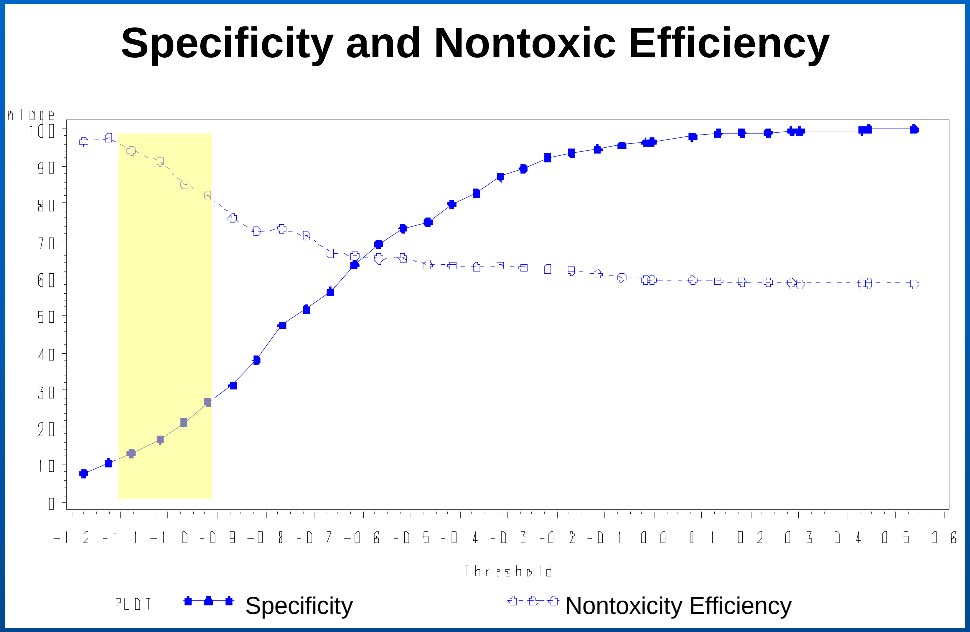
Specificity=C/(C+D) x 100
(percent of all nontoxic samples that are classified as a no hit)

Positive Predictive Value =B/(B+D) x 100
(percent of hits that are toxic)
=**Toxic Efficiency**

Sensitivity=B/(B+A) x 100
(percent of all toxic samples that are classified as a hit)



Predictive Ability Assessment



Calculated mean specificity or sensitivity for a range of efficiency

Higher values for same range indicate better ability to discriminate among affected/unaffected samples

- SQG performance is threshold dependent
- Inverse relationship between efficiency (toxic or nontoxic) and specificity or sensitivity
- Both types of measures are important for SQG application
- Improved SQG utility when greater sensitivity or specificity obtained without sacrificing efficiency

Predictive Ability: Statewide

	Sensitivity	Specificity
CA Pmax	31	26
Kappa_Tox	30	37
Nat. Pmax	16	23
NOAA ERMq	19	19

Mean sensitivity calculated for 60-80% toxic efficiency

Mean specificity calculated for 70-90% nontoxic efficiency

Statewide versions of Kappa and CA Pmax have best ability to distinguish among toxic and nontoxic samples

Predictive Ability: Regional

	North		South	
	Sensitivity	Specificity	Sensitivity	Specificity
CA Pmax_State	56	44	53	34
CA Pmax_Region	46	45	52	28
Kappa_Tox_State	70	53	36	27
Kappa_Tox_Region	77	76	35	29
Nat. Pmax	53	NA	41	32
NOAA ERMq	47	37	28	27

