Optimizing Environmental Monitoring Designs using Uncertainty Analysis

Carrie R. Levine, UC Berkeley

Ruth D. Yanai, SUNY College of Environmental Science and Forestry

Gregory Lampman, NYSERDA

Douglas A. Burns, US Geologic Survey

Charles T. Driscoll, Syracuse University

Gregory B. Lawrence, U.S. Geological Survey

Jason A. Lynch, US Environmental Protection Agency

Nina Schoch, Biodiversity Research Institute

Who needs environmental monitoring?

Gary M Lovett[1*], Douglas A Burns[2], Charles T Driscoll[3], Jennifer C Jenkins[4], Myron J Mitchell[5], Lindsey Rustad[6], James B Shanley[7], Gene E Likens[1], and Richard Haeuber[8]

Environmental monitoring is often criticized as being unscientific, too expensive, and wasteful. While some monitoring studies do suffer from these problems, there are also many highly successful long-term monitoring programs that have provided important scientific advances and crucial information for environmental policy. Here, we discuss the characteristics of effective monitoring programs, and contend that monitoring should be considered a fundamental component of environmental science and policy. We urge scientists who develop monitoring programs to plan in advance to ensure high data quality, accessibility, and cost-effectiveness, and we urge government agencies and other funding institutions to make greater commitments to increasing the amount and long-term stability of funding for environmental monitoring programs.

Front Ecol Environ 2007; 5(5): 253–260

Environmental monitoring consumes resources and can be criticized for being unscientific.

We need an objective way to evaluate monitoring plans, including the spatial and temporal intensity of sampling.

QUANTIFYING UNCERTAINTY IN ECOSYSTEM STUDIES

Evaluating the efficiency of environmental monitoring programs

Carrie R. Levine, Ruth D. Yanai, Gregory G. Lampman, Douglas A. Burns, Charles T. Driscoll, Gregory B. Lawrence, Jason A. Lynch, Nina Schoch

Ecological Indicators 39: 94 – 101 (2014)

This study was supported by the New York State Energy Research and Development Authority, which supports environmental monitoring for air pollutants associated with the electric power industry.

Analysis of Case Studies

Case study	Model	Parametric or Subsampling	Spatial or Temporal	Research Question
streams	regression	subsampling, followed by parametric	temporal	magnitude of detectable change in slope, one point in space
loons	t-test	parametric (we have verified this approach with subsampling)	spatial	detectable change over time, based on one sample in time
forest biomass	mean	subsampling	spatial (we have also done temporal)	magnitude of spatial uncertainty
lakes	repeated measures mixed effects model	subsampling, followed by statistical analysis	spatial and temporal	uncertainty of concentration; magnitude of detectable change in slope, including spatial variability

Uncertainty in Linear Regression

Question:

• How often should stream chemistry samples be collected to detect long-term chemistry trends?

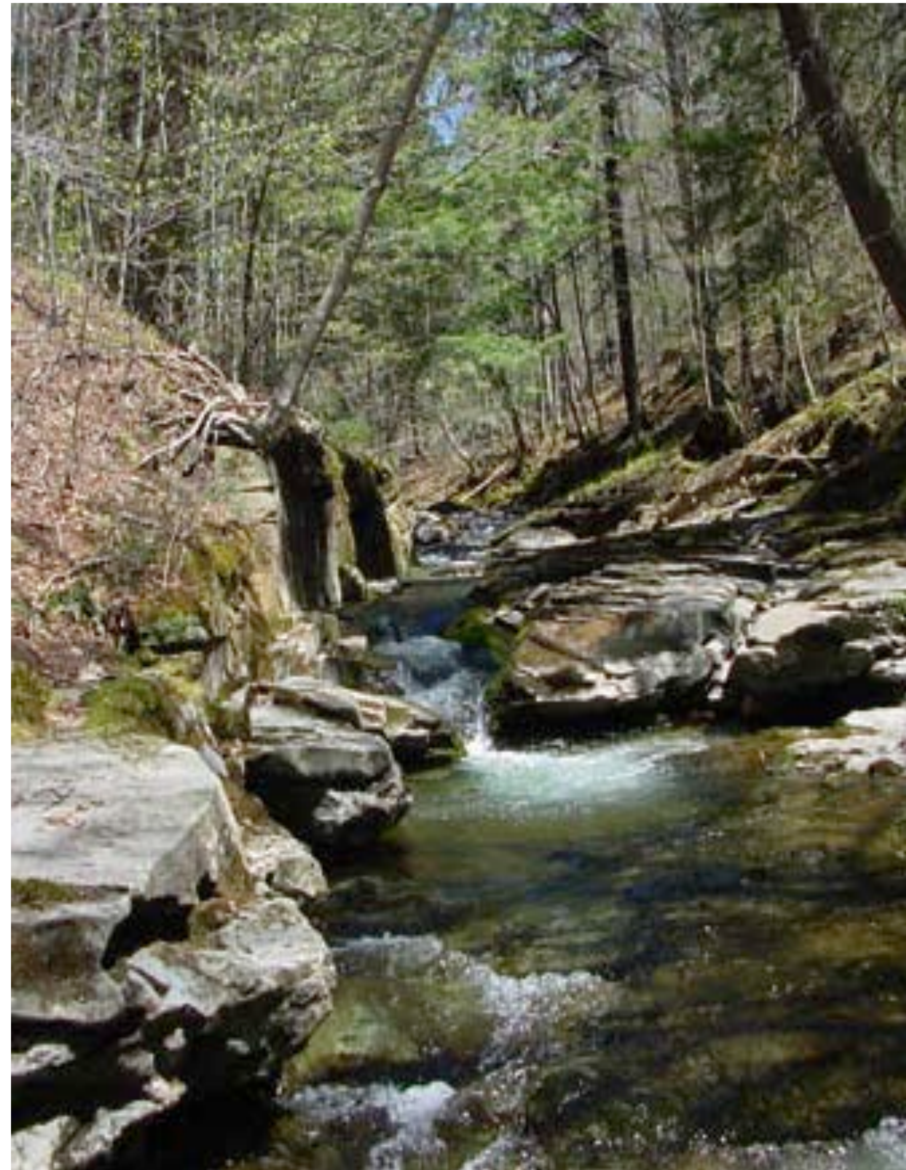
Data sets used in analysis:

• Biscuit Brook weekly stream chemistry (1996-2003).

Analytical approach:

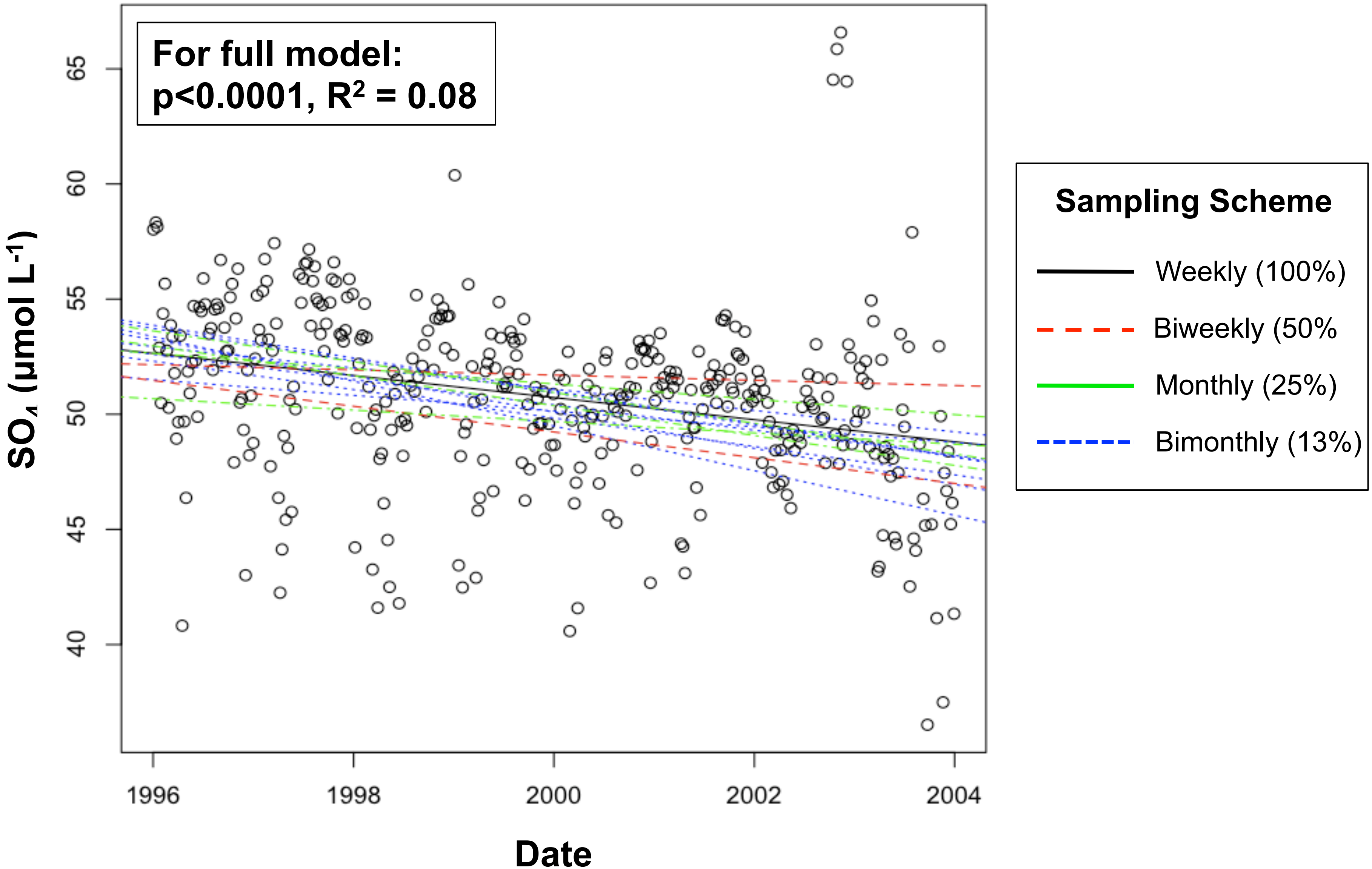
• We simulated reduced sampling efforts and evaluated confidence in the detection of change over time, using linear regression.

