https://doi.org/10.5194/hess-2020-423
Preprint. Discussion started: 14 September 2020

Hydrology and Earth System Sciences
Open Access
Discussions

Seasonal watershed-scale influences on nitrogen concentrations across the Upper Mississippi River Basin

Michael L. Wine[1,2], Heather E. Golden[3], Jay R. Christensen[3], Charles R. Lane[3], Oleg Makhnin[4]

[1]Oak Ridge Institute for Science and Education @ U.S. Environmental Protection Agency, Office of Research and Development, Cincinnati, 45268, United States
[2]Currently Independent Scientist, Denver, 80222, USA
[3]U.S. Environmental Protection Agency, Office of Research and Development, Cincinnati, 45268, United States
[4]Department of Mathematics, New Mexico Institute of Mining and Technology, Socorro, 87801, United States

Correspondence to: Michael L. Wine (mlw63@cornell.edu)

Abstract. Humanity's footprint on Earth systems has engendered water quality impoverishment in streams, lakes, and coastal waters globally. In agricultural areas, stream nitrogen concentrations are often high where excess nitrogen fertilization and wetland loss via artificial drainage degrade water quality. While the watershed-scale influence of fertilization and wetland loss on annual nitrogen loads has been studied, little is known about the watershed-scale effects of these wetland losses at seasonal time scales. Here we apply machine learning and linear statistical analyses in a big data framework to improve understanding of the role wetlands play in influencing the seasonality of down-gradient water quality. We confirm the seasonal role of wetlands in improving water quality at the watershed scale and uncover evidence demonstrating the importance of contemporary watershed nitrogen inputs to in-stream total nitrogen concentrations [TN]. We observe that in the Upper Mississippi River Basin, United States, after the application of spring fertilizers, [TN] drops by 70% from June to September suggesting the importance of seasonal nutrient loading. Our data mining approach affords exploration of the potential influence of numerous landscape and wetland hydrologic processes on [TN], some of which are shown to exert seasonal influence. Our counterfactual analysis—in which wetlands are restored to their historic extent—points to the substantial water quality benefits of wetland restoration, including particular water quality improvements in the spring when [TN] are highest. Water quality benefits due to wetland restoration would make water safer for human consumption and improve the security of aquatic ecosystems.

1 Introduction

Numerous rivers (Meybeck and Helmer, 1989), lentic inland waters (Brooks et al., 2016), and receiving coastal waters (Diaz and Rosenberg, 2008) are in the throes of a water quality crisis largely driven by increasing rates of anthropogenic nutrient loading since the 1850s (Vitousek et al., 1997) to sustain Earth's human population of 7.8 billion. With projections of continued population growth (Gerland et al., 2014) requiring expansion or further intensification of agriculture, in the absence of dramatic measures, already realized environmental degradation will persist or increase in extent or severity (Liu et al., 2012). Water quality degradation due to excess nutrient loads can lead to deterioration of ecosystems, "impoverishment of aquatic biodiversity" (Bogardi et al., 2012), harmful algal blooms (Brooks et al., 2016), and hazards to human health (Falkenmark, 1990;Bouwer, 2000). In coastal waters impacted by eutrophication, hypoxia-driven habitat compression may hinder life cycle functions of pelagic species (Diaz and Rosenberg, 2008). Integrating the economic consequences of those excess nitrogen (N) impacts yields a staggering economic burden—on the order of hundreds of billions USD per year (Houlton et al., 2019;Compton et al., 2011;Sutton et al., 2011;Sobota et al., 2015), a cost that might be partially offset through wetland restoration (Rankinen et al., 2014;Hey, 2002;Houlton et al., 2019).

At the global scale, 64% of inorganic nitrogen export originates from anthropogenic sources, and 54% occurs as a consequence of diffuse agricultural inputs (Seitzinger et al., 2005), though urban contributions are important as well (Chen et al., 2016). At regional scales as much as 78% of N loading can be a consequence of agricultural fertilization (Compton et al., 2019). Riverine export rates have been observed ranging from 5% (David et al., 2010) to 38% (Compton et al., 2019) of N inputs. While some have attributed this observed nitrogen loading in part to legacy effects (Van Meter et al., 2018;Basu et al., 2010), others have attempted to distinguish between legacy versus contemporary loading hypotheses and have found stronger evidence for the role of ongoing large-scale fertilization (Ballard et al., 2019;Stackpoole et al., 2019). However, in the absence of large-scale critical experimentation (Platt, 1964) the relative importance of possible legacy effects across different systems remains uncertain.

https://doi.org/10.5194/hess-2020-423
Preprint. Discussion started: 14 September 2020

Hydrology and Earth System Sciences
Open Access
Discussions

Nitrogen, along with phosphorus, is a key limiting nutrient that, in excess, contributes to observed widespread eutrophication in coastal (Ryther and Dunstan, 1971), riverine (Dodds and Smith, 2016), and lacustrine (Conley et al., 2009) environments. Nitrogen-laden leachate and surface runoff that flows downgradient toward rivers, lakes, and oceans may first intersect connecting floodplain (Noe and Hupp, 2005;Sanchez-Perez et al., 2003) and/or non-floodplain wetlands (Lane et al., 2018;Mushet et al., 2015). In the anaerobic sediments underlying these wetlands, biologically available nitrogen in the form of nitrate (NO_3^-) may—in the presence of labile carbon and anoxic conditions for microbial activity (Soares, 2000)—be converted to nitrogen gas (N_2) by the denitrification process (Sanchez-Perez et al., 2003). Evidence has therefore emerged that wetlands may substantially reduce nitrogen loading of surface waters at the watershed scale (Hansen et al., 2018;Quin et al., 2015;Fisher and Acreman, 2004;Golden et al., 2019). The efficacy of wetlands in reducing nitrogen loading has been observed across a range of flow conditions and seasons (Uuemaa et al., 2018).

Investigations of fertilizer inputs and wetland impacts on nitrogen loading are commonly conducted at an annual timescale (Basu et al., 2010;Thompson et al., 2011;Van Meter et al., 2016;Van Meter et al., 2018, 2019;Golden et al., 2019). Yet temporal variability is critically important because it influences the timing and frequency of events in which nutrient concentrations exceed levels safe for human consumption (e.g., 10 mg l^{-1} in the US; 4.4 mg l^{-1} in Germany) and ecological integrity. Despite recent advances in understanding the role of wetlands in influencing water quality across spatial scales, improved understanding of the seasonal variability remains a priority (Bloschl et al., 2019).

Golden et al. (2019) suggest the possible role of big data in improving understanding of the influence of non-floodplain wetlands on water quality. To deal with the challenges associated with big data, the hydrologic science community has tentatively and successfully tested machine learning (Shen, 2018;Tyralis et al., 2019), a data driven approach that contrasts with and supplements the normative physically based methods. Application of machine learning to seasonal water quality data and process-based drivers of nutrient loads, presents an opportunity for discerning the seasonal dynamics of wetlands as they relate to watershed-scale nitrogen conditions.

In this paper, we explore the seasonal role nitrogen inputs and wetlands play in influencing downgradient water quality at the watershed scale. To advance this goal we ask: 1) What is the role of watershed characteristics in mediating intra-annual stream TN concentrations ([TN]) at the catchment to large watershed scales? 2) To what extent is restoration of historic wetland distributions a viable means to improve intra-annual stream [TN] and how might water quality improvements be seasonally dependent? To answer these questions, we capitalize on recently developed novel databases of wetland metrics that reflect the structural and functional characteristics of hydrological flowpaths into and out of wetlands (Mengistu et al, In Revision; Leibowitz et al., In Review), in addition to distributed measurements of streamflow and [TN]. We combine machine learning (random forest) with a linear mixed effects model to assess seasonal variability of [TN] in watersheds across the Upper Mississippi River Basin (UMRB) in the Midwestern US. The results of our combined machine learning and linear statistical approach points to the key drivers of variations in [TN] across seasons in the UMRB and provide insights on how machine learning may be used in future watershed-scale water quality analyses.

