

Investigating the effect of lab bias on long-term stream chemistry data

Cindi L. Brown^{a,*}, Chelcy F. Miniati^b and Jennifer D. Knoepp^a

^a USDA Forest Service, Southern Research Station, Coweeta Hydrologic Lab, 3160 Coweeta Lab Road, Otto, NC 28763, USA

^b USDA Forest Service, Rocky Mountain Research Station, 333 Broadway SE, Suite Albuquerque, NM 87102, USA

*Corresponding author. E-mail: cindi.l.brown@usda.gov

ABSTRACT

Long-term (LT) stream chemistry studies are used to examine changes in and responses to the environment. Much of the data collected over long periods of time goes through changes in instrumentation, methods, and personnel potentially resulting in changing values. A data user must understand these measures of data quality through quality control (QC) results to know with certainty if trends are real or attributable to other factors. We used the Web of Science database search engine to search for LT stream chemistry studies. For each study, we then determined: record or study length; if QC was reported; and if QC was used. We found that 33% of papers reported QC in the method, and 12% presented the QC in the results. Next, we conducted a case study on 46 years of stream chemistry data to evaluate the data with and without the application of QC protocols from two watersheds (WS) at Coweeta Hydrologic Laboratory; WS 7; clear-cut in 1967–77 and adjacent WS 2 which serves as a reference. We focused on nitrogen and sulfur due to their importance in understanding the forest ecosystem response to disturbance (NO_3) and acid deposition (SO_4). We determined average annual dissolved inorganic nitrogen (DIN) export ($\text{NH}_4 + \text{NO}_3 = \text{DIN}$) using three methods for censoring values below the method detection limit (mdl): (1) the found value, (2) the value of zero, and (3) one-half the mdl value. We found that DIN export for WS 2/WS 7 was (1) 66.9/831.4 ($\text{g ha}^{-1} \text{yr}^{-1}$), (2) 45.4/808.0 ($\text{g ha}^{-1} \text{yr}^{-1}$), and (3) 72.1/823.2 ($\text{g ha}^{-1} \text{yr}^{-1}$) using the three censoring methods, and that the export estimate was significantly different for WS 2 but not for WS 7 ($P = 0.001$). We found that on average stream NH_4 concentrations were below the mdl 58% of the time until an instrument change in 1994 resulted in improved mdls resulting in fewer data points below detection. We found increased bias for stream SO_4 concentration following an instrumentation change from segmented flow analysis to ion chromatography. As a result, stream SO_4 concentration data that were bias-corrected declined more rapidly in WS 2 compared with non-bias-corrected data, but not in WS 7. We conclude that including QC results with LT data is essential to verify data validity and give the data user a full understanding of the results.

Key words: method detection, nitrogen, quality control, sulfate, uncertainty

HIGHLIGHTS

- Long-term (LT) stream chemistry studies are used to examine environmental changes, but few studies report QA/QC information.
- An analysis of 272 LT stream chemistry papers revealed that only 33% reported QC.
- A study of 46 years of stream nitrogen export from Coweeta Hydrologic Laboratory showed that incorporating QA/QC data for values below the detection limit resulted in significantly different export estimates.

INTRODUCTION

Quantifying the effects of changing climate, land use, land management, or policy often employs long-term (LT) data. Additional benefits of LT studies, such as identifying infrequent events, trends, and unexpected findings, have also been established (Whitehead *et al.* 2009; Lindenmayer *et al.* 2010), while noting it may take 20–30 years to recognize the trend (Burt *et al.* 2008). Collecting data over these long periods can come with changes in method, instrumentation, and personnel – all sources of uncertainty in the data (Buso *et al.* 2000; Campbell *et al.* 2016). Uncertainty studies have shown that nutrient budgets can be misrepresented if the error is not recognized and propagated; uncertainty from sample analysis worst case scenario can lead to a 400% error; and identifying the greatest sources of uncertainty can help mitigate the errors (Harmel *et al.* 2006; Lehrter & Cebrian 2010; Yanai *et al.* 2015).

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY-NC-ND 4.0), which permits copying and redistribution for non-commercial purposes with no derivatives, provided the original work is properly cited (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

From [Standard Methods \(2012\)](#) – ‘**Quality assurance (QA)** is a laboratory operations program that specifies the measures required to produce defensible data with known precision and accuracy. This program is defined in a QA manual, written procedures, work instructions, and records. The manual should include a policy that defines the statistical level of confidence used to express data precision and bias, as well as method detection levels (MDLs) and minimum reporting limits (MRLs). The overall system includes all QA policies and **quality control (QC)** processes needed to demonstrate the laboratory’s competence and to ensure and document the quality of its analytical data.’

The quality of data obtained during water sample analyses must be supported by measurements of error which provide a basis for assessing quality ([Taylor 1987](#)). In laboratory analyses, there are three sources of error: data entry, contamination, and analytical error ([Buso *et al.* 2000](#)). Data entry errors and contamination are identified through QA procedures. Analytical error, the uncertainty in the analysis, can be quantified by using certified QC reference samples. Although analytical error is reported by the instrument manufacturer, the actual uncertainty can vary year-to-year and from operator-to-operator. Documenting these changes and how the data were affected (the uncertainty) is an important tool for the data user ([Beard *et al.* 1999](#)) and should be part of the QA/QC program of every laboratory ([Sullivan *et al.* 2012](#)). Reliability of data is assessed using the QC parameters: bias (mean error), precision, (relative standard deviation, %RSD), and the method detection limit (mdl) ([Standard Methods 2012](#)). Bias tells a user how close the mean of observed values is to the actual value while precision allows the user to know how reproducible that value is. The two parameters give the accuracy of the result ([Walther & Moore 2005](#); [González & Herrador 2007](#)). The mdl defines the minimum concentration of a substance that can be measured and reported with 99% confidence that the analyte concentration is greater than zero ([USEPA 2016](#)). Minimizing the analytical error associated with the analysis should be the goal of every lab ([Campbell *et al.* 2016](#)), but keeping records and reporting the error is also crucial to good data management ([Karl *et al.* 1989](#); [Michener *et al.* 2011](#)). Error in data found outside the established range is termed censored data. Much of the censored data in stream chemistry is data found to be below the mdl ([Kroll & Stedinger 1996](#)). Data below the mdl may or may not produce a value depending on the sensitivity of the instrument making the measurement. Censoring data below mdl produces distorted results and the loss of available information ([Farnham *et al.* 2002](#); [Helsel 2005](#)). Helsel investigates three approaches to extract information from censored data: substitution, maximum likelihood estimation, and nonparametric methods. [Farnham *et al.* \(2002\)](#) investigate the differences of assigning the censored data to values of zero, one-half the mdl, and the mdl for measurements involving trace analysis. They found that substituting one-half the mdl was superior to substitution with 0 or mdl for the data set they worked with and that all of the substitution methods behaved poorly when 30% of the observations were below mdl.

