LIMNOLOGY
and
OCEANOGRAPHY: METHODS

Limnol. Oceanogr.: Methods 17, 2019, 639–649

doi: 10.1002/lom3.10338

Comparison of total nitrogen data from direct and Kjeldahl-based approaches in integrated data sets

Emily H. Stanley,[1]* **Shirley Rojas-Salazar**,[2] **Noah R. Lottig**,[3] **Erin M. Schliep**,[2] **Christopher T. Filstrup**,[4] **Sarah M. Collins**[5]

[1]Center for Limnology, University of Wisconsin-Madison, Madison, Wisconsin
[2]Department of Statistics, University of Missouri, Columbia, Missouri
[3]Trout Lake Station, Center for Limnology, University of Wisconsin, Boulder Junction, Wisconsin
[4]Natural Resources Research Institute, University of Minnesota Duluth, Duluth, Minnesota
[5]Department of Zoology and Physiology, University of Wyoming, Laramie, Wyoming

Abstract

There are multiple protocols for determining total nitrogen (TN) in water, but most can be grouped into direct approaches (TN-d) that convert N forms to nitrogen-oxides (NO_x) and combined approaches (TN-c) that combine Kjeldahl N (organic N +NH_3) and nitrite+nitrate (NO_2+NO_3-N). TN concentrations from these two approaches are routinely treated as equal in studies that use data derived from multiple sources (i.e., integrated data sets) despite the distinct chemistries of the two methods. We used two integrated data sets to determine if TN-c and TN-d results were interchangeable. Accuracy, determined as the difference between reported concentrations and the most probable value (MPV) of reference samples, was high and similar in magnitude (within 3.5–4.5% of the MPV) for both methods, although the bias was significantly smaller at low concentrations for TN-d. Detection limits and data flagged as below detection suggested greater sensitivity for TN-d for one data set, while patterns from the other data set were ambiguous. TN-c results were more variable (less precise) by many measures, although TN-d data included a small fraction of notably inaccurate results. Precision of TN-c was further compromised by propagated error, which may not be acknowledged or detectable in integrated data sets unless complete metadata are available and inspected. Finally, concurrent measures of TN-c and TN-d in lake samples were extremely similar. Overall, TN-d tended to be slightly more accurate and precise, but similarities in accuracy and the near 1 : 1 relationship for concurrent TN-d and TN-c measurements support careful use of data interchangeably in analyses of heterogeneous, integrated data sets.

Total nitrogen (TN) determination is a standard activity among diverse aquatic research and monitoring programs because nitrogen influences several processes and reflects prevailing conditions in aquatic ecosystems. A variety of methods are used to measure TN concentrations in water samples, but most commonly used protocols are variations on two basic analytical approaches: (1) quantification of TN as a nitrogen oxide (NO_x) form (most commonly as NO_2+NO_3-N) after sample oxidation, and (2) total Kjeldahl nitrogen (TKN) methods that convert N in the trivalent state (broadly interpreted as organic N + NH_3-N) to NH_4-N for analysis, combined with a separate determination of NO_2+NO_3-N. There are strengths and weaknesses associated with each of these approaches that favor one method over another depending on prevailing environmental conditions or sample type (e.g., *see* D'Elia et al. 1977; Saéz-Plaza et al. 2013). For example, high NO_2+NO_3-N concentrations that may be common in settings such as agriculturally dominated landscapes (Stanley and Maxted 2008) can interfere with the accuracy of TN measurements by the Kjeldahl method (Schluter 1977). Alternatively, methods based on sample oxidation may not be effective in converting some classes of organic N compounds to NO_x, especially in the presence of high sediment concentrations (Bronk et al. 2000; Maher et al. 2002; Rus et al. 2012). Such issues would reasonably lead to an expectation that the use of different methods would reflect prevailing limnological conditions across broach geographic regions. Yet for larger portions of the U.S.A., method choices correspond to state boundaries rather than obvious environmental patterns (Fig. 1).

With increasing access to open data and water quality data sets derived from multiple sources (integrated data sets; e.g., Read et al. 2017; Soranno et al. 2017), results based on these two general approaches are likely to be used interchangeably in secondary data analyses. This led us to ask: are results from these two strategies for determining TN in fact interchangeable? Or, more

---

*Correspondence: ehstanley@wisc.edu

Additional Supporting Information may be found in the online version of this article.

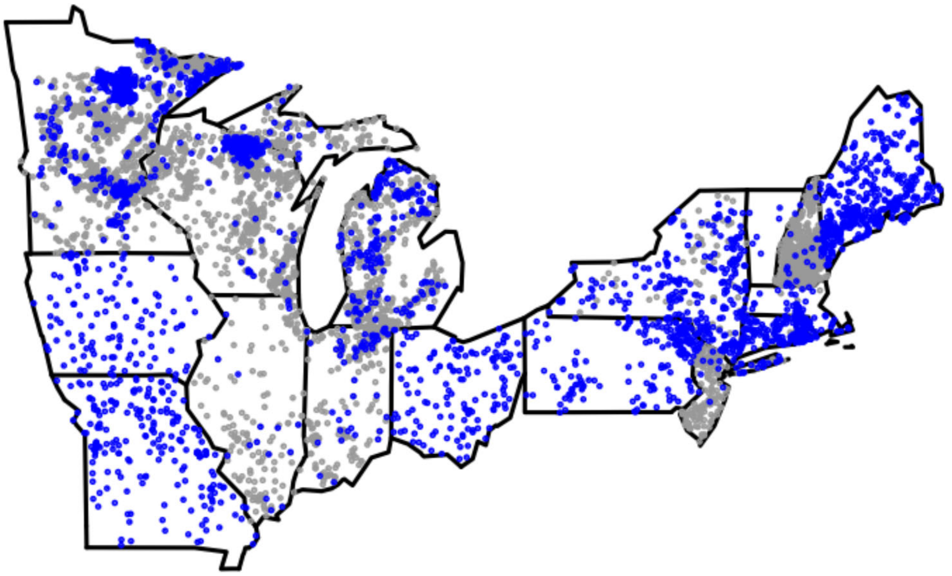

Fig. 1. Distribution of TN data analyzed using oxidation to NO_2+NO_3-N (TN-d; blue) and calculated from Kjeldahl and NO_2+NO_3-N determination (TN-c; gray) from the LAGOS-NE database.

specifically, are there meaningful differences in the accuracy and/or precision between these two approaches present within heterogeneous, integrated data sets? We addressed these questions using data derived from two open-access water quality data sets to evaluate accuracy, precision, and agreement between approaches. Our intent was to compare methodologies not as laboratory analysts, but from the perspective of secondary data users who might ignore or have limited access to information about analytical methods, and thus may unknowingly be introducing error into statistical analyses.