2 Methods

2.1 Study area

The Mississippi is the longest river in the United States, originating at Lake Itasca, Minnesota. The study area consists of the Upper Mississippi River basin (UMRB; 492,000 km^2; Figure 1), the largest contributor of residual nitrogen to the Mississippi River basin (Burkart and James, 1999;Qi et al., 2020). The UMRB consists principally of the Great Plains, northern forests, and eastern temperate forests ecoregions (Omernik and Griffith, 2014). The surficial geology of the Upper Mississippi River basin is dominated by thick silty glacial till sediments interspersed with thinner units (Soller and Reheis, 2004). Additionally, thick coarse-grained proglacial sediments are present, typically toward the basin's northern extent. Precipitation averages 920 mm yr^{-1}, two-thirds of which falls in spring and summer. Precipitation varies spatially from a low of 600 mm yr^{-1} in the northwest to a high of 1200 mm yr^{-1} in the southeast (Daly et al., 2008). Potential evaporation increases from a low in January to a peak in July. Elevation, which serves as a hydraulic driver in topographically driven flow regimes, ranges from 520 m in the northeast to 140 m in the southeast.

Overlaid on climate dynamics and natural physiography, human alterations in the form of conversion of perennial vegetation to seasonal crops (Zhang and Schilling, 2006) and artificial drainage have pervasively amplified the hydrologic cycle by increasing stream discharge (Blann et al., 2009;Belmont et al., 2011;Schottler et al., 2014). The Upper Mississippi River is also influenced by the presence of numerous locks (Gramann et al., 1984). Integrating the natural environmental conditions with human alterations yields a hydrologic regime in which

https://doi.org/10.5194/hess-2020-423
Preprint. Discussion started: 14 September 2020

specific discharge (the quotient of streamflow volume and contributing area) increases throughout the winter to a peak in the spring (May), decreases through the summer, and remains low in the fall.

The UMRB drains some of the continent's most fertile arable land, which is predominantly in either corn or soybean production. Because of these and other land uses a variety of nutrients and contaminants have been observed in the waters of the Mississippi River. The majority of fertilizer is applied in the spring prior to or after planting, though dry fall fertilizer applications range from 0-25% of the total annual amount (Cao et al., 2018). Nutrient loading has been exacerbated by land management practices, such as artificial drainage, that enhance discharge in the Upper Mississippi River (Schilling et al., 2010). River sediment cores have revealed an order of magnitude increase of sediment deposition, since the 1830s (Engstrom et al., 2009). In the UMRB mean N fertilizer use rate in 2015 was 49 (maximum = 173) kg N ha^{-1} $yr^{-1,}$ corresponding to $2.4 \cdot 10^6$ t N yr^{-1} (Cao et al., 2018). The use rate within the UMRB exceeds NO_3-N loads discharged from the Mississippi River into the Gulf of Mexico as reported by Van Meter et al. (2018). Within the Upper Mississippi River, water quality degradation is known to impact aquatic communities (Houser and Richardson, 2010). Excess nutrients—including a tripling of NO_3^-N input to the Gulf of Mexico (Goolsby et al., 2001;McIsaac et al., 2002)—are credited with causing the dead (hypoxic) zone in the Gulf of Mexico, which has been measured at 20,700 km^2 in extent (Rabotyagov et al., 2010) and has raised questions about the current approaches to improving water quality in the Mississippi River (McLellan et al., 2015).

2.2 Water quality data

We obtained quality controlled stream [TN] (unfiltered total nitrogen as N) data (1995-2007) from the SPARROW (Spatially Referenced Regression on Watershed Attributes) Major River Basin 3 water quality modeling group, which compiled data from federal, state, and local government monitoring (Saad et al., 2011). Sites included at least 25 stream [TN] measurements per site (over the course of the 13-year sampling window) distributed throughout the year to ensure representation of all seasons. This resulted in a total of 6895 [TN] measurements divided amongst 82 sites (Table S1), all of which included corresponding discharge measurements. [TN] values in this dataset range from 0.1 to 25.1 mg l^{-1} (median=3.6). These measurements are taken from streams draining watersheds of median area 2,580 km^2 (mean=7,229), ranging from small catchments (45 km^2) to large watersheds (52,048 km^2).

2.3 Derivation of variables

To describe the variation in our response variable [TN], we used static (i.e., time-invariant) watershed-scale predictor variables developed in recent work (Mengistu et al., in revision) that aimed at quantifying watershed and wetland characteristics as well as the structural and functional attributes of flowpaths between wetlands and rivers. The variables (Table S2) reflect potentially important elements of watershed-scale nutrient cycling, including source (e.g., nutrient loading via cultivated areas), sink (e.g., denitrification via wetland and open water areas), and transport processes (e.g., soil types of overland flowpaths).

Specifically, for the 82 watersheds, the variables describe watershed characteristics (e.g., land cover, watershed area) and average watershed-scale wetland-to-stream flowpath characteristics (e.g., wetland-to-stream flowpath Mannings values or maximum soil porosity along that flowpath) that are intended as proxies for hydraulic and hydrologic processes. The variables represent source contributions <u>to</u> wetlands and the structural characteristics of hydrological flowpaths <u>from</u> wetlands to the nearest streams. For example, maximum porosity along the flowpath (a structural characteristic) is a proxy for infiltration and conversion of overland flow to shallow and/or deep subsurface flows, which attenuate TN from reaching the stream (functional characteristics). The derivation and detailed lists of these variables are described in detail in Mengistu et al. (in revision).

The static variables are supplemented by time varying values developed in this study—monthly soil NO_3-N (Wu and Liu, 2012), daily discharge (from USGS gages; see Table S1), monthly wetness index (the quotient of spatially averaged precipitation and potential evaporation) derived from PRISM climate data (Daly et al., 2008), year of [TN] sampling, and day of year of [TN] sampling. Potential evaporation (PET) was estimated following Hargreaves (1994), where daily minimum and maximum temperatures were extracted from PRISM for each watershed. Soil NO_3-N consists of one value for each month, estimated from a Soil and Water Assessment Tool (SWAT) simulation for the Iowa River basin, which is located within UMRB (Wu and Liu, 2012). In the absence of watershed-specific information on temporal evolution of soil NO_3-N (kg ha^{-1}), the secular values from Wu and Liu (2012) were allowed to remain uniform across watersheds. Areal extent calculations were performed in the Albers Equal Area projection for the conterminous US. All geospatial analyses were conducted in ArcGIS 10.7 and R (Mengistu et al. (in revision)).

2.4 Modeling approach

We used a three-phased model approach to investigate the potential watershed-scale wetland and landscape drivers of seasonal [TN] variability: (1) a nonlinear machine learning approach followed by (2) a linear statistical model with seasonal harmonics (i.e., periodic functions) and (3) counterfactual simulations using the statistical model to ascertain the potential influence of wetland restoration on [TN]. The three-phased approach helped us ascertain which hydrologic processes potentially serve as key drivers of [TN] variability across the UMRB, because the relative importance of the hydrologic processes influencing the seasonal evolution of water quality–including their interactive influence and scale dependence on drivers of [TN]–are unknown *a priori*. Hence, we first interrogated watershed metrics (see Table S2) as predictor variables for [TN] using random forest (see *Machine learning*). Based on the scores of explanatory variables in random forest, we then selected the top candidate variables to develop linear mixed effect (LME) models that capture and reproduce the cyclical seasonal nature of [TN] (see *Mixed effects modeling*). We subsequently linked these [TN] LME models to the seasonal influence of wetlands using an additive variable approach (see *Mixed effects modeling*). Using the linked [TN] LME model, we performed the counterfactual simulations (see *Wetland restoration scenarios*).

2.5 Machine learning

We chose the nonlinear random forest machine learning algorithm to analyze watershed proxies explaining [TN] variability. Random forest quantifies individual variable importance and minimizes overfitting and bias, while remaining "competitive" with other methods (Breiman, 2001) and providing the flexibility of fewer assumptions than traditional linear statistical methods. Despite limitations associated with interpretability (Shen, 2018), random forest has numerous perceived advantages for water resource applications—including accounting for interactions among variables (Cutler et al., 2007), appropriateness for big data applications, and computational efficiency (Tyralis et al., 2019). Random forest is increasingly applied to extract information relevant to water resources within a big data context (Cho et al., 2019), including its use in determining the role of competing drivers in mediating observed water quality across spatial scales (Read et al., 2015).

