Ecological Indicators 39 (2014) 94–101

Contents lists available at ScienceDirect

Ecological Indicators

journal homepage: www.elsevier.com/locate/ecolind

Evaluating the efficiency of environmental monitoring programs

Carrie R. Levine[a,*], Ruth D. Yanai[a], Gregory G. Lampman[b], Douglas A. Burns[c], Charles T. Driscoll[d], Gregory B. Lawrence[c], Jason A. Lynch[e], Nina Schoch[f]

[a] Department of Forest and Natural Resources Management, State University of New York, College of Environmental Science and Forestry, 1 Forestry Drive, Syracuse, NY 13210, USA
[b] New York State Energy Research and Development Authority, 17 Columbia Circle, Albany, NY 12203, USA
[c] U.S. Geological Survey, 425 Jordan Road, Troy, NY 12180, USA
[d] Department of Civil and Environmental Engineering, Syracuse University, 151 Link Hall, Syracuse, NY 13244, USA
[e] U.S. Environmental Protection Agency, Clean Air Markets Division, 1200 Pennsylvania Avenue NW, Washington, DC 20460, USA
[f] Biodiversity Research Institute, 652 Main Street, Gorham, ME 04038, USA

ARTICLE INFO

Article history:
Received 27 July 2013
Received in revised form
28 November 2013
Accepted 3 December 2013

Keywords:
Uncertainty
Long-term
Lakes
Monitoring
Streams
Loons
Biomass

ABSTRACT

Statistical uncertainty analyses can be used to improve the efficiency of environmental monitoring, allowing sampling designs to maximize information gained relative to resources required for data collection and analysis. In this paper, we illustrate four methods of data analysis appropriate to four types of environmental monitoring designs. To analyze a long-term record from a single site, we applied a general linear model to weekly stream chemistry data at Biscuit Brook, NY, to simulate the effects of reducing sampling effort and to evaluate statistical confidence in the detection of change over time. To illustrate a detectable difference analysis, we analyzed a one-time survey of mercury concentrations in loon tissues in lakes in the Adirondack Park, NY, demonstrating the effects of sampling intensity on statistical power and the selection of a resampling interval. To illustrate a bootstrapping method, we analyzed the plot-level sampling intensity of forest inventory at the Hubbard Brook Experimental Forest, NH, to quantify the sampling regime needed to achieve a desired confidence interval. Finally, to analyze time-series data from multiple sites, we assessed the number of lakes and the number of samples per year needed to monitor change over time in Adirondack lake chemistry using a repeated-measures mixed-effects model. Evaluations of time series and synoptic long-term monitoring data can help determine whether sampling should be re-allocated in space or time to optimize the use of financial and human resources.

1. Introduction

Environmental monitoring is essential for detecting changes associated with biological invasions, land use change, and stressors such as air pollutants and climate change. Monitoring is also needed to evaluate the effects of past or proposed environmental policies and resource management activities (Lovett et al., 2007). Data from ongoing long-term monitoring programs can be used to assess the efficiency and effectiveness of those programs and can inform the development of other similar monitoring programs.

The intensity of monitoring needed to detect trends over space and time is determined by the natural spatial and temporal variation of the measured parameters, measurement and model error, and the acceptable Type I error rate. A parameter with high natural variation requires more intensive sampling to achieve the same statistical power (Garman et al., 2012). Periodic evaluation of monitoring programs is important because the objectives of a monitoring plan may change, the available technology improves, and the amount of data accumulates over time (Lovett et al., 2007). However, most literature on designing ecological monitoring programs is focused on the initial program design (McDonald, 2012; Urquhart, 2012). There has been less focus on strategies for adapting existing monitoring plans to maximize effectiveness and efficiency using the data already collected. In this paper, four case studies are presented to illustrate the evaluation of existing monitoring schemes.

Long-term monitoring of stream chemistry can be used to assess changes in watershed dynamics and to evaluate environmental effects of atmospheric deposition of air pollutants. In the northeastern US, acidic deposition of sulfuric and nitric acids and ammonium result in elevated leaching of nutrient cations from available soil

* Corresponding author. Present address: Department of Environmental Science, Policy, and Management, University of California, Berkeley, 130 Mulford Hall, Berkeley, CA 94720, USA. Tel.: +1 773307 2583.
E-mail addresses: crlevine@berkeley.edu (C.R. Levine), rdyanai@syr.edu (R.D. Yanai), ggl@nyserda.ny.gov (G.G. Lampman), daburns@usgs.gov (D.A. Burns), ctdrisco@syr.edu (C.T. Driscoll), glawrenc@usgs.gov (G.B. Lawrence), lynch.jason@epamail.epa.gov (J.A. Lynch), nina.schoch@briloon.org (N. Schoch).

1470-160X
http://dx.doi.org/10.1016/j.ecolind.2013.12.010

Table 1
Description of the four case studies and models included in this paper.

Case study	Data set includes spatial replication	Data set includes temporal replication	Model used	Parametric or subsampling approach	Research question
Stream chemistry	No	Yes	Regression	Subsampling, followed by parametric	Magnitude of detectable change in slope for one point in space
Loon tissue chemistry	Yes	No	Detectable difference (*t*-test)	Parametric	Detectable change over time based on one sample in time
Forest biomass	Yes	No	Mean	Subsampling	Magnitude of spatial uncertainty for one point in time
Lake chemistry	Yes	Yes	Repeated measures mixed effects model	Subsampling, followed by parametric	Magnitude of detectable change in model mean, including spatial and temporal variability

pools, resulting in deleterious effects on vegetation (Horsley et al., 2000; Battles et al., 2013). Acidic deposition has also had negative effects on water quality in the region by causing decreases in pH and acid-neutralizing capacity (ANC) and increases in aluminum concentrations (Driscoll et al., 2003a).

We use the long-term stream chemistry monitoring at Biscuit Brook in the Catskill Mountains, NY, to illustrate an approach for selecting a sampling plan for a single site with multiple measurements over time. Biscuit Brook has been sampled weekly for 28 years to evaluate the effects of atmospheric acid deposition on stream chemistry (Murdoch and Shanley, 2006). Using this long-term record, we subsample data sets of stream NO_3 and SO_4 concentrations to assess the effects of sampling frequency on the power to detect changes in stream chemistry.