Methods

We used a variety of strategies to identify potential differences in results generated from combined (TKN + NO_2+NO_3-N; hereafter TN-c) vs. direct (NO_x-N after oxidation or TN-d) approaches within data sets derived from multiple sources. Data from the U.S. Geological Survey's Standard Reference Sample Project (USGS-SRS) were used to examine patterns in biases (accuracy) and variability (precision) for each of the two TN methodologies. Agreement between approaches in the analysis of surface water samples was evaluated using data from LAGOS-NE, a lake and reservoir database covering 17 northeastern and north-central U.S. states (Soranno et al. 2015, 2017). Finally, we compiled information on analytical difficulties for the two approaches from both the USGS-SRS and LAGOS-NE data sets, including reported detection limits (DLs) and cases in which the sum of the N fractions that compose TKN or TN exceeded reported concentrations of either TKN or TN. All analyses and graphics were conducted in R (R Core Team 2018); specific R packages are listed with the corresponding statistical analysis. Data and code used to generate results are available at: https://doi.org/10.5281/zenodo.3252078

Bias and variance analysis

The USGS-SRS project provides investigators and monitoring organizations with an opportunity to evaluate their lab's performance for a variety of different analytes and compare this performance to other labs. Each participating lab measures and reports the concentration of a given water chemistry variable in an unknown standard reference sample (SRS), and these reports are used to compute the "most probable value" (MPV): the median concentration across all reported values. SRS "round robins" occur twice a year, with USGS providing high- and low-concentration reference samples to interested parties. Further details about this program, data, and preparation of the SRS are available via the USGS-SRS project web page (https://bqs.usgs.gov/srs/).

We acquired results for TKN (listed as "Ammonia + Organic Nitrogen as N" by USGS), TN, and NO_2+NO_3-N for high and low nutrient reference sample evaluations between 2012 and 2018. Analyses were restricted to entries that reported the use of colorimetric methods for TKN and TN-d, and cases in which NO_2+NO_3-N (not NO_3-N alone) data were available for labs using the TN-c approach. We also calculated our own TN-c values for all laboratories reporting TKN and NO_2+NO_3-N regardless of whether they also reported TN. A series of logical decisions (Fig. S1) were then used to remove data that could not be analyzed then infer the approach used for TN determination by individual labs: measured in a single analysis (TN-d) or calculated from TKN + NO_2+NO_3-N (TN-c). This workflow was applied to each unique lab-round robin combination. We subsequently examined these assignments and removed several entries prior to statistical assessments based on four criteria: (1) uncertainty regarding the identification of the approach for determining TN (e.g., the inferred approach seemed to switch between TN-c and TN-d among dates for a given lab); (2) apparent data entry errors (e.g., submitted results were ~ 10× greater than the MPV and were consistent with a misplaced decimal point); (3) using TKN concentration as the reported TN value when NO_2+NO_3-N was not reported or vice versa; and (4) data entered as less than some value—that is, the reported result was below the lab's DL.

After the analytical approaches were determined, we followed the example of the USGS Inorganic Blind Sample Project (https://qsb.usgs.gov/ibsp/) and treated the MPV as the "true" concentration to quantify biases for the two approaches. Bias values were determined as the average of deviations from the MPV of TN (i.e., lab TN result—MPV_{TN}). We also generated a second trimmed data set using the DescTools R package in which the top and bottom 10% of the deviations were removed for each method in order to reduce (and thus determine) the leveraging effects of a small number of outliers on our statistical assessments. To evaluate possible differences between the two TN approaches, we compared the bias and variances of deviations for TN-d and TN-c using *t*-tests and *F*-tests, respectively. These comparisons were done for all concentrations combined, and then separately for low- and high-concentration SRS results. Significance was determined based on the Bonferroni correction for multiple comparisons. Finally, deviations from MPV_{TN} for both approaches were regressed against the MPV for NO_2+NO_3-N ($MPV_{NO_2+NO_3}$) to determine if there were biases associated with concentrations of this inorganic N fraction for either TN approach, as has previously been reported for both TN-d (Lee and Westerhoff 2005) and TN-c (Schluter 1977; Patton and Kryskalla 2003).

Since each SRS was assigned an MPV for TN based on the results from all labs that analyzed the same unknown reference sample, biases in TN estimates from each round robin were examined further using a non-parametric approach. Specifically, the median of the absolute deviation (MAD) between the reported TN and MPV_{TN} was computed for each unique SRS. We then compared MADs for TN-d vs. TN-c. MADs for TKN and NO_2+NO_3-N for TN-c analyses were also computed, and we graphically compared the sum of MAD_{TKN} + $MAD_{NO_2+NO_3}$ to $MAD_{TN\text{-}c}$ to assess error propagation.

Concurrent measurements of TN-c and TN-d

To expand the scope of the analysis to include a wider range of natural waters and concentrations, we compared results from individual programs that concurrently analyzed TN via both TN-c and TN-d approaches. These data were extracted from LAGOS-NE_{LIMNO} (Soranno et al. 2017) v1.087.2. The data set included 1711 concurrent measurements of TN-c and TN-d from 261 different lakes by 10 different agencies or research groups. Results flagged as below DLs were excluded from the comparison to avoid introducing new error associated with differential handling of data at/below DLs (Stow et al. 2018). This resulted in 838 cases in which TN-c and TN-d were measured concurrently from samples collected on the same date and at the same within-lake location and depth for 191 different lakes. To further limit possible biases due to unequal sampling effort across lakes (there were between 1 and 85 observations/lake), we randomly selected a single observation from each lake that had more than 1 observation. The relationship between TN-d and TN-c was quantified using robust errors-in-variables regression that accounts for measurement errors in both variables (Passing and Bablok 1983; Manuilova et al. 2014) using the mcr package in R. Confidence intervals of the regression coefficients were quantified by bootstrapping. Slopes and intercepts that contain 1 and 0, respectively, indicate no significant differences between the two methods.

Next, we considered the occurrence of "impossible outcomes" as a means of assessing within-method consistency. "Impossible" results are cases in which the reported concentration of a TN constituent (TKN, NH_4-N, or NO_2+NO_3-N) exceeded that of TN itself, or similarly, in which NH_4-N exceeded TKN. All observations of the associated nitrogen constituents in LAGOS-NE_{LIMNO} v1.087.2 were used to check for occurrence of such cases. The nitrogen data were extracted from this data set instead of the published data set because impossible outcomes had been filtered from the final (published) version of LAGOS-NE 1.087.1 during database assembly and QA/QC checks (Soranno et al. 2015). Impossible outcomes were identified as values > 1 for ratios of NO_2+NO_3-N : TN, NH_4-N : TN, NH_4-N : TKN, and TKN : TN.

