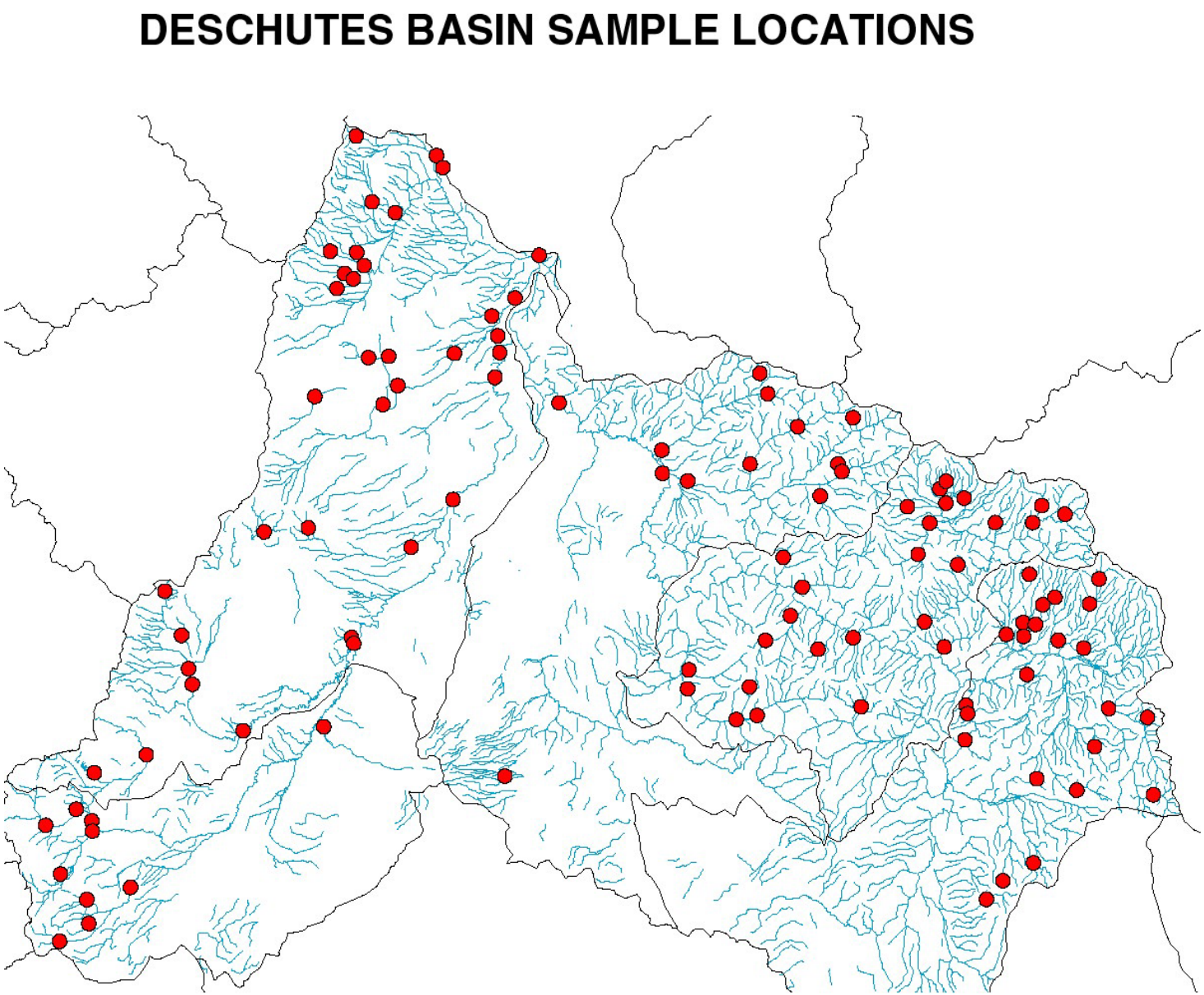
Decomposing variability
Phil Larsen - EMAP

- Phil Larsen
- Scott Urquhart
- Tom Kincaid
- Phil Kaufmann
- Others…

Context: multiple site surveys



Sites include

- Stream reaches or points on a stream
- HUCs
- Lakes
- Wetlands
- Time scale: years

Topics

- Framework for organizing variability
- Precision and the Signal:Noise ratio
- Trend detection

