esa

ECOSPHERE

SPECIAL FEATURE: UNCERTAINTY ANALYSIS

Approaches to stream solute load estimation for solutes with varying dynamics from five diverse small watersheds

Brent T. Aulenbach,[1,]† Douglas A. Burns,[2] James B. Shanley,[3] Ruth D. Yanai,[4] Kikang Bae,[4] Adam D. Wild,[4] Yang Yang,[4] and Dong Yi[4]

[1]U.S. Geological Survey, South Atlantic Water Science Center, Norcross, Georgia 30093 USA
[2]U.S. Geological Survey, New York Water Science Center, Troy, New York 12180 USA
[3]U.S. Geological Survey, New England Water Science Center, Montpelier, Vermont 05601 USA
[4]State University of New York, College of Environmental Science and Forestry, Syracuse, New York 13210 USA

Citation: Aulenbach, B. T., D. A. Burns, J. B. Shanley, R. D. Yanai, K. Bae, A. D. Wild, Y. Yang, and D. Yi. 2016. Approaches to stream solute load estimation for solutes with varying dynamics from five diverse small watersheds. Ecosphere 7(6):e01298. 10.1002/ecs2.1298

Abstract. Estimating streamwater solute loads is a central objective of many water-quality monitoring and research studies, as loads are used to compare with atmospheric inputs, to infer biogeochemical processes, and to assess whether water quality is improving or degrading. In this study, we evaluate loads and associated errors to determine the best load estimation technique among three methods (a period-weighted approach, the regression-model method, and the composite method) based on a solute's concentration dynamics and sampling frequency. We evaluated a broad range of varying concentration dynamics with stream flow and season using four dissolved solutes (sulfate, silica, nitrate, and dissolved organic carbon) at five diverse small watersheds (Sleepers River Research Watershed, VT; Hubbard Brook Experimental Forest, NH; Biscuit Brook Watershed, NY; Panola Mountain Research Watershed, GA; and Río Mameyes Watershed, PR) with fairly high-frequency sampling during a 10- to 11-yr period. Data sets with three different sampling frequencies were derived from the full data set at each site (weekly plus storm/snowmelt events, weekly, and monthly) and errors in loads were assessed for the study period, annually, and monthly. For solutes that had a moderate to strong concentration–discharge relation, the composite method performed best, unless the autocorrelation of the model residuals was <0.2, in which case the regression-model method was most appropriate. For solutes that had a nonexistent or weak concentration–discharge relation (model R^2 < about 0.3), the period-weighted approach was most appropriate. The lowest errors in loads were achieved for solutes with the strongest concentration–discharge relations. Sample and regression model diagnostics could be used to approximate overall accuracies and annual precisions. For the period-weighed approach, errors were lower when the variance in concentrations was lower, the degree of autocorrelation in the concentrations was higher, and sampling frequency was higher. The period-weighted approach was most sensitive to sampling frequency. For the regression-model and composite methods, errors were lower when the variance in model residuals was lower. For the composite method, errors were lower when the autocorrelation in the residuals was higher. Guidelines to determine the best load estimation method based on solute concentration–discharge dynamics and diagnostics are presented, and should be applicable to other studies.

Key words: load methodology; sample design; Special Feature: Uncertainty Analysis; streamwater fluxes; streamwater loads; water quality.

Received 11 May 2015; revised 20 November 2015; accepted 5 December 2015. Corresponding Editor: J. Campbell.

† **E-mail:** btaulenb@usgs.gov

Introduction

Streamwater load represents the mass of solutes or sediment that passes a given location on a stream during a set period. Load estimation is a central objective of many water-quality monitoring programs and research studies. In watershed studies, loads serve as an integrated measure of inputs and biogeochemical processes within a watershed that affect water quality (e.g., Likens et al. 1967, Semkin et al. 1994). With increased emphasis on watershed-based strategies, such as best management practices (BMPs) for the control of nonpoint-source pollutants, reliable measures of loads are needed to assess whether water quality is improving or degrading. In the United States, stream reaches that do not meet water-quality standards are subject to waste-load allocation schemes based on the total maximum daily loads (TMDL; USEPA 2000). Accurate and precise load estimates are essential to relate changes and trends in water-quality to changes in inputs of point- and nonpoint-source pollutants and to evaluate the effects of short-term (e.g., drought) and long-term climatic patterns, to assess the effectiveness of BMPs on water quality, and to quantify the seasonal variability in hydrologic and biogeochemical processes within a watershed.

The total solute load (L) is the product of solute concentration (C) and discharge (Q) integrated over time (t):

$$L = \int C(t)\,Q(t)\mathrm{d}t. \qquad (1)$$

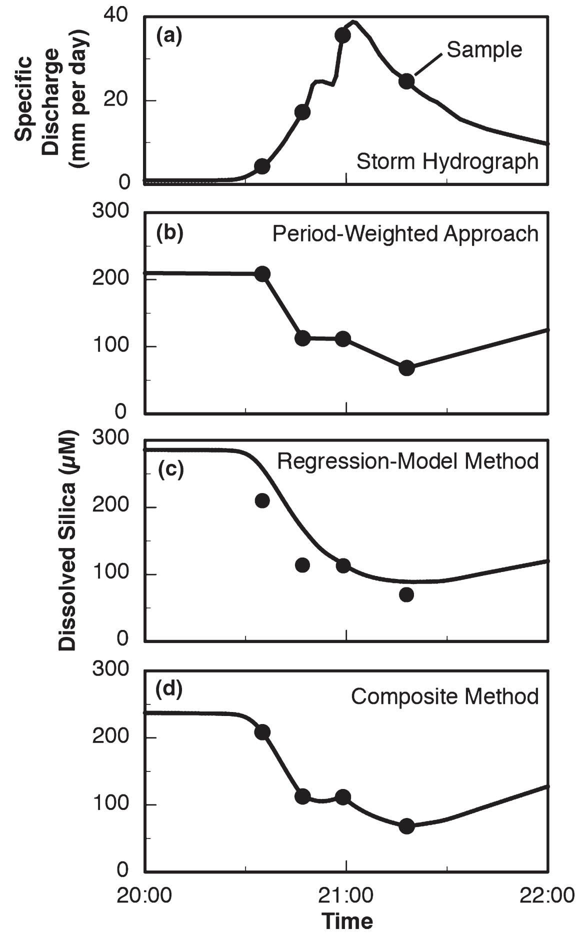

Fig. 1. Example of (a) storm hydrograph and sampling; and dissolved silica concentration functions for: (b) Period Weighted Approach; (c) Regression-Model Method, and; (d) Composite Method for a storm at Panola Mountain.

To evaluate the integral in Eq. 1 requires a continuous record of concentration and discharge. While discharge can readily be measured in a continuous manner, most solutes require the analysis of discrete samples. Various techniques have been developed to estimate loads from discrete concentration observations. For studies that require inter-annual and seasonal load estimates, period-weighted approaches and regression-model (or rating-curve) methods are most appropriate, as they allow estimation of concentrations and loads continuously through time.

In a period-weighted approach, measured concentrations are assumed to represent the concentration in the period around the sample collection, either as a step function of the sample concentrations (e.g., Likens et al. 1977) or as a piecewise linear function between measured concentrations (e.g., Larson et al. 1995, Fig. 1b), which is then used in equation 1 to estimate load. The load estimate is sensitive to sampling frequency and design (Richards and Holloway 1987), particularly if concentrations vary strongly with stream flow. In this situation, baseflow concentrations can be misapplied to adjacent, higher stormflows while storm concentrations can be misapplied to adjacent, lower baseflows. These misrepresentations do not equally counteract each other, resulting in load estimates that are typically more biased toward baseflow concentrations, especially as sampling frequency is decreased. Various sampling designs, such as selection at list time (SALT; Thomas 1985) and time-stratified sampling (Thomas and Lewis 1993) and load estimation methods such as that of Worrall et al. (2013) have been developed to reduce this type of error.

In a regression-model method, $C(t)$ is estimated using a regression model relating concentration to continuous variables such as discharge and day of year (e.g., Johnson 1979, Crawford 1991, Cohn et al. 1992, Fig. 1c), enabling a calculation using Eq. 1. The model predicts the average concentration response for the conditions present, and therefore does not attempt to match the observed concentrations at any given time. Regression-model method estimation errors can be reduced by increasing the sampling frequency (Horowitz 2003, Verma et al. 2012) and by including either event sampling (Preston et al. 1989, Robertson and Roerish 1999, Robertson 2003) or targeted high-flow sampling (Horowitz 2003) to ensure that the concentration–discharge relationship is adequately defined (Smith and Croke 2005).

The composite method is a hybrid load estimation approach that incorporates aspects of the period-weighted approach and the regression-model method (Huntington et al. 1994, Aulenbach and Hooper 2006), and is equivalent to the Q-proportionate method employed by Vanni et al. (2001). In this method, a regression model is used to predict concentrations continuously through time, but the curve is forced through the observations by adjusting the regression model concentrations by the residual concentrations (sample observed minus regression-model predicted concentrations) and interpolating this residual correction between the observed sample concentrations (Fig. 1d). This approach generally improves the precision of regression-model method load estimates at shorter reporting periods when autocorrelation is present in the residual concentrations, which is indicative of persistent temporal deviations from the predicted concentrations (Aulenbach and Hooper 2006, Aulenbach 2013).

Estimating uncertainty in load estimates for these three load methods can be complex, but is important for selecting the best method for a particular data set. For a period-weighted approach, the variance of a load estimate can be derived from a semivariogram calibrated to the data using a cross-validation technique (Shih et al. 1998). For regression-model methods, the uncertainty in the regression is easy to calculate, but this can underestimate errors if the model calibration data set is not representative of all hydrologic conditions or model residuals are autocorrelated (Aulenbach 2013). The uncertainty in the composite method cannot be directly calculated, but has been estimated using a subsampling approach using rich data sets (Aulenbach and Hooper 2006, Aulenbach 2013). Appling et al. (2015) used a delete-one jackknife approach to estimate the uncertainty of loads in all three methods, allowing them to choose the method with the smallest errors.

The appropriate load estimation method to use depends on the sampling frequency and design (e.g., fixed-interval, targeted high flow, and event sampling), watershed size, the variability in flow, and the strength and form of the concentration–discharge relation (Richards and Holloway 1987, Preston et al. 1989, 1992). Moatar and Meybeck (2007) indicated that various solutes had different and somewhat inherent load precisions. Preston et al. (1989) compared results from various load estimation methods for various combinations of fixed-interval and event sampling designs and found that the regression-model approach provides accurate and precise load estimates when the concentration–discharge relationship is strong and consistent. Studies where the concentration–discharge relationship is not strong (R^2s < 0.3) have indicated that the regression-model method is not appropriate (Quilbé et al. 2006) and that a period-weighted approach was a better method (Moatar and Meybeck 2005). Kerr et al. (2015) indicated that solutes that had higher temporal variation in concentrations had poorer precisions when using a period-weighted approach. Aulenbach (2013) indicated that for suspended sediment with a relatively strong concentration–discharge relation, when the autocorrelation of regression-model residual concentrations was >0.15, the composite method could improve upon regression-model method load estimates.

Previous studies have indicated that the best load estimation method is dependent on the solutes considered at a particular study site, along with the sampling strategy employed. In this study, we estimated loads for four solutes at five diverse, small, forested watersheds with a broad range of concentration variability (dynamics) to provide recommendations as to the most accurate and precise method to apply. The effect of sampling frequency was explored through the application to three sub-sampled data sets (weekly plus event, weekly, and monthly). Loads were estimated using three common approaches (period-weighted approach, regression-model method, and composite method) for each solute-watershed combination and for each sampling frequency. The accuracy for the study period and the precisions at annual and monthly reporting periods were determined. In addition to determining the best approach to load estimation for particular characteristic solute dynamics, solute and regression model diagnostics were explored to quantify the accuracy and precision of the three load estimation approaches.