Detection limits

We determined the frequency of reports of DLs and the concentrations of those DLs for both USGS-SRS and LAGOS-NE_{LIMNO} v1.087.2 data sets to identify possible differences between methods. For the USGS-SRS data set, we recorded the number of occurrences of values reported as 0 or qualified as "<" a value for TKN, NO_2+NO_3-N, and TN to determine if DL issues were more frequent for one of the two TN methods (these data were removed from the bias analysis as described above). We also collated the corresponding values for these putative DLs as a qualitative means of examining differences in reported sensitivity between the two approaches. For LAGOS-NE 1.087.2, we determined the percentage of programs that reported a DL for any observation for each N variable in our analysis and documented the ranges of DLs reported by LAGOS-NE data contributors.

Results

Bias and variance analysis

Across the 28 SRSs used in this analysis, TN_{MPV} concentrations ranged from 0.199-3.5 mgN L^{-1} derived from 22 to 36 independent analyses per reference standard (Table S1). The magnitude of the deviations was generally small relative to the MPV for both analytical approaches (median = 3.5% and 4.5% for TN-d and TN-c, respectively for the full data set) and on average, reported concentrations were higher than their respective MPV (positive bias) for both methods (Table 1). The overall bias was 0.017 mgN L^{-1} (SD = 0.222) for TN-d and 0.024 mgN L^{-1} (SD = 0.098) for TN-c. Biases were not significantly different between the two methods (p = 0.56). However, for the trimmed (top and bottom 10% of

Table 1. Summary statistics for deviations in TN (value reported by participating labs—MPV in mgN L^{-1}) using calculated (TN-c) and direct (TN-d) approaches, and reference sample concentration categories (low or high) for the full and trimmed data sets.

	Full data set			Trimmed data set		
Method and concentration category	**Sample size**	**Mean**	**SD**	**Sample size**	**Mean**	**SD**
Calculated	514	0.024	0.098	412	0.016	0.041
Direct	423	0.017	0.222	339	0.008	0.031
Calculated - low	239	0.026	0.068	213	0.013	0.037
Direct - low	237	0.019	0.127	211	0.003	0.026
Calculated - high	275	0.023	0.119	199	0.019	0.045
Direct - high	186	0.015	0.303	128	0.016	0.036

Table 2. Variance comparison for deviations in TN among methods and reference sample concentration categories for the full and trimmed data sets. Significance was assessed using α = 0.0031, which allowed for multiple testing and resulted in a family error rate of 0.05.

	F-test *p*-value	
Variance comparison	**Full data**	**Trimmed data**
Calculated - direct	< 0.001	< 0.001
Calc. low - direct low	< 0.001	< 0.001
Calc. high - direct high	< 0.001	0.007
Calc. low - calc. High	< 0.001	0.004
Direct. low - direct high	< 0.001	< 0.001

values removed) data set, the bias for TN-d (0.008) was significantly lower ($p = 0.0013$, $\alpha = 0.0031$ following adjustment for multiple testing) than for TN-c (0.016). This significant difference in bias was also present for the subset of low TN reference sample results in the trimmed data set ($p = 0.0005$), but not for the high concentration group ($p = 0.53$). Plots of deviations vs. $MPV_{NO_2+NO_3}$ indicated no relationship for TN-c, but a significant, although weak relationship for TN-d (Fig. S2). A negative bias at high concentrations of NO_2+NO_3-N translated to a predicted underestimation of TN-d of 0.096 mgN L^{-1} at 3 mg NO_2+NO_3-N L^{-1}. However, this bias was notably diminished (but not completely eliminated) in similar analyses using the trimmed data set (Fig. S2).

Variance in the deviations of the reported concentrations from the MPVs for the full data set was significantly larger for TN-d (collectively, and when divided into the low and high SRS categories), and for high-concentration SRS results relative to low-concentration results within each method (Table 2). However, trimming the full data set reversed these patterns: variance in deviations for trimmed TN-d was significantly lower than for trimmed TN-c collectively and for the low concentration subset, but not for the high concentration results. These results reflect the presence of a small number of samples with large deviations for TN-d that led to high variance in the full data set (and subsequently, low variance following their removal via trimming). Finally, significant differences in the variance of deviations for low vs. high comparisons persisted for the direct approach but not for the combined approach.

Consistent with the variance analysis of the trimmed data set, MADs from the 28 unique SRS trials were lower for TN-d than for

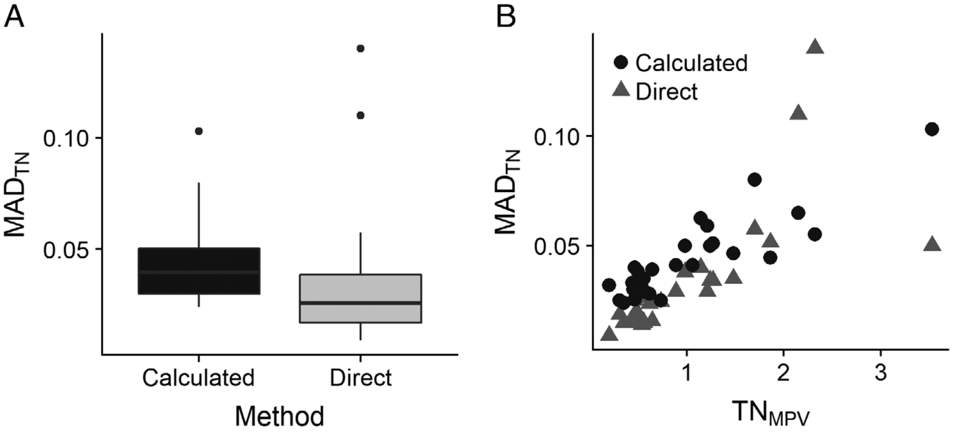

Fig. 2. (**A**) Boxplots of median absolute deviation of TN across all values of MPV_{TN} for each inferred method. (**B**) Median absolute deviation of TN for both methods (direct and calculated) for each MPV_{TN}.