Organizing variability: Hierarchical framework

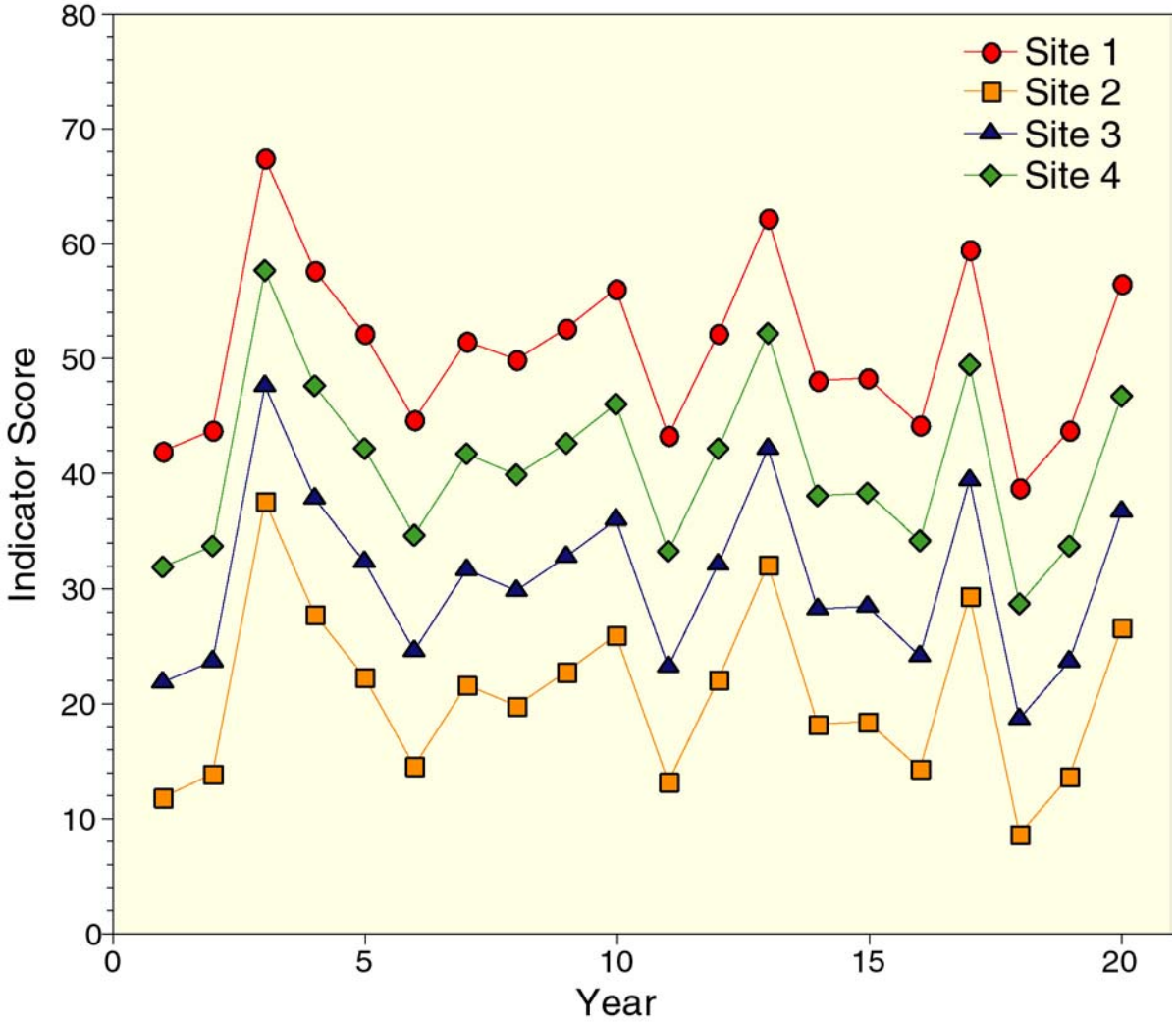
- Spatial (σ^2_s)
  - Site to site variation
  - HUC/site within HUC
- Temporal (years)
  - Concordant (synchronous, "Moran effect") **(σ^2_y)**
  - Interaction (Independent) **(σ^2_i)**
- Residual (σ^2_r)
  - Within season temporal
  - Local site scale
  - Team to team
  - Measurement

Site-to-site variation (σ^2_s)

- Persistent differences among sites due to:
  - Landscape/historical context:
    - Size
    - Gradient
    - Substrate composition
    - Elevation
    - …
  - Human disturbance effects

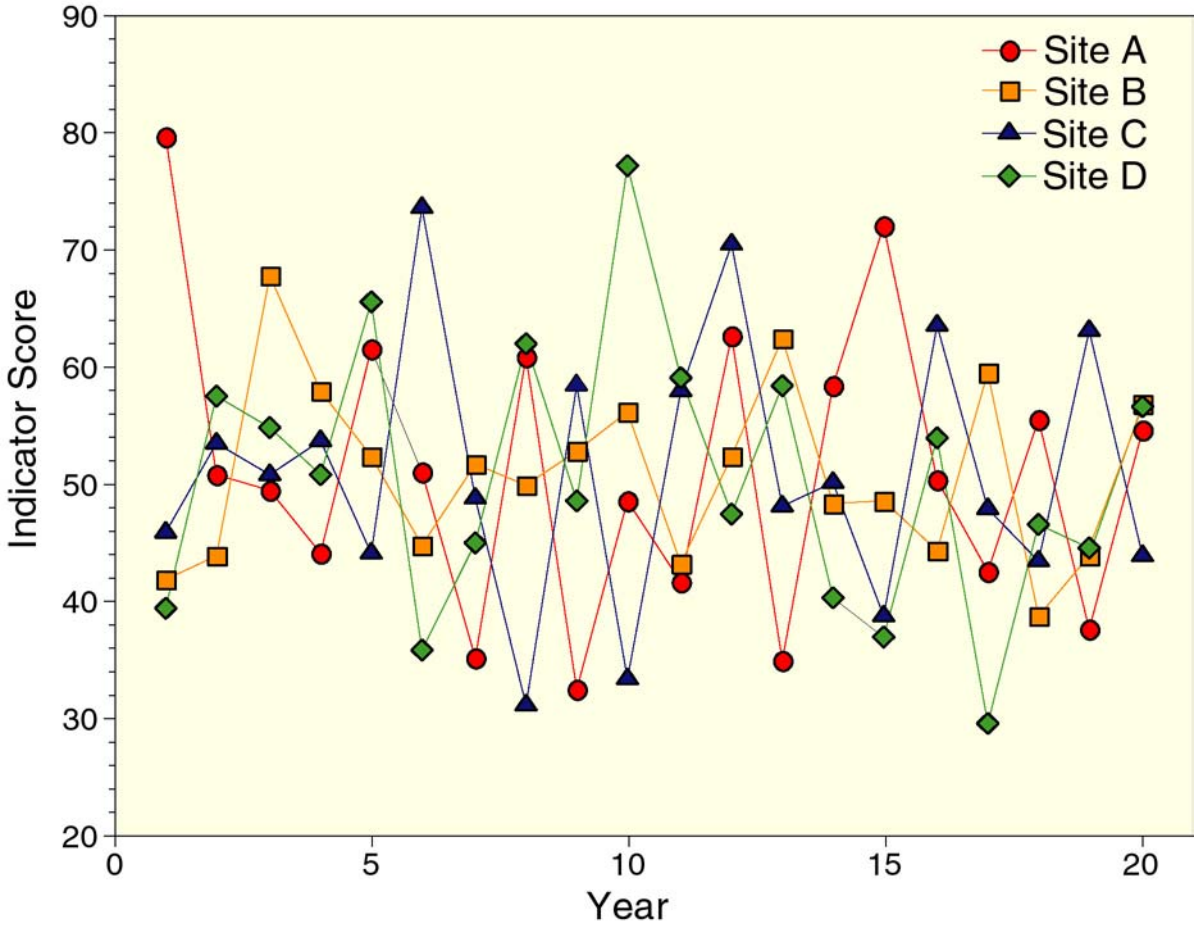
Year variation (σ^2_y)

- Concordant year-to-year variation across all sites
- Caused by regional phenomena such as:
  - Wet/Dry years
  - Ocean conditions


