Inexpensive spot sampling provides unexpectedly effective indicators of watershed nitrogen status

Donald E. Weller † and Thomas E. Jordan

Smithsonian Environmental Research Center, 647 Contees Wharf Rd., Edgewater, MD 21037-3702, USA

† Email: wellerd@si.edu

Abstract. Stream water-quality data are essential for understanding watershed processes and managing water pollution; but the effort and expense of stream monitoring limit how many watersheds can be studied. For 59 small watersheds in the Chesapeake Bay drainage, we compared water quality measurements from inexpensive spot sampling to data from costly automated monitoring that used 1-3 years of continuous flow measurement and weekly, temporally composited water sampling. Mean nitrogen (N) levels ranged from 0.01 to 16 mg N/L among streams. There were important temporal variations in N concentrations at each site, but the differences among sites were much greater. Spot samples were very effective at accurately and precisely placing average stream N levels within the N gradient among streams draining N-enriched watersheds. Among watersheds, nitrate (NO3) and total N concentrations from spot samples were very strongly correlated with means from weekly composite sampling ($R^2 > 97\%$). We confirmed this result for independent data for 85 larger watersheds in the Chesapeake Bay Non-tidal Network. NO3 concentration from a single March spot sample was highly correlated ($R^2 > 92\%$) with flow-weighted average total N concentrations synthesized from five years of monitoring. Spot sampling effectively quantifies average N status across N-enriched watersheds because most N moves as NO3 in subsurface flow, and that flux is much less variable than the episodic surface transport of particulate materials. For questions answered by quantifying average N levels, spot sampling can assess more watersheds at much lower cost than automated sampling, so it should be more widely used to support cost-effective N research and management. For materials that are mainly bound to particulates, like phosphorus, spot sampling is much less effective.

Key words: stream sampling; water quality; synoptic sample; composite sampling; watershed assessment; bootstrapped regression; nitrogen pollution; nitrogen management

INTRODUCTION

Measurements of nitrogen levels in streams and rivers provide critical information for advancing basic ecosystem science as well as quantifying and managing anthropogenic pollution of aquatic systems. Nitrogen is often the nutrient that limits plant production in natural ecosystems (Schlesinger 2009), so information on nitrogen loss in stream water is essential for quantifying watershed nitrogen balances (Jordan and Weller 1996, Boyer et al. 2002) and for better understanding terrestrial plant production and nitrogen cycling (Brookshire et al. 2011). Low nitrogen levels also limit productivity in managed systems, motivating the application of nitrogen fertilizer, especially to croplands (Jordan et al. 1997b, a, Harmel et al. 2006a, Stewart and Lal 2017). The resulting release of nitrogen in land runoff can pollute aquatic systems, causing eutrophication and associated ecological and economic disruption (Nixon 1995, Doney 2010, Sobota et al. 2015, Boesch 2019). Global fertilizer applications have increased roughly fivefold over the past 50 years (Foley et al. 2011), and will likely continue to increase due to population growth and increasing meat consumption (Galloway and Cowling 2002, Abbott et al. 2018). Managing the impacts on aquatic systems demands data on nitrogen levels in streams to identify nitrogen source areas, quantify aquatic nitrogen loading, and assess the value of management efforts to reduce it.

The most accurate methods for measuring stream nitrogen transport employ automated monitoring stations that combine continuous streamflow measurement with frequent samples of nitrogen concentration (Swistock et al. 1997). However, one must balance the high cost of temporally intensive sampling against the acceptable level of uncertainty in water quality characterization. Scientists and engineers have examined the effect of sampling strategy on uncertainty in concentrations or load measurements, and many have concluded that composite sampling is an effective way to balance sampling effort against uncertainty (Harmel and King 2005, Moatar and Meybeck 2005, Harmel et al. 2006b, Harmel et al. 2006c, Schleppi et al. 2006a, Birgand et al. 2010). Volume-integrated composite sampling collects frequent water samples in volumes proportional to the flow rates at the times of collection but combines those samples over time to yield fewer samples requiring chemical analysis. Such sampling schemes yield essentially unbiased material flux estimates without requiring the chemical analysis of many samples (Schleppi et al. 2006a, Schleppi et al. 2006b). They also ensure adequate sampling of particulate materials transported during stormflow (e. g., Jordan et al. 1986, Jordan et al. 1997b, a).

A synoptic survey--in which a single spot sample (also called a grab sample) of water is collected from each study site--is a much simpler and cheaper sampling strategy. The low labor and cost enable relatively larger sample sizes to expand spatial coverage or to include watersheds encompassing greater ranges of land use, geology, or other factors relevant to nitrogen export. Synoptic surveys have been criticized because they cannot characterize temporal dynamics (such as seasonality, storm events, or trends) and can yield unrepresentative estimates for the concentrations of materials that vary strongly with stream discharge rate (Kirchner and Neal 2013). Nevertheless, many studies have concluded that synoptic surveys are effective for applications where temporal dynamics are less important, such as for understanding differences among watersheds in average loads or ranking watersheds by important drivers, such as land use, fertilizer application, human population, sewage output (Messer et al. 1988, Kaufmann et al. 1991, Grayson et al. 1997, Wolock et al. 1997). Our own experience suggests that for materials whose concentrations do not increase greatly during storm events, spot sampling can yield water quality data of sufficient accuracy and precision for many important purposes (Weller et al.

2010). Those include quantifying average differences among watersheds in the levels of materials in stream discharge as well as placing watersheds along the gradient of a driving variable, such as geology, fertilizer application, the proportion of cropland, or the prevalence of nitrogen sinks in a watershed (Correll et al. 1995, Jordan and Weller 1996, Liu et al. 2000, Weller et al. 2011, Weller and Baker 2014).

In this paper, we more formally tested the power of spot sampling as a cost-effective way to characterize nitrogen status among nitrogen-enriched watersheds. We compared estimates of stream nitrogen levels based on seasonal spot sampling of stream nitrogen concentration (Liu et al. 2000) to measurements of average annual nitrate and total nitrogen levels for the same watersheds derived from automated monitoring stations performing volume-integrated composite sampling (Jordan et al. 1997b, a, Jordan et al. 2000, Jordan et al. 2003). We focused on nitrate concentration as a potential indicator of total nitrogen level for several reasons. Nitrate concentration is relatively easy to sample and measure, human intervention in the nitrogen cycle often raises nitrate levels in streams and rivers (Caraco and Cole 1999, Seitzinger et al. 2002), and nitrate is often the dominant form of nitrogen in surface waters (Creed and Band 1998, Boyer et al. 2006) even in forested areas (Campbell et al. 2004, Eshleman et al. 2013). We demonstrate that inexpensive spot sampling provides very a strong indicator of the nitrogen levels measured by the more labor intensive and costly automated sampling methods. We conclude that spot sampling of many watersheds can often be more useful and cost-effective way to explore spatial patterns and broad nitrogen-enrichment gradients than more expensive sampling of fewer watersheds. We recommend that spot sampling should be more widely utilized in such efforts.

METHODS

Overview and study area

We used information from two data sets assembled for watersheds in the 166,000 km2 Chesapeake Bay drainage, which extends over four major physiographic provinces--Coastal Plain, Piedmont, Blue Ridge, and Appalachian (Langland et al. 1995)—within the mid-Atlantic region of the United States. We first analyzed data from our own Smithsonian Environmental Research Center study of watersheds within the Chesapeake Bay drainage (here called the SERC data) to quantify and model the relationships between spot nitrogen measurements and measurements from automated monitoring. Then, we confirmed the results and conclusions from the SERC data with an independent data set assembled by the Chesapeake Bay Program Nontidal Network (CBNTN, Chanat et al. 2016, Moyer et al. 2017).

SERC watershed data

Study sites.--For 59 study watersheds distributed across all four major physiographic provinces of the Bay drainage (see map, Fig. 1), we collected seasonal spot samples from the effluent stream and analyzed for nitrate and total nitrogen. For each site, we also established an automated monitoring station that measured stream depth continuously and controlled samplers that collected volume-integrated weekly water samples, which were also analyzed for nitrate and total nitrogen (Jordan et al. 1997b, a). The 59 sites are the subset of watersheds in which both spot and integrated sampling were done, taken from a larger group of 517 study watersheds (Jordan et al. 1997c, a, Jordan et al. 2000, Liu et al. 2000, Jordan et al. 2003, Weller and Baker 2014).

The watersheds are distributed in 14 clusters across the Chesapeake drainage basin (Fig. 1 and Appendix S1, Table S1). The locations of the clusters represent prevalent geological types in each major physiographic province of the Chesapeake Bay drainage basin (Langland et al. 1995) as described in (Liu et al. 2000). Within each cluster, we sampled streams draining watersheds with strongly contrasting land covers to maximize our ability to observe and quantify the effects of land cover on nitrogen discharges and to detect differences in those effects among geological settings. We delineated the boundary and area of the watershed draining to each sampling point by applying automated watershed delineation to digital elevation and stream maps within a geographic information system (GIS, as described in Baker et al. 2006). To quantify watershed land cover, we used the GIS to intersect the watershed boundaries with the 2001 National Land Cover Data set (Homer et al. 2004). To identify the physiographic province of each study watershed, we intersected the GIS layers of study watershed boundaries and physiographic province boundaries (Langland et al. 1995) as previously described (Weller and Baker 2014).

Stream sampling.--From each of the 59 watersheds, we collected 6-22 seasonal spot samples under baseflow conditions over a period of 1-3 years. The sampling periods varied among the watershed clusters, but all were within 1992-2000 (Appendix S1, Table S1). Spot samples were filtered in the field so that subsequent laboratory analyses quantified only the dissolved fractions of nitrogen species. Correll et al. (1995) and Liu et al. (2000) provide more details on the spot sampling methods. An automated monitoring station measured stream depth continuously for 1.3-2.9 years at the outlet of each watershed. The period of automated monitoring in each watershed overlapped with the period of spot sampling (above), and all automated sampling was within 1990-2000. At the Rhode River cluster (arrow in Fig. 1 and stations 101-111 in Appendix S1, Table S1), seven stations used V-notch weirs, so water depth was converted to flow using published equations (Correll 1977, Correll 1981). At all the other watersheds, the automated station monitored stream depth, and we calculated water flow from rating curves of flow vs. depth. The rating curves were calibrated using measurements of depth, cross-sectional area, and flow rate under a range of stream flow conditions (Jordan et al. 1997c, a). The automated stream stations implemented volume-integrated composite sampling by activating pumps to collect water every time a set volume of flow occurred. Thus, the station pumped water more frequently at higher flow rates, so that the composite samples properly represented materials in the water under all flow conditions as well as the contributions from overland stormflow and groundwater emerging in the stream. We retrieved the composite samples weekly for laboratory analysis. The number of weekly samples ranged from 51 to 144 according to the period sampled at each station. Fig. 2 illustrates the results of automated and spot sampling of nitrate concentration relative to discharge monitoring for one station. Previous papers provide more sampling details (Correll 1977, Correll 1981, Jordan et al. 1997c, a).

Stream nitrogen levels.--We measured nitrate concentrations in spot samples with a Dionex Ion Chromatograph model 1400i. In the automated samples, we measured the sum of nitrate and nitrite concentrations by reducing nitrate to nitrite with cadmium amalgam and analyzing nitrite by reaction with sulfanilamide (APHA, 1989). Nitrite concentrations were always very low relative to nitrate, so we refer to their sum as nitrate throughout the paper. Total Kjeldahl nitrogen was determined using the Kjeldahl digestion (Martin 1972; APHA, 1989) and analysis of the resulting ammonium by distillation and Nesslerization (APHA 1989). Total nitrogen is the sum of Total Kjeldahl nitrogen and nitrate. In the composite water samples, ammonium and organic nitrogen can be bound to particulates as well as dissolved, so the nitrogen analyses for composite samples yielded the total of the particulate and dissolved fractions. Because we filtered the spot samples in the field, the nitrogen analyses assessed only the dissolved fractions. Nitrate is not significantly bound to particulates, so the filtered spot samples capture all the nitrate. Dissolved total nitrogen in the spot samples was measured for only 48 of the 59 study watersheds. Previous papers provide more details of the chemical analyses (Jordan et al. 1997c, a, Liu et al. 2000).