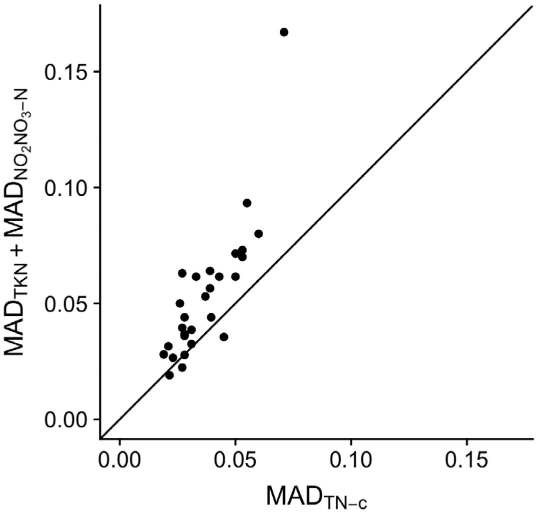

Fig. 3. Comparison of MADs for TN-c vs. the summed MADs for the components TKN and NO_2+NO_3-N for each of the 28 USGS-SRS round-robins, with the 1 : 1 line for reference.

TN-c for the full data set (Fig. 2A). A plot of MAD vs. TN_{MPV} indicated that lower deviations for TN-d compared to TN-c held for low to moderate TN_{MPV} concentrations (~ 0.20–2.0 mgN L^{-1}), but became much noisier for high TN_{MPV} concentrations where data availability was limited (Fig. 2B). The comparison of MADs for TN-c alone vs. summed MADs for the two component N fractions (MAD_{TKN} + $MAD_{NO_2+NO_3}$) from the USGS-SRS round robins revealed consistently higher values for the summed MADs (Fig. 3).

Detection limits

Among the 606 TKN and 447 TN-d measurements available from the USGS-SRS project, 11.9% of the TKN entries and 1.1% of the TN-d entries were reported as 0 or "<" the lab's reporting limit (Table 3). Four of the 47 labs reporting TN-d had such qualified results, compared to 11 of the 47 labs reporting TKN. Less than 0.5% of all NO_2+NO_3-N data were qualified as "<" the lab's reporting limit. These apparent DLs varied from 0.1 to 0.2 mgN L^{-1} for NO_2+NO_3-N, 0.078 to 1.0 mgN L^{-1} for TKN, and 0.036 to 1.0 mgN L^{-1} for TN-d.

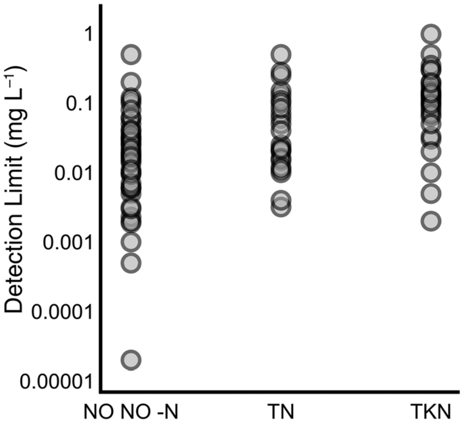

Fig. 4. Detection limits for NO_2+NO_3-N, TN, and TKN data reported in the LAGOS-NE database. Each unique detection limit value in the database is represented by an individual bubble (n = 44 different DLs reported for NO_2+NO_3-N, 23 for TN, and 28 for TKN), and detection limits span several orders of magnitude.

DL information in the LAGOS-NE database is incomplete, both in terms of reporting DLs and consistent flagging of data points at or below these limits. For example, there were no reported DLs for TN for 35% of programs, no TKN DLs for 44% of programs, and a maximum of 62% of the programs did not provide DLs for NO_2+NO_3-N. For programs that did report DLs, TN, and TKN limits were comparable, though slightly narrower for the former: TN DLs ranged from 0.002 to 0.50 mgN L^{-1} with a median of 0.05 mgN L^{-1} while TKN DLs varied from 0.002 to 0.98 mgN L^{-1} with a median of 0.10 mgN L^{-1} (Fig. 4). NO_2+NO_3-N DLs were lower, ranging from 0.00002 to 0.50 mg NO_2+NO_3-N L^{-1} with a median of 0.015 mg NO_2+NO_3-N L^{-1}.

Comparison of simultaneous TN-c and TN-d results among lakes

Concurrent measurements of TN-c and TN-d were highly correlated across a range of concentrations that varied by almost 3 orders of magnitude (r = 0.877; Fig. 5A). The estimates of the intercept and slope were –0.030 and 1.025, respectively. The 95% confidence intervals of the estimated slope (0.994, 1.108) included 1 but the interval for the

Table 3. Number of TKN, TN-calculated (TN-c), and TN-direct (TN-d) results included in the USGS-SRS program between 2012 and 2018, and number of entries below apparent DLs for different N analyses.

	TKN	NO_2+NO_3-N*	TN-c†	TN-d
Number of laboratories	47	77	18	47
Total number of analyses	606	938	177	447
Entries of 0 or with "<"	74	4	1	6
Range of reported DLs ("<" values; mgN L^{-1})	0.078–1.0	0.1–0.2	2.0	0.036–1.0

*Results are for all reported NO_2+NO_3-N results, regardless of approach used to determine TN.

†Results are for labs that reported a TN value (does not include TN-c results calculated for this study).

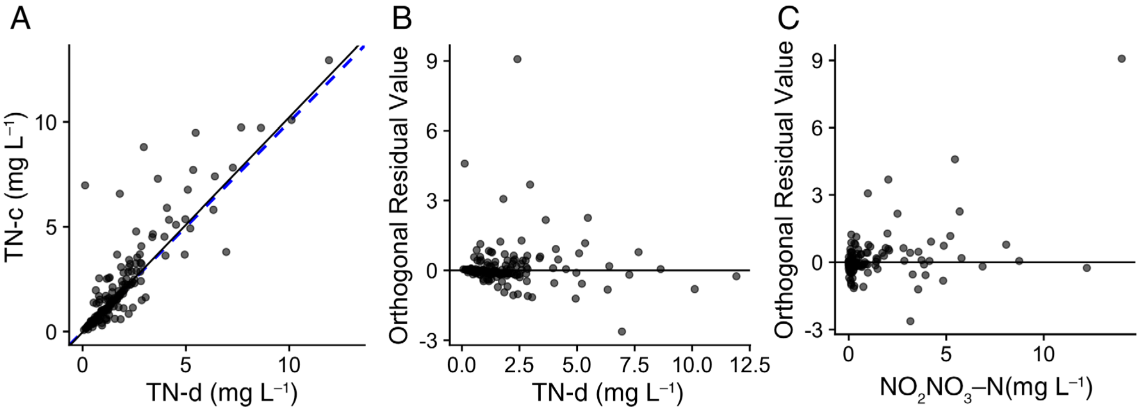

Fig. 5. (**A**) Comparison 191 concurrent measures of TN-d and TN-c taken for discrete lakes located in Connecticut, Iowa, Michigan, Minnesota, New York, and Wisconsin. Dashed line is 1 : 1 and solid line is the best orthogonal fit. Bootstrapped confidence intervals of the intercept excluded 0, but the slope included 1. (**B**) Orthogonal residuals for the regression (represented by the solid reference line at 0) as a function of the measured TN concentration. (**C**) Orthogonal residuals for the regression (represented by the solid reference line at 0) as a function of the measured NO_2+NO_3-N concentration.

intercept (−0.085, −0.005) did not include 0. Examination of the residuals across the TN concentration gradient did not show a consistent pattern, although the largest residuals were associated with very high TN-c values at low TN-d concentrations (Fig. 5B). Similarly, no pattern was apparent in the residuals across the gradient of sample NO_2+NO_3-N concentration, although the highest residuals were associated with TN-c (Fig. 5C).

