RESEARCH ARTICLE

WILEY

Estimating uncertainty in streamflow and solute fluxes at the Hubbard Brook Experimental Forest, New Hampshire, USA

Ge Pu[1,2] | John L. Campbell[3] | Mark B. Green[3,4] | Jeff L. Merriam[3] | David Zietlow[3] | Ruth D. Yanai[5]

[1]Department of Environmental Resources Engineering, State University of New York College of Environmental Science and Forestry, Syracuse, New York, USA

[2]Cleveland Water Alliance, Cleveland, Ohio, USA

[3]USDA Forest Service, Northern Research Station, Durham, New Hampshire, USA

[4]Department of Earth, Environmental, and Planetary Sciences, Case Western Reserve University, Cleveland, Ohio, USA

[5]Department of Sustainable Resources Management, State University of New York College of Environmental Science and Forestry, Syracuse, New York, USA

Correspondence
Ruth D. Yanai, State University of New York College of Environmental Science and Forestry, 1 Forestry Drive, Syracuse, NY 13210, USA.
Email: rdyanai@syr.edu

Abstract

Stream fluxes are commonly reported without a complete accounting for uncertainty in the estimates, which makes it difficult to evaluate the significance of findings or to identify where to direct efforts to improve monitoring programs. At the Hubbard Brook Experimental Forest in the White Mountains of New Hampshire, USA, stream flow has been monitored continuously and solute concentrations have been sampled approximately weekly in small, gaged headwater streams since 1963, yet comprehensive uncertainty analyses have not been reported. We propagated uncertainty in the stage height–discharge relationship, watershed area, analytical chemistry, the concentration–discharge relationship used to interpolate solute concentrations, and the streamflow gap-filling procedure to estimate uncertainty for both streamflow and solute fluxes for a recent 6-year period (2013–2018) using a Monte Carlo approach. As a percentage of solute fluxes, uncertainty was highest for NH_4^+ (34%), total dissolved nitrogen (8.8%), NO_3^- (8.1%), and K^+ (7.4%), and lowest for dissolved organic carbon (3.7%), SO_4^{2-} (4.0%), and Mg^{2+} (4.4%). In units of flux, uncertainties were highest for solutes in highest concentration (Si, DOC, SO_4^{2-}, and Na^+) and lowest for those lowest in concentration (H^+ and NH_4^+). Laboratory analysis of solute concentration was a greater source of uncertainty than streamflow for solute flux, with the exception of DOC. Our results suggest that uncertainty in solute fluxes could be reduced with more precise measurements of solute concentrations. Additionally, more discharge measurements during high flows are needed to better characterize the stage-discharge relationship. Quantifying uncertainty in streamflow and element export is important because it allows for determination of significance of differences in fluxes, which can be used to assess watershed response to disturbance and environmental change.

KEYWORDS
gap filling, Hubbard Brook, Monte Carlo, stream fluxes, streamflow, uncertainty analysis

1 | INTRODUCTION

Characterizing fluxes of solutes in stream water at a watershed outlet provides insight on ecosystem processes occurring within the upslope contributing area. Since stream water solute fluxes are a spatially integrated signal, they are useful for characterizing processes such as plant uptake, mineral weathering, and microbially mediated transformations across broad spatial scales (Holloway & Dahlgren, 2001; Tittel et al., 2022). Additionally, solute fluxes in stream water can be compared with fluxes in precipitation to evaluate whether nutrients are

being retained or lost from an ecosystem (Bormann & Likens, 1967). These nutrient budgets aid in assessing impacts from both natural and human disturbances, such as forest harvesting, air pollution and extreme weather events. Understanding and predicting the impacts of disturbances on stream solute fluxes is important to water resource managers in their efforts to meet water quality objectives such as total maximum daily loads (e.g., Lewis et al., 2006).

Replication is commonly used to establish significance of difference, but watershed studies are difficult to replicate, because individual watersheds have unique characteristics. Even where replicate watersheds can be identified, replicated whole-watershed experiments (e.g., harvesting, nutrient addition) may be prohibitively expensive because of their large scale. Instead, in paired watershed experiments, a treated watershed is compared to an untreated reference watershed without replication of the treatment (Bosch & Hewlett, 1982; Neary, 2016). In these cases, other sources of information on uncertainty are needed to evaluate differences between watersheds or the significance of change over time. In the case of paired-watershed experiments, a pre-treatment comparison of the two watersheds is sometimes included (Bosch & Hewlett, 1982; Hornbeck et al., 1993). Quantifying each source of uncertainty and combining them, either through numerical methods involving Monte Carlo error propagation or through analytical methods, is another option. A comprehensive uncertainty analysis has other benefits to monitoring designs, as it can be used to help identify where best to focus efforts, such as determining the optimal strategy of stream chemistry sampling (Levine et al., 2014). Despite its importance, uncertainty analysis has not been widely adopted in watershed studies because of the challenges involved in its quantification (Pappenberger & Beven, 2006).

Stream solute fluxes are calculated as the product of discharge and solute concentration. It is important to collect, analyse and compute the fluxes in a way that ensures the results are of the desired quality. Stream flux data reflect many sources of uncertainty (Campbell et al., 2016; McMillan et al., 2012; Yanai et al., 2015) that should be characterized for proper interpretation of the data. Uncertainty in streamflow estimates have been investigated using linear regression, generalized likelihood uncertainty estimation, Bayesian approaches, and fuzzy methods (Kiang et al., 2018). Few studies have attempted to combine uncertainty in streamflow with that of solute chemistry to produce an overall estimate of uncertainty in solute flux (but see Campbell et al., 2016; Harmel et al., 2006).

The objective of this study was to quantify uncertainty in stream solute fluxes from a small-watershed ecosystem at the Hubbard Brook Experimental Forest in New Hampshire, USA, using a Monte Carlo approach. The work improves on a previous Monte Carlo analysis of uncertainty in the hydrologic flux of Ca^{2+} at Hubbard Brook (Campbell et al., 2016) by including more measured solutes and improving estimates of uncertainty associated with gaps (based on See et al., 2020), high streamflow, watershed area, and solute flux calculations (i.e., using concentration–discharge relationships to inform interpolation between sampling dates; Aulenbach & Hooper, 2006). We determined the overall uncertainty in stream solute flux and also

quantified the individual sources of uncertainty in discharge and solute concentration to rank the sources of uncertainty in the calculations.

2 | METHODS

2.1 | Study site

The Hubbard Brook Experimental Forest is located in the White Mountain National Forest in central New Hampshire, USA (43°56 'N, 71°45 'W). In this study we focused on Watershed 3 (W3), the hydrologic reference watershed, which has not been experimentally manipulated. The watershed is 42 ha in area with an elevation range of 527–732 m. Vegetation consists mostly of northern hardwoods: sugar maple (*Acer saccharum* Marsh.), American beech (*Fagus grandifolia* Ehrh.), and yellow birch (*Betula alleghaniensis* Britt.), with red spruce (*Picea rubens* Sarg.) and balsam fir (*Abies balsamea* (L.) Mill.) in areas of shallow soils and bedrock outcrops. Soils are derived from glacial drift of sandy loam to loamy sand texture, with thickness varying up to 8 m. Soils are classified mostly as base-poor Spodosols spanning a range of drainage classes (Bailey et al., 2014). The underlying bedrock is sillimanite grade mica schist, quartz schist, and calc-silicate granulite of the Silurian Rangeley formation (Barton et al., 1997).

