Assessing interannual variability in nitrogen sourcing and retention through hybrid Bayesian watershed modeling

Jonathan W. Miller[1], Kimia Karimi[2], Arumugam Sankarasubramanian[1], Daniel R. Obenour[1]

[1] Dept. of Civil, Construction and Environmental Engineering, North Carolina State University, Raleigh, NC, USA
5   [2]Dept. of Geospatial Analytics, North Carolina State University, Raleigh, NC, USA

Correspondence to: Jonathan W. Miller (jwmille7@ncsu.edu)

Abstract. Excessive nutrient loading is a major cause of water quality problems worldwide, often leading to harmful algal blooms and hypoxia in lakes and coastal systems. Efficient nutrient management requires that loading sources are accurately quantified. However, loading rates from various urban and rural non-point sources remain highly uncertain especially with
10   respect to climatological variation. Furthermore, loading models calibrated using statistical techniques (i.e., hybrid models) often have limited capacity to differentiate export rates among various source types, given the noisiness and paucity of observational data common to many locations. To address these issues, we leverage data for two North Carolina Piedmont river basins collected over three decades (1982-2017) using a mechanistically parsimonious watershed loading and transport model calibrated within a Bayesian hierarchical framework. We explore temporal drivers of loading by incorporating annual
15   changes in precipitation, land use, livestock, and point sources within the model formulation. Also, different representations of urban development are compared based on how they constrain model uncertainties. Results show that urban lands built before 1980 are the largest source of nutrients, exporting over twice as much nitrogen per hectare than agricultural and post-1980 urban lands. In addition, pre-1980 urban lands are the most hydrologically constant source of nutrients, while agricultural lands show the most variation among high and low flow years. Finally, undeveloped lands export an order of magnitude (~ 7-
20   13x) less nitrogen than built environments.

1 Introduction

Eutrophication stimulated by anthropogenic nutrient loading is a common cause of water quality problems worldwide (Smith et al., 1999). In North Carolina (NC, USA), watershed-level nutrient management strategies have been developed for major reservoirs like Jordan Lake (JL) and Falls Lake (FL) using various process-based models (NC DWR, 2009; Tetra Tech 2014).
25   Such models can operate on fine temporal scales (i.e., days) and characterize various mechanistic processes related to the transfer of water and nutrients through watersheds. However, due to the large number of uncertain parameters (i.e., rates, coefficients) included in these models, multiple parameter sets may appear to fit the observational data equally well (Beven, 2006) without benefits to predictive performance (Jackson-Blake et al., 2017). Related to these issues, there is critical need for systematic model calibration and uncertainty quantification, if modeling results are to inform management decisions
30   (Reckhow, 1994; NRC, 2001; Reichert, 2020).

Hybrid (empirical-mechanistic) watershed models, which represent nutrient loading, transport, and retention using simple mechanistic relationships and statistical calibration techniques (e.g., nonlinear regression, Bayesian inference), have also been developed for nutrient source apportionment. For example, numerous applications of the SPAtially Referenced Regressions of contaminant transport On Watershed attributes (SPARROW) model have been applied to many large river basins (Preston et al., 2011; Hoos and McMahon, 2009; Garcia et al., 2011). SPARROW is calibrated using nonlinear regression that allows for parameter uncertainty quantification (i.e., confidence intervals). A limitation of SPARROW is that it models long-term average conditions and does not directly consider variability due to changes in precipitation (e.g., wet versus dry years) and watershed development, which have been shown to greatly affect nutrient loading (Howarth et al., 2012; Sinha and Michalak, 2016; Strickling and Obenour, 2018).

Methodological enhancements to SPARROW and similar hybrid watershed models have been proposed over time (Qian et al., 2005; Wellen et al., 2012; Xia et al., 2016). Recently, a Bayesian-hierarchical hybrid watershed model was developed to leverage temporal variability in source distributions and precipitation over multiple decades, providing an assessment of how land use change and hydroclimatological variations have affected nutrient loading over time (Strickling and Obenour, 2018). Additionally, this approach systematically incorporated and updated prior information on nutrient export and retention rates from previous studies through Bayesian inference. At the same time, Bayesian hybrid models often show limited capacity to differentiate loading rates among multiple source types (e.g., different land uses). Previous applications typically included only a small number of source types or had wide posterior credible intervals for export rates (e.g., Qian et al. 2005; Wellen et al., 2012; Strickling & Obenour 2018).

The goal of this study is to improve our understanding of nitrogen export within two highly managed NC basins that feed critical water supply reservoirs. Using a Bayesian hybrid watershed modelling approach, we characterize export rates from several different land uses, livestock types, and point sources. In particular, we explore how nitrogen-loading estimates vary among different types of urban lands based on density and the age of construction, considering how improved regulations and building practices may influence nutrient export. In addition, we demonstrate a novel approach for characterizing interannual variability in both nitrogen export and stream and waterbody retention, based on mean annual precipitation. This study benefits from a relatively high-resolution monitoring network (with a mean watershed monitoring unit of just 321 km^2, compared to 1535 km^2 in Strickling & Obenour, 2018) and over 30 years of loading data (1982-2017). Finally, we characterize instream retention rates and partition reservoir loading into various upstream sources based on varying hydro-climatological conditions to help inform watershed management.

2 Methods

2.1 Study Area

JL and FL, located in the Piedmont region of NC (Fig. 1), were impounded by the US Army Corps of Engineers in the early 1980s. Portions of each reservoir have exceeded NC water quality criteria, particularly for algae (chlorophyll a; NC DWR,

2020). JL watershed planning has been ongoing since the early 2000s, and initial TN reductions were set at 35% for the New Hope (NH) Creek basin and 8% for the Haw River (HR) basin. FL watershed planning was formalized in 2011, and Phase I

65   goals of the Falls Lake Rules included a TN reduction of 40% from major sources in the watershed (http://portal.ncdenr.org/web/fallslake/). JL and FL fall within the Cape Fear River basin and Neuse River basin respectively, but both share similar underlying hydroclimatic and soil conditions and comparable levels of anthropogenic development (Markewich et al., 1990; Strickling & Obenour, 2018).

2.2 Load monitoring sites (LMSs)

70   Nutrient load monitoring sites (LMSs) were identified based on locations that had sufficient flow and nutrient sampling data to calculate yearly TN loads. To be included, a site needed a minimum of five years of daily flow records and at least 50 water quality samples during that period of record. While somewhat context specific, these minimum conditions are generally consistent with previous studies using USGS WRTDS for load estimation (Chanat et al., 2012; Hirsch and De Cicco, 2015). All flow data were obtained from the United States Geological Survey (USGS), whereas nutrient data were obtained from the

75   Water Quality Portal (WQP; Read et al., 2017) as well as local city managers (e.g., city of Durham). The two largest sources of nutrient data (from the WQP) came from the USGS and the NC Department of Environmental Quality (NCDEQ). Sites from these different entities were often located in close proximity. Data from water quality sites with less than 5% deviations in watershed area and no intervening point sources were compiled together (Table 1).

In many cases, ample water quality data were available at the location of the USGS flow monitoring station. However, if little

80   or no water quality data were located at the flow station, a nearby water quality station was used instead, assuming there was less than a 20% change in watershed area between the flow and water quality monitoring stations. If multiple water quality sites were located close to the flow station, only the site with the longest record was chosen. In one exception, two water quality sites utilized the same flow monitoring station (NH1 and NH6; Table 1), which was done to include two substantial data records collected above and below a major point source on Morgan Creek. In such cases, the LMSs were represented at

85   the location of the water quality monitoring sites, and flows were adjusted based on the drainage area ratio between the two sites adjusting for any intervening wastewater flow.

There were 25 LMSs in our study area (Fig. 1; Table 1). Stations were split into three major basins for classification purposes: the HR basin of JL, the NH Creek basin of JL, and the FL basin. LMSs captured 85% of the HR basin, 49% of the NH basin, and 62% of the FL basin. Three LMSs were located directly downstream of major impoundments (HR4, FL6, and FL 9; Fig.

90   1; Table 1).

2.3 Delineation of Incremental Watersheds and Subwatersheds

Watersheds for LMSs were delineated using Spatial Analyst tools in ArcMap 10.6.1 (ESRI, 2018). Watershed drainage areas ranged from 11 km^2 to over 3000 km^2 (Table 1) with a median value of 106 km^2. Often, LMS watersheds had one or more LMS contained within their upstream watershed (Fig. 1). In order to accommodate nested LMS watersheds, we determined

incremental watersheds by subtracting out any upstream LMS watersheds that were contained in a larger (downstream) LMS watershed (Schwarz et al., 2006). If a LMS did not have any upstream LMSs in its watershed, its incremental watershed was equal to its total watershed. Note that the loads associated with these incremental watersheds formed the main response variable in our model (Section 2.7).

To more accurately account for nitrogen transport and retention, incremental watersheds were divided into subwatersheds (Fig. 1). Most data (e.g., land use, precipitation, livestock) were compiled at the subwatershed level. The largest possible subwatershed corresponded to a USGS 12-digit hydraulic unit code (HUC; https://water.usgs.gov/GIS/huc.html). If a LMS was located in the middle of a HUC, the HUC was split into two. Seventy-nine subwatersheds were located within the study area, with a mean drainage area of 63 km^2, minimum of 11 km^2, and maximum of 146 km^2.

2.4 Anthropogenic Factors

2.4.1 Land Uses

Land use variables were derived from the U.S. conterminous Wall-to-wall Anthropogenic Land use Trends (NWALT) dataset (Falcone 2015). We aggregated NWALT land use designations into three major categories: urban (including residential, transportation, industrial, and commercial development), agriculture (pasture and crop), and undeveloped (semi-developed, low use, and wetlands). Semi-developed land was included with undeveloped because it is mostly comprised of forested land in central NC (Miller et al., 2019). We further split urbanization constructed before and after a given date (e.g., 1980, 2000) and between low and high density. To determine when urbanization occurred in the region, we interpolated available NWALT data (1974, 1982, 1992, 2002, and 2012) to obtain year-specific land use values for each subwatershed. Since our study extended beyond 2012, we also used linear extrapolation for years 2013-2017 based on 2002 and 2012 values. Land use trends throughout the study period were generally gradual, such that modest linear extrapolation was considered reasonable (Fig. 2; top row).

2.4.2 Point Sources

Point sources included major (> 0.044 m^3/s) and minor wastewater treatment plants (WWTPs) (Fig. 1, Fig. S1, Table S1). Discharge data were obtained from NC DEQ and included monthly TN and flow values. However, many WWTPs had numerous missing months, so we determined annual loads as the product of yearly median concentrations and flows for each WWTP. LMSs with major WWTPs in their watersheds were only modeled starting in 1994 due to a lack of discharge data before that year (i.e., HR1,3,5, NH1-3, and FL1,3,10), while LMSs without major WWTPs were modeled from 1982 depending on data availability. Only one LMS (HR4) with pre-1994 monitoring data included a minor WWTP in its watershed. Since the minor WWTP represented < 3% of the LMS mean load, we assumed the pre-1994 load was equal to the mean post-1994 load. TN trends of point source discharges aggregated by basin are shown in Fig. 2 (middle row).