Changing methods or instruments used in an LT sampling program can affect the data if not realized and accounted for. Effects can manifest as increased or decreased variance, or step changes. For example, [Evans *et al.* \(2007\)](#) found that, over the course of a 26-year study measuring sulfate in water samples, the data were initially noisy but became stable after a change in method occurred. [Meyer *et al.* \(2014\)](#) used a regression coefficient to adjust dissolved organic carbon data affected by an instrument/method change during a 25-year study. These examples illustrate not only why the goal of a QC program should be to provide the end user with as much information as possible to accept or reject a particular sample value ([Eischeid *et al.* 1995](#)), but also the importance of using and presenting QC data in publishing and presenting LT data. While there are published guidelines for users of LT data regarding QC information (e.g., [Porter & Callahan 1994](#)), ultimately presenting QC information depends on the data user. One study reported that nearly one-third of papers did not adequately describe their QA/QC procedures despite it being a requirement of the journal the papers were published in ([Kervin *et al.* 2013](#)). [Michener \(2015\)](#) highlighted the use of LT data sharing sites where the success or failure of the data was dependent on fully documenting the meta-data, which included measures of QA/QC. Some research studies do present QC methods and QC results (e.g., [Ludtke *et al.* 2000](#)), but one decades-old survey reported that many studies did not then use or report QC ([Bowser 1986](#)). The importance of quantifying uncertainty in lab studies was not fully recognized until 1978 when, due to the lack of reliable and comparable data, Good Laboratory Practice (GLP) regulations were introduced by the Food and Drug Administration and the Environmental Protection Agency (EPA) for toxicological research but later became more widely used ([Vijverberg & Cofino 1987](#); [Libes 1999](#)). Although part of GLP is the establishment of QA/QC protocols, EPA’s Methods for Chemical Analysis of Water and Wastes ([USEPA 1983](#)) did not have QA/QC protocols

until updated in 1993 (USEPA 1993). Because the widespread acceptance and use of QA/QC emerged in the last approximately 30 years, it is conceivable that the start of some LT sampling studies pre-dates this emergence and thus does not include QA/QC for the entire record.

Our aims were to investigate the use of QC data in the LT stream water quality literature and to present and apply QC information to a LT stream water chemistry data record to illustrate potential challenges in interpreting time trends without QC information. We used a database of scientific literature to search for publications that presented LT stream water quality data and determined whether QC information was considered in presenting or interpreting the results. We hypothesized that, despite the accepted importance of utilizing this information, the majority of studies would not use this information, regardless of record length presented. We also examined 46 years of stream water chemistry results at Coweeta Hydrologic Laboratory (CHL) in western North Carolina and applied QC results of bias and mdl obtained from the quarterly QC checks. We tested whether the observed interpretation of trends would change after QC incorporation.

METHODS

Survey of quality control

Using the Web of Science (WOS), a subscription-based website providing access to multiple databases, we conducted a search for papers studying nutrients in freshwater streams using the following search parameters: stream AND [sulfate OR nitrate OR nutrient] NOT [mercury OR sewage OR lake OR pesticide OR biological OR hormone]. We conducted a second search adding the parameter 'long-term' to the above parameters. Using WOS search history tools, we combined the results of the two searches. Publications within the fields of environmental science, forestry, agriculture, ecology, water resources, and multidisciplinary science were used. The time span was 1900 to 2018. The search yielded 3,738 papers. We further refined the subsampled papers to include only LT studies with data collection over 5 years or more involving natural freshwater streams. Given the volume of results, for time and efficiency purposes, we subsampled these results using the order of relevance as assigned by the WOS and examined 1,465 for review. From the 1,465 papers we examined, we found 591 papers that were not relevant and could not obtain a paper for 63 of the citations. For the remaining 811 papers, we found that 539 were not LT studies, and 272 were LT studies of 5 years or greater. Thus, our final sample size was 272. For these papers, we evaluated each for QC information, categorizing them into 'no QC reported', 'QC reported', or 'QC used', and we further noted their record or study length, categorizing them into '5–10 years', '11–15 years', '16–20 years', '21–25 years', '26–30 years', '31–35 years', '36–40 years', or '41+ years'. We used a broad definition for reporting QC results as citing any of the following: mdl – which could be stated as mdl or detection limit (DL), precision, accuracy, determination of ion balance equation, % recovery, or any statement, suggesting the data were evaluated using QC protocols.

To test whether the majority of papers either presented or used QC information, we computed binomial proportions, confidence limits, and tests for the 'no QC reported' category across all study lengths using PROC FREQ in SAS (SAS v9.4, Cary, NC, USA). We tested the null hypothesis that the population proportion of papers reporting no QC information equaled 50% at the $\alpha = 0.05$ level.

To test whether the QC reporting and/or use varied with study length, we modeled the frequency data as the main effect categories of study length and QC reporting status using a generalized logit model with a log-linear response function to transform the cell probabilities (PROC CATMOD, SAS v9.4, Cary, NC, USA). All main effects and interactions (if significant) were tested *post hoc* with custom contrast statements at the $\alpha = 0.05$ level.

Case study

We tested the impact of incorporating QC results on LT trends on stream chemistry data from CHL. CHL is a USDA Forest Service Experimental Forest located in Otto, NC, established in 1934. Stream chemistry measurements began in 1972, yielding 46 years of data during which time there were 19 analytical instruments, quantifying 12 chemical constituents for 15 watersheds. We focus on stream ammonium (NH_4) and nitrate (NO_3) nitrogen, and sulfate (SO_4), both important to understanding the forest ecosystem response to acid deposition as seen in the LT changes in atmospheric sulfur deposition (USEPA 2015) and the significant nitrogen response to clear-cutting in WS 7 (Swank *et al.* 2014). Although we focus on nitrogen and SO_4 , other analytes (Cl^- , PO_4 , K, Na, Ca, Mg, SiO_2 , and H^+) measured at CHL have been examined. The stream concentrations for these analytes (other than PO_4) are at levels in which the mdl was not a factor and when graphed we did

not visually find differences in values of bias-corrected and not-corrected. PO_4 , with an average stream concentration of 0.004 mg L^{-1} and an average mdl of 0.007 mg L^{-1} , was often below the sensitivity of the instrument and not used.

The CHL has a well-established QC program for stream sample analysis that includes daily development of a calibration curve for analyses on each instrument, verified with certified calibrants and calibrant checks (USDA FS 2017). The instrument's performance is further tested using certified QC samples (e.g., NSI Solutions, Raleigh, NC, USA and Environmental Resources Associates, Golden, CO, USA). Certified QC samples were purchased and analyzed annually through the mid-1980s with the frequency of analyses per year increasing until the mid-90s when quarterly QC checks were adopted. Three solutions representing a range of dilutions were made from the purchased QC sample stock; all diluted samples were analyzed in triplicate. The diluted QC sample concentrations closely resemble CHL stream chemistry for all analytes. From these triplicate results, measures of bias, precision, and the mdl were determined. See Table 1, measures of QC, for definitions and formulas. Because so few measurements of mdl were made in the early years, we used the triplicate values from the observed values on the QC sample to calculate the mdl ($n - 1$ observations = 2) until December 2010 when more frequent mdl determinations from 10 observations were made. An averaged mdl was then determined from the multiple mdls determined throughout the year. This is a modified version of the method recommended by the EPA (USEPA 2016).

We used stream chemistry data from WS 2 and WS 7. WS 2 is a low-elevation watershed that serves as a reference for WS 7, which was clear-cut and cable-logged in 1976–77 (Swank *et al.* 2014). We evaluated the effects of changing instrumentation and analytical methods over time by looking at changes in mdl, bias, and precision. We (1) compared export values using three censoring techniques; (2) compared the bias-corrected vs. non-corrected data over the entire time series to see if trends differed; (3) evaluated if the bias was changing over time; and (4) compared the bias-corrected vs. non-corrected data within each analytical method (analytical method refers to either chemical method or instrument used) to see if there was a difference.

