esa

ECOSPHERE

SPECIAL FEATURE: UNCERTAINTY ANALYSIS

Uncertainty in the net hydrologic flux of calcium in a paired-watershed harvesting study

John L. Campbell,[1,†] Ruth D. Yanai,[2] Mark B. Green,[1,3] Gene E. Likens,[4,5] Craig R. See,[2] Amey S. Bailey,[1] Donald C. Buso,[4] and Daqing Yang[6]

[1]*Northern Research Station, USDA Forest Service, Durham, New Hampshire 03824 USA*
[2]*Department of Forest and Natural Resources Management, State University of New York – College of Environmental Science and Forestry, Syracuse, New York 13210 USA*
[3]*Center for the Environment, Plymouth State University, Plymouth, New Hampshire 03264 USA*
[4]*Cary Institute of Ecosystem Studies, Millbrook, New York 12545 USA*
[5]*Department of Ecology and Evolutionary Biology, University of Connecticut, Storrs, Connecticut 06269 USA*
[6]*National Hydrology Research Center, Environment Canada, Saskatoon, Saskatchewan S7N 3H5 Canada*

Citation: Campbell, J. L., R. D. Yanai, M. B. Green, G. E. Likens, C. R. See, A. S. Bailey, D. C. Buso, and D. Yang. 2016. Uncertainty in the net hydrologic flux of calcium in a paired-watershed harvesting study. Ecosphere 7(6):e01299. 10.1002/ecs2.1299

Abstract. Monitoring solutes in precipitation inputs and stream water exports at small watersheds has greatly advanced our understanding of biogeochemical cycling. Surprisingly, although inputs to and outputs from ecosystems are instrumental to understanding sources and sinks of nutrients and other elements, uncertainty in these fluxes is rarely reported in ecosystem budgets. We illustrate error propagation in input–output budgets by comparing the net hydrologic flux of Ca in a harvested and reference watershed at the Hubbard Brook Experimental Forest, New Hampshire. We identify sources of uncertainty and use a Monte Carlo approach to combine many sources of uncertainty to produce an estimate of overall uncertainty. Sources of uncertainty in precipitation inputs included in this study were: rain gage efficiency (undercatch or overcatch), gaps in measurements of precipitation volume, selection of a model for interpolating among rain gages, unusable precipitation chemistry, and chemical analysis. Sources of uncertainty in stream water outputs were: stage height–discharge relationship, watershed area, gaps in the stream flow record, chemical analysis, and the selection of a method for flux calculation. The annual net hydrologic flux of Ca in the harvested and reference watersheds was calculated from 1973 through 2009. Relative to the reference watershed, the harvested watershed showed a marked increase in Ca flux after it was cut in 1983–1984, and slowly declined toward pretreatment levels thereafter. In 2009, the last year evaluated, the 95% confidence intervals for the annual estimates approach the 95% confidence intervals of the pretreatment regression line, suggesting that the increased net loss of Ca in the harvested watershed may soon be indistinguishable from the reference. Identifying the greatest sources of uncertainty can be used to guide improvements, for example in reducing instances of unusable precipitation chemistry and gaps in stream runoff. Our results highlight the value of estimating uncertainty in watershed studies, including those in which replication is impractical.

Key words: calcium; error analysis; Hubbard Brook Experimental Forest; Monte Carlo; precipitation; Special Feature: Uncertainty Analysis; stream water; uncertainty; watershed; whole-tree harvest.

Received 10 June 2015; revised 7 October 2015; accepted 6 November 2015. Corresponding Editor: J. Taylor.

† **E-mail:** jlcampbell@fs.fed.us

Introduction

Small watersheds are hydrologically distinct landscape units that integrate chemical, physical, and biological processes. The basic premise of the small watershed approach is that the flux of water and chemical elements in and out of headwater catchments can be used to evaluate net gains or losses (Bormann and Likens 1967, Likens 2013). Precipitation inputs are calculated as the product of the chemical concentration and volume of precipitation. Outputs are typically calculated as the product of the chemical concentration and stream runoff at the watershed outlet. The net hydrologic flux is the difference between precipitation inputs and stream outputs. For some elements, other fluxes are needed to close the budget, such as gas exchange and chemical weathering (Likens 2013). Quantification of internal element pools (e.g., biomass, soil) and fluxes (e.g., biotic uptake, litterfall, decomposition) can provide additional information and insight for interpreting inputs and outputs.

Surprisingly, although ecosystem inputs and outputs are used to describe sources and sinks of nutrients and other elements, uncertainty in these fluxes is rarely reported in ecosystem budgets (Harmel et al. 2006, 2009). This omission stems in part from the fact that each ecosystem is unique, making it challenging to identify replicate sampling units. In paired-watershed experiments (e.g., Bates and Henry 1928, Bosch and Hewlett 1982, Likens 1985), a stream draining an experimentally manipulated watershed is compared to a stream from a nearby unmanipulated reference watershed; the treatment is not replicated. Even if replication were possible (Likens 1985), monitoring the number of watersheds required for an acceptable level of confidence may be prohibitively expensive. Without replication, it is still possible to describe the uncertainty in the measurements, but multiple sources of uncertainty are involved and the calculations are complex. Methods and tools for quantifying uncertainty are becoming more available and there is a growing recognition of the importance of including estimates of uncertainty at the ecosystem scale of measurement (Beven 2006, Pappenberger and Beven 2006, Rode and Suhr 2007, Harmel et al. 2009, Yanai et al. 2012).

A previous study by Yanai et al. (2015) identified sources of uncertainty in stream solute export at the Hubbard Brook Experimental Forest (HBEF) in New Hampshire, USA, and compared them to other watershed studies at the Coweeta Hydrological Laboratory, North Carolina, USA, and Gomadansan Experimental Forest, Nara Prefecture, Japan. Here, we build on that work by using Monte Carlo methods to combine estimates of individual sources of uncertainty at the HBEF into an overall estimate that includes precipitation as well as stream flow. To our knowledge, this is the first report of uncertainty in net hydrologic flux for a paired-watershed study.

To demonstrate this approach, we evaluate the legacy effects of disturbance on the net hydrologic flux of calcium (Ca) following a whole-tree harvest at the HBEF. Whole-tree harvesting involves the removal of most of the aboveground biomass (boles and branches), raising concerns about nutrient depletion and forest productivity (Likens and Bormann 1974, Thiffault et al. 2011, Vadeboncoeur et al. 2014). In recent years, there has been renewed interest in this management practice due to rising demand for biofuels to meet energy needs. Calcium is an essential nutrient that limits forest productivity at the site (Likens et al. 1998, Battles et al. 2014). The risk of Ca limitation is exacerbated in the northeastern United States, where decades of acidic deposition have depleted essential base cations, such as Ca, from forest soils (e.g., Federer et al. 1989, Likens et al. 1996, Lawrence et al. 1999).

In the year following the experimental whole-tree harvest at the HBEF, the net loss of Ca in the treated watershed was 20 kg Ca ha^{-1} yr^{-1} compared to 5 kg Ca ha^{-1} yr^{-1} in the reference watershed (Fig. 1; Likens et al. 1998). This difference in the Ca flux between watersheds was initially obvious; however, net Ca losses in the treated watershed declined over the subsequent 26 yr, and the fluxes in the two watersheds have been converging over time. We illustrate the application of uncertainty analysis to ecosystem budgets by evaluating whether the difference in net losses of Ca between watersheds is still greater than the uncertainty of the values.