Methods

Chemical solutes

Four dissolved chemical solutes were included in this analysis: sulfate (SO_4^{2-}), silica (Si), nitrate (NO_3^-), and dissolved organic carbon (DOC). These solutes were selected because they differ in their sources, predominant hydrologic flowpaths, and biogeochemical reactivities at these watersheds (Johnson et al. 1969, Lawrence and Driscoll 1990, Stoddard and Murdoch 1991, Shanley and Peters 1993, Huntington et al. 1994, Burns et al. 1998, Shanley et al. 2004, Peters et al. 2006, Sebestyen et al. 2009, Stallard and Murphy 2012, 2014) resulting in a wide range of solute dynamics that are expected to be representative of many solutes at other sites. The most important differentiating characteristics of these solutes are that SO_4^{2-} is atmospherically derived and can be somewhat reactive, Si is a weathering product that behaves conservatively, and NO_3^- and DOC are biogeochemically controlled and have contrasting strengths of concentration–discharge relations.

Watershed descriptions and sampling

This analysis uses data from five small (13–1780 ha), forested watersheds (Fig. 2): Sleepers River Research Watershed (watershed W-9), in northeastern Vermont; Hubbard Brook

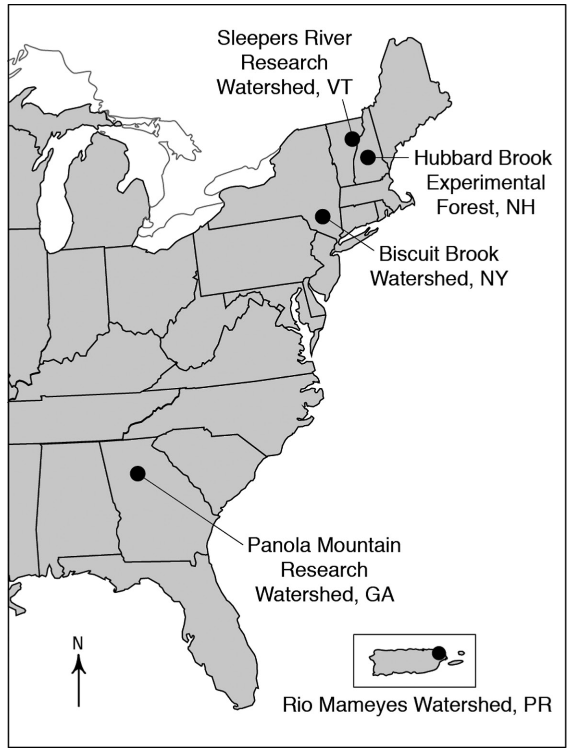

Fig. 2. Locations of the five watersheds used in this study.

Experimental Forest (watershed 6), in the White Mountains of New Hampshire; Biscuit Brook Watershed, in the Catskill Mountains of New York; Panola Mountain Research Watershed, in the Piedmont of Georgia; and the Río Mameyes Watershed, in the Luquillo Experimental Forest in the Luquillo Mountains, Puerto Rico. Watersheds were selected that had at least 10 yr of stream sampling including high-flow sampling. Additionally, the watersheds represent a variety of bedrock types and a wide range in climatic and ecological settings that provide contrasting water-quality dynamics for the four selected solutes (Table 1). The watersheds also vary hydrologically, with snowmelt being important at the three northern watersheds and stream flows at Río Mameyes being an order of magnitude greater than the other watersheds (see Fig. 3 in the *Results* section).

From each watershed, we used 10 or 11 yr of weekly and hydrologic event-based water-quality sampling along with continuous (1- to 15-min interval) stream flow measurements throughout

Table 1. Descriptions of the five watersheds used in this study.

Watershed	Agency; program†	Drainage area (ha)	Geology	Climate‡; average air temperature; average precipitation	Ecoregion§	Vegetation	Elevation range (m)
Sleepers River Research Watershed (Watershed W-9), Danville, Vermont¶	USGS; WEBB	40.5	Calcareous granulite/ quartz-mica phyllite; overlain by glacial till	Humid continental (cool summer); 4.6°C; 1320 mm (25–30% snow)	Northeastern Highlands: Northern Piedmont [581]	Rich northern hardwood forest	519–678
Hubbard Brook Experimental Forest (Watershed 6), West Thornton, New Hampshire#	USDA & NSF; LTER	13.2	Meta-morphosed sedimentary – gneiss and veins of pegmatite; overlain by glacial till	Humid continental (cool summer); 6°C; 1230 mm (25–33% snow)	Northeastern Highlands: White Mountains/ Blue Mountains [58p]	Northern hardwood forest with spruce and fir at higher elevations	549–792
Biscuit Brook Watershed, Frost Valley, Catskill Mountains, New York††	USGS	959	Sedimentary – sandstone; overlain by glacial till	Humid continental (warm summer); 6°C; 1656 mm (20–25% snow)	Northeastern Highlands: Catskill High Peaks [58y]	Northern hardwood forest with red spruce and balsam fir near divide	628–1130
Panola Mountain Research Watershed, Stockbridge, Georgia‡‡	USGS; WEBB	41	Granodiorite with interspersed pods of hornblende biotite gneiss	Humid sub-tropical; 16°C; 1200 mm (<2% snow)	Piedmont: Southern Outer Piedmont [45b]	Equal amounts of deciduous, coniferous and mixed forest; 10% outcrop	224–279
Río Mameyes Watershed, Luquillo Mountains, Puerto Rico§§	NSF & USGS; LTER & WEBB	1780	Volcanoclastic Cretaceous rocks and quartz diorite	Tropical rain forest; 23°C; 2520 mm (no snow)	Puerto Rico Province: Dry-Humid Mountains [M411A]	Mature montane wet evergreen forest	83–1050

† USGS, U.S. Geological Survey; USDA, U.S. Department of Agriculture; NSF, National Science Foundation; WEBB, USGS Water, Entergy and Biogeochemical Budgets Program; LTER, National Science Foundation Long Term Ecological Research Network.

‡ Climate from Koppen–Geiger climate classification (Peel et al. 2007).

§ Ecoregion level III and IV from Bailey 1976, McNab and Avers 1994, Griffith et al. 2001, 2009, and Bryce et al. 2010.

¶ Shanley et al. (2004).

Likens et al. (1967).

†† Stoddard and Murdoch (1991).

‡‡ Peters et al. (2000).

§§ Murphy and Stallard (2012).

the period. The amount and type of event-based sampling varied greatly among watersheds (Table 2). At Sleepers River, event samples were split between snowmelt events and storms at other times of the year. Event sampling was uneven throughout the period, with most of the sampling occurring during 5 yr, 2002–2004 and 2008–2009. Hubbard Brook had the fewest event-based samples; these were a mixture of snowmelt event, storm, and intermittent supplemental baseflow samples. Biscuit Brook event sampling was fairly consistent over the period and focused particularly on snowmelt events, but also included several storms each year that entailed a sample collected on the rising hydrograph limb, near to peak flow, and on the recession limb of each storm. Panola Mountain had the most extensive storm sampling, capturing most of the large storms that occurred during the period. Samples at Panola were not analyzed for DOC for a sufficiently long period to include in this analysis. Storms were sampled at Río Mameyes sporadically throughout the period. Storm samples were analyzed fairly frequently for Si, occasionally for SO_4^{2-} and NO_3^-, and infrequently for DOC. The amount of event sampling at the different watersheds was considered when drawing conclusions about the resulting load estimates. Sample concentrations, along with stream flow and air temperature data used for concentration regression modeling and load estimation, are documented and included in the supporting information.

Load estimation methodology

Loads for each watershed-solute combination were estimated using three schemes: (1) a period-weighted approach; (2) a regression-model method; and (3) the composite method. Each of these load estimation methods was applied using three data sets: (1) monthly; (2) weekly; and (3) complete (weekly sampling plus event sampling). The period-weighted approach was implemented by linearly interpolating the concentrations between samples. Both regression-model and composite methods require concentration regression models to be developed, and these are detailed in the appendix of the supporting information. Loads for all three methods were then estimated using computer code written in 4D database software (4D, San Jose, CA; Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government) and are included in the supporting information. The program is an update to code written in 1993 by the lead author and used in the original implementation of the composite method (Huntington et al. 1994, Aulenbach and Hooper 2006). The code was updated to handle additional concentration regression model variables utilized in this analysis.

Solute variability and diagnostics

Solute concentration attributes and model diagnostics were compiled to classify solute dynamics and to assist in assessing load estimate accuracy and precision. These variables were selected, in part, because they do not require knowledge of the true load estimate, nor require any additional analysis beyond fitting a regression model to calculate model residual concentrations.

The regression model R^2 (adjusted for the number of predictors) represents the variance in concentration explained by the regression model. The standard error (SE) of the concentrations represents the observed sample concentration variance, and the SE of the model residuals represents the remaining unexplained concentration variance after the model fit. Relative SEs were calculated by dividing by the mean of the sample concentrations, and are expressed as a percentage of the mean concentration. Relative SEs should be related to the obtainable accuracy and precision of load estimates, reflecting greater difficulty in estimating loads of solutes that are more variable. Two other useful diagnostics are the autocorrelation (also known as serial correlation) of sample concentrations and of model residuals, which indicate the degree of similarity between adjacent values in time. A period-weighted approach exploits this similarity in concentrations to estimate the concentrations between samples, while the composite method uses the structure of patterns in the model residual concentrations to adjust the model-predicted concentrations to fit the observed concentrations.

Load accuracy and precision

The accuracy and precision of load estimates could be defined by comparison to the true load, if it were known. For this study, we assume that the composite method applied to the complete data set is the "best" estimate of loads that can be made with the samples collected in the analysis. Aulenbach and Hooper (2006) showed that the composite method converged on an apparent "true" load with increased sampling based on a bootstrapping experiment of a rich data set from Panola Mountain for alkalinity and chloride. We used a similar approach herein by comparing the composite method loads using the complete data sets vs. using the complete data set while excluding every other event sample. If load estimates from these two sampling frequencies are similar, this should provide some reassurance that event sampling was sufficient and that convergence on the true load has been nearly reached.