We determined how watershed dissolved inorganic nitrogen export estimates were affected when the mdl was taken into account using three methods for censoring values data below mdl, two of which were recommended by Farnham *et al.* (2002). Export was estimated using the weekly cumulative streamflow calculated from 5-min stage measurements multiplied by the weekly concentrations of NH_4 and NO_3 nitrogen in a stream grab sample. If the weekly stream concentrations were below mdl, we either (1) kept all values; (2) removed the values below mdl and assigned them a zero value; or (3) assigned them a value equal to one-half the mdl value. We then summed weekly export values to annual nitrogen export and tested whether the method for treating data below mdl resulted in a significantly different annual export estimate. Because our inference space was limited to two watersheds (WS 2 and WS 7), we used mdl treatment as the main factor with three levels and year as a replicate. Means were tested using a general linear model with year as a repeated factor (PROC GLM, SAS v9.4, Cary, NC, USA). If main effects were significant, level means were tested using a *post hoc* test with Tukey's adjustment at the $\alpha = 0.05$ level.

We compared the bias-corrected vs. non-corrected data for stream chemistry over the entire time series to see if trends differed. We adjusted weekly stream chemistry values by multiplying the weekly value by the QC bias and then subtracted (which if negative bias would be added) this error from the weekly value. After visually examining

Table 1 | Measures of quality control

Term	Other designations	Definition	Equation
Bias	Measurement error, percent relative error	How close the mean of observed values is to the actual value.	$((\text{true value} - \text{mean of observed values}) / \text{true value}) \times 100$
Precision	Relative standard error (RSD), %RSD	How reproducible the observed value is.	$(\text{standard deviation of observed values} / \text{mean of observed values}) \times 100$
Method detection limit	mdl	The minimum concentration of a substance that can be measured and reported with 99% confidence that the analyte concentration is greater than zero (USEPA 2016).	$\text{standard deviation of observed values} \times \text{student } t \text{ for } n - 1 \text{ at } 99\% \text{ confidence level. } n = \text{number of samples}$

the graph displaying annual means that were both adjusted for bias and unadjusted, sulfate (SO_4) appeared to differ when bias was incorporated vs. not. We thus statistically tested whether the linear trends were significantly different using a general linear model and an explicit test of slopes and intercepts using a dummy variable to denote inclusion of bias or not (PROC GLM, SAS v9.4, Cary, NC, USA) (Zarnoch 2009). We tested for autocorrelation among residuals of this model to lag 4 using the Durban-Watson test (PROC AUTOREG, SAS v9.4, Cary, NC, USA) and verified no significant autocorrelation among residuals. We also tested whether the bias differed among instruments over time (PROC ANOVA, SAS v9.4, Cary, NC, USA). Lastly, we separated the data by the analytical method – segmented flow analysis (SFA) vs. ion chromatograph (IC) – and statistically tested whether the linear trends were significantly different using a general linear model and an explicit test of slopes and intercepts using a dummy variable to denote inclusion of bias or not (PROC GLM, SAS v9.4, Cary, NC, USA) (Zarnoch 2009). For linear trends of the bias-corrected vs. non-corrected data for stream chemistry that were significantly different, we determined at what value of bias and year the points diverge by setting the equations of regression equal to one another.

RESULTS

Survey of quality control

The majority of LT water quality studies did not report any QC information (Figure 1). Of the 272 papers, 33% (90) had some form of QC, with 12% (32) using the QC finding in the data results. The proportion of the sampled papers that did not report or use QC information was 66% (182). This proportion differed significantly from the null hypothesis of 50% ($z = 5.58$, $P < 0.001$), leading us to reject it.

The proportion of papers not reporting QC data was greater than the proportion reporting and/or using QC information (QC category effect, $\chi^2 = 29.86$, $df = 1$, $P < 0.001$), as noted above; and that the frequency of studies using QC information decreased with increasing study length (study length effect, $\chi^2 = 103.11$, $df = 7$, $P < 0.001$). Studies that spanned more than 30 years were exceptionally rare, while those spanning 5–10 years were most common. Overall, the length of study did not affect the number reporting QC (no interaction effect, $\chi^2 = 3.99$, $df = 7$, $P = 0.78$), leading us to accept our hypothesis that the majority of studies would not use this information, regardless of record length presented. Many of the papers citing QC information (50) and those using QC in the presentation of results (24) used mdl as the QC cited. The mdl used in the results was either: (1) shown graphically, (2) modified the measured concentration to one-half the mdl, (3) set the concentration to 0, or (4) the values were excluded. Of the 90 papers with QC results, eight merely referenced QC information with phrases like ‘all QC protocols were followed’. Twenty of the papers reported changes in method, instrumentation, or personnel during the period of measurement.

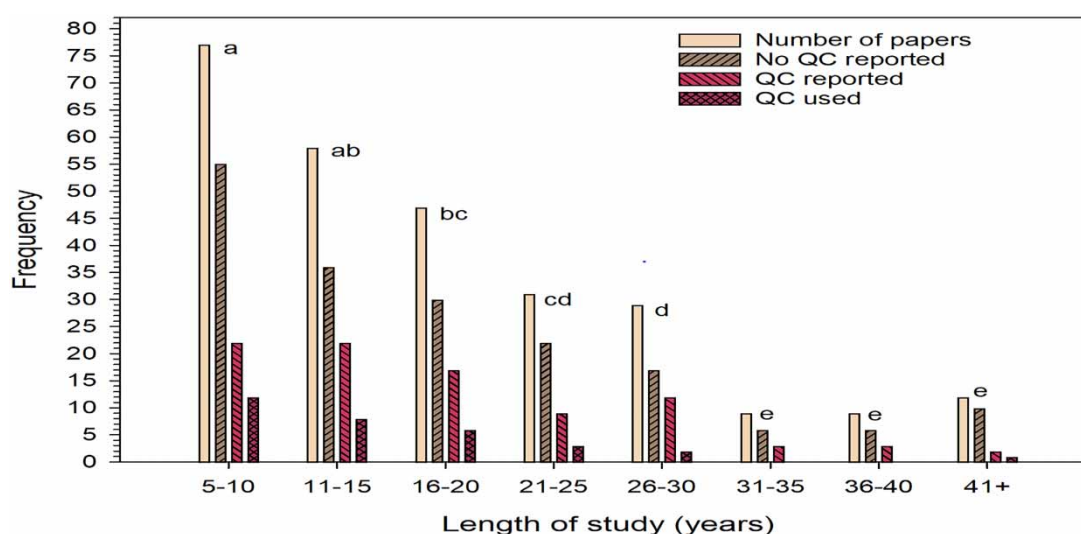


Figure 1 | Results of the literature search. Frequencies of papers found and frequency of those papers that had no QC reported, QC reported, and QC used in the result. Year categories not sharing the same lowercase letters were significantly different at the $\alpha = 0.05$ level.