2.2 | Stream solute flux calculations

At Hubbard Brook, solute fluxes are calculated as the product of solute concentration and discharge, normalized by watershed area. We used data collected during the 2013–2018 water years (i.e., beginning on 1 June 2013 and ending on 31 May 2019), because 2013 marked the advent of digital stage-height recording and a change in the analytical laboratory. Streamflow at the outlet of W3 is measured using a 120-degree v-notch weir for stage heights up to 0.6 m (2 feet); above that height, the rectangular area of the cement structure above the v-notch is used. Stream stage height was recorded in a stilling basin with a float-and-pulley system equipped with a shaft encoder (Campbell Scientific, CS410) until October 2017, when it was replaced with an optical encoder (AMASS Data Technologies, model PSE-SDI \D\LiPO). Weekly grab samples for stream chemistry are collected just upstream from the stilling basin to avoid sample contamination from the cement structure. In most years, additional automated sampling was done periodically during rainstorm and snowmelt events (Table 1).

Chemical analyses were performed at the U.S. Forest Service Laboratory in Durham, New Hampshire, except for pH, which was measured with a benchtop pH meter (Thermo Scientific, Orion 3-Star) at the Hubbard Brook Experimental Forest laboratory on the same day samples were collected. Samples were filtered with a pre-combusted glass-fibre filter (0.7 μm nominal pore size) and stored frozen prior to analysis, except for Ca^{2+}, Mg^{2+}, Na^+, K^+ and Si, for which an aliquot was poured off and refrigerated. Samples were analysed for SO_4^{2-},

TABLE 1 Annual streamflow, annual volume-weighted mean stream water solute concentration, and solute flux at W3 for the years of our study.

Concentration (mg/L)

Water Year	Streamflow (mm)	Number of samples	Ca^{2+}	Mg^{2+}	K^+	Na^+	NH_4^+-N	H^+	SO_4^{2-}-S	NO_3^--N	Cl^-	DOC	TDN	Si
2013	1118	52	0.75	0.24	0.15	1.07	0.004	1.68×10^{-6}	0.96	0.20	0.39	2.07	0.28	2.96
2014	810	101	0.73	0.23	0.12	0.98	0.004	1.52×10^{-6}	0.96	0.09	0.46	2.04	0.18	3.01
2015	984	94	0.68	0.19	0.12	0.95	0.004	1.88×10^{-6}	0.91	0.03	0.45	2.40	0.12	2.84
2016	780	100	0.71	0.24	0.15	1.02	0.004	1.32×10^{-6}	0.99	0.04	0.49	2.22	0.12	3.21
2017	1028	130	0.68	0.21	0.13	0.99	0.004	1.50×10^{-6}	0.93	0.03	0.47	2.40	0.11	2.94
2018	1301	96	0.61	0.19	0.10	0.80	0.007	1.37×10^{-6}	0.83	0.03	0.46	2.56	0.12	2.75

Flux (kg/ha/year)

Water Year	Streamflow (mm)	Ca^{2+}	Mg^{2+}	K^+	Na^+	NH_4^+-N	H^+	SO_4^{2-}-S	NO_3^--N	Cl^-	DOC	TDN	Si
2013	1118	8.13	2.47	1.83	9.83	0.06	4.10×10^{-5}	2.56	0.67	4.36	25.81	4.43	27.89
2014	810	5.55	1.63	1.22	6.33	0.06	2.67×10^{-5}	1.84	0.23	3.90	22.11	1.96	20.22
2015	984	5.95	1.60	1.15	7.60	0.05	3.36×10^{-5}	2.19	0.08	4.49	28.58	1.36	23.22
2016	780	4.54	1.36	1.02	5.47	0.04	2.06×10^{-5}	1.70	0.04	3.43	21.06	1.11	18.52
2017	1028	5.64	1.61	1.53	7.11	0.07	3.23×10^{-5}	2.20	0.07	5.06	33.84	1.37	22.90
2018	1301	6.69	1.94	1.55	7.94	0.11	3.25×10^{-5}	2.51	0.11	5.39	41.17	1.80	28.21

Note: The water year starts on June 1—for example water year 2013 begins on June 1, 2013 and ends on May 31, 2014.

NO_3^- and Cl^- using ion chromatography (Metrohm 761); NH_4^+ with automated colorimetry (SmartChem AQ2 Discrete Analyser); Ca^{2+}, Mg^{2+}, Na^+, K^+, and Si with inductively coupled plasma optical emission spectroscopy (Agilent 730); and DOC and TDN using high-temperature catalytic oxidation with chemiluminescent N detection (Shimadzu TOC-VCSH/TNM-1 analyser).

Some solutes are commonly below the method detection limit (Table 2). In these cases, solute fluxes at Hubbard Brook have traditionally been calculated using half the detection limit (Buso et al., 2000). For the purpose of quantifying uncertainty, however, we used the actual values obtained from the instrument, even for those below zero. We used this approach to avoid the bias that occurs with other methods of handling values below detection, such as eliminating them, setting them to zero, or using half the detection limit (Helsel, 1990). Because we report uncertainty using the 2.5th and 97.5th percentile of Monte Carlo iterations, occasional negative values in solute fluxes were not a problem for the analysis.

2.3 | Monte Carlo uncertainty analysis

We used Monte Carlo simulation to generate many estimates of stream loads by randomly sampling from possible values for each variable used in the calculations, (Figure 1), as described below, resulting in a probability distribution of the estimates (Press et al., 1986). The analyses were performed in the statistical computing language R (v3.5.2), and the documented computer code is included in the archived supplemental material. Confidence intervals were determined from the 2.5 and 97.5 percentiles of the distribution of the estimates, 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 analysis with all the sources of uncertainty combined and then one source of uncertainty at a time, to evaluate the relative importance of each source of uncertainty. We ran three sets of iterations for each combination of sources, increasing the number of iterations (500, 1000, 2000, or 4000) as needed for the three flux estimates to agree within 1%. In the case of NH_4^+, many more iterations (18000) were required to achieve this target, because concentrations were so low. A target agreement in units of concentration would have been attained more easily. The following is an explanation of how each source of uncertainty was estimated, including the determination of input distributions for the Monte Carlo analysis.

2.4 | Chemical analysis

To characterize uncertainty in the chemical analysis of stream samples, we used measured values of check standards run during routine sample analysis (Table 2). The uncertainty in each sample was represented by randomly sampling an error term from the comparison of the expected and measured concentrations of the check standard with the concentration closest to the stream sample in question. The number of different concentrations of check standards ranged from 8 to 12, depending on the element, across the range of stream water concentrations in our data set. The number of observations per concentration of check standards averaged from 2 to 1367, depending on the element and the concentration of the check standard. In the case of pH, we sampled a set of measured pH deviations from two check standards (pH 4 and 7) and used those as the analytical errors, before transforming the data from pH into H^+ concentration for calculation of fluxes.

2.5 | Stage height–discharge relationship

Uncertainty in the relationship between stream stage height and discharge was determined at Watershed 2, the weir adjacent to our

TABLE 2 Distribution of solute concentrations in W3 for 2013–2018 (5th, 50th, and 95th percentiles).