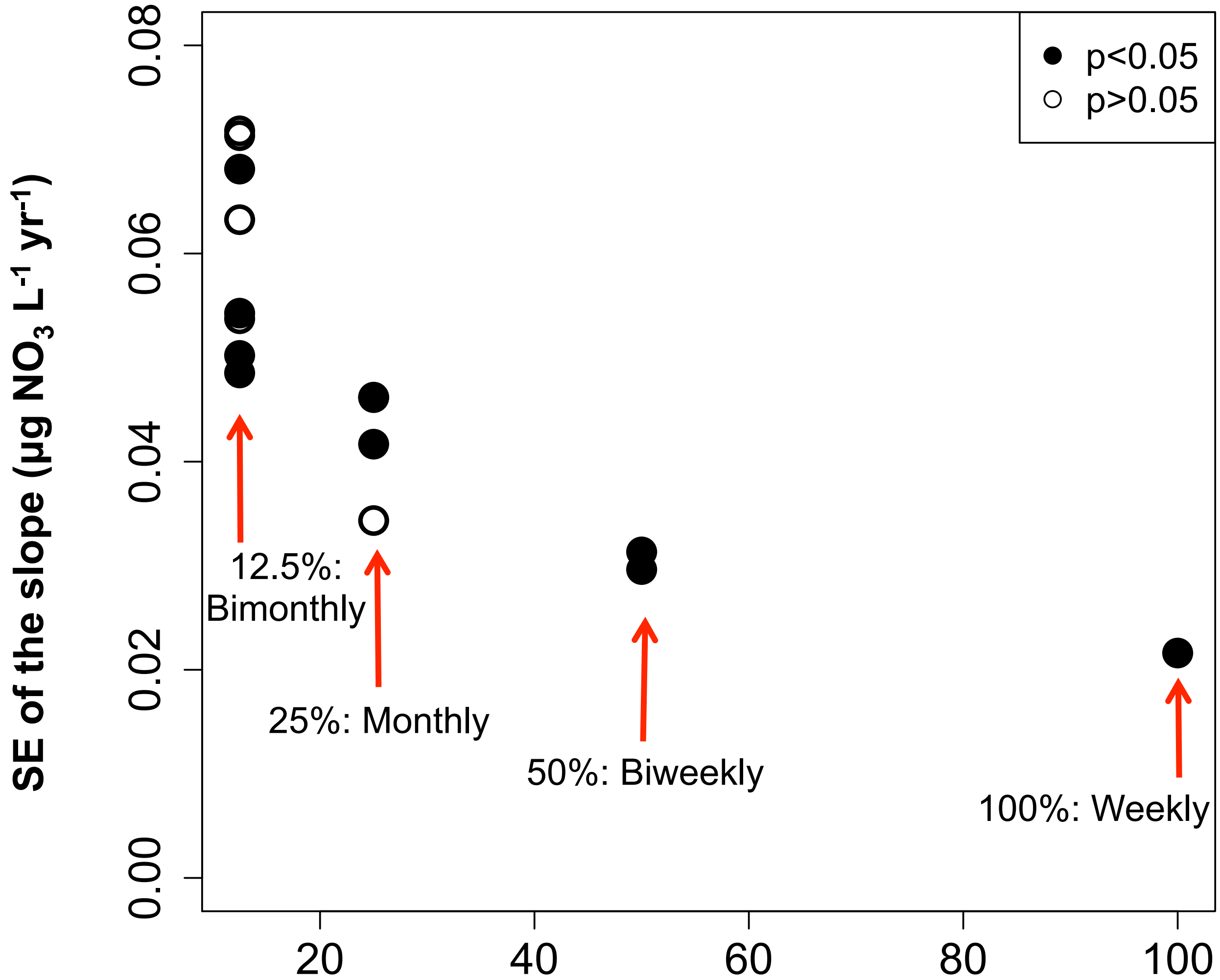
• Weekly, biweekly, monthly, bimonthly.



NYSDEC: http://ny.cf.er.usgs.gov/nyc/site_page.cfm?ID=01434025

Subsampling the data set affects the slope and intercept of the regression of long-term data.





The error in the slope increases as sampling intensity decreases.

Effect of reduced sampling schemes on detectability of long-term trends in stream chemistry at Biscuit Brook (1996-2003)

	# of significant regressions / Total # of possible regressions			
	Weekly	**Biweekly**	**Monthly**	**Bimonthly**
SO_4^{2-}	1/1	2/2	3/4	3/8
NO_3^-	1/1	2/2	3/4	4/8
H^+	1/1	1/2	2/4	2/8
Al	1/1	2/2	4/4	7/8

Detectable Difference (T-test)

Question:

• How many samples would be required to detect a change in mercury in loons at a future sampling date?

