AGU PUBLICATIONS

Water Resources Research

Supporting Information for

Leveraging Spatial and Temporal Variability to Probabilistically Characterize Nutrient Sources and Export Rates in a Developing Watershed

H. L. Strickling and D. R. Obenour

Department of Civil, Construction, & Environmental Engineering, North Carolina State University, Raleigh, NC 27695, USA.

Contents of this file

Page

Introduction

This Supporting Information document presents land use and point source input data (Table S1, Figures S1 and S2), average annual precipitation (Figure S3), WRTDS TN loading estimates (Figures S4 to S7), the procedure for determining load estimate uncertainty (Text S1), model parameter analyses (Tables S2 to S3) and results summary figures (Figures S8 to S13). The RStan model code is provided as Text S2.

Table S1. NWALT land-use classifications with their respective areas (ha) and % of total area in all three river basins (Falcone, 2015).

Land Use Classification	Area (hectares) 1992	2002	2012	% of total area 1992	2002	2012
Water						
Water	21548	21548	21960	1.2	1.2	1.2
Wetlands	130783	130783	130783	7.3	7.3	7.3
Developed						
Major Transportation	26197	28453	30156	1.5	1.6	1.7
Commercial/Services	30030	35111	38592	1.7	2.0	2.2
Industrial/Military	10842	12392	13884	0.6	0.7	0.8
Recreation	6997	8049	8669	0.4	0.5	0.5
Residential, High Density	14185	20450	26935	0.8	1.1	1.5
Residential, Low-Medium Density	64792	75844	90279	3.6	4.3	5.1
Developed, Other	22929	17494	12631	1.3	1.0	0.7
Semi-Developed						
Urban Interface High	25584	23495	21489	1.4	1.3	1.2
Urban Interface Low Medium	129582	217673	267704	7.3	12.2	15.0
Anthropogenic Other	599	531	448	0.0	0.0	0.0
Production						
Mining/Extraction	1031	1060	1136	0.1	0.1	0.1
Crops	335551	318468	318840	18.9	17.9	17.9
Pasture/Hay	165608	165823	156385	9.3	9.3	8.8
Grazing Potential	8403	8258	8363	0.5	0.5	0.5
Low Use						
Low Use	784342	693572	630749	44.1	39.0	35.4
Very Low Use, Conservation						
Very Low Use, Conservation	375	375	375	0.0	0.0	0.0

Table S2. Probabilistic comparison of parameter magnitudes; table values represent the percent probability of row parameter exceeding column parameter.

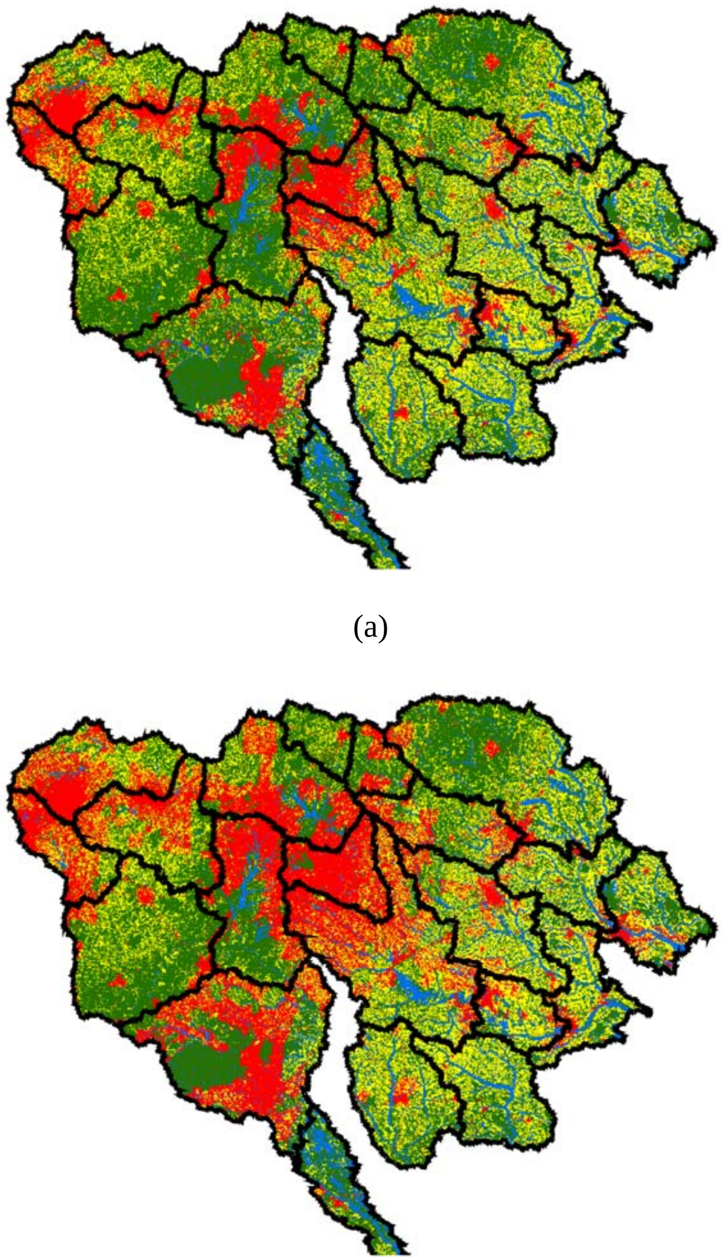
EC	$\beta_{e,a}$	$\beta_{e,d}$	$\beta_{e,w}$
$\beta_{e,a}$	—	87	90
$\beta_{e,d}$	13	—	54
$\beta_{e,w}$	10	46	—