Analyte	W3 stream concentrations (mg/L)			Method detection limit (mg/L)	Samples below MDL (%)
	5	50	95		
Ca^{2+}	0.48	0.64	0.98	0.028	0
Mg^{2+}	0.14	0.19	0.31	0.003	0
K^+	0.07	0.12	0.51	0.011	0
Na^+	0.57	0.82	1.56	0.010	0
NH_4^+–N	0.00006	0.004	0.01	0.003	63
SO_4^{2-}–S	0.59	0.85	1.19	0.050	1
NO_3^-–N	0.0021	0.041	0.31	0.013	21
Cl^-	0.32	0.44	0.70	0.088	0
DOC	1.68	2.46	5.14	0.340	0
TDN	0.08	0.14	0.38	0.099	18
Si	1.74	2.47	3.96	0.045	3

Note: The method detection limits and the proportion of samples below detection are also shown.
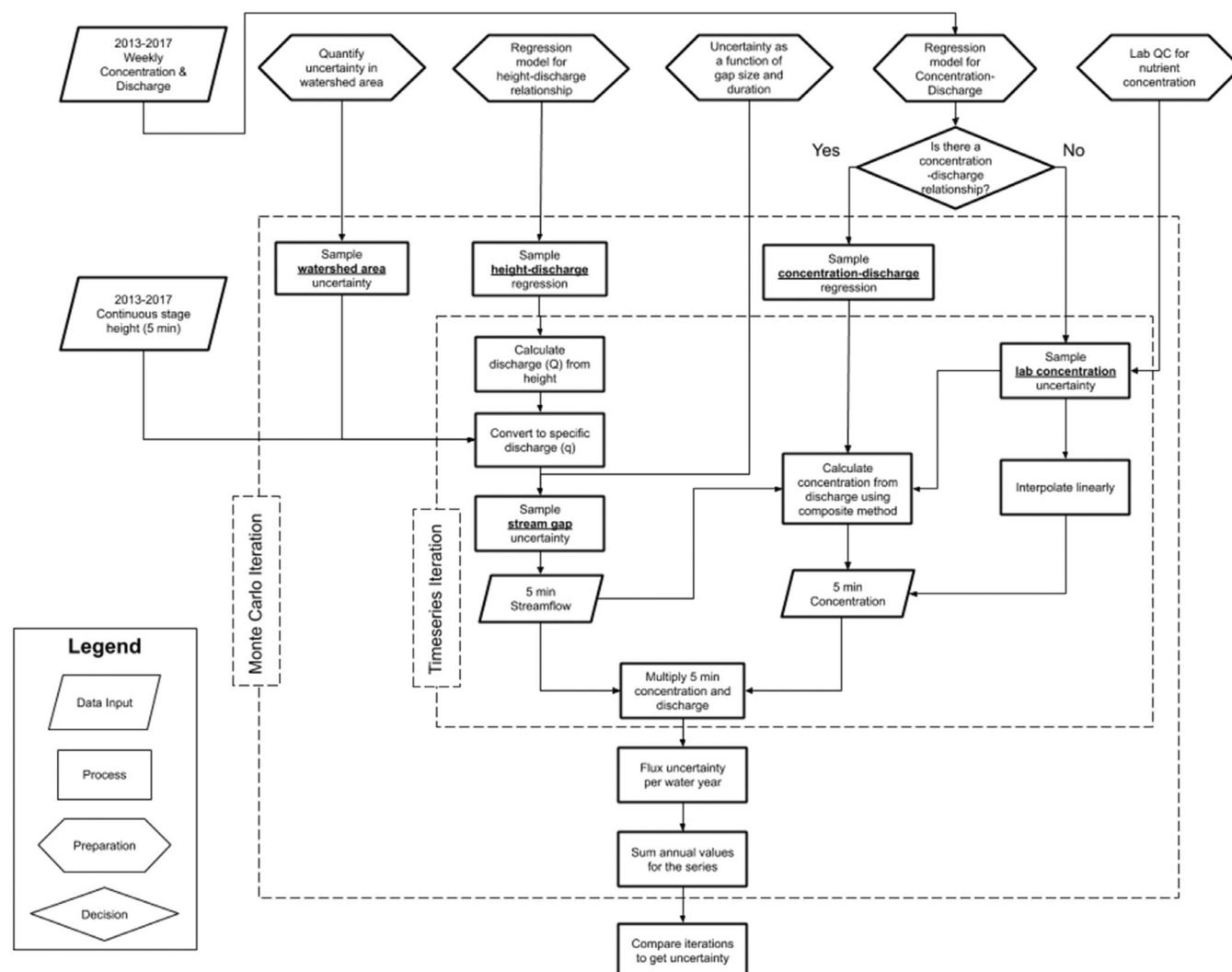
FIGURE 1 Flowchart of the steps in the Monte Carlo calculation of uncertainty in solute flux. Bold and underscore type indicates random sampling from data. Processes between the dotted lines showing the "Monte Carlo Iteration" and "Time series Iteration" involve selecting parameters that are held constant during each time series loop.

focus watershed; such measurements have not been made at W3. The control structures for W3 and W2 are similar sizes and designs, and both are tied into bedrock. Discharge was measured by timing the capture of stream water in a large container at flows up to 4 L/s and by salt-dilution gauging (Moore, 2005) at flows up to 33 L/s (Figure 2). There are 49 observations in the data set, collected on 26 dates between April and December 2017. Although stream discharge in W3 was within the range of these manual measurements 93% of the time (i.e., <32 L/s), higher flows accounted for 50% of the streamflow. The observations were compared to the theoretical weir equation, $Q = 2302 \times H^{2.449}$, where H is in m and Q is in L/s (Lindell et al., 2018) and the deviation of the observations from the discharge predicted by the theoretical curve was described with an equation of form $\Delta Q = a \times H/(b + H)$, where a and b are parameters. Rather than randomly selecting 49 values, we randomly selected 26 dates, with replacement, because measurements within a date were clearly not independent, and we randomly selected one observation from each

date. We fit this relationship 10 000 times (using the minpack.lm non-linear solver implemented in R; Elzhov et al., 2016) (Figure 3). To estimate uncertainty in the stage height–discharge relationship, we randomly selected one of these 10 000 relationships for each iteration of the Monte Carlo. Measurement uncertainty in stage-height and discharge are reflected in the model fit and were not additionally sampled in the Monte Carlo. Uncertainty in stage-height measurement can be characterized by comparing continuous measurements with manual measurements made with a hook gauge (Yanai et al., 2013).

2.6 | Gaps in stream streamflow measurements

We simulated uncertainty due to gaps in the streamflow record during the study period in W3. The most common causes of gaps at Hubbard Brook, in descending frequency, are debris in the v-notch, malfunctioning sensors, ice in the v-notch, weir maintenance and repairs, and
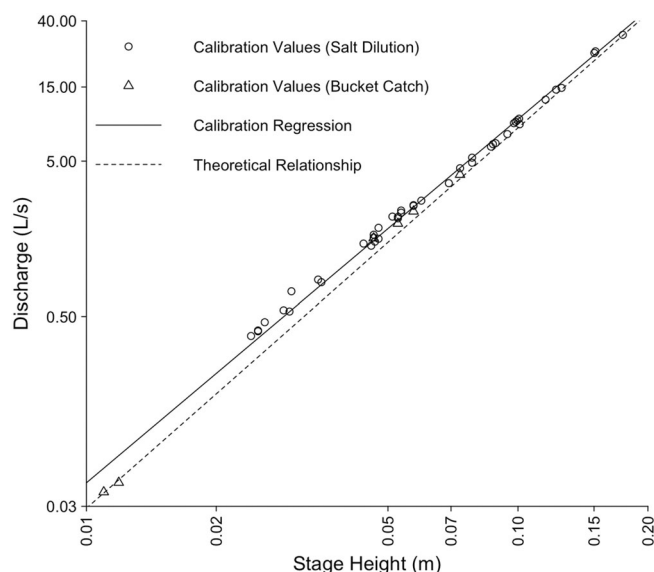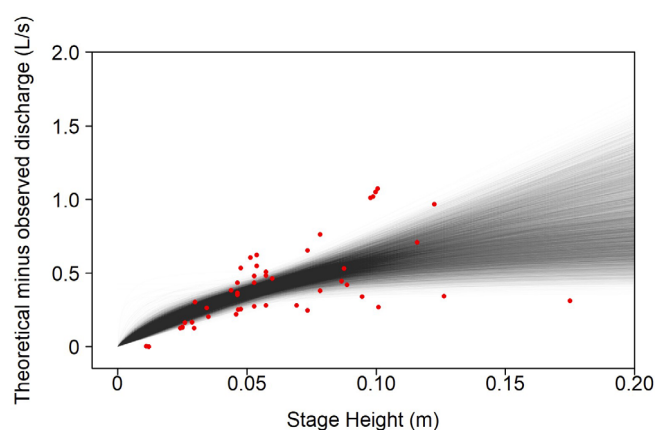
FIGURE 2 Relationship between stage height and discharge at the outlet of Watershed 2, which is adjacent to our focus watershed.