Calculating TN from TKN and NO_2+NO_3-N increased the opportunity for errors in component N species to be incorporated into the results. The percentage of data with impossible outcomes (ratios of component : total > 1) were low for inorganic fractions compared to TN (0.7% for NH_4-N : TN, 1.1% for NO_2+NO_3-N : TN) and slightly higher but still low (2.9%) for NH_4-N : TKN. However, cases in which TKN exceeded TN reached 30%, although there were fewer concurrent measurements of these variables than for the other variable pairs (Fig. 6).

Discussion

Our analyses of USGS-SRS and LAGOS-NE data sets were intended to provide a variety of perspectives on biases and potential interchangeability of TN results derived from direct and combined approaches. Several prior studies have compared TN

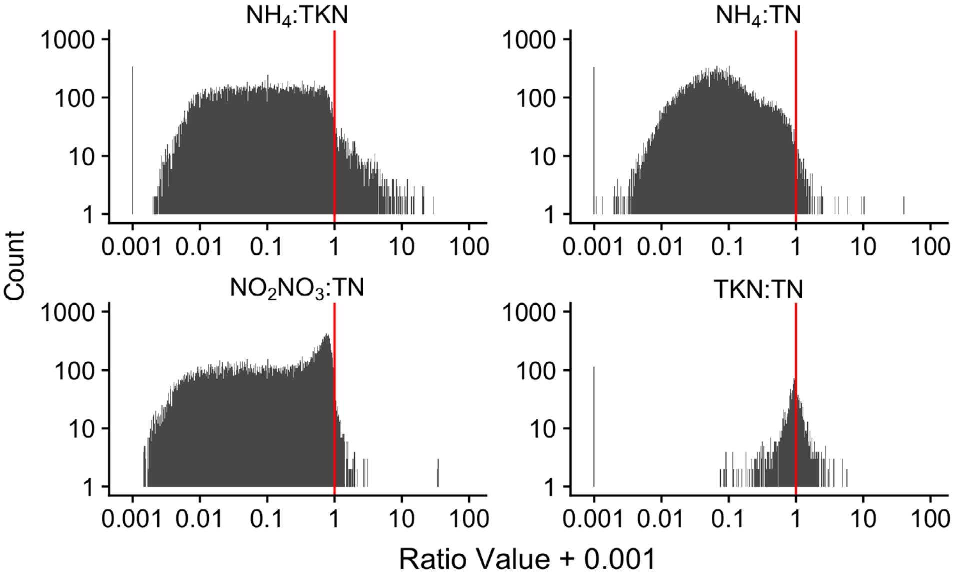

Fig. 6. Impossible outcomes: Ratios histograms of N variable pairs. Histograms exclude 25 observations that have an infinite value (i.e., denominator was zero). Red line is a ratio of 1. Values beyond 1 made up 2.94% for NH_4 : TKN, 0.69% for NH_4 : TN, 1.12% for NO_2NO_3 : TN, and 30.39% for TKN : TN. Note both axes are log10 transformed to visualize the data.

methodologies both within direct (e.g., Bronk et al. 2000; Vandenbruwane et al. 2007) and combined (reviewed by Saéz-Plaza et al. 2013) approaches, as well as between TN-d and TN-c methods (e.g., Smart et al. 1981; Maher et al. 2002; Patton and Kryskalla 2003; Rus et al. 2012). However, these comparisons have all been laboratory investigations that considered a small number of protocols (usually 2 or 3) that were well known and well controlled by the authors. Our intent differed in that we were interested in possible biases or uncertainties that persist in integrated data sets in which protocols, sample sources, and conditions may be neither well known nor well controlled. As integrated data resources become increasingly common in limnological research, checking for these sorts of biases should be a regular part of their use; otherwise, unrecognized sources of error can lead to reduced analytical power or inaccurate conclusions (Stanley et al. 2019). Below, we consider our results in the context of accuracy, precision, and agreement between these two approaches and their implications for data interchangeability.

Accuracy

The USGS SRS inter-laboratory comparison provided a realistic, albeit indirect means of assessing accuracy of the two TN approaches that was well-suited for our consideration of interchangeability of TN-c and TN-d results. This data set is composed of several years of results from several laboratories that inevitably use a variety of protocols for measuring TN-c or TN-d, consistent with the realities of integrated data sets. However, the analysis is indirect in that the true concentration of TN in a reference sample is not known with 100% certainty; instead, the MPV of the true concentration is estimated from the collection of reported results.

Comparison of TN reported by individual labs to MPV_{TN} indicated relatively small overall biases (high accuracy) that were similar in magnitude between methods. However, reanalysis of results after removing observations that deviated the most from the MPV (trimmed data set) indicated a larger bias for TN-c than TN-d, particularly for low-concentration reference samples. Trimming the data set also reduced a significant but small influence of NO_2+NO_3-N on TN-d accuracy in the full data set, in which TN had been overestimated at low NO_2+NO_3-N concentrations and underestimated at higher concentrations. Taken together, these results suggest that the direct approach can generate a small percent of notably inaccurate outcomes, but on average, it is more accurate than the combined approach (Table 4).

Accuracy at low concentrations of an analyte (i.e., sensitivity) has been a persistent motivation for development and modification of TN methods (e.g., Walsh 1989; Frankovich and Jones 1998; Vandenbruwane et al. 2007). In addition to examining accuracy as a function of TN concentration, we used reported DLs as an indicator of potential differences in sensitivity between approaches. For the USGS-SRS data, the greater frequency of concentrations listed as less than some value for TKN highlight a pattern of reduced sensitivity for the TN-c approach. Although Kjeldahl methods are often cited as having higher DLs than direct approaches (e.g., *see* Patton and Kryskalla 2003), this result does

Table 4. Summary of conclusions drawn from comparisons between TN-d and TN-c data sets, and the strength of support for the conclusion.