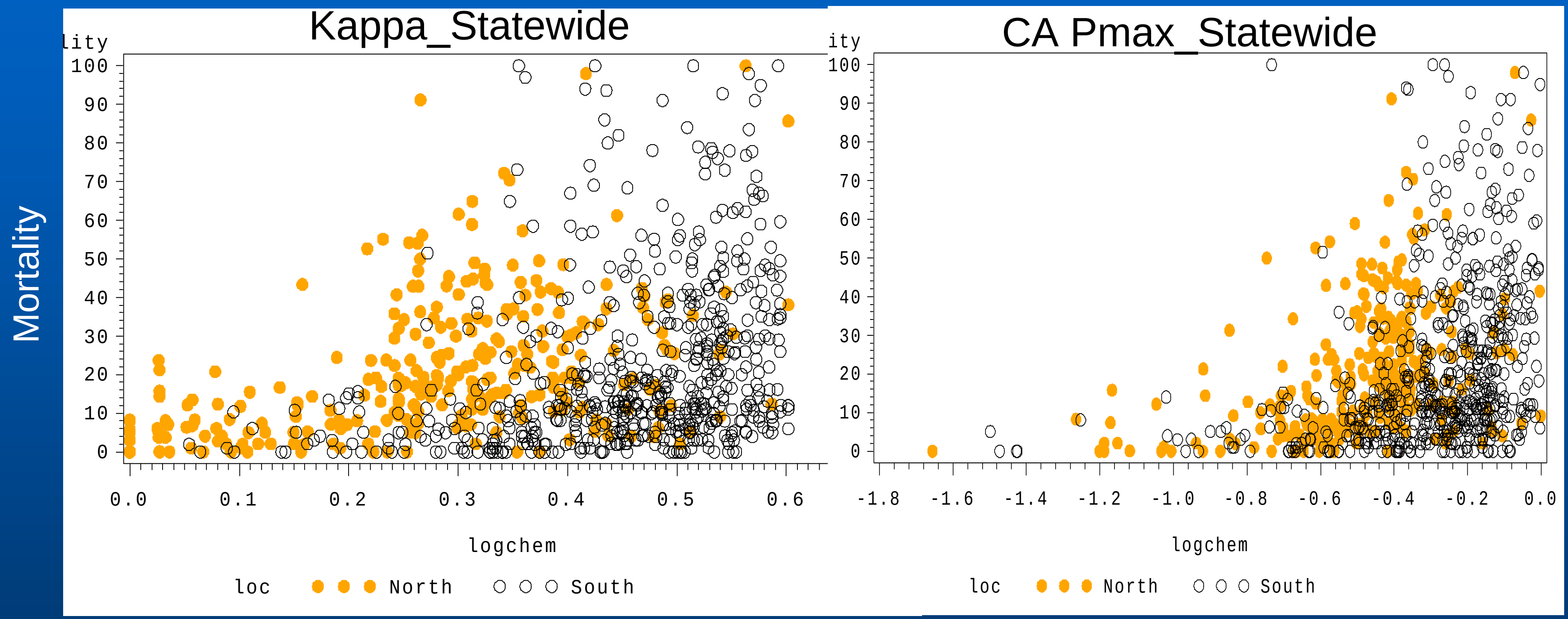
Mean sensitivity calculated for 60-80% toxic efficiency