Site 1
Site 2
Site 3
Site 4
Indicator Score
Year
0
10
20
30
40
50
60
70
80
0
5
10
15
20

Interaction variation (σ^2_i)

- Independent year-to-year variation among sites
- Driven by local factors


Site A
Site B
Site C
Site D
Indicator Score
Year
90
80
70
60
50
40
30
20
0
5
10
15
20

Residual variation (σ^2_r)

- The rest of it including:
  - Temporal or seasonal variation during sampling window
  - Fine scale spatial variation
  - Crew-to-crew differences in applying the protocol
  - Measurement error
  - …

Measurement Precision: 3 versions

- Average standard deviation of repeat measurements during index window: σ_r
  - AKA RMSE (root mean square error)
- Coefficient of variation: 100 σ_r /Mean
  - Average of CV across multiple sites
  - Average σ_r/Regional Mean
- Signal:noise (for classification)
  - σ^2_s / σ^2_r

σ_r

- Advantages:
  - Units of precision are units of measurement
  - Explicit/Unambiguous
- Disadvantages:
  - Sometimes difficult to interpret for “derived” indicators like indices
  - Difficult to compare across indicators measured in different units or that have different potential ranges

Coefficient of Variation

- Advantages:
  - Scales precision so comparison among indicators is possible
- Disadvantages:
  - Dependence on mean can be misleading
  - For example, consider mean canopy cover of 0.1 and 0.9
    - $\sigma_r = 0.1$; CV=100% or 11%

Signal:Noise

- Advantages:
  - Interprets precision in context of the signal of interest
  - Facilitates comparison among different indicators or metrics
  - The higher S:N, the better the indicator or metric is able to discern differences among sites
- Disadvantages:
  - Contextual: same “noise” might give strong S:N or weak S:N depending on magnitude of S
  - Need for adequate sample sizes for both S and N

For more detail:

- Kaufmann, P.R., P. Levine, E. G. Robison, C. Seeliger, and D.V. Peck. 1999. Quantifying Physical Habitat in Wadeable Streams. EPA/620/R-99/003. U.S. Environmental Protection Agency, Washington, D.C.

Estimating Variances

- Sample size: 30 – 50 “replicates”
- Indirect methods:
  - ANOVA
  - Requires solving series of simultaneous equations of the form:
    - Mean Square (Source) = $c_1\sigma^2_s + c_2\sigma^2_y + c_3\sigma^2_i + c_4\sigma^2_r$
- Direct methods:
  - S-Plus: VARCOMP function
  - SAS: PROC MIXED

Reducing σ^2_r

- Reducing measurement component of σ^2_r
  - Additional training
  - Improved indicators
- Increasing "sample size"
  - Revisits within index window

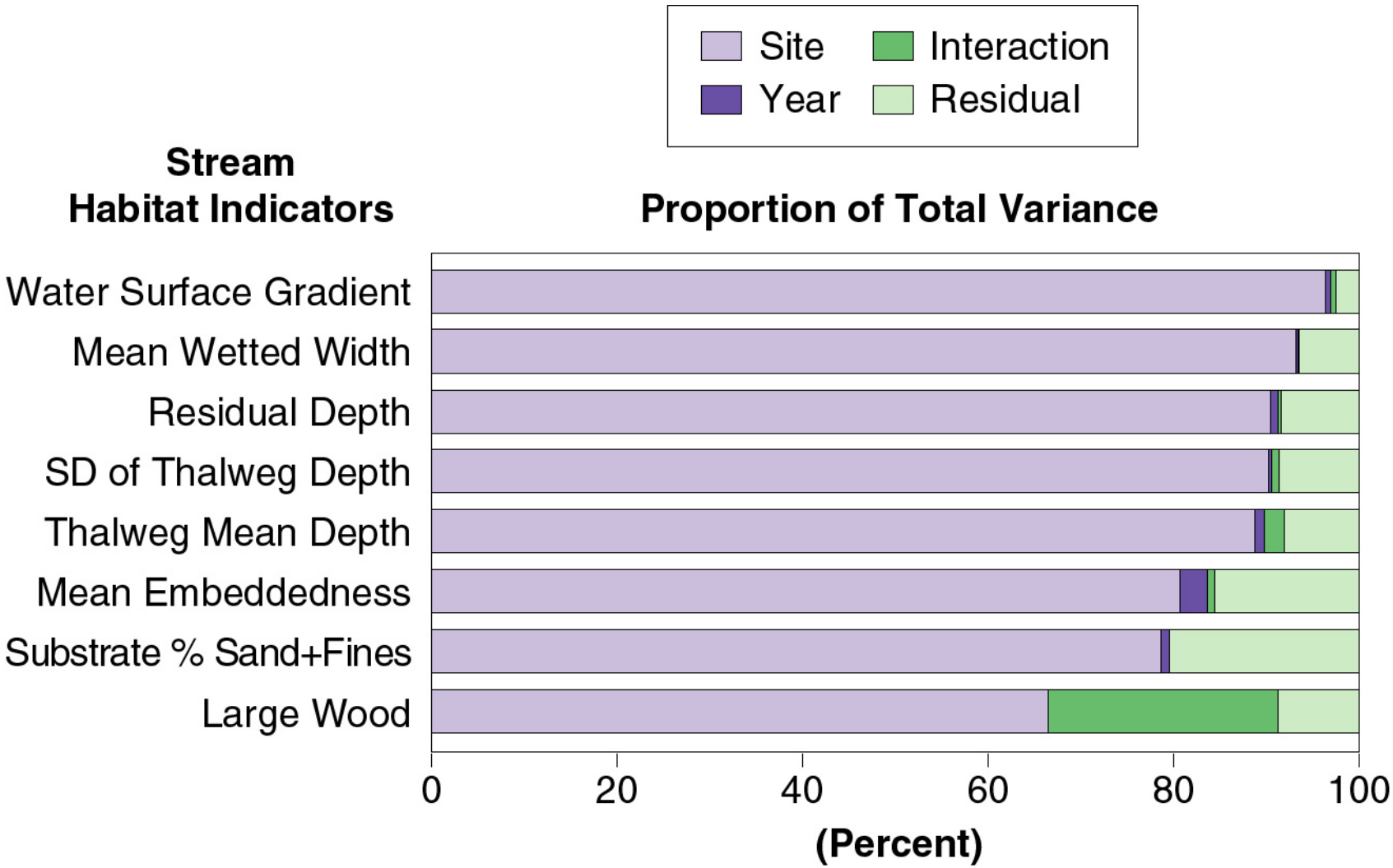
Comparing Precision

Attribute	RMSE		CV		S:N	
	MAHA	Oregon	MAHA	Oregon	MAHA	Oregon
Thalweg depth (cm)	6.4	6.2	22	17	7.3	6.9
Mean Residual Depth (cm)	1.6	2.2	17	19	16	9.0
% Pools	11	16	88	48	1.2	2.1
% Sand and fines	7.7	11	24	36	10	7.1
Log_{10}(Large wood) (m^3/100m)	0.53	0.34	n.a	n.a	2.5	12
Densiometer Canopy cover (%)	5.7	5.8	7.5	8.1	19	15