Comparison			Conclusion	Strength
Accuracy				
	Bias analysis			
		General	Both have low bias (high accuracy), but accuracy is more consistent for TN-d	Moderate
		Effect of TN concentration	TN-d is more accurate at low TN concentrations	Moderate
		Effect of NO_2+NO_3-N concentration	TN-d accuracy is affected by NO_2+NO_3-N concentrations	Weak
	Detection limits		DLs indicate lower sensitivity for TN-c in USGS-SRS data	Strong
			Higher average and maximum DLs for TN-c in LAGOS-NE	Weak
			DLs often unreported for NO_2+NO_3-N, leading to more unknown DL relationships for TN-c	Strong
Precision				
	Variance		TN-d is more precise on average but has a small fraction of extreme outcomes	Moderate
	Median absolute deviation		TN-d is more precise than TN-c. Error propagation in TN-c amplifies this difference	Strong
	Impossible outcomes		TN-c is susceptible to TKN>TN	Strong
Agreement			Good agreement between concurrent TN-d and TN-c analyses of surface water samples	Moderate

not necessarily mean that TN-c is inherently less sensitive than TN-d, but that use of TN-c protocols with low sensitivities is more common. Examination of LAGOS-NE provided weak support for lower sensitivity of TN-c: reported DLs for TN and TKN had substantial overlap, but median and maximum DLs were higher for TKN than for TN. Perhaps the most notable outcomes from this exercise were the often remarkably high DLs and the widespread absence of DL information. There were several cases of DLs in excess of 0.5 mgN L^{-1}, and such a coarse level of sensitivity strongly limits the value of monitoring or research efforts. For example, this threshold is greater than mean TN concentrations typical of oligotrophic and mesotrophic lake states (Wetzel 2001). With respect to missing DLs, the progressive loss of DL information from TN to TKN to NO_2+NO_3-N highlights a larger problem for TN-c than for TN-d with respect to trying to assess method sensitivity or recognizing biases that can occur when DL information is unavailable (Stow et al. 2018).

Precision

The variance in deviations from MPV_{TN} was significantly greater for TN-d than for TN-c. However, trimming the data by removing a subset of observations with the most extreme deviations (both positive and negative) resulted in greater accuracy in estimating TN. The improvement in precision as a result of trimming was larger for TN-d than TN-c, leading to a conclusion of better precision for the former approach. Thus, as with accuracy, greater precision for TN-d was counterbalanced by the occurrence of a fraction of values that were particularly extreme. Although we cannot determine the cause of these outlier results, one possible source of this variability could be the larger diversity of methods used to oxidize TN (e.g., multiple permutation on persulfate, UV, and high temperature methods) compared to Kjeldahl digestions (Saéz-Plaza et al. 2013). The greater precision of TN-d suggested by the trimmed data set agreed with the nonparametric comparison of variability in deviations for TN up to 2 mgN L^{-1} for the entire data set (Table 4). Beyond this concentration, it was difficult to draw any conclusion given the small number of data points available.

Considering the precision of TN-c in terms of its variance and MAD is a conservative strategy, as some of the variability inherent to TN-c may be masked. If errors for TKN and NO_2+NO_3-N are in opposite (offsetting) directions, little to no error would be apparent for a reported TN-c value. Assuming independence, the true variance for TN-c is the sum of the absolute values of variance for each component fraction. Availability of concentration data for both the sum (i.e., TN-c) and the parts (TKN and NO_2+NO_3-N) allowed us to compare the variance of TN-c by itself to the combined variance of its constituents. We elected to consider this issue of error propagation using a robust nonparametric approach because the true concentration of the SRS was not known with 100% certainty and instead was estimated as the MPV. But even with this cautious approach, biases for TN-c alone were notably lower than for TKN and NO_2+NO_3-N in combination (Table 4). Thus, when TN-c alone is reported, its apparent precision is falsely inflated. Diminished precision due to summing components of TN is well known, and error propagation has been identified as a key limitation of the combined approach (e.g., D'Elia et al. 1977; Patton and Kryskalla 2003). Given that the MADs for TN-c alone were greater than for TN-d, the larger actual propagated error is particularly concerning. It is easy to overlook this error when using publicly available data, or it may not even be possible to see the error if TN methods are not known or if the components of TN-c are not included in an integrated data set.

Differences in precision were also considered via impossible outcomes; that is, sample analyses in which component N fractions exceed TN or TKN. The frequency of cases in which TKN concentration exceeded TN was surprisingly high both in general and relative to the other ratios considered, as almost one third of all records had TKN concentrations that exceeded TN. The distinctly high frequency of impossible outcomes for TN-c is similar to problems reported for dissolved organic N (DON). Imprecision in the determination of TN and inorganic N (needed to calculate DON) have led to frequent reports of negative DON concentrations (Hansell 1993, Chen et al. 2015) and again highlight the disadvantage of approaches that require multiple analyses to arrive at the result. Multiple analyses are unavoidable for determining DON; the same is not the case for TN however.

Agreement

If results from the USGS-SRS exercise are used to predict the relationship between concurrent measurements of TN-c and TN-d, the lack of a significant difference in bias between the two analyses for the full data set suggests close adherence to a line with a slope of 1 and an intercept of 0. However, the trimmed data set would shift this prediction slightly to expecting the y-axis (representing TN-c) to have an intercept greater than 0 due to the significantly higher bias for the combined approach. Because this bias was present for low-concentration SRS results only, we would expect the line to approach 1 : 1 and thus have a slope slightly less than 1. In contrast, the actual regression for TN-d vs. TN-c using LAGOS-NE data had a slope that did not differ from 1 and a negative y-intercept, indicating lower TN-c results relative to its paired TN-d value—at least at low concentrations. However, the estimated average y-intercept departure from 0 was relatively small compared to the range of concentrations included in the analysis.

Examples of similar comparisons of simultaneous determination of TN across a range of limnological conditions are limited. However, a rigorous study by Patton and Kryskalla (2003) provides some context for our regression results. These investigators compared Kjeldahl digestion (KN) to alkaline persulfate oxidation minus NO_3-N (N_{AlkP}) for over 2000 surface and groundwater samples across the United States. Their regression of KN vs. N_{AlkP} produced a nearly identical slope (1.023 for their regular linear regression vs. 1.025 for our Passing-Bablok regression), and a y-intercept that was similar in magnitude, though opposite in direction (0.038 for KN vs. −0.030 for TN-c). This difference, along with a larger sample size translated to a statistically significant difference between the two methods in their study. The difference between methods also increased in magnitude as sample NO_3-N concentrations increased, which was attributed to overestimation by the KN method—a tendency we did not observe in the LAGOS-NE data. Comparing between these two studies should be done cautiously given their obvious differences; nonetheless, the similarity in the magnitude of the y-intercepts and slopes is striking, and in the current comparison, the slope did not differ from 1. Thus, we suggest that the comparison of paired results indicates reasonable, though not definitive support for interchangeability of results within integrated data sets (Table 4).