FIGURE 3 The deviation of the observed from the theoretical discharge as a function of stage height (H). There are 49 observations in the data set, collected on 26 dates between April and December 2017 at Watershed 2, which is adjacent to our focus watershed. The 10 000 lines show best-fit curves of the form $a \times H/(b + H)$, based on bootstrap resampling of data from the 26 dates.

technician error (Campbell et al., 2016). Gaps were filled using an ensemble of regressions built with data from 1955 to 2012, which relate discharge at each stream to the five other streams on the south-facing slope (See et al., 2020). Predicted values at the beginning and end of a gap were forced to match the measured values at the beginning and end of each gap. Uncertainty associated with filling gaps was characterized by comparing observed streamflow to streamflow estimated by this approach for a population of 1 000 000 artificial gaps produced for W6, a nearby reference watershed at Hubbard Brook (See et al., 2020). To account for uncertainty in filling gaps in the streamflow record for W3, we randomly selected a value from this population of errors (in quantiles of 1000, each with 1000 possible values) based on streamflow during each gap in the record for W3.

2.7 | Watershed area

To allow comparisons to precipitation inputs and to watersheds of different areas, stream discharge is divided by the area of each watershed. The measured area of the watershed is thus a source of uncertainty in stream flux estimates. We used areas delineated from two independent LiDAR-derived 1-m digital elevation models (DEMs) resampled to 3-m resolution with a low-pass smoothing filter, which gave areas of 41.8 and 42.1 ha (Gillin et al., 2015). The area of the original chain-and-compass survey of the topographic divides, conducted on the ground in the 1950s, produced a similar estimate (42.3 ha) but was not used in our analysis because the method is inferior to the DEM delineations. To characterize uncertainty in the best estimate of the watershed area, we randomly selected (with replacement) two values from our two estimates and used the mean value at each iteration of the Monte Carlo simulation (this is the simplest possible instance of bootstrapping).

## 2.8 | Temporal interpolation of stream chemistry

While stream discharge is measured nearly continuously, stream chemistry is measured at specific points in time; thus temporal interpolation of chemical concentrations is needed (Swistock et al., 1997; Ullrich & Volk, 2010), and the interpolation process is a source of uncertainty. Traditionally, solute fluxes at Hubbard Brook were calculated on a daily basis using daily streamflow and the chemical concentration measured on that day, if it was measured, and for the intervening days, using the average of the previous and subsequent samples (Buso et al., 2000). A better estimate of concentration between sampling dates can be obtained using the concentration–discharge relationship (Figure 4), if one exists (Aulenbach et al., 2016). We used this approach for Ca^{2+}, Mg^{2+}, Na^+, H^+, SO_4^{2-}, DOC, and Si, which had a coefficient of determination (r^2) of at least 0.3 between concentration and streamflow (Aulenbach et al., 2016). For K^+, NH_4^+, Cl^-, NO_3^-, and TDN, the concentration–discharge relationships were not strong ($r^2 < 0.3$), and concentrations were linearly interpolated between sampling dates. The uncertainty associated with linear interpolation of streamwater concentrations was not estimated.

The uncertainty in the concentration–discharge relationship was characterized by randomly selecting observations with replacement and fitting a log–log relationship for each solute for each iteration of the Monte Carlo. We did not add analytical uncertainty to these data, because uncertainty in measurement contributes to the model error and should not be counted twice. The solute concentration was predicted using this concentration–discharge relationship, but forcing the curve through the observed concentrations (including uncertainty in chemical analysis) by adding in the linearly interpolated residual (Figure 5; Aulenbach & Hooper, 2006). For elements that have lower correlations, we used linear interpolation, and uncertainty was estimated only from the uncertainty in chemical analysis (described
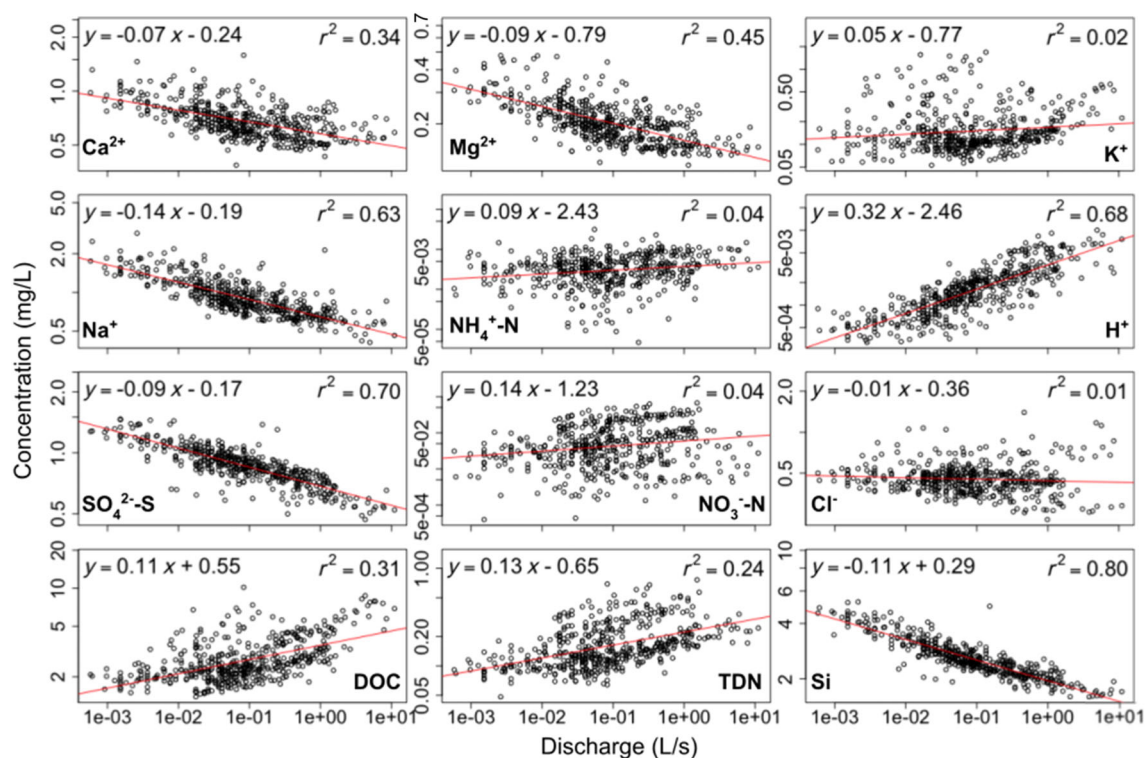
FIGURE 4 Concentration–discharge relationships. For solutes with $r^2 > 0.03$, we used these relationships for interpolating concentration values between sampling dates (illustrated in Figure 5).
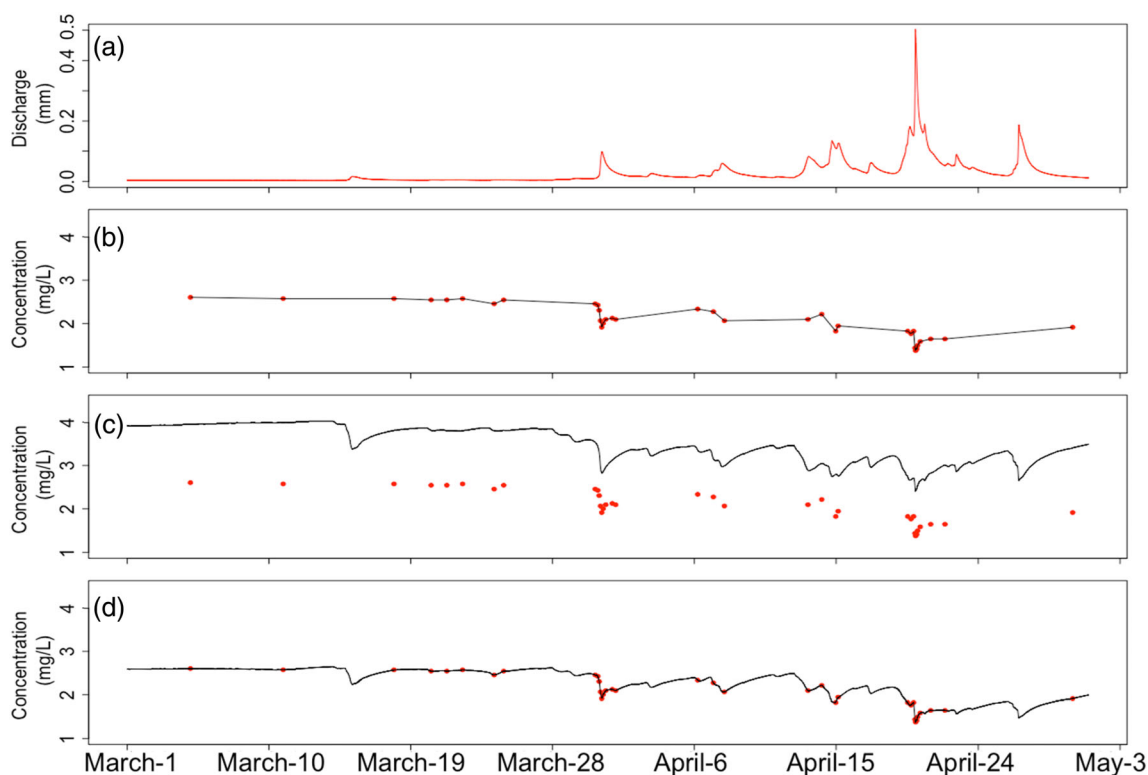
FIGURE 5 Example of interpolation of solute concentration using SO_4^{2-} from March through April 2019 showing: (a) discharge, (b) linear interpolation (in black) between observations (red dots), (c) modelled concentration based on discharge (Figure 4), (d) interpolation based on modelled concentration.