We identified and quantified uncertainty associated with catch efficiency (undercatch or overcatch) of precipitation gages, gaps in precipitation volume, the selection of a model for

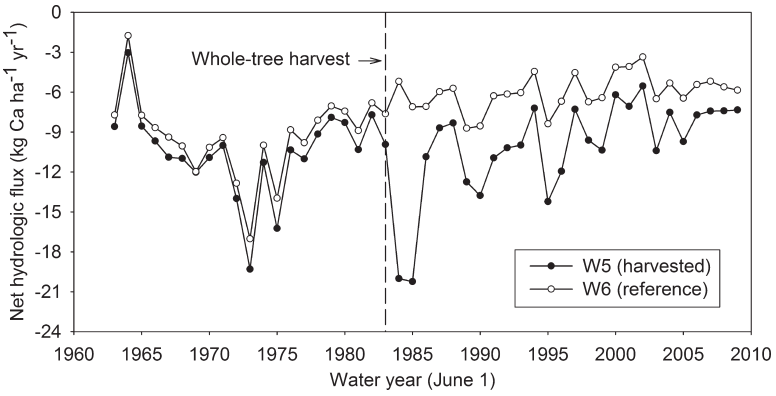

Fig. 1. Net hydrologic flux of Ca (kg Ca ha^{-1} yr^{-1}) for the experimentally harvested watershed (W5) and adjacent reference watershed (W6).

interpolating precipitation gage data, unusable precipitation chemistry, chemical analysis, stream stage height–discharge relationship, watershed area, gaps in the stream flow record, and the selection of a method for stream flux calculation. Additional sources of uncertainty (e.g., sampling frequency, preservation of samples, stage height measurement) have been identified in other studies (Ramsey 1998, Harmel et al. 2006, McMillan et al. 2012), but we focused on those most likely to be important and with sufficient data for analysis. We used a Monte Carlo approach to combine these sources of uncertainty in the net hydrologic flux of Ca at two paired watersheds at the HBEF. The contributions of individual sources of uncertainty were also quantified to identify the importance of each source of uncertainty in our estimates. Recognizing which sources contribute the most uncertainty could help guide improvements to monitoring designs to minimize uncertainty in input–output budgets.

Materials and Methods

Study site and watershed treatment

The Hubbard Brook Experimental Forest is located in the White Mountain National Forest in central New Hampshire, USA (43°56′ N, 71°45′ W). The northern hardwood forest is dominated by sugar maple (*Acer saccharum* Marsh.), American beech (*Fagus grandifolia* Ehrh.), and yellow birch (*Betula alleghaniensis* Britt.), which comprise > 80% of the basal area. At higher elevations, the forest is largely composed of two coniferous species, red spruce (*Picea rubens* Sarg.), and balsam fir (*Abies balsamea* (L.) Mill.), along with the deciduous paper birch (*Betula papyrifera* Marsh.). Soils consist mostly of base-poor Spodosols developed in glacial drift. Depth to bedrock is variable, and is generally 0–3 m in the area of the experimental watersheds (Johnson et al. 2000). Several different whole-watershed manipulations have occurred at the HBEF (Hornbeck et al. 1997, Campbell et al. 2013), including the whole-tree harvest experiment that is the focus of this study.

A commercial whole-tree harvest experiment was conducted in Watershed 5 (W5) at the HBEF to evaluate the effects of intensive harvesting on hydrology, biogeochemistry, and forest productivity (Huntington and Ryan 1990, Dahlgren and Driscoll 1994, Johnson 1995). All trees greater than 5 cm diameter at breast height were cut and removed from the watershed. The lower part of the watershed was harvested between October 1983 and January 1984, and the upper part was harvested the following summer. Approximately 3% of the watershed was not harvested because of inaccessible terrain. An adjacent watershed (W6) remained uncut, and served as a reference for the paired-watershed analysis.

Monte Carlo analysis

The time frame for the uncertainty analysis was 1973–2009 using a June 1 water year (e.g., WY 2004 is from 1 June 2004 through 31 May 2005). Collection and analysis of precipitation

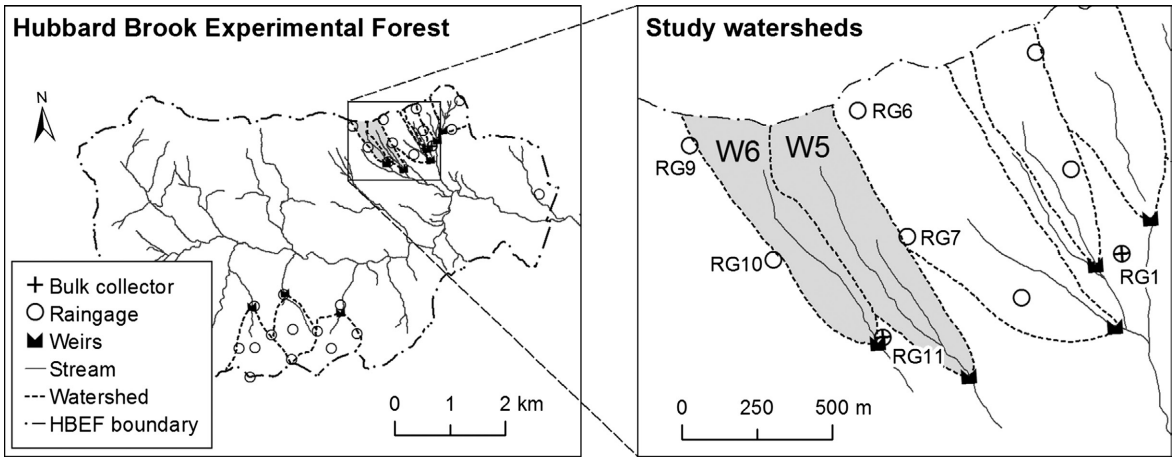

Fig. 2. Map of the Hubbard Brook Experimental Forest and study watersheds showing locations of the harvested (W5) and reference (W6) watersheds, weirs, precipitation gage network, and bulk collectors for measuring precipitation chemistry.

and stream water chemistry began 10 years earlier, with the inception of the Hubbard Brook Ecosystem Study (HBES) in 1963. However, we did not attempt to analyze the uncertainty prior to water WY 1973 because the methods for collecting precipitation chemistry during the first decade of the HBES were slightly different (Buso et al. 2000), which would have added additional complexity to the analysis. The time frame for which uncertainty was estimated included an 11-yr period before the whole-tree harvest (i.e., 1973–1983), during which time both watersheds remained unmanipulated.

The uncertainty analysis was conducted with the Monte Carlo method, which is a formal probabilistic approach to uncertainty analysis that makes repeated estimates by randomly sampling from distributions for each variable used in the calculation (Press et al. 1986). The analyses were performed in the statistical computing language R (v3.2.0), and the documented computer code is included in the archived Appendix S1. In this application, the net hydrologic flux of Ca was calculated using the traditional method at the HBEF, as described in subsequent sections (see also Buso et al. 2000). The uncertainty in these values (i.e., 95% confidence intervals) was then estimated by repeated (10,000 times) random sampling of precipitation and stream water volume and chemistry values from distributions of possible values. Confidence intervals were determined from the 2.5 and 97.5 percentiles of the distribution of possible values, indicating with 95% certainty that the true value falls within this range. With this method, the distribution need not be normally distributed, and the error bars may be asymmetrical. We conducted the Monte Carlo with all the sources of uncertainty combined, and then with only one source of uncertainty at a time to evaluate the relative importance of each source of uncertainty. The following is an explanation of how each source of uncertainty was estimated, including the determination of input distributions for the Monte Carlo.