Load estimate accuracy (systematic error or bias) and precision (random error) were calculated for all watersheds and solutes for each combination of the three load estimation methods and three sampling strategies. There are no error estimates for the composite method applied to

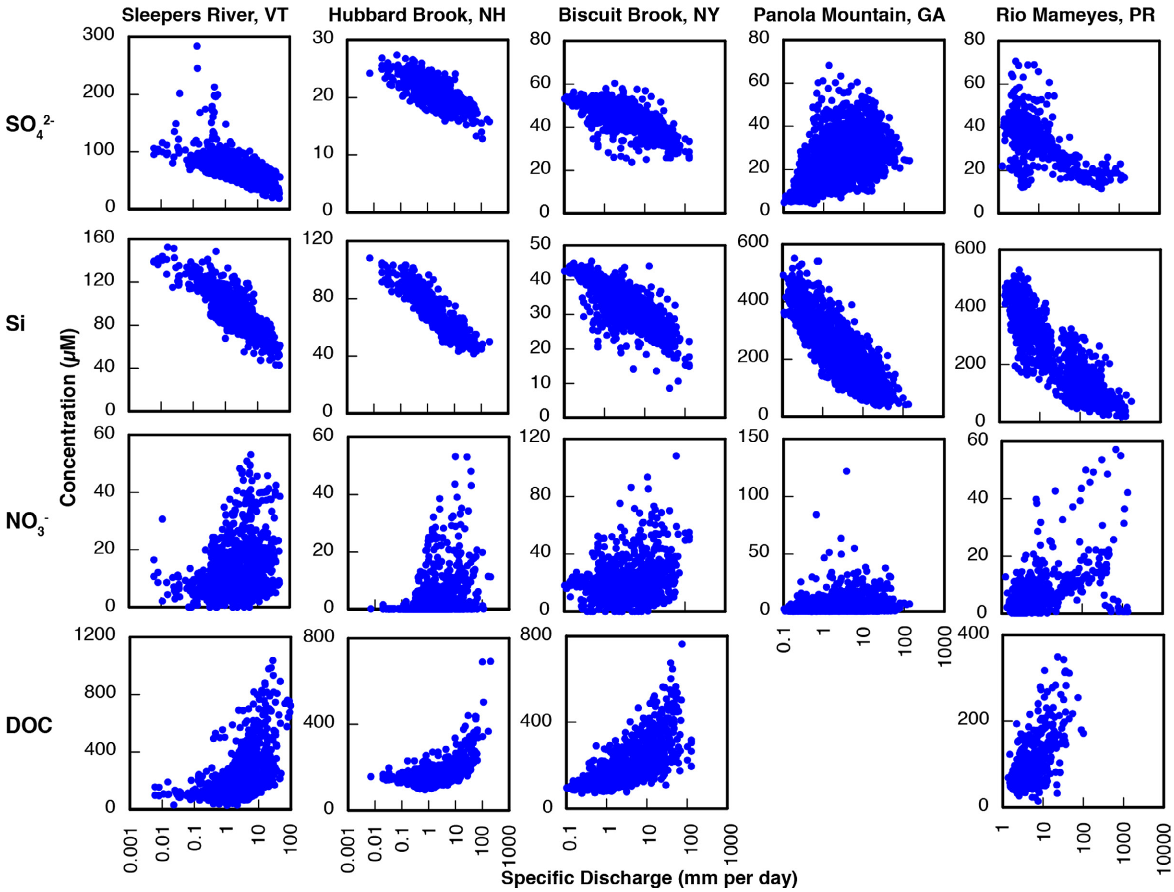

Fig. 3. Concentration vs. specific discharge (stream flow per unit area) relations for the five watersheds and four solutes. Note that the flow at Rio Mameyes is an order of magnitude greater than the other watersheds. [SO_4^{2-}, dissolved sulfate; Si, dissolved silica; NO_3^-, dissolved nitrate; DOC, dissolved organic carbon].

the complete data set, as these were designated as the "best" estimate of load to use for comparisons. Accuracies for the period of record were calculated as a percent error, the difference between a load estimate and the best estimate of load, divided by the best estimate of load and multiplied by 100%. The annual and monthly precisions were calculated similarly by calculating the percent errors at annual and monthly time steps and then determining the SE (variance) of all the percent errors. Note that the error assessment in this analysis is for load estimation methodology, sampling approach, and solute dynamics, and does not consider measurement errors in the data such as in streamflow and sample concentrations (e.g., Yanai et al. 2015). To further simplify the discussion of load errors, we qualified the statistical errors as "good", "fair", and "poor" (Table 3). These categorizations are subjective, based on reasonable expectations of accuracies and precisions in load estimates observed in various studies and are not meant to be a definitive measure of quality. Higher levels of error were expected and accepted in the categorization of NO_3^- loads because NO_3^- had much higher relative SEs in concentrations compared to the other solutes (see Tables 4 and 5 in *Results*).

Results

Streamwater concentration–flow relations

Dissolved SO_4^{2-} concentrations exhibited dilution (decreasing concentrations with increasing stream flow) at four of the five watersheds; concentrations increased with increasing stream flow at Panola Mountain (Fig. 3). The relations

Table 2. Sampling periods and number of event samples at each of the watersheds.

Watershed	Time period	Number of years	Number of event samples: Dissolved sulfate	Dissolved silica	Dissolved nitrate	Dissolved organic carbon
Sleepers River, VT	CY1999–2009	11	563	566	566	555
Hubbard Brook, NH	CY1998–2008	11	164	164	165	142
Biscuit Brook, NY	WY1999–2008	10	442	440	431	462
Panola Mountain, GA	WY2000–2009	10	1559	1547	1595	na
Río Mameyes, PR	WY1992–2002; Silica WY1993–2003; DOC WY1998–2008	11	192	1085	192	23

Note: CY, calendar year; WY, water year (defined as October 1st of previous year to September 30th); na, not applicable.

were fairly strong with some exceptions. Sleepers River exhibited high SO_4^{2-} concentrations during the 9-month period after a long drought and the Río Mameyes exhibited more variable SO_4^{2-} concentrations at lower flows. At Panola Mountain, the SO_4^{2-} relationship was much more variable than at the other watersheds, mainly due to the release of SO_4^{2-} from soils when the watershed was wetting up.

Dissolved Si had a strong, tight dilution relationship at all five watersheds. This relationship was presumably due to the contribution of more dilute water from shallow flowpaths during hydrologic events.

Dissolved NO_3^- concentrations exhibited a weak positive relationship with stream flow at all watersheds except Panola Mountain. Sleepers River and Biscuit Brook plots exhibited a wedge-shaped pattern where NO_3^- concentrations were more variable at higher stream flows. Low NO_3^- concentrations were observed at Hubbard Brook and Panola Mountain across their entire range of discharges. Río Mameyes had a stronger positive relation of NO_3^- concentrations with stream flow than the other watersheds.

Dissolved organic carbon concentrations increased with increasing stream flow at all the watersheds where it was measured (DOC data were not available for Panola Mountain). This increase was likely due to flushing of DOC stored in shallow landscape positions during events. These relations had a high degree of scatter except at Hubbard Brook. Sleepers River and Biscuit Brook DOC concentrations exhibited a "rising wedge" pattern with greater variability at higher discharges.

Model diagnostics and solute yields

The concentration regression model R^2 values, concentration and residual diagnostics, and study period yields (loads per unit area) for the 60 model data sets developed in this analysis are reported in Table 4, and are summarized by solute in Table 5. Additional details on the models are contained in the appendix. Average model R^2 values ranged from a low of 0.38 for NO_3^- to a high of 0.83 for Si, with SO_4^{2-} and DOC having intermediate values of 0.66 and 0.65, respectively. The magnitude of model R^2 reflects both the strength of the concentration–discharge relations observed in Fig. 3

Table 3. Categorization of load estimation bias and precision by solute.

Statistic	Solutes	Categorization of errors: Good	Fair	Poor
Bias	SO_4^{2-}, Si, DOC	±3%	−6% to −3% & 3% to 6%	>±6%
Bias	NO_3^-	±5%	−10% to −5% & 5% to 10%	>±10%
Annual Precision	SO_4^{2-}, Si, DOC	<5%	5 to 10%	>10%
Annual Precision	NO_3^-	<10%	10 to 20%	>20%
Monthly Precision	SO_4^{2-}, Si, DOC	<10%	10 to 20%	>20%
Monthly Precision	NO_3^-	<20%	20 to 40%	>40%

Note: SO_4^{2-}, dissolved sulfate; Si, dissolved silica; DOC, dissolved organic carbon; NO_3^-, dissolved nitrate.

Table 4. Regression model R^2 values, yield estimates, and statistics for solute concentrations (Conc.) and model residual concentrations.

				Relative standard error, in percent		Autocorrelation		Period yield ($\mu M/m^2$/year)		
Solute	Watershed	Sampling frequency	Model R^2 (adjusted)	Conc.	Residuals	Conc.	Residuals	Composite method	Regression-model method	Period-weighted approach
SO_4^{2-}	Sleepers River	Complete	0.661	32.4	18.8	0.919	0.824	148	150	150
SO_4^{2-}	Sleepers River	Weekly	0.829	32.8	13.5	0.913	0.577	149	147	154
SO_4^{2-}	Sleepers River	Monthly	0.849	33.4	12.6	0.841	0.338	150	151	159
SO_4^{2-}	Hubbard Brook	Complete	0.837	11.3	4.5	0.728	0.527	52.7	53.0	53.7
SO_4^{2-}	Hubbard Brook	Weekly	0.851	10.7	4.1	0.767	0.576	53.3	53.5	55.1
SO_4^{2-}	Hubbard Brook	Monthly	0.853	10.4	3.9	0.476	0.222	53.2	53.3	57.1
SO_4^{2-}	Biscuit Brook	Complete	0.739	14.5	7.4	0.764	0.486	121	121	120
SO_4^{2-}	Biscuit Brook	Weekly	0.777	10.8	5.1	0.642	0.280	122	122	124
SO_4^{2-}	Biscuit Brook	Monthly	0.797	11.1	4.9	0.582	0.152	123	123	127
SO_4^{2-}	Panola Mountain	Complete	0.642/0.575†	48.7	30.3	0.833	0.773	18.5	18.8	18.5
SO_4^{2-}	Panola Mountain	Weekly	0.667/0.551†	57.5	33.3	0.735	0.787	19.1	18.8	16.9
SO_4^{2-}	Panola Mountain	Monthly	0.677/0.624†	60.3	31.2	0.608	0.650	20.5	20.0	17.0
SO_4^{2-}	Río Mameyes	Complete	0.462	28.9	21.1	0.535	0.205	245	240	264
SO_4^{2-}	Río Mameyes	Weekly	0.086	23.3	22.1	0.221	0.198	253	250	270
SO_4^{2-}	Río Mameyes	Monthly	0.147	22.2	20.0	0.223	0.126	246	242	267
Si	Sleepers River	Complete	0.826	19.8	8.2	0.890	0.689	201	191	204
Si	Sleepers River	Weekly	0.778	16.4	7.7	0.858	0.645	204	203	208
Si	Sleepers River	Monthly	0.790	15.9	7.2	0.669	0.470	204	204	211
Si	Hubbard Brook	Complete	0.867	19.0	6.9	0.700	0.506	162	165	168
Si	Hubbard Brook	Weekly	0.863	17.5	6.5	0.689	0.503	163	165	175
Si	Hubbard Brook	Monthly	0.871	17.4	6.1	0.312	0.291	165	166	188
Si	Biscuit Brook	Complete	0.707	18.7	10.0	0.750	0.485	84.9	85.2	83.7
Si	Biscuit Brook	Weekly	0.747	15.5	7.7	0.688	0.550	84.6	85.2	87.4
Si	Biscuit Brook	Monthly	0.811	16.3	7.0	0.460	0.219	85.2	85.3	90.6
Si	Panola Mountain	Complete	0.883	43.2	14.7	0.846	0.528	212	211	214
Si	Panola Mountain	Weekly	0.868	24.7	8.9	0.781	0.598	217	216	235
Si	Panola Mountain	Monthly	0.863	26.4	9.5	0.641	0.375	210	211	236
Si	Río Mameyes	Complete	0.888	64.5	21.5	0.786	0.426	1920	1910	2010
Si	Río Mameyes	Weekly	0.529	23.7	16.2	0.199	0.324	2040	2060	2330
Si	Río Mameyes	Monthly	0.544	21.6	14.5	0.122	0.129	2250	2230	2560
NO_3^-	Sleepers River	Complete	0.537	72.9	49.3	0.913	0.800	27.7	29.3	27.7
NO_3^-	Sleepers River	Weekly	0.513	65.8	45.7	0.830	0.662	27.0	26.0	27.0
NO_3^-	Sleepers River	Monthly	0.561	64.7	42.0	0.587	0.362	26.0	25.0	25.9
NO_3^-	Hubbard Brook	Complete	0.408	179	137	0.886	0.846	16.1	17.6	14.8
NO_3^-	Hubbard Brook	Weekly	0.350	205	164	0.875	0.837	16.2	16.4	12.9
NO_3^-	Hubbard Brook	Monthly	0.337	198	159	0.728	0.675	18.6	19.2	10.5
NO_3^-	Biscuit Brook	Complete	0.484	69.2	49.5	0.833	0.732	69.7	69.1	69.7
NO_3^-	Biscuit Brook	Weekly	0.535	68.8	46.5	0.733	0.525	72.4	72.0	66.5
NO_3^-	Biscuit Brook	Monthly	0.474	63.1	45.0	0.682	0.544	63.3	65.0	60.4
NO_3^-	Panola Mountain	Complete	0.170	163	148	0.678	0.625	2.41	3.31	2.44
NO_3^-	Panola Mountain	Weekly	0.223	113	99.6	0.837	0.780	1.60	1.65	2.01
NO_3^-	Panola Mountain	Monthly	0.248	110	94.0	0.555	0.411	1.26	1.39	1.90
NO_3^-	Río Mameyes	Complete	0.301	102	84.1	0.775	0.701	66.8	62.7	56.5

Table 4. Continued.