Data analysis.--We sought to quantify how well simple spot measurements of nitrogen concentration can predict the average nitrogen concentrations from high quality, flow-weighted composite sampling. We summarized two dependent variables from the composite samples at each site: average total nitrogen concentration (TN) and nitrate concentration (NO3). These flow-weighted averages were calculated by weighting each weekly composite concentration measurement by the volume of discharge during that week. As potential predictors (independent variables), we calculated the simple averages of the seasonal spot measurements of dissolved total nitrogen (sDTN) and nitrate (sNO3) concentrations at each site. Single spot samples are often used to characterize stream water chemistry in stream assessments (Stranko et al. 2017); so we also considered as possible predictors the dissolved total nitrogen and nitrate concentrations (fsDTN and fsNO3) in the first spring spot sample collected between March 1 and May 31--the time-of-year when stream assessment surveys typically collect water samples (e. g., the Maryland Biological Stream Survey, Ashton et al. 2014, Stranko et al. 2017). In all, we tested two spot measurements (sNO3 and fsNO3) as estimators of composite sampled NO3 and four spot measurements (sDTN, fsDTN, sNO3, and fsNO3) as estimators of composite-sampled TN.

We evaluated three approaches for predicting the high quality, flow-weighted measurements of nitrogen concentration from the spot concentration measurements. The first approach simply used the average of the spot measurements (or the first spring spot measurement) as the estimate of true average nitrogen concentration. Many studies have interpreted spot measurements this way, including our previous work (Correll et al. 1995, Liu et al. 2000, Weller et al. 2011, Weller and Baker 2014). The second approach exploited contemporaneous spot and flow-weighted composite samples from the same watersheds to calibrate a linear regression model (R lm function, Venables and Ripley 2002, R Core Team 2017) that predicts flow-weighted average concentration from a spot measurement. This approach quantifies the strength of association between spot and flow-weighted concentrations and identifies possible biases in the simpler first approach. Linear regression also yields a prediction equation for estimating nitrogen concentrations from watersheds where only spot measurements are available. Finally, the regression model quantifies the uncertainty in its estimates by providing confidence limits and prediction intervals.

The third approach applied bootstrap resampling (Efron 1982, Efron and Gong 1983) to enhance the statistical rigor of the regression approach. The variance of concentration measurements is typically greater at higher concentrations, and the residuals of our regression models are bigger at higher concentration (see Results). Such patterns violate the assumption of equal variance among residuals (homoscedasticity) underlying linear regression. Bootstrapping (detailed below) can accommodate heterogeneity in the variances of data and residuals, and it can also quantify the effects of including or excluding influential data points in an analysis. Because bootstrapping accounts for heteroscedasticity and sampling uncertainty, we expected the confidence intervals for parameters and predictions of the bootstrapped regression model to be larger than corresponding intervals for the simple regression, but those larger intervals better represent the true uncertainty of the estimates.

Logarithmic transformation is a simpler and more common solution for analyzing heteroscedastic variables for which the variance increases with the mean (Snedecor and Cochran 1989, Draper and Smith 1998), and log-log regression relating one such variable to another is widely applied in water quality analyses (Helsel and Hirsch 2002). However, statisticians have documented general problems with log-transformation, including failure to eliminate heteroscedasticity and difficulty applying parameter estimates or hypothesis tests back to the untransformed variables (Feng et al. 2013, Feng et al. 2014, Choi 2016, Greenacre 2016, Rendevski et al. 2016, Curran-Everett 2018, Ekwaru and Veugelers 2018). We chose the bootstrapping approach instead of log-log regression because it handled our heteroscedastic variables without causing these problems and because bootstrapping gave other benefits, like quantifying sampling uncertainty. Appendix S1 in the Supporting Information provides a more thorough review of the possible problems with log-transformation. For our data, Appendix S1 also shows that log-log transformation failed to homogenize the variance and produced models that performed poorly for predicting the high nitrogen levels that are of greatest concern in addressing management questions.

We implemented the bootstrap approach in two steps. The first step quantified sampling uncertainty using a pairs bootstrap (Wu 1986, Flachaire 2005), in which we created 2000 bootstrap samples with 59 observations by resampling observations with replacement. For each sample, we fit the linear regression model, and then applied the model to predict flow-weighted average concentrations from spot concentrations ranging from 0 to 18 mg N/L in steps of 1 mg N/L. For each of those 19 values of the independent variable, the median prediction across the bootstrap samples provided the bootstrap prediction of flow-weighted average concentration, and the 2.5th and 97.5th percentile values provided the 95% confidence limits for the median predictions. The 19 median values formed a perfect straight line, and we used the slope and intercept of that line as the coefficients of the linear bootstrap prediction model.

We implemented the second bootstrap step to provide prediction intervals for the estimates of flow-weighted average concentration at a particular site. For each of the 2000 pairs bootstrap samples, we used the fitted linear model to predict flow-weighted average concentrations for all 59 study watersheds in the full data set, and then calculated the model residual (observed-predicted) for each watershed. We then implemented a wild bootstrap—a method developed for heteroscedastic data (Wu 1986, Mammen 1993, Flachaire 2005, Davidson and MacKinnon 2006)--by generating 50 bootstrap samples in which we added to the predicted value for each watershed a resampled residual, calculated by multiplying the residual for that watershed by an independent normally-distributed variate with mean 0 and standard deviation 1 (Roodman et al. 2019). The resampling of model residuals accounts for the variability in flow-weighted average concentration that is *not* explained by the prediction model, so that the 2.5th and 97.5th percentiles across the 100,000 bootstrap samples (2000 pairs X 50 wild) estimate the 95% prediction interval for an individual watershed. We applied loess smoothing (R ggplot2 package, Wickham 2016) across the 59 upper limits and 59 lower limits to provide a smoothed visualization of the prediction interval.

For each set of dependent and independent variables, we quantitatively evaluated the performance of the direct, simple linear, and bootstrapped approaches by comparing the predictions of each approach to the observed data using the gof (goodness of fit) function of the R hydroGOF package (Zambrano-Bigiarini 2020). We report five metrics of skill. Mean error (bias $\bar{\varepsilon}$) and percent bias account for accuracy. Root mean squared error (RMSE) accounts for both accuracy and precision. As a measure of precision alone, we calculated unbiased root mean squared error (ubRMSE) from bias and RMSE by rearranging the equation $\mathrm{RMSE}^2 = \bar{\varepsilon}^2 + \mathrm{ubRMSE}^2$ (Jolliff et al. 2009). We also report the percentage of variance in flow-weighted average concentration explained (R^2) by each approach. We used the R statistical package (R Core Team 2017) for all of the analyses.

CBNTN verification data

Study sites.--The independent verification data for our analysis came from data on streamflow and water chemistry assembled for the Chesapeake Bay Program's Non-tidal Monitoring Network (CBNTN). The data have been curated and analyzed by the U.S. Geological Survey (e.g., Moyer et al. 2017). We analyzed five years of data (water years 2012 -2016 from October 2012 through September 2016) from a subset of 85 watershed sampling sites (Fig. 1, Appendix S1, Table S2) for which stream discharge and loads of total nitrogen and nitrate have been summarized (Chanat et al. 2016, Moyer et al. 2017) and for which digitized watershed outlines are available (Ryberg et al. 2017). We summarized the 2013 National Land Cover Data set

(Yang et al. 2018) to characterize human activities in the watersheds by using a GIS to intersect the digital watershed boundaries (Ryberg et al. 2017) with the NLCD data and then tabulating land cover proportions.

Stream nitrogen levels.--The CBNTN does not employ composite sampling like the SERC study (above). Instead, the CBNTN monitors streamflow continuously and measures material concentrations in discrete water samples collected throughout the year and under different flow conditions. These sparse long-term monitoring data are combined with daily discharge, to characterize episodic, seasonal, and long-term dynamics of nutrients and sediments. During water years 2012-2016, the median number of total nitrogen and nitrate concentration measurements per site was 98 (range 55-193, Appendix S1, Table S2). The USGS applies advanced statistical models to the flow and concentration measurements to estimate material loads, flow-weighted concentrations, and other summary quantities (Moyer et al. 2017). The current model (called Weighted Regressions on Time, Discharge, and Seasonality, WRTDS, Hirsch et al. 2010, Chanat et al. 2016) provides unbiased estimates of nitrogen and nitrate loads (Zhang et al. 2019). For the 85 study watersheds in water years 2012-2016, we extracted the 60 monthly estimates of discharge and the average concentrations of total nitrogen and nitrate from a recent WRTDS summary of the CBNTN (Moyer et al. 2017). For each watershed, we calculated the five-year (2012-2016) average nitrate and total nitrogen concentrations as the weighted average of the 60 monthly concentrations weighted by the product of monthly discharge and month length in days.

Data analysis.--Like the SERC composite sample data, the integrated estimates of average concentration from the WRTDS analysis were treated as the dependent variable--flow-weighted average concentration--to be estimated from simpler spot sampling. The independent predictor variables we evaluated were a single, discrete measurement of total nitrogen and nitrate concentration. For each site, we selected from the CBNTN concentration data base the first uncensored TN and NO3 measurements taken in March 2012 (see Moyer et al. 2012) for information on censoring). The month of March begins the period when streams are commonly visited for stream assessment (see above). We call these potential predictors fsTN and fsNO3 in the rest of the paper.

We applied the same data analyses used for the SERC data: first summarizing the watershed characteristics and concentration data and then exploring relationships between the spot concentrations and the flow-weighted average concentrations from the WRTDS analysis of the CBNTN data. We compared WRTDS flow-weighted average nitrate (NO3) concentration to the first spring spot nitrate concentration (fsNO3), and we related average WRTDS total nitrogen (TN) to the first spring spot measures of total nitrogen and nitrate (fsTN and fsNO3). We applied the same three prediction approaches (direct substitution, simple linear regression, and bootstrapped linear regression) and evaluated them with the same metrics of model skill (as described above, but with 85 CBNTN watersheds instead of 59 SERC watersheds). Like the SERC analyses, the CBNTN analyses test how well simple spot samples can predict average nitrogen concentration as measured by much more thorough and expensive sampling and modeling (composite sampling for SERC, advanced WRTDS synthesis for the CBNTN). We evaluated if patterns and performance for the CBNTN data supported findings from SERC data.

RESULTS

SERC study watersheds

Watershed geographic characteristics.--The sizes of the 59 study watersheds range from 0.05 to 324 km^2 (median 9.61, mean 28.0, Fig. 3a, Appendix S1, Table S3). The data set includes watersheds that are entirely natural forest and wetland as well as watersheds that are mostly agricultural or developed land. Land cover percentages range from 3-100% forest (median 44%, mean 50%); 0-2% wetland (median 0.08%, mean 0.25%), 0-61% cropland (median 6.6, mean 11%), 0-70% grassland (median 26%, mean 30%), and 0-80% developed land (median 1%, mean 7%; Appendix S1, Table S3).

A three-dimensional plot of three aggregated land cover categories illustrates the dominant patterns of human land cover disturbance across the data set (Fig. 3b). The three aggregates are cropland plus grassland (agricultural land), forest plus wetland (natural land), and developed land. Rural watersheds lie along the diagonal line in the plane of forest plus wetland vs. cropland plus grassland where the two aggregate categories together cover almost all of the land. Developed watersheds fall off that line and above that plane, reflecting the past replacement of natural and agricultural land with developed land. The data set includes watersheds from all four major physiographic provinces comprising the Chesapeake Bay drainage (Coastal Plain, 25 watersheds; Piedmont, 19; Appalachian Mountain, 8; and Appalachian Plateau, 7).

Stream nitrogen levels.--Flow-weighted average composite-sampled nitrogen concentrations ranged from very low (0.01 mg NO3-N/L and 0.12 mg TN/L) to quite high (16.2 mg NO3-N/L and 17.5 mg TN/L, Fig. 4 and Appendix S1, Table S4), reflecting the range from very low to high levels of human activity (and associated nitrogen enrichment) revealed by the land cover data (Fig. 3b). The distributions of NO3 and TN concentrations were positively (right) skewed, with more low values and fewer high values (Fig. 4). The central values and ranges of the flow-weighted average concentrations and spot sampled concentrations were similar (Fig. 5, Table S5, and figures below), and variability in flow-weighted average nitrogen concentration was heteroscedastic, with variability increasing with the mean of either flow-weighted average NO3 or TN as well as spot NO3 or TN (Fig. 4, Appendix S1, Table S5).