DC	$\beta_{d,c}$	$\beta_{d,h}$
$\beta_{d,c}$	—	57
$\beta_{d,h}$	43	—

PIC	$\beta_{p,a}$	$\beta_{p,d}$	$\beta_{p,w}$	$\beta_{p,c}$	$\beta_{p,h}$
$\beta_{p,a}$	—	80	62	83	65
$\beta_{p,d}$	20	—	30	53	34
$\beta_{p,w}$	38	70	—	70	55
$\beta_{p,c}$	17	47	30	—	34
$\beta_{p,h}$	35	66	45	66	—

Table S3. Mean parameter estimates with 95% credible intervals for the spatial-only model with and without informative priors on EC and DC. Mean parameter estimates are also shown for each fold of the cross validation. Parameter units are provided in Table 1.

Symbol	Full model with all priors		Full model without EC & DC priors		Validation		
					TP & N	N & CF	CF & TP
	Mean	95% CI	Mean	95% CI	Mean	Mean	Mean
$\beta_{e,a}$	7.4	2.2-13	9.7	2.0-17	7.4	5.6	7.6
$\beta_{e,d}$	5.1	1.1-9.8	5.8	0.35-15	6.3	5.2	5.7
$\beta_{e,w}$	2.3	0.19-5.3	4.4	0.33-11	3.3	2.0	2.3
$\beta_{d,c}$	0.04	0.0018-0.13	0.05	0.0013-0.16	0.14	0.050	0.050
$\beta_{d,h}$	0.03	0.0012-0.081	0.03	0.00078-0.090	0.070	0.030	0.040
$\beta_{d,p}$	1.0	0.76-1.2	1.1	0.20-2.2	1.0	0.99	0.99
ρ_s	17	2.3-43	32	6.2-58	33	16	21
ρ_{wb}	14	6.7-22	37	15-58	8.5	14	13
σ_{res}	0.08	0.023-0.18	0.10	0.029-0.20	0.070	0.10	0.13
σ_α	1.30	0.14-2.7	1.2	0.13-2.9	1.0	1.5	1.8
σ_p	0.23	0.012-0.9	0.21	0.0067-0.81	0.18	0.26	0.33



(a)

(b)

Figure S1. Agriculture (yellow), developed (red), wild (green), and water/wetlands (blue) land-use in all three river basins for 1992 (a) and 2012 (b).

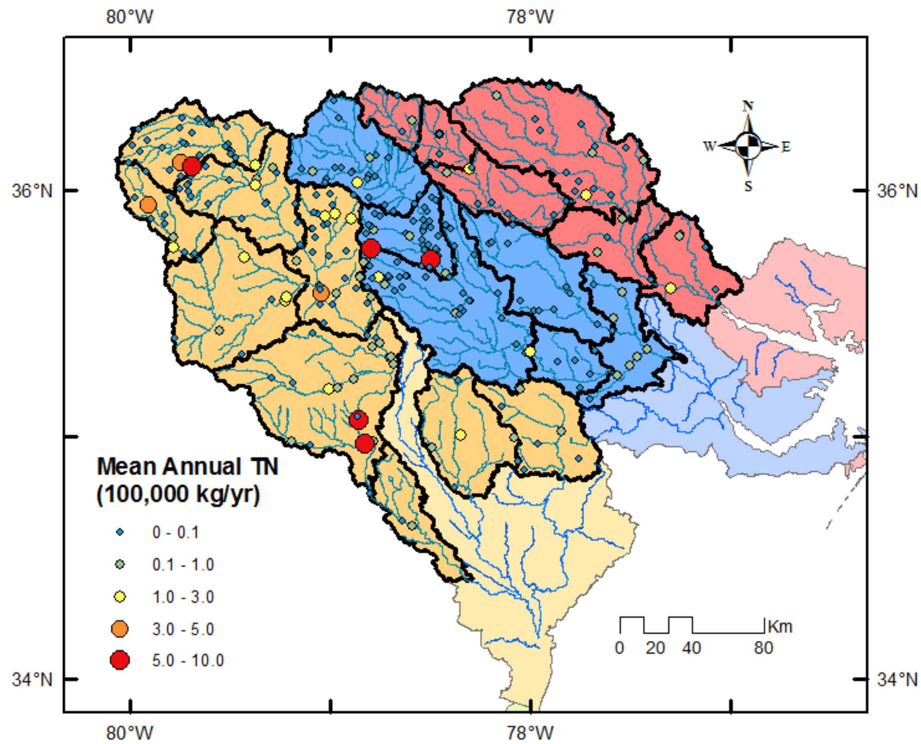


Figure S2. Mean annual TN loads from point source dischargers from 1994 to 2012.