Solute	Watershed	Sampling frequency	Model R^2 (adjusted)	Relative standard error, in percent: Conc.	Residuals	Autocorrelation: Conc.	Residuals	Period yield (μM/m²/year): Composite method	Regression-model method	Period-weighted approach
NO_3^-	Río Mameyes	Weekly	0.284	90.9	80.5	0.631	0.558	60.1	61.3	49.7
NO_3^-	Río Mameyes	Monthly	0.165	97.8	88.0	0.440	0.351	63.0	66.1	49.7
DOC	Sleepers River	Complete	0.631	79.4	48.1	0.796	0.678	383	510	352
DOC	Sleepers River	Weekly	0.239	59.4	51.6	0.273	0.205	339	352	314
DOC	Sleepers River	Monthly	0.254	48.5	39.8	−0.011	−0.132	317	320	276
DOC	Hubbard Brook	Complete	0.781	34.3	16.0	0.442	0.295	580	570	533
DOC	Hubbard Brook	Weekly	0.729	21.8	11.3	0.423	0.407	574	572	484
DOC	Hubbard Brook	Monthly	0.735	23.5	11.7	0.187	0.130	570	594	435
DOC	Biscuit Brook	Complete	0.706	52.5	28.4	0.585	0.465	540	579	576
DOC	Biscuit Brook	Weekly	0.670	41.7	23.9	0.238	0.284	496	519	453
DOC	Biscuit Brook	Monthly	0.764	35.9	17.3	0.005	0.028	482	490	407
DOC	Río Mameyes	Complete	0.491	53.2	37.7	0.110	0.107	1130	1180	897
DOC	Río Mameyes	Weekly	0.494	53.8	38.0	0.100	0.098	1140	1190	894
DOC	Río Mameyes	Monthly	0.575	55.1	34.9	0.023	0.082	1240	1260	846

Notes: The relative standard error of residuals is calculated with respect to the average concentration. SO_4^{2-}, dissolved sulfate; Si, dissolved silica; NO_3^-, dissolved nitrate; DOC, dissolved organic carbon.
† First value for day of year <200, second value for day of year >200.

and the additional variance explained by the seasonal and long-term trend model parameters. The average variance by solute, as expressed by the relative SE of concentrations, was lowest for SO_4^{2-} (27%) and Si (33%), and highest for NO_3^- (117%). The relative SE of the residuals was lowest for Si (12%) and highest for NO_3^- (94%). The percentage reduction in variance of the relative SE between the concentrations and residuals was highest for Si (63%) and lowest for NO_3^- (20%), and was related to the amount of variance explained by their models.

Autocorrelations in the concentrations were fairly high and similar for SO_4^{2-}, Si and NO_3^-, ranging from 0.76 to 0.82, and were lower for DOC (0.48; Table 5). There was still quite a bit of autocorrelation remaining in the residuals, ranging from 0.39 (DOC) to 0.74 (NO_3^-). The amount of autocorrelation removed between the concentrations and residuals was positively related to the amount of variance explained by the models.

Composite method complete data set yields spanned 0.47 (DOC) to 1.46 (NO_3^-) orders of magnitude (Table 5). The large ranges are indicative of the diversity amongst the five watersheds for each solute. Yields were highest at Río Mameyes for all solutes except NO_3^- (Table 4), due

Table 5. Summary of model R^2 values, solute concentration and model residual statistics, and composite method period yields by solute for the complete sampling cases.

Solute	Average model R^2 (adjusted)	Average relative standard error (%): Concentrations	Residuals	Reduction between concentrations and residuals (%)	Average autocorrelation: Concentrations	Residuals	Composite method period yield (μM/m²/year): Minimum	Maximum	Order of magnitude difference
SO_4^{2-}	0.66	27	16	40	0.76	0.56	18.5	245	1.12
Si	0.83	33	12	63	0.79	0.53	84.9	1920	1.35
NO_3^-	0.38	117	94	20	0.82	0.74	2.41	69.7	1.46
DOC	0.65	55	33	41	0.48	0.39	383	1130	0.47

Notes: The relative standard error of residuals is calculated with respect to the average concentration.
SO_4^{2-}, dissolved sulfate; Si, dissolved silica; DOC, dissolved organic carbon; NO_3^-, dissolved nitrate.

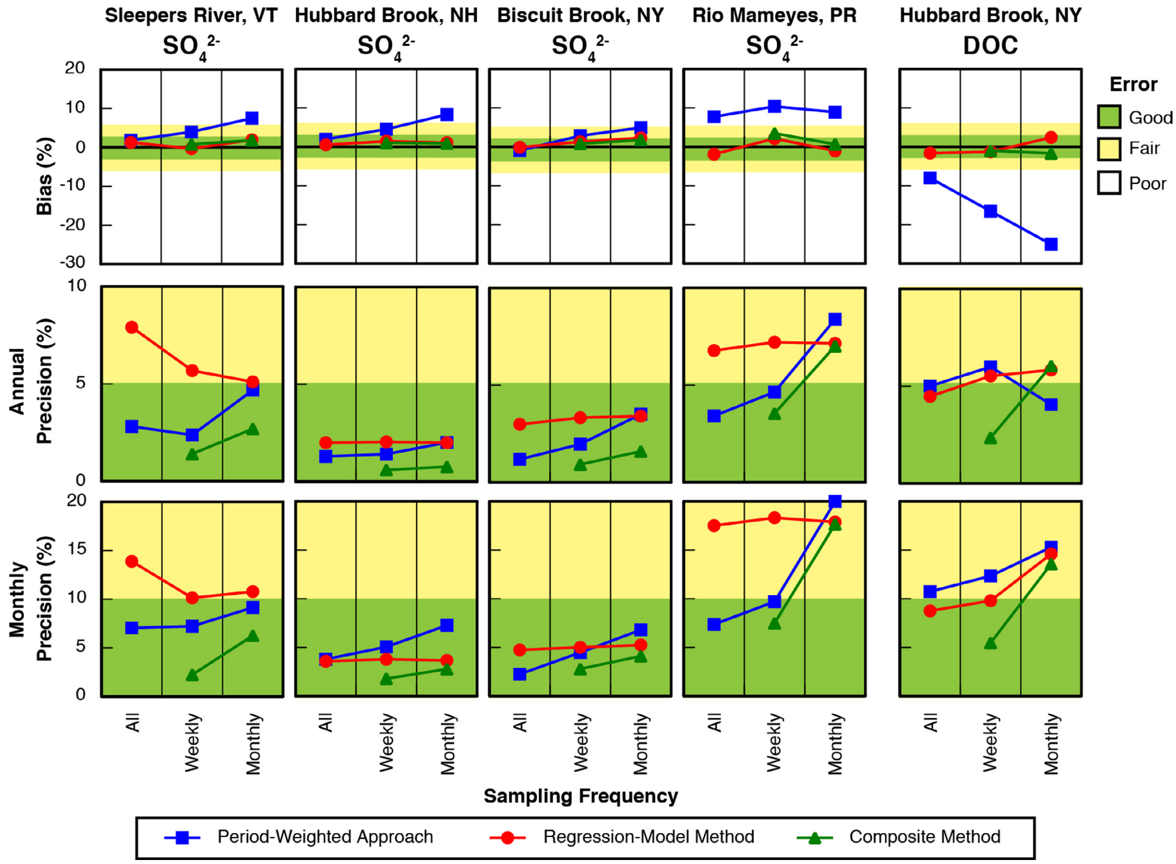

Fig. 4. Load estimation model bias and annual and monthly precision for solutes with strong concentration–discharge relations. [SO_4^{2-}, dissolved sulfate; DOC, dissolved organic carbon].

largely to the order of magnitude higher specific discharge (stream flow per unit area) observed at this watershed (Fig. 3).

Categorization of solute responses and load errors

The various solute-watershed combinations can be categorized into three groups, based on aspects of their concentration–discharge relationships (Fig. 3): (1) solutes with strong, tight concentration–discharge relations where the solute was conservatively transported (Si) or was relatively unreactive, such as SO_4^{2-} (except at Panola Mountain) and DOC at Hubbard Brook; (2) solutes where the concentration–discharge relation was more moderate and variable, such as DOC (except at Hubbard Brook) and SO_4^{2-} at Panola Mountain, and; (3) NO_3^-, which exhibits a weak to nonexistent concentration–discharge relation at all sites. These categorizations were helpful in describing the load errors due to similarities in the results within these categories.

Load errors for solutes with strong concentration–discharge relations

For solutes with strong, tight concentration–discharge relations (Figs. 4 and 5), regression-model and composite methods load accuracies and precisions were mostly categorized as good for all three sampling cases. All three sampling cases were characterized by low levels of unexplained variability, with the relative SE of residuals ≤21.5% (Table 4). Composite method annual and monthly precisions categorized as fair and poor typically had autocorrelation in residuals of <~0.15, which is an indication that the sampling frequencies were too low for the sample concentrations to be representative between samples.