We used the randomForest package (v. 4.6-14) in R (v. 3.6.1) and tuned the m_{try} parameter to minimize out of bag error, a procedure described in detail by Tyralis et al. (2019) and citations therein. Specifically, out of bag refers to those samples that were withheld from model training for verification purposes. We principally considered candidate variables for the LME models amongst the top five highest ranking in the random forest model results. However, noting the large number of variables considered (n=53; Table S2) and substantial cross-correlation among variables, we supplemented the random forest approach with expert knowledge by removing variables deemed important (by random forest) but not known in the scientific literature as important sources or sinks of nitrogen. (Removed variables are reported in the Results section, below.) In the interest of incisively determining the limits of random forest for the application at hand we ran the algorithm on the dataset as a whole, a random subset of 70% of observations at each measurement location, and on all measurements at a random subset using 70% of sampling locations.

Our final selected variables based on random forest and expert judgement were input into the LME model. Specifically, to build an LME, we relied on a subset of the most important predictors that emerged from random forest, our system understanding, and common metrics of model performance—the Akaike and Bayesian Information Criteria (AIC and BIC, respectively; Helsel et al., 2020).

2.6 Mixed effects modeling

Mixed effects modeling is an extension of simple linear modeling and is widely used in investigating complex water resource problems (Bart, 2016;Wine et al., 2018a;Wine et al., 2018b;Bywater-Reyes et al., 2018;Hurley and Mazumder, 2013;Araujo et al., 2012;Ahearn et al., 2005). This is particularly true for those situations in which part of the natural variability is associated with measured phenomena (i.e., fixed effects) and part of the natural variability results from complex phenomena (i.e., random effects) such as site-specific characteristics. LME also offers tools to overcome heteroscedasticity. In this way, LME relaxes certain assumptions commonly associated with application of simpler methods (Zuur, 2009). Assumptions of LME models include homogeneity of variance and correct model specification—that all relevant terms and interactions are included.

In developing an LME model, we sequentially added variables starting from those assigned highest importance by random forest, ensuring that variables representing key concepts from the advection diffusion reaction equation (ADRE)—including nitrogen sources, fluid advection, reactions (i.e., denitrification), and scale are represented. While we did not intend to solve the transient ADRE (Clairambault, 2013;Oldham et al., 2013) here, we nonetheless considered it briefly as a lens into the physical processes underlying the temporal dynamics of non-conservative solute concentrations (*c*):

https://doi.org/10.5194/hess-2020-423
Preprint. Discussion started: 14 September 2020

Hydrology and Earth System Sciences Discussions

$$\frac{\partial c}{\partial t} = -\nabla \cdot (cu) + \nabla \cdot (D\nabla c) - \frac{\partial q}{\partial t} + S, \qquad \text{Eq. 1}$$

which are controlled by velocity (u), the diffusion coefficient (D), concentration change due to reactions (q), and sources and sinks (S). These dynamics are transient in time (t) and distributed across space. This transient equation allows for seasonal variations in [TN] as a consequence of advective ($\nabla \cdot (cu)$), diffusive ($\nabla \cdot (D\nabla c)$), reactive ($\frac{\partial q}{\partial t}$), or additive ($S$) processes. The quotient of the respective rates of advective transport and diffusion, the Peclet number, is directly related to velocity and length scale and inversely related to diffusivity. Empirically, increasingly chemostatic behavior has indeed been observed as catchment scale increases (Creed et al., 2015), consistent with ADRE. With respect to the source term, the timing of fertilization is expected to yield a key seasonal source whereas crop growth and denitrification serve sink functions. In UMRB it is reasonable to expect that [TN] is influenced by coupled surface and subsurface flow and transport dynamics. Though a physically based approach to understanding seasonal nitrogen dynamics is beyond the scope of this work, this approach nonetheless remains as a reference.

To build our LME models, we considered first order interactions for those variables perceived as having an interactive influence on [TN], i.e., in cases where a predictor's influence on [TN] was expected to depend on the value of another predictor. Model development was ceased—following successive AIC improvements—when the aforementioned key concepts and interactions had been represented.

We anticipated seasonally cyclic behavior in [TN] across the UMRB, particularly because seasonal nutrient loading from agricultural fertilization, as well as from other diffuse nutrient sources and point discharges, is prominent across the UMRB. Helsel et al. (2020) suggest representing this cyclic behavior with periodic functions or harmonics (i.e., sine and cosine) and identifying and addressing cases where cycles shorter than one year may occur. We used this approach for our intra-annual [TN] LME models. However, we did not anticipate that all 82 watersheds would exhibit identical periodic behavior (i.e., amplitude and timing of peak). Hence, we included these periodic functions as random effects, which allows the parameters defining each harmonic to vary by watershed. To ensure that modeling assumptions were met we examined the model residuals of each watershed for heteroscedasticity at the completion of the modeling analyses.

Once the LME models were developed and to further explore factors that may be driving the seasonality in [TN], we calculated the amplitudes of the first harmonic in the final selected LME models and applied Spearman rank correlations between these amplitudes and the watershed variables derived from random forest. This nonparametric approach minimizes distributional assumptions and sensitivity to extreme values and provides further insights into how [TN] varies seasonally with our random forest-based watershed variables.

2.7 Wetland restoration scenarios

To evaluate the extent to which wetland restoration might serve as an effective means of enhancing water quality by decreasing [TN], we engaged in counterfactual modeling. Our counterfactual modeling used the final selected LME model and we altered two variables to create our various wetland restoration scenarios: the proportion of the watershed covered by wetland and the proportion of the watershed covered by cultivated land-cover types. (Note that since wetland area is not strictly independent of cultivated area—due to issues of complementarity—the results must be interpreted in the context of this limitation.) For our counterfactuals, we inferred historic wetland areas from Horvath et al. (2017), which provides fine resolution (30 m) estimates of potential wetland restoration on agricultural lands using soils and topography. In contrast to the widely referenced GLWD, derived from small-scale maps at 1:1,000,000 to 1:3,000,000 spatial resolution, Horvath et al. (2017) relies in part on SSURGO (USDA NRCS Soil Survey Geographic database), which is based principally on 1:24,000 to 1:12,000 spatial resolution. We developed two counterfactual scenarios: (1) 50% wetland restoration to historic conditions and (2) 100% wetland restoration to historic conditions. We assumed that increasing wetland area proportionally decreased cultivated area. Further, potential [TN] reductions were subject to the natural limit that concentration reductions cannot exceed observed concentrations. For each scenario, we predicted [TN] across the 82 UMRB watersheds.

3. Results

3.1 Seasonal variability of [TN] across UMRB

[TN] varies strongly in the UMRB—by a factor of 250— among all sites and across the 13-year study period. Concentrations range from 0.1 mg l^{-1} to 25.1 mg l^{-1}. In 12% of the measurements, [TN] exceeds 10 mg l^{-1}, the maximum contaminant level goal for nitrate as N in the US, and in 44% of the measurements [TN] exceeds 4.4 mg l^{-}

https://doi.org/10.5194/hess-2020-423
Preprint. Discussion started: 14 September 2020

Hydrology and Earth System Sciences
Open Access
Discussions

^{1}, the drinking water standard in Germany (as an example of more stringent global water quality standards). Considering all watersheds, [TN] is lowest in September, increases during the fall when fertilizer is sometimes applied, remains steady through the winter, and increases further in the spring when fertilizer is commonly applied (Figure 2). Between June and September, median [TN] drops by 70% (Figure 2). Though this pattern is generally exhibited by the dataset as a whole, the relatively small watersheds (<350 km^2) have a wide range of [TN] seasonal patterns (Figure 3).

3.2 Random forest

Random forest predicted, with low bias, all training datasets as well as the verification dataset that consisted of a subset of observations from each site (Figure 4). Prior to running random forest, the optimization procedure incrementally reduced out of bag error for [TN] predictions. However, when predictions were attempted by a random forest trained on a random subset of sites, substantial positive biases (i.e., over-prediction) were observed on sites that had been excluded from the training dataset. This is a consistent and known challenge in predictive modeling of complex systems. Ultimately, random forest assigned candidate predictor variables a range of importance values, calculated as the percent increase in mean square error when a variable is withheld.