125   **2.4.3 Livestock**

The livestock in subwatersheds were estimated from county-level US Department of Agriculture (USDA) census and survey reports (https://www.nass.usda.gov/). Cow and swine data covered our entire study period (1982-2017) while chicken data were available every five years beginning in 1997. For missing years between census dates, chicken counts were interpolated, whereas chicken counts before 1997 were assumed to be equal to 1997 values. Only two incremental watersheds (HR1, 3) had

130   large chicken counts (>1,000,000 and >150,000, respectively) and these watersheds were not modeled before 1994 (as they were also missing major WWTP discharge data).

To represent the locations of livestock throughout the region, county-level data were assigned to incremental watersheds based on an area ratio. Major urban areas were excluded when calculating these proportions, as livestock were assumed to be located outside cities. Livestock counts were then divided into the subwatersheds (using area ratios). However, chickens in Chatham

135   County were accounted for differently because a majority of Chatham's chicken farms (>90%) are located outside of the JL basin, and the county has a relatively high chicken count (>3,000,000; USDA). Chatham County records (*opendata-chathamncgis.opendata.arcgis.com*) were used to estimate that 8.2% of the county's chickens were within the JL basin. Basin level trends of livestock are shown in Fig. 2 (bottom row).

2.5 Precipitation

140   Monthly precipitation estimates for this study were obtained from the PRISM Climate Group (http://www.prism.oregonstate.edu/). These data were processed using the R package "raster" (Hijmans et al., 2015; R Core Team, 2019) to determine mean annual precipitation for each subwatershed. There was substantial variation in precipitation among years (0.82 – 1.59 m/yr) and among different subwatersheds within the same year (Fig. S2).

2.6 Nutrient Load Calculations

145   Our model required yearly TN loadings at each LMS for Bayesian inference (i.e., calibration). Most riverine monitoring programs measure streamflow daily, whereas nutrient concentrations are sampled less frequently (e.g., monthly). In this study, daily TN concentrations were estimated using WRTDS (Hirsch et al., 2010). WRTDS develops a semi-parametric regression for each day in the estimation period where observations that are collected under similar conditions to the estimation date (in terms of time, discharge, and season) are more heavily weighted. For some LMS sites, there were abrupt temporal changes in

150   nutrient loading associated with WWTP upgrades. In these cases, WRTDS was run separately before and after the upgrade date to avoid smoothing out these transitions (Table S2).

Some LMSs had incomplete monitoring data (daily flow or water quality samples) during our study period (Table 1). If a LMS was missing flow data for a given year, that year was omitted. However, WRTDS is able to calculate loads for years with missing water quality data. Gaps in water quality samples of up to one year were considered acceptable in our study, as

155   preliminary analysis (removing single years of observational data at random) showed that a one-year gap affected loading

estimates by less than 1% and this is more conservative than WRTDS guidelines (Hirsch and De Cicco, 2015). In addition, at least 6 samples were required in both the beginning and ending year of each loading record.

Uncertainties in loading estimates were determined through subsampling of three NC stations that had nearly daily TN observations for at least 7 consecutive years (Strickling and Obenour, 2018). By comparing TN loads based on the full dataset and different subsets, we estimated the coefficient of variation (and standard deviation, SD) of WRTDS estimates based on the number of water quality samples available for a given year (Fig. S3). Accounting for these uncertainties allowed us to give more weight to loading estimates based on a larger number of nutrient samples (Section 2.7).

The response variable in our model was the change in nutrient load across an incremental watershed, defined as the difference between the load at an incremental watershed's downstream LMS and the cumulative load from any upstream LMSs. For sites with no upstream LMSs, the incremental load was equal to its total load. The uncertainties of incremental loads for LMSs with upstream LMSs were calculated based on the relationship between correlated random variables (Eq. (1); Kottegoda & Rosso, 2008):

$$\tilde{\sigma}_{i,t}^2 = \sigma_{i,t}^2 - 2\sum_{k=1}^{n} \rho_{i,k}\sigma_{i,t}\sigma_{k,t} + \sum_{k=1}^{n}\sum_{l=1}^{n} \rho_{k,l}\sigma_{k,t}\sigma_{l,t} \qquad (1)$$

where $\tilde{\sigma}_{i,t}^2$ is the incremental load variance for a given incremental watershed (i) in year (t), with n upstream LMSs (max n=3 for HR3; Fig. 1; Table 1). Here, $\sigma_{i,t}^2$ is the error variance at the downstream LMS, $\sigma_{k,t}$ and $\sigma_{l,t}$ are the WRTDS SDs at upstream LMSs, and $\rho_{i,k}$ and $\rho_{k,l}$ are correlation coefficients between LMS loadings.

2.7 Model Construction

Our model is formulated similar to Strickling and Obenour (2018). Within a Bayesian framework, we relate deterministically predicted incremental loads ($\hat{y}_{i,t}$; Eq. (2)) to an inferred incremental load ($y_{i,t}$). The watershed-level random effect (α_i; Gelman et al., 2014) accounts for spatial variability not explained by the deterministic prediction ($\hat{y}_{i,t}$), and the residual error (with SD σ_ε) primarily accounts for temporal variability unexplained by the deterministic prediction. The hyperdistribution of the normally distributed watershed-level random effect is centered on zero, with variance σ_{LMS}^2.

$$L(y_{i,t}) \sim N(L(\hat{y}_{i,t +}\alpha_i), \sigma_\varepsilon) \qquad (2)$$
$$\alpha_i \sim N(0, \sigma_{LMS})$$

$L(y)$ is the natural log transformation of $y + 10^5$ (kg/yr of TN). This transformation reduces heteroscedasticity in residuals while accounting for any negative incremental loads that would produce non-real values when log transformed. Negative incremental loads are possible in this model, especially for incremental watersheds with large impoundments that retain a substantial portion of the load from upstream LMSs. The largest negative loads were on the order of 5×10^4 kg/yr.

The inferred incremental load ($y_{i,t}$; Eq. (2)) is related to the WRTDS incremental estimates ($\tilde{y}_{i,t}$; Eq. (3)) by taking into account the uncertainty of those loading estimates ($\tilde{\sigma}_{i,t}$; Eq. (1)).

$$\tilde{y}_{i,t} \sim N(y_{i,t}, \tilde{\sigma}_{i,t}) \qquad (3)$$

Within the model, the deterministic prediction ($\hat{y}_{i,t}$) is calculated by aggregating incremental watershed source contributions and subtracting in-stream losses from upstream LMS loads (Eq. (4)).

$$\hat{y}_{i,t} = L_{i,t,ur1} + L_{i,t,ur2} + L_{i,t,ag} + L_{i,t,und} + L_{i,t,ps} + L_{i,t,ch} + L_{i,t,sw} + L_{i,t,cw} - U_{i,t} * r_{t,z} \qquad (4)$$

Contributions are calculated for two urban ($L_{i,t,ur1}$; $L_{i,t,ur2}$), agricultural ($L_{i,t,ag}$), and undeveloped ($L_{i,t,und}$) lands, point sources ($L_{i,t,ps}$), chickens ($L_{i,t,ch}$), swines ($L_{i,t,sw}$), and cows ($L_{i,t,cw}$). Loads from upstream incremental watersheds ($U_{i,t}$) are reduced by their expected in-stream and reservoir losses ($r_{t,z}$; Eq. (6); see *Nitrogen Retention*). Each source-specific load is calculated as follows:

$$L_{i,t,x} = \beta_x (\tilde{p}_{i,t}{}^{\gamma_x}) * \mathbf{a^T}_{i,t,x} * (1 - \mathbf{r}_{i,t}) \qquad (5a)\ \text{(land use)}$$

$$L_{i,t,x} = \beta_x (\tilde{p}_{i,t}{}^{\gamma_x}) * \mathbf{h^T}_{i,t,x} * (1 - \mathbf{r}_{i,t}) \qquad (5b)\ \text{(livestock)}$$

$$L_{i,t,x} = \beta_{ps} * \mathbf{w^T}_{i,t} * (1 - \mathbf{r}_{i,t}) \qquad (5c)\ \text{(point source)}$$

where $L_{i,t,x}$ (kg/yr) represents the total contributed load for a given LMS (i), source (x), and year (t). Parameter β_x represents a land or livestock export coefficient (EC; kg/ha/yr or kg/an/yr) or the point source (i.e., WWTP) delivery coefficient (unitless, 0-1). Parameter γ_x is the precipitation impact coefficient (PIC, unitless) for a given nonpoint source, which is parameterized as a power relationship with the export coefficient (Eq. (5)). PICs differ by source but are related to each other through a common hyperdistribution with mean μ_γ and SD σ_γ (Table S3). This formulation differs from the linear relationship between precipitation and loading used in Strickling and Obenour (2018) and avoids potentially negative loading values during extremely low flow years. Point sources do not have a PIC term (Eq.(5c)) as the WWTP data already account for yearly variation. Scaled annual precipitation ($\tilde{p}_{i,t}$) for each incremental watershed is determined by dividing by the mean precipitation of the study area. Often, a given source type was distributed among multiple locations (e.g., subwatersheds) within an incremental watershed. To account for this, $\mathbf{a^T}_{i,t,x}$, $\mathbf{h^T}_{i,t,x}$, and $\mathbf{w^T}_{i,t,x}$ are transposed vectors of sources (i.e., ha of land use, counts of livestock, and load from WWTPs, respectively) across different locations that are multiplied by a vector ($\mathbf{r}_{i,t}$) of location-specific stream and reservoir retention losses.

2.8 Nutrient Retention

Nitrogen retention in streams is represented based on a first-order decay rate ($-\kappa$, d^{-1}) and mean stream residence time (τ_z, d) for each path (z) from a given source (subwatershed or point source) to its downstream LMS. Estimated mean stream velocities are from the National Hydrography Dataset Plus (NHD+ ; Moore and Dewald, 2016). Travel distance was estimated as half the distance of the longest flow path within the source subwatershed plus the distance from the subwatershed to the downstream LMS. Nitrogen retention in reservoirs is modeled as a function of hydraulic loading rates (ratio of flow to surface area, q_w, m/yr) for each waterbody (w) and a mass transfer coefficient, (ω; m/yr.; Kelly, 1987). An overall retention rate ($r_{t,z}$), combining streams and reservoirs, for each path (z) and year (t) is determined as:

$$r_{t,z} = 1 - \exp(-\kappa * \tau'_{t,z}) * \prod_w \exp\left(\frac{-\omega}{q'_{t,w}}\right) \qquad (6)$$

allowing for multiple waterbodies along each flow path (i.e., the product function). While Strickling and Obenour (2018) used constant retention rates, here we allow rates to vary interannually by relating travel time and hydraulic loading to annual precipitation. Specifically, annual stream travel times ($\tau'_{t,z}$) and reservoir hydraulic loading rates ($q'_{t,w}$) are determined based on a retention PIC (γ_{ret}) and normalized yearly precipitation (p_t; yearly precipitation minus mean precipitation divided by SD), specific to each incremental watershed and year (Eq. (7a,b)):

$$\tau'_{t,z} = \frac{\tau_z}{(1+\gamma_{,ret}*\,p_t)} \tag{7a}$$

$$q'_{t,w} = q_w * (1+ \gamma_{ret} * p_t) \tag{7b}$$

2.9 Bayesian Inference

All model parameters were assigned prior probability distributions (Table S3). Informative priors were used when previous studies reporting similar parameters were available. Prior distributions for land export rates were taken from Dodd (1992), while stream retention rates were adapted from previous SPARROW studies (Hoos and McHahon, 2009; Garcia et al., 2011). Prior distributions for chicken and swine TN export coefficients were adapted from Strickling and Obenour (2018) to represent kilograms of TN per animal per year. Essentially uninformative priors (i.e., wide uniform priors) were used for the remaining parameters.