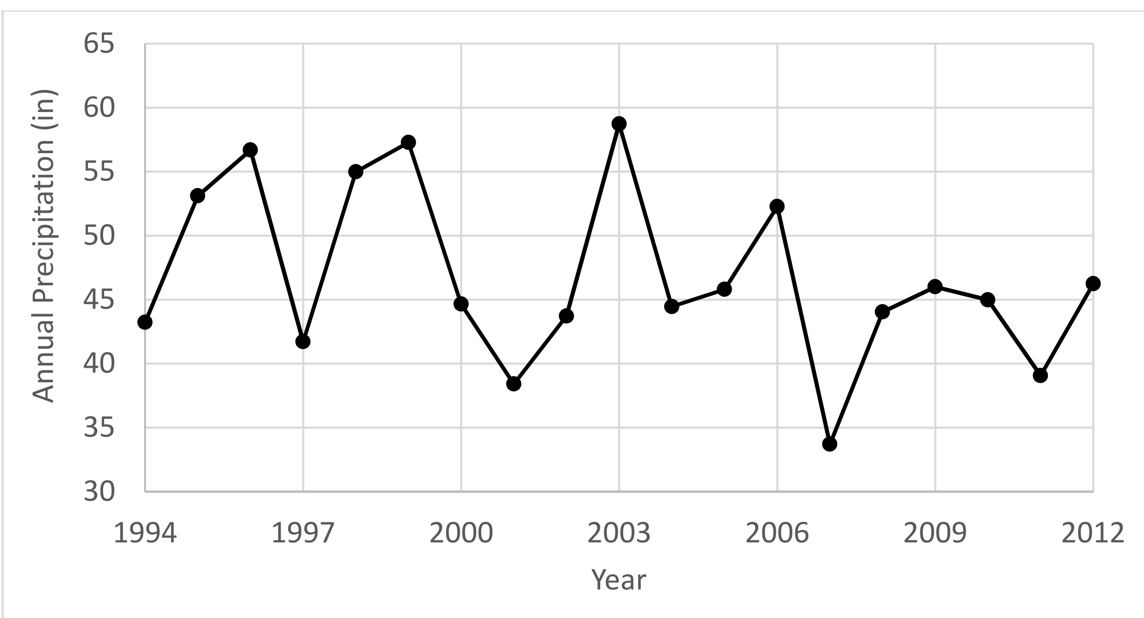


Figure S3. Average annual precipitation over the study area from 1994 to 2012.

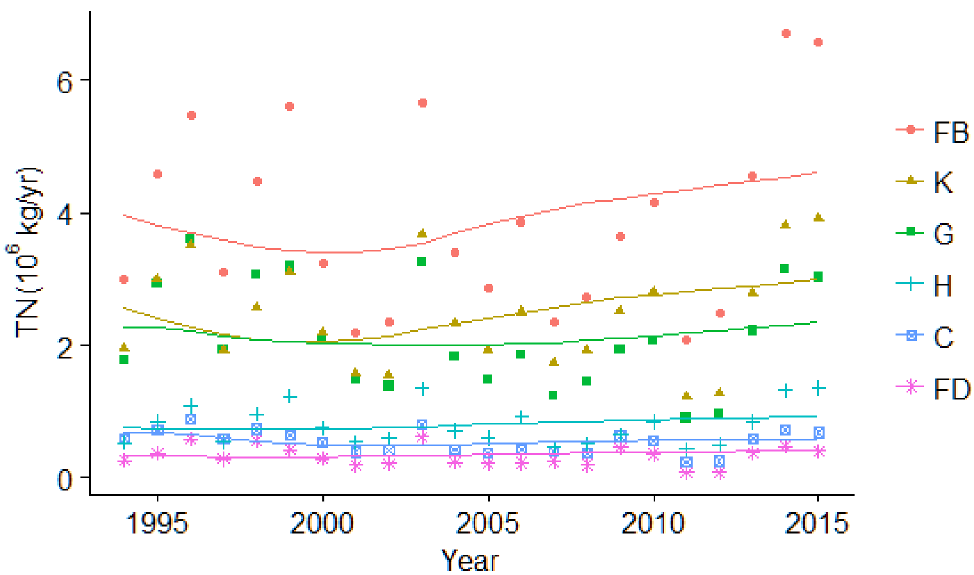


Figure S4. WRTDS TN estimates for stations along the Neuse River for 1994 to 2015. Stations include Falls Dam (FD), Clayton (C), Goldsboro (G), Kinston (K), and Ft. Barnwell (FB). Points are annual estimates and lines are flow-normalized estimates.

