From Raw Data to Report Cards

A brief review of methods for transforming, aggregating and binning monitoring results

Why turn data into indices?

- Turn various measurements into thematic scores
- Evaluate overall trends
- Simplify communication with decision makers and the public
- Prioritize management efforts

Key Upfront Decisions

- What do I need the data to tell me?
- Is the management decision binary, such as list / don't list?
- Is a gradient scale needed instead, such as excellent, good, fair, poor, very poor?
- Is different information needed for different beneficial uses, or for high quality waters?

Key Points for the HSP

What should be consistent statewide,
and what needs local and regional flexibility?

- Assessment approach (index components, scoring)
- Assessment standards or thresholds
- Categorical (grades) versus quantitative (scores)
- Math, algorithms, grading break points

Two Main Approaches

Quantitative (math based)

Data ➡ number scores ➡ index scores ➡ letter grade

Categorical (rule based)

Data ➡ letter grade ➡ index letter grade

Four Steps for Quantitative Scores

1. Group indicators into thematic indices
2. Transform measured data into unit-less scores
3. Aggregate scores from multiple indicators into a summary index score
4. Breakpoints to bin index scores into grades

Examples of Indices & Beneficial Uses

- Human Health WQ Index (safe to drink, swim)
- Aquatic Life WQ Index (aquatic life)
- Toxicity Index
- Biostimulatory Risk Index
- Habitat Index
- Riparian Index

Example Indicators within Indices

Human Health Water Quality Index (surface water)

- Fecal Coliform
- E. coli
- Pathogens
- Nitrate

- Toxic chemicals

Aquatic Life Water Quality Index (surface water)

- Ammonia
- Nitrate
- Ortho-phosphate
- DO departure
- Turbidity
- Total dissolve solids
- Total susp. sediment
- pH departure

- Toxic chemicals

Many ways to convert measured data into numerical scores

- Comparison to thresholds (standards)
- Using data distributions
- Empirical equations
- Tolerance Values: observed over expected
- Cumulative distribution frequencies
- Best professional judgment (& Delphi methods)
- Bioassessment indices (ask Pete)

Comparison to thresholds

Quotient Method

Divide measurement by a standard

Pro: + scientific consensus behind standards

\+ capture magnitude of exceedence

Con: - no fixed upper end to scale

\- appropriate standards must be available

Canadian CCME WQ Index (3 parts)

Factor 1: Scope
(# indicators)

$$F_1 = \left(\frac{\text{Number of failed variables}}{\text{Total number of variables}} \right) \times 100$$

Factor 2: Frequency
(# samples)

$$F_2 = \left(\frac{\text{Number of failed tests}}{\text{Total number of tests}} \right) \times 100$$

Factor 3: Amplitude
(standard quotient)

$$excursion_i = \left(\frac{FailedTestValue_i}{Objective_j} \right) - 1$$

CCME WQ Index

$$= 100 - \left(\frac{\sqrt{F_1^2 + F_2^2 + F_3^2}}{1.732} \right)$$

Quadratic mean (or root mean square) gives greater weight to larger values, thus emphasizing excursions.

Example of Canadian CCME WQI

North Saskatchewan River at Devon - 1997

DATE	DO Mg/L	pH	TP mg/L	TN mg/L	FC #/dL	As mg/L	Pb Mg/L	Hg g/L	2,4-D g/L	Lindane g/L
7-Jan-97	11.4	8.0	0.006	0.160	4	0.0002	0.0004	L0.05	L0.005	L0.005
4-Feb-97	11.0	7.9	0.005	0.170	L4[2]	L0.0002	**0.0094**	L0.05		
4-Mar-97	11.5	7.9	0.006	0.132	4	L0.0002	L0.0003	L0.05		
8-Apr-97	12.5	7.9	**0.058**[1]	0.428	L4	L0.0002	0.0008	L0.05	0.004	L0.005
6-May-97	10.4	8.1	0.042	0.250	L4	0.0002	0.0008	L0.05		
3-Jun-97	8.9	8.2	**0.108**	0.707	26	0.0006	0.0013	L0.05		
8-Jul-97	8.5	8.3	0.017	0.153	9	0.0002	0.0004			
5-Aug-97	7.5	8.2	0.008	0.153	8	L0.0002	L0.0003	L0.05	L0.005	L0.005
2-Sep-97	9.2	8.2	0.006	0.130	12	0.0003	0.0018	L0.05		
7-Oct-97	11.0	8.1	0.008	0.093	12	L0.0002	0.0011	L0.05	L0.005	L0.005
4-Nov-97	12.1	8.0	0.006	0.296	8	L0.0002	**0.0051**	L0.05		
1-Dec-97	13.3	8.0	0.004	0.054	4	L0.0002	L0.0003	L0.05		
OBJECTIVE:	*5*	*6.5 - 9.0*	*0.05*	*1*	*400*	*0.05*	*0.004*	*0.1*	*4*	*0.01*