Discharge was identified as the single best correlate of [TN] (Table 1), though this correlation cannot be interpreted independently of other terms in ADRE. Other important variables that emerged included *year of sampling*, a proxy for interannual climate or land-management variability; *day of year of sampling*, a proxy for intra-annual variability in nitrogen loading and the hydrologic cycle; and *monthly wetness index*, a proxy for drivers of hydrologic dynamics. While wetland metrics tended to achieve low importance ranks, this does not necessarily indicate an inability on their part to influence water quality. Rather, this may instead point to the limited remaining distribution, or historic loss, of wetlands across large areas of the UMRB (Figure 5).

Forest emerged as the most important watershed metric in predicting [TN]. (As anticipated, the nature of the correlation was inverse.) Forests are not expected to serve as a major source of nitrogen, and forests are typically nitrogen sinks only in locations where atmospheric deposition is the most important nitrogen source (e.g., in forest of the Northeastern United States (Goodale et al., 2002)). Therefore, based on expert opinion that forests in the UMRB do not match either criteria, we removed the forest predictor. Its ranking in the random forest model was subsequently replaced by total watershed N inputs, i.e., total annual agricultural inputs of TN plus annual atmospheric deposition of TN (Mengistu et al, in revision), along with average watershed Manning's roughness coefficients for wetland-to-stream flowpaths. Manning's coefficients were estimated based on land cover values, with the highest values occurring in forested areas (Table 2).

3.3 Linear mixed effects: sequential results

Our linear mixed effects modeling aimed initially to reproduce the cyclical variability in [TN] and then to link [TN] to the seasonal influence of wetlands. The resulting idealized model (Table 3, equation 14) represents cyclical trends in [TN]. To determine the final model, we fit 14 sequential, increasingly complex models.

In the first four models (Table 3), our goal was to reproduce the seasonally cyclic behavior in [TN] seen across the study area (e.g., see Figure 2 and Figure 3, and also large watersheds, as seen in Figure 6) with the first and second harmonics (Figure 7, Table 3 equations 1-4). We next included discharge as a random effect (Table 3, Eq. 6), noting the importance of discharge in random forest (Tables 1 and 2) and that concentration-discharge relationships may be direct, inverse, or weak (Figure 8). (Fitting discharge as a random effect allows LME to assign positive or negative coefficients of appropriate magnitudes.) Most commonly, higher [TN] was observed at higher stream stage (particularly in midsized, untiled watersheds), though the strength of this relationship was variable. It is also important to note that spring fertilization occurs coincident with spring rains, which thereby reflects a high source availability during a high flow period—a combination anticipated to enhance solute transport. Attempts at inclusion of additional random effects increased time to model convergence and the possibility of instability.

Several fixed effect terms further improved the model, including year, monthly soil nitrate concentration, cultivated area, wetland area, watershed area, and monthly wetness index. In our final step, interactions between monthly wetness index and monthly soil nitrate concentrations, as well as wetness index and cultivated area, were added and improved model fit. The final model form (Table 3, Equation 14) explained the greatest amount of the observed variability in [TN] (Figure 9). We accounted for observed heteroscedasticity in model residuals by quantifying the variance of model residuals by watershed in an initial model run (of Eq. 14 in Table 3) and assigning these variances in a fixed variance structure for the final model.

Spearman rank correlations of the first harmonic, representing seasonal [TN] dynamics, with watershed variables revealed small amplitudes in association with low watershed nutrient loading (e.g., the presence of grassland (r=-0.39)) and watershed area (r=-0.21), where the convolution of many flowpaths across large watersheds

https://doi.org/10.5194/hess-2020-423
Preprint. Discussion started: 14 September 2020

Hydrology and Earth System Sciences Discussions

is expected to dampen peaks and troughs in [TN] temporal variability (Table 4). The total number and number of wetlands per watershed area (wetland density) were inversely related with the [TN] first harmonic amplitude (r=-0.26). A wetland-to-stream structural characteristic that attenuates flow (Manning's coefficient along the flowpath) was also negatively related to the [TN] harmonic (r=-0.25). Because Manning's values are based on land cover and forested areas are assigned the highest Manning's values, this result might simply reflect smaller amplitudes in predominantly forested watersheds. Impervious areas in wetland drainage areas were positively correlated with the [TN] amplitude (r=0.24). Collectivity, the seasonal results from our LME model and correlations with the first harmonic correspond to our process understanding as derived from first principles, i.e., those embedded in the ADRE. Specifically, we see that the LME model and the associated amplitude of the harmonics captures seasonality in the sources and sinks represented in dynamic water quality models.

3.4 Counterfactual models

The results of our counterfactual model suggest that the potential water quality gains (i.e., [TN] reductions) associated with 50% and 100% wetland restoration are substantial, even in large watersheds of the UMRB (Figure 10). Reductions reach 25 mg l^{-1} with Horvath et al. (2017) wetland estimates. For example, median measured [TN] was 3.6 mg l^{-1}, which decreases to 1.4 mg l^{-1} following simulated (100%) wetland restoration. The decrease in the third quartile of [TN] is more dramatic, dropping from 7 mg l^{-1} to 2.3 mg l^{-1}. If complementarity of cultivated areas with wetland loss is neglected, and full wetland restoration alone is considered, the third quartile decreases only to 5.4 mg l^{-1}, implying that denitrification by wetlands is secondary to the reduction in fertilization, which is of primary importance. The [TN] reductions are largest during the spring when water quality degradation is most pronounced (Figure 10).

It is clear that [TN] reductions are most prevalent when the seasonal [TN] model developed herein is forced by historic wetland distribution from Horvath et al. (2017), which accounts for wetland areas not detected by GLWD and therefore captures a more complete spatial coverage of potential wetlands across the UMRB. Further, while we assume wetland and cultivated areas are complementary, this is typically reasonable given that most wetland loss in the UMRB has been a consequence of drainage to facilitate expansion of cultivated lands.

4 Discussion

4.1 Seasonality of TN

In the presence of recurrent anthropogenic nitrogen loading from fertilizer across the UMRB, which coincides in time with peak discharges (Figure 2), flowpaths were likely activated proximal to the nitrogen source at the land surface, resulting in relatively high [TN] during high discharge periods (Domagalski et al., 2008). Lower flows tended to occur from July through February in the UMRB, out of phase with spring fertilizer application.

Artificial tile drainage decreases the residence time of water in the vadose zone (Danesh-Yazdi et al., 2016) in many of the watersheds throughout the UMRB, thereby facilitating rapid transport of recurrently applied N and engendering high [TN] under baseflow conditions. This is clear in inverse concentration-discharge relationships, particularly in smaller watersheds of the UMRB (see example in Figure 8b). As flows increase in watersheds that are heavily artificially drained, concentrations rapidly decrease, suggesting [TN] is source-limited across a wide range of flow conditions in artificial drained watersheds.

4.2 Wetland restoration and decreased [TN]

While we are unable to separate the relative effects of reduced loading from wetland restoration, our counterfactual modeling of potential effects of wetland restoration (Figure 10) further supports this axiom: the presence of wetlands occurs in association with enhanced water quality (Mitsch et al., 2001;Carpenter et al., 1998;Creed et al., 2017;Cohen et al., 2016;Vymazal, 2007;Jordan et al., 2011;Lane et al., 2018;Lane et al., 2015;Golden et al., 2019;Marton et al., 2015). Our work comports with McLellan et al. (2015) that wetlands, together with other measures, can dramatically improve [TN] in the Mississippi River basin. Whereas much past nutrient removal work has focused on the annual time step (Basu et al., 2010;Cheng and Basu, 2017) or snapshots in time (Hansen et al., 2018), here we show that those water quality improvements associated with the absence of loading and nitrogen removal by wetlands correspond to large reductions in stream [TN] during the spring when concentrations are highest (Figure 10).

Our results regarding the potential of wetland restoration to improve water quality contrast with past assertions that Gulf of Mexico water quality goals may not be achieved because of legacy nitrogen (Van Meter et al., 2018). Our observations of a 70% drop in [TN] between June and September and [TN] increases that tend to occur coincident with fertilization—during the spring and fall—instead appear to be indicative of the importance of contemporary basin-scale nitrogen loading (Van Meter et al., 2020). Based on empirical observations (Figure 2), if excess fertilizer applications ceased, we speculate that a large portion of the observed water quality impairment would resolve at the time scale of months.