above). In both approaches, we predicted concentrations at 5-min time intervals, consistent with the resolution of the streamflow record.

# 3 | RESULTS

3.1 | Streamflow

Uncertainty in annual streamflow from W3 at Hubbard Brook varied somewhat from year to year (Table 1; Figure 6). In most years, the most important source of uncertainty in streamflow, represented as the 95% CI of Monte Carlo iterations, was the stage height–discharge relationship, which accounted for uncertainties of 1.9%–2.2%, depending on the year. Median values of the parameters for the relationship between the Q residual and stage height were $a = 1.32$ and $b = 0.13$, resulting in a range of bootstrapped curves used in the Monte Carlo analysis (Figure 3).

In 2 years, the most important source of uncertainty in streamflow was the filling of gaps in the record of stage heights (Figure 6a). Overall, there were 1100 gaps of at least 5 min duration in the streamflow record from W3 during the 6 years of this study. The average gap duration was 0.3 days; the longest gap was 20 days.

Figure 7 shows the distribution of all gap durations against their corresponding simulated cumulative streamflow. Half of the gaps fell between 10 and 20 min (interquartile range), and cumulative
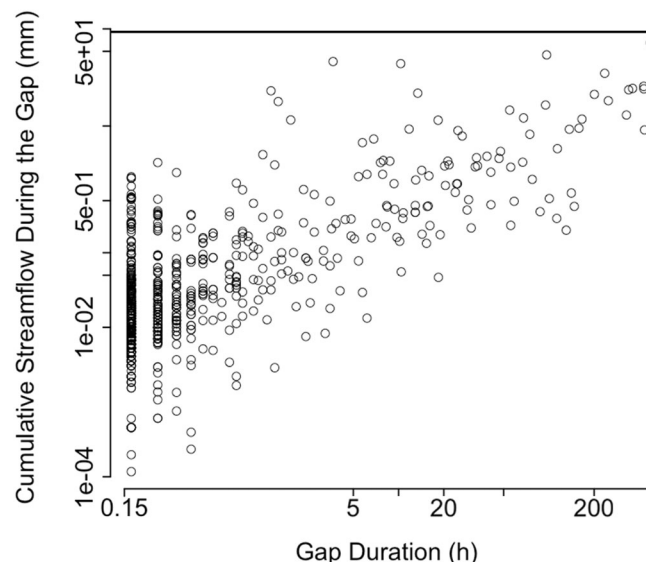
FIGURE 7 The distribution of all gap durations against their corresponding simulated cumulative runoff.
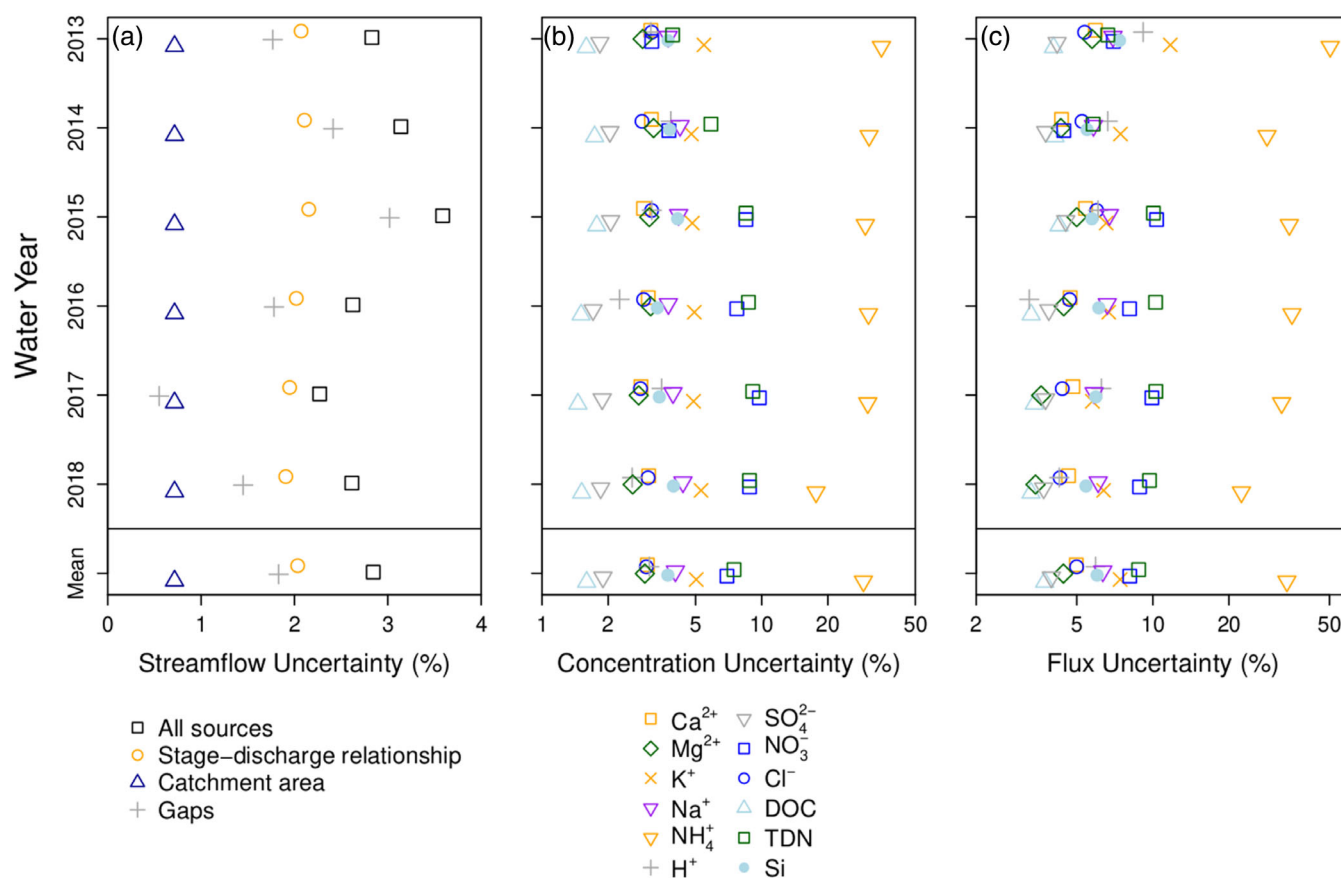
FIGURE 6 Uncertainty (95% CI relative to the median) in (a) annual streamflow, (b) annual volume-weighted average concentration, and (c) annual solute flux for each solute, for each year and for the 6-year study period.