Estimating average concentration from spot measurements.--Spot concentration measurements were very strong predictors of flow-weighted average nitrate concentration regardless of prediction method, but the method did affect bias and confidence limits for the predictions (Table 1). The simplest method used the spot measurements as a direct estimator of flow-weighted average concentration (Fig. 5). For predicting flow-weighted average NO3 from average spot sNO3, these direct estimates explained 98.3% of the variability among watersheds in flow-weighted average NO3 (R^2 in Table 1), but tended to overestimate (positive percent bias of 11.0%, Table 1) because nitrate in baseflow is often higher than in stormflow or overall (see Discussion). Implementing a simple linear regression did not change the amount of variability in flow-weighted average NO3 explained, but it did eliminate the overestimation bias by fitting a regression slope less than one (0.970, percent bias 0%, Table 1, Fig. 6a). Unlike direct substitution (Fig. 5), the linear model also provided confidence and prediction intervals, which were quite narrow (Fig. 6a), reflecting the high R^2 and low residual variation of the regression (Table 1). Unlike the simple regression, the bootstrap model accounted for heteroscedasticity in the concentration measurements (Fig. 6b) as well as for sampling uncertainty in the predictions, especially uncertainty arising from including or excluding watersheds. The bootstrap method achieved a slightly higher proportion of variance explained (R^2=98.7%, Table 1, but had a small negative bias (percent bias=-0.04%, Table 1) and a slightly shallower regression slope (0.952). More importantly, the 95% confidence and prediction intervals of the bootstrap model (Fig. 6b) were wider than those of the simple regression (Fig. 6a), especially at higher nitrate levels. This reflects the ability of the bootstrap method to account for sampling uncertainty and heteroscedasticity.

Not surprisingly, the watershed with the highest observed nitrogen concentrations (uppermost point in Fig. 5 and Fig. 6a-f; watershed 522 in Fig. 3b and Appendix S1, Table S2) had a strong influence on the regression results. The distributions of parameters and predictions of the bootstrapped NO3 vs. sNO3 model (Fig. 7) reveal that influence. The distribution of regression slope estimates is bimodal (Fig. 7a). The left mode has a median slope of 0.904 and summarizes bootstrap samples omitting watershed 522. The right mode for bootstrap samples including station 522 has a steeper median slope of 0.985 (the median of all bootstrap samples is 0.970, Table 1). Bimodality in the slope estimates yields bimodal predictions of flow-weighted average nitrate at high levels of spot nitrate (Fig. 7b), but not at low levels of spot nitrate (Fig. 7c). The ability of the bootstrap model to account for the sampling uncertainty arising from including or excluding influential observations like watershed 522 demonstrates one advantage of the bootstrap approach. The high uncertainty at high nitrogen levels also indicates a need to sample more high-nitrogen watersheds to reduce the sensitivity of the results to influential observations like watershed 522.

Despite the advantages of bootstrapped estimates over the simple regressions (Table 1), the associations between flow-weighted average concentration and spot concentrations are so strong that even the simple regressions yield very good predictions. The simple regression predictions might be adequate for applications that require only predictions of mean concentration; however, for applications that also need uncertainty estimates, the ability of the bootstrap regression to account for uncertainties from heteroscedasticity and sampling error becomes more important.

The single first spring spot nitrate sample (fsNO3) was almost as good a predictor of flow-weighted average NO3 concentration as the average based on 6-22 spot nitrate samples per station (sNO3). The percent of variability in flow-weighted average concentration explained by the bootstrapped model for fsNO3 (R^2=97.2%) was slightly lower than the bootstrapped model based on average sNO3 (R^2=98.7%), and the 95% confidence and prediction limits for the first spot model (Fig. 6c) were wider than were those of the average spot model (Fig. 6b).

Spot concentration measurements were also very effective predictors of the flow-weighted average total nitrogen concentration from composite sampling. We explored four possible predictors of composite TN: average spot dissolved nitrogen (sDTN): first spot dissolved nitrogen (fsDTN), sNO3, and fsNO3. For all four predictors, we saw the same bias in the direct method and the same enhancements with the simple regression and bootstrap methods as reported above for fsNO3 (Fig. 6, Table 1). For the bootstrapped models, average spot total nitrogen concentration (sDTN) was a slightly better predictor (R^2=98.6%; Fig. 6d; Table 1) than average spot nitrate (sNO3, R^2=97.8%, Fig. 6e, Table 1), and first spot concentrations were slightly weaker predictors than their corresponding average spot concentrations (fsDTN, R^2=98.0%; fsNO3, R^2=96.7%, Table 1, Fig. 6f). Importantly, even the single first spring spot nitrate sample provided a very strong indication of total nitrogen concentration (R^2=96.7%, Fig. 6f).

CBNTN study watersheds

Watershed geographic characteristics.--The 85 watersheds in the CBNTN verification dataset were substantially larger than watersheds in the SERC study (Fig. 1 and Fig. 3a). The CBNTN watershed areas ranged from 19.7 to 70,162 km^2. The median (666 km^2) and mean (4,854 km^2) for CBNTN watersheds were 69 and 173 times larger, respectively, than for the SERC watersheds (Fig. 3a, Appendix S1, Table S6). Land cover combinations across the CBNTN and SERC data sets were generally similar (Fig. 3b), but the SERC set does include more watersheds with extreme land cover proportions (high agricultural land, near 100% natural land, or high developed land, Fig. 3b; Tables S3 and S6). Among the CBNTN watersheds, land cover (Appendix S1, Table S6) ranged from 23-91% forest (median 61%, mean 58%), 0-0.79% wetland (median 0.07%, mean 0.15%), 0 to 63% cropland (median 5%, mean 11%), 0.4 to 51% grassland (median 17%, mean 18%), and 2-73% developed land (median 7%, mean 12%).

Stream nitrogen levels.--Among the CBNTN watersheds, the five-year, flow-weighted averages of monthly nitrogen concentrations estimated by WRTDS ranged from very low (0.03 mg NO3-N/L and 0.295 mg TN/L) to high (7.20 mg NO3-N/L and 7.89 mg TN/L, see Appendix S1, Table S7). As with the SERC data, the central values and ranges of the flow-weighted average concentrations and spot sampled concentrations were similar (Fig. 4, Fig. 6g-i), and variability in flow-weighted average nitrate or spot nitrate concentration was heteroscedastic (Fig. 4, Appendix S1, Table S7). The distributions of flow-weighted average nitrate and total nitrogen concentrations among the CBNTN watersheds are roughly like the distributions for the SERC watersheds (Fig. 4). All the distributions are skewed right with many low values and few high values, but the SERC data set includes four watersheds with nitrate and total nitrogen values above the maxima in the CBNTN data (Fig. 4). Those four SERC watersheds all had high levels of agricultural land (Tables S3 and S4) and lie close to the apex representing high percentages of cleared land in the graph of land cover proportions (Fig. 3b).

Estimating average concentration from spot measurements.--For the CBNTN watersheds, the first spot concentration measurements were very strong predictors of the five-year, flow weighted nitrogen concentrations from WRTDS synthesis (Table 1, Fig. 6g-i). As with the SERC data, the direct method produced biased estimates of flow-weighted average concentration, but the regression method removed the bias (Table 1). The bootstrap method again explained the most variation in flow-weighted average concentration while also accounting for sampling error and heteroscedasticity. Compared to the SERC results, the proportions of variance explained were slightly lower and the regression slopes were shallower (Table 1). For example, to predict flow-weighted average nitrate concentration from first spot nitrate, the bootstrapped SERC model had R^2=97% and slope=0.993 while the CBNTN model had R^2=94.5% and slope=0.843. Importantly, a single spring spot sample of nitrate concentration was again a remarkably effective predictor of the five-year average total nitrogen level estimated by advanced statistical synthesis (WRTDS) of daily flow data and 55-193 (median 98) individual TN measurements per station (Fig. 6i, Table 1).

DISCUSSION

Central findings

Our main conclusions are that simple spot sampling provides a surprisingly effective way to estimate average nitrogen levels in streams (Table 1, Fig. 5 and Fig. 6) and that, for some purposes, more costly and laborious sampling programs may not be needed (see Applications section below). We demonstrated the effectiveness of spot sampling with two independent sets of study watersheds: relatively small watersheds from the SERC study and much larger watersheds from the CBNTN sampling network. The two data sets gave slightly different slopes relating flow-weighted average concentrations to spot measurements (Table 1), likely because of differences in methods of sampling, laboratory analysis, data synthesis (see Methods section) and the ranges of nitrogen concentrations actually sampled (Fig. 4). However, the differences in the relationships between the two data sets are small in a combined plot of the two data sets (Fig. 8). In both data sets, just one spring spot sample was a strong predictor of flow-weighted average nitrogen levels. Importantly, each data set shows that one relationship between flow-weighted average concentration and spot measurements works well for all the study watersheds (Fig. 6), despite strong differences among physiographic provinces in how land use affects stream nitrogen levels (Jordan et al. 1997c, Liu et al. 2000, Jordan et al. 2003, Weller et al. 2003, Weller et al. 2011, Weller and Baker 2014).

Spot surveys have long been conducted to complement to automated watershed sampling (Messer et al. 1988, Kaufmann et al. 1991, Grayson et al. 1997, Wolock et al. 1997), and several studies have reported strong correlations of spot measurements with better measurements (Schleppi et al. 2006a, Schleppi et al. 2006b, Rozemeijer et al. 2010, Abbott et al. 2018). More recently, McCarthy and Haggard (2016) recommended that spot sampling alone may be sufficient for many nutrient management purposes. Schlelppi et al. (2006b) recommended using parallel measurements of spot and flow-weighted samples to calibrate the first against the second. We extended that idea by calibrating spot measurements against multiyear, flow-weighted measurements to estimate nitrogen levels for many watersheds spanning a gradient from pristine to strongly agricultural or developed. Our analyses more formally tested the ability of spot samples to estimate multiyear, flow-weighted average concentrations. Our results rigorously demonstrate and quantify the very high efficiency of spot sampling for estimating multiyear, flow-weighted average nitrogen concentrations for nitrogen-enriched watersheds (Table 1, Fig. 6).

Our results for watersheds in the Chesapeake Bay drainage should be relevant in other regions with significant rainfall and nitrogen enrichment from human population or agricultural activities. Our findings are less relevant for areas like the western United States, where human population, nitrogen fertilization, and rainfall are all low and nitrate is not the dominant component of stream nitrogen (Scott et al. 2007).

Why does this work so well?

There are several reasons why spot sampling is such a surprisingly effective predictor of flow-weighted average nitrate and total nitrogen concentrations. One key factor is the way nitrate is transported through watersheds and streams. Sediment and nutrients that are primarily bound to particles (like phosphorus) are mobilized during storms and transported to streams by surface flow; therefore, their stream concentrations during storms can be orders of magnitude greater than in baseflow (Correll et al. 1999c). In contrast, nitrate is not strongly bound to soils or to suspended sediments, so it moves freely in dissolved form. In many watersheds, nitrate is transported toward streams primarily in subsurface flow and groundwater, and is often somewhat diluted during storm events so that stream nitrate concentrations are lower during storms than during baseflow (Jordan et al. 1997c, Correll et al. 1999c, Rozemeijer et al. 2010, McCarty and Haggard 2016). Because nitrate concentrations are not wildly amplified during storms, baseflow nitrate concentration is much more representative of stormflow concentration and of overall average nitrate concentration than are the baseflow concentrations of materials that are mainly transported on particles.

Secondly, nitrate is the dominant chemical form of total nitrogen in major rivers (Caraco and Cole 1999, Seitzinger et al. 2002) and in in streams draining smaller watersheds (Fig. 7 and Creed and Band 1998, Boyer et al. 2006), even many forested ones (Campbell et al. 2004, Eshleman et al. 2013). Stream nitrate levels increase much more strongly with increasing human impacts from agriculture and land development than do other forms of nitrogen (Fig. 7 and Jordan et al. 1997b, a, Liu et al. 2000, Jordan et al. 2003, Golden et al. 2009). Among the watersheds we examined, nitrate becomes the majority of TN when TN reaches 0.6 mg N/L for the CBNTN watersheds and 2 mg N/L for the SERC watersheds (Fig. 9). Above those levels, nitrate increasingly dominates TN as TN levels rise further. The strong dominance of total nitrogen levels by nitrate, especially in streams draining human-impacted watersheds, means that dissolved nitrate in baseflow spot samples is strongly associated with total nitrogen concentrations as well as total nitrate concentrations.

Finally, among watersheds ranging from low to high levels nitrogen enrichment, both nitrate and total nitrogen show more spatial variation among watersheds than temporal variation within watersheds. We quantified the fraction of total variability among watersheds and weeks that is due to differences among watersheds for the weekly SERC and monthly CBNTN concentration data. We used a linear model with site number as a categorical random variable (R lmer function, Bates et al. 2015). Heteroscadasticy in the concentration measurements was not a concern for this model because we used it only to estimate the among-watershed fraction of total variability, *not* to estimate P values for hypothesis tests. For nitrate and total nitrogen concentrations in both data sets, the linear model explained 93% (NO3) and 87% (TN) of the total variation among all watersheds and weeks in the SERC data as well as 97% (both NO3 and TN) among all watersheds and months in the CBNTN data. Thus, 87% or more of the total variation can be attributed to differences among watersheds, leaving only 13% or less of the total variation to be attributable to temporal variation and error. Spot sampling does not effectively account for temporal variation (Kirchner and Neal 2013), but that did not much limit the ability of spot samples to predict flow-weighted average nitrate and total nitrogen concentrations because temporal variation in those concentrations was much smaller than the differences in concentration among watersheds.