Mean specificity calculated for 70-90% nontoxic efficiency

North: Kappa has best predictive ability

South: CA Pmax has best predictive ability

Region Comparison



CA Pmax has less regional sensitivity than Kappa

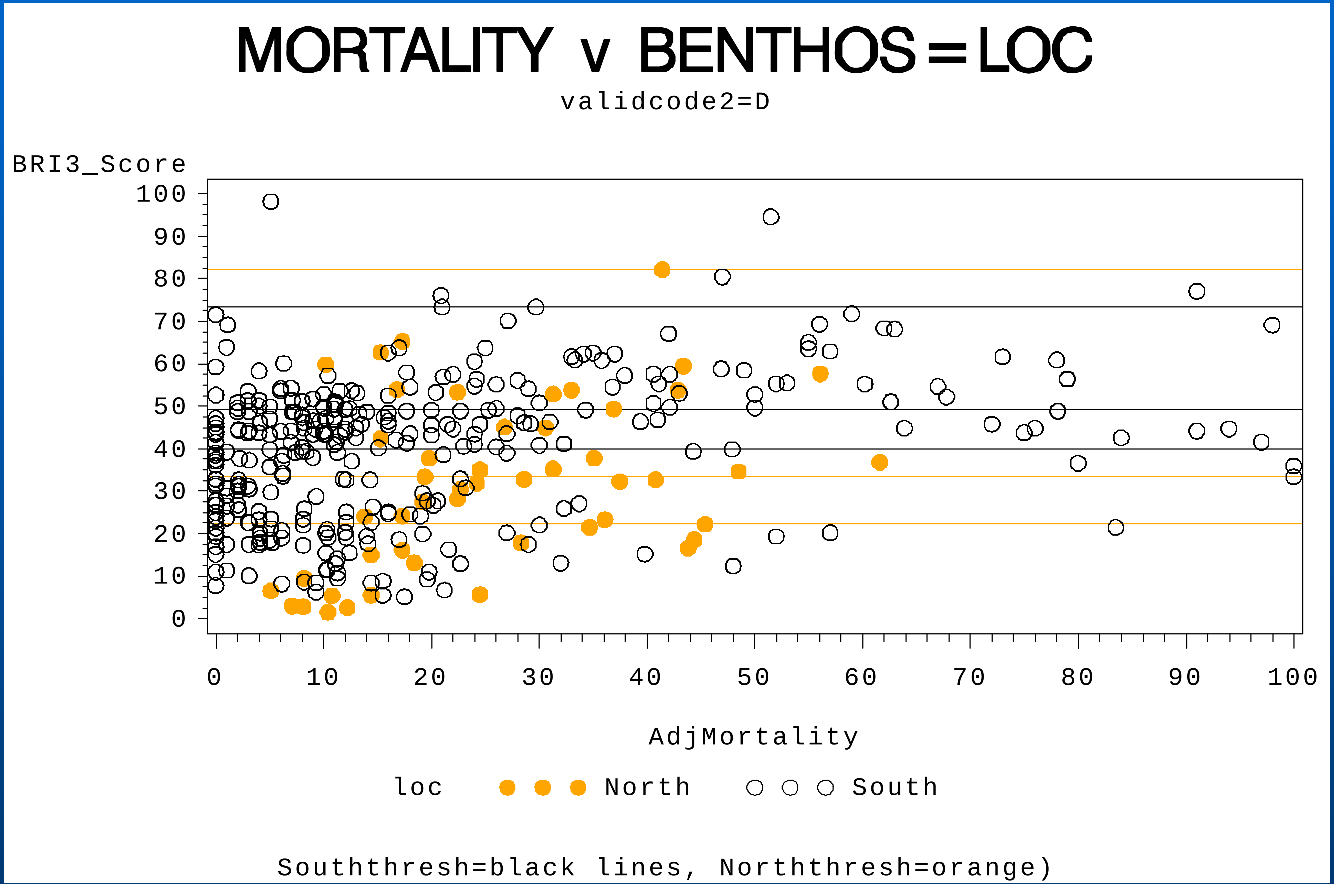
Chem:Toxicity Evaluation Summary

- Calibrated (CA Pmax) and New (Kappa) SQGs have the best performance overall
  - Highest correlations with mortality
  - Highest sensitivity and specificity
- Some regional effects are evident
  - Best performing SQG varies by region
  - Predictive ability greater for data within a region than statewide
  - Effect varies with SQG

Chem:Benthos SQG Evaluation

- Similar to approach for Chem:Tox except:
- Used BRI and preliminary thresholds
- Less data available
- Focus on regional evaluation
  - Two different assemblages represented in the results
  - Cannot combine BRI scores across state
- Limited development of SQGs for benthos
  - Applied national and calibrated SQGs from toxicity evaluations
  - Developed benthos version of kappa

Chem:Benthos SQG



Few samples with high benthic impact category

North samples are fewer and have less defined relationship with toxicity

Correlation with Benthic Index

	SQG	North	South
Regional	Kappa_Benthos	0.51	0.36
Statewide	Kappa_Tox	0.30	0.30
	CA Pmax	0.29	0.42
	Nat. Pmax	0.04	0.44
	NOAA ERMq	0.39	0.28
	Consensus	0.48	0.28
	SQGQ1	0.36	0.34

Short list of SQGs for further evaluation:

Kappa_Benthos, CA Pmax_State, National Pmax, NOAA ERMq, Consensus

Predictive Ability: Regional

	North		South	
	Sensitivity	**Specificity**	**Sensitivity**	**Specificity**
Kappa_Benthos	**65**	**51**	**62**	**64**
CA Pmax_State	**38**	**43**	**56**	**56**
Nat. Pmax	**47**	**28**	**58**	**66**
NOAA ERMq	**<56**	**44**	**44**	**44**
Consensus	**63**	**48**	**52**	**46**

Mean sensitivity calculated for 55-75% toxic efficiency

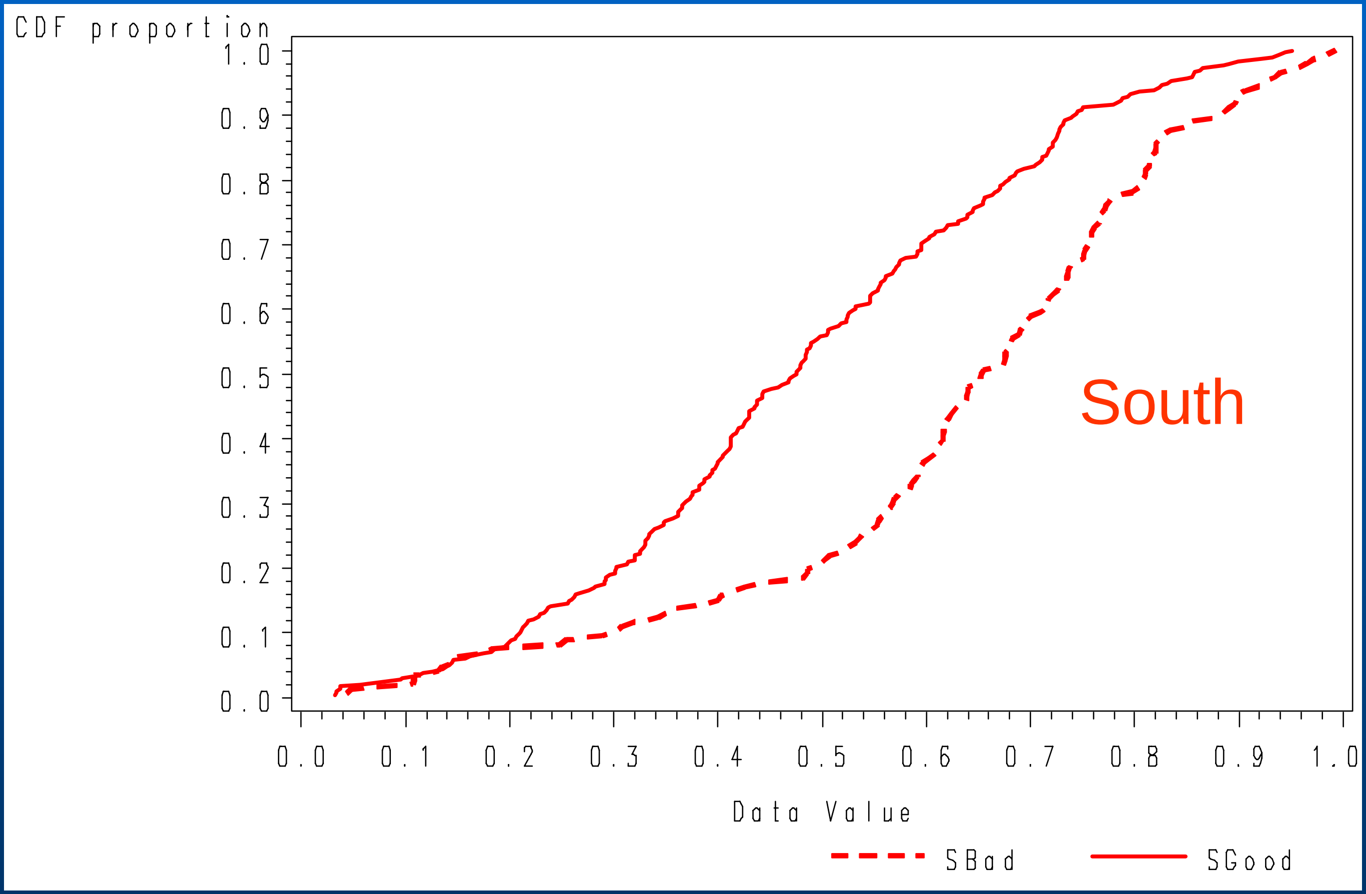
Mean specificity calculated for 55-70% nontoxic efficiency