Uncertainty in precipitation Ca fluxes

Gaps in precipitation volume.—Gaps in the record of precipitation volume are infrequent at the HBEF; however, on rare occasions data can be lost for various reasons, which contributes to the uncertainty in inputs of Ca in precipitation. We used field notes from WY 1973–2009 to identify and characterize gaps in precipitation volume. Across all 24 precipitation gage stations (Rain Gage-RG) at the HBEF (Fig. 2), the proportion of gaps in precipitation volume attributable to technician error was 53%, equipment failure 40%, animal activity 5%, weather-related phenomena 1%, and vandalism 1%. For the precipitation gages used to calculate precipitation at W5 and W6 (i.e., RG6, 7, 9, 10,

11), the average number of gaps was 0.09 per precipitation gage per year, totaling 0.63 d of missing data per precipitation gage per year. For W5, which is calculated based on five precipitation gages (i.e., RG6, 7, 9, 10, 11), there were 9 years (of 37) that contained gaps; for W6, where precipitation volume is calculated based on three precipitation gages (i.e., RG9, 10, 11), there were 7 years that contained gaps. Gaps generally last a week or less because problems are discovered and addressed during routine weekly collections.

When precipitation volume data are missing, the gaps are normally filled in one of two ways. For RG6 and 10, missing values are filled by substituting values from co-located recording precipitation gages. For the other gages, which do not have co-located recording precipitation gages, missing values are filled by averaging the values from several adjacent precipitation gages (RG7 is filled with the mean of RG1, 2, 4, and 8; RG9 with the mean of RG10 and 11; and RG11 with the mean of RG9 and 10; Fig. 2).

To quantify the uncertainty due to filling gaps in precipitation volume for each precipitation gage, we created artificial gaps by randomly sampling 16 week-long periods from WY 1973 through 2009 across all precipitation gages used in the analysis (i.e., RG6, 7, 9, 10, and 11). The number of gaps created (i.e., 16) was determined from the actual number of gaps in the record, and weeks that contained actual gaps were excluded from the pool of potential weeks to sample. For each gap, an error term was added to each day by randomly sampling from a distribution of the differences between the actual daily precipitation and daily precipitation calculated using the gap filling method (Fig. 3a). Days without measureable precipitation were not included in the distribution.

Precipitation gage efficiency.—At the HBEF, precipitation volume is measured weekly with National Weather Service 8-inch (20-cm) standard precipitation gages. Even though these gages are equipped with single Alter wind shields, they are prone to wind-induced catch error (Yang et al. 1998). To estimate the uncertainty associated with precipitation catch efficiency, we used data from an intercomparison of precipitation measurement (Goodison et al. 1998, Yang et al. 1998) conducted by the World Meteorological Organization. This study included a comparison of a National Weather Service 8-inch standard precipitation gage equipped with an Alter shield, such as those used at the HBEF, with a manual Tretyakov gage equipped with an octagonal vertical double fence shield (the Double Fence Intercomparison Reference, DFIR), which provides more accurate measurements of precipitation volume (Yang et al. 1995). Catch efficiency results from the Sleepers River Watershed in Danville, Vermont, USA, were used in this analysis because among sites included in the intercomparison (Goodison et al. 1998), it was the station closest to the HBEF (60 km northwest) and has a similar climate.

Uncertainty due to precipitation gage catch efficiency was evaluated by comparing precipitation volume collected with the 8-inch standard precipitation gage and the DFIR. In this analysis, we used precipitation volume (> 1 mm) recorded for 173 events during December 1986 through April 1992. Separate sampling distributions for rain and snow were generated by calculating the percent difference in precipitation volume between the two precipitation gage types for each event (Fig. 3b). When air temperature at the HBEF was > 0°C, the uncertainty was estimated by multiplying the daily precipitation volume from an individual rain gage by a percentage that was randomly selected from the distribution for rain, and when the air temperature was < 0°C, percentages from the distribution for snow were used.

Spatial interpolation of precipitation volume.—Because precipitation volume is measured at a network of precipitation gages distributed throughout the area of the experimental watersheds, spatial interpolation is required to convert these point measurements to an areal estimate for the watersheds. To quantify the uncertainty due to the selection of a spatial interpolation method, we compared the Thiessen polygon method, which has traditionally been used at the HBEF, with four common interpolation methods: inverse distance weighting, spline, regression, and kriging (Yanai et al. 2012). Based on this comparison, the uncertainty in precipitation volume for each method (expressed as a percent of the Thiessen polygon method) ranged from 99.5 to 100.3% for W5 and 100.5 to 101.6% for W6 (Fig. 3c). For each iteration of the

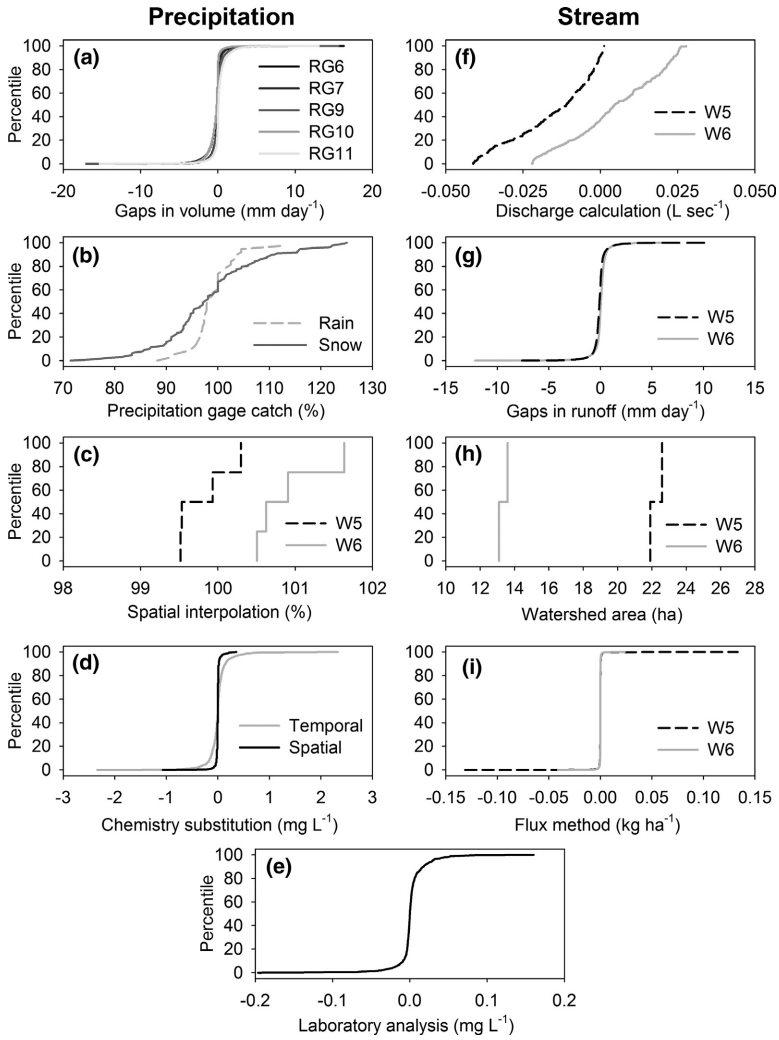

Fig. 3. Distributions for each source of uncertainty in precipitation and stream water that was included in the Monte Carlo analysis.

Monte Carlo, the percent uncertainty from one of these methods was randomly selected and multiplied by the precipitation volume.

Unusable precipitation chemistry.—In addition to uncertainty in precipitation volume, there is also uncertainty in precipitation chemistry, including uncertainty associated with precipitation chemistry samples that are rejected (e.g., because of contamination). The most common reason for unusable precipitation chemistry data is sample contamination due to pollen, insect bodies, bird droppings, and unidentified particulate matter (Buso et al. 2000). Precipitation samples for chemical analyses are collected with bulk collectors mounted 2 m above the ground in areas that are cleared of vegetation (Buso et al. 2000). During the warmer months when precipitation falls mostly as rain, these collectors consist of a high-density polyethylene (HDPE) funnel connected by tubing to a sample bottle. During winter, when precipitation falls mostly as snow, an open-top HDPE bucket is used.