Case study

We found that on average, the stream NH_4 concentrations in the weekly grab samples for both WS 2 and 7 were approximately equal (0.004 mg L^{-1}) and were below the mdl (average 0.006 mg L^{-1}) 58% of the time. We found that on average, the stream NO_3 concentration for WS 2 and 7 (average 0.008 and 0.069 mg L^{-1}) was below the mdl (average 0.006 mg L^{-1}) 28 and 7% of the time. An instrument change in 1994 for measuring NH_4 resulted in improved mdl (average 0.004 mg L^{-1}) and decreased the number of data points below the mdl. Prior to 1994, 78% (WS 2) and 76% (WS 7) of NH_4 values were below the mdl; after 1994, 38% (WS 2) and 41% (WS 7) of NH_4 values were below the mdl. The measurement of NO_3 also experienced a decrease in mdl with an instrument change (average before the change was 0.007 and 0.005 mg L^{-1} after the change). The percent below the mdl, however, increased for WS 2 (from 23 to 31%) and decreased for WS 7 (from 9.8 to 5.2%). The common methods for treating data below the mdl – keeping all instrument-recorded values, assigning the value of zero, and assigning values below the mdl to one-half the calculated mdl value – resulted in significantly different estimates of total annual nitrogen export for WS 2 (treatment effect, $F_{2,132} = 6.99$, $P = 0.001$), with the method of replacing values below the mdl with zero resulting in significantly lower export values than the other two methods, which did not differ from one another. The annual export of nitrogen for WS 2 averaged (SE) 66.9 (5.5) $\text{g ha}^{-1} \text{ yr}^{-1}$ for all data included (Figure 2(a)), 45.4 (4.9) $\text{g ha}^{-1} \text{ yr}^{-1}$ when values below detection were assigned the value of zero (Figure 2(b)), and 72.1 (7.1) $\text{g ha}^{-1} \text{ yr}^{-1}$ when one-half mdl was substituted for values below detection (Figure 2(c)). A different result was found when nitrogen export from WS 7 was examined (NS treatment effect), with all three methods producing similar export values regardless of the censoring method used for data below the mdl. The annual export of nitrogen for WS 7 averaged (SE) 831.4 (91.5) $\text{g ha}^{-1} \text{ yr}^{-1}$ with all data included (Figure 3(a)), 808.0 (92.4) $\text{g ha}^{-1} \text{ yr}^{-1}$ when values below the mdl were assigned the value of zero (Figure 3(b)), and 823.2 (92.1) $\text{g ha}^{-1} \text{ yr}^{-1}$ when one-half mdl was substituted for values less than mdl (Figure 3(c)).

Quarterly QC bias for stream SO_4 concentration increased when going from the SFA to the IC ($F_{3,111} = 10.60$, $P < 0.001$, Figure 4). The percent bias for the SO_4 QC (absolute error) increased from 0.22% (2.67%) to 3.98% (5.07%) but improved the analytical precision (%RSD) from 4.12 to 1.10. We found that SO_4 concentration decreased over the entire period of record from 1973 to 2017 in both watersheds (WS 2: $F_{3,86} = 3.58$, $P = 0.02$, $R^2 = 0.11$; WS 7: $F_{3,86} = 8.75$, $P < 0.001$, $R^2 = 0.20$) (Figure 5). When bias was applied, SO_4 concentration appeared to decline at a greater rate over time compared with data with no bias applied for both watersheds, but this was not statistically significant (comparison of slopes, WS 2: $F_{1,86} = 0.14$, $P = 0.71$; WS 7: $F_{1,86} = 0.52$, $P = 0.47$). When evaluating changes in the SO_4 concentration for data analyzed only on the segmented flow analyzer (SFA) using bias-corrected compared with non-bias-corrected data, the time trend did not differ for either watershed (comparison of slopes, WS 2: $F_{1,32} = 0.01$, $P = 0.92$; WS 7: $F_{1,32} = 0.06$, $P = 0.88$). However, when the ion chromatograph (IC) analysis was used, SO_4 concentration that was bias-corrected declined more rapidly in WS 2 compared with non-bias-corrected data, but not in WS 7 (comparison of slopes, WS 2: $F_{1,50} = 4.66$, $P = 0.04$; WS 7: $F_{1,50} = 3.14$, $P = 0.08$, Figure 6).

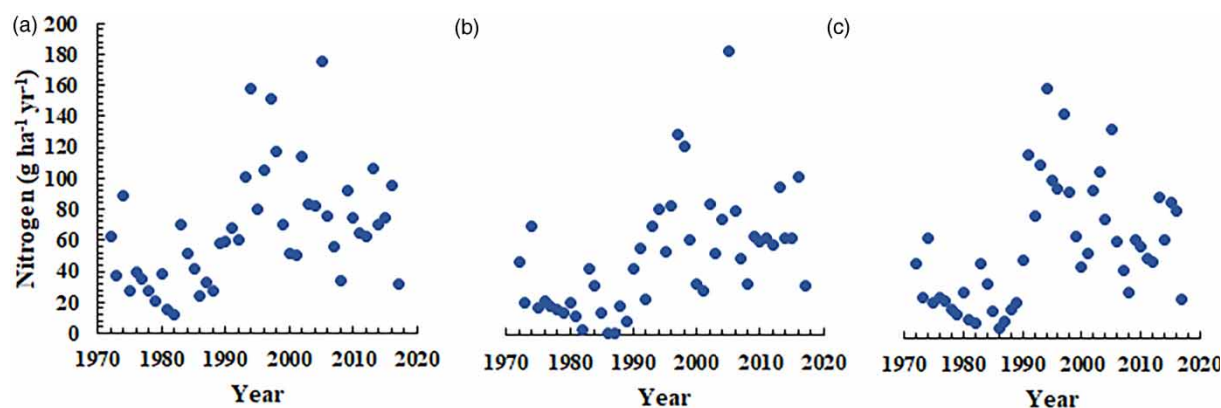


Figure 2 | Watershed 2 nitrogen yearly export showing: all values (a), values below detection assigned the value of zero (b), and assigning one-half mdl to values below detection (c). Mean annual export for nitrogen ($\text{g ha}^{-1} \text{ yr}^{-1}$): 66.9 (a), 45.4 (b), and 72.1 (c).

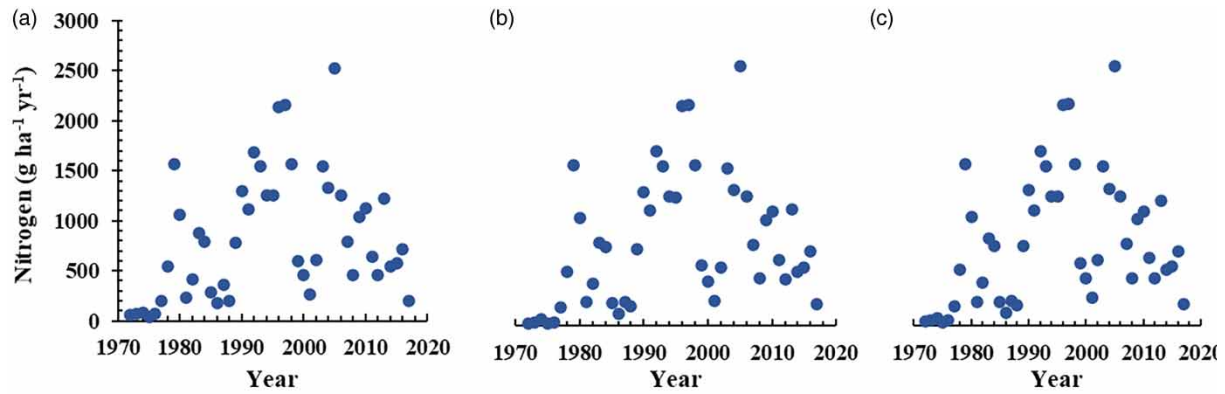


Figure 3 | Watershed 7 nitrogen yearly export showing: all values (a), values below detection assigned the value of zero (b), and assigning one-half md to values below detection (c). Mean annual export for nitrogen ($\text{g ha}^{-1} \text{yr}^{-1}$): 831.4 (a), 808.0 (b), and 823.2 (c).