We emphasize that the dominance of spatial variation among watersheds relative to temporal variation at a watershed does not mean that the temporal variation is unimportant. To the contrary, we observed substantial temporal variation at each site in both data sets (Fig. 2, Fig. 5, Tables S4 and S7), and nitrogen levels are known to vary among years, seasons, and storm events (Correll et al. 1999b, c, a, Kirchner and Neal 2013, Abbott et al. 2018). Measuring that variability and understanding its causes are critical to addressing many questions in nitrogen cycling and nitrogen management, but not so critical to the task of placing the temporally averaged nitrogen levels for watersheds across a broad gradient of nitrogen enrichment.

Other water quality constituents

This paper is focused on testing the ability of spot samples to match the average nitrate and total nitrogen concentrations sampled by more costly and labor-intensive methods, but the SERC and CBNTN programs also measured other water quality constituents. These were dissolved silicate (Si), total ammonium (NH4), total Kjeldahl nitrogen (TKN), total phosphorus (TP), total ortho-phosphate (PO4), and total organic carbon (TOC) in the SERC study (Jordan et al. 1997b, a) and total phosphorus (TP), dissolved ortho phosphate (PO4), and total suspended sediment (TSS) in the CBNTN program (Chanat et al. 2016, Moyer et al. 2017). We again used linear regression to quantify the ability of spot samples to predict the higher quality concentration estimates for these additional constituents (see Methods). We also again used a linear model (R lmer function, Bates et al. 2015) with site number as a random categorical variable to assess amount of the total variation among high quality measurements attributable to differences among watersheds rather than to temporal variability within watersheds (Table 2).

For dissolved silicate (SERC watersheds), the average spot sample concentration was a very strong predictor (R^2=97%) of the average concentration from flow-weighted composite samples (Table 2). Like nitrate, Si is diluted rather than amplified during storm events, as are other mostly dissolved constituents, (such as Ca, Mg, K, Na, SO4, Cl, NO3, and conductivity, Schleppi et al. 2006b). As with nitrate and total nitrogen, most of the total variability in dissolved silicate among weeks and watersheds is explained by differences among watersheds (84.9%), with much less variability potentially due to temporal variation within watersheds (Table 2). In contrast, the other additional constituents from both data sets are materials that are transported mostly on particles (Jordan et al. 1997b, a). Comparted to nitrate, total nitrogen, and dissolved silicate; spot samples are much less effective at predicting flow-weighted average concentration for the materials transported on particles (R^2>92% for dissolved materials, R^2<52% for particulates, Table 2). Furthermore, the proportion of the total variability in flow-weighted average concentration due to differences among watersheds is much lower for materials transported mostly on particles than for dissolved materials, so that the importance of temporal variability within watersheds is greater for particulate-transported materials. Temporal variability appears to be more dominant in the SERC data (>85% of total variability) than in the CBNTN data (>43% of total variability, Table 2) due to differences in watershed size (smaller watersheds are more temporally variable (Abbott et al. 2018) and data frequency--the weekly SERC data inherently capture more temporal variation than the monthly CBNTN estimates.

The SERC and CBNTN data sets both support the conclusion that spot measurements are very good predictors of flow-weighted average concentration for materials transported in dissolved form, but much less effective for estimating flow-weighted average concentrations of materials that bind to particles . This is consistent with other reports of much higher correlations for nitrogen than phosphorus when comparing spot samples to composite samples (Schleppi et al. 2006a, Schleppi et al. 2006b) or baseflow spot samples to storm samples (McCarty and Haggard 2016). Table 2 also supports ranking nitrogen>phosphorus>sediment in order of predictability as reported for a variety of modeling approaches (Weller et al. 2003, Brakebill et al. 2010, Preston et al. 2011, Boomer et al. 2013).

Our analyses relate to the idea of spatial stability presented by Abbott et al. (2018). They developed concepts and methods to quantify patterns of spatial and temporal variability in water quality within stream networks, and they discussed the ecological and hydrological implications of those patterns. They proposed the correlation between instantaneous and longer-term concentrations (as in our Tables 1 and 2) as a direct measure of spatial stability of water chemistry patterns, and they suggested that temporal synchrony among watersheds promotes spatial stability. Our analysis of the proportion of total variability among watersheds and sampling times due to spatial differences among watershed (Table 2) provides another measure of spatial stability, and the results suggest that the domination of total variability by differences among stations also promotes spatial stability. Abbott et al. (2018) argue that spatial stability determines the sampling frequency needed to identify and evaluate critical source areas and that synoptic sampling can be useful for those purposes when water quality patterns are spatially stable. In our data, the very high spatial stability of nitrate and total nitrogen levels across a broad nitrogen enrichment gradient (Table 2) suggests that just one spot sample may be adequate for such evaluations of those materials.

Application to science and management

Synoptic spot sampling is already widely used in reconnaissance efforts to measure baseline levels, identify water quality problems, target critical source areas, or measure compliance (NRCS 2003), often as a complement to more frequent automated sampling at a few selected locations (Messer et al. 1988, Kaufmann et al. 1991, Grayson et al. 1997, Wolock et al. 1997). Synoptic sampling provides data for more locations, helps assess relative importance of sources throughout a watershed, and is often interpreted to identify landscape parameters and ecosystem processes correlated with water chemistry (Liu et al. 2000).

The relatively low costs for labor and laboratory analysis are a prime advantage of synoptic sampling over frequent automated monitoring. (Harmel et al. 2006c) note that success of monitoring projects depends on careful attention to the tradeoff between the resources available for data collection and adequate characterization of water quality. Automated samplers typically yield better data but are especially expensive compared to manual sampling. The cost of automated monitoring is a significant obstacle to assessing large numbers of watersheds and restricts data available for analysis.

We demonstrate statistically that spot sampling is even more effective than previously reported, especially for placing average nitrogen levels in watershed discharges within broad enrichment gradient (Fig. 6). For this purpose, the SERC data revealed that a single spot sample was almost as effective as the far greater and more costly effort of monitoring flow continuously and collecting and analyzing 52 weekly composite samples for 1-3 years (Fig. 6c, e). Similarly, the

CBNTN analysis showed that a single spot sample was almost as effective for assessing averaging nitrogen concentration as monitoring flow continuously, collecting and analyzing an average of 98 water samples per site, and integrating the flow and concentration data with an advanced statistical model (Fig. 6i). Of course, the more detailed CBNTN protocols remain necessary to meet the CBNTN goal of characterizing nutrient and sediment dynamics at multiple temporal scales, including events, seasons, years, and multiyear trends.

Spot sampling may be adequate to meet some purposes for which more expensive sampling methods are now recommended. Current recommendations suggest automated or composite sampling for measuring fate and transport, program effectiveness, and research (NRCS 2003) as well as for predicting longer term longer term water quality, especially for smaller systems with high temporal variability (Kirchner and Neal 2013). Cassidy and Jordan (2011) state that only near-continuous monitoring is adequate for comparative monitoring and evaluation. However, many research and management issues lead to questions about how *average* nitrogen levels compare among watersheds or before and after management interventions. Our results suggest that, for nitrogen, spot sampling can be adequate for answering those questions (Fig. 6), even given high temporal variability in nitrogen levels in individual watersheds (Fig. 2 and Fig. 5). When the focus is on differences in average nitrogen levels among watersheds driven by different amounts of nitrogen enrichment, frequent sampling may not be needed. Our results also support stream assessment protocols that collect one spring nitrate sample to assess watershed and stream nitrogen status (Ashton et al. 2014, Stranko et al. 2017).

Given the effectiveness of spot sampling (Table 1, Fig. 6), we support its more widespread application in nitrogen assessment and management. McCarty and Haggard (2016) made a similar recommendation. They argued for a revolution in allocating water quality monitoring resources by using spot sampling of baseflow to assess nitrogen and phosphorus pollution and to target management actions, thus freeing resources to examine water-quality at finer spatial scales and to provide a more complete information on spatial variability in water quality across watersheds. Our analysis strongly supports their recommendation for nitrogen management and assessment, but less so for phosphorus (see below).

Other authors have also emphasized the need for better spatial coverage in water sampling. (Abbott et al. 2018) highlighted the need to understand sources and sinks in headwater catchments where the vast majority of water and solutes enter aquatic ecosystems (Alexander et al. 2007, Baker et al. 2007, Bishop et al. 2008, McDonnell and Beven 2014). Those headwater systems are where water quality problems originate, yet they are too numerous (thousands or more in large river systems) to monitor frequently, presenting a "headwater conundrum", which can be resolved with synoptic sampling (Abbott et al. 2018).

Spot sampling of stream nitrate could be especially useful in citizen science efforts to assess water quality. Such efforts engage citizen volunteers to expand the capabilities of research or assessment teams and to educate citizens about science and management issues. Nitrate monitoring with baseflow sampling could be a part of a citizen monitoring program, requiring only minimal training in sample collecting, sample storage, and using smartphone global positioning to locate and document sampling sites.

Enthusiasm for the success of spot sampling in predicting flow-weighted average nitrogen levels (Fig. 6) should be tempered when considering phosphorus or other materials transported mainly on particles. McCarty and Haggard (2016) suggested using baseflow sampling for assessing other materials, such as phosphorus. We did find statistically significant correlations between spot measurements of phosphorus and flow-weighted average levels in composite measurements, but those relationships have much lower explanatory power ($R^2 < 42\%$) than the relationships for nitrate and total nitrogen ($R^2 >82\%$, Table 2). For nitrogen levels, spatial differences among watersheds explain more of the observed variability than does temporal variation within watersheds, but the opposite is true for phosphorus and other particulates (Table 2). Nor does good information on nitrogen levels help much with assessing phosphorus levels. The correlation between flow-weighted average total phosphorus and total nitrogen is weak and not significant in both data sets (R^2=8%, P=.07 for SERC composite samples and R^2=0.1%, P=0.7, for CBNTN estimates from WRTDS synthesis). Successful assessment of phosphorus levels and other particulates continues to demand monitoring methods that capture episodic, high concentrations occurring during storm events.

Is bootstrapping really necessary?

We fit linear relationships using a two-step bootstrapping procedure. Many practitioners may not have the time or interest to implement bootstrapping, and they will seek easier ways to calibrate relationships predicting multiyear average nitrogen levels from spot sample measurements. In our analyses of nine linear relationships (six for SERC data and three for CBNTN), the slopes, intercepts, and R^2 values from simple linear regression closely match those from bootstrapping (Table 1). The two approaches give very similar predictions of multiyear average nitrogen levels, but bootstrapping gives wider confidence limits (compare Fig. 6a to Fig. 6b) because bootstrapping accounts for heteroscedasticity and sampling uncertainty while a simple linear model does not. These results suggest that a simple linear model might be adequate for applications that need to predict average nitrogen levels but do not need estimates of confidence limits. In contrast, simple linear models fit to log-log transformed variables did not perform well for our data. The transformation did not eliminate heteroscedasticity, and the models underpredicted for watersheds with high nitrogen levels--the most important watersheds for many research and management questions (see SERC Data Analysis section in Methods and Appendix S1, Supporting Information). Analyses of our data suggest that simple linear regression using untransformed data would provide the most accurate shortcut for avoiding the bootstrapping method. However, when needs include confidence limits or significance tests, not just predictions, a procedure like bootstrapping should be included to account for heteroscedasticity and sampling uncertainty.

CONCLUSION

The key findings of this study include:

- Spot sample measurements estimate average nitrate and total nitrogen concentration in streams draining nitrogen-enriched watersheds almost as effectively as multiyear data from flow-weighted composite sampling or from WRTDS synthesis of continuous flow measurements and frequent water samples.
- Estimates from spot samples are unbiased when implemented using calibrated relationships between spot measurements and flow-weighted composite or WRTDS measurements.
- A simple linear regression works very well for fitting the calibrated relationships, but bootstrapping can make the analysis more rigorous by accounting for sampling error and heteroscedasticity.
- Even a single spring spot sample can efficiently place watersheds within a broad gradient of anthropogenic watershed nitrogen loading.