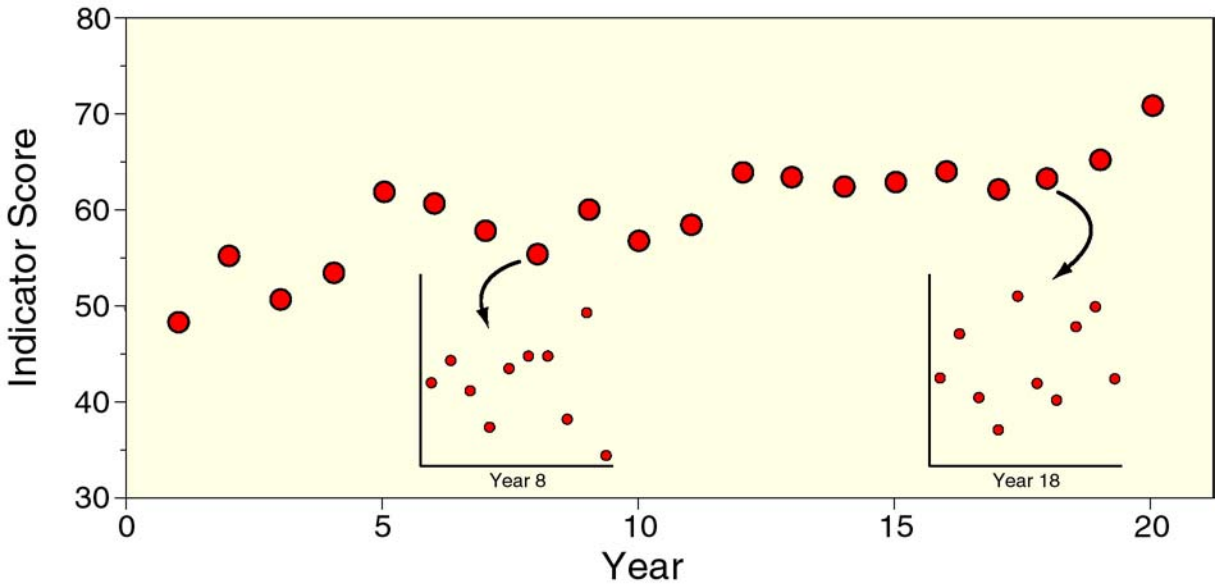
Comparing Precision

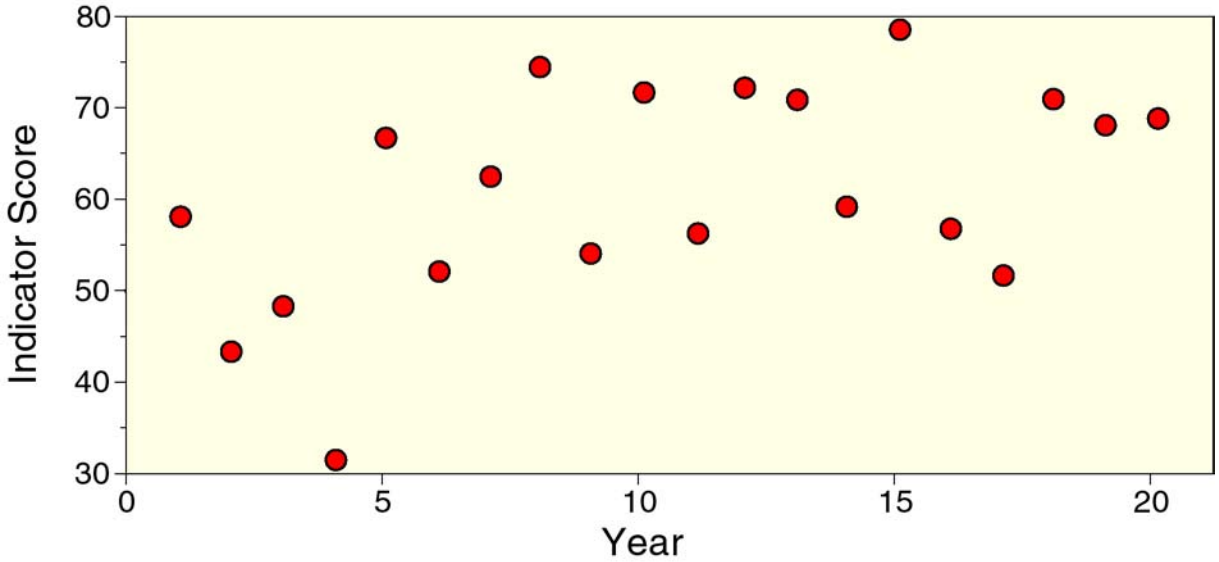
Attribute	RMSE		
	EMAP MAHA	EMAP Oregon	ODFW Oregon
Residual Depth (cm)	1.6	2.2	13.9
% Pools	11	16	11.8
% Sand and fines	7.7	11	10
Log_{10}(large wood) (m^3/100m)	0.53	0.34	0.66
Canopy Cover (%)	5.7	5.8	8.9