North: Kappa_Benthos has best predictive ability

South: Kappa_Benthos and Nat. Pmax have best predictive ability

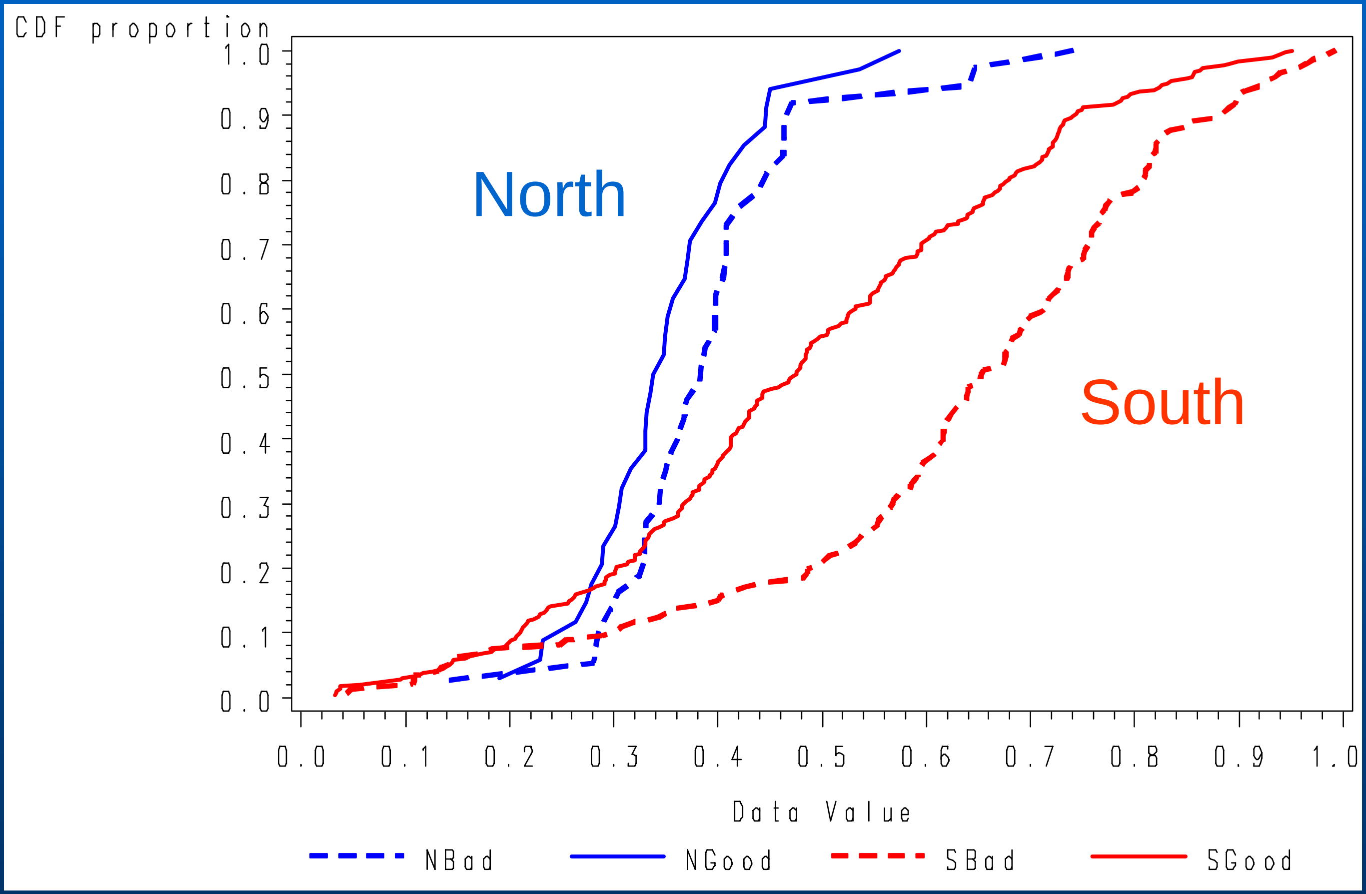
Regional Effects
Cumulative Distribution Frequency

