SUPPLEMENTAL INFORMATION

S1. Watershed delineation method

Watersheds associated with sampling points were delineated using the approach in the Stream Catchment dataset (Hill et al., 2016) which is based on NHDPlus Version 2 [NHDPlusV2(101)]. NHDPlusV2 defines a local drainage catchment for every stream reach, and the watershed is the combination of the local catchment for a stream reach and all upstream catchments. The local catchment is adjusted to only include the area upstream of the sampling point and then combined with upstream catchments.

S2 Calculating Maximum r^2

N concentrations are temporally variable, and static predictor variables used in the model cannot be expected to account for this variability. The maximum r^2 value for the regression models depends on the amount of variability attributable to temporal variability relative to total variability in the data. The larger the amount of variability attributable to temporal variability, the lower the maximum r^2 possible (Olson and Hawkins, 2013):

Maximum r^2 = Var(Site)/(Var(Site) + Var(Rep)

where Var(Site) is the variance associated with spatial variability (across sites), and Var(Rep) is the within-site variance over time. We used the approach outlined by John Olson (personal communication), to calculate the maximum r^2 values possible for our TN, DIN, and TON models. This approach uses a linear mixed model, with site as a random factor, to isolate the within-site variability, which we obtain from the ~9% of sites that were more than once. The R code and output are below.

Signal-Noise

Calculate the signal to noise ratio for TN, DIN, and DON in NRSA data set.

```
library(lme4)

## Loading required package: Matrix
```

Read in data

Linear mixed models for TN, DIN, TON with site as random factor:

```
# note: Values are log-transformed so results are relevant to regressions, which are also based on log-transformed data.

## TN "Null" model
TN.S2N.null <- lmer(log(TN) ~ 1 + (1|SITE_ID), data=site.chem, REML=T, verbose=F)

TN.Site.var <- as.numeric(VarCorr(TN.S2N.null))
TN.Rep.var <- attr(VarCorr(TN.S2N.null), "sc")^2
TN.SN <- TN.Site.var/TN.Rep.var
TN.maxR2 <- TN.SN/(TN.SN+1)

## DIN "Null" model
DIN.S2N.null <- lmer(log(DIN) ~ 1 + (1|SITE_ID), data=site.chem, REML=T, verbose=F)

DIN.Site.var <- as.numeric(VarCorr(DIN.S2N.null))
DIN.Rep.var <- attr(VarCorr(DIN.S2N.null), "sc")^2
DIN.SN <- DIN.Site.var/DIN.Rep.var
DIN.maxR2 <- DIN.SN/(DIN.SN+1)
```

```
## DON "Null" model
DON.S2N.null <- lmer(log(DON+0.005) ~ 1 + (1|SITE_ID), data=site.chem, REML=T
, verbose=F)

DON.Site.var <- as.numeric(VarCorr(DON.S2N.null))
DON.Rep.var <- attr(VarCorr(DON.S2N.null), "sc")^2
DON.SN <- DON.Site.var/DON.Rep.var
DON.maxR2 <- DON.SN/(DON.SN+1)
```