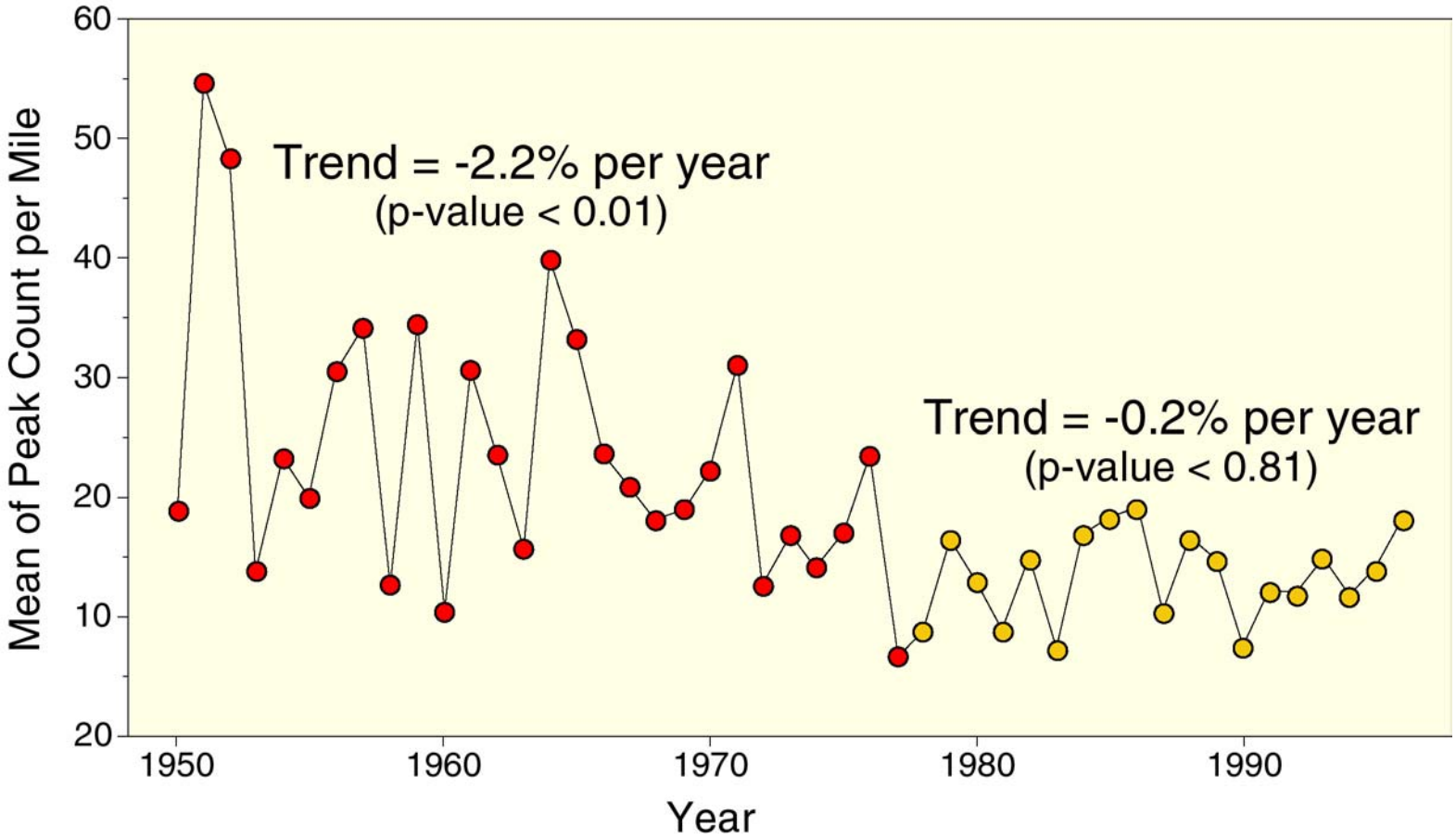
Variance Summary



Trend detection


Indicator Score
80
70
60
50
40
30
0
5
10
15
20
Year
Year 8
Year 18


Indicator Score
80
70
60
50
40
30
0
5
10
15
20
Year


Mean of Peak Count per Mile
60
50
40
30
20
10
20
Trend = -2.2% per year
(p-value < 0.01)
Trend = -0.2% per year
(p-value < 0.81)
1950
1960
1970
1980
1990
Year

Linear trend detection

- Hypothesis test: Slope = 0?
- Power: If a trend is present, what is the likelihood of detecting it?

Variance of a slope: How precisely can we estimate the slope?

$$\text{var}(slope) = \frac{\sigma^2}{\sum (X_i - \overline{X})^2}$$

Variance of a slope (Expanded version)