- Spot sampling of nitrogen works well because it is transported to streams primarily as nitrate dissolved in subsurface flow rather attached to particles in surface flow during storms.
- For nitrogen levels in the data sets we examined, more of the total variability across places and times was due to spatial differences among study watersheds than to temporal variation within watersheds.
- Spot measurement of stream nitrate is a low cost, low labor way to quantify average nitrogen status.
- Spot sampling can be a powerful tool for identifying nitrogen source areas and monitoring the results of nitrogen management actions.
- Spot sampling should be more widely applied to make nitrogen assessment and management programs more expansive and cost effective.
- Spot sampling is much less effective for materials that are mainly transported on particles, like phosphorus, so spot samples of such materials should be interpreted cautiously.

ACKNOWLEDGMENTS

The SERC stream data were collected with support from NSF (BSR-89-05219, DEB-92-06811, and DEB-93-17968), NOAA (NA66RG0129), the Governor's Research Council of Maryland, the government of Charles County Maryland, and the Smithsonian Institution Environmental Sciences Program. David L. Correll initiated the SERC watershed study. We thank U.S.G.S. scientists Douglas L. Moyer, and Jeffrey G. Chanat for help accessing and interpreting the data from the Chesapeake Bay Nontidal Network (CBNTN).

LITERATURE CITED

Abbott, B. W., G. Gruau, J. P. Zarnetske, F. Moatar, L. Barbe, Z. Thomas, O. Fovet, T. Kolbe, S. Gu, A. C. Pierson-Wickmann, P. Davy, and G. Pinay. 2018. Unexpected spatial stability of water chemistry in headwater stream networks. Ecology Letters 21:296-308.

Alexander, R. B., E. W. Boyer, R. A. Smith, G. E. Schwarz, and R. B. Moore. 2007. The role of headwater streams in downstream water quality. Journal of the American Water Resources Association 43:41-59.

Ashton, M. J., R. P. Morgan, 2nd, and S. Stranko. 2014. Relations between macroinvertebrates, nutrients, and water quality criteria in wadeable streams of Maryland, USA. Environmental Monitoring and Assessment 186:1167-1182.

Baker, M. E., D. E. Weller, and T. E. Jordan. 2006. Comparison of automated watershed delineations: Effects on land cover areas, percentages, and relationships to nutrient discharge. Photogrammetric Engineering and Remote Sensing 72:159-168.

Baker, M. E., D. E. Weller, and T. E. Jordan. 2007. Effects of stream map resolution on measures of riparian buffer distribution and nutrient retention potential. Landscape Ecology 22:973-992.

Bates, D., M. Mächler, B. Bolker, and S. Walker. 2015. Fitting linear mixed-effects models using lme4. Journal of Statistical Software 67:1-48.

Birgand, F., C. Faucheux, G. Gruau, B. Augeard, F. Moatar, and P. Bordenave. 2010. Uncertainties in assessing annual nitrate loads and concentration indicators: part 1. Impact of sampling frequency and load estimation algorithms. Transactions of the ASABE 53:437-446.

Bishop, K., I. Buffam, M. Erlandsson, J. Folster, H. Laudon, J. Seibert, and J. Temnerud. 2008. Aqua incognita: the unknown headwaters. Hydrological Processes 22:1239-1242.

Boesch, D. F. 2019. Barriers and bridges in abating coastal eutrophication. Frontiers in Marine Science 6:123.

Boomer, K. M. B., D. E. Weller, T. E. Jordan, L. Linker, Z.-J. Liu, J. Reilly, G. Shenk, and A. A. Voinov. 2013. Using multiple watershed models to predict water, nitrogen, and phosphorus discharges to the Patuxent estuary. Journal of the American Water Resources Association 49:15-39.

Boyer, E. W., R. B. Alexander, W. J. Parton, C. Li, K. Butterbach-Bahl, S. D. Donner, R. W. Skaggs, and S. J. D. Grosso. 2006. Modeling denitrification in terrestrial and aquatic ecosystems at regional scales. Ecological Applications 16:2123-2142.

Boyer, E. W., C. L. Goodale, N. A. Jaworski, and R. W. Howarth. 2002. Anthropogenic nitrogen sources and relationships to riverine nitrogen export in the northeastern U.S.A. Biogeochemistry 57/58:137-169.

Brakebill, J. W., S. W. Ator, and G. E. Schwarz. 2010. Sources of suspended-sediment flux in streams of the Chesapeake Bay Watershed: A regional application of the SPARROW model. Journal of the American Water Resources Association 46:757-776.

Brookshire, E. N. J., S. Gerber, J. R. Webster, J. M. Vose, and W. T. Swank. 2011. Direct effects of temperature on forest nitrogen cycling revealed through analysis of long-term watershed records. Global Change Biology 17:297-308.

Campbell, J. L., J. W. Hornbeck, M. J. Mitchell, M. B. Adams, M. S. Castro, C. T. Driscoll, J. S. Kahl, J. N. Kochenderfer, G. E. Likens, J. A. Lynch, P. S. Murdoch, S. J. Nelson, and J. B. Shanley. 2004. Input-output budgets of inorganic nitrogen for 24 forest watersheds in the northeastern United States: A review. Water Air and Soil Pollution 151:373-396.

Caraco, N. F., and J. J. Cole. 1999. Human impact on nitrate export: An analysis using major world rivers. Ambio 28:167-170.

Cassidy, R., and P. Jordan. 2011. Limitations of instantaneous water quality sampling in surface-water catchments: Comparison with near-continuous phosphorus time-series data. Journal of Hydrology 405:182-193.

Chanat, J. G., D. L. Moyer, J. D. Blomquist, K. E. Hyer, and M. J. Langland. 2016. Application of a weighted regression model for reporting nutrient and sediment concentrations, fluxes, and trends in concentration and flux for the Chesapeake Bay nontidal water-quality monitoring network, results through water year 2012. Scientific Investigations Report 2015-5133. U.S. Geological Survey, Reston, Virginia USA.

Choi, S. 2016. Life is lognormal! What to do when your data does not follow a normal distribution. Anaesthesia 71:1363-1366.

Correll, D. L. 1977. An overview of the Rhode River watershed research program. Pages 105-123 *in* D. L. Correll, editor. Watershed research in eastern North America - a workshop to compare results. Smithsonian Institution, Edgewater, Md.

Correll, D. L. 1981. Nutrient mass balances for the watershed, headwaters intertidal zone, and basin of the Rhode River Estuary. Limnology and Oceanography 26:1142-1149.
Correll, D. L., T. E. Jordan, and D. E. Weller. 1995. The Chesapeake Bay watershed: effects of land use and geology on dissolved nitrogen concentrations. Pages 639-648 *in* P. Hill and S. Nelson, editors. Toward a sustainable coastal watershed: The Chesapeake experiment. Chesapeake Research Consortium, Solomons, Maryland USA.
Correll, D. L., T. E. Jordan, and D. E. Weller. 1999a. Effects of precipitation and air temperature on nitrogen discharges from Rhode River watersheds. Water, Air, and Soil Pollution 115:547-575.
Correll, D. L., T. E. Jordan, and D. E. Weller. 1999b. Precipitation Effects on Sediment and Associated Nutrient Discharges from Rhode River Watersheds. Journal of Environmental Quality 28:1897-1907.
Correll, D. L., T. E. Jordan, and D. E. Weller. 1999c. Transport of nitrogen and phosphorus from rhode river watersheds during storm events. Water Resources Research 35:2513-2521.
Creed, I. F., and L. E. Band. 1998. Export of nitrogen from catchments within a temperate forest: Evidence for a unifying mechanism regulated by variable source area dynamics. Water Resources Research 34:3105-3120.
Curran-Everett, D. 2018. Explorations in statistics: the log transformation. Advances in Physiology Education 42:343-347.
Davidson, R., and J. G. MacKinnon. 2006. Bootstrap methods in econometrics. Page 1097 *in* H. Hassani, T. C. Mills, and K. Patterson, editors. Palgrave Handbook of Econometrics Volume 1: Econometric Theory. Palgrave Macmillan UK, London UK.
Doney, S. C. 2010. The growing human footprint on coastal and open-ocean biogeochemistry. Science 328:1512-1516.
Draper, N. R., and H. Smith. 1998. Applied regression analysis. John Wiley & Sons, New York, New York, USA.
Efron, B. 1982. The jackknife, the bootstrap and other resampling plans. J. W. Arrowsmith Ltd., Bristol, England.
Efron, B., and G. Gong. 1983. A Leisurely Look at the Bootstrap, the Jackknife, and Cross-Validation. The American Statistician 37:36-48.
Ekwaru, J. P., and P. J. Veugelers. 2018. The overlooked importance of constants added in log transformation of independent variables with zero values: A proposed approach for determining an optimal constant. Statistics in Biopharmaceutical Research 10:26-29.
Eshleman, K. N., R. D. Sabo, and K. M. Kline. 2013. Surface water quality is improving due to declining atmospheric N deposition. Environmental Science & Technology 47:12193-12200.
Feng, C., H. Wang, N. Lu, T. Chen, H. He, Y. Lu, and X. M. Tu. 2014. Log-transformation and its implications for data analysis. Shanghai Archives of Psychiatry 26:105-109.
Feng, C., H. Wang, N. Lu, and X. M. Tu. 2013. Log transformation: application and interpretation in biomedical research. Statistics in Medicine 32:230-239.
Flachaire, E. 2005. Bootstrapping heteroskedastic regression models: wild bootstrap vs. pairs bootstrap. Computational Statistics & Data Analysis 49:361-376.
Foley, J. A., N. Ramankutty, K. A. Brauman, E. S. Cassidy, J. S. Gerber, M. Johnston, N. D. Mueller, C. O'Connell, D. K. Ray, P. C. West, C. Balzer, E. M. Bennett, S. R. Carpenter, J. Hill, C. Monfreda, S. Polasky, J. Rockstrom, J. Sheehan, S. Siebert, D. Tilman, and D. P. Zaks. 2011. Solutions for a cultivated planet. Nature 478:337-342.
Follmi, K. B. 1996. The phosphorus cycle, phosphogenesis and marine phosphate-rich deposits. Earth Science Reviews 40:55-124.
Froelich, P. N. 1988. Kinetic control of dissolved phosphate in natural rivers and estuaries: A primer on the phosphate buffer mechanism1. Limnology and Oceanography 33:649-668.
Galloway, J. N., and E. B. Cowling. 2002. Reactive nitrogen and the world: 200 years of change. Ambio 31:64-71.
Golden, H. E., E. W. Boyer, M. G. Brown, S. T. Purucker, and R. H. Germain. 2009. Spatial variability of nitrate concentrations under diverse conditions in tributaries to a lake watershed. Journal of the American Water Resources Association 45:945-962.
Grayson, R. B., C. J. Gippel, B. L. Finlayson, and B. T. Hart. 1997. Catchment-wide impacts on water quality: the use of 'snapshot' sampling during stable flow. Journal of Hydrology 199:121-134.
Greenacre, M. 2016. Data reporting and visualization in ecology. Polar Biology 39:2189-2205.
Harmel, D., S. Potter, P. Casebolt, K. Reckhow, C. Green, and R. Haney. 2006a. Compilation of measured nutrient load data for agricultural land uses in the United States. Journal of the American Water Resources Association 42:1163-1178.
Harmel, R. D., R. J. Cooper, R. M. Slade, R. L. Haney, and J. G. Arnold. 2006b. Cumulative uncertainty in measured streamflow and water quality data for small watersheds. Transactions of the ASABE 49:689-701.

Harmel, R. D., and K. W. King. 2005. Uncertainty in measured sediment and nutrient flux in runoff from small agricultural watersheds. Transactions of the ASAE 48:1713-1721.

Harmel, R. D., K. W. King, B. E. Haggard, D. G. Wren, and J. M. Sheridan. 2006c. Practical guidance for discharge and water quality data collection on small watersheds. Transactions of the ASABE 49:937-948.

Helsel, D. R., and R. M. Hirsch. 2002. Statistical methods in water resources. Report 04-A3, U. S. Geological Survey, Reston, Virginia USA.

Hirsch, R. M., D. L. Moyer, and S. A. Archfield. 2010. Weighted Regressions on Time, Discharge, and Season (WRTDS), with an Application to Chesapeake Bay River Inputs. Journal of the American Water Resources Association 46:857-880.

Homer, C., C. Huang, L. Yang, B. Wylie, and M. Coan. 2004. Development of a 2001 national land-cover database for the United States. Photogrammetric Engineering and Remote Sensing 70:829-840.

Jolliff, J. K., J. C. Kindle, I. Shulman, B. Penta, M. A. M. Friedrichs, R. Helber, and R. A. Arnone. 2009. Summary diagrams for coupled hydrodynamic-ecosystem model skill assessment. Journal of Marine Systems 76:64-82.