Precipitation samples for chemical analysis have been collected at multiple locations for varying time periods at the HBEF. For WY 1973–2009, data from RG11 are used to calculate the chemical inputs in precipitation to W5 and W6. A nearby precipitation collector, RG1 (Fig. 2),

has been in operation from 1987 to 2010 and is used when data from RG11 are unavailable (referred to as *spatial* infilling in Fig. 3d). When RG1 data are unavailable, missing chemistry values for RG11 are filled using the next uncontaminated RG11 value in the database (referred to as *temporal* infilling in Fig. 3d). We evaluated the error due to this substitution by comparing the Ca concentrations for RG11 and RG1 for dates where both are reported and created a distribution of differences (Fig. 3d). Similarly, we created a distribution of differences between two weekly values for all the weekly values in the data set to evaluate the error associated with the gap filling approach that was used. We sampled randomly from this distribution for the number of occasions per year that samples were substituted in the long-term precipitation chemistry record.

Chemical analysis.—The measurement of Ca concentration is a source of uncertainty in both precipitation and stream fluxes. After water samples were collected, they were shipped to the Cary Institute of Ecosystem Studies in Millbrook, New York for chemical analysis. Precipitation and stream samples from the HBEF have been analyzed for Ca by flame atomic absorption spectrophotometry for the period reported here (i.e., WY 1973 through 2009; Buso et al. 2000). The uncertainty due to chemical analysis was determined by comparing quality control standards of a known concentration with the value measured in the laboratory (Buso et al. 2000, Yanai et al. 2015). We used quality control samples from 1999 to 2011 for concentrations up to 10 mg L^{-1}, which covers the range of observed Ca concentrations in stream water and precipitation at the HBEF. There were 961 observations in this data set, and the differences between the certified and measured concentrations had a mean absolute value of 9.3 μg L^{-1}. An error term for each precipitation and stream chemistry sample was determined by randomly sampling from the distribution of differences between the certified and measured concentrations of quality control standards (Fig. 3e).

Uncertainty in stream Ca fluxes

Stage height–discharge relationship.—Much like the flux of Ca in precipitation, the flux of Ca in stream water is calculated as the product of the volume of stream water and Ca concentration. Stream runoff is measured at a v-notch weir located at the watershed outlet (Fig. 2; Bailey et al. 2003). The stage height (water level) is recorded continuously in a stilling well adjacent to the weir, with a float that is connected to a chart recorder. The charts are digitized (unequal intervals) to produce a continuous record of stage height. The conversion of stream stage height to discharge introduces a source of uncertainty that we estimated by comparing two methods: (1) a theoretical equation for a 90° v-notch weir (Brater et al. 1996); and (2) field measurements that were made by timing the capture of stream water in a large container. This method is impractical at high flow, so the comparison was limited to periods when discharge was less than 1.3 L s^{-1} in W5 and less than 2.5 L s^{-1} in W6. Although stream discharge in W5 was below the 1.3 L s^{-1} threshold on 32% of the days, it comprised only 2% of the runoff. In W6, discharge was below the 2.5 L s^{-1} threshold on 69% of the days and comprised 16% of the runoff. To estimate uncertainty at low flow, we calculated discharge for each stage height value using both methods (i.e., theoretical equation and field measurements) and then randomly sampled a distribution of the differences between methods (Fig. 3f). The individual stage height values were then summed to daily totals.

Gaps in stream runoff measurements.—Gaps in stream runoff data are more common than gaps in precipitation volume data. We evaluated stream flow gaps in W5 and W6 at the HBEF from 1996 through 2009. Prior to 1996, gaps were filled, but not documented, and thus would be difficult to categorize. For the 14-yr period when gaps were well documented, we analyzed the causes of gaps and evaluated the importance of each type of gap weighted by duration. Gaps in stream runoff were caused by debris in the v-notch (22% in W5; 44% in W6), malfunctioning chart recorders (39% in W5; 21% in W6), ice in the v-notch (30% in W5; 25% in W6), weir maintenance and repairs (5% in W5; 7% in W6), and technician error (4% in W5; 3% in W6). Most of the gaps due to malfunctioning chart recorders were attributed to faulty clocks (79% in W5; 68% in W6), which are difficult and time consuming to repair.

We characterized the uncertainty due to stream flow gaps by randomly generating gaps in the daily data based on the number of gaps per year from 1996 to 2009 (670 gaps over 37 yr or 18.1 gaps yr^{-1} for W5; 714 gaps over 37 yr or 19.3 gaps yr^{-1} for W6). The length of the gap was randomly selected from a distribution of actual gap lengths recorded (W5 gap lengths ranged from 1 to 613 h with a mean of 50; W6 gap lengths ranged from 1 to 294 h with a mean of 26). Gaps in the stream flow record were filled using a regression model describing the relationship in discharge between Watersheds 5 and 6. The relationship was established with daily stream flow data from 1996 to 2009 (R^2 = 0.98), excluding dates with missing data. Gaps were filled by randomly sampling from a distribution of the differences between the actual and modeled stream flow (Fig. 3g).

Watershed area.—Watershed area is needed to convert stream discharge (L s^{-1}) to stream flow (mm d^{-1}), which can be compared to the precipitation in units of depth, and can be used to compare stream flow among watersheds of different sizes. Consequently, watershed area is a source of uncertainty in stream runoff. When the watersheds were first established at the HBEF, the boundaries were delineated by a land survey based on field observations of the topographic divides. More recently, watershed boundaries have been mapped with a LIDAR-based digital elevation model (1 m) in a geographic information system (ArcGIS v10.1). The areas estimated for W5 and W6 were 21.9 and 13.6 ha by survey and 22.6 and 13.1 ha by LIDAR, respectively. For each watershed, we calculated stream flow in mm by randomly dividing discharge (L s^{-1}) by one of the two different estimates of watershed area (ground survey vs. LIDAR; Fig. 3h).

Temporal interpolation of stream chemistry.—Many different algorithms have been developed to calculate the flux of elements in stream water (e.g., Swistock et al. 1997, Ullrich and Volk 2010). These algorithms combine the continuous record of stream flow with measurements of chemistry collected at discrete intervals (weekly intervals at the HBEF). Because stream solute flux is calculated as the product of concentration at one point in time and discharge over an interval, there is uncertainty associated with the temporal pattern of the concentration between measurement dates. To quantify the uncertainty due to the algorithm used, we compared two different methods for calculating stream element fluxes. The flux calculation method traditionally used at the HBEF is a period-weighted approach that involves multiplying the chemical concentration (mg L^{-1}) for each sample date by the daily stream flow (mm d^{-1}) for that day. For days between sampling dates, the average of the previous and subsequent concentrations is multiplied by the daily stream flow (Buso et al. 2000). The second approach applied was linear interpolation, which is similar to the method described previously, with the exception that the daily chemical concentration between sampling dates is calculated by linear interpolation rather than using the mean. In the Monte Carlo analysis, uncertainty due to the selection of a flux calculation method was determined by randomly sampling from a distribution of differences between the two methods considered (Fig. 3i).