For comparison, models were calibrated with urban lands split in four different ways: 1) pre and post-1980 urban lands, 2) pre and post-2000 urban lands, 3) low versus high density urban (high-density residential only), and 4) low vs. high density urban (high density residential, industrial, and commercial). Cut-offs of 1980 and 2000 were chosen to represent urban areas built before and after changing NC environmental regulations related to erosion and sediment control (1980) and stormwater quality control measures (2000) that have come into effect over the past 50 years (Howell 1990; USEPA 2005, NC DEQ, 2017). In order to evaluate the best representation, we compared model fit and the degree of overlap between the marginal posterior parameter distributions of the two different urban export coefficients within each model.

Models were parameterized within the Bayesian framework using RStan software in R (R Core Team, 2019; Stan Development Team, 2020). RStan uses Hamiltonian Monte Carlo sampling of the posterior distribution and often converges faster than other samplers (Gelman et al., 2015). Three parallel chains of 20,000 iterations with a burn-in period of 5,000 iterations (that were discarded), creating 9,000 posterior samples after thinning (accepting every fifth iteration). Parameters were considered to have converged when their scale reduction coefficient ($\hat{R}$) was below 1.1 (Gelman and Rubin, 1992).

2.10 Model Assessment and Validation

Predictive performance was assessed using the coefficient of determination (i.e., variance explained, R^2, Faraway, 2016) for incremental nutrient loads. Predicted incremental loading estimates ($\hat{y}_{i,t}$) were derived using the Bayesian mean posterior values and compared to WRTDS loading estimates ($\tilde{y}_{i,t}$). Model performance was assessed for LMSs in the HR, NH, and FL watersheds with and without their watershed-level random effects. To test the ability of the model to make out-of-sample

predictions, we performed a 3-fold cross-validation (Elsner and Schmertmann, 1994). The data were split into three groups by major basin (HR, NH, and FL) and the model was trained on 2 of the 3 watersheds, in turn. Predictions were then made on the excluded basin (in turn).

3 Results

3.1 In-stream Nutrient Loading Estimates

WRTDS-derived annual loading estimates are quite noisy (Fig. 3; Fig. S4) due largely to hydrologic variability, while flow-normalized estimates help illustrate long-term trends (Hirsch et al., 2010). TN loads in all basins decrease substantially from 1980 to the late 1990s. However, post-2000 loading patterns are inconsistent both within and across basins (Fig. 3). In the HR watershed, TN loading has steadily increased since 2000. In the NH watershed, substantial increases in loads are seen after 1995, though subsequent WWTP improvements reduced loading in some tributaries (Fig. 3; Fig. S4). In FL, large post-2000 TN reductions appear in FL1 and FL10 (Fig. 3), both of which have major WWTP discharges, while loadings in other FL watersheds have remained constant or trended upwards.

3.2 Comparing Different Urban Land Classifications

The hybrid watershed model explains the spatial and temporal variability of in-stream (WRTDS-derived) TN loads based on precipitation and nutrient source distributions. To explore urban TN sources in more detail, we compare the posterior parameter distributions of different classifications of urban land use that consider the age, density, and type of urban development (Table 2). We find that a classification based on a pre/post 1980 split results in significantly different export rates, with the pre-1980 urban lands exporting more than twice the amount of post-1980 urban lands. Here, "significantly different" implies a >95% probability based on samples from the joint posterior parameter distribution. In addition, the pre/post 1980 division leads to the highest R^2 values for individual LMSs. None of the other urban splits result in significantly different parameter estimates. However, export rates from high-density and older urbanization are consistently higher than the less dense and newer urban lands. Among the two splits based on density, combining high-density residential, industrial, and commercial lands has higher predictive power than just separating high-density residential from other urbanization (R^2= 0.47 vs. 0.44; Table 2).

3.3 Model Posterior Parameter Estimates

The posterior ECs (β_x) of the preferred model (with the pre/post 1980 urban split) show that urban and agricultural lands both contribute substantial TN per hectare (Table 3; Fig. 4). In particular, pre-1980 urban development exports 9.4 kg/ha/yr of TN (coefficient of variation (CV) of 11%; Table 3), while post-1980 development exports 3.9 kg/ha/yr (CV=41%). Agriculture also contributes a substantial 4.0 kg/ha/yr (CV=21%), while undeveloped lands export a relatively low 0.7 kg/ha/yr. In addition to pre-1980 urban export being significantly greater than other forms of development, undeveloped land has significantly less export than all developed lands, despite having a high CV (50%; Table 3). Model posterior distributions indicate that parameter

uncertainties are reduced substantially relative to the prior distributions (Fig. 4), consistent with Strickling & Obenour (2018),
280 who found that priors had a relatively small influence on parameter estimates relative to the data (i.e., the likelihood).

Land use ECs represent expected nutrient export for a year with mean annual precipitation (i.e., $\tilde{p}_{i,t} = 1$; Eq. (5a)). Because the relationship between export and precipitation is nonlinear, the export coefficients represent median (but not mean) loading rates. The precipitation impact coefficients (PICs) can be used to calculate TN export during low and high flow years. Agriculture has the largest PIC (4.0; Table 3) implying that export from agricultural lands (crop and pasture) vary the most
285 due to rainfall. During a high flow year (90th percentile $\tilde{p}_{i,t}$=1.18), nutrient export for agriculture would almost double from 4.0 to 7.7 kg/ha/yr. For a low flow year (10th percentile $\tilde{p}_{i,t}$=0.81), nutrient export for agriculture (1.7 kg/ha/yr) is less than half the median export. Pre-1980 urban lands show the lowest variation due to precipitation, ranging from 81% of median export in low flow years to 122% in high flow years.

3.4 Spatial Variation in Nutrient Export and Retention

290 The TN export from nonpoint sources is calculated for each subwatershed (Fig. 5a) using mean precipitation, and mean posterior land and livestock ECs (β_{ec}; Table 3). Since the most intensive nutrient export comes from pre-1980 urban lands, subwatersheds intersecting the urban cores of major cities (Fig. 5a) have the largest expected export. Predominantly rural watersheds export between 1-3 kg/ha/yr of TN, while urban cores export over 6 kg/ha/yr.

On average, 13% of TN is retained within streams and waterbodies (Fig. 5b). Little TN is retained in subwatersheds close to
295 JL and FL and along higher order streams, while large TN removal rates (> 70%) occur for subwatersheds located upstream of reservoirs in the upper northwest portions of the JL basin. Overall, more TN retention occurs in reservoirs than in streams. Residence times and hydraulic loading rates are also affected by precipitation as modulated by the PIC for stream retention (γ_{ret}; 0.07; Table 3). For one SD increase in yearly precipitation (17.1 cm), expected stream residence times and hydraulic loading rates decrease roughly 7% (Table 3; Eq. (7a,b)). During low precipitation years (lower 33%), 15% of TN is retained
300 in the JL and FL networks, while 12% is retained during normal and high precipitation years (upper 67%).

Watershed-level random effects account for unexplained spatial variations in nutrient loading. For example, the negative random effects for small watersheds comprised of mostly pre-1980 urban development (NH7, NH8, FL2) imply these watersheds export less TN (-2.5, -1.8, and -1.2 kg/ha/yr, respectively; Fig. S5) than typical pre-1980 urban lands (9.4 kg/ha/yr; Table 3). Similarly, two LMSs (FL6, FL9) located directly downstream of large impoundments had negative random effects,
305 implying these impoundments may be particularly efficient at trapping nutrients. On the other hand, three watersheds (NH1, HR5, and FL1) located just downstream of major WWTP discharges have elevated TN watershed-level effects, suggesting loads may be underestimated by the available point source data (Fig. S5).

3.5 Nutrient Source Allocations Over Time

Yearly TN loadings from 1994-2017 are reported based on mean model parameters, land use, livestock counts, and precipitation. Only the NH watershed shows a clear downward trend in TN loading, which appears to be largely driven by WWTP discharge reductions (Fig. 6). In both the HR and FL watersheds, annual loadings nearly tripled from the lowest to highest precipitation years due to high levels of agricultural lands which substantially increase export during wet years (Fig. 6). This high-level of interannual variation, makes it difficult to distinguish any positive or negative trends in these basins over the study period. Additionally, model residuals show no noteworthy temporal trends (Fig. S6).

3.6 Model Skill Assessment

The full hybrid model, including random effects, explains 95% of the variation in the WRTDS loading estimates at LMSs (Fig. S7). Discounting the random effects, the model still explains 93% of TN loading variability. The model (with watershed-level random effects) explains 96% of the variation in the HR, 92% in NH, and 83% in FL (Fig. S7), suggesting some spatial variability in model performance. In cross validation, the predictive ability of the model remains high, with the R^2 of the full TN model (without watershed-level random effects) dropping slightly from 93% to 90%. R^2 values are also reported for individual LMSs to characterize the model's ability to predict the temporal variability. These R^2 values range greatly from below 0 (FL2, JL2, JL7) to above 0.80 (JL4, FL7, FL10) with a mean of 0.48 (Table 2).

4 Discussion

4.1 Nutrient Export Rates and Discharge Coefficients

In this study, we aim to enhance our understanding of nitrogen export, especially as it relates to land use and different types of urbanization. Variation in urban TN export has often been associated directly with population density (Bales et al., 1999; Burns et al., 2005; Line, 2013; Tetra Tech, 2014) or with proxies for density like net food imports (Hong et al., 2011; Sinha and Michalak, 2016). In this study, we compare variations in urban export due to the age of the urbanization versus different urban land covers (e.g., high and low-density residential, commercial). Pre-1980 urban lands and high-density residential are moderately correlated in our study area (r^2=0.64), yet Bayesian posterior parameter estimates show that export from low and high density urban areas are not statistically different from each other. However, TN export from pre and post-1980 urban areas are significantly different (Table 3). This suggests that urban infrastructure age and historical development practices are more important than population density.