lmer summaries for each null model, followed by extracted variances, S:N, and max r2 for each form of N:

```
summary(TN.S2N.null)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: log(TN) ~ 1 + (1 | SITE_ID)
##    Data: site.chem
##
## REML criterion at convergence: 6419.2
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -3.5435 -0.2069  0.0029  0.2110  3.9523
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  SITE_ID  (Intercept) 1.1910   1.091
##  Residual             0.1347   0.367
## Number of obs: 2155, groups:  SITE_ID, 1968
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept) -0.52223    0.02589  -20.17
```

```
summary(DIN.S2N.null)
```

```
## Linear mixed model fit by REML ['lmerMod']
## Formula: log(DIN) ~ 1 + (1 | SITE_ID)
##    Data: site.chem
##
## REML criterion at convergence: 8115.9
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.8196 -0.3023 -0.0268  0.2498  5.0606
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  SITE_ID  (Intercept) 2.5687   1.6027
##  Residual             0.3232   0.5685
## Number of obs: 2155, groups:  SITE_ID, 1968
```

```
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept) -1.88982    0.03823  -49.44

summary(DON.S2N.null)

## Linear mixed model fit by REML ['lmerMod']
## Formula: log(DON + 0.005) ~ 1 + (1 | SITE_ID)
##    Data: site.chem
##
## REML criterion at convergence: 6046.1
##
## Scaled residuals:
##     Min      1Q  Median      3Q     Max
## -4.3262 -0.2487  0.0248  0.2734  3.7988
##
## Random effects:
##  Groups   Name        Variance Std.Dev.
##  SITE_ID  (Intercept) 0.8903   0.9436
##  Residual             0.1801   0.4244
## Number of obs: 2155, groups:  SITE_ID, 1968
##
## Fixed effects:
##             Estimate Std. Error t value
## (Intercept) -1.19277    0.02322  -51.36

# table of results
results <- cbind(c(TN.Site.var, TN.Rep.var,TN.SN, TN.maxR2),
                 c(DIN.Site.var, DIN.Rep.var,DIN.SN, DIN.maxR2),
                 c(DON.Site.var, DON.Rep.var,DON.SN, DON.maxR2))
row.names(results) <- c("Site.var", "Rep.va", "S:N", "max R2")
colnames(results) <- c("TN", "DIN", "DON")

results

##                 TN       DIN       DON
## Site.var 1.1910458 2.5686720 0.8903351
## Rep.va   0.1347144 0.3231632 0.1800859
## S:N      8.8412690 7.9485295 4.9439474
## max R2   0.8983871 0.8882498 0.8317616
```

S3. Residual spatial autocorrelation analysis

Stream nutrient concentration measurements from sites with nested watersheds are not truly independent because the sampling sites share drainage areas and stream flow. However, if spatial factors that control the variability in nutrient concentrations are modeled well, the model residuals will not be spatially correlated and coefficients should not be biased (Kuhn and Dormann 2012). The majority of NRSA sites (87%) are nested within other watersheds and/or have sites nested within their watersheds, so we assess the potential effect of this lack of independence on model results.

Because nested sampling sites are physically connected via water flow, flow distance is a more appropriate measure of spatial relatedness than Euclidean distance. Additionally, the larger the fraction of flow at the downstream site that is contributed by the upstream site, the more correlated the sites are likely to be. Therefore, we used semivariograms to qualitatively examine residual variance as a function of flow distance, fraction watershed uniqueness (100 - % overlap)/100, and a combined uniqueness index that is calculated as the product of those two values (with flow distance normalized to the maximum distance). Pairs were binned every 10% decrease in spatial-relatedness measure, so the number of pairs varies from bin to bin. The nugget (0 difference in space) was calculated by using the difference in residuals from site that were visited twice. Increasing semi-variance with decreasing spatial relatedness based on all three measures indicated that there was some residual spatial autocorrelation in the data (Fig. S1), although values in bins with few pairs should be interpreted with caution.

To test whether parameter coefficients from the model based on all data were biased due to the inclusion of nested sampling sites, we subsetted the data into 200 sets of 200 sites with non-nested watersheds: for each subset, a sampling site was randomly selected, any sites connected via upstream or downstream flow were removed from possible selection, then another watershed was randomly selected, and so on, until either 200 sites were obtained, or no more independent watersheds were available. TN, DIN, and TON were modeled, and the average model coefficients from all runs were compared to those obtained when all data are used simultaneously. All average coefficient values, plus or minus 2 standard errors, fell within 95% confidence intervals of coefficients from the model using all data (Figures S2-S4), suggesting that coefficients estimated using all data are not strongly affected by included sampling sites from nested watersheds. Therefore, we report results from the model based on all available sites.