Mercury contamination in water and plant and animal tissues is a major environmental concern. Mercury deposited on the landscape from the atmosphere can be converted into methylmercury, a neurotoxin that bioaccumulates up aquatic and terrestrial food chains and impairs animal and human health (Driscoll et al., 2007). There have been several recent studies of Hg in surface waters and in tissues of aquatic fauna (Yu et al., 2011; Turnquist et al., 2011; Simonin et al., 2006); detecting change over time will require repeating these measurements in the future. To explore sampling strategies for trend detection, we apply a simple power analysis to assess the sampling intensity and interval needed for future surveys, based on expected rates of change. We use the example of a survey of Hg concentrations in feathers and blood in the Common Loon (*Gavia immer*) in the Adirondacks (Schoch et al., 2011) to illustrate this approach.

Forests are routinely monitored to assess stand dynamics, forest health and productivity, and timber and non-timber values (Fahey and Knapp, 2007; Smith, 2002). Methods for accurately measuring forest biomass accumulation have important implications for quantifying carbon sequestration in emerging carbon markets and for evaluating alternatives to fossil fuels (Woodbury et al., 2007; Stupak et al., 2007; Raciti et al., 2012). By assessing the uncertainty due to sample size, we can determine how many plots must be sampled to achieve a desired confidence. To illustrate this approach, we use the example of the reference watershed at the Hubbard Brook Experimental Forest, NH. Because all the plots on the watershed are measured in each census, we can subsample plots using a bootstrapping approach to assess the effect of sampling intensity on the confidence in biomass estimates.

Lake ecosystems have been used to monitor the effects of environmental policies affecting acidic deposition (Driscoll et al., 2003b). The Adirondack region of NY exhibits some of the most severe impacts from acidic deposition in the US, and lake systems there have been monitored to assess recovery of these ecological systems over time (Civerolo et al., 2011). The long-term monitoring of Adirondack lake chemistry is similar to many long-term environmental monitoring programs that involve sampling many sites at regular intervals. It may be the case that the sampling regime is deficient or excessive in either space or time. We apply a repeated measures mixed-effects statistical model to data sets that simulated reduced sampling schemes in space and time.

These four case studies were chosen because they represent a variety of experimental designs common to environmental monitoring programs (Table 1). The objective of this paper is to illustrate approaches for reviewing the efficiency and effectiveness of environmental monitoring schemes. These approaches can be applied to a wide range of sampling designs. We describe best practices for evaluating monitoring programs and give examples of additional approaches that may be useful for analyzing complex long-term monitoring data.

2. Methods

2.1. General linear model error applied to stream sampling for long-term trends at Biscuit Brook

We used a general linear regression and the standard error of the slope to assess uncertainty in long-term temporal trends for Biscuit Brook. Biscuit Brook (latitude 41°59′43″, longitude 74°30′05″) is a water-quality station in the Neversink Reservoir watershed in the Catskill Mountains, NY. Biscuit Brook drains a 963-ha forested watershed with a gauging station at an elevation of 628 m (McHale and Siemion, 2010). The Biscuit Brook station has been in operation since 1983 and has been sampled weekly by the US Geological Survey since 1991. Field and laboratory methods for streamflow and water quality data collection are described by McHale and Siemion (2010).

We subsampled the weekly stream chemistry data from Biscuit Brook (1996–2003) to assess the effects of sampling intensity on the regression standard error and the ability to detect change over time. Subsampled data sets representing 50% of current effort (measurements taken every other week), 25% of current effort (measurements taken every month), or 12.5% of current effort (measurements taken every other month) were used in this analysis.

We used this approach to examine the effects of subsampling on SO_4 and NO_3 concentrations, as these are commonly monitored to assess the effects of atmospheric deposition on long-term stream chemical trends. Sulfate has shown a consistent decreasing long-term trend at Biscuit Brook with little seasonal or flow-related variability (Kerr et al., 2012). In contrast, NO_3 shows a more variable long-term trend with high seasonal and flow-related variability for the period we considered. Thus, these solutes illustrate scenarios for which long-term trends in solute concentration may be more or less difficult to detect due to within-year variation and the strength of the trend over time.

We compared the standard error of the regression across the subsampling scenarios. We also compared the significance of the slopes of the subsampling scenarios to assess the ability of reduced

sampling schemes to detect a significant change in stream chemistry over time. All regression analyses were done in R64 (R Core Team, 2013).

2.2. Detectable difference analysis applied to loon tissue sampling in the Adirondacks

The Biodiversity Research Institute (BRI) has conducted studies on the trophic transfer of Hg in Adirondack lakes and the impact of Hg deposition on loon health and reproductive rates (Yu et al., 2011; Schoch et al., 2011). In addition to annual surveys of Hg concentrations in loon tissues, BRI also conducted a one-time survey of Hg concentrations in loon tissues. The survey was conducted on 44 lakes within the Adirondack State Park, New York from 2003 to 2004 (Schoch et al., 2011). In this analysis, we used Hg concentrations in loon blood and feathers derived from this one-time survey. The sample size in this analysis was the number of lakes, as the loons within a lake are not independent samples of the population and the individual loons may not persist over time.

We used a detectable difference approach to assess the sample size required to detect a difference of a particular magnitude of change in Hg concentrations of loon blood and feathers. In a detectable difference analysis, the input variables include the sample size and standard deviation of the original survey data and the α and power level $(1-\beta)$. We used the commonly accepted values of 0.05 for α and 0.8 for power (Lenth, 2001). The detectable difference δ for a two-sample t-test (Eq. (1)) is calculated as follows:

$$\delta = \left(\frac{S}{\sqrt{n/2}}\right)(t_{\alpha,\upsilon} + t_{\beta,\upsilon}) \quad (1)$$

where s is the standard deviation of the paired differences, n is the sample size (number of lakes where loon individuals were sampled), $t_{\alpha,\upsilon}$ is the $(1-\alpha/2)\times 100$ percentile of the t-distribution, $t_{\beta,\upsilon}$ is the $100\times$(power) percentile of the t-distribution, $\upsilon = 2n-2$ degrees of freedom, α is the probability of a Type I error, and β is the probability of a Type II error (Zar, 1996, p. 109). When using a paired test, we replace $\sqrt{(n/2)}$ with $\sqrt{n}$ (Yanai et al., 2003).