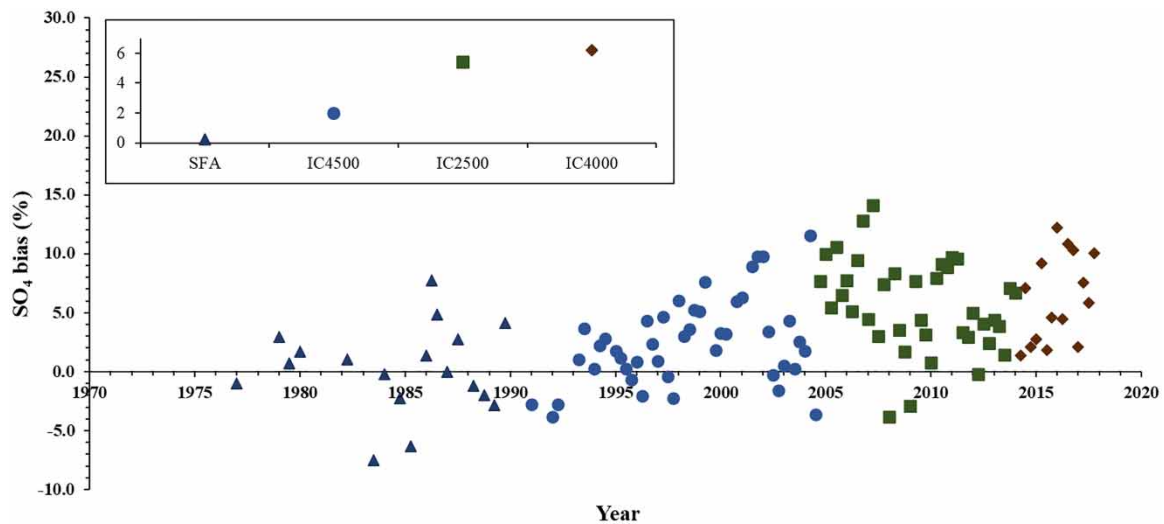


Figure 4 | Percent bias for SO_4 QC samples over time. Chart inset shows averaged bias by the instrument.

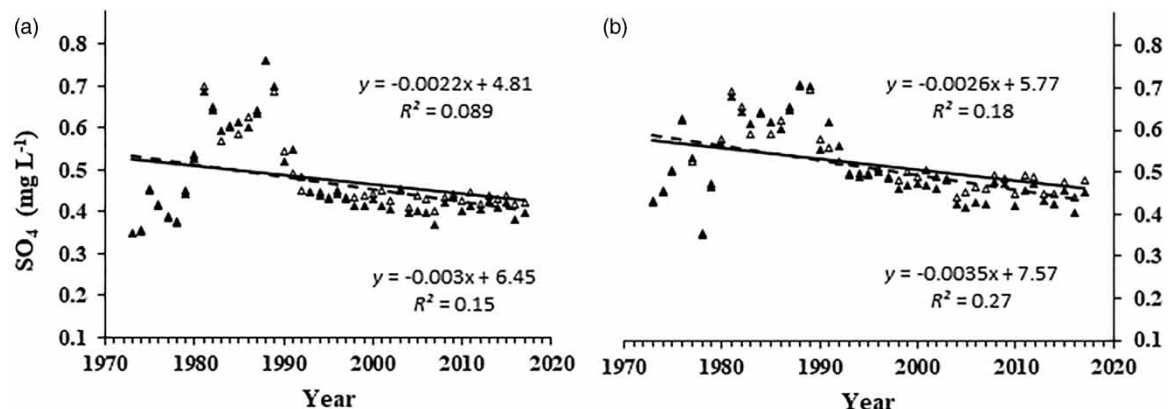


Figure 5 | For watershed 2 (a) and watershed 7 (b), average annual stream SO_4 concentration with (solid symbols) and without (no fill symbols) bias applied. Linear regression fit to stream SO_4 concentration with (dash line, regression on bottom) and without (solid line, regression on top) bias applied over time.

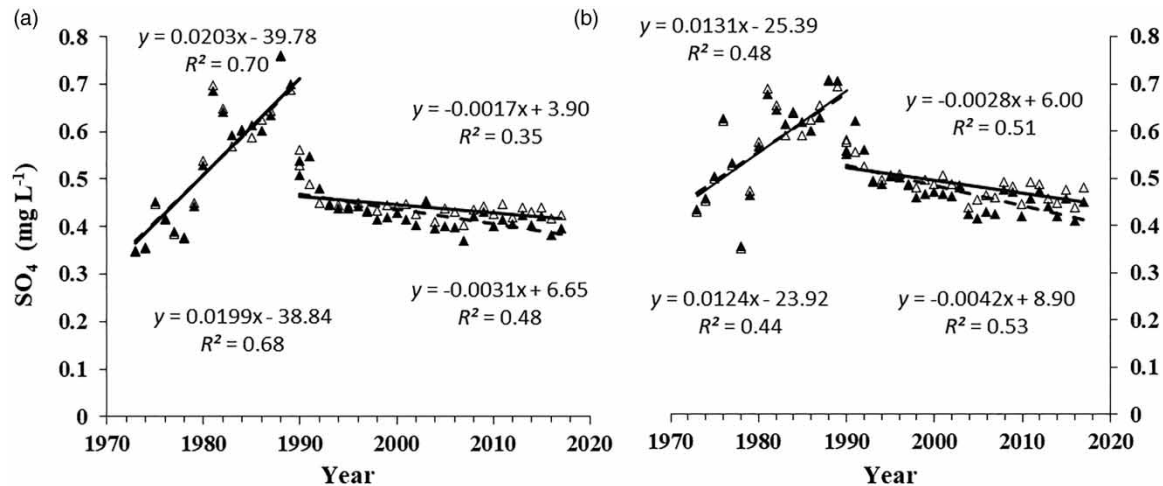


Figure 6 | Comparison of results when measuring SO_4 on the SFA (1973 to 1990) and on the IC (1990 to 2018). Shown are average annual stream SO_4 concentration WS 2 (a) and 7 (b) without (open symbols) and with (solid symbols) bias applied. Linear regression fit to stream SO_4 concentration without (solid line, regression on top) and with (dashed line, regression on bottom) bias applied.

Using the trendlines from [Figure 6\(a\)](#), we determined the point of intersect, IC only, to be at the year 1998 with a corresponding absolute bias value of 4.42% for SO_4 QC. We found the average absolute bias after 1998 was 5.64%. The absolute bias for all SO_4 QC determinations was above 5.0% (and above 10%) for the IC 40% (8%) of the time and 17% (0%) of the time for the SFA ([Figure 7](#)).

DISCUSSION

For small watersheds, [Harmel et al. \(2006\)](#) found four procedural categories for the uncertainty in measured water quality data: streamflow measurement, sample collection, sample preservation/storage, and laboratory analysis. We focused on laboratory analysis to investigate possible effects changing methods might have on the data. Our aim was to bring to the attention of LT data users the need to recognize how these changes over time can affect the trends established from the data. We assert that including QC results with LT data is essential to verify data validity and give the data user a full understanding of the results. We recognize there are other factors which can affect the analysis such as sample storage/preservation, filtration, and washing protocols. We did not investigate these for this study.