These two period-weighted flux calculation methods used are similar and tend to be biased if there is a relationship between concentration and stream flow: the calculated flux is too low when there is a positive relative relationship, and too high when there is a negative relationship. In such cases, estimates of chemical concentrations between sampling dates can be improved with regression models that incorporate stream flow and possibly other explanatory variables, such as time of year or turbidity (e.g., Johnson et al. 1969). The composite method for interpolating concentration data has gained popularity in recent years because it is a hybrid technique that uses observed concentrations to improve the concentrations predicted by regression methods (Aulenbach and Hooper 2006). These discharge-based methods can improve estimates of stream concentrations, but were deemed unsuitable for our analysis because of the poor concentration–discharge relationship for Ca at the HBEF (Johnson et al. 1969). Weekly data from the reference watershed (W6) during the period analyzed (WY 1973–2009) showed only a very weak positive relationship between log Ca concentration and log discharge at the time the sample was collected (N = 2195; R^2 = 0.06). Aulenbach et al. (2016) suggest that period-weighted approaches are most appropriate in cases such as ours, in which a concentration–discharge relationship is absent or weak (i.e., model R^2 < ~0.3); therefore, only

period-weighted approaches were considered in our analysis.

Results

The annual precipitation Ca flux was lower and had a narrower range (0.6 to 2.0 kg Ca ha^{-1} yr^{-1} in both watersheds) than the stream water flux (6.6 to 21.3 kg Ca ha^{-1} yr^{-1} in the harvested watershed [W5] and 4.4 to 19.0 kg Ca ha^{-1} yr^{-1} in the reference watershed [W6]). Subtracting annual Ca inputs in precipitation from exports in stream water yields a long-term average net hydrologic loss of 10.4 kg Ca ha^{-1} yr^{-1} in W5 and 7.1 kg Ca ha^{-1} yr^{-1} in W6 for the entire period considered (1973–2009; Table 1). Differences in the net hydrologic flux of Ca between watersheds are due to differences in stream water losses, as annual differences in precipitation inputs of Ca between watersheds were minimal during the 37-yr period (< 0.2 kg Ca ha^{-1} yr^{-1}, relative to outputs of Ca in stream water [ranging from 0.8 to 14.8 kg Ca ha^{-1} yr^{-1}]).

For the 20 years prior to the whole-tree harvest (1963–1982), the annual net hydrologic flux of Ca at W5 was consistently greater than the flux at W6 by an annual average of 1.1 kg Ca ha^{-1} yr^{-1} (Fig. 1). Regression analysis showed that there was a strong linear relationship in the annual net hydrologic flux of Ca between W5 and W6 before the whole-tree harvest ($R^2 = 0.98$, P = < 0.001; Fig. 4). After the whole-tree harvest, stream water Ca losses in the cut watershed increased markedly relative to the reference. The annual difference in Ca export (14.8 kg Ca ha^{-1} yr^{-1}) was greatest the year immediately after the cut was completed (WY 1984) and declined thereafter. The minimum difference after the cut was 1.5 kg Ca ha^{-1} yr^{-1} in the last year evaluated (2009). The average difference in the net hydrologic flux of Ca between the two watersheds was 1.1 kg Ca ha^{-1} yr^{-1} for the 20-yr preharvest period (1963–1982), which was 3.0 kg Ca ha^{-1} yr^{-1} less than the average difference in Ca flux for the 26-yr postharvest period (4.1 kg Ca ha^{-1} yr^{-1} for WY 1984–2009). Assuming a background difference of 1.1 kg Ca ha^{-1} yr^{-1} based on pretreatment data, the harvest resulted in a net loss of 78.2 kg Ca ha^{-1} during the posttreatment period.

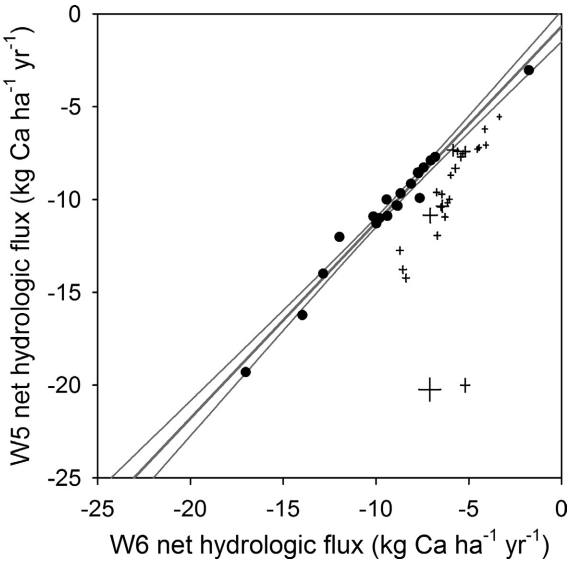

Fig. 4. Annual net hydrologic flux of Ca (kg Ca ha^{-1} yr^{-1}) in the harvested (W5) vs. reference (W6) watersheds. Solid black circles indicate the period before the whole-tree harvest. The regression line for the preharvest period is shown, with thinner lines representing the 95% confidence intervals for the regression. Crossed bars indicate the posttreatment period with the size of the bars showing the 95% confidence intervals determined with the Monte Carlo-based uncertainty analysis.

Table 1. The long-term (WY 1973–2009) average annual precipitation, stream water, and net hydrologic flux of Ca (kg Ca ha^{-1} yr^{-1}), 95% confidence interval surrounding the flux (kg Ca ha^{-1} yr^{-1}), and coefficient of variation (standard deviation divided by the mean) for Watershed 5 and 6.

	Watershed 5			Watershed 6		
Source	Flux	95% CI	CV (%)	Flux	95% CI	CV (%)
Precipitation	1.1	0.4	10.1	1.1	0.4	10.0
Stream water	11.5	0.3	0.6	8.2	0.2	0.6
Net	−10.4	0.5	1.3	−7.1	0.5	1.8

To characterize uncertainty in the net hydrologic flux of Ca, we identified and quantified five sources of uncertainty in precipitation inputs: precipitation gage efficiency (undercatch or overcatch), gaps in precipitation volume, selection of a model for interpolating among precipitation gages, unusable precipitation chemistry, and chemical analysis. During the period for which uncertainty was evaluated (i.e., WY 1973–2009), the combined uncertainty (i.e., 95% confidence intervals) in the annual precipitation flux ranged from 0.2 (WY 2002) to 1.4 (WY 1977) with a long-term mean of 0.4 kg Ca ha^{-1} yr^{-1} (Table 1). In stream water, we also identified and quantified five sources of uncertainty: stage height–discharge relationship, watershed area, gaps in the stream flow record, chemical analysis, and the selection of a method for flux calculation. Based on this evaluation for the same period as precipitation (WY 1973–2009), combined uncertainty in the stream flux ranged from 0.2 (WY 2009) to 0.6 (WY 1984) with a mean of 0.3 kg Ca ha^{-1} yr^{-1} for W5 and 0.1 (WY 2002) to 0.3 (WY 1973) with a mean of 0.2 kg Ca ha^{-1} yr^{-1} for W6. The range of uncertainty in the net hydrologic flux was 0.3 (WY 2002) to 1.4 (WY 1977) with a mean of 0.5 kg Ca ha^{-1} yr^{-1} for W5 and 0.2 (WY 2002) to 1.5 (WY 1077) with a mean of 0.5 kg Ca ha^{-1} yr^{-1} for W6. The postharvest differences in the annual net hydrologic flux between watersheds were consistently greater than the uncertainty in the values. In WY 2009, which was the last year evaluated, the 95% confidence intervals for the annual estimates approach the 95% confidence intervals of the pretreatment regression line (Fig. 5), suggesting that the increased net loss of Ca in response to the harvest in W5 may soon no longer be distinguishable from the reference W6.