The detectable difference tests require several assumptions. The first is that the standard deviation (s) of the initial sampled population remains the same between sampling periods. A different s could be assumed, but often there is no basis for predicting a change toward higher or lower variance. Second, the future sample is either paired or unpaired, depending on whether the same sampling units are re-measured at the second sampling date. The paired test can detect smaller differences, and we favor this approach when it is applicable. Clearly, if samples are added, they cannot be paired with previous measurements and an unpaired analysis is needed. In our examples, a normal t distribution is assumed, but a non-parametric subsampling approach should be used in the case of a non-normal distribution. In the case of paired tests, we tested sample sizes of 10%, 25%, 50%, and 100% of the sampling effort of the original survey. We used unpaired tests to estimate detectable differences for sampling efforts representing 200% and 300% of the sampling effort of the original survey. These analyses were done in R (R Core Team, 2013).

2.3. Bootstrapping approach applied to tree biomass at Hubbard Brook

We used a bootstrapping approach to estimate the sampling uncertainty in aboveground biomass for a reference watershed, Watershed 6, at the Hubbard Brook Experimental Forest in NH. The watershed is divided into 208 plots, each 25 m × 25 m (Fig. 1). The diameter of all trees >10 cm in diameter at breast height (1.37 m) were measured in all 208 plots (Siccama et al., 1970). We used the

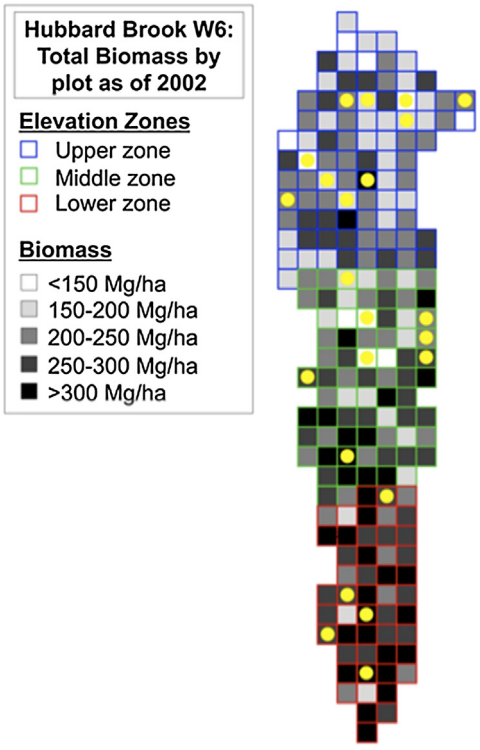

Fig. 1. Map of Hubbard Brook Watershed 6 divided into 208 plots (25 m × 25 m); biomass for each plot is from 2002. Yellow circles indicate a random sample of 20% of the plots, stratified by elevation zone.

data from the 2002 survey in this analysis. We described sampling uncertainty by randomly selecting a subset of plots without replacement, stratified by three elevational bands (upper, middle, and lower elevation; Fig. 1). Because the elevation bands are not equal in size, we subsampled according to the proportions represented in the full data set. In this exercise, we did not include sources of uncertainty other than sampling intensity, such as the error of the allometric equations used to estimate tree biomass or the error in measuring tree diameters (Yanai et al., 2012). We repeated this procedure 1000 times for subsampling intensities from 1 to 208 plots. We determined how many plots would be necessary to produce confidence intervals that were 10%, 5%, and 1% of the mean; the number of plots was rounded up to the nearest integer. The bootstrapping analysis was done in R (R Foundation for Statistical Computing, Vienna, Austria).

2.4. Repeated-measures mixed effects model applied to long-term trends in lake chemistry in the Adirondacks

We used a repeated-measures mixed-effects model to assess the effect of subsampling on the model mean and standard error of a long-term study of lake chemistry in the Adirondack Park. We used data collected by the Adirondack Lake Survey Corporation's Adirondack Long Term Monitoring (ALTM) program. The ALTM program has collected water samples from 52 lakes monthly since June 1992 to assess effects of acidic deposition on water quality in the Adirondack region (Roy et al., 2010). These lakes are located throughout the Adirondack Park and were selected to encompass a range of hydrologic and chemical characteristics. In this analysis, we used concentrations of NO_3 and SO_4 from 48 lakes (4 experimentally limed lakes were not included) from June 1992 to December 2010.

The mixed-effects model is a generalized linear model that can include random and fixed effects. The lakes were considered a random effect. This model did not include any fixed effects because

Table 2
Subsampling scenarios of ALTM lakes and ratios of model SE based on reduced data sets to those using all the data. A ratio >1.00 indicates that the reduced model has a higher SE than the full model. For the simulated sampling schemes at 50%, 33% and 8% temporal effort, the values listed below are the average of 2, 4, and 12 scenarios, respectively.

	% of current effort	Simulated sampling scheme	Ratio of model SE of the reduced to the full data set	
			SO_4	NO_3
Subsampling lakes	90%	43 lakes	1.23	1.46
	80%	38 lakes	1.30	1.57
	60%	29 lakes	1.41	1.64
	40%	19 lakes	1.60	1.93
	20%	10 lakes	2.64	3.18
	10%	5 lakes	3.30	4.18
Subsampling months	67%	March–October	1.00	1.00
	58%	March–September	1.00	1.00
	50%	Every other month[a]	1.01	1.01
	33%	Quarterly[b]	1.00	1.01
	8%	Yearly[c]	1.01	1.01

[a] The two scenarios averaged here are sampling of even months and sampling of odd months.
[b] The four scenarios of quarterly sampling are (1) January, May, September; (2) February, June, October; (3) March, July, November; and (4) April, August and December.
[c] The twelve examples of annual sampling include one sample for each month.

there were no measurements nested within the sites. The time series of measurements from each lake was treated as a repeated measure. Because samples were uniformly spaced (monthly sampling), a first-order autoregressive structure was used. Chemical analytes were analyzed in separate models.