Our findings agree with retrospective analyses observing rapid recovery of groundwater nitrate concentrations from initial values exceeding 10 mg l^{-1} to values below this threshold at the timescale of months with cropland to grassland conversion−and if environmental protection was assigned a high priority (Van Meter and Basu, 2015). Specifically, our data analyses suggest measures to reduce [TN] in streams in the UMRB, such as wetland restoration and associated decreases in fertilization, may not be substantially confounded by legacy effects. For these reasons, we suggest that additional critical experiments (Platt, 1964) interrogating the legacy effects of nitrogen are needed.

4.3 Big data, machine learning, and uncertainty

When we engage in statistical modeling or data mining, we do so with two implicit assumptions—that variables are independent of one another and the amount of measurement error is similar across variables. Our analysis questions these widespread assumptions. Instead we observe the risk in data mining of incorrectly selecting a feature (e.g., forest) whose proportional cover in the watershed is correlated—for reasons of complementarity—with the true cause of nitrogen loading (i.e., surplus nutrient inputs in cultivated areas). Why is a correlated variable assigned—by random forest—higher importance than the causal driver? Perhaps, forested areas are simply more readily quantified (i.e., by satellite remote sensing) with lower error (Wickham et al., 2017) than nitrogen loading—whose estimation is subject to greater uncertainties. To be sure (inverse) dependence of nitrogen export on forest cover is well known (Wickham and Wade, 2002). However, the complexity inherent in coupled human-natural systems means that the question of how to best infer cause-effect relations will persist as a formidable challenge (Ferraro et al., 2019;Muller and Levy, 2019).

Following our foray into big data, we agree with McCabe et al. (2017) that big data presents "our community with a unique opportunity to develop new insights that advance fundamental aspects of the hydrological sciences", though big data and machine learning present a double-edged sword (Karpatne et al., 2019). On the one hand an unprecedented quantity of information is available, with the clear potential to improve hydrologic process understanding. On the other hand, this treasure trove of information raises the risk that any number of spurious variables will be assigned high importance by non-physical machine learning algorithms. For this reason, shallow learning algorithms such as random forest are best applied to hypothesis generation and gap filling (of incomplete time series).

Among the most informative aspects of this work was the importance assigned to each of scores of candidate predictor variables. In contrast to the typical approach of testing a single hypothesis, we show here that (if we think of each predictor variable as an alternative hypothesis) scores of candidate hypotheses find some level of support. As an example of the implications of the challenge this presents, our framework does not allow for distinguishing between the effects of denitrification versus the simple absence of fertilization. Whereas structural uncertainty is acknowledged as a "major scientific and engineering challenge" (Renard et al., 2010), the level of support assigned to scores of different processes relevant to hydrology suggests the possibility that conventional modeling approaches may underestimate structural uncertainty.

When the challenges of model structure are combined with those of the parameterization of those models with complex structures, Beven (1993) concludes that "application of distributed hydrological models is more an exercise in prophecy than prediction". However, we agree with Sivapalan (2009) that by changing the question that we pose we can circumvent the uncertainty that arises when uncertain model predictions are used to interrogate the possible effects of secondary, tertiary, or quaternary drivers—often a convenient focus of modern hydrology. For example, in the case of this study, we shifted our research question from a focus on observed wetland influence to wetland restoration scenarios. By asking questions related to the main drivers of unprecedented degradation of natural water resources (e.g., the environmental impacts of growth of humanity's nutrient footprint), we can minimize uncertainty and provide the hydrologic science basis required to remedy the degradation of natural water resources observed to threaten human wellbeing and ecological integrity today.

4.4 Study Implications

We find ourselves at a point in the Anthropocene in which the measures humans take to secure our well-being simultaneously threaten our health and the health of the environment. For example, as we observed here, widespread nitrogen loading and the increase in agricultural land at the expense of wetlands is expected to improve agricultural

https://doi.org/10.5194/hess-2020-423
Preprint. Discussion started: 14 September 2020

Hydrology and Earth System Sciences Discussions
Open Access
EGU

productivity, though it simultaneously has impaired water quality. For this reason, humanity's environmental footprint remains unsustainable (Hoekstra and Wiedmann, 2014;Ehrlich and Holdren, 1971) and diverse approaches are needed to address ever-expanding nutrient-driven water quality issues.

Recently, greater emphasis has been placed on approaching complex water security challenges from an integrative approach that considers perspectives from all relevant disciplines (Zeitoun et al., 2016;Melsen et al., 2018). In this case the challenge involves securing high quality freshwater for humans and aquatic ecosystems in the face of powerful agricultural interests together with an unprecedented human population. Indeed, there is growing acknowledgement within the water resource community of the role of water resource securitization—the characterization of an issue as an existential threat requiring implementation of extraordinary measures—in attaining hydrologic process understanding (Wine, 2020, 2019;Wine and Laronne, 2020;Goulden et al., 2009;Brooks and Trottier, 2010;Yu et al., 2015;Schmeier and Shubber, 2018;Farnum, 2018;Grech-Madin et al., 2018). With respect to water quality in the UMRB, there is a need to examine the role played by water resource securitization in hydrologic process understanding, including how it influences the relative importance of such foci as legacy effects, climate change, or the uncertainty associated with non-point source pollution origin or best management practice siting.

5 Conclusion

Principally as a consequence of ongoing seasonal nitrogen loading in agriculture, [TN] in the UMRB is elevated to the extent that it regularly exceeds the US federal maximum contaminant level for nitrate-nitrogen, with exceedance most likely in June. Expansion of agriculture into former wetland areas together with widespread contemporary nitrogen fertilization (in excess of crop uptake) are primary drivers of the observed degraded conditions whose hazards to human health and biotic security are well known. Here we show that spring rains incident on fertilized agricultural fields increase [TN]. Our model results suggest that restoring historic wetland extent would substantially reduce [TN]—particularly in the spring and early summer when water quality is most severely degraded. This improved water quality, with anticipated benefits for human and biotic health, is caused by the denitrification that takes place in wetland sediments as well as the decrease in fertilization.

Data availability

Data presented in figures will be posted to SciHub following article acceptance. Precipitation and temperature data were obtained from the PRISM Climate group (https://prism.oregonstate.edu/). Solute concentration data are available from the SPARROW (Spatially Referenced Regression on Watershed Attributes) Major River Basin 3 water quality modeling group.

Competing interests: The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Author contribution: Conceptualization, MLW, HEG, JRC, CRL, and OM; Data curation, MLW; Formal analysis, MLW; Investigation, MLW; Methodology, MLW, HEG, JRC, CRL, and OM; Software, MLW and OM; Supervision, HEG, JRC, CRL, and OM; Writing—original draft, MLW; Writing—review & editing, MLW, HEG, JRC, CRL, and OM.

Acknowledgements

We appreciate nutrient concentration data and assistance from Dale Robertson and David Saad (USGS) and spatial data assistance from Ellen D'Amico and Amy Prues. We also thank Grey Evenson and Brent Johnson for helpful suggestions on the paper. This project was supported in part by an appointment to the Science Education Programs at the U.S. Environmental Protection Agency, administered by ORAU through the U.S. Department of Energy Oak Ridge Institute for Science and Education. Mention of trade names or commercial products does not constitute endorsement or recommendation for use. Statements in this publication reflect the authors' professional views and opinions and should not be construed to represent any determination or policy of the U.S. Environmental Protection Agency. This paper has been reviewed in accordance with the U.S. Environmental Protection Agency's peer and administrative review policies and approved for publication.

References

https://doi.org/10.5194/hess-2020-423
Preprint. Discussion started: 14 September 2020

Hydrology and Earth System Sciences Discussions

Ahearn, D. S., Sheibley, R. W., Dahlgren, R. A., Anderson, M., Johnson, J., and Tate, K. W.: Land use and land cover influence on water quality in the last free-flowing river draining the western Sierra Nevada, California, J Hydrol, 313, 234-247, 10.1016/j.jhydrol.2005.02.038, 2005.

Araujo, H. A., Cooper, A. B., Hassan, M. A., and Venditti, J.: Estimating suspended sediment concentrations in areas with limited hydrological data using a mixed-effects model, Hydrol Process, 26, 3678-3688, 10.1002/hyp.8462, 2012.

Ballard, T. C., Michalak, A. M., McIsaac, G. F., Rabalais, N. N., and Turner, R. E.: Comment on "Legacy nitrogen may prevent achievement of water quality goals in the Gulf of Mexico", Science, 365, ARTN eaau8401 10.1126/science.aau8401, 2019.