Benthos vs. CA Pmax_State



Regional Effects
Cumulative Distribution Frequency

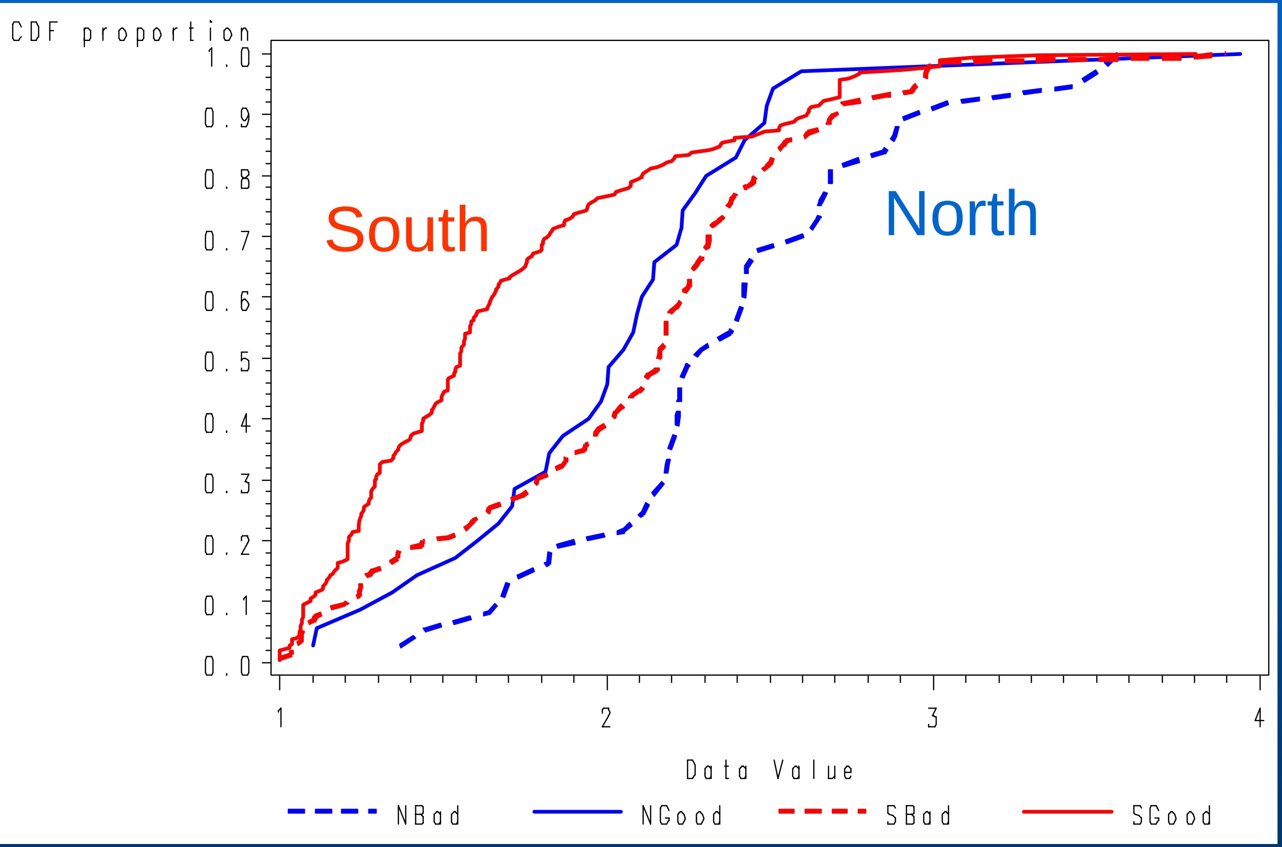
Benthos vs. CA Pmax_State



Strong regional difference in relationship between benthic effects and statewide SQG values