To evaluate the effect of sampling intensity on the standard error of the slope, subsamples of sites or sampling dates were randomly selected. A bootstrap routine of 500 iterations was used to simulate sampling 90% (43 lakes), 79% (38 lakes), 60% (29 lakes), 40% (19 lakes), 21% (10 lakes), and 10% (5 lakes) of the lakes (Table 2). When randomly selecting lakes for simulated scenarios, lakes were stratified by acid neutralizing capacity (ANC) class to ensure that models included a representative subsample. The ANC categories used here were <0 $\mu eq\,L^{-1}$ (7 lakes), 0–50 $\mu eq\,L^{-1}$ (27 lakes), and >50 $\mu eq\,L^{-1}$ (14 lakes). The ANC value of each lake was based on the average of three years of recent sampling (2008–2010). We also simulated scenarios in which sampling was reduced by systematically eliminating months within years, rather than by eliminating lakes (Table 2). When generating subsamples, data were randomly sampled without replacement.

To compare the effect of sampling intensity on the uncertainty in estimates, we used the Least Squares Means (LSM) model standard error. When the LSM standard error is higher under reduced sampling, this indicates a decreased ability to detect significant changes over time. Repeated measures mixed effects tests were done in SAS 9.3 (SAS Institute, 2008).

3. Results

3.1. General linear model error applied to stream sampling at Biscuit Brook

The simulation of reduced sampling schemes at Biscuit Brook (1996–2003) showed that as sampling was reduced, the uncertainty in the slope increased. The uncertainty in the change over time in SO_4 concentrations based on weekly samples was 0.02 $\mu g\,SO_4\,L^{-1}\,yr^{-1}$; for biweekly samples it was 0.03 $\mu g\,SO_4\,L^{-1}\,yr^{-1}$, for monthly samples 0.04 $\mu g\,SO_4\,L^{-1}\,yr^{-1}$, and for bimonthly samples, it was 0.06 $\mu g\,SO_4\,L^{-1}\,yr^{-1}$, averaging the standard errors of the slopes for the multiple possible subsamples (Fig. 2).

Higher uncertainty in the slope of the regression of solute concentration over time means that the slope is less likely to be statistically distinguishable from zero. Reducing sampling intensity to bimonthly (12.5% of a weekly effort), for example, resulted in significant declines ($p < 0.05$) in SO_4 concentrations in only four of the eight data subsets (Fig. 2).

The long-term trend in NO_3 shows higher variability than SO_4 due to seasonal variation within years related to snowmelt. The uncertainty for NO_3 concentrations in the slopes of weekly, biweekly, monthly, and bimonthly sampling schemes was 0.03, 0.05, 0.07 and 0.10 $\mu g\,NO_3\,L^{-1}\,yr^{-1}$, respectively (Fig. 2). Declines in NO_3 concentrations were significant for a weekly sampling scheme, for one of two biweekly sampling schemes, for one of four monthly simulations, and for one of eight bimonthly simulations (Fig. 2).

3.2. Detectable difference analysis applied to loon tissue sampling in the Adirondacks

Changes in total Hg concentrations in Adirondack loons would be detectable at $\alpha = 0.05$ if they were at least 38% in adult loon blood and 31% in adult loon feathers (Fig. 3). Thus, if there were reason to expect that loon Hg concentrations would change at the rate of 1% per year, for example, it would be reasonable to repeat the survey in 30–35 years. Alternatively, a change would be detectable at an earlier date by sampling a greater number of lakes. If 140 lakes were sampled, a change of approximately 20% in Hg concentrations in loon blood and feathers could be detected in 10 years.

3.3. Bootstrapping approach applied to tree biomass at Hubbard Brook

Using bootstrapping to describe the relationship between sampling intensity and confidence in the mean, we found that only 25 plots were required to estimate the aboveground biomass of a Hubbard Brook watershed within 10% of the true mean. To achieve a 5% sampling uncertainty, 77 plots were required, and a 1% sampling uncertainty required 194 plots (Fig. 4).

The true mean is known, in this case, because all the trees in each of the 208 grid cells in the watershed are measured, and thus the sampling uncertainty in the bootstrap simulation converges on zero. This bootstrapping approach can also be applied to data sets with a sample, rather than a census, of a population. In these cases, the confidence interval will still be minimal at the maximum number of samples, reflecting the sampling uncertainty in the data set analyzed.

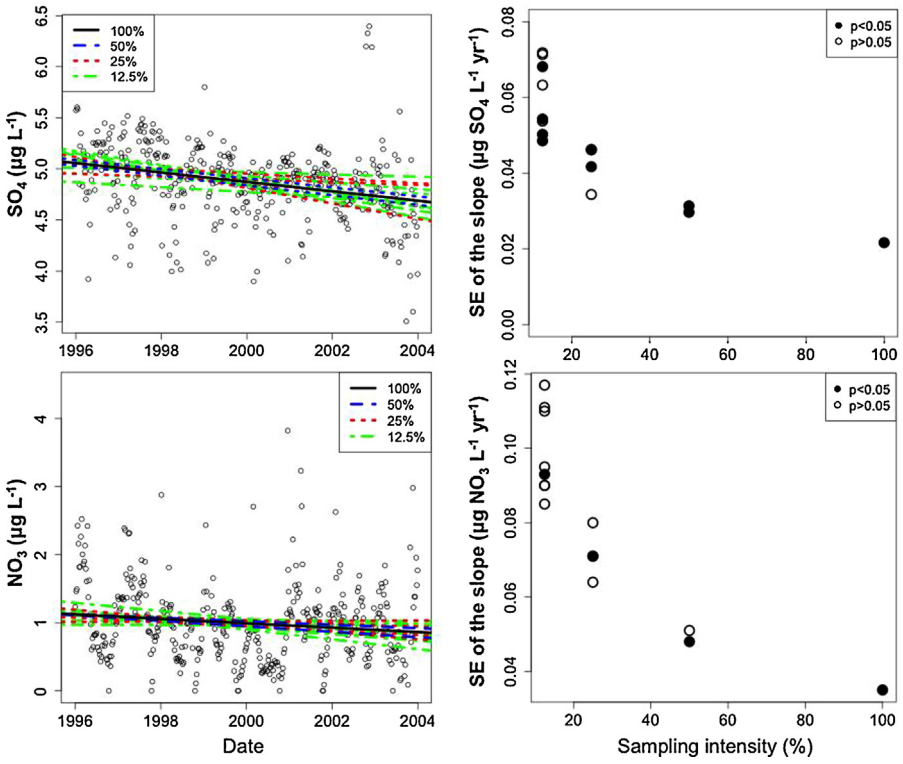

Fig. 2. Panels on the left show the long-term record of SO_4 and NO_3 concentrations at Biscuit Brook, measured weekly from 1996 to 2003. Lines show the regressions based on the full data set (100%), two biweekly subsamples (50%), four monthly subsamples (25%), and 8 bimonthly subsamples (12.5%) of the full data set. Panels on the right show the uncertainty in the rate of change over time for each sampling intensity.