Jordan, T. E., D. L. Correll, W. T. Peterjohn, and D. E. Weller. 1986. Nutrient flux in a landscape: the Rhode River watershed and receiving waters. Pages 57-76 *in* D. L. Correll, editor. Watershed research perspectives. Smithsonian Institution Press, Washington, DC USA.

Jordan, T. E., D. L. Correll, and D. E. Weller. 1997a. Effects of Agriculture on Discharges of Nutrients from Coastal Plain Watersheds of Chesapeake Bay. Journal of Environmental Quality 26:836-848.

Jordan, T. E., D. L. Correll, and D. E. Weller. 1997b. Nonpoint Source Discharges of Nutrients from Piedmont Watersheds of Chesapeake Bay. Journal of the American Water Resources Association 33:631-645.

Jordan, T. E., D. L. Correll, and D. E. Weller. 1997c. Relating nutrient discharges from watersheds to land use and streamflow variability. Water Resources Research 33:2579-2590.

Jordan, T. E., D. L. Correll, and D. E. Weller. 2000. Mattawoman creek watershed nutrient and sediment dynamics: Final contract report to Charles County, Maryland. Smithsonian Environmental Research Center, Edgewater, Maryland USA.

Jordan, T. E., and D. E. Weller. 1996. Human contributions to terrestrial nitrogen flux. Bioscience 46:655-664.

Jordan, T. E., D. E. Weller, and D. L. Correll. 2003. Sources of nutrient inputs to the Patuxent River estuary. Estuaries 26:226-243.

Kaufmann, P. R., A. T. Herlihy, M. E. Mitch, J. J. Messer, and W. S. Overton. 1991. Stream chemistry in the eastern United States: 1. Synoptic survey design, acid-base status, and regional patterns. Water Resources Research 27:611-627.

Kirchner, J. W., and C. Neal. 2013. Universal fractal scaling in stream chemistry and its implications for solute transport and water quality trend detection. Proceedings of the National Academy of Sciences, USA 110:12213-12218.

Langland, M. J., P. L. Lietman, and S. Hoffman. 1995. Synthesis of nutrient and sediment data for watersheds within the Chesapeake Bay drainage basin. U.S. Geological Survey, Lemoyne, Pennsylvania USA.

Liu, Z.-J., D. E. Weller, D. L. Correll, and T. E. Jordan. 2000. Effects of Land Cover and Geology on Stream Chemistry in Watersheds of Chesapeake Bay. Journal of the American Water Resources Association 36:1349-1365.

Mammen, E. 1993. Bootstrap and wild bootstrap for high-dimensional linear-models. Annals of Statistics 21:255-285.

McCarty, J. A., and B. E. Haggard. 2016. Can we manage nonpoint-source pollution using nutrient concentrations during seasonal baseflow? Agricultural & Environmental Letters 1:1-5.

McDonnell, J. J., and K. Beven. 2014. Debates—The future of hydrological sciences: A (common) path forward? A call to action aimed at understanding velocities, celerities and residence time distributions of the headwater hydrograph. Water Resources Research 50:5342-5350.

Messer, J. J., C. W. Ariss, J. R. Baker, S. K. Drousé, K. N. Eshleman, A. J. Kinney, W. S. Overton, M. J. Sale, and R. D. Schonbrod. 1988. Stream chemistry in the southern Blue Ridge: Feasibility of a regional synoptic sampling approach. Journal of the American Water Resources Association 24:821-829.

Moatar, F., and M. Meybeck. 2005. Compared performances of different algorithms for estimating annual nutrient loads discharged by the eutrophic River Loire. Hydrological Processes 19:429-444.

Moyer, D. L., R. M. Hirsch, and K. E. Hyer. 2012. Comparison of two regression-based approaches for determining nutrient and sediment fluxes and trends in the Chesapeake Bay Watershed. Scientific Investigations Report 2012–5244. U. S. Geological Survey, Reston, Virginia USA.

Moyer, D. L., M. J. Langland, J. D. Blomquist, and G. Yang. 2017. Nitrogen, phosphorus, and suspended-sediment loads and trends measured at the Chesapeake Bay Nontidal Network stations: Water years 1985-2016. U.S. Geological Survey data release https://doi.org/10.5066/F7RR1X68

Nixon, S. W. 1995. Coastal marine eutrophication: a definition, social causes, and future consequences. Ophelia 41:199-219.

NRCS. 2003. National water quality handbook. Natural Resources Conservation Service, U.S. Department of Agriculture, Washington, DC USA.

Preston, S. D., R. B. Alexander, G. E. Schwarz, and C. G. Crawford. 2011. Factors affecting stream nutrient loads: a synthesis of regional SPARROW model results for the continental United States. Journal of the American Water Resources Association 47:891-915.

R Core Team. 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

Rendevski, V., B. Aleksovski, M. Kolevska, D. Stojanov, K. Dimitrovski, A. M. Rendevska, V. Aleksovski, A. Petlickovski, D. Trajkov, and K. Stojanoski. 2016. Effects of data transformation on multivariate analyses in intracerebral hemorrhage. Macedonian Pharmaceutical Bulletin 62:37-42.

Roodman, D., M. Ø. Nielsen, J. G. MacKinnon, and M. D. Webb. 2019. Fast and wild: Bootstrap inference in Stata using boottest. The Stata Journal 19:4-60.

Rozemeijer, J., Y. P. van der Velde, H. de Jonge, F. van Geer, H. P. Broers, and M. Bierkens. 2010. Application and evaluation of a new passive sampler for measuring average solute concentrations in a catchment scale water quality monitoring study. Environmental Science & Technology 44:1353-1359.

Ryberg, K. R., A. J. Sekellick, and J. A. Falcone. 2017. Ancillary data related to nutrients in the Chesapeake Bay—Data supporting structural equation modeling of nutrient loads. U.S. Geological Survey data release https://doi.org/10.5066/F7NG4NQ4

Schleppi, P., P. A. Waldner, and B. Fritschi. 2006a. Accuracy and precision of different sampling strategies and flux integration methods for runoff water: comparisons based on measurements of the electrical conductivity. Hydrological Processes 20:395-410.

Schleppi, P., P. A. Waldner, and M. Stahli. 2006b. Errors of flux integration methods for solutes in grab samples of runoff water, as compared to flow-proportional sampling. Journal of Hydrology 319:266-281.

Schlesinger, W. H. 2009. On the fate of anthropogenic nitrogen. Proceedings of the National Academy of Sciences, USA 106:203-208.

Scott, D., J. Harvey, R. Alexander, and G. Schwarz. 2007. Dominance of organic nitrogen from headwater streams to large rivers across the conterminous United States. Global Biogeochemical Cycles 21.

Seitzinger, S. P., R. V. Styles, E. W. Boyer, R. B. Alexander, G. Billen, R. W. Howarth, B. Mayer, and N. van Breeman. 2002. Nitrogen retention in rivers: model development and application to watersheds in the Eastern U.S.A. Biogeochemistry 57/58:199-237.

Snedecor, G. W., and W. G. Cochran. 1989. Statistical methods, 8th Edition. Iowa State University Press, Ames, Iowa USA.

Sobota, D. J., J. E. Compton, M. L. McCrackin, and S. Singh. 2015. Cost of reactive nitrogen release from human activities to the environment in the United States. Environmental Research Letters 10:025006.

Stewart, B. A., and R. Lal. 2017. The nitrogen dilemma: Food or the environment. Journal of Soil and Water Conservation 72:124A-128A.

Stranko, S., D. Boward, J. Kilian, A. Becker, M. J. Ashton, M. Southerland, B. Franks, W. Harbold, and J. Cessna. 2017. Maryland biological stream survey: Round four field sampling manual. Maryland Department of Natural Resources, Annapolis, Maryland USA.

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

Venables, W. N., and B. D. Ripley. 2002. Modern applied statistics with S. Fourth edition. Springer, New York, New York USA.

Weller, D. E., and M. E. Baker. 2014. Cropland Riparian Buffers throughout Chesapeake Bay Watershed: Spatial Patterns and Effects on Nitrate Loads Delivered to Streams. Journal of the American Water Resources Association 50:696-712.

Weller, D. E., M. E. Baker, and T. E. Jordan. 2011. Effects of riparian buffers on nitrate concentrations in watershed discharges: new models and management implications. Ecological Applications 21:1679-1695.

Weller, D. E., T. E. Jordan, D. L. Correll, and Z. J. Liu. 2003. Effects of land-use change on nutrient discharges from the Patuxent River watershed. Estuaries 26:244-266.

Weller, D. E., T. E. Jordan, K. G. Sellner, K. Foreman, K. Shenk, P. Tango, S. W. Phillips, and M. Dubin. 2010. Small watershed monitoring designs. Chesapeake Bay Program Scientific and Technical Advisory Program, Edgewater, Maryland USA.

Wickham, H. 2016. ggplot2: elegant graphics for data analysis. Springer, New York, New York USA.

Wolock, D. M., J. Fan, and G. B. Lawrence. 1997. Effects of basin size on low-flow stream chemistry and subsurface contact time in the Neversink River Watershed, New York. Hydrological Processes 11:1273-1286.

Wu, C. F. J. 1986. Jackknife, bootstrap and other resampling methods in regression analysis. The Annals of Statistics 14:1261-1295.

Yang, L. M., S. M. Jin, P. Danielson, C. Homer, L. Gass, S. M. Bender, A. Case, C. Costello, J. Dewitz, J. Fry, M. Funk, B. Granneman, G. C. Liknes, M. Rigge, and G. Xian. 2018. A new generation of the United States National Land Cover Database: Requirements, research priorities, design, and implementation strategies. ISPRS Journal of Photogrammetry and Remote Sensing 146:108-123.

Zambrano-Bigiarini, M. 2020. hydroGOF: Goodness-of-fit functions for comparison of simulated and observed hydrological time series, R package version 0.4-0. https://doi.org/10.5281/zenodo.839854

Zhang, Q., J. D. Blomquist, D. L. Moyer, and J. G. Chanat. 2019. Estimation bias in water-quality constituent concentrations and fluxes: A synthesis for Chesapeake Bay rivers and streams. Frontiers in Ecology and Evolution 7.

Supporting Information

Additional supporting information may be found online at:

Data Availability

The SERC data are available in the Supporting Information (see Data S1 and Metadata S1), and the CBNTN data are available in the references cited.

Table 1. Measures of skill for three methods of estimating multiyear average nitrate or total nitrogen concentration from spot concentration measurements applied to the SERC and CBNTN data sets.

Relationship	Method	Intercept	Slope	Mean Error	% Bias	RMSE	ubRMSE	R^2(%)
SERC								
NO3 ~ sNO3	Direct			0.213	10.1	0.455	0.403	98.3
	Simple linear	-0.142	0.970	0.000	0.0	0.392	0.392	98.3
	Bootstrap	-0.109	0.952	-0.008	-0.4	0.353	0.353	98.7
NO3 ~ fsNO3	Direct			0.058	2.7	0.553	0.550	96.6
	Simple linear	-0.065	1.003	0.000	0.0	0.550	0.550	96.6
	Bootstrap	-0.041	0.993	0.002	0.1	0.503	0.503	97.2
TN ~ sDTN	Direct			-0.223	-6.3	0.506	0.454	98.2
	Simple linear	0.170	1.019	0.000	0.0	0.450	0.450	98.2
	Bootstrap	0.174	1.011	-0.018	-0.5	0.403	0.402	98.6
TN ~ fsDTN	Direct			-0.354	-10.0	0.652	0.548	97.4
	Simple linear	0.337	1.006	0.000	0.0	0.547	0.547	97.4
	Bootstrap	0.358	0.988	-0.029	-0.8	0.487	0.486	98.0
TN ~ sNO3	Direct			-0.565	-19.6	0.742	0.481	97.8
	Simple linear	0.503	1.027	0.000	0.0	0.474	0.474	97.8
	Bootstrap	0.511	1.018	-0.014	-0.5	0.433	0.433	98.2
TN ~ fsNO3	Direct			-0.720	-24.9	0.966	0.644	96.2
	Simple linear	0.584	1.063	0.000	0.0	0.617	0.617	96.2
	Bootstrap	0.587	1.057	-0.008	-0.3	0.578	0.578	96.7
CBNTN								
NO3 ~ fsNO3	Direct			0.084	6.1	0.468	0.461	94.1
	Simple linear	0.144	0.843	0.000	0.0	0.370	0.370	94.1
	Bootstrap	0.143	0.843	-0.002	-0.2	0.356	0.356	94.5
TN ~ fsTN	Direct			0.141	8.0	0.533	0.514	92.1
	Simple linear	0.124	0.861	0.000	0.0	0.450	0.450	92.1
	Bootstrap	0.133	0.855	-0.001	-0.1	0.440	0.440	92.5
TN ~ fsNO3	Direct			-0.311	-17.6	0.572	0.481	92.7
	Simple linear	0.481	0.883	0.000	0.0	0.435	0.435	92.7
	Bootstrap	0.482	0.880	-0.003	-0.2	0.420	0.420	93.2

Table 2. Analyses relating multiyear average concentration to spot concentration for additional water quality constituents sampled by SERC and CBNTN.