Various mechanisms, beyond density, could explain why older urban areas export more TN than recently constructed urban areas. Increased impervious connectivity in pre-1980 urban areas may lead to elevated runoff and nutrient washoff (Wolheim et al., 2005). In addition, pre-1980 urban development generally lacked stormwater management and erosion control measures (Howells, 1990; NC DEMLR, 2019). Therefore, streams in these areas have legacy sediments and often exhibit the urban

stream syndrome with altered geomorphology and low biological health (Paul and Meyer, 2001; Bernhardt and Palmer, 2007; Miller et al., 2019) which can affect downstream nutrient loads and uptake (Meyer et al. 2005). Older neighborhoods are also

340   likely to have larger trees and thus more leaf litter over impervious surfaces, which can also increase nutrient export (Janke et al. 2017). Finally, leaky sewer infrastructure in pre-1980 urban areas might be a substantial source of nutrients (Kaushal et al., 2011; Pennino et al., 2016) as compared to newer and more reliable infrastructure in post-1980 urban areas.

Model results indicate that post-1980 urban and agricultural lands exported similar amounts of nitrogen (3.9 and 4.0 kg/ha/yr, respectively; Table 3). Our estimated agricultural export rate is lower than previous studies (Dodd et al. 1992; Strickling and

345   Obenour 2018), which may be related to the fact that over 90% of agricultural lands in our study area are pasturelands, rather than croplands (Falcone 2015). Post-1980 urban export shows the most uncertainty in model posteriors (Table 3; Fig. 4). This might be due to the inconsistency of regulations being applied both temporally and spatially from 1980 to the present, and it might also indicate variation in the best management practices (BMPs) used in the region. Finally, undeveloped lands have very low export (0.7 kg/ha/yr; Table 3) with moderate uncertainty (0.1-1.5 95% interval; Table 3). This mean value is roughly

350   three times lower than previous studies in the region (Tetra Tech 2014; Strickling and Obenour 2018).

Livestock export coefficients for chickens, swine, and cows (0.01, 0.04, and 0.50 kg/yr, respectively) imply that less than 1% of the TN produced by these animals (i.e., 0.6, 9.9, and 54.8 kg/an/yr, respectively; Ruddy et al., 2006) results in excess TN pollution in our study area. Overall, livestock-related nutrient export appears to account for <2% of nutrient loading to the downstream reservoirs (Fig. 6; Table S4). Per hectare of agricultural land, livestock export averages just 0.5 kg/ha/yr, about

355   an order of magnitude less than the baseline agricultural land use export. It is important to note that livestock waste used to replace other (e.g., synthetic) fertilizers is generally represented by the agricultural land export. The livestock rates reported here represent TN export in excess of typical pasture and cropland export.

Our point source coefficient discounts WWTP loads by nearly 20% (β_{ps}=0.83; Table 3). One potential explanation for this result is that TN from WWTPs (primarily nitrate) is processed and retained in stream networks more efficiently than TN from

360   nonpoint sources. For example, increased denitrification rates have been observed downstream of WWTPs due to altered biochemical conditions (Wakelin et al. 2008; Rahm et al. 2016).

4.2 Interannual Variability

Our analysis of interannual variability is facilitated through two enhancements to the hybrid watershed modeling approach of Strickling and Obenour (2018). First, we allow for retention to vary across years due to changes in precipitation, represented

365   parsimoniously by a PIC for stream residence times and reservoir hydraulic loading rates ($\gamma_{,ret}$; Eq. (7a,b)). This modification produces a +/-8% variation in retention for -/+1 SD change in annual precipitation. In this region, we find that the majority of retention occurs in reservoirs. Our mean stream retention rate (0.04 d^{-1}; Table 3) is comparable in magnitude to regional hybrid models (Strickling and Obenour, 2018; Gurley et al., 2019), but lower than previous watershed process models (Tetra Tech, 2014). Accurately quantifying stream retention rates is important to operators of WWTPs and regulators in order to accurately

370 determine nutrient offsets for projects, which are often valued in millions of dollars (personal communication Jim Hawhee, NC DEQ, 2020).

Second, we model the effects of precipitation variability on land use export using power functions ($\gamma_{,x}$; Eq. (5)) instead of linear relationships (Strickling and Obenour, 2018). The new formulation recognizes that as precipitation increases, its marginal effect on nutrient loading may intensify, as infiltration and evapo-transpiration rates are exceeded (Chin, 2013).

375 Consistent with this explanation, agriculture has the highest PIC (4.0; Table 3), indicating that when annual precipitation is 20% higher than the mean, TN export from agricultural lands will approximately double. On the other hand, pre-1980 urban lands, which are substantially impervious, have the lowest PIC values and only export 24% more TN for a 20% increase in annual precipitation.

By compiling nitrogen loads throughout the major basins (1994-2017; Fig. 6), we can analyse interannual trends and variability

380 in TN sourcing. Interannual variation is large (~ 3x difference from lowest to highest year) in both HR and FL, yet NH loadings show relatively little variation. This is due to high percentages of agricultural lands in HR and FL, which have the highest PICs. NH loadings, in contrast, are dominated by pre-1980 urbanization (lowest PIC) and point sources. By separating TN loadings into their sources, we are able to identify trends in specific nitrogen sources that could help inform management of individual watersheds. For example, though loadings in the NH watershed appear to decline over the study period from 6 x

385 10^5 kg/yr to 4 x 10^5 kg/yr, this trend was driven by a nearly 50% reduction in point sources (4 x 10^5 down to 2.1 x 10^5 kg/yr). At the same time, loading attributable to post-1980 urbanization increased almost two-fold, from 0.2 x 10^5 up to 0.4 x 10^5 kg/yr (Fig. 6).

4.3 Potential Nutrient Reductions

Model results strongly indicate that the majority of nutrient inputs to JL and FL are from anthropogenic sources. Based on

390 identified sources of TN in the watershed, four management strategies would potentially lead to large nutrient load reductions: 1) reduction of point source loadings (i.e., WWTPs), which remain the largest individual source of TN, 2) retrofitting or replacing infrastructure in older urban environments (i.e. pre-1980 urban), which are the largest nonpoint source of TN per unit area, 3) mitigating TN loading from agricultural lands, especially during wet conditions, and 4) limiting, reducing, or offsetting the removal of undeveloped land, which have the lowest export.

395 The effectiveness of these nutrient control strategies will vary across different hydrologic conditions. For example, point sources are responsible for 44% of TN loadings to JL and 14% to FL from 1994 to 2017. However, these percentages rise to 55% and 24%, respectively, during dry years (Table S4). On the other hand, agricultural TN export accounts for 18% of JL loadings and 30% of FL loadings during normal flow years, but those numbers increase to 27% and 37% during high flow years (Table S4). Therefore, strategies aimed at mitigating point source loadings will affect lake loadings more in low flow

400 years, while agricultural strategies will have a larger effect in high flow years.

Undeveloped lands have the lowest nutrient export of all land-use sources (Table 3), which is consistent with findings from targeted water quality monitoring recently done in JL (NC Policy Collaboratory, 2019). However, undeveloped areas are

decreasing in this region (1.5 km^2/yr, 3 km^2/yr, and 2 km^2/yr in HR, NH, and FL basins, respectively, from 1995 to 2015). Even though more recent development (post-1980) has significantly lower TN export when compared to pre-1980

405 development, it still exports approximately six times more TN than undeveloped lands (3.9 vs. 0.7 kg/ha/yr; Table 3). At the same time, post-1980 urban export rates are similar to agricultural rates, implying that new urban construction in agricultural areas may have limited impact on total nutrient loading.

4.4 Summary and Future Directions

To efficiently manage watersheds, it is critical to identify the major sources and locations of pollutant loading under varying

410 hydroclimatological conditions. We enhanced and applied a hybrid watershed modeling approach within a Bayesian framework to characterize TN loading rates from point and nonpoint sources and tested different classifications of urban land. By modeling interannual variability, we assessed how land use change and hydroclimatological variations have affected nutrient loading over time. Process-based and hybrid modeling approaches (e.g., SPARROW; Gurley et al. 2019) have been widely used to determine nitrogen loading rates, but these applications are often limited by an inability to capture interannual

415 variations in loading and/or provide holistic parameter calibration with uncertainty quantification. Compared to previous Bayesian hybrid watershed modeling studies (Qian et al., 2005; Wellen et al., 2012; Strickling & Obenour, 2018), this study advances our ability to account for interannual variability in both export and retention. The study also discriminates how export rates vary across a relatively large number of sources (4 land uses, 3 livestock types, and point sources). Our ability to resolve 18 process-based parameters within the Bayesian framework is facilitated, in part, by a relatively dense stream monitoring

420 network and modern tools for Bayesian inference (Monnahan et al., 2017). The efficacy of the proposed approach for characterizing nutrient export in larger but less densely monitored watersheds could be explored in future research.

In this study, we identify areas of elevated TN export. In particular, we find that pre-1980 urban areas are hot spots for nonpoint TN loading. In addition, watershed-level random effects help identify outlier watersheds that export significantly more or less TN than the source distribution data would otherwise imply. Great costs have been incurred to protect waterways in the last

425 30-40 years without a clear understanding of how effective current policies have been in reducing nutrient loading (Parr et al., 2016; Utz et al., 2016). Our results suggest that post-1980 construction and land development BMPs have helped to reduce TN loadings from the built environment. We hope these findings will stimulate further research into the specific mechanisms that result in lower TN export from newer development. Enhancing the hybrid model with BMP and wastewater infrastructure data, in addition to more detailed land use and hydrography data, could be one approach for refining our understanding of how

430 specific development practices influence watershed-scale nutrient loading. Given that even newer (post-1980) development is found to increase TN export by a factor of five or more, relative to undeveloped lands, further efforts are required to understand and mitigate the adverse impacts of urban development on nutrient loading.

Code/Data Availability

Hydrometeorological and water quality data can be obtained from the public sources described in the methods (e.g., USGS, Water Quality Portal). RStan code is included in the SI (Text S1). A complete input dataset for the code is available here: https://github.com/obelab/FallsJordanWshHESS

Author Contribution

Jonathan Miller and Daniel R. Obenour designed the research. Jonathan Miller and Kimia Karimi developed the codes and created tables and figures. All authors analysed the research. Jonathan Miller and Daniel R. Obenour prepared the manuscript with contributions from all co-authors.

Competing interest

The authors declare that they have no conflict of interest.

Acknowledgements

We would like to thank Hayden Strickling for assistance with preliminary modelling files. We would also like to thank Bridget Munger (NC DEQ) for data assistance with WWTP data.

References

Bales, J., Weaver, J. C., & Robinson, J. B.: Relation of Land Use to Streamflow and Water Quality at Selected Sites in the City of Charlotte and Mecklenburg County, North Carolina, 1993-98 (Vol. 99, No. 4180), USDI, US Geol. Surv., https://doi.org/10.3133/wri994180, 1999.