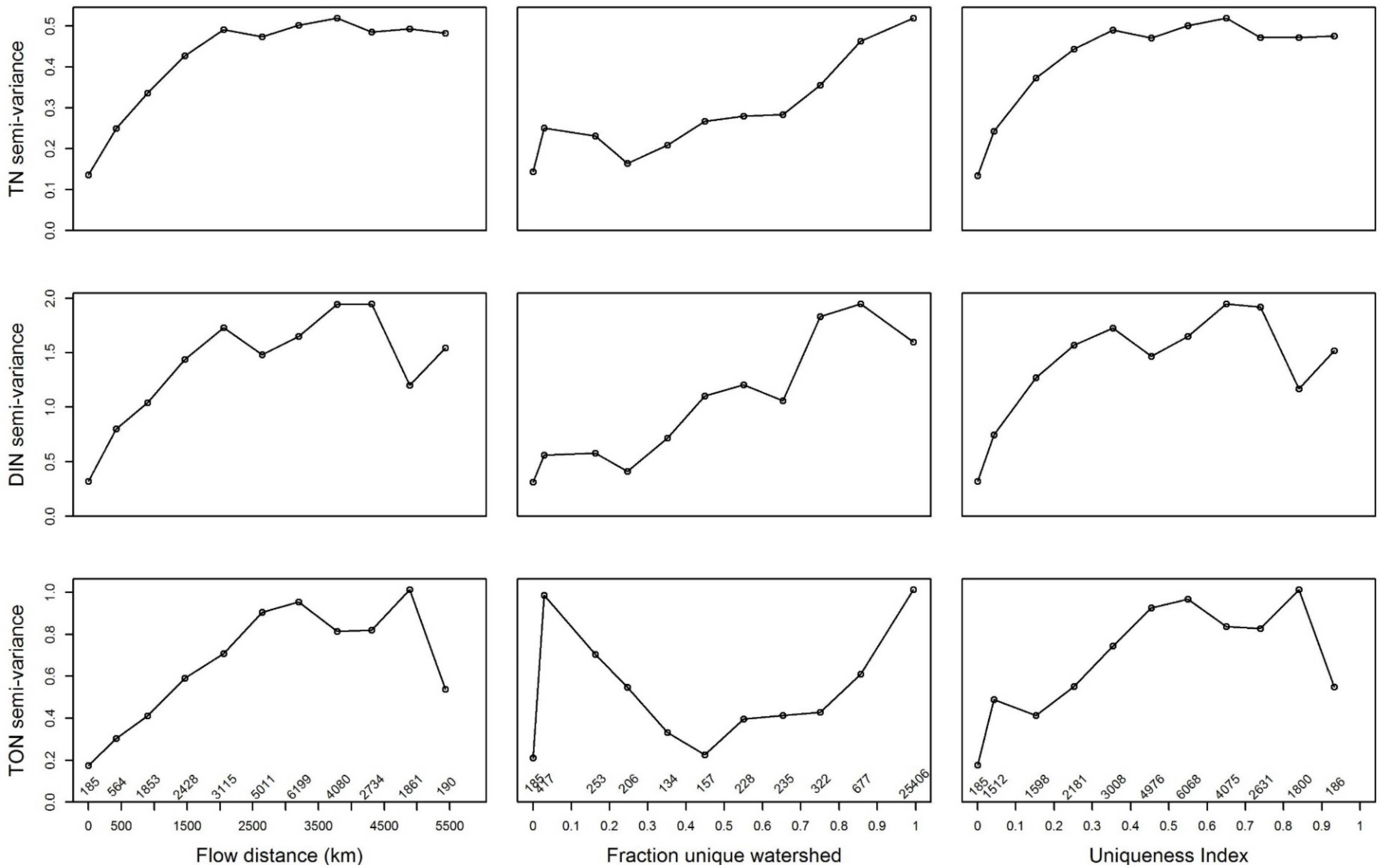

Figure S1. Semi-variograms for residuals of TN (top row), DIN (middle row), and TON (bottom row) models (unstandardized data) based on flow distance between sampling sites (top row), Fraction uniqueness of watersheds [(100 - % overlap of smaller in larger)/100], and a uniqueness index that is the product of flow distance normalized to the maximum distance and fraction uniqueness (see text for more explanation). The value at 0 (nugget) reflects semi-variance across time in sites with two visits. Values just above the x-axis in the bottom row correspond to the number of pairs of watersheds in each bin.