[1] Bolded values do not meet the objective
[2] L = less than

Requires at least 4 variables, sampled at least 4 times

$$F_1 = \left(\frac{2}{10}\right) \times 100 \ = \mathbf{20}$$

$$F_2 = \left(\frac{4}{103}\right) \times 100 \ = \mathbf{3.9}$$

$$excursion = \left(\frac{0.058}{0.05}\right) - 1 \ = 0.16, \text{ etc.}$$

$$nse = \frac{(0.16 + 1.16 + 1.35 + 0.275)}{103} = 0.029$$

$$F_3 = \left(\frac{0.029}{0.01(0.029) + 0.01}\right) = \mathbf{2.8}$$

$$CCMEWQI = 100 - \left(\frac{\sqrt{20^2 + 3.9^2 + 2.8^2}}{1.732}\right) = \mathbf{88}$$

Excellent: 95 – 100, Good: 80 – 94, Fair: 65 – 79,
Marginal: 45 – 64, Poor: 0 - 44

Canadian CCME WQ Index

Pro:
+ widely excepted (esp. for drinking water)
+ considers scope, frequency, amplitude

Con:
- resolution depends heavily on standards selected
- final scoring “somewhat subjective,” based on BPJ
- less useful for “healthy;” only looks at failed tests

Rule-based Grades

(Example: NO_3 in mg/L)

Dark Red	If the mean for all samples at a site > 20
Red	If the mean > 10 and the mean <= 20
Orange	If the mean > 5 and the mean <= 10
Yellow	If the mean <= 5
Green	If the mean <= 1 and the max < 5
Blue	If the mean <= 0.3 and the max < 1

Rule-based Grades
(using percentiles of data from the site)

Dark Red	If the median for all samples at a site > 10
Red	If the 75th % > 10 and the median <= 1
Orange	If the 75th % <= 10 and the 90th % > 10
Yellow	If the 90th % > 1 and the 90th % <= 10
Green	If the 90th % <= 1

Rule-based Grades

Pro:
- \+ beneficial use-relevant break points
- \+ can be tailored to local conditions
- \+ can use percentiles to deal with outliers
- \+ incorporates logic and BPJ

Con:
- \- incorporates BPJ
- \- cannot be aggregated numerically
- \- may require different rules for each indicator

Minimum Maximum Approaches

$$\frac{(X - Min)}{(Max - Min)} \qquad \frac{(56 - 2)}{(93 - 2)} = \mathbf{59}$$

Pro: + Simple

+ Local context

Con: - No fixed range: max and min can change

- Strongly affected by outliers

- Not necessarily reflective of problems (e.g., if all values are in good condition)

Transformations based on Distributions of Data from All Sites

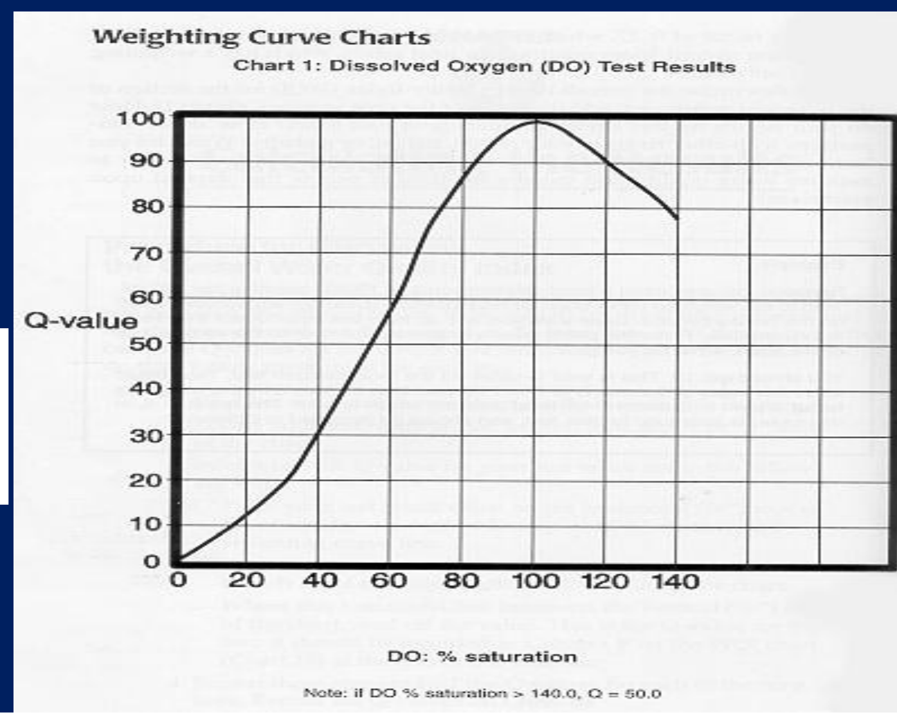
- Distribution percentiles
  (e.g., 10^{th} or 90^{th} percentile of all sites)
- Cumulative distribution frequencies
  (esp., with probabilistic designs)
- Arbitrary distances above or below means