Bernhardt, E. S., and Palmer M.A.: Restoring streams in an urbanizing world, Freshwater Biol., 52(4), 738-751, https://doi.org/10.1111/j.1365-2427.2006.01718.x, 2007.

Burns, D., Vitvar, T., McDonnell, J., Hassett, J., Duncan, J., and Kendall, C.: Effects of suburban development on runoff generation in the Croton River basin, New York, USA, J. Hydrol., 311(1-4), 266-281, https://doi.org/10.1016/j.jhydrol.2005.01.022, 2005.

Chanat, J. G., Moyer, D. L., Blomquist, J. D., Hyer, K. E., and Langland, M. J.: Application of a weighted regression model for reporting nutrient and sediment concentrations, fluxes, and trends in concentration and flux for the Chesapeake Bay Nontidal Water-Quality Monitoring Network, results through water year 2012 (No. 2015-5133), USGS, https://doi.org/10.3133/sir20155133, 2012.

Chin, D.: Water-resources engineering, 3rd Edition, Pearson Education Inc, Upper Saddle River, New Jersey, 2013

Dodd, R. C., Cunningham, P. A., Curry, R. J., Liddle, S. K., McMahon, G., Stichter, S., and Tippett, J. P.: Watershed planning in the Albemarle-Pamlico estuarine system (No. 3), Albemarle-Pamlico Estuarine Study, NC Dept. of Environment, Health, and Natural Resources, 1992.

Elsner, J. B., and Schmertmann C.P..: Assessing forecast skill through cross validation, Weather Forecast., 9(4), 619-624, https://doi.org/10.1175/1520-0434(1994)009<0619:AFSTCV>2.0.CO;2, 1994

Falcone, J. A.: US conterminous wall-to-wall anthropogenic land use trends (NWALT), 1974–2012 (No. 948),US Geol. Surv., https://doi.org/10.3133/ds948, 2015.

Faraway, J. J.: Linear models with R. CRC press, https://doi.org/10.1201/b17144, 2016

Garcia, A. M., A. B. Hoos, and Terziotti S.: A Regional Modeling Framework of Phosphorus Sources and Transport in Streams of the Southeastern United States, J. Am. Water Resour. A., 47(5), 991-1010, https://doi.org/10.1111/j.1752-1688.2010.00517.x, 2011.

Gelman, A. and Rubin D. B.: Inference from iterative simulation using multiple sequences, Stat. Sci., 7(4), 457-472. https://doi.org/10.1214/ss/1177011136 , 1992.

Gelman, A. and Hill, J.: Data analysis using regression and multilevel/hierarchical models. Cambridge university press, https://doi.org/10.1017/CBO9780511790942, 2006.

Gelman, A., Carlin, J. B., Stern, H. S., and Rubin, D. B.: Bayesian data analysis (Vol. 2), Boca Raton, FL, USA, Chapman & Hall/CRC, 2014.

Gelman, A., D. Lee, and J. Guo J.: Stan: A probabilistic programming language for Bayesian inference and optimization, J. Educ. Behav. Stat., 40(5), 530-543, https://doi.org/10.3102/1076998615606113, 2015.

Gurley, L.N., A. M. Garcia, S. Terziotti, and Hoos, A.B..: SPARROW model datasets for total nitrogen and total phosphorus in North Carolina, including simulated streamloads: USGS data release, https://doi.org/10.5066/P9UUT74V, 2019.

Hijmans, R. J., et al.: Package 'raster', R package, 2015.

Hirsch, R. M., D.L. Moyer, and Archfield, S.A.: Weighted regressions on time, discharge, and season (WRTDS), with an application to chesapeake bay river inputs, J. Am. Water Resour. A.,46(5), 857–880, https://doi.org/10.1111/j.1752-1688.2010.00482.x, 2010.

Hirsch, R.M., and De Cicco, L.A.: User guide to Exploration and Graphics for RivEr Trends (EGRET) and dataRetrieval: R packages for hydrologic data (version 2.0): USGS Techniques and Methods book 4, chap. A10, https://doi.org/10.3133/tm4A10, 2015.

Hong, B., Swaney, D. P., and Howarth, R. W.: A toolbox for calculating net anthropogenic nitrogen inputs (NANI), Environ. Modell. Softw., 26(5), 623-633, https://doi.org/10.1016/j.envsoft.2010.11.012, 2011.

Hoos, A. B. and McMahon G.: Spatial analysis of instream nitrogen loads and factors controlling nitrogen delivery to streams in the southeastern United States using spatially referenced regression on watershed attributes (SPARROW) and regional classification frameworks, Hydrol. Process., 23(16), 2275-2294, https://doi.org/10.1002/hyp.7323, 2009.

Howarth, R., Swaney, D., Billen, G., Garnier, J., Hong, B., Humborg, C., Johnes, P., Morth, C.M, and Marino, R.: Nitrogen fluxes from the landscape are controlled by net anthropogenic nitrogen inputs and by climate, Front. Ecol. Environ., 10(1), 37-43, https://doi.org/10.1890/100178, 2009.

Howells, D. H.: Quest for clean streams in North Carolina: An historical account of stream pollution control in North Carolina, Water Resources Research Institute of the University of North Carolina, 1990.

Jackson, C. R., Martin, J. K., Leigh, D. S., and West, L. T.: A southeastern piedmont watershed sediment budget: Evidence for a multi-millennial agricultural legacy, J. Soil Water Conserv., 60(6), 298-310, 2005.

Jackson-Blake, L. A., Sample, J. E., Wade, A. J., Helliwell, R. C., and Skeffington, R. A.: Are our dynamic water quality models too complex? A comparison of a new parsimonious phosphorus model, S imply P, and INCA-P, Water Resour. Res., 53(7), 5382-5399, https://doi.org/10.1002/2016WR020132, 2017.

Janke, B. D., Finlay, J. C., and Hobbie, S. E.: Trees and streets as drivers of urban stormwater nutrient pollution, Environ. Sci. Technol., 51(17), 9569-9579, https://doi.org/10.1021/acs.est.7b02225, 2017.

Kaushal, S. S., Groffman, P. M., Band, L. E., Elliott, E. M., Shields, C. A., and Kendall C.: Tracking nonpoint source nitrogen pollution in human-impacted watersheds, Environ. Sci. Technol.,45(19): 8225-8232. https://doi.org/10.1021/es200779e, 2011.

Kelly, C. A., Rudd, J. W. M., Hesslein, R. H., Schindler, D. W., Dillon, P. J., Driscoll, C. T., Gherini, S.A., and Hecky R.E.: Prediction of biological acid neutralization in acid-sensitive lakes, Biogeochemistry, 3(1-3), 129-140, https://doi.org/10.1007/BF02185189, 1987.

Kottegoda, N. T. and Rosso, R.: Applied statistics for civil and environmental engineers, Malden, MA: Blackwell, 2008.

Line, D. E.: Effect of development on water quality for seven streams in North Carolina, Environ. Monit. Assess., 185(8), 6277-6289, https://doi.org/10.1007/s10661-012-3024-z, 2013.

Markewich, H. W., Pavich, M. J., and Buell, G. R.: Contrasting soils and landscapes of the Piedmont and Coastal Plain, eastern United States, Geomorphology, 3(3-4), 417-447, https://doi.org/10.1016/0169-555X(90)90015-I, 1990.

Meyer, J. L., Paul, M. J., and W. K. Taulbee. Stream ecosystem function in urbanizing landscapes. Journal of the North American Benthological Society, *24*(3), 602-612, https://doi.org/10.2307/1468410, 2005.

Miller, J. W., Paul, M. J., and Obenour, D. R..: Assessing potential anthropogenic drivers of ecological health in Piedmont streams through hierarchical modeling, Freshw. Sci., 38(4): 771-789, https://doi.org/10.1086/705963, 2019.

Monnahan, C. C., Thorson, J. T., and Branch, T. A.: Faster estimation of Bayesian models in ecology using Hamiltonian Monte Carlo, Method. Ecol. Evol., 8(3), 339-348, https://doi.org/10.1111/2041-210X.12681, 2017.

Moore, R. B. and Dewald, T. G.:The Road to NHDPlus—Advancements in Digital Stream Networks and Associated Catchments, J. Am. Water Resour. A., 52(4), 890-900, https://doi.org/10.1111/1752-1688.12389, 2016.

NC DEMLR (NC Division of Energy, Minerals and Land Resources).: Erosion and sediment control planning and design manual. NC Department of Environmental Quality, https://deq.nc.gov/about/divisions/energy-mineral-land-resources/energy-mineral-land-permit-guidance/erosion-sediment-control-planning-design-manual, 2019.

NC DWR (NC Division of Water Resources).: Falls Lake Watershed Analysis Rick Management Framework (WARMF) Development, NC Department of Environmental Quality, Modeling/TMDL Unit, 2009.

NC DWR (NC Division of Water Resources).: Integrated Report 303(d) 305(b) Files, NC Department of Environmental Quality,https://deq.nc.gov/about/divisions/water-resources/planning/modeling-assessment/water-quality-data-assessment/integrated-report-files, 2020

NC Geol. Surv.: NC Department of Environmental Quality. https://deq.nc.gov/about/divisions/energy-mineral-land-resources/north-carolina-geological-survey, 2019.

NC Policy Collaboratory.: The University of North Carolina Jordan Lake Study. Final Report to the North Carolina General Assembly. "Stream Monitoring and Nutrient Loading," pp. 57-70, https://collaboratory.unc.edu/wp-content/uploads/sites/476/2020/01/2019-jordan-lake-final-report.pdf, 2019.

NRC (National Research Council). Assessing the TMDL approach to water quality management, National Academies Press, Washington, D.C., 2001.

Parr, T. B., Smucker, N. J., Bentsen, C. N., and Neale, M. W.: Potential roles of past, present, and future urbanization characteristics in producing varied stream responses, Freshw. Sci., 35(1), 436-443, https://doi.org/10.1086/685030, 2016.

Paul, M. J., and Meyer, J. L.: Streams in the urban landscape, Annual review of Ecol. and Systematics, 32(1), 333-365, https://doi.org/10.1146/annurev.ecolsys.32.081501.114040, 2001.

Pennino, M. J., Kaushal, S. S., Mayer, P. M., Utz, R. M., and Cooper. C. A.: Stream restoration and sewers impact sources and fluxes of water, carbon, and nutrients in urban watersheds, Hydrol. Earth Sys. Sc., 20(8), 3419-3439, doi:10.5194/hess-20-3419-2016, 2016.

Preston, S. D., R.B. Alexander, G.E. Schwarz, and Crawford, C. G.: Factors Affecting Stream Nutrient Loads: A Synthesis of Regional SPARROW Model Results for the Continental United States, J. Am. Water Resour. A., 47(5), 891-915, https://doi.org/10.1111/j.1752-1688.2011.00577.x, 2011.