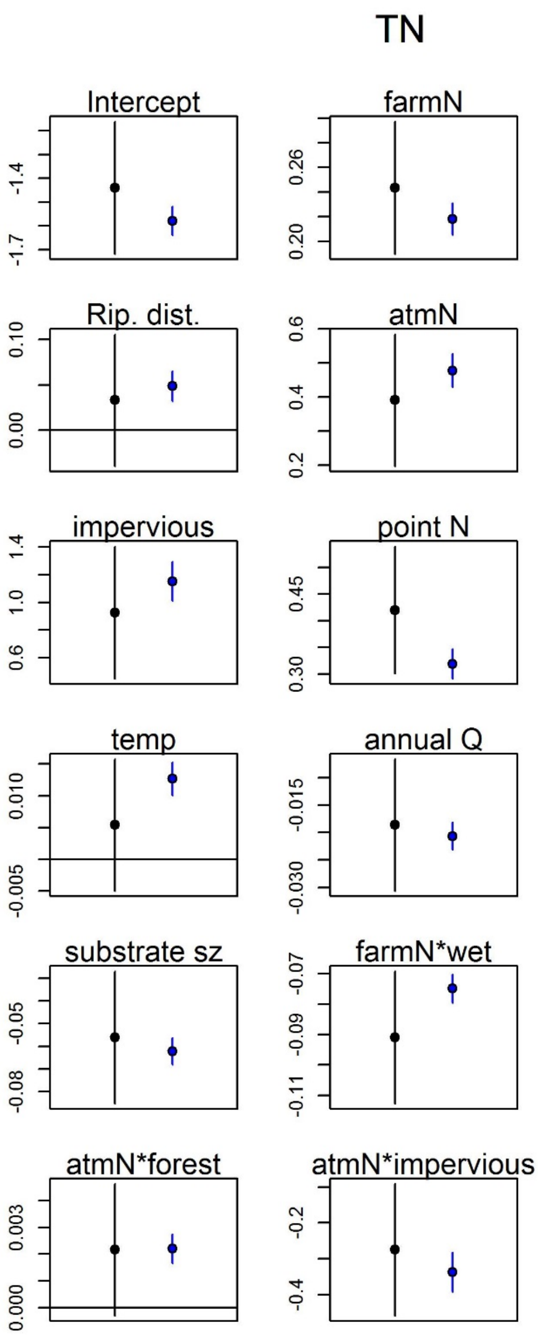

Figure S2. Comparison of the coefficient values from the TN model based on all sampling sites (black) and the average of coefficient values from the 200 model runs, each with an different set of truly independent watersheds (blue). Vertical lines represent +/- 2 standard errors.

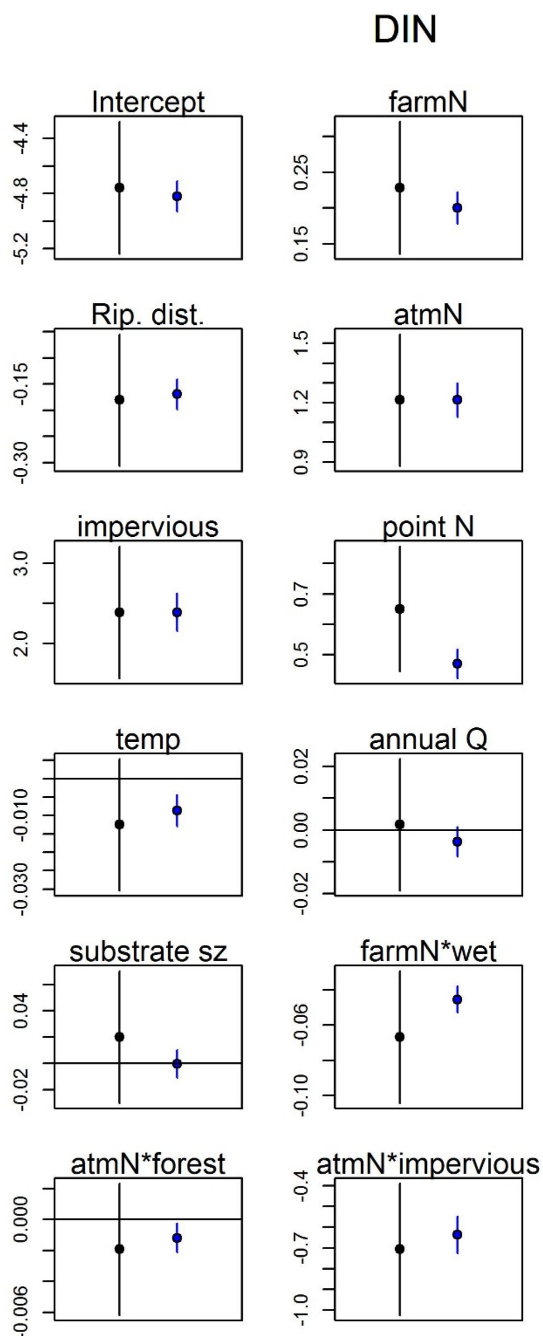

Figure S3. Comparison of the coefficient values from the DIN model based on all sampling sites (black) and the average of coefficient values from the 200 model runs, each with an different set of truly independent watersheds (blue). Vertical lines represent +/- 2 standard errors.