Context and conclusion

Monitoring analytical accuracy and precision is a regular activity for most water quality laboratories and is typically done by comparing measurements of one or more standards and/or spiked samples to their known concentrations. We did not use this approach here, but instead took a step back and considered tendencies in composite data sets derived from multiple labs, multiple sites, and multiple dates, as well as using protocols that inevitably differed within each of the two analytical approaches. As such, it is essential to recognize some of the advantages and disadvantages of our study.

Labs participating in the USGS-SRS comparison study include a variety of metropolitan, regional, state, and federal agencies, private analytical labs, and university research groups. Many of these organizations are required to engage in regular proficiency testing such as involvement in the USGS-SRS (e.g., municipal and state facilities) and labs that voluntarily engage in these exercises may be particularly concerned with their performance. Thus, it is reasonable to assume that participants are experienced and reliable analysts. If so, then USGS-SRS results provide a conservative view on biases and other pitfalls associated with TN analysis regardless of the method.

We assume that the two approaches to TN determination considered here contain a variety of specific methods within each category. For example, conversion of TN to NO_x in the direct approach can be done using various different persulfate digestion protocols (e.g., D'Elia et al. 1977; Valderrama 1981; Crumpton et al. 1992), photo-oxidation (e.g., Armstrong et al. 1966), or high-temperature combustion/oxidation (Suzuki and Sugimura 1985; Walsh 1989) to convert organic N and NH_4 to NO_x. Similarly, this latter N fraction can also be quantified using several different instruments and protocols (APHA 2017). Kjeldahl methods include macro- and micro-digestions, and again with multiple protocols for analyzing the NH_3 produced by these digestions (Saéz-Plaza et al. 2013). Thus, biases and variability described here inevitably include uncertainty attributable to this finer-scale methodological diversity. Unfortunately, such analytical information is often not available in integrated data sets. For example, methods were reported in terms of general categories (e.g., colorimetric, elecrometric, AA, or other for TN) in the USGS-SRS data, eliminating the opportunity to evaluate this potential source of error. This problem is by no means unique to TN; most water quality parameters have more than one standard method. Using data collated from multiple sources thus comes with the assumption of data equality—that is, that all methods and all data sources are equally reliable and thus are interchangeable. Or perhaps more accurately that meaningful environmental signals are larger than any noise that results from different analytical practices.

To answer our original question: are TN-d and TN-c data interchangeable? Our answer based on three lines of evidence is: mostly. First, the accuracy of both methods was high and similar to prior laboratory studies using a limited number of well-controlled procedures (Rus et al. 2012). Second, evidence for one method or another consistently over- or underestimating TN was inconsistent; and finally, regressions of samples simultaneously analyzed by both approaches were difficult to distinguish from a 1 : 1 relationship. On this basis, it is hard to argue against using combined and direct data interchangeably. However, we recognize that on balance, most of our comparisons suggested better or more consistent accuracy and precision for the direct approach (Table 4), but these results were not clear cut and we would argue, not sufficiently distinct to warrant the separation of data or dismissal of existing TN-c results. As with our analyses, laboratory studies comparing various Kjeldahl and direct methods provide a similarly ambiguous perspective on this issue, with reports of similar (Smart et al. 1981; Bachmann and Canfield 1996; Doval et al. 1997) as well as greater (Patton and Kryskalla 2003) accuracy of TN-d relative to TN-c. Hence, this ambiguity coupled with reasonable evidence of similar accuracy and good agreement between approaches has led to our endorsement of interchangeability.

Despite this recommendation, we offer two caveats when using results from these two approaches interchangeably in data analyses. First, interchangeability may be precarious at low TN concentrations, as indicated by greater biases and higher DLs for TN-c. Indeed, limited accuracy and sensitivity of Kjeldahl digestions has led the marine community to effectively abandon this approach for DON and TN determinations (*see* Sharp et al. 2002) due to often low concentrations in ocean waters. Second, TN-c data were less precise by many (though not all) measures, a result consistent with most laboratory comparisons (e.g., D'Elia et al. 1977, Smart et al. 1981, Bachmann and Canfield 1996). The combination of lower precision, much of which may not be apparent when viewing TN data by itself, and pervasive absence of information about DLs for the combined approach may be problematic and could lead to reduced statistical power or susceptibility to type I errors. Under most circumstances, TN-d approaches are vastly preferable because they avoid this error propagation issue as well as being less hazardous and less labor-intensive, among other criticisms (Frankovich and Jones 1998; Patton and Kryskalla 2003). Indeed, if a laboratory has the opportunity to select an analytical approach, these caveats lead to a clear recommendation for the use of direct approaches to determine

TN. However, given the reality that data generated from TN-c approaches are widespread among studies of inland waters (Fig. 1), we urge careful examination of the data, recognizing the reduced sensitivity and higher variance of TN-c results. But these issues should not preclude the use of data from both approaches for data synthesis studies.

References

APHA. 2017. Standard Methods for the Examination of Water and Wastewater, 23rd Edition. Washington, D.C: American Public Health Association, American Water Works Association, American Public Health Association.

Armstrong, F. A. J., P. M. Williams, and J. D. H. Strickland. 1966. Photo-oxidation of organic matter in sea water by ultra-violet radiation, analytical and other applications. Nature **211**: 481–483. doi:10.1038/211481a0

Bachmann, R. W., and D. E. Canfield Jr. 1996. Use of alternative methods for monitoring total nitrogen concentrations in Florida lakes. Hydrobiologia **323**: 1–8. doi:10.1007/BF00020542

Bronk, D. A., M. W. Lomas, P. M. Glibert, K. J. Schukert, and M. P. Sanderson. 2000. Total dissolved nitrogen analysis: Comparisons between the persulfate, UV and high temperature oxidation methods. Mar. Chem. **69**: 163–178. doi:10.1016/S0304-4203(99)00103-6

Chen, B., P. Westerhoff, L. Zhang, A. Zhu, X. Yang, and C. Wang. 2015. Application of pretreatment methods for reliable dissolved organic nitrogen analysis in water—A review. Crit. Rev. Env. Sci. Tec. **45**: 249–276. doi:10.1080/10643389.2013.852400

Crumpton, W. G., T. M. Isenhart, and P. D. Mitchell. 1992. Nitrate and organic N analysis with second-derivative spectroscopy. Limnol. Oceanogr. **37**: 907–913. doi:10.4319/lo.1992.37.4.0907

D'Elia, C. F., P. A. Steudler, and N. Corwin. 1977. Determination of total nitrogen in aqueous samples using persulfate digestion. Limnol. Oceanogr. **22**: 760–764. doi:10.4319/lo.1977.22.4.0760

Doval, M. D., F. Fraga, and F. F. Perez. 1997. Determination of dissolved organic nitrogen in seawater using Kjeldahl digestion after inorganic nitrogen removal. Oceanol. Acta **20**: 71–720.