Qian, S. S., Reckhow, K. H., Zhai, J., & McMahon, G.: Nonlinear regression modeling of nutrient loads in streams: A Bayes ian approach, Water Resour. Res., 41(7), https://doi.org/10.1029/2005WR003986, 2005.

R Core Team.: R: A language and environment for statistical computing, R Foundation for Statistical Computing, Vienna, A ustria, https://www.R-project.org/, 2019.

Rahm, B. G., Hill, N. B., Shaw, S. B., and Riha, S.J.: Nitrate dynamics in two streams impacted by wastewater treatment plant discharge: Point sources or sinks? J. Am. Water Resour. A., 52(3), 592-604, https://doi.org/10.1111/1752-1688.12410, 2016.

Read, E. K., Carr, L., De Cicco, L., Dugan, H. A., Hanson, P. C., Hart, J. A., Kreft, J., Read, J. S., and Winslow, L. A.: Water quality data for national-scale aquatic research: The Water Quality Portal, Water Resour. Res., 53(2), 1735-1745, https://doi.org/10.1002/2016WR019993, 2017.

Reckhow, K. H.: Water quality simulation modeling and uncertainty analysis for risk assessment and decision making. Ecol. Model., 72(1-2): 1-20, https://doi.org/10.1016/0304-3800(94)90143-0, 1994.

Reichert, P.: Towards a comprehensive uncertainty assessment in environmental research and decision support. Water Science and Technology, 81(8), 1588-1596, https://doi.org/10.2166/wst.2020.032, 2020.

Ruddy, B. C., Lorenz, D. L., and Mueller, D. K.: County-level estimates of nutrient inputs to the land surface of the conterminous United States, 1982-2001, https://doi.org/10.3133/sir20065012, 2006.

Schwarz, G. E., Hoos, A. B., Alexander, R. B., and Smith, R. A. The SPARROW surface water-quality model: theory, application and user documentation. US geological survey techniques and methods report, book, 6(10), 248. 2006.

Sinha, E., and Michalak, A. M.: Precipitation dominates interannual variability of riverine nitrogen loading across the continental United States, Environ. Sci. Technol., 50(23), 12874-12884, https://doi.org/10.1021/acs.est.6b04455, 2016.

Smith, V. H., Tilman, G. D., and Nekola, J. C.: Eutrophication: impacts of excess nutrient inputs on freshwater, marine, and terrestrial ecosystems. Environ. Pollut., 100(1-3), 179-196, https://doi.org/10.1016/S0269-7491(99)00091-3, 1999.

Strickling, H.L. & Obenour, D.R.: Leveraging Spatial and Temporal Variability to Probabilistically Characterize Nutrient Sources and Export Rates in a Developing Watershed, Water Resour. Res, 54(7), 5143-5162, https://doi.org/10.1029/2017WR022220, 2018.

Stan Development Team.: "RStan: the R interface to Stan." R package version 2.19.3, http://mc-stan.org/, 2020

575 Tetra Tech.: Lake B. Everett Jordan Watershed Model Report. NC Nutrient Science Advisory Board, NC Division of Water Resources, and Triange J. Council of Governments, 2014.

Utz, R. M., Hopkins, K. G., Beesley, L., Booth, D. B., Hawley, R. J., Baker, M. E., Freeman, M. C., and Jones, K. L.: Ecological resistance in urban streams: the role of natural and legacy attributes, Freshw. Sci., 35(1), 380-397, https://doi.org/10.1086/684839, 2016.

580 Voli, M. T., Wegmann, K. W., Bohnenstiehl, D. R., Leithold, E., Osburn, C. L., and Polyakov, V.: Fingerprinting the sources of suspended sediment delivery to a large municipal drinking water reservoir: Falls Lake, Neuse River, North Carolina, USA. J. Soils Sediments, 13(10), 1692-1707, https://doi.org/10.1007/s11368-013-0758-3, 2014.

Wakelin, S. A., Colloff, M. J., and Kookana, R. S.: Effect of wastewater treatment plant effluent on microbial function and community structure in the sediment of a freshwater stream with variable seasonal flow. Appl. Environ. Microb., 74(9), 585 2659-2668, https://doi.org/10.1128/aem.02348-07 , 2008.

Wellen, C., Arhonditsis, G. B., Labencki, T., and Boyd, D.: A Bayesian methodological framework for accommodating interannual variability of nutrient loading with the SPARROW model. Water Resour. Res., 48(10), https://doi.org/10.1029/2012WR011821, 2012.

Wollheim, W. M., Pellerin, B. A., Vörösmarty, C. J., and Hopkinson, C.S.: N retention in urbanizing headwater 590 catchments, Ecosystems, 8(8), 871-884, https://doi.org/10.1007/s10021-005-0178-3, 2005.

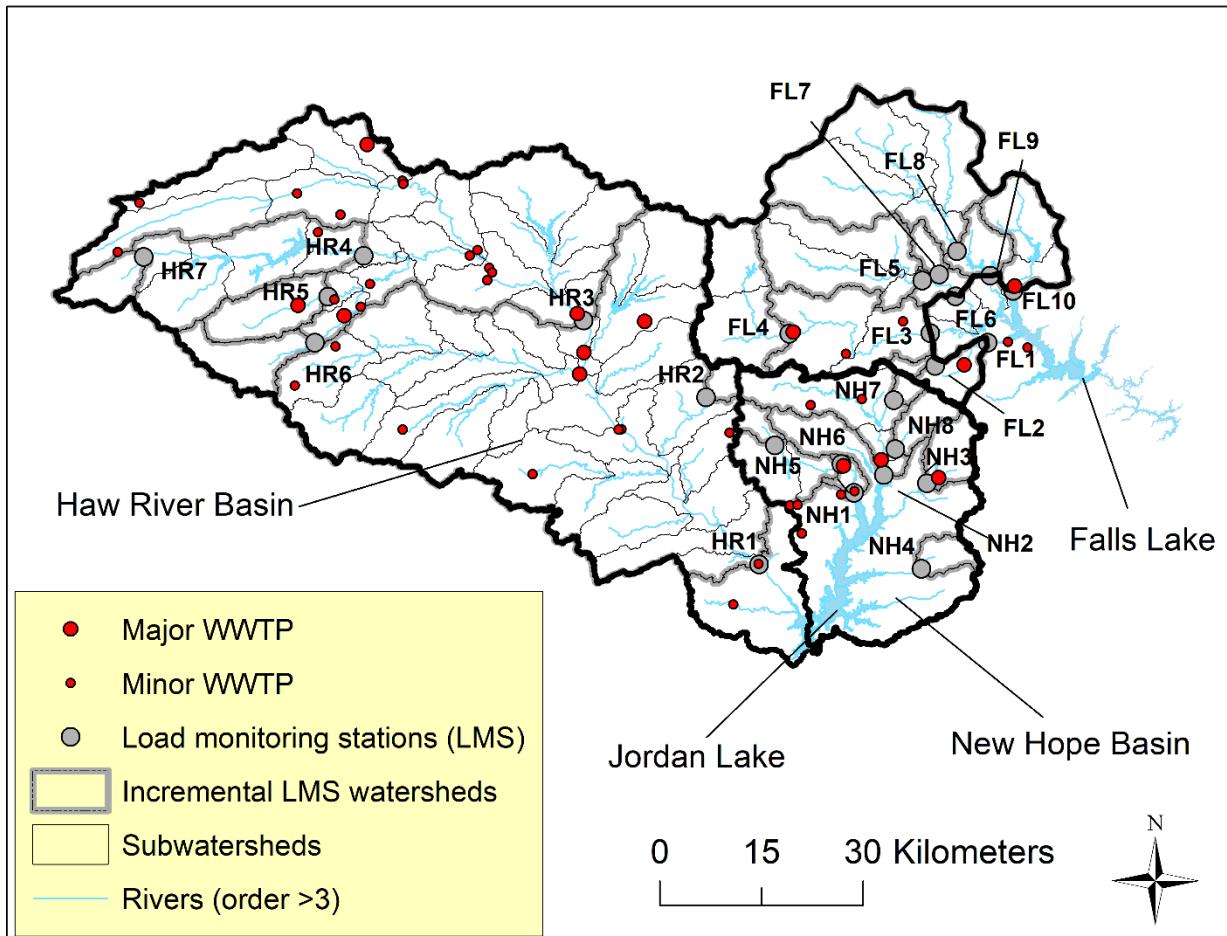
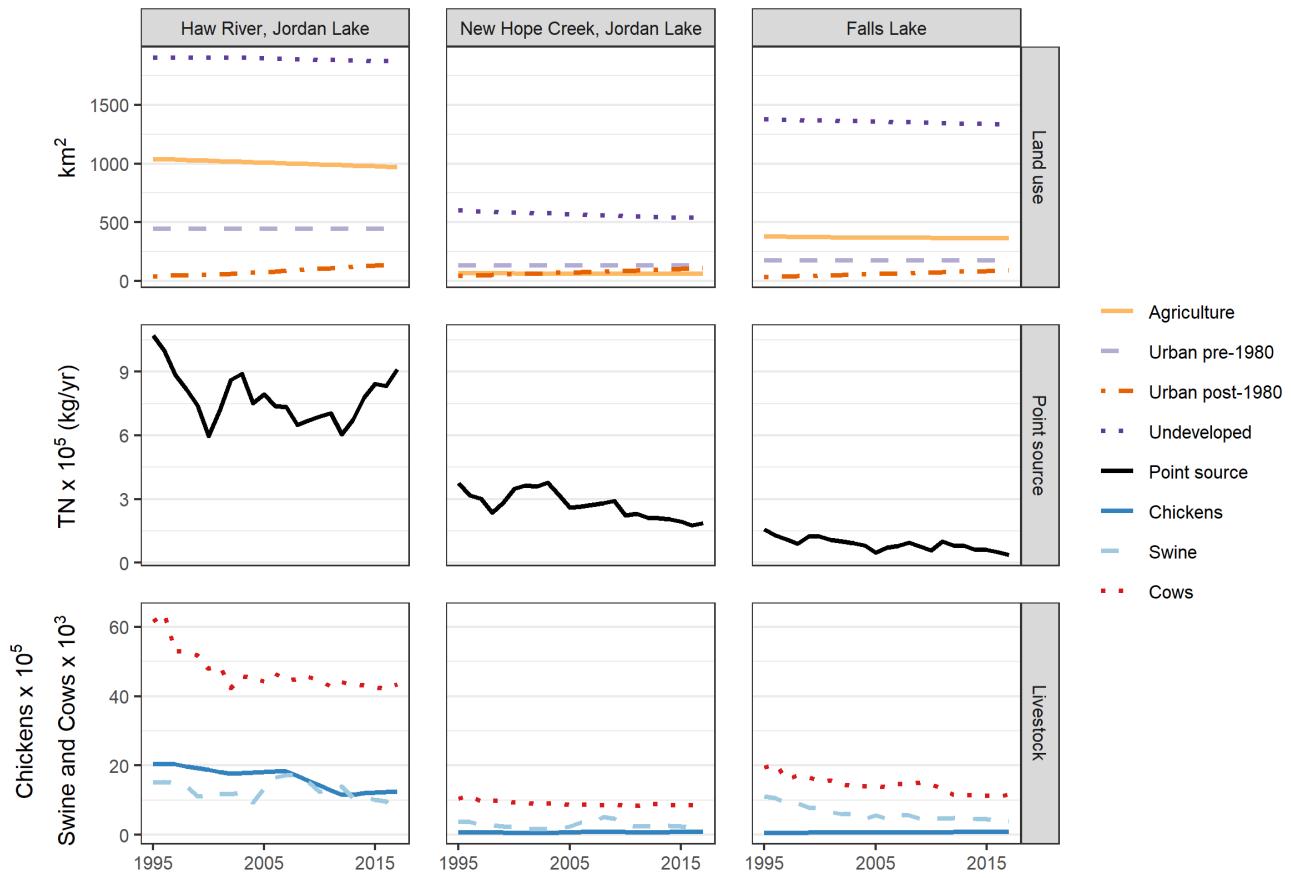
Xia, Y., Weller, D. E., Williams, M. N., Jordan, T. E., & Yan, X.: Using Bayesian hierarchical models to better understand nitrate sources and sinks in agricultural watersheds, Water Res., 105, 527-539, https://doi.org/10.1016/j.watres.2016.09.033, 2016.