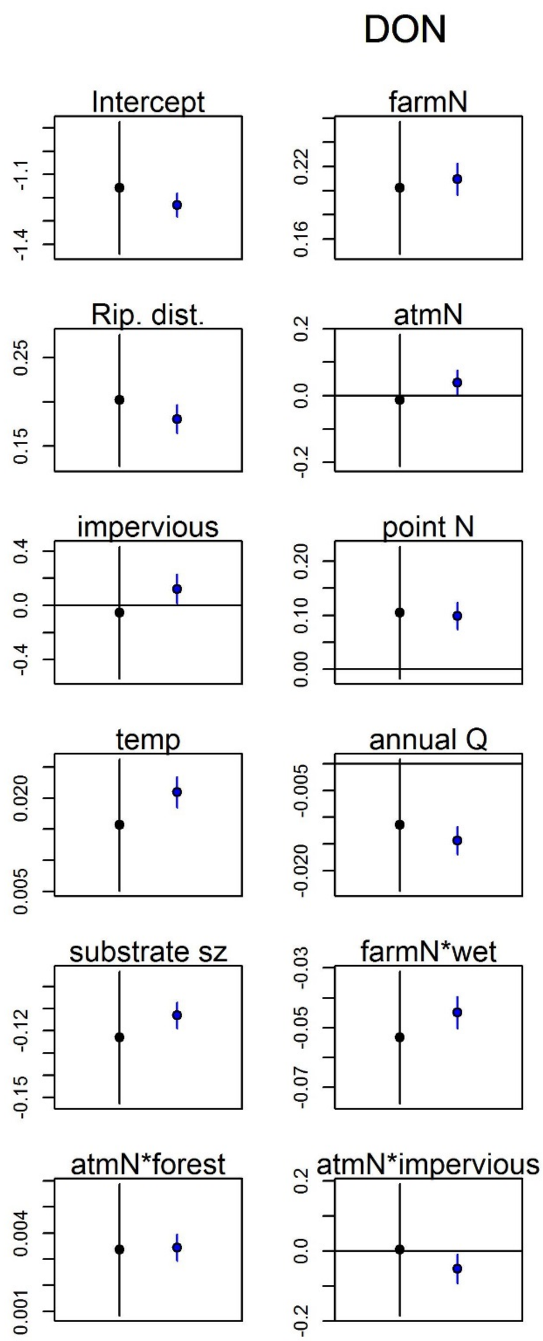

Figure S4. Comparison of the coefficient values from the TON model based on all sampling sites (black) and the average of coefficient values from the 200 model runs, each with an different set of truly independent watersheds (blue). Vertical lines represent +/- 2 standard errors.

S4. Model Results

Table S1. Parameter coefficients and AIC values for the top TN model, and models within 2 AIC of the top model.

Intercept	% Forest	Atm N dep	Agricultural N	% Impervious	Point source N	Discharge	% Riparian disturbance	% Wetland	% Open water	Precipitation	Substrate Size	Temperature	Atm N * % Forest	Atm N * % Impervious	Agricultural N * % Riparian disturbance	Agricultural N * % Wetland	AIC
-1.44	-0.013	0.42	0.24	0.95	0.42	-0.018	0.026	0.22	NA	-0.00047	-0.059	NA	0.0021	-0.28	0.023	-0.089	4326.58
-1.62	-0.010	0.54	0.24	1.00	0.43	-0.018	0.036	0.21	NA	-0.00049	-0.059	NA	NA	-0.31	0.017	-0.088	4327.47
-1.44	-0.013	0.40	0.24	0.91	0.42	-0.018	0.030	0.22	NA	-0.00050	-0.056	0.0046	0.0021	-0.27	0.025	-0.091	4327.78
-1.44	-0.013	0.42	0.24	0.95	0.42	-0.018	0.026	0.21	0.00224	-0.00047	-0.059	NA	0.0022	-0.28	0.023	-0.088	4328.52

Table S2. Parameter coefficients and AIC values for the top DIN model, and models within to AIC of the top model.