TABLE 3 Routine check standards were used to characterize uncertainty in solute chemistry.

Analyte	Check standard (mg/L)	Number of replicates	Accuracy (mg/L)	Bias (mg/L)	Precision (mg/L)	Number of stream samples
Ca^{2+}	0.5	148	0.024	0.015	0.033	229
	1	1307	0.031	−0.001	0.050	104
Mg^{2+}	0.1	1367	0.004	0.002	0.005	46
	0.25	148	0.009	0.005	0.012	281
	0.5	148	0.017	0.001	0.025	6
K^{+}	0.005	150	0.006	0.006	0.012	1
	0.05	161	0.006	0.000	0.009	20
	0.1	1347	0.007	0.001	0.010	228
	0.25	148	0.014	0.002	0.023	52
	0.5	148	0.024	0.000	0.036	23
	1	1306	0.044	−0.009	0.072	9
Na^{+}	0.5	148	0.031	0.000	0.058	60
	1	1307	0.048	−0.018	0.069	264
NH_4^{+}-N	0.005	396	0.002	0.001	0.003	229
	0.01	313	0.002	0.002	0.003	90
	0.025	476	0.003	0.002	0.003	14
	0.05	226	0.003	0.003	0.003	1
SO_4^{2-}-S	1.5	30	0.019	−0.003	0.025	16
	2.5	785	0.020	−0.002	0.029	306
	5	798	0.031	0.010	0.045	12
NO_3^{-}-N	0.05	775	0.005	−0.001	0.006	75
	0.1	837	0.004	0.000	0.006	107
	0.5	889	0.006	−0.003	0.007	58
	1	820	0.006	−0.002	0.012	80
	2.5	795	0.010	0.003	0.015	13
	3.07	23	0.021	−0.009	0.028	1
Cl^{-}	0.3	30	0.017	0.011	0.019	90
	0.5	843	0.023	−0.008	0.029	234
	1	743	0.032	−0.003	0.041	10
DOC	1	431	0.053	0.004	0.072	47
	2.5	451	0.060	−0.007	0.079	243
	5	521	0.089	−0.007	0.120	39
	7.5	430	0.122	−0.001	0.163	4
TDN	0.05	291	0.002	0.001	0.011	61
	0.1	523	0.028	−0.006	0.033	124
	0.25	582	0.014	−0.001	0.018	115
	0.5	431	0.020	−0.010	0.024	30
	0.75	448	0.027	−0.018	0.029	4
Si	0.01	172	0.038	0.038	0.019	10
	1	1242	0.033	−0.008	0.048	4
	2.5	114	0.162	0.156	0.211	296
	5	113	0.155	0.075	0.206	26
H^{+}	0.0001	44	1.91×10^{-6}	-1.27×10^{-6}	2.37×10^{-6}	286
	0.1	44	2.50×10^{-3}	-5.76×10^{-4}	3.20×10^{-3}	47

Note: Each stream water sample was assigned a random error term taken from the population of check standards closest to it in concentration. Accuracy is the difference between the observed and target concentration, reported as the average of the absolute values of the errors. Bias is the average of the positive and negative errors. Precision is the standard deviation of the difference between the laboratory check standard and measured concentration. The number of stream samples from this study associated with each check standard is also shown.

streamflow during the gaps had an interquartile range of 0.008–0.04 mm. The uncertainty in streamflow due to filling gaps in the record averaged 1.8% (17.3 mm) and ranged from 0.6% to 3.0%, depending on the year. The years with highest uncertainty due to gaps were those with the most streamflow during the gap, which ranged from 18 to 107 mm.

The other source of uncertainty in streamflow was the watershed area (Figure 6a). This was only 0.7% of annual streamflow, as the two independent LiDAR-derived 1-m digital elevation models gave similar results. The uncertainties were the same every year because the inputs to the calculation did not vary.

With all sources combined, uncertainty in annual streamflow for water years 2013–2018 ranged from 20 to 35 mm, depending on the year. Uncertainties were consistent as a percentage of annual streamflow (2.3%–3.6%, Figure 6), with no relationship to wet or dry years (Table 1).

3.2 | Solute concentrations

For most solutes, the greatest source of uncertainty in stream fluxes was the determination of concentrations in the laboratory. Based on variation in concentrations of quality control standards (Table 3), NH_4^+, which was the most dilute solute in our dataset (Table 1) had the highest uncertainty—18%–35%, depending on the year. Nitrate had similar uncertainties in units of concentration, but smaller uncertainties as a percentage of concentration—3%–10%—because NO_3^- concentrations were higher than NH_4^+ concentrations (Table 1). Total dissolved N had higher uncertainties in units of concentration, but concentrations were higher, so percentage uncertainties ranged from 4% to 9%. Base cations are analysed by ICP, and they had similar uncertainties in laboratory determination of concentration: the average annual volume-weighted uncertainties were 2.8% for Mg^{2+}, 2.9% for Ca^{2+}, 3.1% for Cl^-, 3.7% for Si, 3.8% for Na^+, and 5.1% for K^+. Uncertainties were lowest for DOC (1.3%), H^+ (1.4%), and SO_4^{2-} (1.6%).

The concentration–discharge relationships used to interpolate concentration between observations (where these relationships had $r^2 > 0.3$) produced uncertainties of 0.1% for SO_4^{2-} and Si, 0.4% for Ca^{2+}, Mg^{2+}, and H^+, 0.5% for DOC, and 0.6% for Na^+.

The combined effect of uncertainty from both laboratory analyses and the concentration–discharge relationship produced average uncertainties in annual volume-weighted concentrations ranging from 1.6% to 29%, with DOC having the lowest and NH_4^+ the highest uncertainty (Figure 6b). In units of concentration, the solutes with high uncertainty (Si, SO_4^{2-}, Na^+, DOC, Ca^{2+}, in descending order) were the ones with greatest concentrations (Table 1) (Si, DOC, SO_4^{2-}, Na^+, Ca^{2+}); uncertainties varied less across solutes as a percentage of the mean (Figure 6b).

3.3 | Solute fluxes

Uncertainties in solute fluxes (Figure 6c) were dominated by uncertainty in concentration (Figure 6b), which exceeded uncertainty in

streamflow (Figure 6a) for all solutes with the exception of DOC, which had the lowest concentration uncertainty. Combining streamflow with solute concentrations resulted in uncertainties in annual stream fluxes ranging from 3.7% for DOC in water year 2017 to 34% for NH_4^+ in water year 2013 (Figure 6c). Excluding NH_4^+, for which uncertainties are high in units of % but low in kg/ha/year, the highest solute flux uncertainty was 9% for TDN. In units of flux, uncertainties were highest for Si (1.4 kg/ha/year), DOC (1.1 kg/ha/year), SO_4^{2-} (0.9 kg/ha/year), and Na^+ (0.5 kg/ha/year) and lowest for H^+ (0.00003 kg/ha/year) and NH_4^+ (0.02 kg/ha/year), reflecting the magnitudes of the fluxes.