Data sets used in analysis:

• One-time survey, 42 lakes, different numbers of loons per lake

Analytical approach:

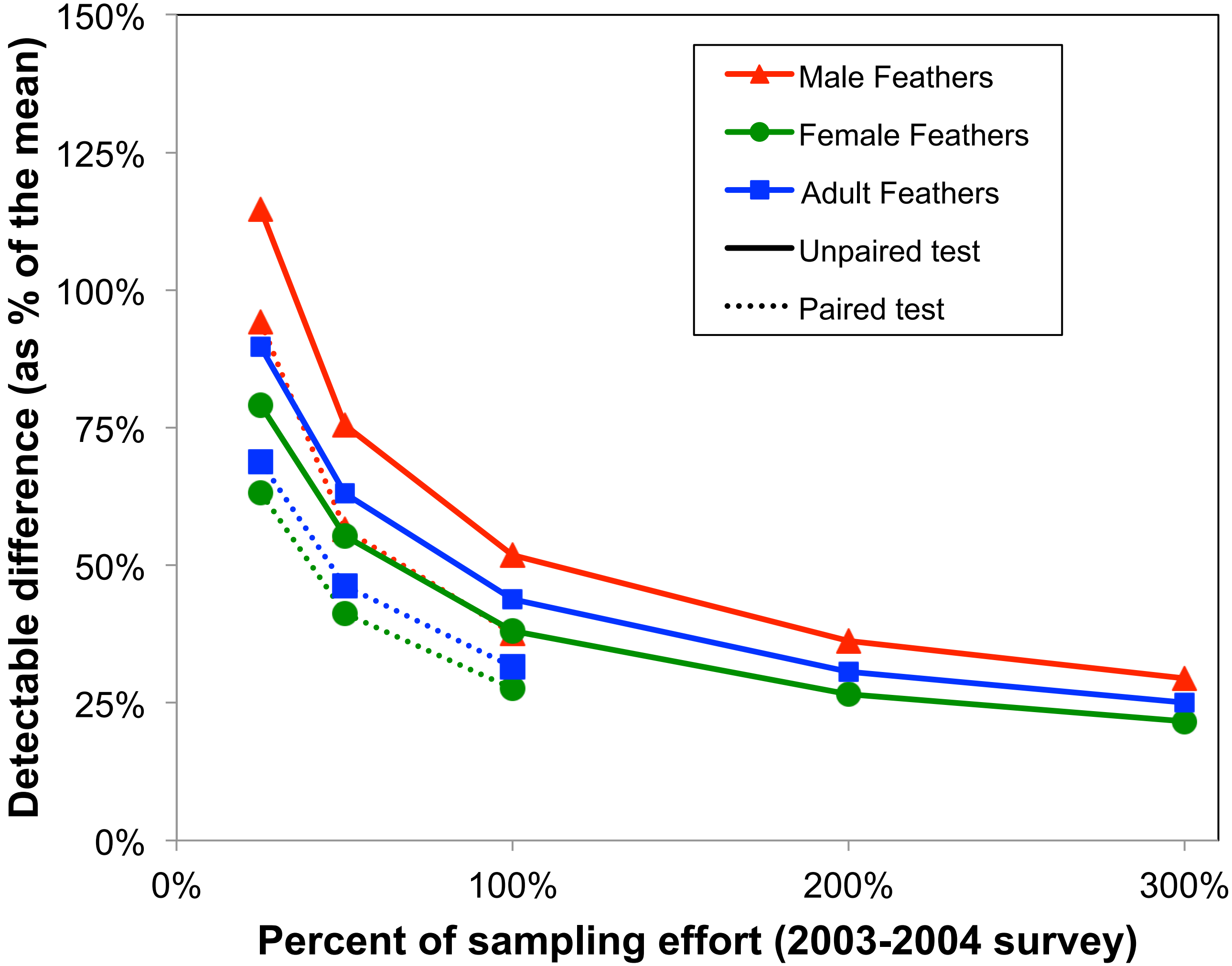
The detectable difference δ for a two-sample t-test is:

$$\delta = (s / \sqrt{n/2})(t_{\alpha,\nu} + t_{\beta,\nu})$$

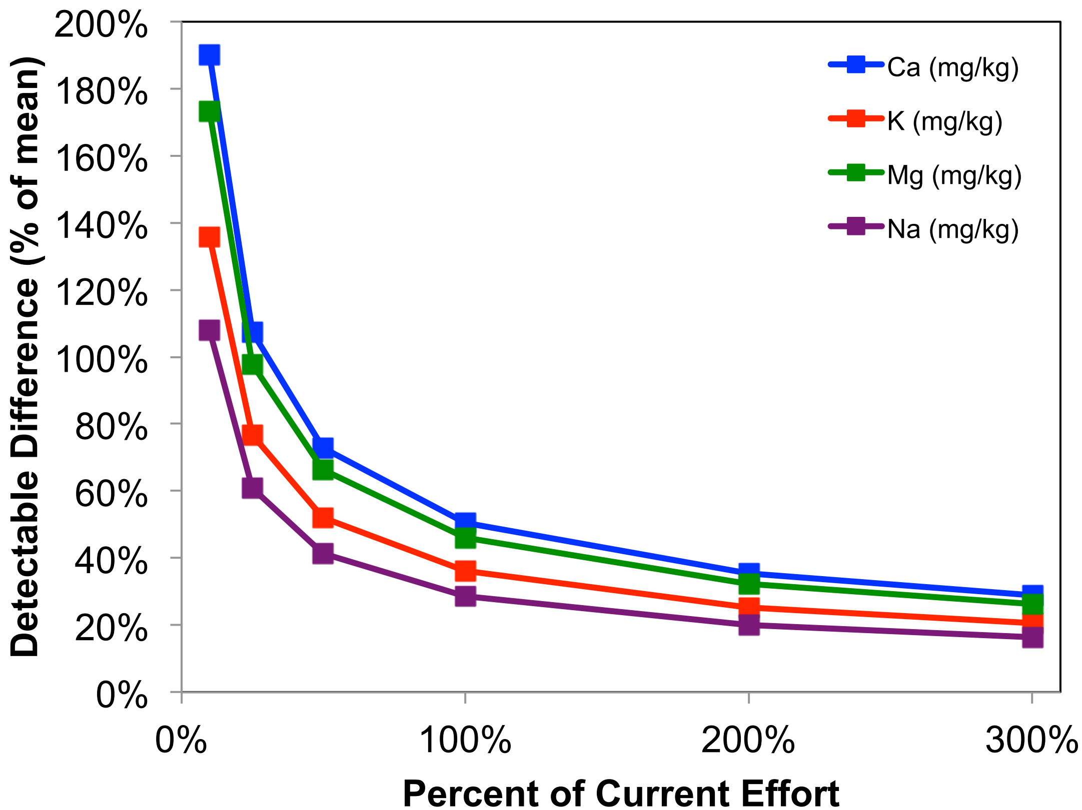
where *s* is the standard deviation of the paired differences, n= sample size, $t_{\alpha,\nu}$ is the (1- α/2) x 100 percentile of the *t*-distribution, $t_{\beta,\nu}$ is the 100 x (power) percentile of the *t*-distribution, ν = 2n-2 degrees of freedom, α is the probability of a Type I error, and β is the probability of a Type II error.

http://images.nationalgeographic.com/wpf/media-live/photos/000/007/cache/common-loon_794_600x450.jpg

Detectable difference of THg in loon blood for females (n=36 lakes), males (n= 37 lakes), all adults (n= 42 lakes) and juveniles (n= 34 lakes).



Detectable difference of exchangeable cation concentrations (mg kg^{-1} dry soil) in mineral soil samples collected by the FIA in 56 plots in the Adirondack region.



These case studies illustrate the effect of sampling intensity on statistical power and the selection of a sampling interval likely to detect an expected change over time

Subsampling

Question:

• How many plots should be sampled to report forest biomass with known confidence?

Data sets used in analysis:

• Hubbard Brook Watershed 6, where every tree is measured on each of 208 plots (each 25m x 25 m) every 5 years. We used data from 2002.

Analytical approach:

• We randomly selected subsets of plots and reported uncertainty in the estimates of forest biomass.

www.plymouth.edu

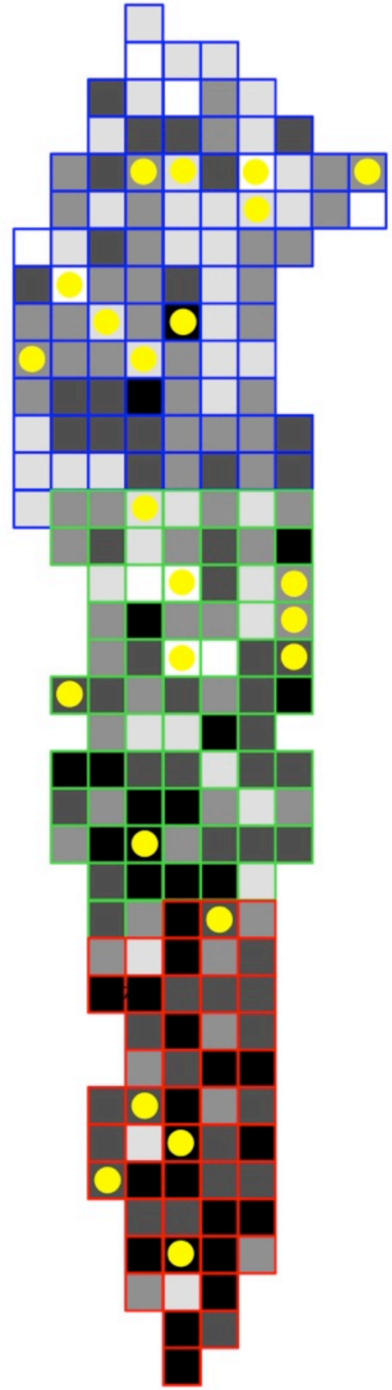
Hubbard Brook W6: Total Biomass by plot as of 2002