The individual sources of uncertainty were ranked according to their contribution to the overall uncertainty in the net hydrologic flux (Fig. 5). The ranking of uncertainty sources was the same for W5 and W6. For the precipitation flux, precipitation chemistry substitution was the major source of uncertainty, contributing ~0.30 kg Ca ha^{-1} yr^{-1} in both watersheds. For stream runoff, gaps in the record were the greatest source of uncertainty, contributing 0.21 kg Ca ha^{-1} yr^{-1} in W5 and 0.15 kg Ca ha^{-1} yr^{-1} in W6. Chemical analysis was the second greatest source of uncertainty for both precipitation and stream water, with chemical analysis of precipitation contributing slightly more (0.22 kg Ca ha^{-1} yr^{-1} in W5 and W6) to the uncertainty than the chemical analysis of stream water (0.15 kg Ca ha^{-1} yr^{-1} in W5 and 0.12 kg Ca ha^{-1} yr^{-1} in W6). Some sources of uncertainty that we quantified were negligible. The precipitation interpolation method, gaps in precipitation volume, and stream stage–discharge relationship were relatively minor sources of uncertainty, each comprising less than 0.5% of the sum of all sources of uncertainty.

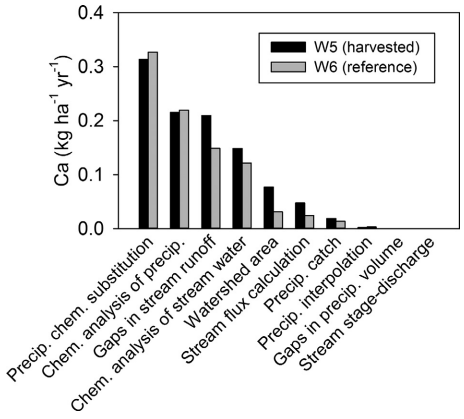

Fig. 5. Contribution of each source of uncertainty (kg Ca ha^{-1} yr^{-1}) to the overall error in the net hydrologic flux of Ca at the harvested (W5) and reference watersheds (W6) shown in order of importance.

Discussion

Benefits of uncertainty analysis

Estimating uncertainty is essential to evaluating the significance of results from cross-site comparisons, as we need to distinguish fundamental differences in ecosystem function from differences due to sampling protocols or computational methods. Even within sites, estimates of uncertainty are needed to evaluate effects with confidence. Without estimates of uncertainty it may be difficult to determine whether responses are due solely to the experimental manipulation or are the result of some other influence (e.g., natural variability). Whole-watershed experiments are normally unreplicated because of the insurmountable challenge in finding suitable replicates and expense of watershed-scale manipulations. Estimates of uncertainty can be used to report

statistical confidence in cases where replication is not possible (e.g., small watershed experiments).

In this study, uncertainty analysis was used to evaluate whether elevated annual net losses of Ca due to whole-tree harvesting are greater than the uncertainty in the values in the years after the cut. The regression approach used (Fig. 4) accounts for pretreatment conditions, which in this case showed higher net losses in W6 compared to W5. Possible causes for the pretreatment difference between watersheds include differences in hydrologic flowpaths, groundwater inputs, mineralogy, or weathering rates (Likens et al. 1998). Results of our analysis indicate that the harvest-induced net Ca loss has lasted for 26 years, but soon may no longer be distinguished from the uncertainty in the measurements. This example demonstrates the usefulness of combining the uncertainty from various sources into a single estimate.

Knowing the individual contributions of different sources of uncertainty is another important benefit of uncertainty analysis because it can be used to determine how best to allocate limited resources to improve data collections and analyses. It should be noted that the sum of the individual uncertainties (Fig. 5) exceeded the estimate of the overall uncertainty in the net hydrologic flux (Table 1) because when the overall estimate of uncertainty is determined with Monte Carlo, a portion of the individual sources of uncertainty cancels out. Thus, the individual sources of uncertainty are not additive when computed independently. The same is true for analytical uncertainty propagation methods.

Uncertainty in precipitation

Based on our analysis, the greatest source of uncertainty was precipitation chemistry substitution (Fig. 5). When a weekly precipitation chemistry sample is missing or deemed unusable because of contamination, it is filled with a value from a nearby precipitation collector. If that sample were also unavailable, the gap would be filled with a value from the subsequent week. The sample from the nearby precipitation collector is typically more similar to the actual value than the sample from the subsequent week, as indicated by the narrower distribution in Fig. 3d. To reduce the uncertainty due to unusable precipitation chemistry data, co-located collectors could be installed so that the missing values are filled with better estimates. Of course the benefit of reduced uncertainty has to be weighed against the cost of collecting and analyzing additional precipitation samples.

Improving the accuracy of the chemical analysis of Ca would also help reduce uncertainty, as this was the second greatest source of uncertainty for both precipitation and stream water fluxes. Stricter laboratory protocols could potentially improve estimates (e.g., more stringent requirements for blanks and standards, running every sample multiple times), but would be labor-intensive and costly. For some analytes, improvements in the methods used for chemical analysis have increased accuracy, as with the change from a persulfate digestion method to high temperature combustion in the case of dissolved organic carbon at the HBEF (Buso et al. 2000).

Uncertainty in precipitation chemistry was greater than the uncertainty in precipitation volume, even though precipitation chemistry is more spatially homogeneous at the HBEF (Buso et al. 2000, Likens et al. 2002). The low uncertainty in precipitation volume is largely the result of the higher spatial density of precipitation gages for volume compared to chemistry collectors. Multiple standard precipitation gages (five in W5 and three in W6) in relatively close proximity are used to estimate precipitation volume, and two of these standard precipitation gages (RG6 and 10) have co-located recording precipitation gages. Thus, the quantity and quality of precipitation gage data used to fill gaps in precipitation volume results in low uncertainty in gap-filled estimates. Additionally, the records from these precipitation gages contain few gaps so imputation is rarely required. The density of precipitation gages for volume also affects the uncertainty associated with the precipitation interpolation method used. The uncertainty due to precipitation interpolation method selection was small (< 0.4% of the total), primarily because the precipitation gage network is spatially dense. Greater uncertainty resulted from interpolation methods in areas distant from precipitation gages.

In addition to uncertainty associated with the precipitation gage network, the individual precipitation gages can over- or underestimate precipitation volume, which is a source of measurement error. Precipitation gages typically

underestimate precipitation volume, causing a systematic bias (Yang et al. 1998). Undercatch is mainly caused by precipitation blowing horizontally over the orifice of the gage and is therefore more of a problem for snowfall than for rain. As expected, the estimated average undercatch for snow in this study (1.8%) was greater than the undercatch for rain (1.2%) and had a broader distribution (Fig. 3b). As undercatch is primarily caused by wind, these values for forested regions of the northeastern United States are much lower than for more open, windier and colder locations where precipitation gages may underestimate water volume by as much as 50% (see Yang et al. 2005, Sieck et al. 2007).

Uncertainty in stream water flux

Unlike precipitation flux, which is calculated from measurements made at multiple locations throughout a watershed, stream water flux is determined from measurements made at the point along a stream that defines the watershed. Consequently, uncertainty due to spatial variability is not an issue with stream fluxes as it is with precipitation fluxes. Instead, uncertainty in estimating stream export derives primarily from temporal variability. Stream stage height is measured continuously, whereas samples for stream chemistry are collected at points in time, requiring some form of interpolation for calculating fluxes. Methods of interpolation range from simple calculations based on the mean of surrounding values, to more complex approaches that utilize concentration–discharge relationships (e.g., Johnson et al. 1969, Cohn et al. 1992, Aulenbach and Hooper 2006). Uncertainty due to the selection of stream flux calculation method in this study was relatively low (4.6% in W5 and 2.7% in W6), indicating that both methods (average and linear interpolation) produce similar results. Stream-flow based flux-calculation methods may be superior in cases where there is a strong concentration–discharge relationship (Aulenbach et al. 2016), but Ca concentrations at the HBEF are fairly constant and not well predicted by discharge (Likens 2013). The difference between regression-based and period-weighted approaches is important for solutes that have a strong relationship with discharge, leading to greater model selection uncertainty (Aulenbach et al. 2016). We should also note that we have not attempted here to quantify the uncertainty incurred by using a particular model, as the period-weighted "models" we used do not allow for comparison of predicted and observed values.