Intercept	% Forest	Atm N dep	Agricultural N	% Impervious	Point source N	Discharge	% Riparian disturbance	% Wetland	% Open water	Precipitation	Substrate Size	Temperature	Atm N * % Forest	Atm N * % Impervious	Agricultural N * % Riparian disturbance	Agricultural N * % Wetland	AIC
-4.57	-0.0038	1.11	0.23	2.29	0.64	NA	-0.19	-0.012	-0.047	-0.00016	NA	-0.015	NA	-0.66	0.087	-0.064	6453.53
-4.59	-0.0052	1.11	0.23	2.32	0.65	NA	-0.18	-0.021	-0.047	NA	NA	-0.020	NA	-0.68	0.083	-0.063	6453.70
-4.56	-0.0026	1.02	0.22	2.15	0.64	NA	-0.18	0.003	-0.041	-0.00023	NA	NA	NA	-0.60	0.092	-0.070	6454.28
-4.73	-0.0005	1.21	0.23	2.33	0.66	NA	-0.18	-0.013	-0.047	-0.00017	NA	-0.015	-0.00183	-0.68	0.082	-0.064	6454.82
-4.58	-0.0040	1.11	0.23	2.30	0.64	NA	-0.19	-0.005	-0.046	-0.00017	0.020	-0.013	NA	-0.67	0.088	-0.064	6454.88
-4.57	-0.0031	1.03	0.22	2.19	0.64	NA	-0.18	0.011	-0.041	-0.00023	0.028	NA	NA	-0.62	0.092	-0.070	6455.01
-4.54	NA	0.95	0.25	2.04	0.62	NA	-0.18	0.017	-0.044	-0.00031	NA	NA	NA	-0.55	0.095	-0.072	6455.05
-4.60	-0.0054	1.11	0.22	2.34	0.65	NA	-0.18	-0.015	-0.046	NA	0.016	-0.019	NA	-0.69	0.083	-0.063	6455.30
-4.71	-0.0029	1.18	0.23	2.36	0.66	NA	-0.17	-0.022	-0.047	NA	NA	-0.020	-0.00133	-0.70	0.079	-0.063	6455.31
-4.57	-0.0038	1.11	0.23	2.28	0.64	0.0030	-0.19	-0.013	-0.047	-0.00016	NA	-0.015	NA	-0.66	0.088	-0.064	6455.44

Table S3. Parameter coefficients and AIC values for the top TON model, and models within to AIC of the top model.

Intercept	% Forest	Atm N dep	Agricultural N	% Impervious	Point source N	Discharge	% Riparian disturbance	% Wetland	% Open water	Precipitation	Substrate Size	Temperature	Atm N * % Forest	Atm N * % Impervious	Agricultural N * % Riparian disturbance	Agricultural N * % Wetland	AIC
-1.12	-0.015	-0.040	0.20	NA	0.082	-0.011	0.18	0.26	0.046	-0.00062	-0.12	0.016	0.0036	NA	-0.034	-0.054	4411.33
-1.14	-0.015	-0.030	0.20	NA	NA	-0.009	0.19	0.26	0.046	-0.00061	-0.12	0.015	0.0036	NA	-0.035	-0.053	4411.49
-1.15	-0.015	-0.005	0.19	NA	NA	NA	0.18	0.26	0.044	-0.00062	-0.13	0.015	0.0035	NA	-0.033	-0.053	4411.62
-1.16	-0.014	-0.011	0.20	-0.043	0.106	-0.011	0.20	0.26	0.046	-0.00061	-0.12	0.016	0.0033	NA	-0.038	-0.053	4412.58
-1.14	-0.015	-0.007	0.19	NA	0.054	NA	0.17	0.26	0.043	-0.00062	-0.13	0.015	0.0035	NA	-0.032	-0.053	4412.61

Table S4. Results of the final mutiple regression TN models based on forward and backward stepwise regression, with selection based on the lowest AIC values. The models use either land cover variables (%wetland, %forest, %impervious) and N input variables corresponding to the whole watershed or 100 m riparian buffer. Coefficients are unstandardized. Partial r^2 value for each variable was calculated as the average r^2 contribution over all orderings of regressors, based on the method from Lindeman, Merenda, and Gold (1980) (R package relaimpo, "lmg" approach).