Bart, R. R.: A regional estimate of postfire streamflow change in California, Water Resour Res, 52, 1465-1478, 10.1002/2014wr016553, 2016.

Basu, N. B., Destouni, G., Jawitz, J. W., Thompson, S. E., Loukinova, N. V., Darracq, A., Zanardo, S., Yaeger, M., Sivapalan, M., Rinaldo, A., and Rao, P. S. C.: Nutrient loads exported from managed catchments reveal emergent biogeochemical stationarity, Geophys Res Lett, 37, Artn L23404 10.1029/2010gl045168, 2010.

Belmont, P., Gran, K. B., Schottler, S. P., Wilcock, P. R., Day, S. S., Jennings, C., Lauer, J. W., Viparelli, E., Willenbring, J. K., Engstrom, D. R., and Parker, G.: Large Shift in Source of Fine Sediment in the Upper Mississippi River, Environ Sci Technol, 45, 8804-8810, 10.1021/es2019109, 2011.

Beven, K.: Prophecy, reality and uncertainty in distributed hydrological modelling, Adv. Water Resour., 16, 41-51, 1993.

Blann, K. L., Anderson, J. L., Sands, G. R., and Vondracek, B.: Effects of Agricultural Drainage on Aquatic Ecosystems: A Review, Crit Rev Env Sci Tec, 39, 909-1001, 10.1080/10643380801977966, 2009.

Bloschl, G., Bierkens, M. F. P., Chambel, A., Cudennec, C., Destouni, G., Fiori, A., Kirchner, J. W., McDonnell, J. J., Savenije, H. H. G., Sivapalan, M., Stumpp, C., Toth, E., Volpi, E., Carr, G., Lupton, C., Salinas, J., Szeles, B., Viglione, A., Aksoy, H., Allen, S. T., Amin, A., Andreassian, V., Arheimer, B., Aryal, S. K., Baker, V., Bardsley, E., Barendrecht, M. H., Bartosova, A., Batelaan, O., Berghuijs, W. R., Beven, K., Blume, T., Bogaard, T., de Amorim, P. B., Bottcher, M. E., Boulet, G., Breinl, K., Brilly, M., Brocca, L., Buytaert, W., Castellarin, A., Castelletti, A., Chen, X. H., Chen, Y. B., Chen, Y. F., Chifflard, P., Claps, P., Clark, M. P., Collins, A. L., Croke, B., Dathe, A., David, P. C., de Barros, F. P. J., de Rooij, G., Di Baldassarre, G., Driscoll, J. M., Duethmann, D., Dwivedi, R., Eris, E., Farmer, W. H., Feiccabrino, J., Ferguson, G., Ferrari, E., Ferraris, S., Fersch, B., Finger, D., Foglia, L., Fowler, K., Gartsman, B., Gascoin, S., Gaume, E., Gelfan, A., Geris, J., Gharari, S., Gleeson, T., Glendell, M., Bevacqua, A. G., Gonzalez-Dugo, M. P., Grimaldi, S., Gupta, A. B., Guse, B., Han, D. W., Hannah, D., Harpold, A., Haun, S., Heal, K., Helfricht, K., Herrnegger, M., Hipsey, M., Hlavacikova, H., Hohmann, C., Holko, L., Hopkinson, C., Hrachowitz, M., Illangasekare, T. H., Inam, A., Innocente, C., Istanbulluoglu, E., Jarihani, B., Kalantari, Z., Kalvans, A., Khanal, S., Khatami, S., Kiesel, J., Kirkby, M., Knoben, W., Kochanek, K., Kohnova, S., Kolechkina, A., Krause, S., Kreamer, D., Kreibich, H., Kunstmann, H., Lange, H., Liberato, M. L. R., Lindquist, E., Link, T., Liu, J. G., Loucks, D. P., Luce, C., Mahe, G., Makarieva, O., Malard, J., Mashtayeva, S., Maskey, S., Mas-Pla, J., Mavrova-Guirguinova, M., Mazzoleni, M., Mernild, S., Misstear, B. D., Montanari, A., Muller-Thomy, H., Nabizadeh, A., Nardi, F., Neale, C., Nesterova, N., Nurtaev, B., Odongo, V. O., Panda, S., Pande, S., Pang, Z. H., Papacharalampous, G., Perrin, C., Pfister, L., Pimentel, R., Polo, M. J., Post, D., Prieto, C., Ramos, M. H., Renner, M., Reynolds, J. E., Ridolfi, E., Rigon, R., Riva, M., Robertson, D. E., Rosso, R., Roy, T., Sa, J. H. M., Salvadori, G., Sandells, M., Schaefli, B., Schumann, A., Scolobig, A., Seibert, J., Servat, E., Shafiei, M., Sharma, A., Sidibe, M., Sidle, R. C., Skaugen, T., Smith, H., Spiessl, S. M., Stein, L., Steinsland, I., Strasser, U., Su, B., Szolgay, J., Tarboton, D., Tauro, F., Thirel, G., Tian, F. Q., Tong, R., Tussupova, K., Tyralis, H., Uijlenhoet, R., van Beek, R., van der Ent, R. J., van der Ploeg, M., Van Loon, A. F., van Meerveld, I., van Nooijen, R., van Oel, P. R., Vidal, J. P., von Freyberg, J., Vorogushyn, S., Wachniew, P., Wade, A. J., Ward, P., Westerberg, I. K., White, C., Wood, E. F., Woods, R., Xu, Z. X., Yilmaz, K. K., and Zhang, Y. Q.: Twenty-three unsolved problems in hydrology (UPH) - a community perspective, Hydrolog Sci J, 64, 1141-1158, 10.1080/02626667.2019.1620507, 2019.

Bogardi, J. J., Dudgeon, D., Lawford, R., Flinkerbusch, E., Meyn, A., Pahl-Wostl, C., Vielhauer, K., and Vörösmarty, C.: Water security for a planet under pressure: interconnected challenges of a changing world call for sustainable solutions, Current Opinion in Environmental Sustainability, 4, 35-43, https://doi.org/10.1016/j.cosust.2011.12.002, 2012.

Bouwer, H.: Integrated water management: emerging issues and challenges, Agric. Water Manage., 45, 217-228, https://doi.org/10.1016/S0378-3774(00)00092-5, 2000.

https://doi.org/10.5194/hess-2020-423
Preprint. Discussion started: 14 September 2020

Breiman, L.: Random forests, Mach Learn, 45, 5-32, Doi 10.1023/A:1010933404324, 2001.

Brooks, B. W., Lazorchak, J. M., Howard, M. D. A., Johnson, M. V. V., Morton, S. L., Perkins, D. A. K., Reavie, E. D., Scott, G. I., Smith, S. A., and Steevens, J. A.: Are Harmful Algal Blooms Becoming the Greatest Inland Water Quality Threat to Public Health and Aquatic Ecosystems?, Environ Toxicol Chem, 35, 6-13, 10.1002/etc.3220, 2016.

Brooks, D., and Trottier, J.: Confronting water in an Israeli-Palestinian peace agreement, J Hydrol, 382, 103-114, 10.1016/j.jhydrol.2009.12.021, 2010.

Burkart, M. R., and James, D. E.: Agricultural-nitrogen contributions to hypoxia in the Gulf of Mexico, J Environ Qual, 28, 850-859, DOI 10.2134/jeq1999.00472425002800030016x, 1999.

Bywater-Reyes, S., Bladon, K. D., and Segura, C.: Relative Influence of Landscape Variables and Discharge on Suspended Sediment Yields in Temperate Mountain Catchments, Water Resour Res, 54, 5126-5142, 10.1029/2017wr021728, 2018.

Cao, P. Y., Lu, C. Q., and Yu, Z.: Historical nitrogen fertilizer use in agricultural ecosystems of the contiguous United States during 1850-2015: application rate, timing, and fertilizer types, Earth Syst Sci Data, 10, 969-984, 10.5194/essd-10-969-2018, 2018.

Carpenter, S. R., Caraco, N. F., Correll, D. L., Howarth, R. W., Sharpley, A. N., and Smith, V. H.: Nonpoint pollution of surface waters with phosphorus and nitrogen, Ecol. Appl., 8, 559-568, 10.2307/2641247, 1998.