Gaps in stream runoff are one of the largest sources of uncertainty. By analyzing past records, we determined that most of the gaps were caused by organic debris in the v-notch weir and malfunctioning chart recorders. To reduce gaps in stream runoff due to debris, technicians now install floating barriers in the ponding basin during the growing season to keep the v-notch clear of floating leaves and twigs that cause debris dams. Improvements in the instruments used to measure stream flow can also reduce gaps in stream runoff. The chart recorders that have been in use since the inception of the Hubbard Brook Experimental Forest in 1955 were recently replaced with shaft encoders connected to dataloggers, which produce fewer gaps in stream runoff, largely because the digital clocks are more reliable.

The uncertainty in stream runoff can also be reduced with an improved estimate of watershed area, as it is used to normalize discharge. The two watershed delineation methods used in this study were ground survey and automated mapping with a 1-m resolution digital elevation model. Ground survey was the best available method for delineating watersheds at the time they were established. Automated delineation with a LIDAR-based digital elevation model is a newer, alternative method, but it is not clear if it is better than the ground survey. These methods were selected for the uncertainty analysis because they are the best currently available. Including additional methods that are known to be worse would incorrectly inflate the uncertainty. Ground survey and LIDAR are based on surface topography, which may not reflect the true hydrological divides that are belowground and can change with groundwater depth. Therefore, the uncertainty due to watershed area may be greater than what is reported here; however, improved estimates are not yet available. Advances in measurement methodologies, especially those that characterize belowground topography, such as ground penetrating radar, may lead to better future estimates of true hydrologic watershed area.

Uncertainty analysis considerations

The list of sources of uncertainty in this study was not exhaustive, and other sources have been identified, and in some cases quantified, in the literature (e.g., Harmel et al. 2006, Rode and Suhr 2007, McMillan et al. 2012). We were unable to quantify some sources of uncertainty due to lack of data (e.g., water leaking around a weir), and other sources contribute little to the flux estimates and are therefore probably not very important to include (e.g., wetting loss, which refers to precipitation that evaporates from the inner walls of the precipitation gage [Yang et al. 1998]). When all the sources of uncertainty considered were combined in a Monte Carlo framework, our calculations suggest that the resultant overall uncertainty in the estimates of the net hydrologic flux of Ca at the HBEF is quite small (5.1% of the total net hydrologic flux for W5 and 6.9% for W6). Including additional sources in the Monte Carlo analysis would increase the estimates of uncertainty; we hope to make measurements to estimate uncertainty in stage-discharge relationships at high flow.

Uncertainty in net hydrologic flux is likely to be higher for other solutes and other sites. Low uncertainty at the HBEF results in part from the intensity of monitoring, which from the onset was designed to reduce uncertainty. Quality control in laboratory analyses is especially good for Ca at the HBEF, with other solutes having uncertainties two or three times higher (Yanai et al. 2015). Much higher uncertainties, in some cases exceeding 100%, have been reported for other small watershed studies, which may be due to factors such as the solutes evaluated, high flow conditions, and the method of uncertainty analysis, which used the "maximum" values of the component errors (Harmel et al. 2009).

Uncertainty analysis is an imperfect science and there are many different ways to estimate uncertainty. In this example, we used a Monte Carlo approach that involves repeated random sampling to obtain results. Alternatively, uncertainty can be calculated statistically by combining the variances of each variable used in the calculation of the net hydrologic flux. One of the advantages of the Monte Carlo method is that, unlike the analytical approach, it does not require assumptions about the nature of the distributions. In this study, we sampled from the actual distributions (Fig. 3). One of the criticisms of Monte Carlo error propagation is that it can be computationally burdensome, but that is becoming less of an issue as computing resources improve. More sophisticated approaches to uncertainty analysis, such as Bayesian modeling, may be warranted when the added complexity improves results.

Variation in approaches to uncertainty analysis makes comparisons among studies difficult. Different sources of uncertainty are addressed in different studies, and different methods of error propagation are used. For this reason, it is important to explain clearly what sources were included in an analysis and how the uncertainty was calculated. Almost any estimate of uncertainty is better than none, even though every estimate is imperfect. Quantifying uncertainty is essential for evaluating confidence in results and identifying areas for improvement and should be a goal for continuous improvement in ecosystem studies.

Acknowledgments

We thank Mary Martin for help with computer code. John Lombardi enhanced the efficiency of the program, and along with Brent Aulenbach and two anonymous reviewers, provided constructive comments on an earlier draft of this manuscript. This paper is a product of QUEST (Quantifying Uncertainty in Ecosystem Studies, funded as a Research Coordination Network by the National Science Foundation (NSF; DEB-1257906) and Long-Term Ecological Research (LTER) Network Office. QUEST supports uncertainty analysis by sharing papers, presentations, tutorials, and sample code (http://www.quantifyinguncertainty.org). This manuscript is a contribution of the Hubbard Brook Ecosystem Study. Hubbard Brook is part of the LTER network, which is supported by the NSF (DEB-1114804). Calcium data were obtained through funding to G.E. Likens from The A.W. Mellon Foundation and the NSF, including the LTER and Long Term Research in Environmental Biology (DEB-1256696) programs. Hydrologic data were provided by the USDA Forest Service. The Hubbard Brook Experimental Forest is operated and maintained by the USDA Forest Service, Northern Research Station, Newtown Square, Pennsylvania.

Literature Cited

Aulenbach, B. T., and R. P. Hooper. 2006. The composite method: an improved method for stream-water solute load estimation. Hydrological Processes 20:3029–3047.

Aulenbach, B. T., D. A. Burns, J. B. Shanley, R. D. Yanai, K. Bae, A. Wild, Y. Yang, and D. Yi. 2016. Approaches to stream solute load estimation for solutes with varying dynamics from five diverse small watersheds. Ecosphere. http://dx.doi.org/10.1002/ecs2.1298

Bailey, A. S., J. W. Hornbeck, J. L. Campbell, and C. Eagar. 2003. The hydrometeorological database for Hubbard Brook Experimental Forest: 1955–2000. NE GTR-305. USDA Forest Service, Northeastern Research Station, Newtown Square, Pennsylvania, USA.

Bates, C. G., and A. J. Henry. 1928. Forest and streamflow experiment at Wagon Wheel Gap, Colorado, US. Monthly Weather Review, Supplement 30.

Battles, J. J., T. J. Fahey, C. T. Driscoll, J. D. Blum, and C. E. Johnson. 2014. Restoring soil calcium reverses forest decline. Environmental Science & Technology Letters 1:15–19.

Beven, K. 2006. On undermining the science? Hydrological Processes 20:3141–3146.

Bormann, F. H., and G. E. Likens. 1967. Nutrient cycling. Science 155:424–429.

Bosch, J. M., and J. D. Hewlett. 1982. A review of catchment experiments to determine the effect of vegetation changes on water yield and evapotranspiration. Journal of Hydrology 55:3–23.

Brater, E. F., H. W. King, J. E. Lindell, and C. Y. Wei. 1996. Handbook of hydraulics. Seventh edition. McGraw-Hill, Boston, Massachusetts, USA.

Buso, D. C., G. E. Likens, and J. S. Eaton. 2000. Chemistry of precipitation, streamwater, and lakewater from the Hubbard Brook Ecosystem Study: a record of sampling protocols and analytical procedures. NE GTR-275. USDA Forest Service, Northeastern Research Station, Newtown Square, Pennsylvania, USA.