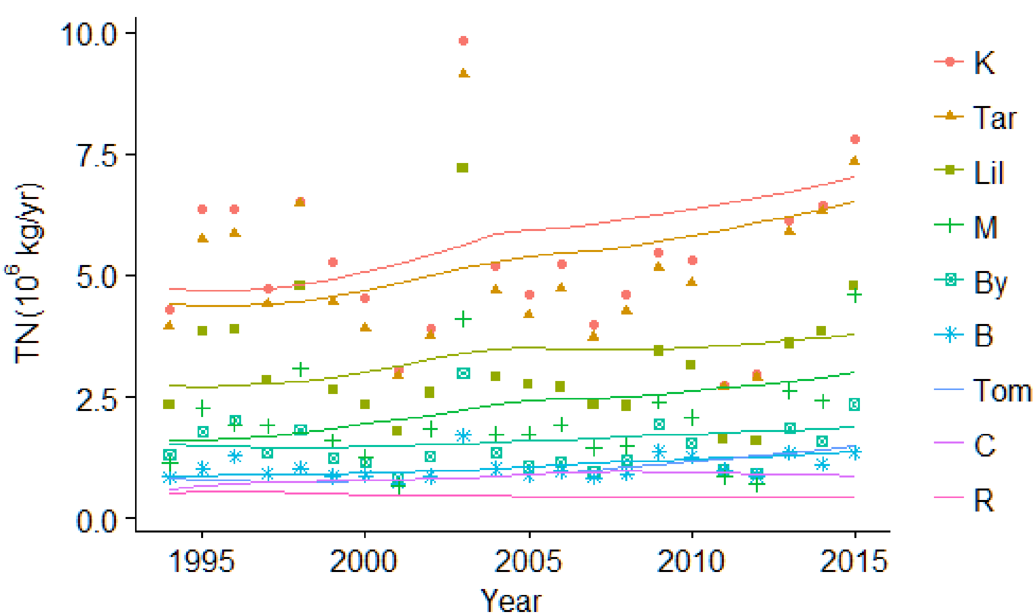


Figure S5. WRTDS TN estimates for stations in the CFRB for 1994 to 2015. Stations include Kelly (K), Tarheel (Tar), Lillington (Lil), Moncure (M), Bynum (By), Burlington (B), Tomahawk (Tom), Chinquapin (C), and Ramseur (R). Points are annual estimates and lines are flow-normalized estimates.

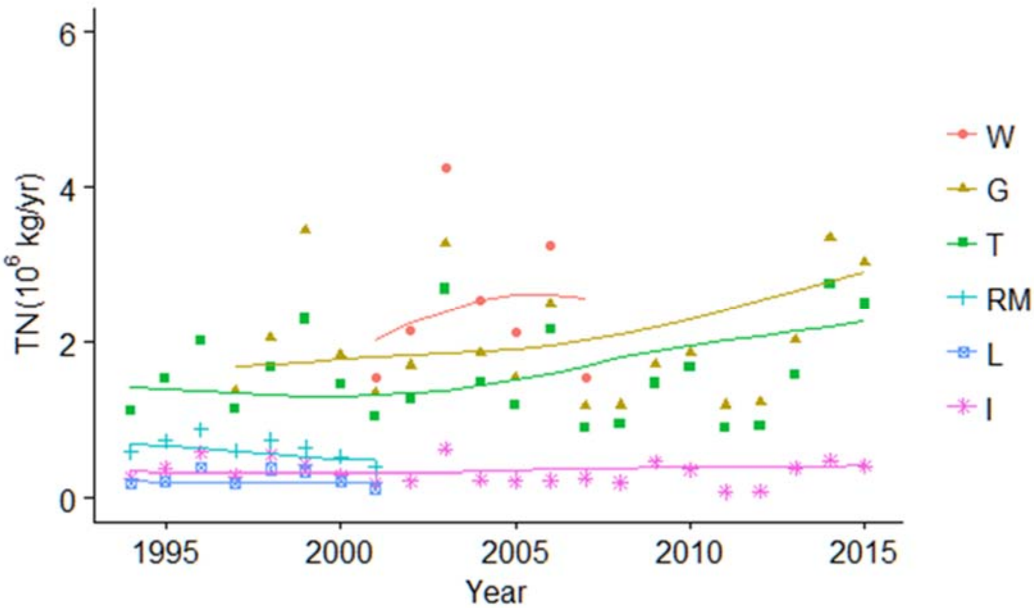


Figure S6. WRTDS TN estimates for stations in the TPRB for 1994 to 2015. Stations include Washington (W), Tarboro (T), Rocky Mount (RM), Louisburg (L), and I96 (I). Points are annual estimates and lines are flow-normalized estimates.