Period-weighted approach accuracies were typically good for the complete data set, but

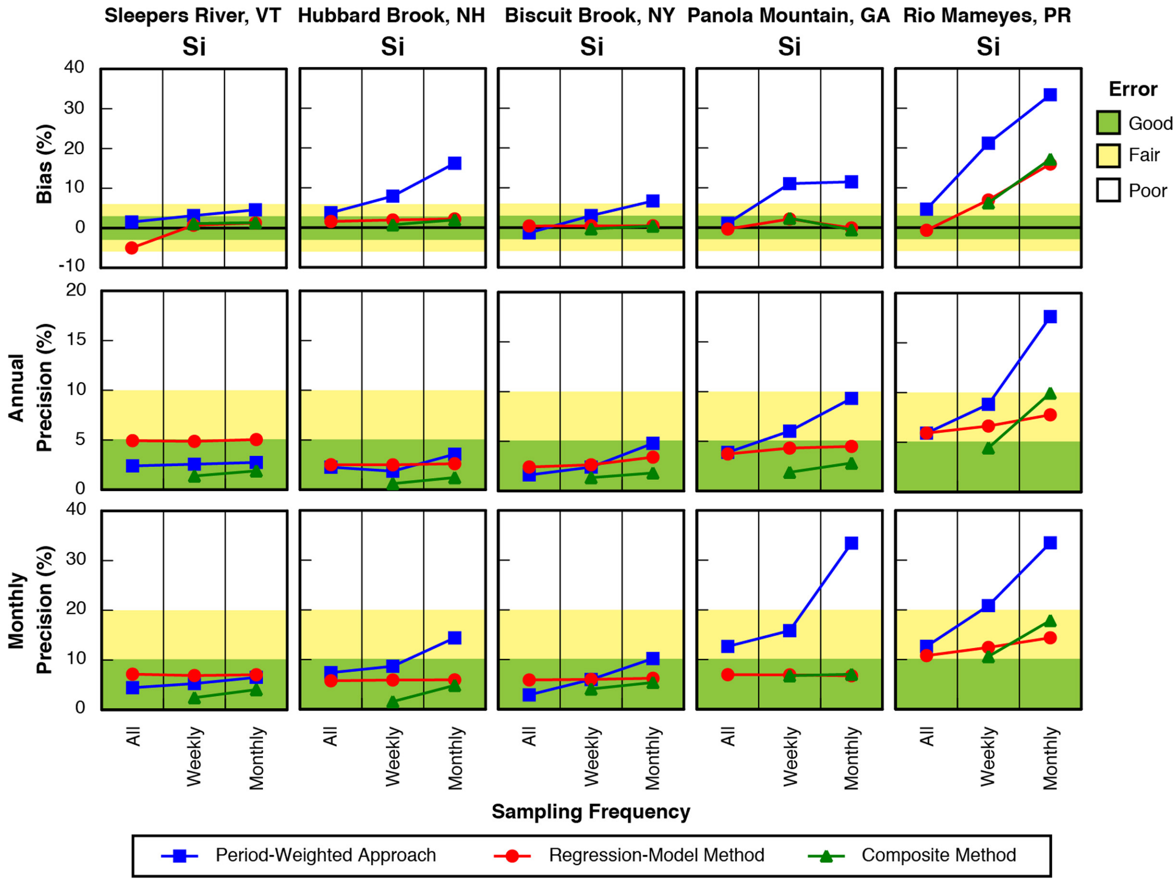

Fig. 5. Load estimation model bias and annual and monthly precision for solutes with strong concentration–discharge relations. [Si, dissolved silica].

were more biased as sampling frequency decreased. Positive biases were observed for the period-weighted approach weekly and monthly data sets for solutes that exhibited dilution and negative biases were observed for solutes that showed increasing concentrations with increasing stream flow. These patterns of bias result when baseflow concentrations are inappropriately applied to intervening storm periods that contribute a large portion of flow and total load. The converse—storm sample concentrations inappropriately applied to adjacent baseflow periods—may also occur, but it typically does not offset the former effect because it is associated with lower stream flows. Period-weighted approach precisions were good, except when the relative SEs of the concentrations were >~20% and the autocorrelation of the concentrations was <~0.3–0.4 (Table 4).

Load errors for solutes with moderate concentration–discharge relations

For solutes with moderate and more variable concentration–discharge relations (Fig. 6), regression-model and composite methods load accuracies ranged from good to poor, composite method precisions typically ranged from good to fair, and regression-model method precisions typically ranged from fair to poor. Errors were generally greater with less frequent sampling except for regression-model method precisions, which were relatively unaffected by sampling frequency. The greater regression-model method biases observed with decreased sampling frequencies is an indication that the sampling was less than adequate to properly define the concentration–discharge relationship (Smith and Croke 2005). Period-weighted approach accuracies were typically poor and were more negatively biased

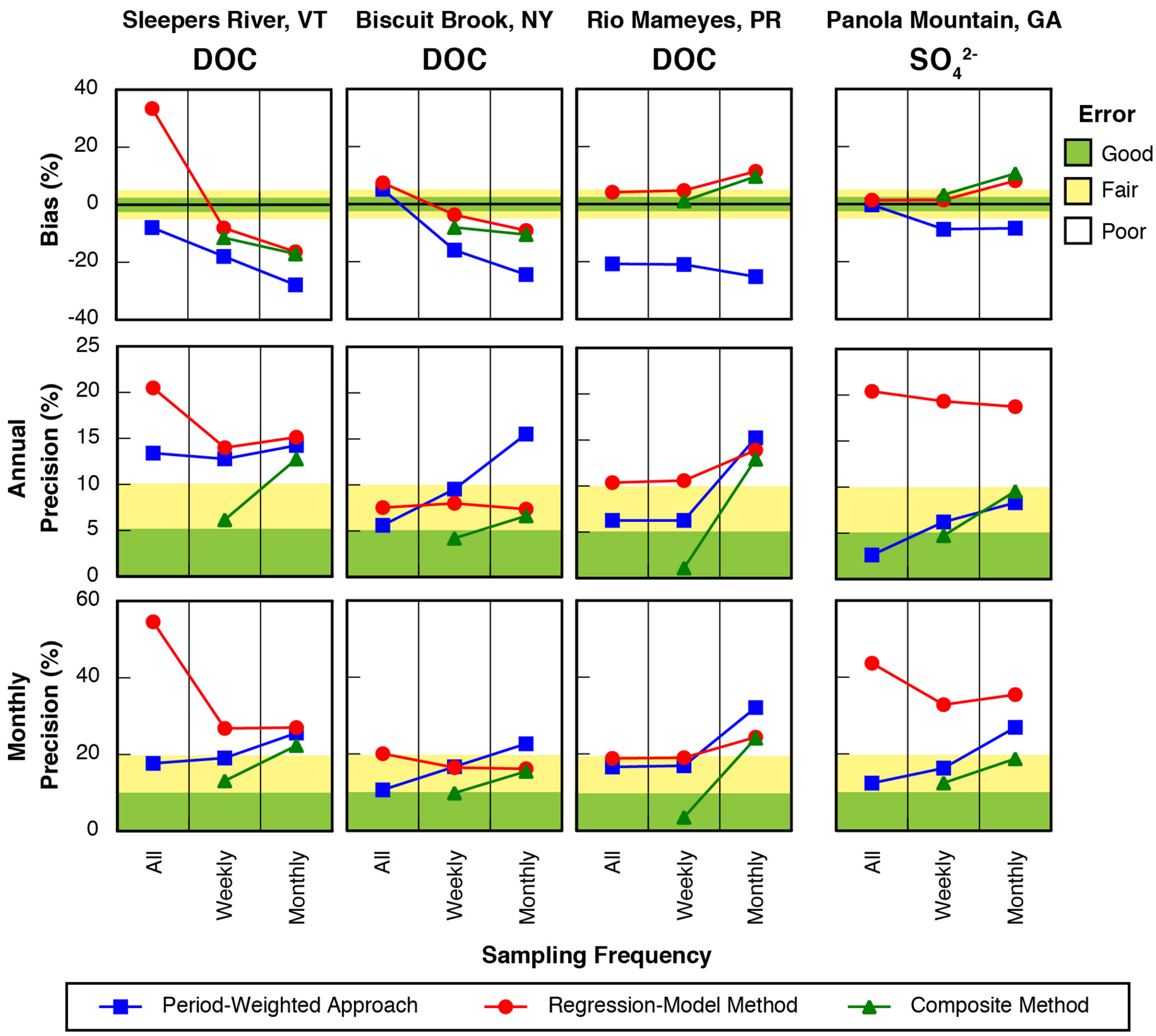

Fig. 6. Load estimation model bias and annual and monthly precision for solutes with moderate concentration–discharge relations. [DOC, dissolved organic carbon; SO_4^{2-}, dissolved sulfate].

with less frequent sampling, as these solutes at these sites exhibited increasing concentrations with increasing stream flow. While average model R^2 values were similar for DOC and SO_4^{2-} (Table 5), SO_4^{2-} precisions were better than DOC precisions. Some of these differences are likely related to higher average relative SEs of concentrations and residuals and lower average autocorrelations of concentrations and residuals for DOC compared to SO_4^{2-}. Period-weighted approach precisions typically ranged from fair to poor, and were often better than regression-model method precisions, especially for the complete and weekly data sets.

Load errors for solutes with weak concentration–discharge relations

For NO_3^-, which exhibited at best a weak concentration–discharge relation at all the watersheds, errors were much larger than for the other solutes (Fig. 7), as corroborated by the high relative SEs in concentrations and residuals (Table 4). Period-weighted approach accuracies and annual precisions ranged from good to poor, and were better at Sleepers River and Biscuit Brook where the relative SE of the concentrations were lower (<100%), and for cases when the autocorrelation of the concentrations

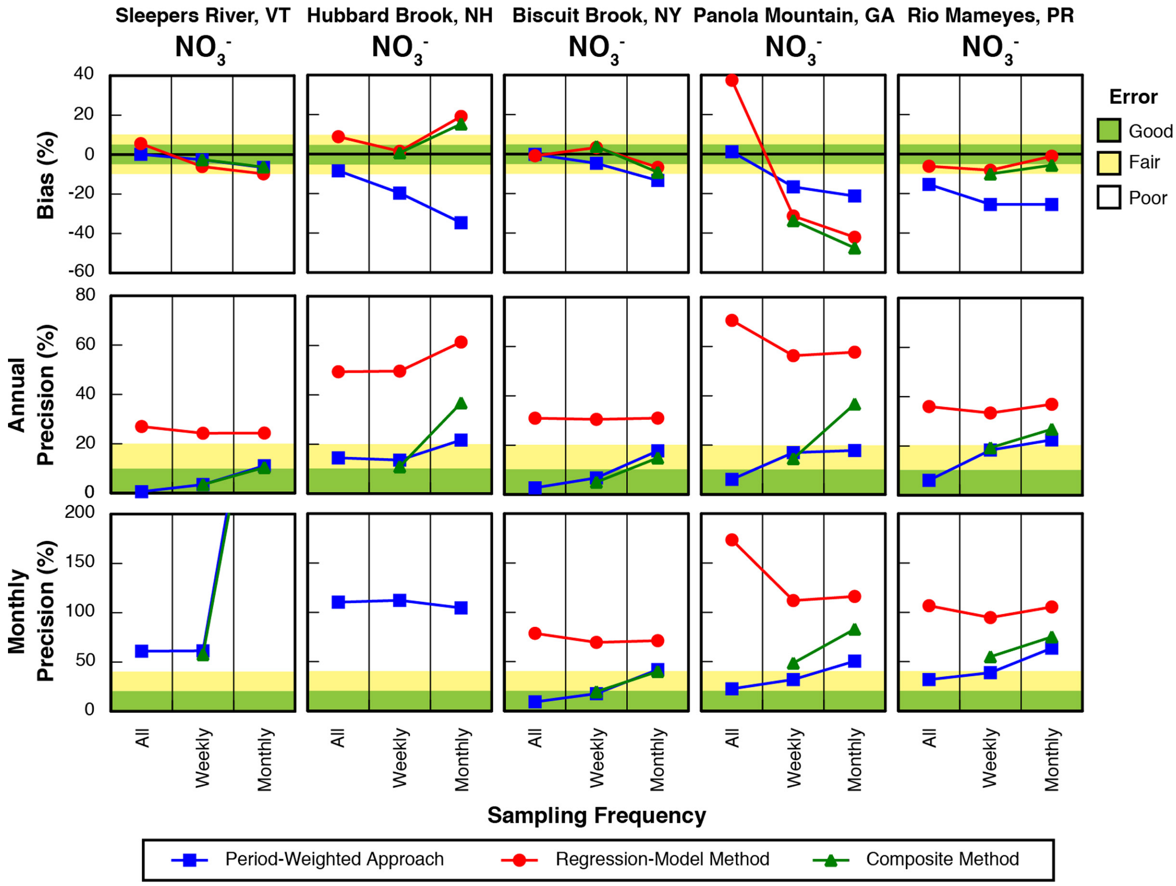

Fig. 7. Load estimation model bias and annual and monthly precision for solutes with weak to nonexistent concentration–discharge relations. Monthly precisions >200% are not shown. [NO_3^-, dissolved nitrate].

were higher and accompanied by more frequent sampling. While composite method accuracies and precisions were similar to period-weighted approach loads for the complete data set at watersheds that had good event sampling (Sleepers River, Biscuit Brook, and Panola Mountain), the composite method was sometimes worse than the period-weighted approach at the reduced sampling frequencies. Regression-model method accuracies ranged from good to poor and were typically fair or poor, and were fairly similar to composite method accuracies at weekly and monthly sampling frequencies. Regression-model method precisions were particularly poor, indicating that the models had little ability to predict short-term variations in concentration. Percentage errors in monthly precisions were quite large in some cases (errors >200% not shown in figures). One reason for these large errors is that small differences in monthly loads with very low observed NO_3^- concentrations often resulted in large errors when reported on a percentage basis.