Elevation Zones

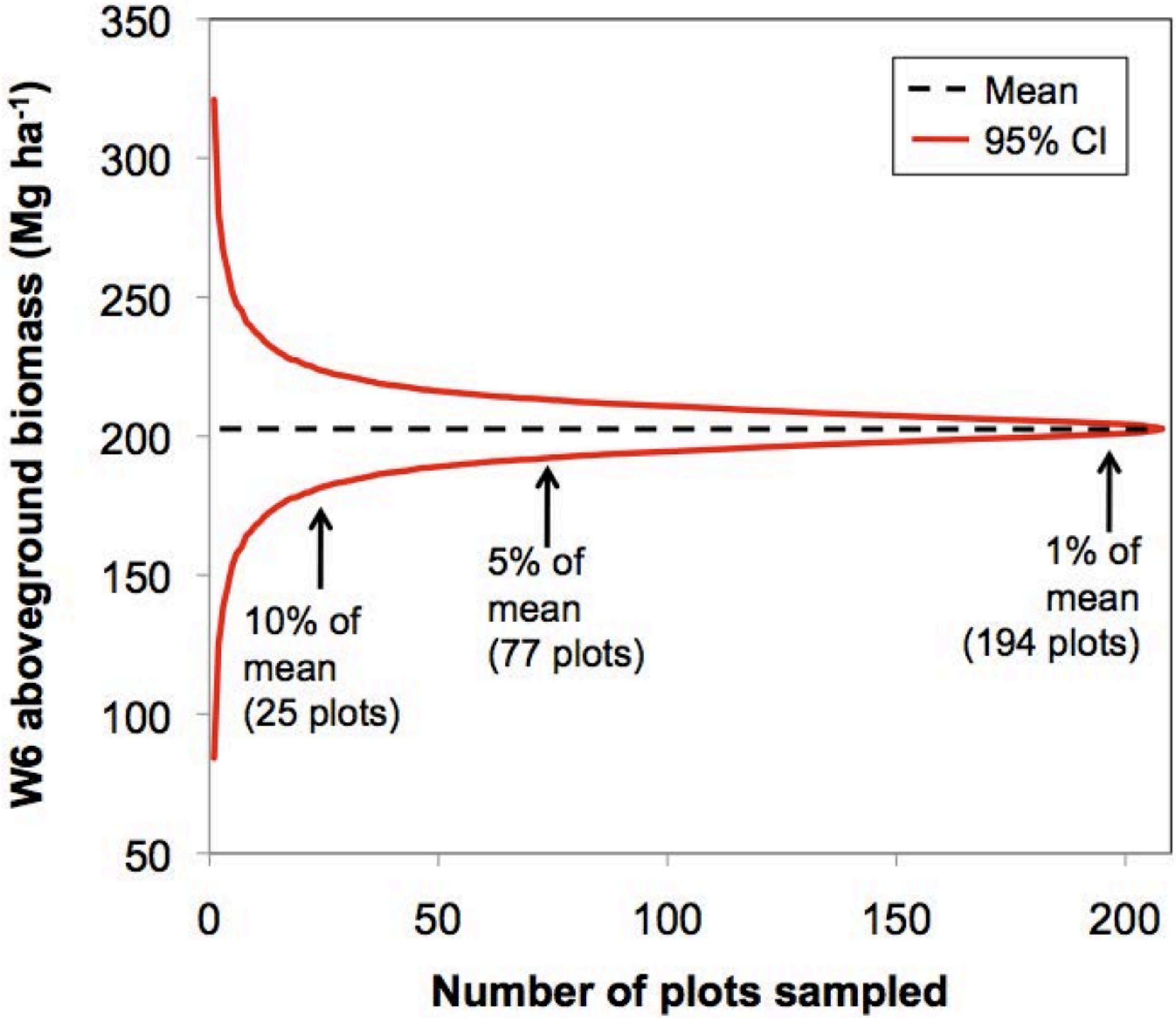
- Upper zone
- Middle zone
- Lower zone

Biomass

- <150 Mg/ha
- 150-200 Mg/ha
- 200-250 Mg/ha
- 250-300 Mg/ha
- >300 Mg/ha



The range in elevation is 550-700 m, with significant vegetation change. Biomass equations were developed for three elevational bands. We used these three bands as strata when subsampling.