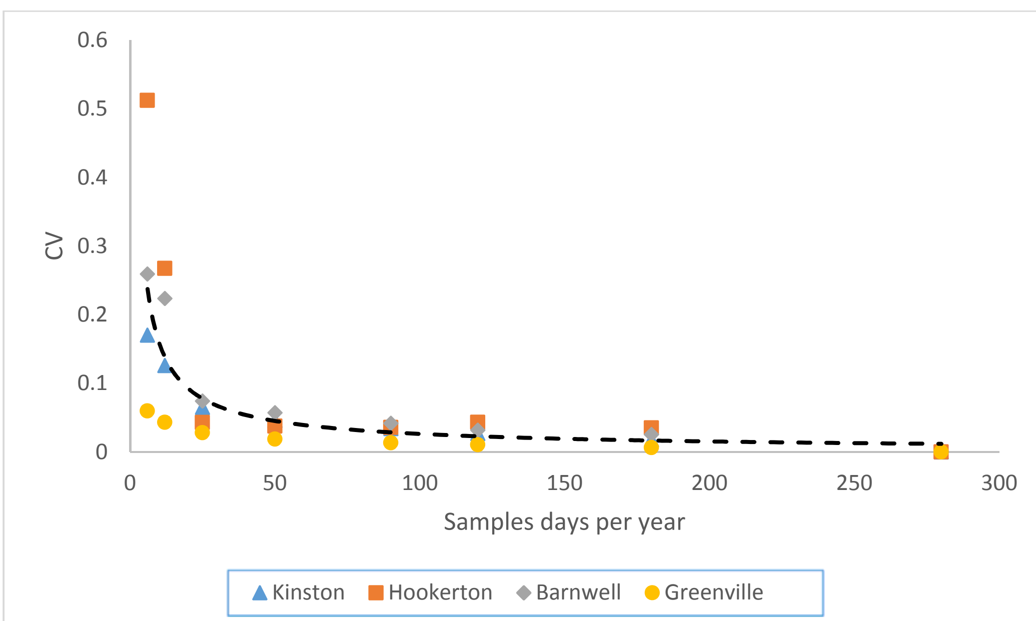


Figure S7. Coefficient of Variation (CV) values for the 8 sampling schedules used in the WRTDS uncertainty analysis by station location. Dashed line is the fit power regression used in this study.

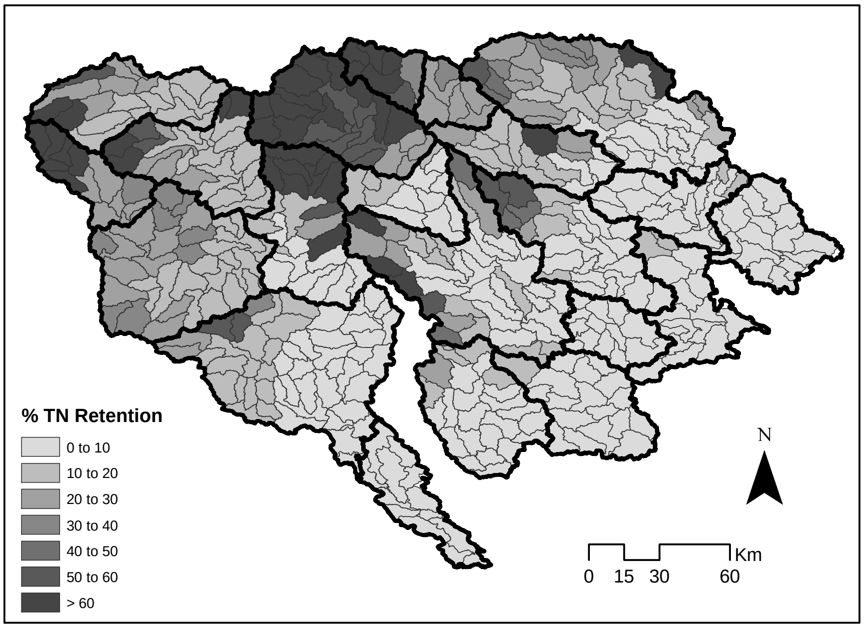


Figure S8. Percent TN retention from each HUC12 to downstream points for each basin (Kelly for all CFRB watersheds except Chinquapin and Tomahawk, Ft. Barnwell for NRB, and Washington for TPRB).

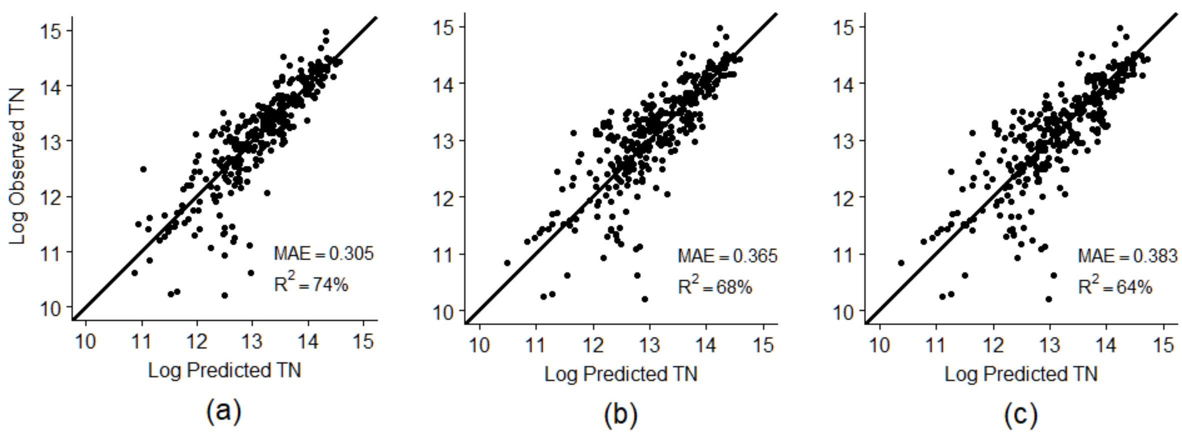


Figure S9. Log observed vs. predicted plots of incremental TN loadings using (a) full model with fixed and random effects, (b) full model with fixed effects only, and (c) cross validation with fixed effects only.