Assessment of true loads

For solutes that have moderate to strong concentration–discharge relations (SO_4^{2-}, Si, and DOC), the differences in composite method accuracies and precisions between the complete sampling case and a sampling frequency where every other event sample was included were generally small, ≤±1.5% for accuracy and <3.0% for annual and monthly precisions in most cases (Table 6). There were three exceptions with higher monthly precision differences ranging between 4.0 and 6.1%. The smaller differences indicate that the amount of event sampling was sufficient for the composite method estimates to be close to the true loads. Río Mameyes had very few event DOC samples, so while

Table 6. Percentage difference between composite method loads all sampling case and sampling case with every other event sample.

Solute	Statistic	Sleepers River, VT	Hubbard Brook, NH	Biscuit Brook, NY	Panola Mountain, GA	Río Mameyes, PR
SO_4^{2-}	Accuracy	0.1	0.1	0.3	−0.5	0.5
SO_4^{2-}	Annual Precision	0.8	0.3	0.5	1.1	1.6
SO_4^{2-}	Monthly Precision	1.1	0.9	1.2	5.8	2.6
Si	Accuracy	0.0	0.1	0.1	0.2	−0.4
Si	Annual Precision	0.1	0.5	0.5	1.0	1.0
Si	Monthly Precision	0.3	1.1	1.6	4.0	2.5
DOC	Accuracy	0.4	−0.1	−1.5	na	−0.1
DOC	Annual Precision	0.8	1.9	2.8	na	0.3
DOC	Monthly Precision	1.6	2.7	6.1	na	2.4
NO_3^-	Accuracy	0.0	0.3	0.0	−0.6	1.9
NO_3^-	Annual Precision	1.1	3.9	1.2	8.0	4.8
NO_3^-	Monthly Precision	2.2	367.2	6.5	24.9	10.7

Note: SO_4^{2-}, dissolved sulfate; Si, dissolved silica; DOC, dissolved organic carbon; NO_3^-, dissolved nitrate; na, not applicable.

the differences in loads were small for this case, there was little difference between sampling data sets, which makes this comparison for determining convergence inconclusive. Load errors determined for this case are likely less reliable.

For NO_3^-, with its poor concentration–discharge relation, the differences in composite method accuracies were small, ≤±1.5%, indicating that the composite method complete sampling case likely approached the period true load. Annual precision differences were low at Sleepers River (1.1%) and Biscuit Brook (1.2%), but were higher at the other three watersheds (3.9–8.0%), indicating approximations of the true loads at only two of the five watersheds on an annual basis. Monthly precision differences were high at all but Sleepers River, and likely indicate that the composite method complete data sets used herein were not sufficient to represent the true loads on a monthly time-step.

Discussion

Solute dynamics and load estimation approach

The best overall load accuracies and precisions were achieved for solutes with a strong concentration–discharge relation. For solutes with a moderate to strong concentration–discharge relation, composite method accuracies and precisions for weekly and monthly data sets were almost always as good or better than the other two load estimation approaches, regardless of sampling frequency. Fair and poor annual and monthly precisions with the composite method were related to weak temporal patterns in the residual concentrations (autocorrelation in residuals <~0.15). In these cases the residuals were not particularly useful in improving the load estimates at annual and monthly time scales; similar to findings for suspended sediment with a fairly strong concentration–discharge relation (Aulenbach 2013).

For the regression-model method, accuracies and precisions were often good for strong concentration–discharge relations, as reported by Preston et al. (1989), but only where the relative SEs of residuals were ≤21.5% (Table 4). For moderate relations where the SEs of residuals were higher, loads calculated by regression were typically more biased, while precision was typically fair or poor and relatively unaffected by sampling frequency. Regression models predict the mean concentration response, and short-term variability is not sufficiently captured in the models to predict loads well at annual and monthly time scales (Aulenbach 2013).

The period-weighted approach loads were susceptible to bias and imprecision when there was a moderate to strong concentration–discharge relation. Good accuracies were sometimes achieved for the all-sampling frequency cases. Good precisions could be achieved except when concentrations were variable (relative SEs of the concentrations were >~20%) and when there were not strong temporal patterns in concentrations (autocorrelation of the concentrations was <~0.3–0.4); and are consistent with the findings of Kerr et al. (2015).

For solutes with a poor concentration–discharge relation (R^2 < about 0.3), the period-weighted approach may be the best. The regression-model method often exhibited the worst accuracies and precisions of the three load methods. This result is consistent with the findings of Moatar and Meybeck (2005) and Quilbé et al. (2006) who indicated that other methods should be used when regression models are not effective. In this situation, the observed concentrations in the period-weighted approach are preferable to attempting to correct flawed predicted concentrations using the composite method. This approach is supported in that the precision of the composite method for weekly and monthly sampling frequencies is better than that of the regression-model method, but not always as good as that of the period-weighted approach (Fig. 7). The period-weighted approach has better accuracy and precision when sampling frequency is high and when concentrations are less variable (lower residual SEs) and have stronger temporal patterns (higher autocorrelations).

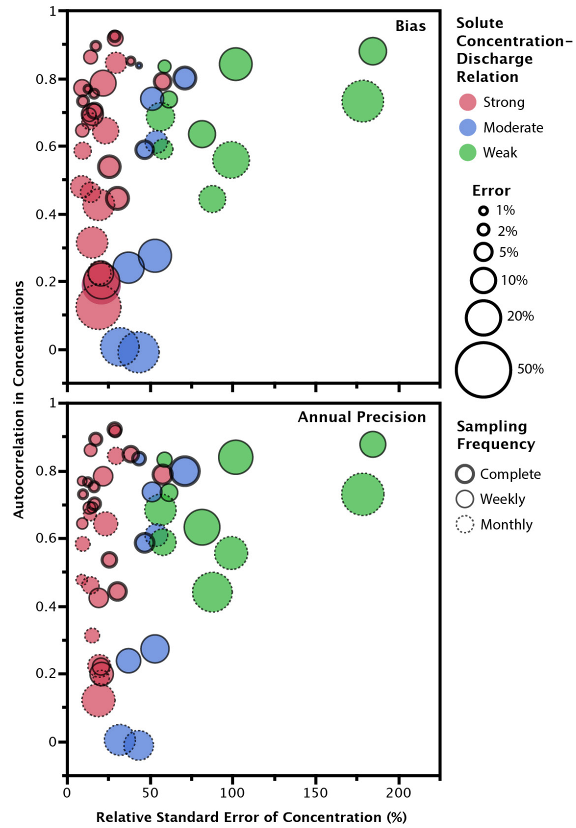

Fig. 8. Bubble plots of period-weighted approach load biases (magnitude) and annual precisions for the various combinations of solutes, watersheds, and sampling frequencies on plots of autocorrelation of concentrations vs. relative standard errors of concentrations. Bubble size indicates magnitude of error, bubble color indicates strength of solute concentration–discharge relation, and bubble border line-style indicates sampling frequency.

Diagnostics for estimating errors

Errors were assessed by plotting their magnitude as a function of various pairs of diagnostics. Diagnostics explored included the relative SEs of the concentrations and residuals, the autocorrelations of concentrations and residuals, and the model R^2 values (Table 4). Errors were assessed only for bias and annual precision, as monthly precisions were not always considered legitimate (Table 6). Biases and annual precisions were recalculated for NO_3^- using the complete data set period-weighted approach, as these loads were considered more likely to represent the "true" loads, as previously discussed. Load error estimates for DOC at Río Mameyes were excluded from this analysis, as there were too few event samples to evaluate convergence on the true loads.

For the period-weighted approach, errors were smaller when autocorrelations and sampling frequencies were high, and when relative SEs were low, as indicated by the size of the error bubbles (Fig. 8). Solutes with strong concentration–discharge relations generally had the best accuracy and precision, as they typically exhibited low relative SEs in their concentrations, and were lower when autocorrelation was higher. Dissolved NO_3^-, with weak concentration–discharge relations, typically had the largest errors, which were associated with high relative standard errors in NO_3^- concentrations, despite having relatively high autocorrelations. The overall effect of autocorrelation makes sense, as the ability for period weighting to estimate load is based on the premise that the concentrations between samples are similar to those of the adjacent samples, and autocorrelation is a measure of similarity between samples. Therefore, the sampling frequency is ac-