	Watershed predictor variables				Riparian buffer predictor variables			
	coefficient	SE	p	partial r^2	coefficient	SE	p	partial r^2
Intercept	-1.440	0.142	<0.001	NA	-1.330	0.122	<0.001	NA
Agricultural N	0.244	0.027	<0.001	0.160	0.128	0.014	<0.001	0.159
Wetland	0.216	0.028	<0.001	0.006	0.124	0.024	<0.001	0.005
Riparian disturbance	0.026	0.037	0.478	0.115	0.013	0.044	0.705	0.124
Atm N deposition	0.422	0.094	<0.001	0.072	0.345	0.083	<0.001	0.069
% Forest	-0.013	0.002	<0.001	0.120	-0.014	0.003	<0.001	0.121
% Impervious	0.953	0.239	<0.001	0.038	0.209	0.056	0.318	0.017
Point source N	0.420	0.061	<0.001	0.020	0.476	0.061	<0.001	0.024
Precipitation	-0.00047	0.00006	<0.001	0.028	-0.00048	0.00007	<0.001	0.025
Temperature					0.015	0.005	0.005	0.008
Discharge	-0.018	0.006	0.003	0.002	-0.017	0.006	0.008	0.002
% Open water					0.016	0.010	0.101	0.002
Substrate size	-0.059	0.015	<0.001	0.027	-0.059	0.015	<0.001	0.027
Agricultural N X % wetland	-0.089	0.011	<0.001	0.015	-0.031	0.005	<0.001	0.010
Agricultural N X % riparian disturbance	0.023	0.010	0.020	0.001	0.013	0.006	0.016	0.001
Atm N dep X % forest	0.002	0.001	0.090	0.001	0.002	0.001	0.097	0.001
Atm N dep X % impervious	-0.285	0.093	0.002	0.003				
Model	$r^2 = 0.607$, adj $r^2 = 0.604$, $p < 0.001$				$r^2 = 0.597$, adj $r^2 = 0.593$, $p < 0.001$			

Table S5. Results of the final mutiple regression DIN models based on forward and backward stepwise regression, with selection based on the lowest AIC values. The models use either land cover variables (%wetland, %forest, %impervious) and N input variables corresponding to the whole watershed or 100 m riparian buffer. Coefficients are unstandardized. Partial r^2 value for each variable was calculated as the average r^2 contribution over all orderings of regressors, based on the method from Lindeman, Merenda, and Gold (1980) (R package relaimpo, "lmg" approach).

	Watershed predictor variables				Riparian buffer predictor variables			
	coefficient	SE	p	lmg	coefficient	SE	p	lmg
Intercept	-4.569	0.155	<0.001	NA	-4.739	0.240	<0.001	NA
Agricultural N	0.227	0.046	<0.001	0.136	0.121	0.023	0.000	0.120
Wetland	-0.012	0.052	0.813	0.005	0.008	0.040	0.838	0.003
Riparian disturbance	-0.192	0.063	0.002	0.088	-0.226	0.075	0.003	0.084
Atm N deposition	1.108	0.119	<0.001	0.102	1.235	0.153	<0.001	0.128
% Forest	-0.004	0.002	0.028	0.016	0.000	0.004	0.953	0.020
% Impervious	2.289	0.413	<0.001	0.063	1.452	0.355	<0.001	0.048
Point source N	0.643	0.100	<0.001	0.030	0.749	0.101	<0.001	0.032
Precipitation	-0.00016	0.00110	0.142	0.005	-0.00020	0.00010	0.052	0.005
Temperature	-0.015	0.009	0.099	0.009				
Discharge								
% Open water	0.047	0.016	0.005	0.002	-0.050	0.016	0.002	0.002
Substrate size								
Agricultural N X % wetland	-0.064	0.019	0.001	0.004	-0.024	0.008	0.002	0.003
Agricultural N X % riparian disturbance	0.087	0.016	<0.001	0.007	0.044	0.009	<0.001	0.007
Atm N dep X % forest					-0.003	0.002	0.143	0.001
Atm N dep X % impervious	-0.659	0.159	<0.001	0.004	-0.333	0.142	0.019	0.001
Model	r^2 = 0.470, adj r^2 = 0.467, p < 0.001				r^2 = 0.456, adj r^2 = 0.452, p < 0.001			

Table S6. Results of the final mutiple regression TON models based on forward and backward stepwise regression, with selection based on the lowest AIC values. The models use either land cover variables (%wetland, %forest, %impervious) and N input variables corresponding to the whole watershed or 100 m riparian buffer. Coefficients are unstandardized. Partial r^2 value for each variable was calculated as the average r^2 contribution over all orderings of regressors, based on the method from Lindeman, Merenda, and Gold (1980) (R package relaimpo, “lmg” approach).