3.4. Repeated-measures mixed effects model applied to lake chemistry in the Adirondacks

In the analysis of NO_3 and SO_4 concentrations in Adirondack lakes, we found that the frequency of sampling could be reduced from monthly to annual with little loss in confidence in quantifying long-term trends (Fig. 5). However, monitoring fewer lakes increased uncertainty in the concentration estimates. Reducing monitoring effort to 90% of lakes increased the uncertainty in estimates of concentration trends by 23% for SO_4 and 46% for NO_3, and the uncertainty increased further as the sampling intensity decreased (Table 2). This increase in uncertainty when lakes are subsampled is not surprising because the spatial variation among lakes is much greater than the temporal variation among months over the sampling period (1992–2010, Table 2).

Sampling only one month per year could be a viable monitoring option. These model simulations had low standard errors due to the low variation in concentrations among months relative to the variation observed among lakes. For Adirondack lakes, many solutes have seasonal variation and there may be important reasons for characterizing this variation. For example, ANC tends to be lowest in the spring coinciding with in the release of inorganic monomeric aluminum in lakes with low ANC values. In such cases, detecting changes in seasonal patterns may be important and sampling seasonally (once in every three months) might be worthwhile.

4. Discussion

Most monitoring designs were not developed based on formal statistical uncertainty analyses. As a result, researchers and managers resist scaling back their monitoring efforts for fear of losing valuable information. Identifying the sources of uncertainty in any estimate can provide a basis for prioritizing possible improvements.

There are several reasons why monitoring programs tend to continue as they were executed historically, despite changes in research objectives and environmental conditions over time. Monitoring programs are not always designed to be flexible, though this is an important characteristic of effective monitoring programs (Lindenmayer and Likens, 2009; Lovett et al., 2007). Researchers may fear that erratic or inconsistent sampling will limit approaches that can be used to analyze the data collected. Fortunately, currently available statistical methods can accommodate irregular data sets.

Another argument for perpetuating intensive sampling schemes is that unlike many research products, the application of monitoring data is not always known. Thus, there is a tendency to collect as much data as possible in case it should prove necessary for future analyses. Fortunately, the longer a record is maintained, the more power is available to detect changes following a reduction in sampling effort.

In this study, we simulated reduced sampling designs using data from studies with intensive long-term sampling records. This is not the same as reducing sampling after a period of more intensive sampling. When sampling is reduced, there is statistical power remaining from the previous record that would contribute to future assessments. When considering reducing monitoring programs, simulated data sets based on projections of future sampling scenarios should be used to assess whether it is possible to reduce sampling and still have enough power to detect trends in the future (Garman et al., 2012).

4.1. Analytical approaches illustrated in these case studies

General linear models can be used to assess the standard error of the slope for data sets that include time series at a single site. These tests are commonly used to assess trends over time. Although we illustrate an analysis that used linear regressions, the same approach is applicable to nonlinear regressions.

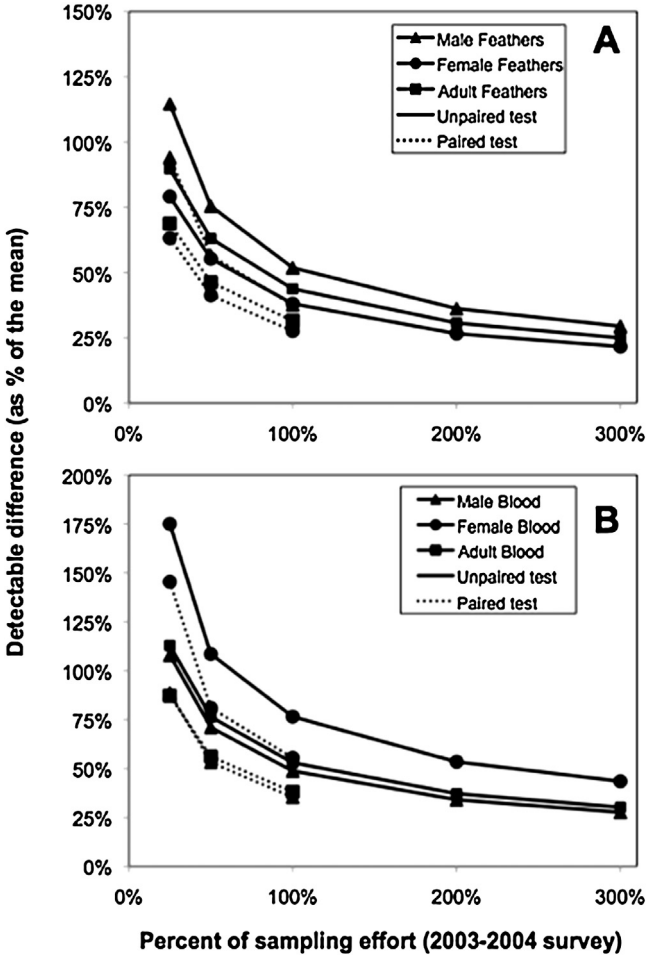

Fig. 3. A. Detectable difference of total Hg (ppm) in feathers of male loons (n = 21 lakes), female loons (n = 22 lakes), and adult loons (male and female combined; n = 30 lakes). B. Detectable difference of total Hg (ppm) in blood of male loons (n = 21 lakes), female loons (n = 21 lakes), and adult loons (male and female combined; n = 40 lakes). Solid lines indicate unpaired resampling and dotted lines indicate paired resampling.