Empirical relationships

- National Sanitation Foundation, Oregon, Washington
- Delphi expert consensus process
- Agreed-upon hand-drawn curves
- Regression and Quadratic equations

$$3.3\ mg/L < DO_C < 10.5\ mg/L$$

$$SI_{DO} = -80.29 + 31.88*DO_C - 1.401*DO_C^2$$



Tolerance Indicator Values (TIVs)

- Based on national data sets of fish and invertebrate taxon-specific sensitivities to individual stressors
- Observed TIV_O: mean for each taxon at the site
- Expected TIV_E: mean for each taxon expected at the site
- TIV_O / TIV_E indicates potential impact of each indicator

Methods to Aggregate Indicator Scores into an Index Score

1. Number of standard exceedences per site
2. Means:
   arithmetic, geometric, harmonic, quadratic
3. Weights and penalty factors
4. Regression of simple to comprehensive indices
5. Empirical formulae

Means

Arithmetic

- Simple
- Easily communicated

Means

Geometric

$$G = \sqrt[n]{x_1 x_2 \cdots x_n},$$

- Suitable for a mix of different types of metrics
- Normalizes the ranges being averaged
- A given percent change in any metric has the same effect as the same change in another
  (e.g., 5 to 8 has same effect as 50 to 80)

Means

Harmonic

$$= \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \cdots + \frac{1}{x_n}}.$$

- Preferable for averaging multiples, ratios or quotients
- High values are given greater weights than low values

Means

Quadratic

$$=100-\left(\frac{\sqrt{F_1^{\,2}+F_2^{\,2}+F_3^{\,2}}}{1.732}\right)$$

- High values are given greater weights than low values

Regression of simple indices against comprehensive indices

- Calculate index with many metrics
- Calculate index with a few easy metrics
- Regress one against the other
- Use simple index and regression equation for final index value

Empirical formulae

$$WQI = \log\left[\frac{(DO)^{1.5}}{(3.8)^{TP}(Turb)^{0.15}(15)^{FCol/10000} + 0.14(SC)^{0.5}}\right]$$

- Original data are not transformed
- Calibrated against existing indices

Set breakpoints to bin index scores

- Magnitude of standards exceedence
- Comparison to biological data (IBI, toxicity)
- Bayesian breakpoint analysis
- Breakpoint regression
- Percentiles of index distribution
- Percentiles of reference site distribution
- Cumulative distribution functions
- Delphi best professional judgment

No consensus approach in the literature

Region 3 Draft Scorecard Rules for Ambient Toxicity Tests (considering 303d)

Is any sample from the site significantly worse than control? (using lab's statistical determination)

No → Site is coded green, no/low toxicity — Good — **A**

Yes ↓

Is the number of toxic samples great enough to place the site on the 303[d] list (Table 3.1)
<u>OR</u>
Are the control –normalized result s from 10% or more of the samples from the site more toxic than the high toxicity threshold*?

No → Site is coded yellow, some significant toxicity. — Fair — **B**

Yes ↓

For all samples from the site for any one species, is the mean control-normalized survival lower than the high toxicity threshold*? (This includes growth endpoint for algae.)

No → Site is coded orange, moderate toxicity. — Evidence of Impact — **C**

Yes ↓

For all samples from the site for a second species, is the mean control-normalized survival lower than the high toxicity threshold*
<u>OR</u>
Is the mean control-normalized survival for any species lower than 25%? (This includes growth endpoint for algae.)

No → Site is coded red, high toxicity. — Impacted — **D**

Yes → Site is coded purple, very high toxicity. — Highly Impacted — **F**

* Using the control-normalized high toxicity thresholds