595

Figure 1: The load monitoring sites (LMSs) in Jordan and Falls Lake watersheds shown with their incremental watersheds. Also shown are the 79 subwatersheds along with point sources (major and minor wastewater treatment plants (WWTP)). Major basins are delineated by thick black lines.

600   **Figure 2: Land use, point sources, and livestock trends from 1994-2017 in the Haw River, New Hope, and Falls Lake basins.**

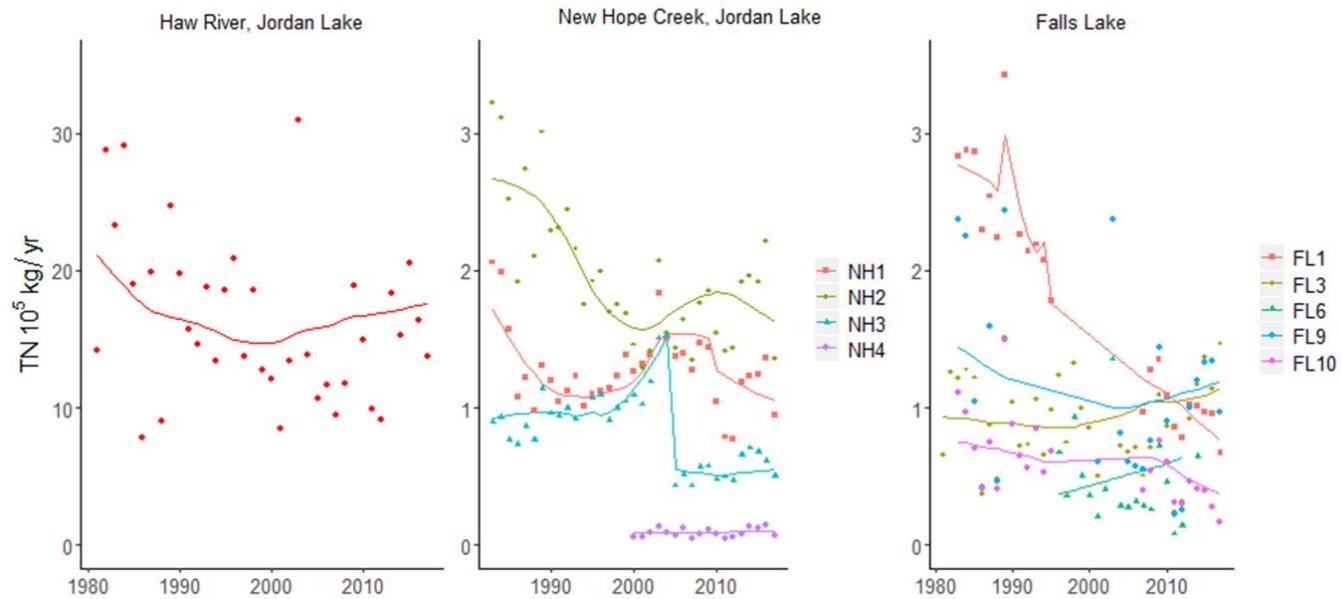



Figure 3: Weighted-regression on Time, Discharge, and Season (WRTDS) annual nutrient loading estimates (points) and flow-normalized estimates (lines) for TN in the Haw River (HR) and New Hope Creek (NH) basins of Jordan Lake (JL) and Falls Lake (FL). For clarity, results are only shown for the most downstream load monitoring site (LMS) of each tributary to Jordan and Falls Lake. The only downstream LMS for Haw River is HR1. WRTDS loading estimates for other LMSs are provided in supporting information (Fig. S4).

605

610

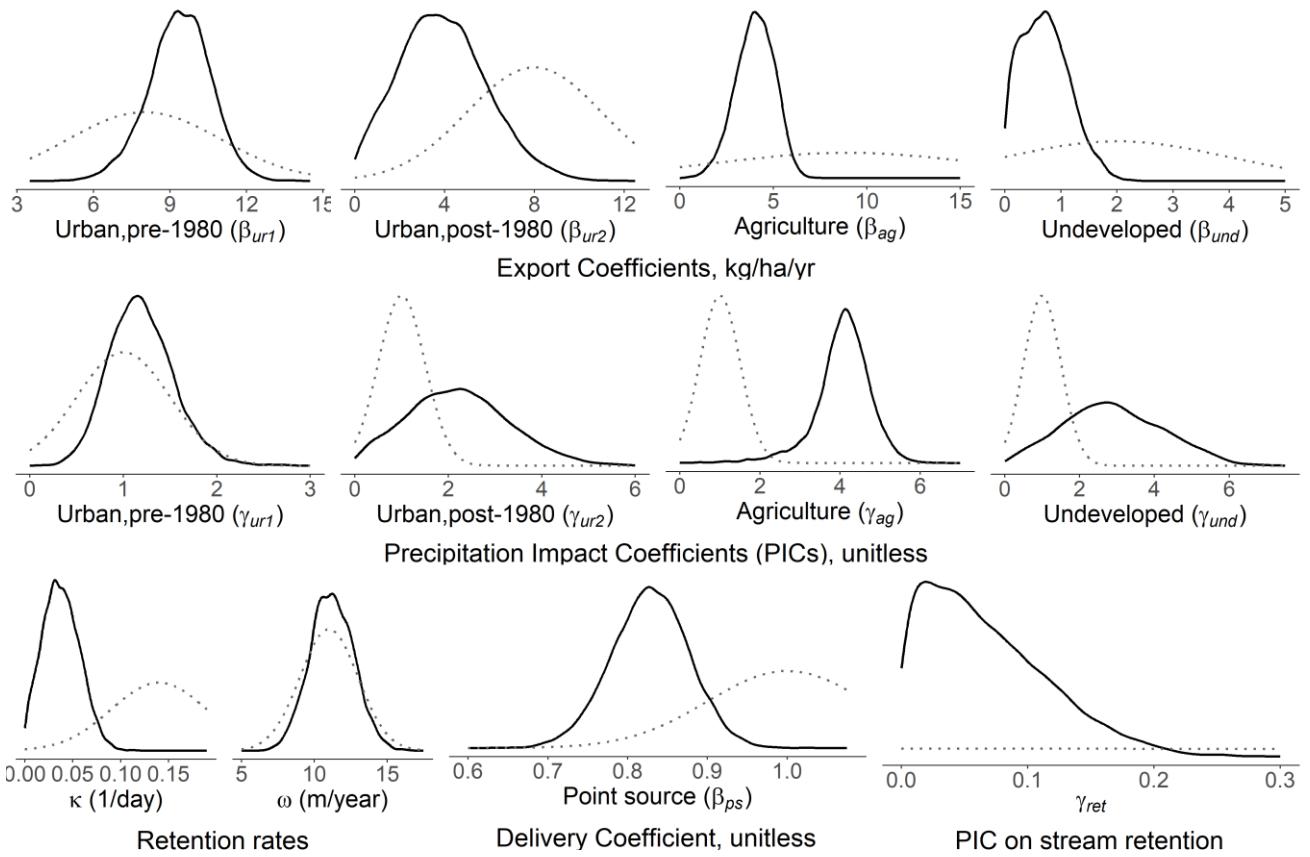



Figure 4: Prior (dotted lines) and posterior (solid lines) distributions for selected model parameters. Note that priors and posteriors are provided for all parameters in Tables 4 and S3.