When analyzing data sets that include multiple sites at a single point in time, such as one-time survey data, a two-sample t-test is appropriate for determining a detectable difference when data pass tests for normality and equal variance. In the case of a non-normal distribution, a non-parametric subsampling approach should be used. This type of analysis can be used to guide decisions about future monitoring efforts by indicating the number of samples needed to detect a change in the mean of a specified magnitude. For example, if a large number of samples are needed to detect a significant change within a short time period, it may be desirable to wait to resample until larger change is expected, rather than increasing the sampling effort.

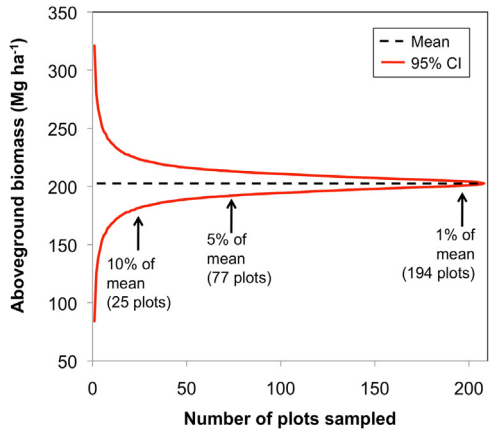

Fig. 4. Uncertainty in estimates of aboveground biomass for Hubbard Brook W6 in 2002 as a function of the number of plots sampled. Estimates are based on 1000 random samples of the number of plots indicated, stratified by elevation as shown in the map (Fig. 1).

We used a bootstrapping approach to subsample vegetation plots at Hubbard Brook and lake chemistry in the Adirondacks in the two examples described above. In these cases, we chose to sample without replacement because this represents a realistic example of a reduced sampling scheme when a proposed reduced scheme would measure only those sampling units (e.g., plots, lakes) that are monitored under the current program. When a proposed change would include sampling units that have not previously been monitored, sampling with replacement could be used, assuming that the variation of the original sample is representative of future sampling units.

Repeated-measures mixed-effects models provide a model error associated with a group of sites over time. In mixed models, it can be difficult to identify which effects are fixed and which are random. In this example, site was a random effect because sites included in the study were chosen from a larger population of potential sites, and thus are considered random. In a study in which all lakes greater than 0.5 ha in the Adirondack Park were sampled (Driscoll and Van Dreason, 1993), site could be considered a fixed effect. Typically though, site will be a random effect because it is rare to sample an entire population of sites when monitoring environmental variables. In some monitoring programs, sites may be chosen for particular reasons; in that case, it may not be appropriate to treat them as random. Applying the proper correlation structure to the model is also important. The correlation structure should be specified based on the sampling scheme. In cases in which samples are not uniformly spaced (for example, sampling hydrologic events), the autoregressive structure should be specified in the model.

4.2. Additional statistical approaches

In this paper, we demonstrate the use of a few relatively simple tests that are widely applicable and easily implemented. However, ecological monitoring often produces unbalanced data sets because equipment fails, funding is inconsistent, and complex sampling schemes are used. In these cases, more sophisticated non-parametric models may be appropriate. Due to the variety of available statistical approaches, there should be few limitations to evaluating monitoring schemes in terms of the suitability of the data structure for analysis.

Parametric models require assumptions about the probability distribution of the observations. Non-parametric analyses such as bootstrapping should be considered when data are not normally distributed. However, they may not be able to detect changes as small as those detected by the analogous parametric tests (Lawrence et al., 2004).

Hierarchical models are well suited for assessing environmental monitoring approaches because they can account for both measurement error and natural variation (Royle and Dorazio, 2008). These models can accommodate nested random effects to account for correlations in measurements, and can also account for irregular sampling schemes by treating the timing of sampling as a covariate (Ver Hoef and Frost, 2003). This is applicable to monitoring individual plants or animals when probability of detection is relevant. Though there remain challenges to the practical implementation of hierarchical models for complex data sets, these models can be a powerful tool for analyzing ecological data (Cressie et al., 2009).

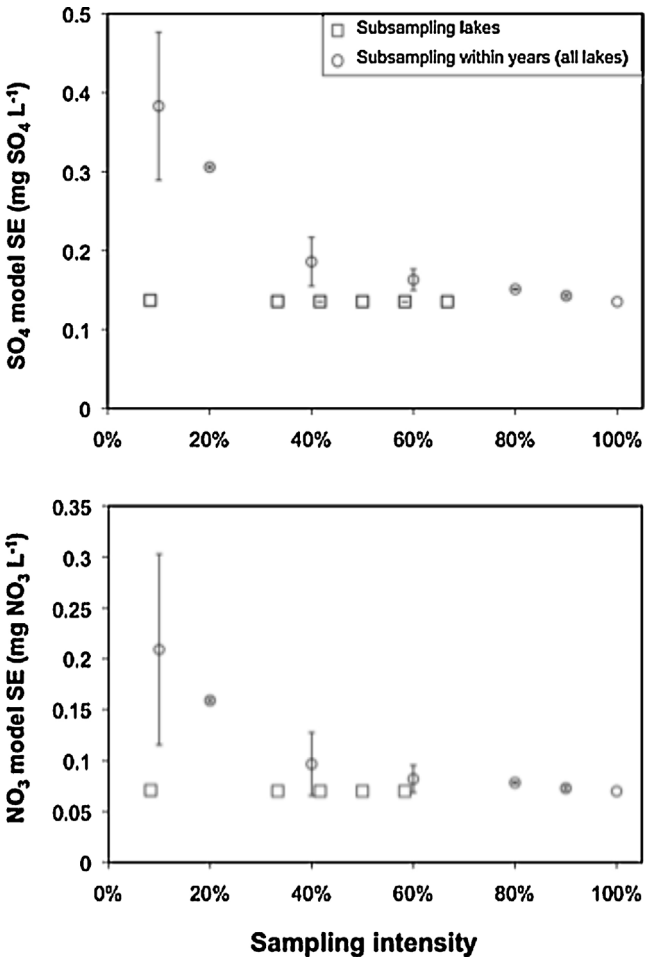

Fig. 5. Model standard error of long-term average concentrations of SO_4 (mg SO_4 L^{-1}) and NO_3 (mg NO_3 L^{-1}) based on a repeated-measures mixed-effects model using 500 random iterations for each simulated subsample size.