	Linear Regression			% of Total Variation	
Constituent	Slope	R^2 (%)	R_s^2 (%)	Station	Residual
SERC					
NO3	0.97	98.3	97.1	92.9	7.1
TN	1.02	98.2	89.3	86.6	13.4
SI	0.78	96.9	86.8	84.9	15.1
NH4	1.33	51.8	78.0	14.6	85.4
TKN	1.77	44.1	44.4	10.0	90.0
TP	4.47	35.1	62.9	8.3	91.7
TOC	0.95	17.6	33.6	5.6	94.4
PO4	2.28	14.4	41.5	8.7	91.3
CBNTN					
NO3	0.84	94.1	93.1	93.9	6.1
TN	0.86	92.1	82.2	94.0	6.0
PO4 †	1.06	26.4	37.9	56.2	43.8
TP	0.05	0.8	16.3	56.4	43.6
TSS	0.03	1.8	0.4	14.2	85.8

Slope and R^2 from linear regressions of flow-weighted average concentration in weekly composite samples vs. average spot sample concentration (SERC), or of flow-weighted average concentration from WRTDS synthesis vs. the first spring spot sample (CBNTN). R_s^2 is the squared Spearman rank-order correlations (R cor function, R Core Team 2017). The Station column is the percentage of total variation among weeks and watersheds (SERC) or among months and watersheds (CBNTN) attributable to differences among watersheds. The Residual column is the remainder due to temporal variation within watersheds and to error. For each data set, constituents listed above the dashed line are transported in dissolved form while constituents below the dashed line are primarily transported on particles.

† We placed CBNTN PO4 below the dotted line even though the CBNTN measures dissolved PO4 on filtered samples. PO4 is transported in streams and rivers mostly on particles (Follmi

1996, Jordan et al. 1997b, a), and dissolved PO4 exchanges with that particulate PO4 (Froelich 1988). Therefore, the factors that drive high temporal variability in particulate PO4 concentration can also affect dissolved PO4 measurements.

Figure Legends

Fig. 1. Boundaries for 59 SERC (blue outlines) and 85 CBNTN (red outlines) watersheds within the Chesapeake Bay drainage (outer black boundary). Shaded areas within that boundary are four major physiographic provinces (Langland et al. 1995). The underlying base map of the U.S. mid-Atlantic region (ESRI 2019) shows the coastline and boundaries of six states (NY, PA, MD, DE, WV, and VA) intersected by the Chesapeake watershed. The diagonal arrow points to the Rhode River watershed cluster.

Fig. 2. SERC stream sampling scheme for one station. Top, stream discharge measured continuously (solid line), with the timing of spot samples marked by ticks on the horizontal axis. Bottom, nitrate concentration in weekly flow-weighted, composite water samples (solid line), seasonal spot samples (blue dots), and the first spring spot sample (red square). Horizontal lines mark the average nitrate concentrations for the weekly composites (black solid line) and the seasonal spots (blue dashed line), as well as the concentration in the first spring spot (red, dot-dash line).

Fig. 3. Geographic characteristics of the study watersheds. (a) distributions of watershed area (SERC, left bars, blue; CBNTN, right bars, red). Note log10 scale on horizontal axis. (b) land cover proportions: SERC (blue circles and square) and CBNTN (red triangles). The aggregated categories shown on the three axes together cover more than 95% of the land in every watershed. The blue square is SERC station 522, which had the highest TN and NO3 concentrations across both data sets.

Fig. 4. Frequency distributions of flow-weighted measurements of nitrogen concentrations. (a,b) nitrate concentration. (c,d) total nitrogen concentration. (a,c; blue) averages from automated composite sampling of 59 SERC watersheds. (b,d; red) averages from WRTDS synthesis for 85 CBNTN watersheds.

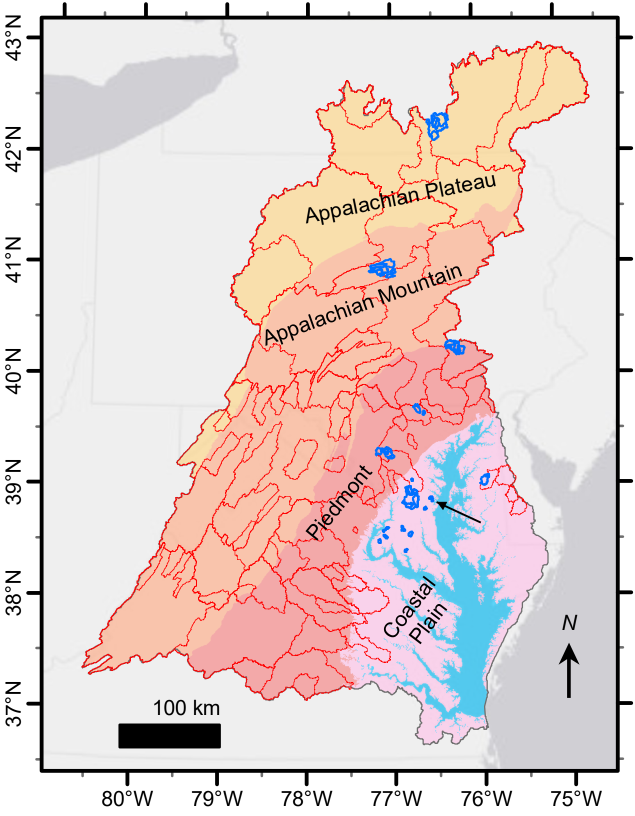
Fig. 5. Average flow-weighted nitrate concentration versus average spot nitrate concentration for 59 SERC watersheds and variability in those measurements. Light blue lines mark the ranges for both variables. For composite measurements only, dark gray lines and whiskers mark one standard deviation while black lines and whiskers mark one standard error of the mean. The dashed line is the 1:1 line. The uppermost point is watershed 522.

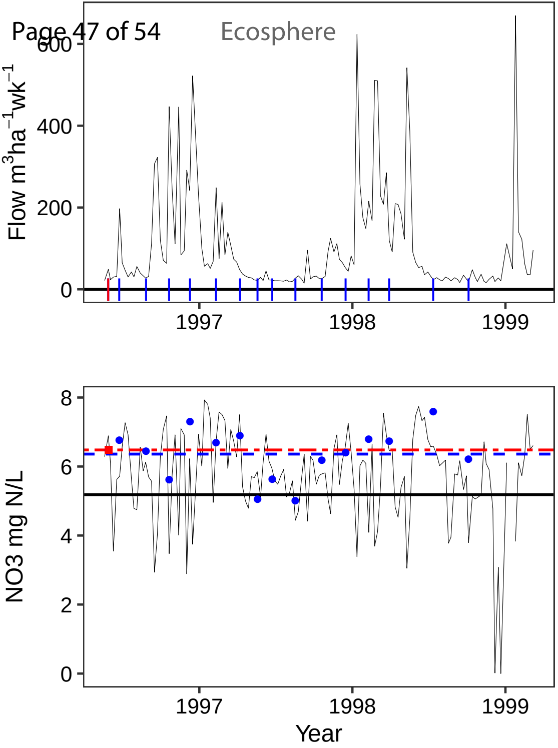
Fig. 6. Regression models for predicting average flow-weighted concentration from spot measurements: (a-c) flow-weighted average NO3 vs. spot measurements for SERC data. (a) linear regression of flow-weighted average NO3 vs. average spot sNO3. (b) bootstrapped linear regression of the same variables. (c) bootstrapped regression of NO3 vs. first spot fsNO3. (d-f) bootstrapped regressions of flow-weighted average total nitrogen concentration (TN) vs. spot measurements for SERC data: (d) average spot dissolved TN, (e) average spot NO3, (f) first spot NO3. (g-i) bootstrapped regressions of average flow-weighted concentration from WRTDS synthesis vs. spot measurements for CBNTN data: (g) WRTDS average NO3 vs. first spot fsNO3, (h) WRTDS TN vs. first spot fsTN, (i) WRTDS TN vs.fsNO3. Note differences in axis scaling between SERC (a-f) and CBNTN (g-i) data. All panels show the 1:1 line (long-short dashed), the regression line (solid), the 95% confidence interval (dark gray shading), and the 95% prediction interval (light gray shading). For bootstrapped models, the outer dashed lines are loess-smoothed representations of bootstrap prediction intervals.

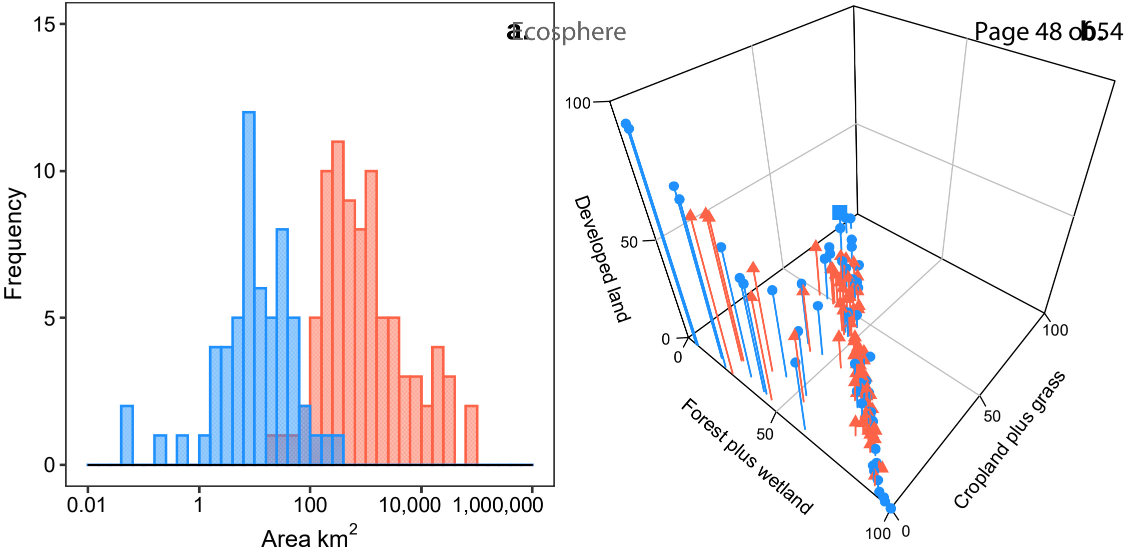
Fig. 7. Distributions of the regression slope (a) and predictions (b,c) for the bootstrapped model relating average composite-sampled nitrate (NO3) to average spot-sampled nitrate (sNO3).

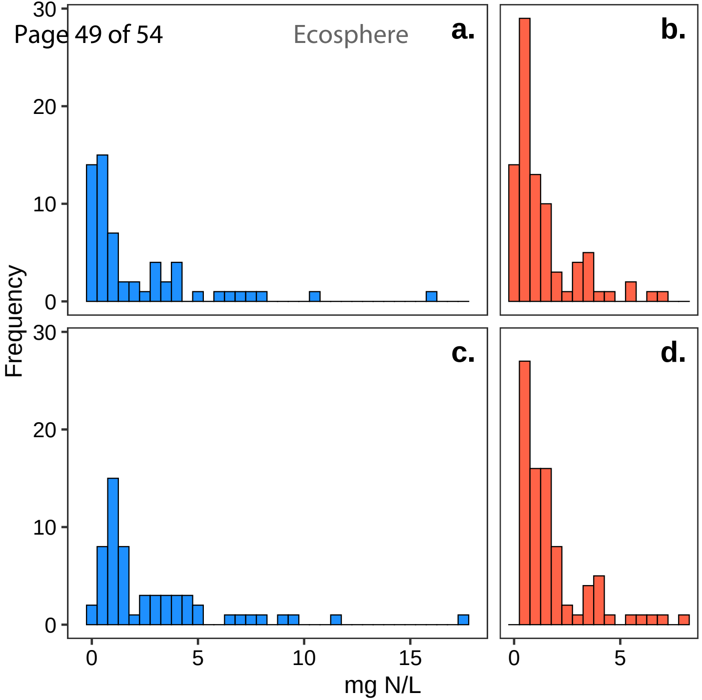
Fig. 8. Relationships of flow-weighted average nitrate (NO3) vs. first spot nitrate (fsNO3) for SERC composite samples (blue circles) and CBNTN WRTDS estimates (red triangles) watersheds. Solid lines are the bootstrapped regression models (Table 1), and the dashed line is the 1:1 line.