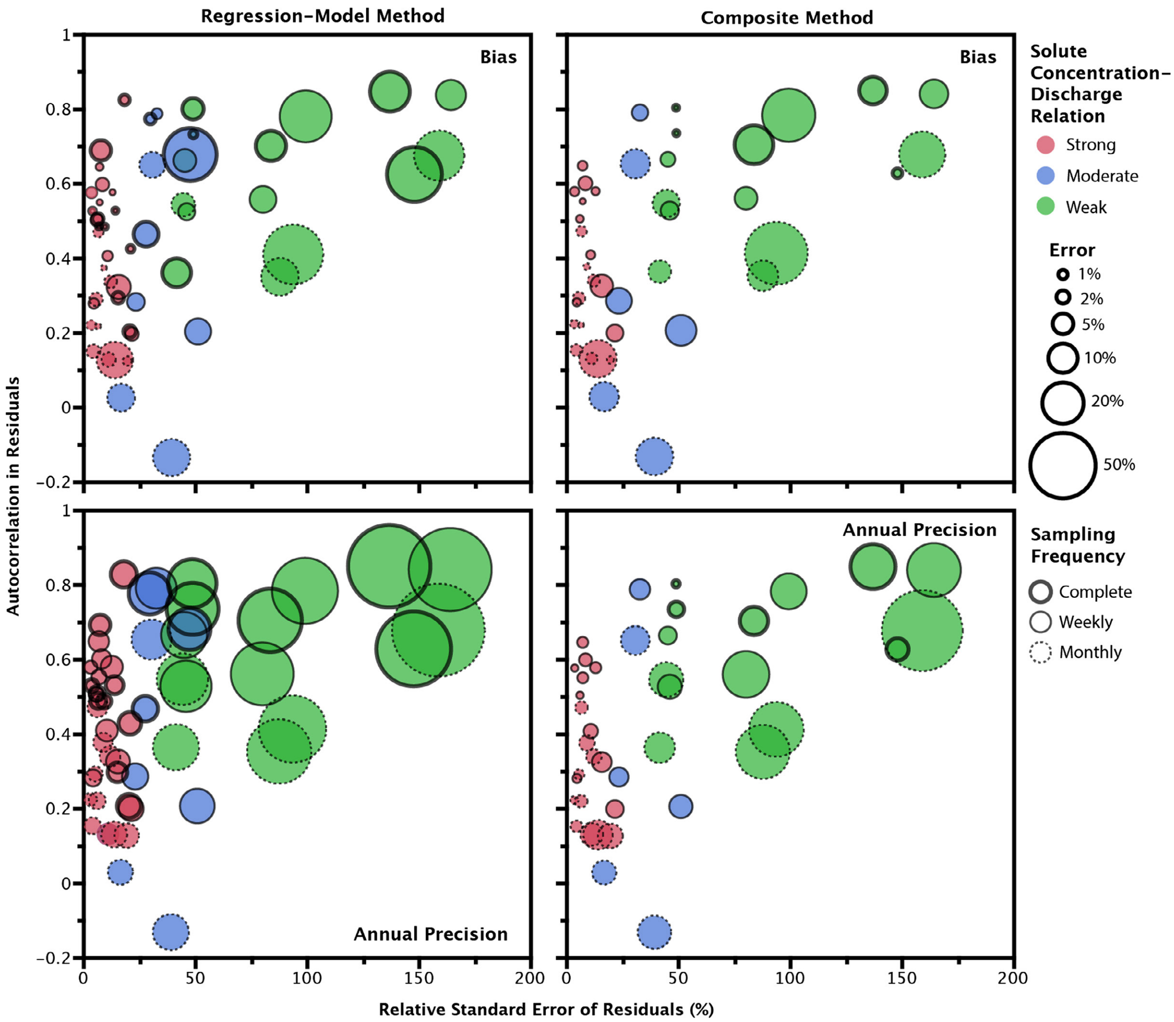

Fig. 9. Bubble plots of regression-model method and composite method load biases (magnitude) and annual precisions for the various combinations of solutes, watersheds, and sampling frequencies on plots of autocorrelation of residuals vs. relative standard errors of residuals. The relative standard error of residuals is calculated with respect to the average concentration. Bubble size indicates magnitude of error, bubble color indicates strength of solute concentration–discharge relation, and bubble border line-style indicates sampling frequency.

counted for to some extent in the autocorrelation diagnostic. Smaller errors were associated with higher autocorrelations in concentrations and reflected higher sampling frequencies, as indicated by the line styles of the error bubbles. Loads were also more difficult to estimate accurately for solutes that are more variable (e.g., have higher relative SEs such as NO_3^-).

For the two load estimation approaches that employ regression models, the two diagnostics that most influenced bias and precision were the autocorrelation in the residual concentrations and the relative SE of the residuals. Bias patterns for the regression-model and composite methods were similar, and were typically lowest for solutes with strong concentration–discharge relations and highest for solutes with weak concentration–discharge relations (Fig. 9). These biases were lower than the biases observed for the period-weighted approach (Fig. 8), except for solutes that had weak concentration–discharge relations (NO_3^-), which exhibited somewhat higher biases. Biases were larger when the relative SE of the residuals was higher. The relative SE of the residuals represents the remaining variance not explained by the model, and the probability of observing larger biases increases when variance is greater. Biases were relatively uninfluenced by the amount of autocorrelation in the residuals, because deviations from the model tended to cancel out over the period of record.

For precision, regression-model method errors were unaffected by the autocorrelation of the residuals, but had poorer precisions than the composite method. The composite method did better when residual autocorrelations were higher, as the stronger patterns in the residuals were used to adjust concentrations based on the observations, thereby improving the load estimates in the short-term. But when autocorrelations were low (<0.2), precisions were similar for the regression-model and composite methods, as there was little structure in the residuals by which the composite method could improve the load estimates. Better composite method precisions for individual solute-watershed combinations were associated with higher autocorrelations in their residuals that reflected their higher sampling frequencies.

While the strength of the concentration–discharge relation of each solute was related to its relative SE, resulting in solutes with similar concentration dynamics clustering in particular regions of the diagnostic plots, the magnitude of errors, indicated by the area of the bubbles, show greater dependence on the diagnostics. Therefore, it should be possible to use these diagnostics and the plots in Figs. 8 and 9 to approximate the error for solutes at other watersheds. The approximated errors could also be utilized in sampling design of future studies if expectations of the solute diagnostics can be reasonably approximated. Furthermore, these results can help determine the appropriate load estimation method—the method with the lowest errors. Combining the results from solutes with both strong and moderate concentration–discharge relations shows that the composite method was the best method when autocorrelations in residuals was >0.2, and the regression-model method was preferable when autocorrelations in residuals was <0.2. In contrast, for solutes with weak concentration–discharge relations, the period-weighted approach had the lowest errors. One shortcoming of these diagnostic plots is that no solutes in this study had high variances and low autocorrelations. It is not known how common this combination of conditions might be.

Caveats regarding diagnostics

The diagnostics are a useful guide for approximating uncertainties, but they can sometimes be misleading. While the relative SE of concentrations can be used to estimate the error of the period-weighted approach, this metric is also influenced by the sampling scheme. For instance, if concentration variance was greater at higher stream flows, having fewer event samples would decrease the relative SE of concentrations and suggest a lower level of error. In reality, the underlying relative SE of concentrations that relates to the errors in the period-weighted approach is better characterized with a sufficient number of event samples to represent the concentration variations that exist. While the effect of sampling frequency on the relative SE of concentrations was observed (Table 4), the effects were generally not large.

Similarly, autocorrelations of concentrations used to assess the errors in the period-weighted approach can also be influenced by the sampling scheme. One might expect autocorrelations to be higher without event sampling because there would be more consistency in concentrations between adjacent baseflow samples than between a baseflow and an event sample, or between adjacent event samples at different stream flows. But the inclusion of event sampling should improve the period-weighted approach load estimates even though the autocorrelation diagnostic might suggest the opposite. In the current analysis this issue was not apparent, as autocorrelations were typically greater for the complete data set than the weekly data set (Table 4).

Load estimation method selection

Guidelines for selecting the appropriate load estimation method can be made based on the relations observed between the load errors and the various solute dynamics, sampling frequencies, and diagnostics in this analysis (Fig. 10). The model R^2 cutoff of 0.3 determined in this analysis is an approximation. Hence, when model R^2 values are near this cutoff, values of various diagnostics and Figs. 8 and 9 should

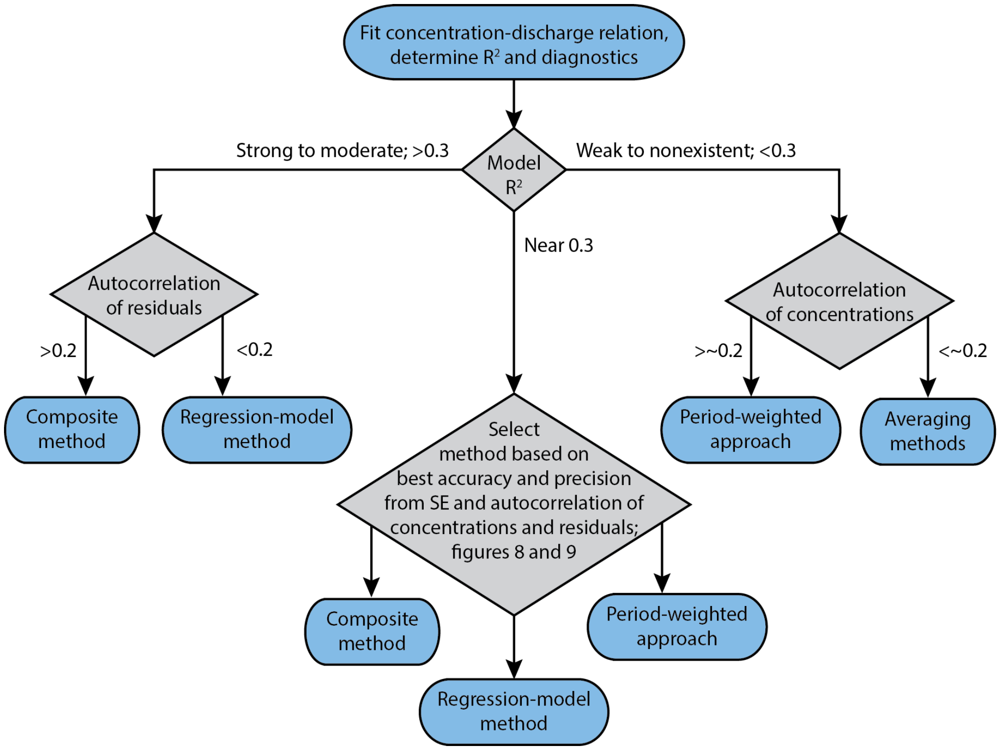

Fig. 10. Flowchart for selecting most appropriate load estimation method using model diagnostics and Figs. 8 and 9. [SE, standard error].

be relied upon to estimate errors and determine the method based on the lowest errors. While not assessed in this analysis, we expect that the period-weighted approach is not useful for solutes with weak concentration–discharge relations and low autocorrelations in concentrations, similar to what was found for the composite method and low autocorrelations in residuals. A 0.2 autocorrelation cutoff similar to that of the composite method is likely, and below this cutoff, an averaging method is more likely appropriate for estimating loads (e.g., a flow-weighted average).

Summary and Conclusions

We evaluated solute load estimation methods over a wide range of concentration dynamics for four solutes (SO_4^{2-}, Si, NO_3^-, and DOC) at five diverse small watersheds. The best approach to load estimation and the corresponding accuracy and precision depended upon specific solute concentration dynamics. By assessing three common load estimation methods (period-weighted approach, regression-model method, and composite method) with various subsets of the data (weekly plus events, weekly, and monthly), it was generally possible to identify the best approach for each solute for the data available.

No load estimation method was always superior. The composite method was the best approach for solutes with moderate (SO_4^{2-} at Panola Mountain and DOC at all watersheds except Hubbard Brook) to strong (Si, SO_4^{2-} at all watersheds except Panola Mountain, and DOC at Hubbard Brook) concentration–discharge relationships (model R^2 values > about 0.3). In these cases, bias and precision were typically categorized as good except where the relative SEs of the residuals were about 15% or more. For solutes with a moderate concentration–discharge relationship, biases ranged from good to poor, and precisions typically ranged from good to fair, with errors generally worse with less frequent sampling. The regression-model method performed almost as well as the composite method when concentration–discharge relationships were strong, but did not perform as well when the concentration–discharge relation was only moderate. The period-weighted approach was typically biased without sufficient event sampling, and precisions were categorized as good only when the relative SE of concentrations was below about 20%.

The period-weighted approach was the best approach for solutes with nonexistent or weak concentration–discharge relationships (NO_3^-; model R^2 values < about 0.3). Period-weighted approach biases and annual precisions ranged from good to poor, and were better when the relative SE of the concentrations was lower and for cases when the autocorrelation of the concentrations was higher and accompanied by more frequent sampling. The regression-model method was often biased, an indication of either weak or poorly fit models. Regression-model precisions were particularly poor as the regression models had little ability to predict short-term variations in concentration.

Sample and model diagnostics can be used to select the most appropriate load estimation method and approximate the bias and precision expected without the need for more complex over- and sub-sampling analyses. For the period-weighted approach, the relative SE of the concentrations and autocorrelation of the concentrations were useful diagnostics, with more accurate and precise estimates when relative SEs were lower and autocorrelations were higher. For methods that used regression models, the relative SE of the model residuals was a good indication of errors, while autocorrelation in the residuals indicated how useful the composite method would be at improving precision at shorter reporting intervals compared to a regression-model method. But when autocorrelations were <0.2, precisions were similar for both methods and the simpler regression-model method should be used. These guidelines should be applicable to other load estimation studies.