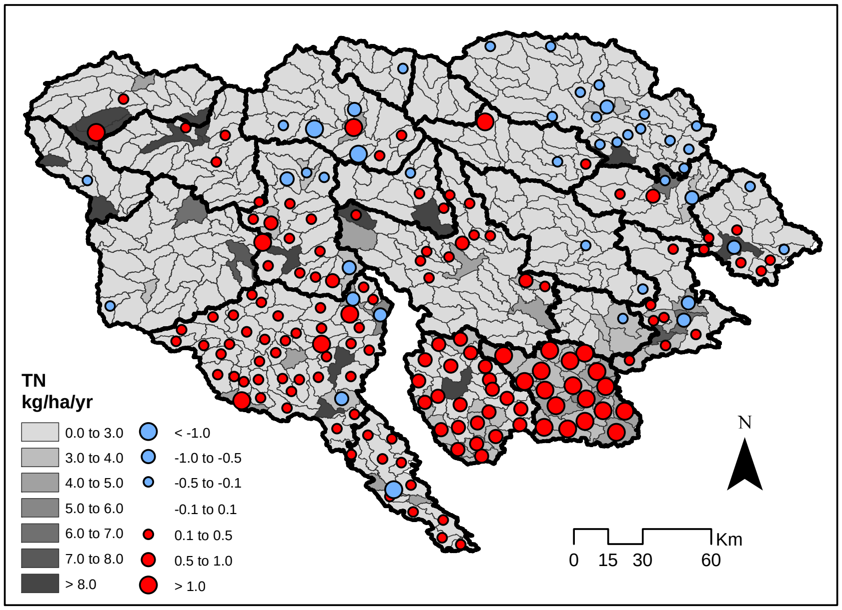


Figure S10. TN loading rates (kg/ha/yr) using 5th percentile precipitation values for each HUC 12. Loading rates for the year 2002 are in grey shading and points reflect change in rates from the year 1994 to 2012. Stream and reservoir retention is included and rates reflect loadings to downstream locations for each basin (Kelly for CFRB, Ft. Barnwell for NRB, and Washington for TPRB).

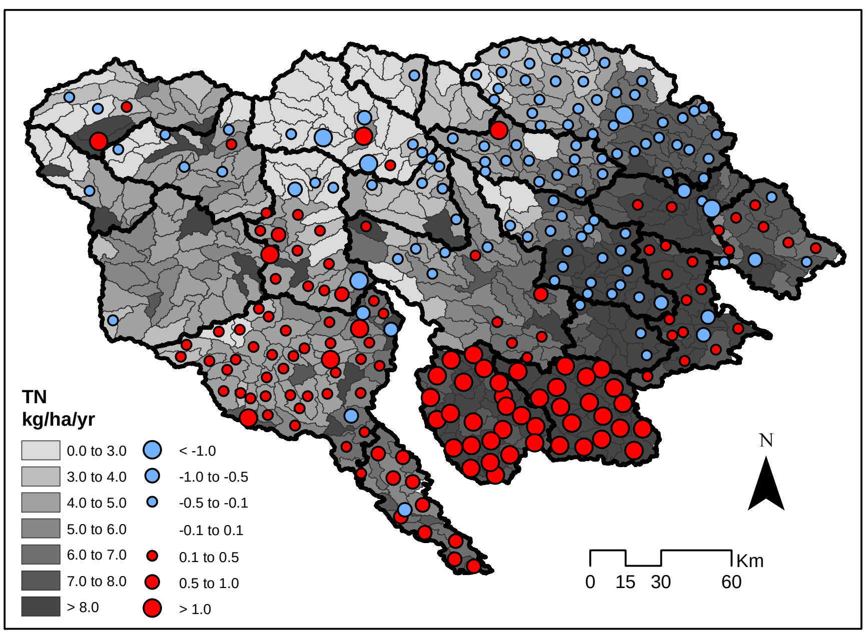


Figure S11. TN loading rates (kg/ha/yr) using 95th percentile precipitation values for each HUC 12. Loading rates for the year 2002 are in grey shading and points reflect change in rates from the year 1994 to 2012. Stream and reservoir retention is included and rates reflect loadings to downstream locations for each basin (Kelly for CFRB, Ft. Barnwell for NRB, and Washington for TPRB).