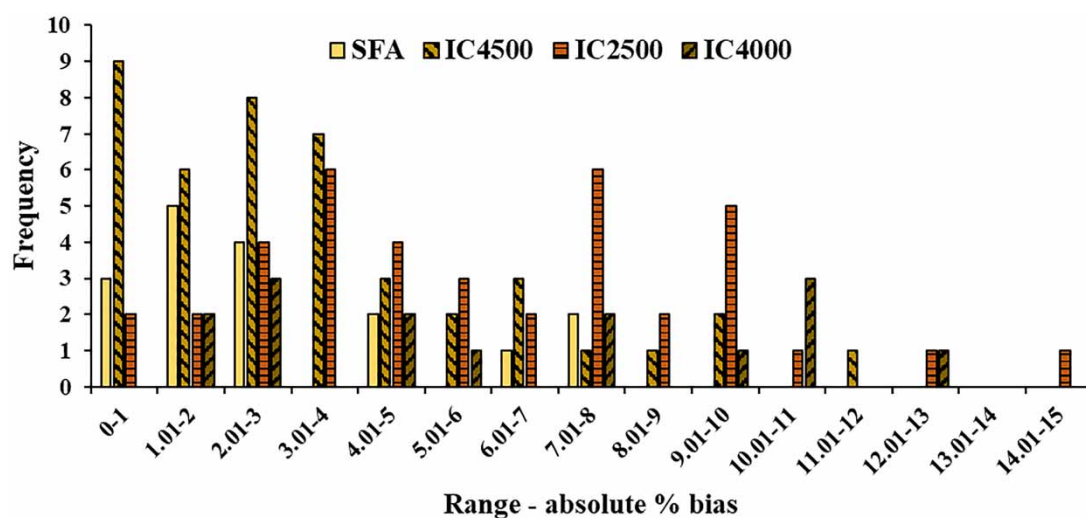


Figure 7 | Frequency of absolute bias for SO_4 QC samples by the instrument.

Survey of quality control

Our findings showed that QC results are not utilized in most published papers on LT stream chemistry, even though longer-term studies are more likely to go through changes in methods and instrumentation (e.g., [Alewell et al. 1999](#); [Bergfur et al. 2012](#); [Coble et al. 2018](#)). Our findings that only one-third of papers had some form of QC differed from [Kervin et al. \(2013\)](#) who found that nearly two-thirds of authors did describe QA/QC in the meta-data for ESA (Ecological Society of America) data archive. However, their study focused only on data papers submitted to ESA in which meta-data description is a submission requirement.

We found that papers most often reported mdl as a QC result. Studies in which the mdl was used in the result either showed mdl graphically, changed the measured concentration to one-half the mdl, used mdl to modify the measured concentration to a value of 0, or used it to exclude values (e.g., [Boy et al. 2008](#); [Snelder et al. 2017](#); [Stets et al. 2018](#)). Previous studies suggest that, for inclusion of censored data (data below the mdl), users either assign values an arbitrary fraction of the mdl, determine a maximum likelihood estimation, or used nonparametric methods (ranking the data) when the data does not follow the normal distribution ([Helsel 2005](#)). The United States Geological Survey National Water Quality Laboratory (NWQL) uses two concentration markers for reporting low concentration data: a long-term mdl (LT-mdl) and the laboratory reporting level (LRL, two times the LT-mdl) for minimizing the risk of critical measurement errors ([Childress et al. 1999](#)). The LT-mdl controls the false-positive error and the LRL controls the false-negative error. Of the LT papers, we reviewed some reported changes in instrumentation and/or method throughout the course of the study and found that some results were affected. Notably, [Rogora et al. \(2001\)](#), in a 21-year study, found that a change in the instrumentation/method affected chloride and sulfate results which resulted in some data not being used. [Coats et al. \(2016\)](#) worked with data that spanned 42 years and noted that 27 years of NO_3 data may have been affected by chemical interference from divalent cations resulting in inefficient recoveries. Others have reported that changes in instrumentation did not affect the data (e.g., [Hruska et al. 2009](#); [Stackpoole et al. 2017](#)).

Case study

We found that a change in method and instrumentation affected the NH_4 and SO_4 stream chemistry data measured at CHL: the mdl for ammonium decreased over time, and sulfate data had an increase in bias. For CHL data we found in 1994 an instrument change in ammonium analysis resulted in a lower mdl, with a subsequent decrease in the number of reported values below detection. A change in method and instrumentation in 1990 for measuring sulfate resulted in increased bias but greater precision for sulfate data. Although there were measurable differences due to the change in instrumentation for both NH_4 and SO_4 , the data trends did not change.

CHL data showed that censored data affected WS 2 nitrogen export but not WS 7, even though NH_4 was often below detection for both watersheds. The WS 7 export contribution from NO_3 minimized the NH_4 contribution resulting in essentially no effect on the nitrogen export calculated from censored data. Though the total nitrogen export is small and therefore may not be biologically or ecologically important, our result illustrates the relevance of QC in LT studies involving trace contaminants such as lead. For instance, over a 10-year period lead had been measured in the San Francisco Bay with changes in instrumentation leading to an mdl for lead of $0.77 \pm 0.29 \text{ ng kg}^{-1}$ ([Squire et al. 2002](#)). Because data provided to investigators from CHL are not censored (unless obvious contamination is found, or there is a known issue with the analysis), a data user must examine the QC results to fully understand the trends in concentration and the export loads that are derived from the data. In a cross-site comparison of experimental forest stream chemistry data, which included CHL data, [Argerich et al. \(2013\)](#) found that the mdl changed over time. To remove the influence of changing mdl, the data were censored, using the highest mdl for the full period, so the changing mdl did not influence the trend. This seems far from ideal since we have shown that the mdl for NH_4 and NO_3 changed for the better with changing instrumentation. This may represent a limitation and perhaps a caution to investigators using the archived CHL database. Although CHL data report yearly mdl values, data below detection are not flagged.

We found that CHL SO_4 data were affected by changes in the analytical method. Our results showed an increasing bias in SO_4 data over time, particularly when changing from SFA to ion chromatography analysis. Our results showed that after 1998, the non-bias-corrected data and the bias-corrected data diverge. The manufacturer of the certified QC sample lists the range at $\pm 10\%$ determined through performance proficiency testing ([ERA 2021](#)). With an average absolute bias of 5.1% for QC samples analyzed on the ion chromatograph (IC), it

shows that although the SO_4 concentration that was bias-corrected declined more rapidly in WS 2 compared with non-bias-corrected data, these data are within CHL QC parameters. We also found the instruments at CHL tended to measure error in one direction. Because the bias was in a positive direction, this suggests that the data were under reported (observed values less than actual). For example, a bias of 5% suggests that a result of 0.95 mg L^{-1} would have had an actual value of 1.00 mg L^{-1} , whereas a bias of -5% suggests that a result of 1.05 mg L^{-1} would have had an actual value of 1.00 mg L^{-1} . This can be helpful information to a data user, especially if the bias is higher.

SFA is a colorimetric analysis method using methyl thymol blue (MTB) and BaCl_2 in which BaSO_4 is formed and the excess Ba reacts with MTB. The uncomplexed MTB is grey which is detected by a spectrophotometer. The uncomplexed MTB is equal to the amount of SO_4 present. The IC utilizes suppression chemistry, the separation capacity of an ion exchange column, and the detection signal from a conductivity detector. Compared with the SFA, the IC method change gave a smoother data series, but the error in bias increased with the instrument change and showed a greater decline in SO_4 when the bias was applied. This was the most significant change in instrumentation and method over the 46-year period of record. Other studies have reported smoother data series, better results, and improved accuracy when switching to IC analysis from SFA (e.g., Fishman & Pyen 1979; Alewell 1993). In a study on sulfate sources at Hubbard Brook experimental forest, Alewell *et al.* (1999), working with SO_4 data measured over a time period that used three different methods, used only data measured by IC to avoid potential confounding factors associated with the difference in analytical technique; and Reynolds *et al.* (2004) excluded SO_4 data obtained pre-ion chromatograph because of lack of sensitivity using the BaCl_2 turbidimetric method. However, McSwain *et al.* (1974), former CHL lab manager, reported a bias of only 1.7% with the BaCl_2 SFA method, and Bachman (1987) concluded that changing methods did not cause a significant difference when comparing major anions (includes SO_4) analyzed on the SFA and the IC.