Regional Effects

Benthos vs. Kappa_Benthos



Regional difference in discrimination between affected samples is less

Chem:Benthos Evaluation Summary

- Kappa_Benthos SQG has better overall relationship to benthic community impacts than the other SQGs evaluated
  - Highest correlation and predictive ability in the north
  - Similar to CA Pmax or Nat. Pmax in the south
- Are pronounced regional effects when statewide SQGs are applied
  - Species assemblages are different
  - Regional SQG can accommodate regional factors

Chemistry Indicator Recommendations

- Use a combination of SQGs based on toxicity and benthos
  - Balance strengths and weaknesses of each
  - Incorporate different approaches to provide a more robust assessment
  - Address both statewide and regional factors
  - Represents two important measures of biological effect
- Use CA Pmax (Statewide SQG)
  - Relatively good overall performance for toxicity
  - Established approach
  - Less sensitivity to regional factors
- Use Kappa_Benthos (regional SQG)
  - Good correspondence to benthos
  - Reflects regional differences in assemblage responses
  - Only SQG to directly incorporate magnitude of biological response, strength of association with individual chemicals, and ordinal classification of effects

Presentation Overview

- Summary of previous results
- SSC recommendations
- Indicator development and evaluation results
- Results interpretation
- Data integration

Results Interpretation

- Multiple categories of chemical condition needed for each SQG
  - Describe potential for a chemically-mediated biological effect
  - Reflect variations in strength of association with effects
  - Provide ordinal information for use in MLOE assessment
- Four categories desired
  - Provides ability to rank stations and prioritize additional investigations or actions

Chemistry Categories

- **Minimal potential:** Sediment-associated contamination may be present, but unlikely to result in effects. Conditions essentially equivalent to uncontaminated reference areas.
- **Low potential:** Slight or marginal increase in contamination that may be associated with increased potential for effects, but magnitude or frequency of occurrence of significant biological impacts is low.
- **Moderate potential:** Clear evidence of sediment contamination that is likely to result in biological effects; an intermediate response category.
- **High effect:** Contamination that is highly likely to result in consistent and possibly severe biological effects; generally present in a small percentage of the samples.

Thresholds

Three thresholds are needed to classify the test results into one of four chemistry categories