4 | DISCUSSION

Overall, uncertainty in streamflow (Figure 6a), solute concentrations (Figure 6b), and solute fluxes (Figure 6c) were low, generally <10%. Uncertainty in streamflow or river discharge would obviously be much higher if measured in a natural channel with an empirical rating curve instead of a weir where hydraulic conditions are better controlled. Uncertainty in solute concentrations in units of concentration are likely similar across sites if analytical methods are similar, but differences in stream water solute concentrations across sites will result in differences in uncertainty as a percentage of solute flux. For this analysis, we selected a period of consistent methods for laboratory analyses and stream discharge measurements. Thus our analysis does not reflect changes in methods over time. Earlier methods included chart recorders, hook gage readings, and different analytical methods (Yanai et al., 2015). It is not common to collect and retain all the information necessary to conduct a complete uncertainty analysis; we hope that this paper helps increase awareness of the value of full uncertainty accounting.

The uncertainties we report differ slightly from those in our previous analysis of Ca^{2+} flux in a nearby watershed at Hubbard Brook (Campbell et al., 2016). The uncertainty in the stage height–discharge relationship was the smallest source of uncertainty in the earlier analysis. The current study quantified uncertainty in the stage height–discharge relationship at flows up to 33 L/s; the previous rating curves applied to flows up to 1.3 or 2.5 L/s. Gaps were a more prominent source of uncertainty compared to this analysis, because the current study used a better method for interpolating streamflow between collection dates (See et al., 2020). The previous method (Campbell et al., 2016), like this one, filled gaps with a regression model but did not force the predictions through the observations, as illustrated in this paper for concentration interpolation (Figure 5). The current study also used a better method for interpolating stream chemistry.

Some sources of uncertainty are widely reported and some are generally overlooked. Sampling error, characterized by replicate measurements, is the source most commonly reported in the ecological literature (Yanai et al., 2021). At Hubbard Brook, there are six similar south-facing headwater watersheds and three larger north-facing watersheds that can be considered replicates (Yanai et al., 2015). However, replication is not an option when characterizing streamflow

at a single point or solute export from a particular watershed. This is why propagating the sources of error involved in the calculation is often the best way to evaluate the uncertainty in watershed studies.

For all solutes except for DOC, the greatest source of uncertainty was the measurement of solute concentrations. This source of uncertainty is generally easy to characterize, when the chemical analysis of solutes is conducted in laboratories that follow standard procedures for quality assurance and quality control (e.g., APHA, 2017). Since stream water is so dilute at Hubbard Brook (Likens & Buso, 2012), the uncertainty in lab analysis (Table 3) constitutes a larger fraction of the total uncertainty than at sites with higher solute concentrations. Although uncertainty in solute concentrations is easily determined, it is not always reported; only 16 of 45 papers reporting solute concentrations reported on uncertainty of chemical analyses in a random sample of the literature (Yanai et al., 2021).

Watershed area is a source of uncertainty in annual streamflow that is rarely addressed. Although there have been eight DEM-based estimates of the area of Watershed 3, they were based on only two independent DEMs, each processed four different ways (Gillin et al., 2015). We used only the best estimate based on each DEM for our Monte Carlo, rather than sample from eight observations as if they were independent estimates. Obviously, with only two independent DEMs, we cannot be confident that they characterize the true uncertainty in watershed area at our site. A potentially more important source of uncertainty is whether the topographic divide describes the watershed, due to subsurface drainage patterns that may not follow surface topography. Consistently low evapotranspiration calculated for the watershed adjacent to our study site, using mean annual precipitation minus streamflow (Bailey, 2003), suggests that the magnitude of this effect may be on the order of 11% of streamflow. The hydrologic divide could even change with the hydrologic conditions, depending on the difference between the topography of the impermeable layer and the land surface. This possible discrepancy between the topographic and hydrologic watershed area is not addressed in our approach to quantifying uncertainty.

Uncertainty in annual streamflow is difficult to evaluate due to the challenge of estimating uncertainty at high flows. Measuring discharge is relatively easy at low flow but difficult at high flow (Hornbeck, 1965). In this case study, we have good data at low flow measured with a bucket and stopwatch, but few measurements of high flows with salt dilution (Moore, 2005); high flows are hard to capture because they occur infrequently. Unfortunately, they are also important: half of the streamflow volume occurred at flows higher than we had data for the relationship of stage height to discharge, although these high flows took place in just 7% of the record, in units of time. We hope to improve our knowledge of this relationship at Hubbard Brook with additional measurements, as our uncertainty in this uncertainty source is high.

Our Monte Carlo estimate of uncertainty does not account for bias, which is important to acknowledge. Bias is more difficult to quantify than precision. In the case of laboratory analyses, precision can be calculated based on replicate measurements of the same sample, but evaluating bias depends on an external standard of known concentration. Based on external reference samples, the median bias in Hubbard Brook solute concentrations for the time period of our study was -0.1% for NH_4^+, -2.0% for NO_3^-, 2.9% for Cl, and -5.0% for TDN (Table 3). The stage height–discharge relationship has a theoretical basis, but can be affected by factors such as the sharpness of the weir blade and the velocity of the water as it approaches the v-notch (Hornbeck, 1965); calibrating this relationship is a means of describing bias. Errors in quantifying watershed area could be a source of bias in estimates of streamflow and stream fluxes. Specifically, if watershed area is underestimated, streamflow per unit area is overestimated, and vice versa. This bias would be constant over time, unless the contributing area changes with hydrologic condition, as described above. The solute sampling regime can also be a source of bias for concentration and flux estimation, insofar as the samples do not represent the true distribution of concentrations. One way to address this a priori could be to leverage basic understanding of solute sources and their travel time distributions to determine how to time stream water sampling across flow conditions and seasons. Scrutinizing the assumptions of an uncertainty analysis can help to identify potential sources of bias. Even though they may not be included in the analysis, it is now more likely that we will conduct quality assurance measurements that help identify them.

Uncertainty analyses are useful in setting priorities for making improvements in environmental monitoring. In this case study, the greatest source of uncertainty was in measurement of chemical concentrations, especially as a percentage of concentration or flux when these were very low. If quantifying these low values is important, rather than simply knowing that they are low (uncertainties were low in units of fluxes), then other methods would be needed for measuring them. More independent measurements of discharge at high flows are needed to better characterize the uncertainty in annual streamflow at Hubbard Brook, which is challenging because these events are rare and difficult to capture. When data such as these are not available, providing an estimate of uncertainty and the methods used to quantify it makes it possible to evaluate confidence in the values.

DATA AVAILABILITY STATEMENT

ORCID

Ge Pu https://orcid.org/0000-0002-0818-8795

John L. Campbell https://orcid.org/0000-0003-4956-1696

Mark B. Green https://orcid.org/0000-0002-7415-7209

Ruth D. Yanai https://orcid.org/0000-0001-6987-2489

REFERENCES

APHA. (2017). *Standard methods for the examination of water and wastewater* (23rd ed.). American Public Health Association.

Aulenbach, B. T., Burns, D. A., Shanley, J. B., Yanai, R. D., Bae, K., Wild, A. D., Yang, Y., & Yi, D. (2016). Approaches to stream solute load estimation for solutes with varying dynamics from five diverse small watersheds. *Ecosphere, 7*, e01298. https://doi.org/10.1002/ecs2.1298

Aulenbach, B. T., & Hooper, R. P. (2006). The composite method: An improved method for stream-water solute load estimation. *Hydrological Processes, 20*, 3029–3047. https://doi.org/10.1002/hyp.6147

Bailey, A. S. (2003). *Hydrometeorological database for Hubbard brook experimental Forest, 1955–2000* (Vol. 305). US Department of Agriculture, Forest Service, Northeastern Research Station.