Changing methods or instrumentation during the course of an LT study are examples of why a data user should require the results of QC accompany the data. Overall, the effect of QC results on CHL data may be small but that fact was verified through those results. The examples from our case study illustrate that there may be more information needed than just the data to fully evaluate the findings. A requirement to include QC data with the meta-data would be a solution. Argerich *et al.* (2013) note that it is necessary for good management and protection of water resources to realize whether or not nutrients are changing over time. These possibly small changes might not be realized depending on if and how the data were censored.

CONCLUSION

Overall, we have shown that data can be affected by instrument and method changes that occur during LT studies, and we have shown that not enough authors are reporting data quality regardless of study length. It is not our suggestion to apply bias to the data, or how to work with censored data, but for users of LT data to recognize how these metrics affect the data and the data trends. It is our contention that more published works should include measures of quality to give the reader a fuller understanding of the conclusion(s) drawn from the study and validate the data used. These measures should include yearly values of: bias, mdl, and precision.

ACKNOWLEDGEMENTS

We are grateful to Drs. Carl Trettin and Jack R. Webster for providing helpful comments on a previous version of this manuscript. This study was supported by the U.S. Department of Agriculture (USDA) Forest Service, Southern Research Station, and the National Science Foundation (NSF) awards, DEB-0218001, DEB-0823293, DEB-1226983, DEB-1440485, and DEB-1637522 from the Long-Term Ecological Research (LTER) Program to the Coweeta LTER. Any opinions, findings, conclusion, or recommendations expressed in the material are those of the authors and do not necessarily reflect the views of the USDA or NSF.

AUTHOR CONTRIBUTIONS

C.L.B.: conceptualization, methodology, investigation, formal analysis, writing – original draft, supervision.
C.F.M.: formal analysis, writing – review and editing. J.D.K.: resources, writing – review and editing.

COMPETING INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

DATA AVAILABILITY STATEMENT

All relevant data are available from <https://www.srs.fs.usda.gov/coweeta/tools-and-data/>.

REFERENCES

- Alewell, C. 1993 Effects of organic sulfur compounds on extraction and determination of inorganic sulfate. *Plant and Soil* **149**, 141–144.
- Alewell, C., Mitchell, M. J., Likens, G. E. & Krouse, H. R. 1999 Sources of stream sulfate at the Hubbard Brook Experimental Forest: long-term analyses using stable isotopes. *Biogeochemistry* **44** (3), 281–299.
- APHA/AWWA/WEF 2012 *Standard Methods for the Examination of Water and Wastewater*, 22nd edn. American Public Health Association/American Water Works Association/Water Environment Federation, Washington DC, USA.
- Argerich, A., Johnson, S. L., Sebestyen, S. D., Rhoades, C. C., Greathouse, E., Knoepp, J. D., Adams, M. B., Likens, G. E., Campbell, J. L., McDowell, W. H., Scatena, F. N. & Ice, G. G. 2013 Trends in stream nitrogen concentrations for forested reference catchments across the USA. *Environmental Research Letters* **8**, 014039.
- Bachman, S. R. 1987 *A Comparison of ion Chromatography and Automated Colorimetry for the Determination of Major Anions in Precipitation*. Illinois State Water Survey Division, University of Illinois, IL, USA.
- Beard, G. R., Scott, W. A. & Adamson, J. K. 1999 The value of consistent methodology in long-term environmental monitoring. *Environmental Monitoring and Assessment* **54**, 239–258.
- Bergfur, J., Demars, B. O. L., Stutter, M. I., Langan, S. J. & Friberg, N. 2012 The Tarland Catchment Initiative and its effect on stream water quality and macroinvertebrate indices. *Journal of Environmental Quality* **41** (2), 314–321.
- Bowser, C. J. 1986 Historic data sets: Lessons from the past, lessons from the future. In: *Research Data Management in the Ecological Sciences* (Michener, W. K., ed.), Belle W. Baruch Library in Marine Science No. 16. University of South Carolina Press, Columbia, SC, USA, pp. 155–179.
- Boy, J., Valarezo, C. & Wilcke, W. 2008 Water flow paths in soil control element exports in an Andean tropical Montane forest. *European Journal of Soil Science* **59** (6), 1209–1227.
- Burt, T. P., Howden, N. J. K., Worrall, F. & Whelan, M. J. 2008 Importance of long-term monitoring for detecting environmental change: lessons from a lowland river in south east England. *Biogeosciences* **5**, 1529–1535.
- 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*. Gen. Tech. Rep. NE-275. U.S. Department of Agriculture, Forest Service, Northeastern Research Station, Newtown Square, PA, USA, p. 52.
- Campbell, J. L., Yanai, R. D., Green, M. B., Likens, G. E., See, C. S., 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**, e01299.
- Childress, C. J. O., Foreman, W. T., Connor, B. F. & Maloney, T. J. 1999 *New Reporting Procedures Based on Long-Term Method Detection Levels and Some Considerations for Interpretations of Water-Quality Data Provided by the US Geological Survey National Water Quality Laboratory*. U.S. Geological Survey Open-File Report 99193, p. 19.
- Coats, R., Lewis, J., Alvarez, N. & Arneson, P. 2016 Temporal and spatial trends in nutrient and sediment loading to Lake Tahoe, California-Nevada, USA. *JAWRA Journal of the American Water Resources Association* **52** (6), 1347–1365.
- Coble, A. A., Wymore, A. S., Shattuck, M. D., Potter, J. D. & McDowell, W. H. 2018 Multiyear trends in solute concentrations and fluxes from a suburban watershed: evaluating effects of 100-year flood events. *Journal of Geophysical Research: Biogeosciences* **123** (9), 3072–3087.
- Eischeid, J. K., Baker, C. B., Karl, T. R. & Diaz, H. F. 1995 The quality control of long-term climatological data using objective data analysis. *Journal of Applied Meteorology* **34**, 2787–2795.
- Environmental Resource Associates, Incorporated (ERA) 2021 Golden Colorado. Available from: <https://www.eraqc.com/> (accessed 2 March 2021).
- Evans, C. D., Reynolds, B., Hinton, C., Hughes, S., Norris, D., Grant, S. & Williams, B. 2007 Effects of decreasing acid deposition and climate change on acid extremes in an upland stream. *Hydrology and Earth System Sciences Discussions* **4**, 2901–2944.
- Farnham, I. M., Singh, A. K., Stetzenbach, K. J. & Johannesson, K. H. 2002 Treatment of nondetects in multivariate analysis of groundwater geochemistry data. *Chemometrics and Intelligent Laboratory Systems* **60** (1–2), 265–281.
- Fishman, M. J. & Pyen, G. 1979 *Determination of Selected Anions in Water by Ion Chromatography* (No. 79–101). U.S. Geological Survey, Lakewood, CO, US.
- González, A. G. & Herrador, M. Á. 2007 A practical guide to analytical method validation, including measurement uncertainty and accuracy profiles. *TrAC Trends in Analytical Chemistry* **26** (3), 227–238.
- 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**, 689–701.