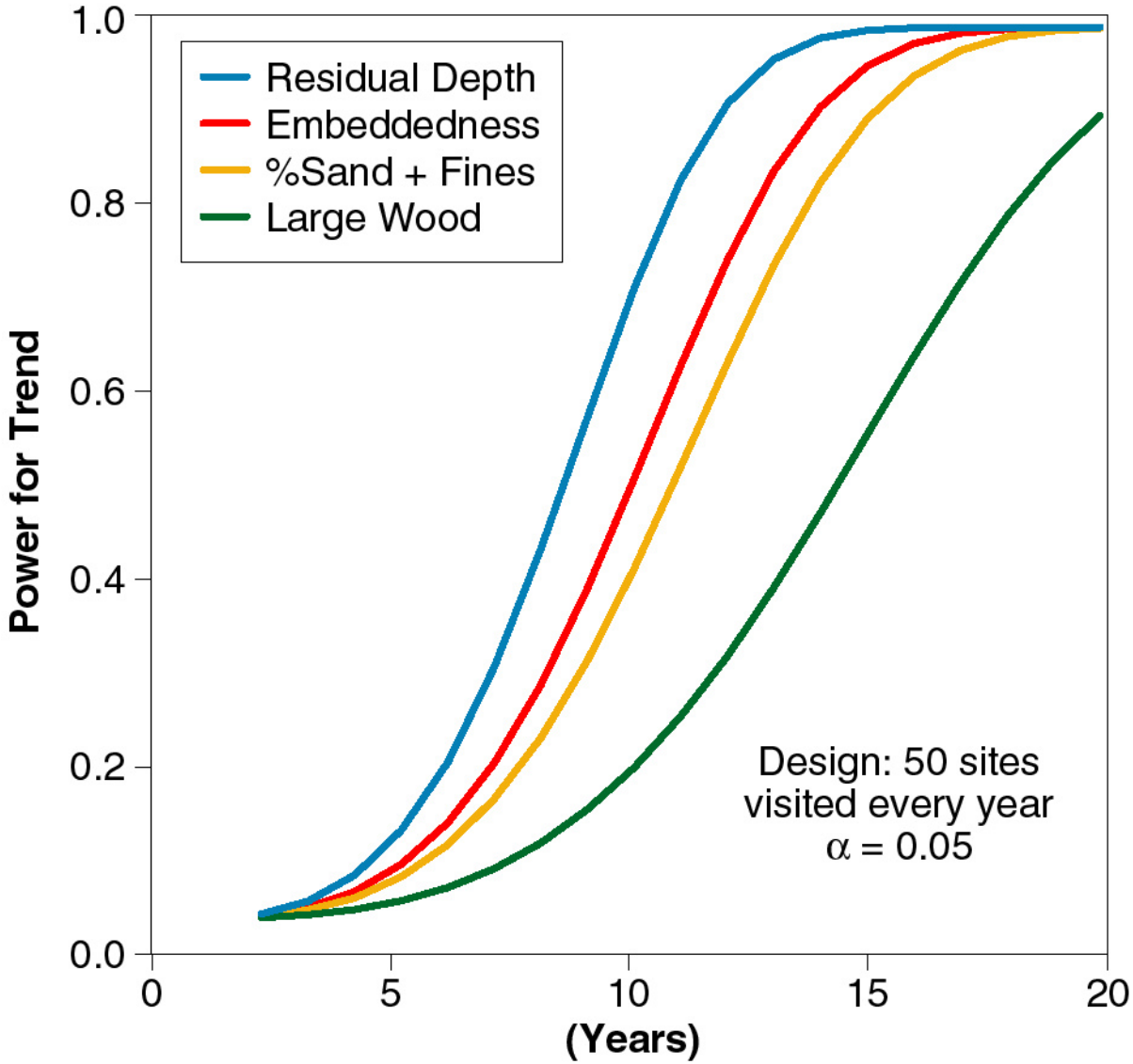
$$\text{var}(slope) = \frac{\overbrace{\frac{\sigma_s^2}{N_s}}^{\text{site}} + \overbrace{\sigma_y^2}^{\text{year}} + \overbrace{\frac{\sigma_i^2}{N_s}}^{\text{interaction}} + \overbrace{\frac{\frac{\sigma_r^2}{N_v}}{N_s}}^{\text{residual}}}{\sum (Y_i - \bar{Y})^2}$$

N_s= Number of sites; N_v= Number of w/in year revisits

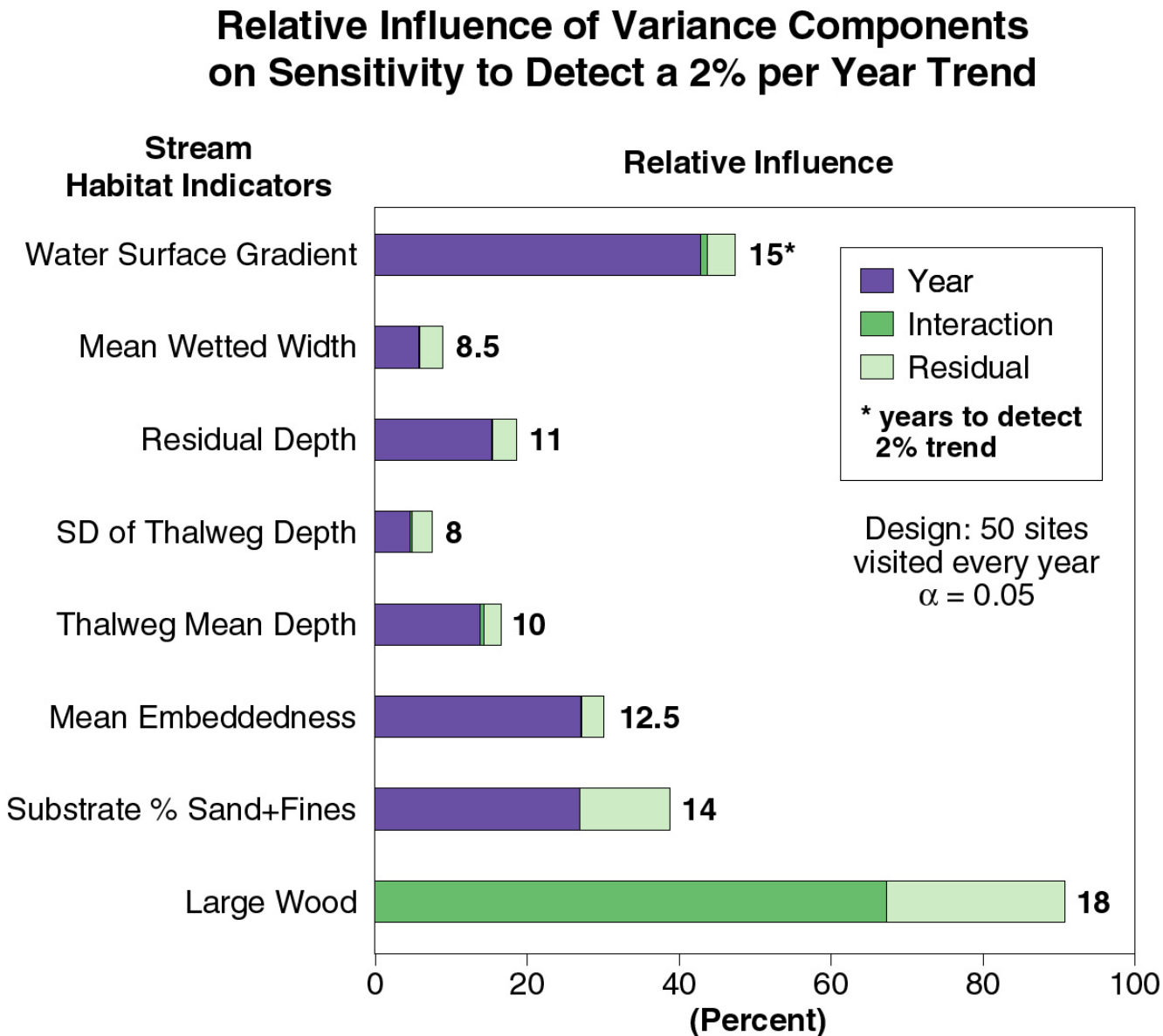
Implications

- Site = 0 if sites are revisited across years
- Year is not sensitive to “sample size”and its effect can become dominant
- Residual is affected by within year revisits
- Interaction and residual are affected by number of sites in survey, therefore other factors being equal, better to add sites to the survey rather than revisit sites