The availability of tools for the statistical analysis of complex data means that long-term monitoring schemes need not follow balanced, static designs. We recommend approaches that sample at mixed frequencies, which can incorporate temporal and spatial sampling after an initial monitoring period. The results of our mixed-model analysis of lake sampling demonstrated that there was much greater variability between lakes than there was within individual lakes each year. Thus an improved sampling scheme might decrease the number of times each lake is sampled, and add lakes in future sampling. Since high-frequency data might be useful in the future, a mixed sampling scheme could include monthly samples at a subset of lakes to capture seasonal patterns.

4.3. Advice for revising monitoring programs

Often, monitoring programs have multiple objectives that need to be considered collectively in decisions to alter the temporal or spatial extent of measurements. For example, many monitoring programs track multiple variables in the samples collected. Commonly, the sampling intensity will be consistent across these variables, although the requirements for detecting long-term trends will be different. The appropriate interval for sampling will depend on many factors, including cost, expected rates of change, desired confidence intervals, and the importance of the variables being monitored.

Identifying possible reductions in sampling intensity may provide opportunities to decrease expenses or redirect monitoring efforts to meet other needs. For example, in the long-term lake study we analyzed, characterizing variation across lakes was more important than characterizing variation within years; thus, an improved sampling regime might include additional lakes while decreasing the temporal sampling within lakes.

A comparison of sources of uncertainty can be useful for identifying sampling priorities. It may not pay off to minimize sampling uncertainty if there are other sources that contribute greater uncertainty to the overall estimate. For example, uncertainty in biomass estimates at Hubbard Brook includes the error in the allometric equations. Because the error in the allometric equations are approximately 5% (Yanai et al., 2012), sampling every plot to achieve a 0% spatial error does not improve the overall estimate and requires a large effort.

Regular evaluation of long-term monitoring programs is an important part of the monitoring process (Lindenmayer and Likens, 2009). Uncertainty analysis can help provide a basis for making difficult decisions about competing demands for labor or other costs. Of course, there is a cost associated with evaluating a monitoring program and implementing changes, and this cost should be balanced with the potential savings that could accrue from modifying the approach. The appropriate frequency for evaluation will be shorter for high-frequency monitoring programs (perhaps every five years) than for those conducted less frequently (perhaps every few sampling rounds). This allows for sufficient data to be collected to support an analysis, but would also minimize the cost of oversampling for too long a time.

Acknowledgements

Support for this study was provided by the New York State Energy Research and Development Authority. We appreciate the support of the Adirondack Lakes Survey Corporation and the New York State Department of Environmental Conservation in providing lake chemistry data from the Adirondack Long Term Monitoring program. We thank Kevin Civerolo, Alan Domaracki, and Gary Lovett for providing valuable feedback on an earlier version of this work and Eddie Bevilacqua for advising us on mixed model analyses. Thanks also to the Journal of Forestry for permission to reprint the map of Hubbard Brook W6. The Hubbard Brook Experimental Forest is operated and maintained by the USDA Forest Service, Northern Research Station, Newtown Square, PA. Hubbard Brook is part of the Long-Term Ecological Research (LTER) network, which is supported by the National Science Foundation. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

References

Battles, J.J., Fahey, T.J., Driscoll, C.T., Blum, J.D., Johnson, C.E., 2013. Restoring soil calcium reverses forest decline. Environ. Sci. Technol. Lett., http://dx.doi.org/10.1021/ez400033.

Civerolo, K.L., Roy, K.M., Stoddard, J.L., Sistla, G., 2011. A comparison of the temporally integrated monitoring of ecosystems and Adirondack long-term monitoring programs in the Adirondack Mountain region of New York. Water Air Soil Pollut. 222 (1–4), 285–296.

Cressie, N., Calder, C.A., Clark, J.S., Ver Hoef, J.M., Wikle, C.K., 2009. Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling. Ecol. Appl. 19 (3), 553–570.

Driscoll, C.T., Van Dreason, R., 1993. Seasonal and long-term temporal patterns in the chemistry of Adirondack lakes. Water Air Soil Pollut. 67, 319–344.

Driscoll, C.T., Driscoll, K.M., Mitchell, M.J., Raynal, D.J., 2003a. Effects of acidic deposition on forest and aquatic ecosystems in New York State. Environ. Pollut. 123, 327–336.

Driscoll, C.T., Driscoll, K.M., Roy, K.M., Mitchell, M.J., 2003b. Chemical response of lakes in the Adirondack Region of New York to declines in acidic deposition. Environ. Sci. Technol. 37 (10), 2036–2042.

Driscoll, C.T., Han, Y.-J., Chen, C.Y., Evers, D.C., Lambert, K.F., Holsen, T.M., Kamman, N.C., Munson, R.K., 2007. Mercury contamination in forest and freshwater ecosystems in the Northeastern United States. Bioscience 57 (1), 17–28.