	Watershed predictor variables				Riparian buffer predictor variables			
	coefficient	SE	p	lmg	coefficient	SE	p	lmg
Intercept	-1.122	0.117	<0.001	NA	-1.036	0.141	<0.001	NA
Agricultural N	0.204	0.027	<0.001	0.081	0.109	0.014	<0.001	0.078
Wetland	0.265	0.031	<0.001	0.041	0.156	0.025	<0.001	0.031
Riparian disturbance	0.183	0.032	0.000	0.065	0.178	0.044	<0.001	0.063
Atm N deposition	-0.040	0.080	0.616	0.017	-0.070	0.095	0.459	0.019
% Forest	-0.015	0.003	<0.001	0.129	-0.016	0.003	<0.001	0.148
% Impervious					-0.467	0.209	0.026	0.003
Point source N	0.082	0.056	0.143	0.002	0.118	0.061	0.054	0.002
Precipitation	-0.0006	0.0001	<0.001	0.055	-0.0006	0.0001	<0.001	0.050
Temperature	0.016	0.005	0.004	0.005	0.017	0.005	0.002	0.005
Discharge	-0.011	0.006	0.071	0.001	-0.012	0.006	0.055	0.002
% Open water	0.046	0.010	<0.001	0.014	0.060	0.010	<0.001	0.016
Substrate size	-0.123	0.015	<0.001	0.061	-0.123	0.015	<0.001	0.059
Agricultural N X % wetland	-0.054	0.011	<0.001	0.007	-0.018	0.005	<0.001	0.005
Agricultural N X % riparian disturbance	-0.034	0.009	<0.001	0.006	-0.017	0.006	0.002	0.005
Atm N dep X % forest	0.004	0.001	0.003	0.005	0.004	0.001	0.005	0.004
Atm N dep X % impervious					0.165	0.084	0.050	0.001
Model	$r^2 = 0.489$, adj $r^2 = 0.485$, $p < 0.001$				$r^2 = 0.491$, adj $r^2 = 0.487$, $p < 0.001$			

S5. Correlations among predictor variables

Figure S5. Correlations among N concentrations and predictor variables. The panels in the lower left contain r values corresponding to relationships shown in the upper right. Variables are, from top left to bottom right, agricultural N inputs, atmospheric N deposition, point source N, % riparian disturbance, % wetland, % impervious cover, % forest, substrate size, % open water, annual mean discharge, precipitation, and temperature. R values larger than 0.4 are shown with proportionally larger text. Values are transformed as indicated in Table 1 of the manuscript.

Table S4. P-values corresponding to linear correlations among variables shown in figure S5. Values less than 0.05 are shown in bold

	Agricultural N	Atmospheric N deposition	Point source N	Riparian disturbance	Wetland cover	Impervious cover	Forest cover	Substrate size	Open water	Annual mean	Mean annual Precipitation	Mean annual
Agricultural N	NA											
Atmospheric N deposition	**<0.001**	NA										
Point Source N	**<0.001**	**<0.001**	NA									
Riparian disturbance	**<0.001**	**<0.001**	**<0.001**	NA								
Wetland cover	**<0.001**	**<0.001**	**<0.001**	**<0.001**	NA							
Impervious cover	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	NA						
Forest cover	**<0.001**	0.047	0.599	**<0.001**	**0.001**	**<0.001**	NA					
Substrate size	**<0.001**	**<0.001**	0.664	**<0.001**	**<0.001**	**0.004**	**<0.001**	NA				
Open water	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	0.978	**<0.001**	NA			
Annual mean discharge	**<0.001**	0.640	**<0.001**	**<0.001**	**<0.001**	**<0.001**	0.596	**<0.001**	**<0.001**	NA		
Mean annual precipitation	0.205	**<0.001**	**<0.001**	**0.008**	**<0.001**	**<0.001**	**<0.001**	**<0.001**	**0.019**	0.043	NA	
Mean annual temperature	**<0.001**	**<0.001**	**<0.001**	**0.001**	0.322	**<0.001**	0.151	**<0.001**	**<0.001**	0.087	**<0.001**	NA