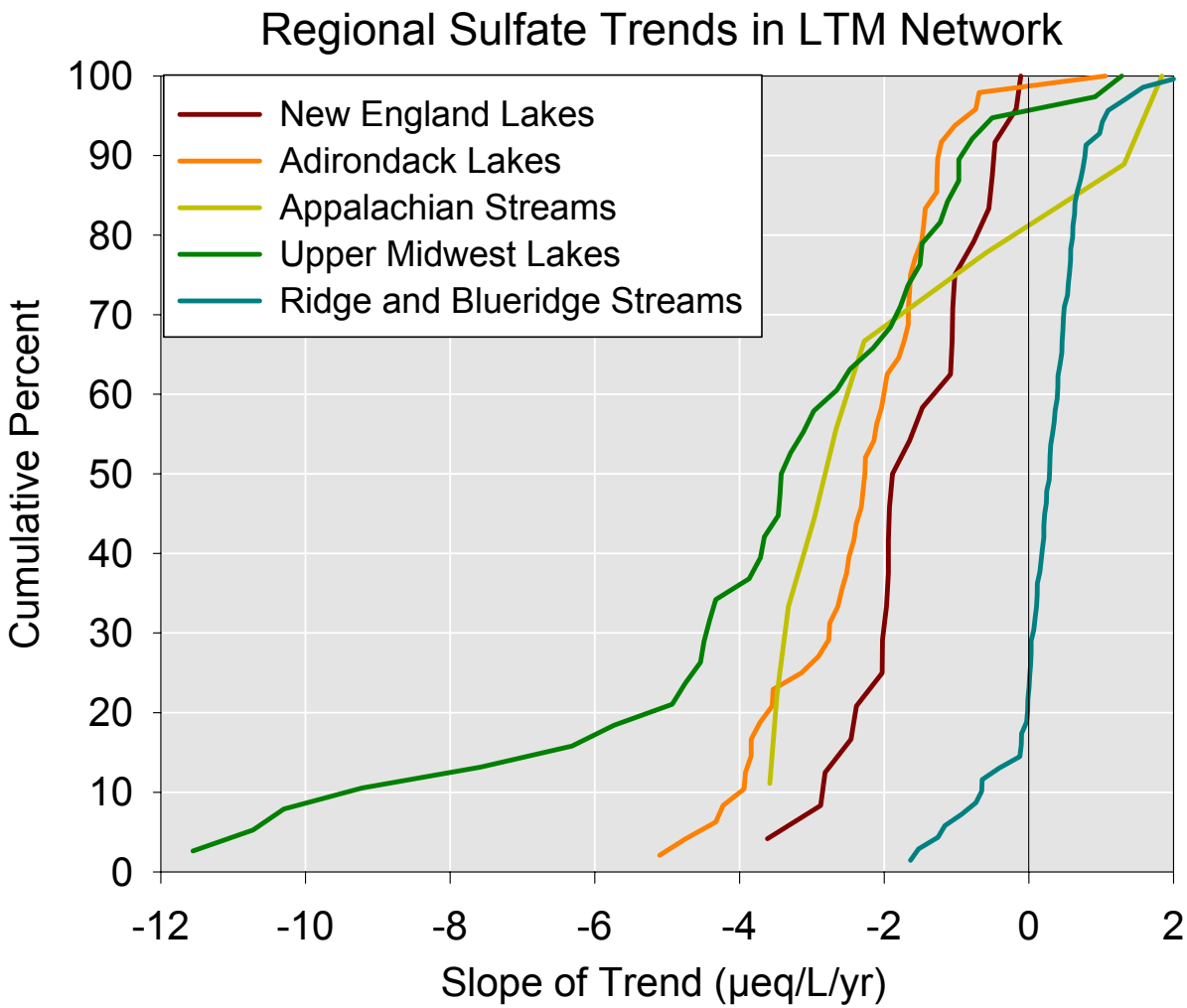
Power to Detect a 2% per Year Trend for Habitat Indicators



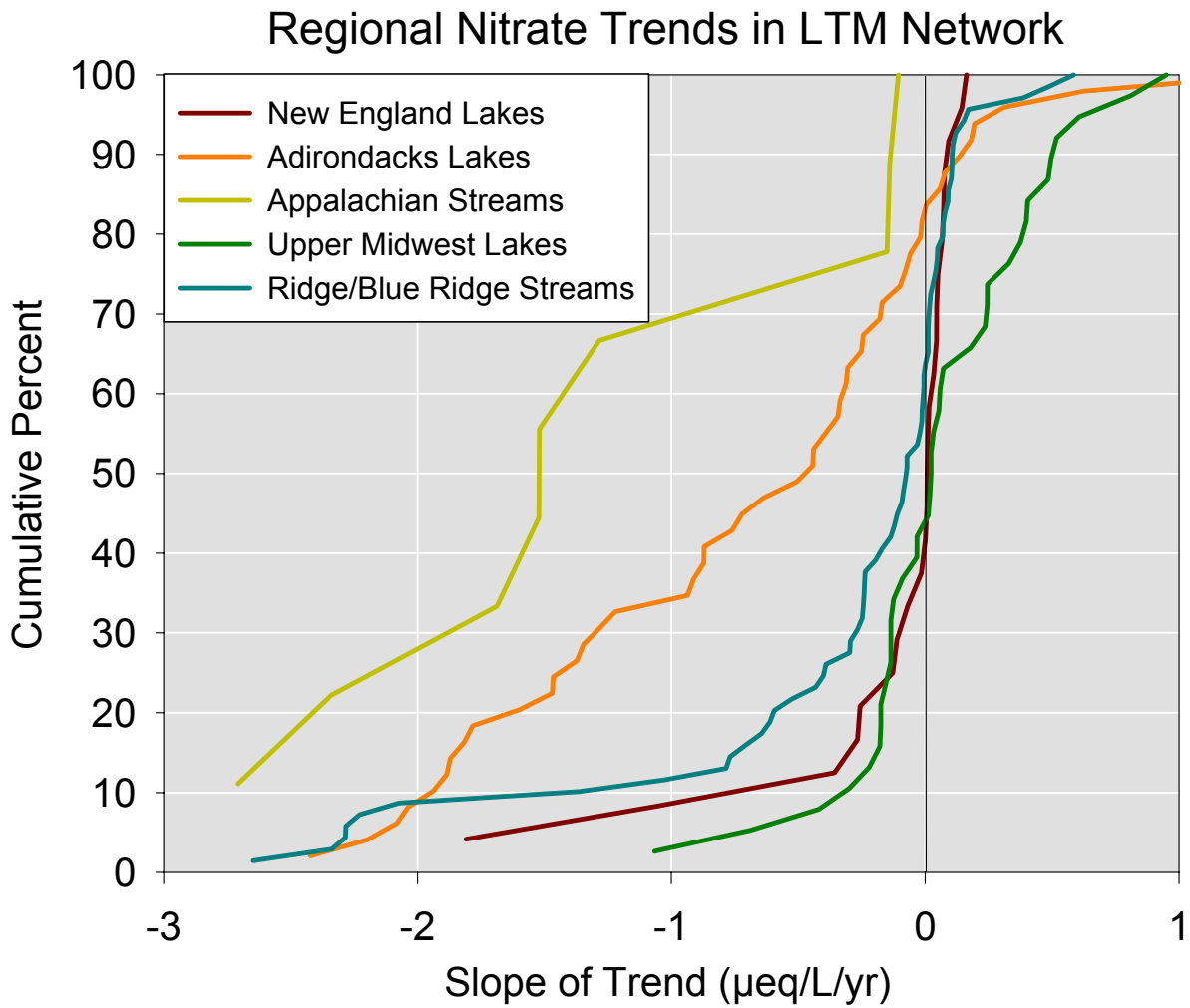
How long will it take to detect a 2%/yr trend with power = 0.8?