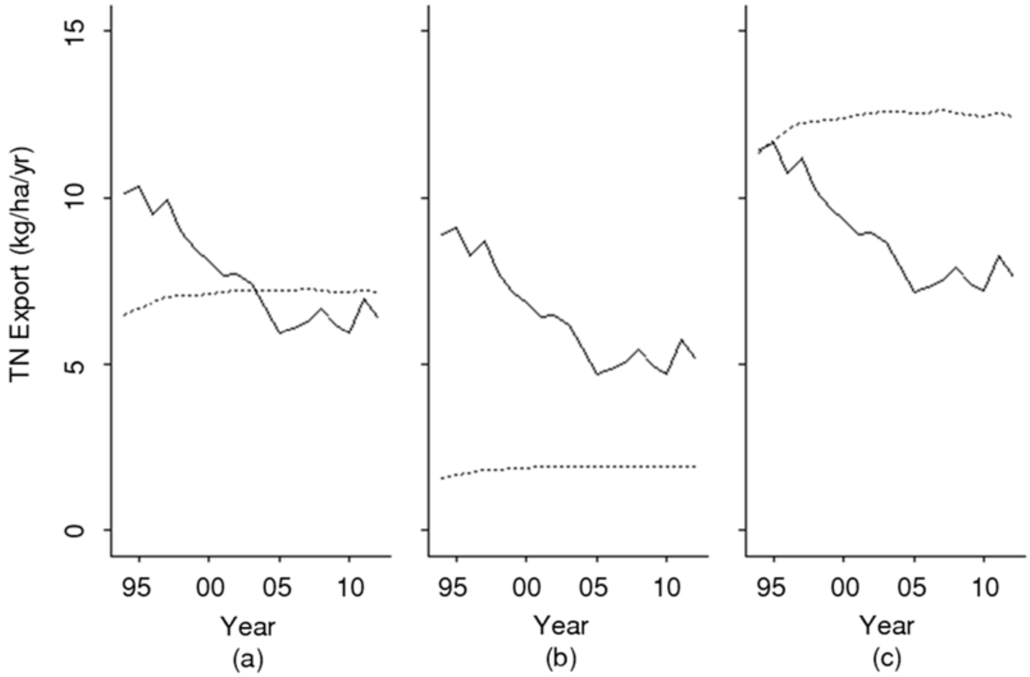


Figure S12. Effective total developed (solid) and total agriculture (dashed) export rates (kg/ha/yr) for (a) mean, (b) 5th, and (c) 95th percentile precipitation for all three basins combined.

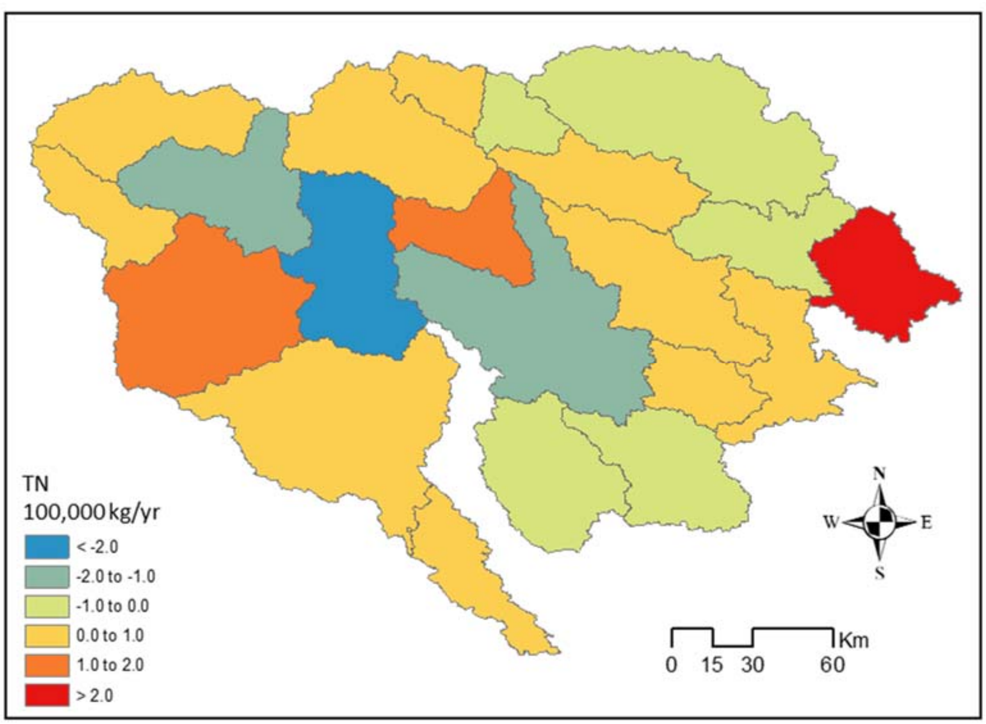


Figure S13. Values of random intercept term α by watershed.

Text S1. Load Estimation Uncertainty

Uncertainty in WRTDS estimates was determined by relating error variance to sampling frequency at monitoring stations. This relationship was developed by sub-sampling at four stations, selected for having nitrogen concentration samples taken nearly every day over a span of at least 7 years. Subsampling of concentration data was implemented based on different sampling schedules: samples taken every two months, 1 month, 2 weeks, 1 week, and 4, 3, and 2 days. Samples were selected at random within a sliding window corresponding to each sampling schedule to simulate real-life variability in sample collection. For each sampling schedule, WRTDS was run 300 times (simulations) with a different randomly selected subsample used for each simulation. A 3-year half-window width was used in the WRTDS estimation, and annual loading estimates were taken from year(s) within 3 years from the start and finish of the simulation. The window width refers to the time-span for the distribution of weights used for the WRTDS regression analysis, where data outside this window will be assigned a weight nearly zero (Hirsch et al., 2010). Residuals were determined by subtracting WRTDS loading estimates based on subsamples from those obtained using the complete dataset. For each year (*t*) of each simulation, a coefficients of variation (CV) was calculated:

$$CV_t = \frac{\sigma_t}{\mu_t} \tag{1}$$

where $\sigma_t$is the standard deviation of residuals, μ_t is the mean loading estimate (for a simulation in year *t*). Results from subsampling are shown in Figure S6, where CV values are plotted as a function of number of samples per year. A power function was fit to this distribution and used to estimate CV based on number of samples taken per year ($R^2 = 0.95$):

$$CV_{i,t} = 0.9662 * X_{i,t}{}^{-0.783} \tag{2}$$

where $CV_{i,t}$ is the coefficient of variation for watershed *i* and year *t*, and $X_{i,t}$ is the number of samples taken. Equation 2 was applied to calculate the CV for each annual WRTDS loading estimate for each of the 21 sampling stations used in this study. These CVs were then used to determine the standard deviation of each loading estimate, consistent with equation 1. The incremental error variance, representing the uncertainty in the change in loading across an incremental watershed, is calculated based on the relationship between two correlated random variables (Kottegoda & Rosso, 2008):