Minimal

Low Threshold

Low

Moderate Threshold

Moderate

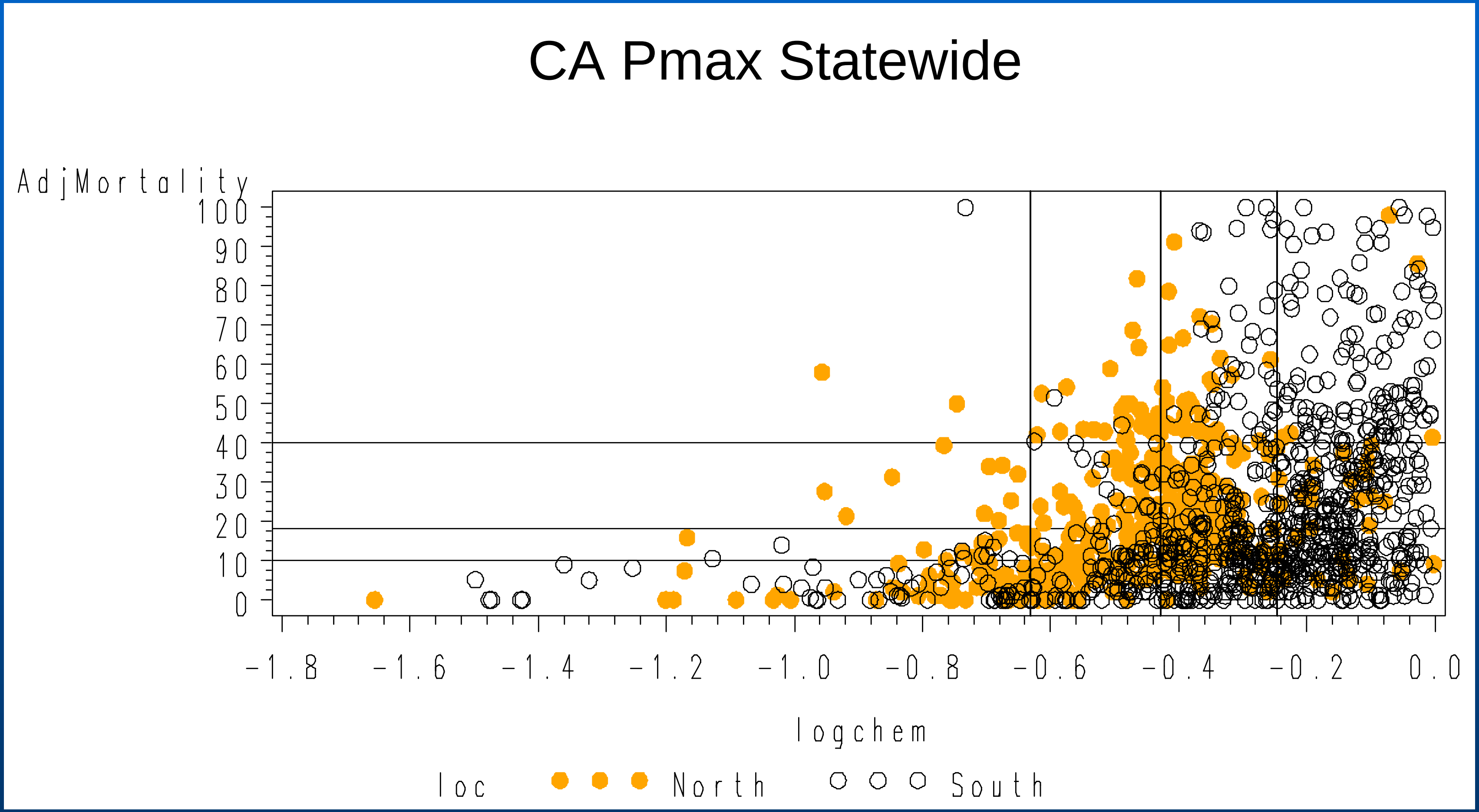
High Threshold

High

Threshold Development Strategy

- Statewide SQG
  - Use same thresholds for entire state
  - Provides consistency in application
- Regional SQG
  - Use thresholds calibrated to each region
  - Provides optimum performance
- Threshold values
  - Based on kappa optimization to biological effects categories
  - Objective and relevant to CA conditions

Thresholds



Thresholds (Pmax; no log):

Low = 0.23 **Moderate = 0.37** **High = 0.57**

Presentation Overview

- Summary of previous results
- SSC recommendations
- Indicator development and evaluation results
- Results interpretation
- Data integration

Data Integration

Objective:

- **Combine multiple SQG results to produce a chemistry LOE classification**

Integration Strategy

Similar to toxicity indicator strategy:

- Weight each SQQ equally
- If both SQGs agree, then LOE category is the same
- If categories differ, then assign category corresponding to the median

Data Integration

SQG 1	SQG 2	LOE Category
Minimal	Minimal	Minimal
Minimal	Low	Low
Minimal	Moderate	Low
Minimal	High	Moderate
Low	Low	Low
Low	Moderate	Moderate
Low	High	Moderate
Moderate	Moderate	Moderate
Moderate	High	High
High	High	High

Research Needs

- Develop and include chemical measurements that effectively describe contaminant bioavailability or dose
  - Empirical relationships based on routine chemistry data are variable and limit utility of SQGs
  - Black carbon, sulfides, contaminant desorption
- Develop chemistry indicators that reflect contaminants of current and emerging concern
  - Insufficient data to address newer pesticides
  - Current SQGs may fail to identify some areas of chemical impact
- Refine benthos-based SQGs as more data become available
  - Limited data for some habitats