Frankovich, T. A., and R. D. Jones. 1998. A rapid, precise and sensitive method for the determination of total nitrogen in natural waters. Mar. Chem. **60**: 227–224. doi:10.1016/S0304-4203(97)00100-X

Hansell, D. A. 1993. Results and observations from the measurement of DOC and DON in seawater using a high-temperature catalytic oxidation technique. Mar. Chem. **41**: 195–202. doi:10.1016/0304-4203(93)90119-9

Lee, W., and P. Westerhoff. 2005. Dissolved organic nitrogen measurement using dialysis pretreatment. Environ. Sci. Technol. **39**: 879–884. doi:10.1021/es048818y

Maher, W., F. Krikowa, D. Wruck, H. Louie, T. Nguyen, and W. Y. Huang. 2002. Determination of total phosphorus and nitrogen in turbid waters by oxidation with alkaline potassium peroxodisulfate and low pressure microwave digestion, autoclave heating or the use of closed vessels in a hot water bath: Comparison with Kjeldahl digestion. Anal. Chim. Acta **463**: 283–229. doi:10.1016/S0003-2670(02)00346-X

Manuilova, E., A. Schuetzenmeister, and F. Model. 2014. mcr: Method comparison regression. R package version 1.2.1. https://CRAN.R-project.org/package=mcr

Passing, H., and W. Bablok. 1983. A new biometrical procedure for testing the equality of measurements from two different analytical methods. Application of linear regression procedures for method comparison studies in clinical chemistry, Part 1. J. Clin. Chem. Clin. Biochem. **21**: 709–720. doi:10.1515/cclm.1983.21.11.709

Patton, C.J. and J. R. Kryskalla. 2003. Methods of analysis by the U.S. Geological Survey National Water Quality Laboratory—Evaluation of alkaline persulfate digestion as an alternative to Kjeldahl digestion for determination of total and dissolved nitrogen and phosphorus in water. U.S. Geological Survey Water Resources Investigations Report 03-4174, Denver, CO.

Read, E. K., and others. 2017. Water quality data for national-scale aquatic research: The Water Quality Portal. Water Resour. Res. **53**: 1735–1745. doi:10.1002/2016WR019993

Rus, D. L., C. J. Patton, D. K. Mueller, and C. G. Crawford. 2012. Assessing total nitrogen in surface-water samples—Precision and bias of analytical and computational methods. U.S. Geological Survey Scientific Investigations Report 2012–5281, Reston, VA.

Saéz-Plaza, P., M. J. Navas, S. Wybraniec, T. Michalowski, and A. G. Asuero. 2013. An overview of the Kjeldahl method of nitrogen determination. Part II. Sample preparation, working scale, instrument finish, and quality control. Crit. Rev. Anal. Chem. **43**: 224–272. doi:10.1080/10408347.2012.751787

Schluter, A. 1977. Nitrate interference in total Kjeldahl nitrogen determinations and its removal by anion exchange resins. U.S. Environmental Protection Agency, EPA-600/7-77-017, Cincinnati, OH, USA.

Sharp, J. H., and others. 2002. A preliminary methods comparison for measurement of dissolved organic nitrogen in seawater. Mar. Chem. **78**: 171–184. doi:10.1016/S0304-4203(02)00020-8

Smart, M. M., F. A. Reid, and J. R. Jones. 1981. A comparison of a persulfate digestion and the Kjeldahl procedure for determination of total nitrogen in freshwater samples. Water Res. **15**: 919–921. doi:10.1016/0043-1354(81)90148-2

Soranno, P. A., and others. 2017. LAGOS-NE: A multi-scaled geospatial and temporal database of lake ecological context and water quality for thousands of U.S. lakes. GigaScience **6**: 1–22. doi:10.1093/gigascience/gix101

Soranno, P. A., and others. 2015. Building a multi-scaled geospatial temporal ecology database from disparate data sources: Fostering open science through data reuse. GigaScience **4**: 28. doi:10.1186/s13742-015-0067-4

Stanley, E. H., S. M. Collins, N. R. Lottig, S. K. Oliver, K. E. Webster, K. S. Cheruvelil, and P. A. Soranno. 2019. Biases in lake water quality sampling and implications for macroscale research. Limnol. Oceanogr. **64**: 1572–1585. doi:10.1002/lno.11136

Stanley, E. H., and J. T. Maxted. 2008. Changes in the dissolved nitrogen pool across land cover gradients in Wisconsin streams. Ecol. Appl. **18**: 1579–1590. doi:10.1890/07-1379.1

Stow, C. A., K. E. Webster, T. Wagner, N. Lottig, and P. A. Soranno. 2018. Small vales in big data: The continuing need for appropriate metadata. Ecol. Inform. **45**: 26–30. doi:10.1016/j.ecoinf.2018.03.002

Suzuki, Y., and I. T. Sugimura. 1985. A catalytic-oxidation method for the determination of total nitrogen dissolved in seawater. Mar. Chem. **16**: 83–97. doi:10.1016/0304-4203(85)90029-5

Valderrama, J. C. 1981. The simultaneous analysis of total nitrogen and total phosphorus in natural waters. Mar. Chem. **10**: 109–122. doi:10.1016/0304-4203(81)90027-X

Vandenbruwane, J., S. De Neve, R. G. Qualls, J. Salomez, and G. Hofman. 2007. Optimization of dissolved organic nitrogen (DON) measurements in aqueous samples with high inorganic nitrogen concentrations. Sci. Tot. Environ. **386**: 103–113. doi:10.1016/j.scitotenv.2007.06.025

Walsh, T. W. 1989. Total dissolved nitrogen in seawater: a new high-temperature combustion method and a comparison with photo-oxidation. Mar. Chem. **26**: 295–311. doi:10.1016/0304-4203(89)90036-4

Wetzel, R. G. 2001. Limnology: Lake and river ecosystems. San Diego, CA: Academic Press.

Acknowledgments

Funding was provided by the National Science Foundation Macrosystems Biology Program through collaborative grants EF-1685534 and EF-1638550 and DBI-1401954 to S.M.C. DEB-1440297, NTL LTER, provided further support to E.H.S. and N.R.L.

Conflict of Interest

None declared.

Submitted 21 June 2019
Revised 21 September 2019
Accepted 07 October 2019

Associate editor: Michael Beman