$$\tilde{\sigma}^2_{i,t} = \sigma^2_{z,i+1,t} + \sigma^2_{z,i,t} - 2\rho_{i+1,i}\sigma_{z,i+1}\sigma_{z,i} \tag{3}$$

where $\tilde{\sigma}^2_{i,t}$ is the incremental error variance for incremental watershed *i* and year *t*, $\sigma^2_{z,i,t}$ is the WRTDS error variance for station i, $\sigma^2_{z,i+1,t}$ is the error variance for the upstream station to station *i*, $\rho_{i+1,i}$ is the correlation of loadings between watershed *i* and the upstream watershed, $\sigma_{z,i+1}$ is the WRTDS standard deviation of station *i*, and $\sigma_{z,i}$ is the WRTDS standard deviation of the upstream station.

Text S2. RStan Code

```
parameters {
real<lower =0> Be_a;                        //Export Coeffiicent for
Agriculture
real<lower =0> Be_d;                        //Export Coeffiicent for
Developed
real<lower =0> Be_w;                        //Export Coeffiicent for Wild
real<lower =0, upper = 1> Bd_c;             //Delivery Export Coefficient
for Chickens
real<lower =0, upper = 1> Bd_h;             //Delivery Export Coefficient
for Hogs
real <lower=0> Bd_p;                        //Delivery Export Coefficient
for Point-Sources
real <lower =1, upper = 100> Ro_str;        //Mass Transfer Coefficient for
streams
real <lower =1, upper = 100> Ro_wb;         //Mass Transfer Coefficient for
water bodies
vector<lower =0, upper = 2> [5] Bp;         //Precipitation Impact
Coefficient
real<lower=0, upper = 2> sigma_res;         //Residual sigma
real <lower=0, upper = 5> sigma_alpha;      //Alpha sigma
vector [wshed_size] alpha;                  //Alpha Random Effects
real<lower = 0, upper = 2> sigma_Bp;        //Sigma for Precipitation
Impact Coefficient
vector [nr] ly;                              //Log unknown true loads
}

model {
vector [nr] tot;
vector [nr] sigma;
vector [nr] y_hat;
vector [nr] ag;
vector [nr] dev;
vector [nr] wild;
vector [nr] ch;
vector [nr] h;
vector [nd] disch_mult;
vector [nr] tot_loss;
vector [nr] disch_loss;
vector [nd] d_vals;
vector [nl] l_vals;
int w;
vector [nr] alpha_vals;
vector [nl] a;
vector [nl] d;
vector [nl] wld;
real t;
vector [nr] ly_hat;
vector [nr] y;

//Calculate stream and water body losses for dischagers
disch_loss = exp(-Ro_str ./ d_str_loss) .* exp(-Ro_wb ./ d_res_loss);

//Calculate stream and water body losses for livestock
liv_loss = exp(-Ro_str ./ tot_loss1) .* exp(-Ro_wb ./ tot_loss2);
```

```
//Calculate loadings for sources
ag = Be_a*(1+Bp[1]*av_prec) .* ag_lst .* exp(-Ro_str ./ ag_loss1) .*
exp(-Ro_wb ./ ag_loss2);
dev = Be_d*(1+Bp[2]*av_prec) .* dev_lst .* exp(-Ro_str ./ dev_loss1) .*
exp(-Ro_wb ./ dev_loss2);
wild = Be_w*(1+Bp[3]*av_prec) .* wild_lst .* exp(-Ro_str ./ wild_loss1)
.* exp(-Ro_wb ./ wild_loss2);
ch = Bd_c*(1+Bp[4]*av_prec) .* chick .* liv_loss;
h = Bd_h*(1+Bp[5]*av_prec) .* hog .* liv_loss;

//Calculate random effects term alpha for each watershed
for (i in 1:nr){
w= wshed[i];
alpha_vals[i] = alpha[w]*100000;
}

//Sum loadings from sources
tot =  ag + dev + wild + ch + h + disch * disch_loss .* (point);

//Add alpha values to source loadings and subtract loss from upstream
load
y_hat = tot + alpha_vals - up_t_load .* (1-(exp(-Ro_str ./
str_loss_load) .* exp(-Ro_wb ./ res_loss_load)));

//priors
Be_a ~ normal(9,7.116);
Be_d ~ normal(7.5,3.41);
Be_w ~ normal(2.33,2.17);
Bd_c ~ normal(0,1);
Bd_h ~ normal(0,1);
sigma_res ~ normal(0,2);
sigma_alpha ~ normal(0,5);
alpha ~ normal(0,sigma_alpha);
sigma_Bp ~ normal(0,2);
Bp ~ normal(0,sigma_B1);
Ro_str ~ lognormal(3.19,1.36);
Ro_wb ~ normal(6.8,5.8);
disch ~ normal(1,.125);

//Likelihood formulation
ly_hat=log((y_hat ./ 100000)+10);
ly ~ normal(ly_hat,sigma_res);
y=exp(ly)-10;
load ~ normal(y,SD);           // load = WRTDS estimate, SD = WRTDS sd
```