Chen, Q., Mei, K., Dahlgren, R. A., Wang, T., Gong, J., and Zhang, M. H.: Impacts of land use and population density on seasonal surface water quality using a modified geographically weighted regression, Sci Total Environ, 572, 450-466, 10.1016/j.scitotenv.2016.08.052, 2016.

Cheng, F. Y., and Basu, N. B.: Biogeochemical hotspots: Role of small water bodies in landscape nutrient processing, Water Resour Res, 53, 5038-5056, 10.1002/2016wr020102, 2017.

Cho, E., Jacobs, J. M., Jia, X. H., and Kraatz, S.: Identifying Subsurface Drainage using Satellite Big Data and Machine Learning via Google Earth Engine, Water Resour Res, 55, 8028-8045, 10.1029/2019wr024892, 2019.

Clairambault, J.: Reaction-Diffusion-Advection Equation, in: Encyclopedia of Systems Biology, edited by: Dubitzky, W., Wolkenhauer, O., Cho, K.-H., and Yokota, H., Springer New York, New York, NY, 1817-1817, 2013.

Cohen, M. J., Creed, I. F., Alexander, L., Basu, N. B., Calhoun, A. J. K., Craft, C., D'Amico, E., DeKeyser, E., Fowler, L., Golden, H. E., Jawitz, J. W., Kalla, P., Kirkman, L. K., Lane, C. R., Lang, M., Leibowitz, S. G., Lewis, D. B., Marton, J., McLaughlin, D. L., Mushet, D. M., Raanan-Kiperwas, H., Rains, M. C., Smith, L., and Walls, S. C.: Do geographically isolated wetlands influence landscape functions?, P Natl Acad Sci USA, 113, 1978-1986, 10.1073/pnas.1512650113, 2016.

Compton, J. E., Harrison, J. A., Dennis, R. L., Greaver, T. L., Hill, B. H., Jordan, S. J., Walker, H., and Campbell, H. V.: Ecosystem services altered by human changes in the nitrogen cycle: a new perspective for US decision making, Ecol Lett, 14, 804-815, 10.1111/j.1461-0248.2011.01631.x, 2011.

Compton, J. E., Goodwin, K. E., Sobota, D. J., and Lin, J.: Seasonal Disconnect Between Streamflow and Retention Shapes Riverine Nitrogen Export in the Willamette River Basin, Oregon, Ecosystems, 10.1007/s10021-019-00383-9, 2019.

Conley, D. J., Paerl, H. W., Howarth, R. W., Boesch, D. F., Seitzinger, S. P., Havens, K. E., Lancelot, C., and Likens, G. E.: Controlling Eutrophication: Nitrogen and Phosphorus, Science, 323, 1014, 10.1126/science.1167755, 2009.

Creed, I. F., McKnight, D. M., Pellerin, B. A., Green, M. B., Bergamaschi, B. A., Aiken, G. R., Burns, D. A., Findlay, S. E. G., Shanley, J. B., Striegl, R. G., Aulenbach, B. T., Clow, D. W., Laudon, H., McGlynn, B. L., McGuire, K. J., Smith, R. A., and Stackpoole, S. M.: The river as a chemostat: fresh perspectives on dissolved organic matter flowing down the river continuum, Can J Fish Aquat Sci, 72, 1272-1285, 10.1139/cjfas-2014-0400, 2015.

Creed, I. F., Lane, C. R., Serran, J. N., Alexander, L. C., Basu, N. B., Calhoun, A. J. K., Christensen, J. R., Cohen, M. J., Craft, C., D'Amico, E., DeKeyser, E., Fowler, L., Golden, H. E., Jawitz, J. W., Kalla, P., Kirkman, L. K., Lang, M., Leibowitz, S. G., Lewis, D. B., Marton, J., McLaughlin, D. L., Raanan-Kiperwas, H., Rains, M. C., Rains, K. C., and Smith, L.: Enhancing protection for vulnerable waters, Nat Geosci, 10, 809-815, 10.1038/ngeo3041, 2017.

Cutler, D. R., Edwards, T. C., Beard, K. H., Cutler, A., and Hess, K. T.: Random forests for classification in ecology, Ecology, 88, 2783-2792, Doi 10.1890/07-0539.1, 2007.

Daly, C., Halbleib, M., Smith, J. I., Gibson, W. P., Doggett, M. K., Taylor, G. H., Curtis, J., and Pasteris, P. P.: Physiographically sensitive mapping of climatological temperature and precipitation across the conterminous United States, Int J Climatol, 28, 2031-2064, 10.1002/joc.1688, 2008.

Danesh-Yazdi, M., Foufoula-Georgiou, E., Karwan, D. L., and Botter, G.: Inferring changes in water cycle dynamics of intensively managed landscapes via the theory of time-variant travel time distributions, Water Resour Res, 52, 7593-7614, 10.1002/2016wr019091, 2016.

David, M. B., Drinkwater, L. E., and Mclsaac, G. F.: Sources of Nitrate Yields in the Mississippi River Basin, J Environ Qual, 39, 1657-1667, 10.2134/jeq2010.0115, 2010.

Diaz, R. J., and Rosenberg, R.: Spreading dead zones and consequences for marine ecosystems, Science, 321, 926-929, 10.1126/science.1156401, 2008.

Dodds, W., and Smith, V.: Nitrogen, phosphorus, and eutrophication in streams, Inland Waters, 6, 155-164, 10.5268/iw-6.2.909, 2016.

Domagalski, J. L., Ator, S., Coupe, R., McCarthy, K., Lampe, D., Sandstrom, M., and Baker, N.: Comparative study of transport processes of nitrogen, phosphorus, and herbicides to streams in five agricultural basins, USA, J Environ Qual, 37, 1158-1169, 10.2134/jeq2007.0408, 2008.

Ehrlich, P. R., and Holdren, J. P.: Impact of Population Growth, Science, 171, 1212-&, DOI 10.1126/science.171.3977.1212, 1971.

Engstrom, D. R., Almendinger, J. E., and Wolin, J. A.: Historical changes in sediment and phosphorus loading to the upper Mississippi River: mass-balance reconstructions from the sediments of Lake Pepin, J Paleolimnol, 41, 563-588, 10.1007/s10933-008-9292-5, 2009.

Falkenmark, M.: Global Water Issues Confronting Humanity, Journal of Peace Research, 27, 177-190, 10.1177/0022343390027002007, 1990.

Farnum, R. L.: Drops of diplomacy: Questioning the scale of hydro-diplomacy through fogharvesting, J Hydrol, 562, 446-454, 10.1016/j.jhydrol.2018.05.012, 2018.

Ferraro, P. J., Sanchirico, J. N., and Smith, M. D.: Causal inference in coupled human and natural systems, P Natl Acad Sci USA, 116, 5311-5318, 10.1073/pnas.1805563115, 2019.

Fisher, J., and Acreman, M. C.: Wetland nutrient removal: a review of the evidence, Hydrol Earth Syst Sc, 8, 673-685, DOI 10.5194/hess-8-673-2004, 2004.

Gerland, P., Raftery, A. E., Ševčíková, H., Li, N., Gu, D., Spoorenberg, T., Alkema, L., Fosdick, B. K., Chunn, J., Lalic, N., Bay, G., Buettner, T., Heilig, G. K., and Wilmoth, J.: World population stabilization unlikely this century, Science, 346, 234, 10.1126/science.1257469, 2014.

Golden, H. E., Rajib, A., Lane, C. R., Christensen, J. R., Wu, Q. S., and Mengistu, S.: Non-floodplain Wetlands Affect Watershed Nutrient Dynamics: A Critical Review, Environ Sci Technol, 53, 7203-7214, 10.1021/acs.est.8b07270, 2019.

Goodale, C. L., Lajtha, K., Nadelhoffer, K. J., Boyer, E. W., and Jaworski, N. A.: Forest nitrogen sinks in large eastern U.S. watersheds: estimates from forest inventory and an ecosystem model, in: The Nitrogen Cycle at Regional to Global Scales, edited by: Boyer, E. W., and Howarth, R. W., Springer Netherlands, Dordrecht, 239-266, 2002.

Goolsby, D. A., Battaglin, W. A., Aulenbach, B. T., and Hooper, R. P.: Nitrogen input to the Gulf of Mexico, J Environ Qual, 30, 329-336, DOI 10.2134/jeq2001.302329x, 2001.