Bailey, S. W., Brousseau, P. A., McGuire, K. J., & Ross, D. S. (2014). Influence of landscape position and transient water table on soil development and carbon distribution in a steep, headwater catchment. *Geoderma, 226*, 279–289.

Barton, C. C., Camerlo, R. H., & Bailey, S. W. (1997). *Bedrock geologic map of Hubbard Brook Experimental Forest, miscellaneous investigation series map I-2562*. U.S. Geological Survey.

Bormann, F. H., & Likens, G. E. (1967). Nutrient cycling: Small watersheds can provide invaluable information about ecosystems. *Science, 155*(3761), 424–429. https://doi.org/10.1126/science.155.3761.424

Bosch, J. M., & Hewlett, J. D. (1982). A review of catchment experiments to determine the effect of vegetation changes on water yield and evapotranspiration. *Journal of Hydrology, 55*(1–4), 3–23. https://doi.org/10.1016/0022-1694(82)90117-2

Buso, D. C., Likens, G. E., & Eaton, J. S. (2000). *Chemistry of precipitation, streamwater, and lakewater from the Hubbard Brook Ecosystem study: A record of sampling protocols and analytical procedures* (p. 52). Northeastern Research Station.

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

Elzhov, T. V., Mullen, K. M., Spiess, A. N., Bolker, B., Mullen, M. K. M., & Suggests, M. A. S. S. (2016). Package 'MINPACK. Lm'. Title R interface Levenberg-Marquardt nonlinear least-sq. algorithm found MINPACK plus support bounds.

Gillin, C. P., Bailey, S. W., McGuire, K. J., & Prisley, S. P. (2015). Evaluation of LiDAR-derived DEMs through terrain analysis and field comparison. *Photogrammetric Engineering & Remote Sensing, 81*(5), 387–396. https://doi.org/10.14358/PERS.81.5.387

Harmel, R. D., Cooper, R. J., Slade, R. M., Haney, R. L., & Arnold, J. G. (2006). Cumulative uncertainty in measured streamflow and water quality data for small watersheds. *Transactions of the ASABE, 49*(3), 689–701.

Helsel, D. R. (1990). Less than obvious: Statistical treatment of data below the detection limit. *Environmental Science & Technology, 24*, 1766–17740.

Holloway, J. M., & Dahlgren, R. A. (2001). Seasonal and event-scale variations in solute chemistry for four Sierra Nevada catchments. *Journal of Hydrology, 250*(1–4), 106–121. https://doi.org/10.1016/S0022-1694(01)00424-3

Hornbeck, J. W. (1965). *Accuracy in streamflow measurements on the Fernow experimental Forest. Research note NE-29* (p. 8). US Department of Agriculture, Forest Service, Northeastern Forest Experiment Station.

Hornbeck, J. W., Adams, M. B., Corbett, E. S., Verry, E. S., & Lynch, J. A. (1993). Long-term impacts of forest treatments on water yield: A summary for northeastern USA. *Journal of Hydrology, 150*(2–4), 323–344.

Kiang, J. E., Gazoorian, C., McMillan, H., Coxon, G., Le Coz, J., Westerberg, I. K., & Mason, R. (2018). A comparison of methods for streamflow uncertainty estimation. *Water Resources Research, 54*(10), 7149–7176.

Levine, C. R., Yanai, R. D., Lampman, G. G., Burns, D. A., Driscoll, C. T., Lawrence, G. B., Lynch, J. A., & Schoch, N. (2014). Evaluating the efficiency of environmental monitoring programs. *Ecological Indicators, 39*, 94–101. https://doi.org/10.1016/j.ecolind.2013.12.010

Lewis, D. J., Singer, M. J., Dahlgren, R. A., & Tate, K. W. (2006). Nitrate and sediment fluxes from a California rangeland watershed. *Journal of Environmental Quality, 35*(6), 2202–2211. https://doi.org/10.2134/jeq2006.0042

Likens, G. E., & Buso, D. C. (2012). Dilution and the elusive baseline. *Environmental Science & Technology, 46*(8), 4382–4387.

Lindell, J. E., Moore, W. P., & King, H. W. (2018). *Handbook of hydraulics* (8th ed., p. 416). McGraw-Hill Education.

McMillan, H., Krueger, T., & Freer, J. (2012). Benchmarking observational uncertainties for hydrology: Rainfall, river discharge and water quality. *Hydrological Processes, 26*(26), 4078–4111. https://doi.org/10.1002/hyp.9384

Moore, R. D. (2005). Slug injection using salt in solution. *Streamline Watershed Management Bulletin, 8*(2), 1–6.

Neary, D. G. (2016). Long-term forest paired catchment studies: What do they tell us that landscape-level monitoring does not? *Forests, 7*(8), 164. https://doi.org/10.3390/f7080164

Pappenberger, F., & Beven, K. J. (2006). Ignorance is bliss: Or seven reasons not to use uncertainty analysis. *Water Resources Research, 42*(5), W05302. https://doi.org/10.1029/2005WR004820

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. (1986). *Numerical Recipes* (p. 818). Cambridge University Press.

See, C. R., Green, M. B., Yanai, R. D., Bailey, A. S., Campbell, J. L., & Hayward, J. (2020). Quantifying uncertainty in annual runoff due to missing data. *PeerJ, 8*, e9531. https://doi.org/10.7717/peerj.9531

Swistock, B. R., Edwards, P. J., Wood, F., & Dewalle, D. R. (1997). Comparison of methods for calculating annual solute exports from six forested Appalachian watersheds. *Hydrological Processes, 11*(7), 655–669. https://doi.org/10.1002/(SICI)1099-1085(199706)11:7<655::AID-HYP525>3.0.CO;2-4

Tittel, J., Büttner, O., Friese, K., Lechtenfeld, O. J., Schuth, S., von Tümpling, W., & Musolff, A. (2022). Iron exports from catchments are constrained by redox status and topography. *Global Biogeochemical Cycles, 36*(1), e2021GB007056. https://doi.org/10.1029/2021GB007056

Ullrich, A., & Volk, M. (2010). Influence of different nitrate–N monitoring strategies on load estimation as a base for model calibration and evaluation. *Environmental Monitoring and Assessment, 171*(1), 513–527. https://doi.org/10.1007/s10661-009-1296-8

Yanai, R. D., Mann, T. A., Hong, S. D., Pu, G., & Zukswert, J. M. (2021). The current state of uncertainty reporting in ecosystem studies: A systematic evaluation of peer-reviewed literature. *Ecosphere, 12*(6), e03535. https://doi.org/10.1002/ecs2.3535

Yanai, R. D., Tokuchi, N., Campbell, J. L., Green, M. B., Matsuzaki, E., Laseter, S. N., Brown, C. L., Bailey, A. S., Lyons, P., Levine, C. R.,

Buso, D. C., Likens, G. E., Knoepp, J. D., & Fukushima, K. (2015). Sources of uncertainty in estimating stream solute export from headwater catchments at three sites. *Hydrological Processes*, *29*(7), 1793–1805. https://doi.org/10.1002/hyp.10265

Yanai, R. D., Vadeboncoeur, M. A., Hamburg, S. P., Arthur, M. A., Fuss, C. B., Groffman, P. M., Siccama, T. G., & Driscoll, C. T. (2013). From missing source to missing sink: Long-term changes in the nitrogen budget of a northern hardwood forest. *Environmental Science & Technology*, *47*(20), 11440–11448. https://doi.org/10.1021/es402572