- Helsel, D. R. 2005 *Nondetects and Data Analysis. Statistics for Censored Environmental Data*. Wiley-Interscience, Denver, CO, US.
- Hruska, J., Krám, P., McDowell, W. H. & Oulehle, F. 2009 Increased dissolved organic carbon (DOC) in Central European streams is driven by reductions in ionic strength rather than climate change or decreasing acidity. *Environmental Science & Technology* **43** (12), 4320–4326.
- Karl, T. R., Tarpley, J. D., Quayle, R. G., Diaz, H. F., Robinson, D. A. & Bradley, R. S. 1989 The recent climate record: what it can and cannot tell us. *Reviews of Geophysics* **27** (3), 405–430.
- Kervin, K., Michener, W. & Cook, R. 2013 Common errors in ecological data sharing. *Journal of eScience Librarianship* **2** (2), 3–16.
- Kroll, C. N. & Stedinger, J. R. 1996 Estimation of moments and quantiles using censored data. *Water Resources Research* **32** (4), 1005–1012.
- Lehrter, J. C. & Cebrian, J. 2010 Uncertainty propagation in an ecosystem nutrient budget. *Ecological Applications* **20**, 508–524.
- Libes, S. M. 1999 Learning quality assurance/quality control using US EPA techniques. An undergraduate course for environmental chemistry majors. *Journal of Chemical Education* **76** (12), 1642.
- Lindenmayer, D. B., Likens, G. E., Krebs, C. J. & Hobbs, R. J. 2010 Improved probability of detection of ecological “surprises”. *Proceedings of the National Academy of Sciences* **107**, 21957–21962.
- Ludtke, A. S., Woodworth, M. T. & Marsh, P. S. 2000 *Quality-Assurance Results for Routine Water Analyses in US Geological Survey Laboratories, Water Year 1998. Water-Resources Investigations Report 00-4176*. U.S. Department of the Interior, U.S. Geological Survey, Water Resources Division, Denver, CO, USA, p. 198.
- McSwain, M. R., Watrous, R. J. & Douglass, J. E. 1974 Improved methyl thymol blue procedure for automated sulfate determinations. *Analytical Chemistry* **46** (9), 1329–1331.
- Meyer, J. L., Webster, J., Knoepp, J. & Benfield, E. F. 2014 Dynamics of dissolved organic carbon in a stream during a quarter century of forest succession. In: *Long-term Response of a Forest Watershed Ecosystem: Clearcutting in the Southern Appalachian* (Swank, W. T. & Webster, J., eds). Oxford University Press, New York, NY, USA, pp. 102–117.
- Michener, W. K. 2015 Ecological data sharing. *Ecological Informatics* **29**, 33–44.
- Michener, W. K., Porter, J., Servilla, M. & Vanderbilt, K. 2011 Long term ecological research and information management. *Ecological Informatics* **6** (1), 13–24.
- Porter, J. H. & Callahan, J. T. 1994 Circumventing a dilemma: Historical approaches to data sharing in ecological research. In: *Environmental Information Management and Analysis: Ecosystem to Global Scales* (Michener, W. K., Brunt, J. W. & Stafford, S. G., eds). Taylor & Francis, Bristol, PA, USA, pp. 193–202.
- Reynolds, B., Stevens, P. A., Brittain, S. A., Norris, D. A., Hughes, S. & Woods, C. 2004 Long-term changes in precipitation and stream water chemistry in small forest and moorland catchments at Beddgelert Forest, north Wales. Available from: <https://hal.archives-ouvertes.fr/hal-00304935/> (accessed 06 July 2020).
- Rogora, M., Marchetto, A. & Mosello, R. 2001 Trends in the chemistry of atmospheric deposition and surface waters in the Lake Maggiore catchment. *Hydrology and Earth System Sciences Discussions* **5** (3), 379–390.
- Snelder, T. H., McDowell, R. W. & Fraser, C. E. 2017 Estimation of catchment nutrient loads in New Zealand using monthly water quality monitoring data. *Journal of the American Water Resources Association* **53** (1), 158–178.
- Squire, S., Scelfo, G. M., Revenaugh, J. & Flegal, A. R. 2002 Decadal trends of silver and lead contamination in San Francisco Bay surface waters. *Environmental Science & Technology* **36** (11), 2379–2386.
- Stackpoole, S. M., Stets, E. G., Clow, D. W., Burns, D. A., Aiken, G. R., Aulenbach, B. T., Creed, I. F., Hirsch, R. M., Laudon, H., Pellerin, B. A. & Striegl, R. G. 2017 Spatial and temporal patterns of dissolved organic matter quantity and quality in the Mississippi River Basin, 1997–2013. *Hydrological Processes* **31** (4), 902–915.
- Stets, E. G., Lee, C. J., Lytle, D. A. & Schock, M. R. 2018 Increasing chloride in rivers of the conterminous US and linkages to potential corrosivity and lead action level exceedances in drinking water. *Science of the Total Environment* **613**, 1498–1509.
- Sullivan, T. J., Herlihy, A. T., Lawrence, G. B. & Webb, J. R. 2012 *USDA Forest Service National Protocols for Sampling Air Pollution- Sensitive Waters*. Gen. Tech. Rep. RMRS-GTR-278WWW. U.S. Department of Agriculture, Forest Service, Rocky Mountain Research Station, Fort Collins, CO, USA, p. 334.
- Swank, W. T., Knoepp, J. D., Vose, J. M., Laseter, S. N. & Webster, J. R. 2014 Response and recovery of water yield and timing, stream sediment, abiotic parameters, and stream chemistry following logging. In: *Long-Term Response of a Forest Watershed Ecosystem: Clearcutting in the Southern Appalachians (The Long-Term Ecological Research Network Series)* (Swank, W. T. & Webster, J. R., eds). Oxford University Press, Oxford, UK, pp. 36–56.
- Taylor, J. K. 1987 *Quality Assurance of Chemical Measurements*. CRC Press, Boca Raton, FL, US.
- U.S. Department of Agriculture Forest Service (USDA FS), Coweeta Hydrologic Laboratory 2017 *Coweeta Quality Assurance Protocol*, p. 21. Unpublished manual. Available from: <https://www.srs.fs.usda.gov/coweeta/tools-and-data/> (accessed 29 September 2020).
- U.S. Environmental Protection Agency (USEPA) 1983 *Methods for Chemical Analysis of Water and Wastes*. EPA 600/4-79-020. Office of Research and Development, Washington, DC, USA, p. 491.
- U.S. Environmental Protection Agency (USEPA) 1993 *Methods for the Determination of Inorganic Substances in Environmental Samples*. EPA 600/R93-100. Environmental Monitoring Systems Laboratory Office of Research and Development, Cincinnati, OH, USA, p. 172.

- U.S. Environmental Protection Agency (USEPA) 2015 *Sulfur Dioxide Trends*. Air Quality Analysis Group, U.S. EPA Office of Air Quality Planning and Standards. Available from: <https://www.epa.gov/air-trends/sulfur-dioxide-trends> (accessed 18 February 2021).
- U.S. Environmental Protection Agency (USEPA) 2016 *Definition and Procedure for the Determination of the Method Detection Limit, Revision 2*. EPA 821-R-16-006. U.S. EPA Office of Water, Washington, DC, USA, p. 6. EPA 600/4-79-020. Office of Research and Development, Washington, DC, 491 p.
- Vijverberg, F. A. J. M. & Cofino, W. P. 1987 *Control Procedures: Good Laboratory Practice and Quality Assurance*. Public Works Department, The Hague, The Netherlands, p. 37.
- Walther, B. A. & Moore, J. L. 2005 [The concepts of bias, precision and accuracy, and their use in testing the performance of species richness estimators, with a literature review of estimator performance](#). *Ecography* **28**, 815–829.
- Whitehead, P. G., Wilby, R. L., Battarbee, R. W., Kernan, M. & Wade, J. 2009 [A review of the potential impacts of climate change on surface water quality](#). *Hydrological Sciences Journal* **54**, 101–123.
- 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**, 1793–1805.
- Zarnoch, S. J. 2009 *Testing Hypotheses for Differences Between Linear Regression Lines*. Res. Note SRS-17. U.S. Department of Agriculture, Forest Service, Southeastern Forest Experiment Station, Asheville, NC, USA, p. 20.

First received 2 November 2020; accepted in revised form 14 May 2021. Available online 28 May 2021