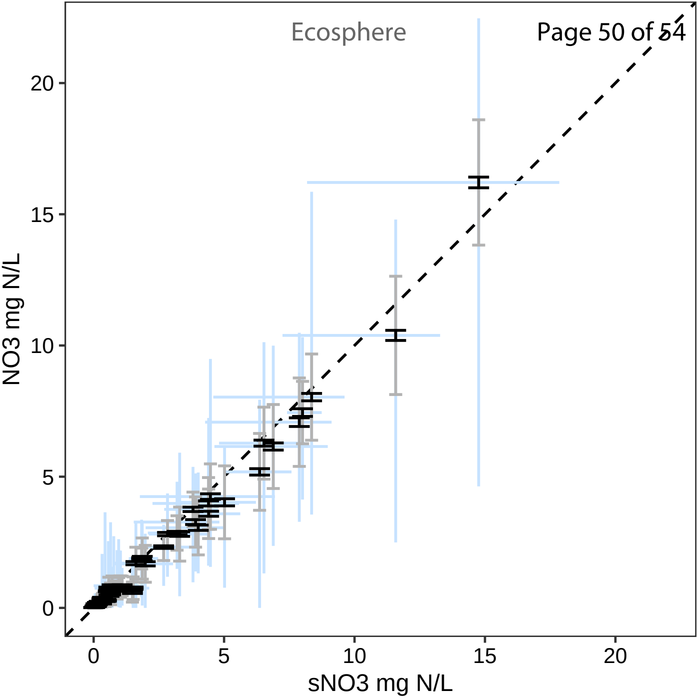
Fig. 9. The fraction of total nitrogen (TN) as nitrate (NO3) versus TN concentration. Left, flow-weighted averages from composite samples at 59 SERC watersheds. Right, flow-weighted average WRTDS estimates for 85 CBNTN watersheds. The black line is a smoothed curve (R loess function, R Core Team 2017)) through the NO3 data (black points). The (red) shaded area below that line is the smoothed fraction of NO3 at any level of TN. The (blue) shaded area above that line is the fraction of other nitrogen components (essentially ammonium plus organic nitrogen). Above the dotted line, more than half of the TN is NO3.

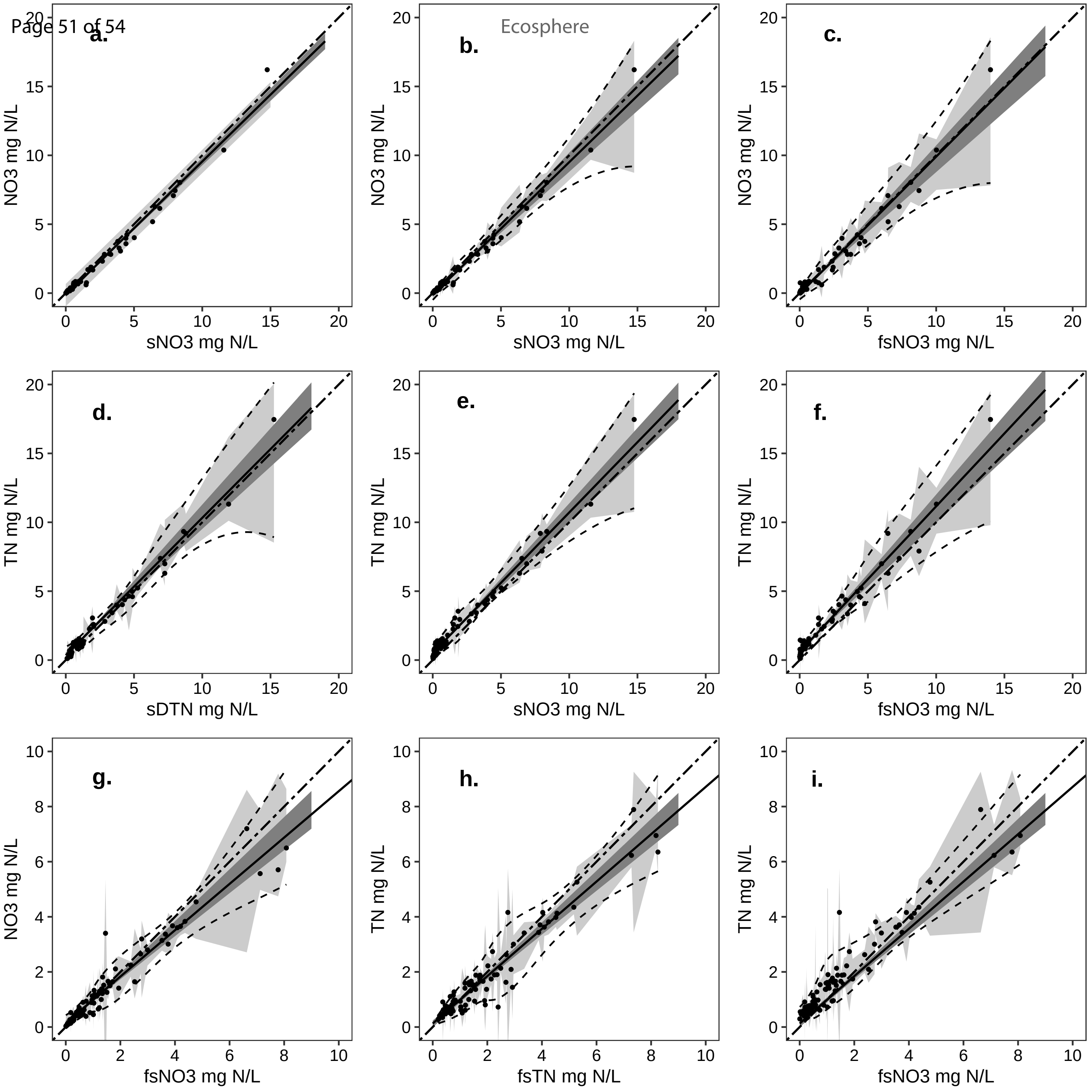

43°N
42°N
41°N
40°N
39°N
38°N
37°N
80°W
79°W
78°W
77°W
76°W
75°W
Appalachian Plateau
Appalachian Mountain
Piedmont
Coastal Plain
100 km
N


Flow m^3 ha^-1 wk^-1
600
400
200
0
1997
1998
1999
NO3 mg N/L
8
6
4
2
0
1997
1998
1999
Year

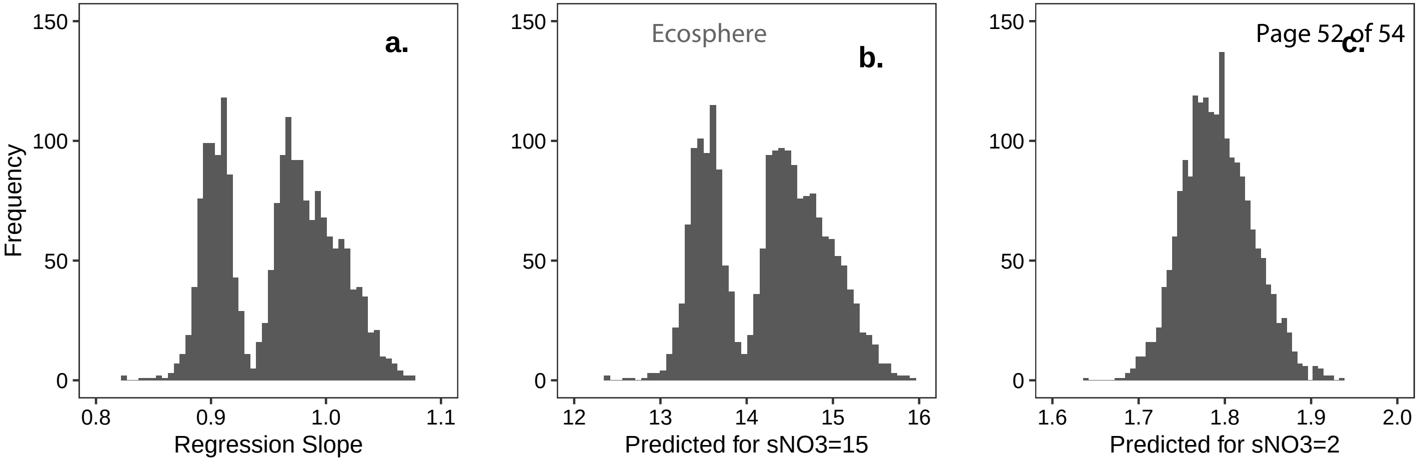

a.
Frequency
15
10
5
0
0.01
1
100
10,000
1,000,000
Area km2
b.
Developed land
100
50
0
Forest plus wetland
0
50
100
Cropland plus grass
0
50
100

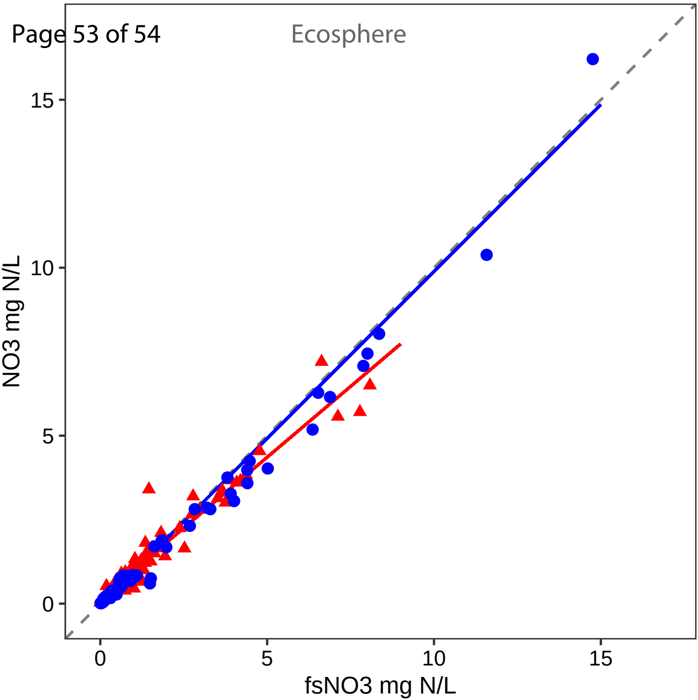

a.
b.
c.
d.
Frequency
0
10
20
30
0
5
10
15
mg N/L


NO3 mg N/L
sNO3 mg N/L
0
5
10
15
20

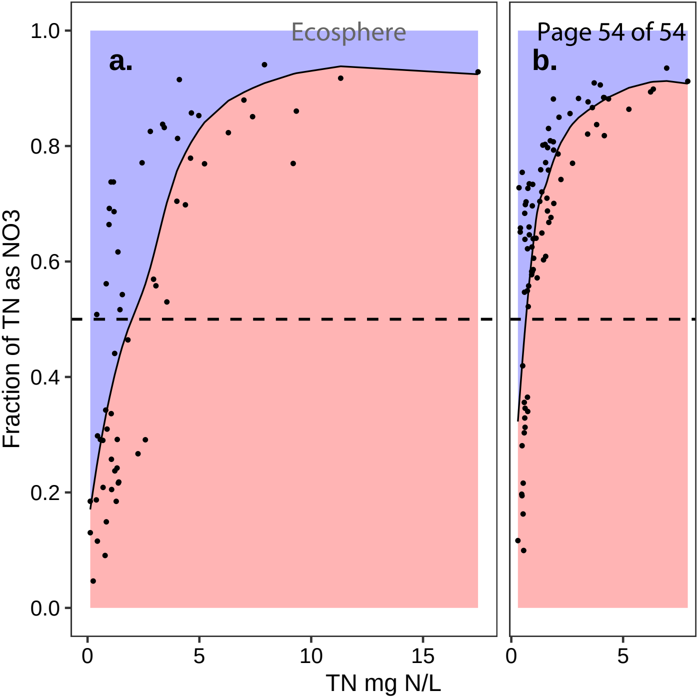


a.
b.
c.
d.
e.
f.
g.
h.
i.
NO3 mg N/L
TN mg N/L
sNO3 mg N/L
fsNO3 mg N/L
sDTN mg N/L
fsTN mg N/L


a.
b.
c.
Frequency
0
50
100
150
0.8
0.9
1.0
1.1
Regression Slope
12
13
14
15
16
Predicted for sNO3=15
1.6
1.7
1.8
1.9
2.0
Predicted for sNO3=2



15
10
5
0
NO3 mg N/L
0
5
10
15
fsNO3 mg N/L


a.
b.
Fraction of TN as NO3
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
0
5
TN mg N/L