Campbell, J. L., A. S. Bailey, C. Eagar, M. B. Green, and J. J. Battles. 2013. Vegetation treatments and hydrologic responses at the Hubbard Brook Experimental Forest, New Hampshire. Pages 1–9. *in* A. E. Camp, L. C. Irland, and C. J. W. Carroll, editors. Long-term silvicultural & ecological studies: results for science and management (Vol. 2). Research Paper 013. Yale University, Global Institute of Sustainable Forestry, New Haven, Connecticut, USA.

Cohn, T. A., D. L. Caulder, E. J. Gilroy, L. D. Zynjuk, and R. M. Summers. 1992. The validity of a simple statistical model for estimating fluvial constituent loads: an empirical study involving nutrient loads entering Chesapeake Bay. Water Resources Research 28:2353–2363.

Dahlgren, R. A., and C. T. Driscoll. 1994. The effects of whole-tree clear-cutting on soil processes at the Hubbard Brook Experimental Forest, New Hampshire, USA. Plant and Soil 158:239–262.

Federer, C. A., J. W. Hornbeck, L. M. Tritton, C. W. Martin, and R. S. Pirece. 1989. Long-term depletion of calcium and other nutrients in eastern US forests. Environmental Management 13:593–601.

Goodison, B. E., P. Y. T. Louie, and D. Yang. 1998. WMO solid precipitation measurement intercomparison. Instruments and Observing Methods Report No. 67, World Meteorlogical Organization, Geneva, Switzerland.

Harmel, R. D., R. J. Cooper, R. M. Slade, R. L. Haney, and J. G. Arnold. 2006. Cumulative uncertainty in measured streamflow and water quality data for small watersheds. Transactions of ASABE 49:689–701.

Harmel, R. D., D. R. Smith, K. W. King, and R. M. Slade. 2009. Estimating storm discharge and water quality data uncertainty: a software tool for monitoring and modeling applications. Environmental Modelling and Software 24:832–842.

Hornbeck, J. W., C. W. Martin, and C. Eagar. 1997. Summary of water yield experiments at Hubbard Brook Experimental Forest, New Hampshire. Canadian Journal of Forest Research 27:2043–2052.

Huntington, T. G., and D. F. Ryan. 1990. Whole-tree harvesting effects on soil nitrogen and carbon. Forest Ecology and Management 31:193–204.

Johnson, C. E. 1995. Soil nitrogen status eight years after whole-tree clear-cutting. Canadian Journal of Forest Research 25:1346–1355.

Johnson, N. M., G. E. Likens, F. H. Bormann, D. W. Fisher, and R. S. Pierce. 1969. A working model for the variation in stream water chemistry at the Hubbard Brook Experimental Forest, New Hampshire. Water Resources Research 5:1353–1363.

Johnson, C. E., C. T. Driscoll, T. G. Siccama, and G. E. Likens. 2000. Element fluxes and landscape position in a northern hardwood forest watershed ecosystem. Ecosystems 3:159–184.

Lawrence, G. B., M. B. David, G. M. Lovett, P. S. Murdoch, D. A. Burns, J. L. Stoddard, B. P. Baldigo, J. H. Porter, and A. W. Thompson. 1999. Soil calcium status and the response of stream chemistry to changing acidic deposition rates. Ecological Applications 9:1059–1072.

Likens, G. E. 1985. An experimental approach for the study of ecosystems. Journal of Ecology 73:381–396.

Likens, G. E. 2013. Biogeochemistry of a forested ecosystem, Third edition. Springer, New York, New York, USA.

Likens, G. E., and F. H. Bormann. 1974. Effects of forest clearing on the northern hardwood forest ecosystem and its biogeochemistry. Pages 330–335 *in* Proceedings of the first international congress of ecology, the Hague, the Netherlands, September 8–14, 1974. Centre for Agricultural Publishing and Documentation, Wageningen, The Netherlands.

Likens, G. E., C. T. Driscoll, and D. C. Buso. 1996. Long-term effects of acid rain: response and recovery of a forest ecosystem. Science 272:244–246.

Likens, G. E., et al. 1998. The biogeochemistry of calcium at Hubbard Brook. Biogeochemistry 41:89–173.

Likens, G. E., D. C. Buso, and J. W. Hornbeck. 2002. Variation in chemistry of stream water and bulk deposition across the Hubbard Brook Valley, New Hampshire, USA. Internationale Vereinigung fur Theoretische und Angewandte Limnologie Verhandlungen 28:402–409.

McMillan, H., T. Krueger, and J. Freer. 2012. Benchmarking observational uncertainties for hydrology: rainfall, river discharge and water quality. Hydrological Processes 26:4078–4111.

Pappenberger, F., and K. J. Beven. 2006. Ignorance is bliss: or seven reasons not to use uncertainty analysis. Water Resources Research 42:W05302. doi: 10.1029/2005WR004820.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery. 1986. Numerical recipes–the art of scientific computing. Cambridge University Press, New York, New York, USA.

Ramsey, M. H. 1998. Sampling as a source of measurement uncertainty: techniques for quantification and comparison with analytical sources. Journal of Analytical Atomic Spectrometry 13:97–104.

Rode, M., and U. Suhr. 2007. Uncertainties in selected river water quality data. Hydrology and Earth System Sciences 11:863–874.

Sieck, L. C., S. J. Burges, and M. Steiner. 2007. Challenges in obtaining reliable measurements of point rainfall. Water Resources Research 43:W01420. http://dx.doi.org/ 01410.01029/02005WR004519.

Swistock, B. R., P. J. Edwards, F. Wood, and D. R. DeWalle. 1997. Comparison of methods for calculating annual solute exports from six forested Appalachian watersheds. Hydrological Processes 11:655–669.

Thiffault, E., K. D. Hannam, D. Paré, B. D. Titus, P. W. Hazlett, D. G. Maynard, and S. Brais. 2011. Effects of forest biomass harvesting on soil productivity in boreal and temperate forests – A review. Environmental Reviews 19:278–309.

Ullrich, A., and M. Volk. 2010. Influence of different nitrate–N monitoring strategies on load estimation as a base for model calibration and evaluation. Environmental Monitoring and Assessment 171:513–527.

Vadeboncoeur, M. A., S. P. Hamburg, R. D. Yanai, and J. D. Blum. 2014. Rates of sustainable forest harvest depend on rotation length and weathering of soil minerals. Forest Ecology and Management 318:194–205.

Yanai, R. D., C. R. Levine, M. B. Green, and J. L. Campbell. 2012. Quantifying uncertainty in forest nutrient budgets. Journal of Forestry 110:448–456.

Yanai, R. D., et al. 2015. Sources of uncertainty in estimating stream solute export from headwater catchments at three sites. Hydrological Processes 29:1793–1805.

Yang, D., et al. 1995. Accuracy of Tretyakov precipitation gauge: result of WMO Intercomparison. Hydrological Processes 9:877–895.

Yang, D., B. E. Goodison, J. R. Metcalfe, V. S. Golubev, R. Bates, T. Pangburn, and C. L. Hanson. 1998. Accuracy of NWS 8" standard nonrecording precipitation gauge: results and application of WMO intercomparison. Journal of Atmospheric and Oceanic Technology 15:54–68.

Yang, D., D. Kane, Z. Zhang, D. Legates, and B. Goodison. 2005. Bias-corrections of long-term (1973–2004) daily precipitation data over the northern regions. Geophysical Research Letters 32:L19501. http://dx.doi.org/ 19510.11029/12005GL 024057.

Supporting Information

Additional Supporting Information may be found online at: http://onlinelibrary.wiley.com/doi/10.1002/ecs2.1299/supinfo