615

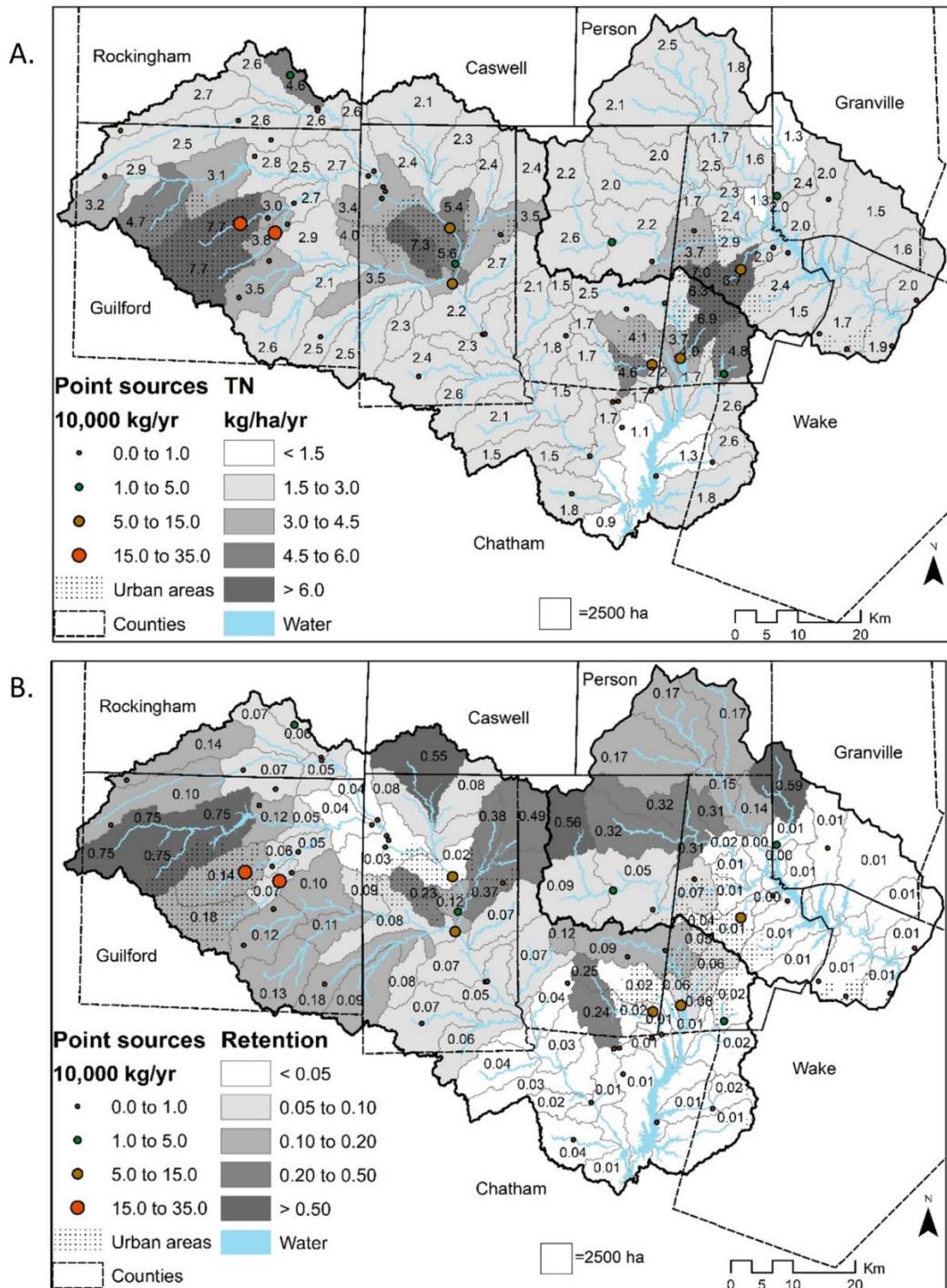



Figure 5: TN export (A) from land use and livestock by subwatershed; fraction of TN export from each subwatershed that is retained (B) in streams and reservoirs prior to reaching Jordan and Falls Lakes. Point source loads are shown separately as dots.

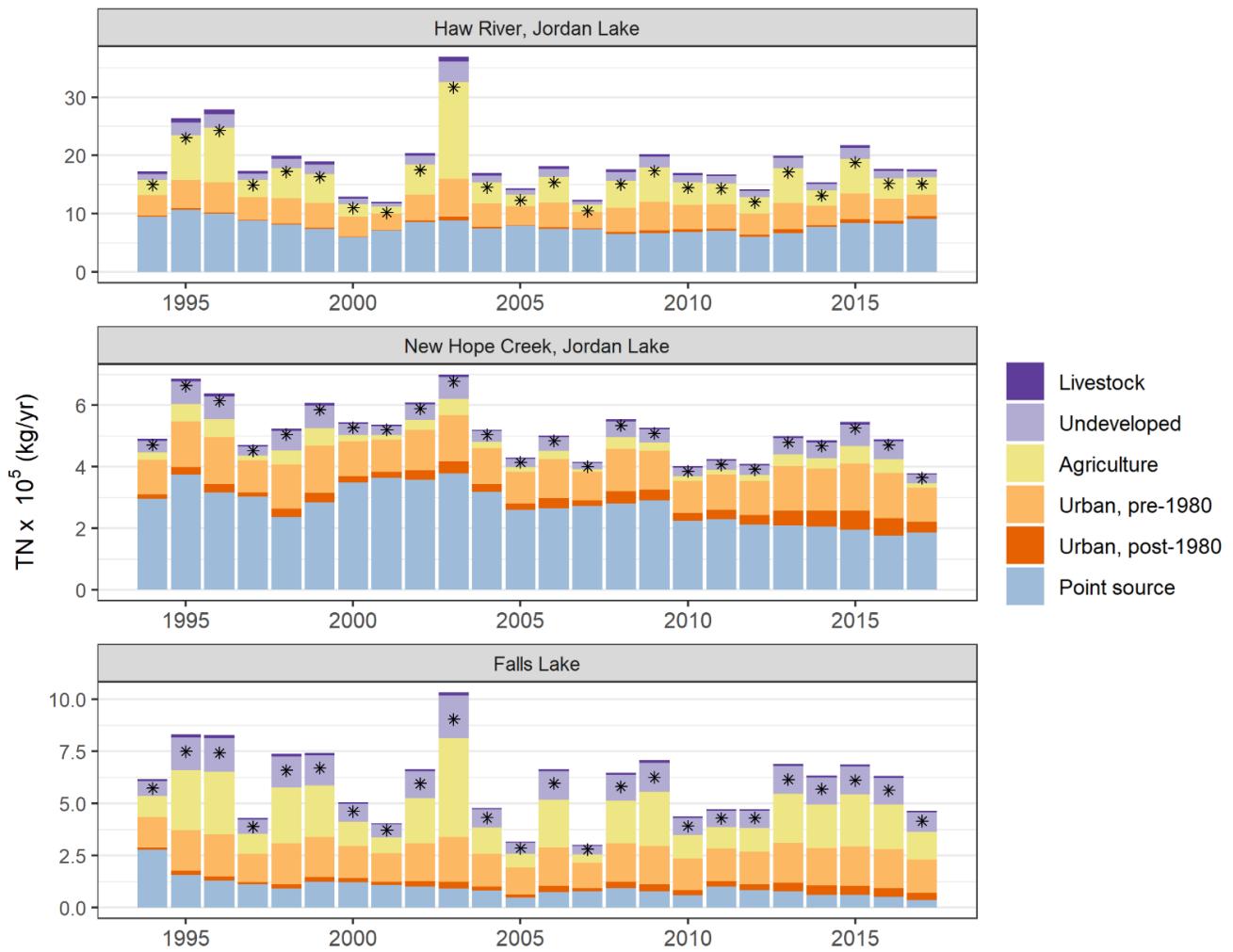



Figure 6: Total nitrogen export by year and major basin separated by source. The star (*) represents the total TN load that reached Jordan or Falls Lake.

620

Table 1: Load monitoring stations (LMSs) located in the Jordan (JL) and Falls Lake (FL) basins along with their complete drainage areas. LMSs belong to either New Hope Creek (NH) or Haw River basins of JL or FL. Years of record corresponds to time that loadings could be estimated (i.e., when daily flow and monthly water quality sampling was performed). The number of total nitrogen (TN) samples available is also shown.

LMS	Name	Res	Drainage area (km^2)	Years of record	# TN samples
NH1	Morgan Creek, Jordan Lake	JL	121.4	1994-2017	578
NH2	New Hope Creek	JL	203.9	1994-2017	575
NH3	Northeast Creek	JL	53.6	1996-2017	430
NH4	White Oak Creek	JL	31.1	2000-2017	106
NH5	Morgan Creek, White Cross	JL	21.4	2000-2017	116
NH6	Morgan Creek, Chapel Hill	JL	103.2	2001-2013	141
NH7	Sandy Creek, Cornwallis	JL	12.1	2009-2017	133
NH8	Third Fork Creek	JL	41.2	2009-2017	107
HR1	Haw River, Bynum	JL	3296.4	1994-2017	590
HR2	Cane Creek	JL	19.6	1989-2017	227
HR3	Haw River, Burlington	JL	1562.1	1994-2017	268
HR4	Reedy Fork, Gibsonville	JL	316.6	1981-1986 2001-2017	341
HR5	N. Buffalo Creek	JL	96.2	1999-2017	394
HR6	S. Buffalo Creek	JL	88.6	2000-2017	343
HR7	Reedy Fork, Oak Ridge	JL	53.4	2001-2017	255
FL1	Ellerbe Creek, Gorman	FL	54.8	2006-2017	280
FL2	Ellerbe Creek, Murray	FL	11.2	2009-2013	100
FL3	Eno River, Durham	FL	367.2	1994-2000 2004-2017	375
FL4	Eno River, Hillsborough	FL	171.0	1990-2017	223
FL5	Little River, Orange Factory	FL	202.7	1988-2000 2005-2017	381
FL6	Little River, Fairntosh	FL	246.4	1996-2011	196
FL7	Mountain Creek	FL	20.8	1995-2011	156
FL8	Flat River, Bahama	FL	385.9	1981-2011	472
FL9	Flat River, Dam	FL	434.4	1983-1990 2003-2017	225
FL10	Knap of Reeds Creek	FL	111.4	2006-2017	142

Table 2: Posterior distributions and 95% credible intervals (CI) of urban export coefficients (EC) split by age and development density (low-density (LD) vs. high-density (HD) urbanization). The probability (P) that older urban lands (or HD urbanization) export more nitrogen than other urban lands was calculated by comparing Bayesian posterior draws. R2 represents the ability of the model to predict temporal variability of loading at each LMS. Mean R2 was determined by averaging the R2 of all 25 LMSs. HD vs. LD (1) compares high-density residential vs. other urban lands while HD vs. LD (2) defines HD as high-density residential, industrial, and commercial lands.

Case	Export Coefficient (EC)	Mean (TN/ha/yr)	95% CI	P(EC1 > EC2)	Mean R^2
A	1. Pre 1980 urban	9.4	7.3-11.3	98%	0.476
	2. Post 1980 urban	3.9	0.9-7.3		
B	1. Pre 2000 urban	8.1	6.0-9.9	71%	0.471
	2. Post 2000 urban	6.5	2.3-10.8		
C	1. HD residential	8.2	4.5-12.0	57%	0.439
	2. Other urban	7.6	4.9-10.1		
D	1. All HD urban	7.9	5.6-10.0	72%	0.468
	2. Other urban	6.5	3.0-10.0		

Table 3: Mean posterior parameter estimates for export and delivery coefficients (β; EC, DC), retention rates (κ, ω), and precipitation impact coefficients (γ; PIC) along with 95% credible intervals (CI). Note that subscripts are same as defined in Eq. (4).

EC, DC, and retention			PIC		
Parameter	Mean	95% CI	Parameter	Mean	95% CI
β_{ag}	4.0	2.3-5.6	γ_a	4.0	2.8-5.1
β_{ur1}	9.4	7.3-11.3	γ_{ur1}	1.2	0.7-1.8
β_{ur2}	3.9	0.9-7.3	γ_{ur2}	2.2	0.5-4.1
β_{und}	0.7	0.1-1.5	γ_{und}	2.9	0.8-5.2
β_{ch}	0.01	0-0.02	γ_{ch}	2.0	0.4-3.9
β_{sw}	0.04	0.01-0.07	γ_{sw}	2.0	0.3-3.8
β_{cw}	0.52	0.06-0.95	γ_{cw}	1.9	0.3-3.8
β_{ps}	0.83	0.73-0.92	γ_{ret}	0.07	0-0.16
κ	0.04	0.01-0.07	μ_γ	1.8	1.1-2.4
ω	11.2	8.7-13.7	σ_γ	1.1	0.7-1.6
σ_ε	0.07	0.07-0.08			
σ_{LMS}	1.34	0.90-1.91			