Acknowledgments

We thank the Quantifying Uncertainty in Ecosystem Studies (QUEST) Research Coordination Network (www.quantifyinguncertainty.org), which provided financial support for this analysis and for publication charges. QUEST is hosted by the State University of New York College of Environmental Science and Forestry and is funded by the National Science Foundation. Many thanks go to all the individuals and projects that helped collect, analyze, and share the data necessary for this analysis. Data for this analysis were provided by the Hubbard Brook Ecosystem Study (HBES), the Luquillo Long Term Ecological Research (LTER) Network, and the U.S. Geological Survey (USGS), including data from the Water, Energy and Biogeochemical Budgets (WEBB) program. The HBES is a collaborative effort at the Hubbard Brook Experimental Forest, which is operated and maintained by the U.S Department of Agriculture (USDA) Forest Service, Northern Research Station, Newtown Square, PA. Significant funding for collection of the HBES data was provided by grants from the National Science Foundation (NSF) LTER Network (NSF grant number DEB-1114804). Significant funding for collection of the Río Mameyes Luquillo LTER data was provided by grants from the NSF to the Department of Environmental Science, University of Puerto Rico, and to the International Institute of Tropical Forestry USDA Forest Service (NSF grant numbers BSR-8811902, DEB 9411973, DEB 0080538, DEB 0218039, DEB 0620910, DEB 0963447, and DEB-129764). The USDA Forest Service, the University of Puerto Rico, and the Luquillo Critical Zone Observatory (CZO; EAR EAR-1331841) gave additional support. Studies at Panola Mountain Research Watershed were conducted in cooperation with the Georgia Department of Natural Resources.

Literature Cited

Appling, A. P., M. C. Leon, and W. H. McDowell. 2015. Reducing bias and quantifying uncertainty in watershed flux estimates: the R package loadflex. Ecosphere 6:269.

Aulenbach, B. T. 2013. Improving regression-model-based streamwater constituent load estimates derived from serially correlated data. Journal of Hydrology 503:55–66.

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

Bailey, R. G. 1976. Ecoregions of the United States (map). U.S. Department of Agriculture, Forest Service, Ogden, UT.

Bryce, S. A., G. E. Griffith, J. M. Omernik, G. Edinger, S. Indrick, O. Vargas, and D. Carlson. 2010. Ecoregions of New York (color poster with map, descriptive text, summary tables, and photographs; map scale 1:1,250,000). U.S. Geological Survey, Reston, Virginia, USA.

Burns, D. A., P. S. Murdoch, G. B. Lawrence, and R. L. Michel. 1998. The effect of groundwater springs on NO_3-concentrations during summer in Catskill Mountain streams. Water Resources Research 34:1987–1996.

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

Crawford, C. G. 1991. Estimation of suspended-sediment rating curves and mean suspended-sediment loads. Journal of Hydrology 129:331–348.

Griffith, G. E., J. M. Omernik, J. A. Comstock, S. Lawrence, G. Martin, A. Goddard, V. J. Hulcher, and T. Foster. 2001. Ecoregions of Alabama and Georgia, (color poster with map, descriptive text, summary tables, and photographs; map scale 1:1,700,000). U.S. Geological Survey, Reston, Virginia, USA.

Griffith, G. E., J. M. Omernik, S. A. Bryce, J. Royte, W. D. Hoar, J. W. Homer, D. Keirstead, K. J. Metzler, and G. Hellyer. 2009. Ecoregions of New England (color poster with map, descriptive text, summary tables, and photographs; map scale 1:1,325,000). U.S. Geological Survey, Reston, Virginia, USA.

Horowitz, A. J. 2003. An evaluation of sediment rating curves for estimating suspended sediment concentrations for subsequent flux calculations. Hydrological Processes 17:3387–3409.

Huntington, T. G., R. P. Hooper, and B. T. Aulenbach. 1994. Hydrologic processes controlling sulfate mobility in a small forested watershed. Water Resources Research 30:283–295.

Johnson, A. H. 1979. Estimating solute transport in streams from grab samples. Water Resources Research 15:1224–1228.

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

Kerr, J. G., M. C. Eimers, and H. Yao. 2015. Estimating stream solute loads from fixed frequency sampling regimes: the importance of considering multiple solutes and seasonal fluxes in the design of long-term monitoring networks. Hydrological Processes. http://dx.doi.org/10.1002/hyp.10733

Larson, S. J., P. D. Capel, D. A. Goolsby, S. D. Zaugg, and M. W. Sandstrom. 1995. Relations between pesticide use and riverine flux in the Mississippi River basin. Chemosphere 31:3305–3321.

Lawrence, G. B., and C. T. Driscoll. 1990. Longitudinal patterns of concentration-discharge relationships in stream water draining the Hubbard Brook Experimental Forest, New Hampshire. Journal of Hydrology 116:147–165.

Likens, G. E., F. H. Borman, N. M. Johnson, and R. S. Pierce. 1967. The calcium, magnesium, potassium, and sodium budgets for a small forested ecosystem. Ecology 48:772–785.

Likens, G. E., F. H. Bormann, R. S. Pierce, J. S. Eaton, and N. M. Johnson. 1977. Biogeochemistry of a forested ecosystem. Springer-Verlag, New York, New York, USA.

McNab, W. H. and P. E. Avers. 1994. Ecological Subregions of the United States: Section Descriptions. WO-WSA-5. USDA Forest Service, Washington, D.C., USA.

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

Moatar, F., and M. Meybeck. 2007. Riverine fluxes of pollutants: towards predictions of uncertainties by flux duration indicators. Comptes Rendus Geoscience 339:367–382.

Murphy, S. F., and R. F. Stallard, eds. 2012. Water quality and landscape processes of four watersheds in eastern Puerto Rico. U.S. Geological Survey Professional Paper 1789. U.S. Geological Survey, Reston, Virginia, USA.

Peel, M. C., B. L. Finlayson, and T. A. McMahon. 2007. Updated world map of the Koppen-Geiger climate classification. Hydrology and Earth System Science 11:1633–1647.

Peters, N. E., R. P. Hooper, T. G. Huntington, and B. T. Aulenbach. 2000. Panola mountain research watershed – water, energy, and biogeochemical budgets program. U.S. Geological Survey Fact Sheet 162-99. U.S. Geological Survey, Reston, Virginia, USA.

Peters, N. E., J. B. Shanley, B. T. Aulenbach, R. M. Webb, D. H. Campbell, R. Hunt, M. C. Larsen, R. F. Stallard, J. W. Troester, and J. F. Walker. 2006. Water and solute mass balance of five small, relatively undisturbed watersheds in the U.S. Science of the Total Environment 358:221–242.

Preston, S. E., V. J. Bierman Jr., and S. E. Silliman. 1989. An evaluation of methods for the estimation of tributary mass loads. Water Resources Research 25:1379–1389.

Preston, S. E., V. J. Bierman Jr., and S. E. Silliman. 1992. Impact of flow variability on error in estimation of tributary mass loads. Journal of Environmental Engineering (American Society of Civil Engineers) 118:402–419.

Quilbé, R., A. N. Rousseau, M. Duchemin, A. Poulin, G. Gangbazo, and J. Villeneuve. 2006. Selecting a calculation method to estimate sediment and nutrient loads in streams: application to the Beaurivate River (Québec, Canada). Journal of Hydrology 326:295–310.

Richards, R. P., and J. Holloway. 1987. Monte Carlo studies of sampling strategies for estimating tributary loads. Water Resources Research 23:1939–1948.

Robertson, D. M. 2003. Influence of different temporal sampling strategies on estimating total phosphorus and suspended sediment concentration and transport in small streams. The Journal of the American Water Resources Association 39:1281–1308.

Robertson, D. M., and E. D. Roerish. 1999. Influence of various water quality sampling strategies on load estimates for small streams. Water Resources Research 35:3747–3759.

Sebestyen, S. D., E. W. Boyer, and J. B. Shanley. 2009. Responses of stream nitrate and DOC loadings to hydrological forcing and climate change in an upland forest of the northeastern United States. Journal of Geophysical Research: Biogeosciences 114:G02002.

Semkin, R. G., D. S. Jefferies, and T. A. Clair. 1994. Hydrochemical methods and relationships for study of stream output from small catchments. Pages 163–187 *in* B. Moldan, and J. Cerny, editors. Biogeochemistry of small catchments: a tool for environmental research. John Wiley and Sons, New York.

Shanley, J. B., and N. E. Peters. 1993. Variations in aqueous sulfate concentrations at Panola Mountain, Georgia. Journal of Hydrology 146:361–382.

Shanley, J. B., P. Krám, J. Hruska, and T. D. Bullen. 2004. A biogeochemical comparison of two well-buffered catchments with contrasting histories of acid deposition. Water, Air & Soil Pollution 4:325–342.

Shih, G., X. Wang, H. J. Grimshaw, and J. VanArman. 1998. Variance of load estimates derived by piecewise linear interpolation. Journal of Environmental Engineering (American Society of Civil Engineers) 124:1114–1120.

Smith, C., and B. Croke. 2005. Sources of uncertainty in estimating suspended sediment load *in* Proceedings of symposium S1 (Sediment Budgets 2) held during the Seventh International Association of Hydrological Sciences Scientific Assembly at Foz do Iguaçu, Brazil. International Association of Hydrological Sciences Publication 292.

Stallard, R. F., and S. F. Murphy. 2012. Water quality and mass transport in four watersheds in eastern Puerto Rico. Pages 113–151 *in* S. F. Murphy, and R. F. Stallard, editors. Water quality and landscape processes of four watersheds in eastern Puerto Rico. U.S. Geological Survey Professional Paper 1789–E. U.S. Geological Survey, Reston, Virginia, USA.

Stallard, R. F., and S. F. Murphy. 2014. A unified assessment of hydrologic and biogeochemical responses in research watersheds in Easter Puerto Rico using runoff-concentration relations. Aquatic Geochemistry 20:115–139.

Stoddard, J. L., and P. S. Murdoch. 1991. Catskill Mountains. Pages 237–271 *in* D. F. Charles, editor. Acidic deposition and aquatic ecosystems: regional case studies. Springer-Verlag, New York.

Thomas, R. B. 1985. Estimating total suspended sediment yield with probability sampling. Water Resources Research 21:1381–1388.

Thomas, R. B., and J. Lewis. 1993. A comparison of selection at list time and time-stratified sampling for estimating suspended sediment loads. Water Resources Research 29:1247–1256.

U.S. Environmental Protection Agency. 2000. Overview of current total maximum daily load – TMDL – program and regulations. EPA841-F-00-009:4. U.S. Environmental Protection Agency, Washington, D.C., USA.

Vanni, M. J., W. H. Renwick, J. L. Headworth, J. D. Auch, and M. H. Schaus. 2001. Dissolved and particulate nutrient flux from three adjacent agricultural watersheds: a five-year study. Biogeochemistry 54:85–114.

Verma, S., M. Markus, and R. A. Cooke. 2012. Development of error correction techniques for nitrate-N load estimation methods. Journal of Hydrology 432–433:12–25.

Worrall, F., M. J. K. Howden, and T. P. Burt. 2013. Assessment of sample frequency bias and precision in fluvial flux calculations – An improved low bias estimation method. Journal of Hydrology 503:101–110.

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

Supporting Information

Additional Supporting Information may be found online at: http://onlinelibrary.wiley.com/doi/10.1002/ecs2.1298/supinfo