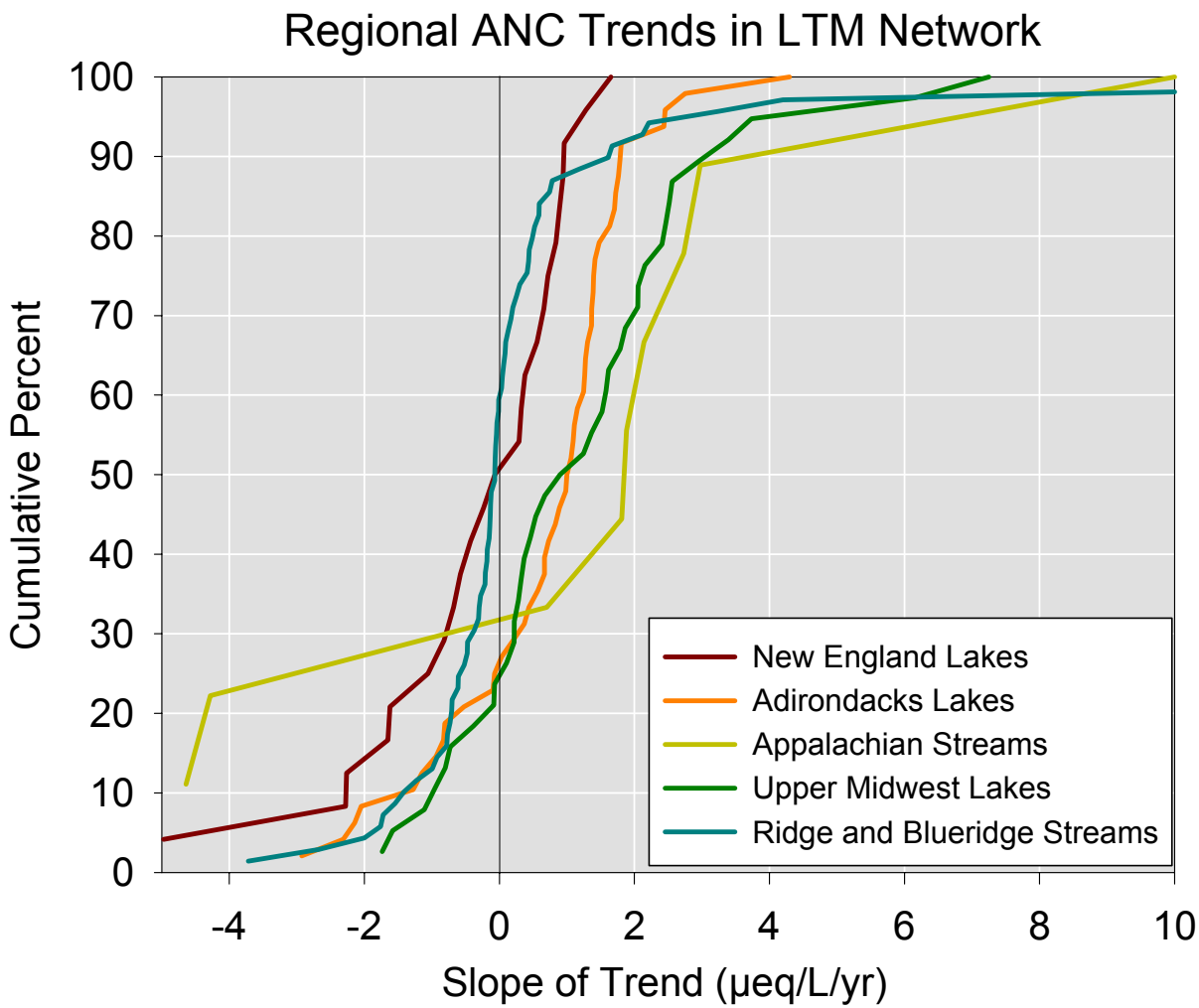
Sulfate Trends in TIME/LTM Regions



Nitrate Trends in TIME/LTM Regions



ANC Trends in TIME/LTM Regions



Some options

- Extend survey interval
- Focus on subpopulations to manage variance
- Monitor hypothesized covariates controlling "year"
- Develop expected trends given management scenarios and options: What can we expect?

Postscript

- EMAP Design Website (Aquatic Resources Monitoring):
  - www.epa.gov/nheerl/arm/

of sites/HUC

Number of Sites	Frequency	
	ODFW	AREMP
1	235	0
2	116	0
3	43	0
4	18	3
5	9	3
6	1	6
7	1	3
8	1	1