W6 aboveground biomass (Mg ha-1)
350
300
250
200
150
100
50
0
50
100
150
200
Number of plots sampled
Mean
95% CI
10% of mean (25 plots)
5% of mean (77 plots)
1% of mean (194 plots)

Yanai et al. (2010, Ecosystems) estimated uncertainty in the Hubbard Brook Valley, including measurement error of tree diameters, uncertainty in allometric equations, and sampling error (with varying numbers of plots)

Table 1. Uncertainty in Estimation of N Content of Trees at Hubbard Brook, Reported as (a) the Coefficient of Variation (the Standard Deviation Divided by the Mean) and (b) kg N/ha of 100 Monte Carlo Iterations

	Stem wood	Stem bark	Branches	Leaves and twigs	Roots	Total biomass
(a) Coefficient of variation (%)						
Diameter measurement	0.03	0.02	0.03	0.02	0.02	0.02
Height equations	3	3	3	2	3	3
Allometric equations	2	5	14	7	6	4
N concentration	5	3	4	2	7	3
Sampling error (15 plots)	8	7	7	5	6	6
All sources combined	9	8	14	9	11	8
All sources, 5 plots	18	13	22	12	18	15
All sources, 10 plots	11	9	20	10	12	10
All sources, 20 plots	8	8	16	10	11	8
All sources, 30 plots	8	7	16	10	10	7
All sources, 40 plots	8	7	15	10	9	7
All sources, 60 plots	7	8	17	10	10	7

Overall uncertainty does not decrease as the number of plots increases above 20, as this source becomes insignificant (the others amount to 7%).

Repeated Measures Mixed Effects Model

Question:

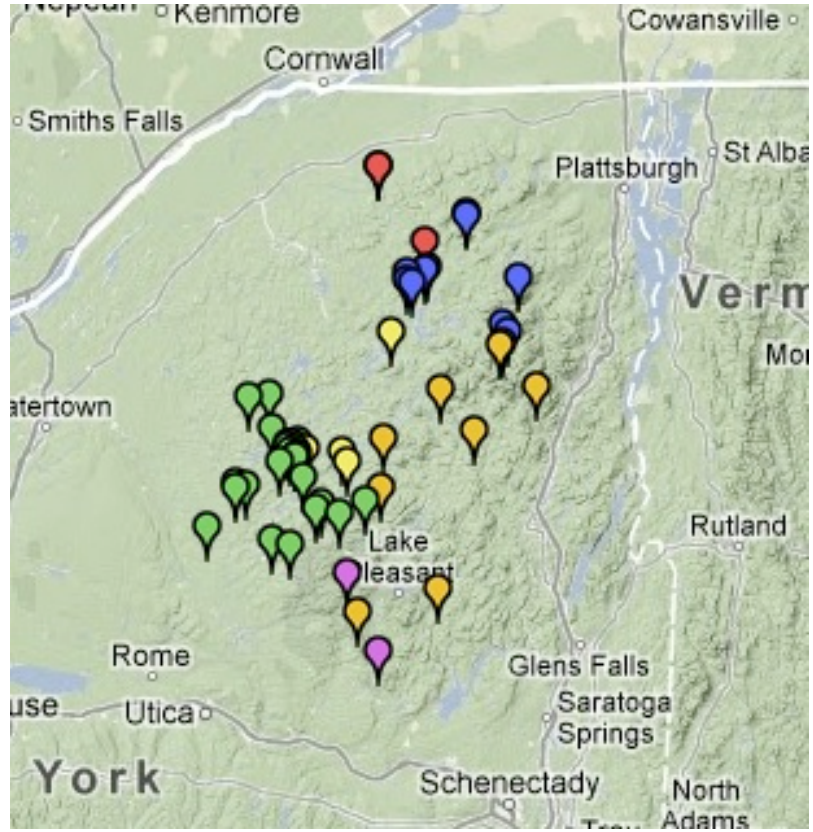
• When monitoring Adirondack lakes, how many lakes should be monitored, and how often?

Data sets used in analysis:

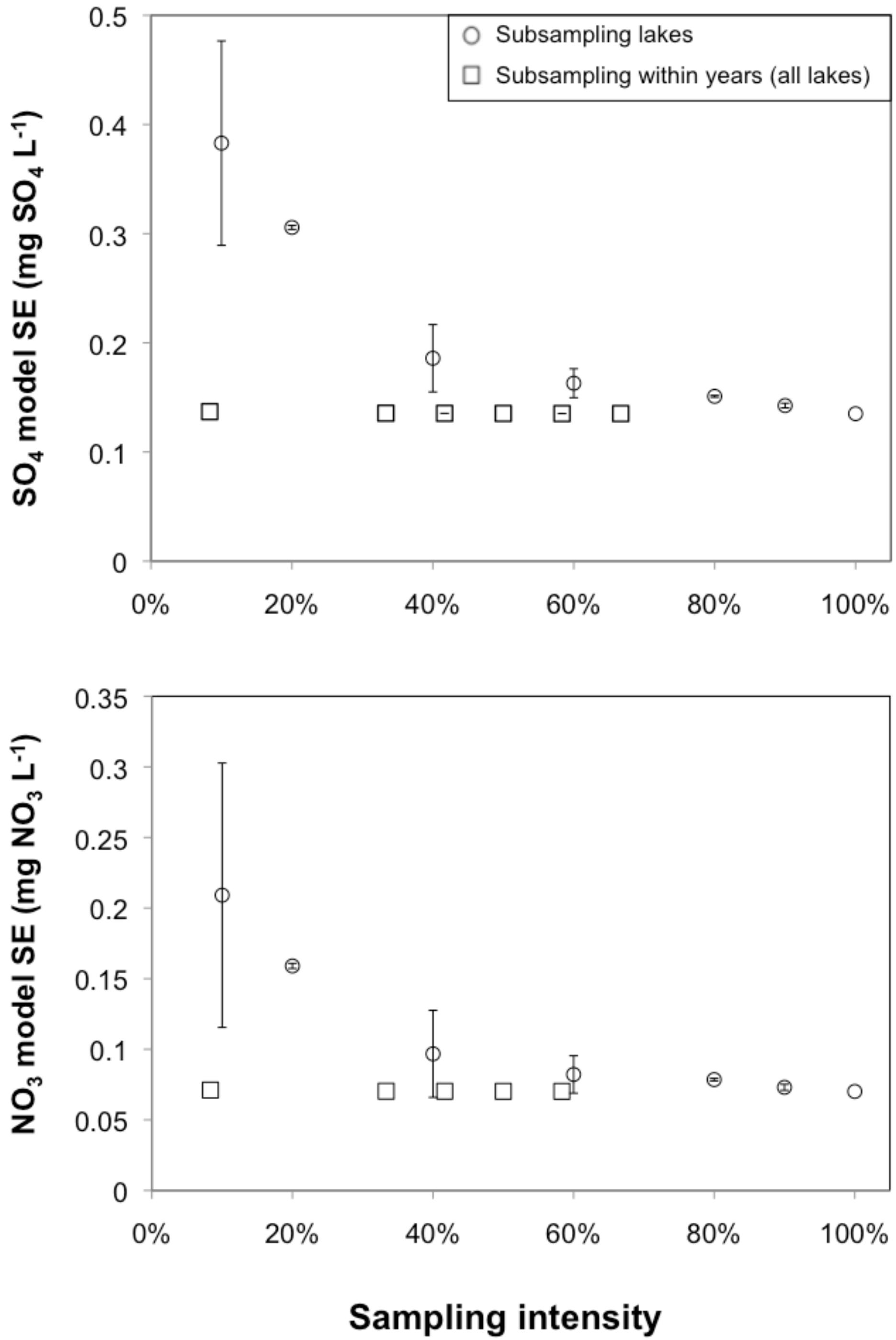
• The Adirondack Lake Survey Corporation monthly lake water samples for a full suite of chemistry analyses from 48 lakes from 1992-2010.

Analytical approach:

• We randomly selected subsets of the data and applied a repeated-measures mixed-effects model to describe uncertainty in the estimates.



http://www.adirondacklakessurvey.org/



0.5
0.4
0.3
0.2
0.1
0
SO4 model SE (mg SO4 L-1)
Subsampling lakes
Subsampling within years (all lakes)
0%
20%
40%
60%
80%
100%
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
NO3 model SE (mg NO3 L-1)
0%
20%
40%
60%
80%
100%
Sampling intensity

The number of lakes showing significant trends over time in mixed model tests decreases as sampling effort decreases

		Number of Lakes Showing Significant Trends Over Time (of a total of 48)						
Percent of Current Sampling Effort	**Sampling Scheme**	**SO_4**	**NO_3**	**NH_4**	**Ca^{2+}**	**ANC**	**H^+**	**SUM**
100	All months*	48	42	26	45	43	36	240
67	Mar-Oct	48	15	0	36	27	15	141
58	Mar-Sept	48	14	0	33	25	17	137
50	Even months	48	9	0	34	23	11	125
50	Odd months	48	9	0	36	25	13	131
42	Mar-Apr, June, Sept-Oct	47	6	0	31	22	9	115
33	Seasonal (Feb, May, Aug, Nov)	46	6	0	27	18	10	107
33	Seasonal (Jan, Apr, July, Oct)	48	6	0	29	15	7	105
33	Seasonal (Mar, Jun, Sept, Dec)	46	5	0	29	22	9	111
33	Mar, Apr, Sept, Oct	47	5	0	31	17	6	106

Summary and Recommendations

Uncertainty analysis can provide an objective way to evaluate monitoring plans, including the spatial and temporal intensity of sampling.

Comparing sources of uncertainty can help identify where best to direct effort to improve knowledge.

Statistical models can handle complex designs, including mixed intensities and unbalanced designs.

When reducing sampling intensity, the information from past sampling is not lost or wasted.

It is important to provide enough information that other researchers can represent the uncertainty in your results.

Join QUEST!

Find more information at:
www.quantifyinguncertainty.org

Read papers, share sample code, stay updated with QUEST News

Email us at
quantifyinguncertainty@gmail.com

Follow us on LinkedIn and Twitter:
@QUEST_RCN