Fahey, T.J., Knapp, A.K. (Eds.), 2007. Principles and Standards for Measuring Primary Production. Oxford University Press, New York, NY, USA.
Garman, S.L., Schweiger, E.W., Manier, D.J., 2012. Simulating future uncertainty to guide the selection of survey designers for long-term monitoring. In: Gitzen, R.A., Millspaugh, J.J., Cooper, A.B., Licht, D.S. (Eds.), Design and Analysis of Long-term Ecological Monitoring Studies. Cambridge University Press, New York, NY, USA, pp. 228–252.
Horsley, S.B., Long, R.P., Bailey, S.W., Hall, T.J., 2000. Factors associated with the decline of sugar maple on the Allegheny Plateau. Can. J. For. Res. 30, 1365–1378.
Kerr, J.G., Eimers, M.C., Creed, I.F., Adams, M.B., Beall, F., Burns, D., Campbell, J.L., Christopher, S.F., Clair, T.A., Courchesne, F., Duchesne, L., Fernandez, I., Houle, D., Jeffries, D.S., Likens, G.E., Mitchell, M.J., Shanley, J., Yao, H., 2012. The effect of seasonal drying on sulphate dynamics in streams across southeastern Canada and the northeastern USA. Biogeochemistry 111, 393–409.
Lawrence, G.B., Momen, B., Roy, K.M., 2004. Use of stream chemistry for monitoring acidic deposition effects in the Adirondack Region of New York. J. Environ. Qual. 33, 1002–1009.
Lenth, R.V., 2001. Some practical guidelines for effective sample size determination. Am. Stat. 55, 187–193.
Lindenmayer, D.B., Likens, G.E., 2009. Adaptive monitoring: a new paradigm for long-term research and monitoring. Trends Ecol. Evol. 24 (9), 482–486.
Lovett, G.M., Burns, D.A., Driscoll, C.T., Jenkins, J.C., Mitchell, M.J., Rustad, L., Shanley, J.B., Likens, G.E., Haeuber, R., 2007. Who needs environmental monitoring? Front. Ecol. Environ. 5 (5), 253–260.
McDonald, T., 2012. Spatial sampling designs for long-term ecological monitoring. In: Gitzen, R.A., Millspaugh, J.J., Cooper, A.B., Licht, D.S. (Eds.), Design and Analysis of Long-term Ecological Monitoring Studies. Cambridge University Press, New York, NY, USA, pp. 101–125.
McHale, M.R., Siemion, J., 2010. US Geological Survey Catskill/Delaware Water-quality Network: Water-quality Report Water Year 2006: U.S. Geological Survey Data Series 497., pp. 36.
Murdoch, P.S., Shanley, J.B., 2006. Detection of water quality trends at high, median, and low flow in a Catskill Mountain stream, New York, through a new statistical method. Water Resour. Res. 42 (8), W08407.
R Core Team, 2008. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. SAS Institute Inc., Cary, NC, USA http://www.R-project.org
Raciti, S.M., Fahey, T.J., Thomas, R.Q., Woodbury, P.B., Driscoll, C.T., Carranti, F.J., Foster, D.R., Gwyther, P.S., Hall, B.R., Hamburg, S.P., Jenkins, J.C., Neill, C., Peery, B.W., Quigley, E.E., Sherman, R., Vadeboncoeur, M.A., Weinstein, D.A., Wilson, G., 2012. Local-scale carbon budgets and mitigation opportunities for the northeastern United States. BioScience 62, 23–38.
Roy, K.M., Dukett, J., Houck, N., Lawrence, G., 2011. A long-term monitoring program for evaluating changes in water quality in selected Adirondack waters: Core Program 2007-2011, New York State Energy Research and Development Authority Report 10-21. NYSERDA, Albany, NY, 51 pp.
Royle, J.A., Dorazio, R.M., 2008. Hierarchical Modeling and Inference in Ecology: the Analysis of Data from Populations, Metapopulations, and Communities. Academic Press, San Diego, CA.
SAS Institute, 2008. Version 9.2. SAS Institute, Cary, NC, USA.
Schoch, N., Glennon, M., Evers, D., Duron, M., Jackson, A., Driscoll, C.T., Yu, X., Simonin, H., 2011. Long-term monitoring and assessment of mercury based on integrated sampling efforts using the Common Loon, prey fish, water, and sediment, New York State Energy Research and Development Authority Report 12-06. NYSERDA, Albany, NY, 116 pp.
Siccama, T.G., Bormann, F.H., Likens, G.E., 1970. The Hubbard Brook ecosystem study: productivity, nutrients, and phytosociology of the herbaceous layer. Ecol. Monogr. 40 (4), 389–402.
Simonin, H.A., Loukmas, J.J., Skinner, L.C., Roy, K.M., 2006. Lake variability: key factors controlling mercury concentrations in New York State fish. Environ. Pollut. 154, 107–115.
Smith, W.B., 2002. Forest inventory and analysis: a national inventory and monitoring program. Environ. Pollut. 116 (1), S233–S242.
Stupak, I., Asikainen, A., Jonsell, M., Karltun, E., Lunnan, A., Mizaraite, D., Pasanen, K., Parn, H., Raulund-Rasmussen, K., Roser, D., Schroeder, M., Varnagiryte, I., Vilkriste, L., Callesen, I., Clarke, N., Gaitnieks, T., Ingerslev, M., Mandre, M., Ozolincius, R., Saarsalmi, A., Armolaitis, K., Helmisaari, H.S., Indriksons, A., Kairiukstis, L., Katzensteiner, K., Kukkola, M., Ots, K., Ravn, H.P., Tamminen, P., 2007. Sustainable utilisation of forest biomass for energy – possibilities and problems: policy, legislation, certification, and recommendations and guidelines in the Nordic, Baltic, and other European countries. Biomass Bioenergy 31, 666–684.
Turnquist, M.A., Driscoll, C.T., Schulz, K.L., Schlaepfer, M.A., 2011. Mercury concentrations in snapping turtles (*Chelydra serpentina*) correlate with environmental and landscape characteristics. Ecotoxicology 20 (7), 1599–1608.
Urquhart, N.S., 2012. The role of monitoring design in detecting trend in long-term ecological monitoring studies. In: Gitzen, R.A., Millspaugh, J.J., Cooper, A.B., Licht, D.S. (Eds.), Design and Analysis of Long-term Ecological Monitoring Studies. Cambridge University Press, New York, NY, USA, pp. 151–173.
Ver Hoef, J.M., Frost, K., 2003. A Bayesian hierarchical model for monitoring harbor seal changes in Prince William Sound, Alaska. Environ. Ecol. Stat. 10, 201–209.
Woodbury, P.B., Smith, J.E., Heath, L.S., 2007. Carbon sequestration in the U.S. forest sector from 1990 to 2010. For. Ecol. Manage. 241, 14–27.
Yanai, R.D., Stehman, S.V., Arthur, M.A., Prescott, C.E., Friedland, A.J., Siccama, T.G., Binkley, D., 2003. Detecting change in forest floor carbon. Soil Sci. Soc. Am. J. 67, 1583–1593.
Yanai, R.D., Levine, C.R., Campbell, J.L., Green, M.B., 2012. Ecosystem budgets have no error: a progress report on quantifying uncertainty in forest ecosystem studies. J. Forest. 110 (8), 448–456.
Yu, X., Driscoll, C.T., Montesdeoca, M., Evers, D., Duron, M., Williams, K., Schoch, N., Kamman, N.C., 2011. Spatial patterns of mercury in biota of Adirondack, New York lakes. Ecotoxicology 20, 1543–1554.
Zar, J.H., 1996. Biostatistical Analysis, 3rd ed. Prentice Hall, Upper Saddle River, NJ.