Goulden, M., Conway, D., and Persechino, A.: Adaptation to climate change in international river basins in Africa: a review, Hydrolog Sci J, 54, 805-828, DOI 10.1623/hysj.54.5.805, 2009.

Gramann, J. H., Mcavoy, L., Absher, J. D., and Burdge, R. J.: The Relationship between Commercial and Recreational Use of Navigation Locks on the Upper Mississippi River, Water Resour Bull, 20, 577-582, 1984.

Grech-Madin, C., Doring, S., Kim, K., and Swain, A.: Negotiating water across levels: A peace and conflict "Toolbox" for water diplomacy, J Hydrol, 559, 100-109, 10.1016/j.jhydrol.2018.02.008, 2018.

Hansen, A. T., Dolph, C. L., Foufoula-Georgiou, E., and Finlay, J. C.: Contribution of wetlands to nitrate removal at the watershed scale, Nat Geosci, 11, 127-+, 10.1038/s41561-017-0056-6, 2018.

Hargreaves, G. H.: Defining and Using Reference Evapotranspiration, J Irrig Drain Eng, 120, 1132-1139, Doi 10.1061/(Asce)0733-9437(1994)120:6(1132), 1994.

Helsel, D. R., Hirsch, R. M., Ryberg, K. R., Archfield, S. A., and Gilroy, E. J.: Statistical methods in water resources, Hydrologic Analysis and Interpretation, edited by: USGS, USGS, Reston, Virginia, 458 pp., 2020.

Hey, D. L.: Nitrogen farming: Harvesting a different crop, Restor Ecol, 10, 1-10, DOI 10.1046/j.1526-100X.2002.10100.x, 2002.

https://doi.org/10.5194/hess-2020-423
Preprint. Discussion started: 14 September 2020

Hoekstra, A. Y., and Wiedmann, T. O.: Humanity's unsustainable environmental footprint, Science, 344, 1114-1117, 10.1126/science.1248365, 2014.

Horvath, E. K., Christensen, J. R., Mehaffey, M. H., and Neale, A. C.: Building a potential wetland restoration indicator for the contiguous United States, Ecol Indic, 83, 463-473, 10.1016/j.ecolind.2017.07.026, 2017.

Houlton, B. Z., Almaraz, M., Aneja, V., Austin, A. T., Bai, E., Cassman, K. G., Compton, J. E., Davidson, E. A., Erisman, J. W., Galloway, J. N., Gu, B. J., Yao, G. L., Martinelli, L. A., Scow, K., Schlesinger, W. H., Tomich, T. P., Wang, C., and Zhang, X.: A World of Cobenefits: Solving the Global Nitrogen Challenge, Earths Future, 7, 865-872, 10.1029/2019ef001222, 2019.

Houser, J. N., and Richardson, W. B.: Nitrogen and phosphorus in the Upper Mississippi River: transport, processing, and effects on the river ecosystem, Hydrobiologia, 640, 71-88, 10.1007/s10750-009-0067-4, 2010.

Hurley, T., and Mazumder, A.: Spatial scale of land-use impacts on riverine drinking source water quality, Water Resour Res, 49, 1591-1601, 10.1002/wrcr.20154, 2013.

Jordan, S. J., Stoffer, J., and Nestlerode, J. A.: Wetlands as Sinks for Reactive Nitrogen at Continental and Global Scales: A Meta-Analysis, Ecosystems, 14, 144-155, 10.1007/s10021-010-9400-z, 2011.

Karpatne, A., Ebert-Uphoff, I., Ravela, S., Babaie, H. A., and Kumar, V.: Machine Learning for the Geosciences: Challenges and Opportunities, Ieee T Knowl Data En, 31, 1544-1554, 10.1109/Tkde.2018.2861006, 2019.

Lane, C. R., Autrey, B. C., Jicha, T., Lehto, L., Elonen, C., and Seifert-Monson, L.: Denitrification Potential in Geographically Isolated Wetlands of North Carolina and Florida, USA, Wetlands, 35, 459-471, 10.1007/s13157-015-0633-7, 2015.

Lane, C. R., Leibowitz, S. G., Autrey, B. C., LeDuc, S. D., and Alexander, L. C.: Hydrological, Physical, and Chemical Functions and Connectivity of Non-Floodplain Wetlands to Downstream Waters: A Review, J Am Water Resour As, 54, 346-371, 10.1111/1752-1688.12633, 2018.

Lehner, B., and Doll, P.: Development and validation of a global database of lakes, reservoirs and wetlands, J Hydrol, 296, 1-22, 10.1016/j.jhydrol.2004.03.028, 2004.

Liu, C., Kroeze, C., Hoekstra, A. Y., and Gerbens-Leenes, W.: Past and future trends in grey water footprints of anthropogenic nitrogen and phosphorus inputs to major world rivers, Ecol. Indicators, 18, 42-49, https://doi.org/10.1016/j.ecolind.2011.10.005, 2012.

Marton, J. M., Creed, I. F., Lewis, D. B., Lane, C. R., Basu, N. B., Cohen, M. J., and Craft, C. B.: Geographically Isolated Wetlands are Important Biogeochemical Reactors on the Landscape, Bioscience, 65, 408-418, 10.1093/biosci/biv009, 2015.

McCabe, M. F., Rodell, M., Alsdorf, D. E., Miralles, D. G., Uijlenhoet, R., Wagner, W., Lucieer, A., Houborg, R., Verhoest, N. E. C., Franz, T. E., Shi, J. C., Gao, H. L., and Wood, E. F.: The future of Earth observation in hydrology, Hydrol Earth Syst Sc, 21, 3879-3914, 10.5194/hess-21-3879-2017, 2017.

McIsaac, G. F., David, M. B., Gertner, G. Z., and Goolsby, D. A.: Relating net nitrogen input in the Mississippi River basin to nitrate flux in the lower Mississippi River: A comparison of approaches, J Environ Qual, 31, 1610-1622, DOI 10.2134/jeq2002.1610, 2002.

McLellan, E., Robertson, D., Schilling, K., Tomer, M., Kostel, J., Smith, D., and King, K.: Reducing Nitrogen Export from the Corn Belt to the Gulf of Mexico: Agricultural Strategies for Remediating Hypoxia, J Am Water Resour As, 51, 263-289, 10.1111/jawr.12246, 2015.

Melsen, L. A., Vos, J., and Boelens, R.: What is the role of the model in socio-hydrology? Discussion of "Prediction in a socio- hydrological world", Hydrolog Sci J, 63, 1435-1443, 10.1080/02626667.2018.1499025, 2018.

Meybeck, M., and Helmer, R.: The Quality of Rivers - from Pristine Stage to Global Pollution, Global Planet Change, 75, 283-309, Doi 10.1016/0921-8181(89)90007-6, 1989.

Mitsch, W. J., Day, J. W., Gilliam, J. W., Groffman, P. M., Hey, D. L., Randall, G. W., and Wang, N. M.: Reducing nitrogen loading to the Gulf of Mexico from the Mississippi River Basin: Strategies to counter a persistent ecological problem, Bioscience, 51, 373-388, Doi 10.1641/0006-3568(2001)051[0373:Rnlttg]2.0.Co;2, 2001.

Muller, M. F., and Levy, M. C.: Complementary Vantage Points: Integrating Hydrology and Economics for Sociohydrologic Knowledge Generation, Water Resour Res, 55, 2549-2571, 10.1029/2019wr024786, 2019.

Mushet, D. M., Calhoun, A. J. K., Alexander, L. C., Cohen, M. J., DeKeyser, E. S., Fowler, L., Lane, C. R., Lang, M. W., Rains, M. C., and Walls, S. C.: Geographically Isolated Wetlands: Rethinking a Misnomer, Wetlands, 35, 423-431, 10.1007/s13157-015-0631-9, 2015.

Noe, G. B., and Hupp, C. R.: Carbon, nitrogen, and phosphorus accumulation in floodplains of Atlantic Coastal Plain rivers, USA, Ecol Appl, 15, 1178-1190, Doi 10.1890/04-1677, 2005.

https://doi.org/10.5194/hess-2020-423
Preprint. Discussion started: 14 September 2020

Oldham, C. E., Farrow, D. E., and Peiffer, S.: A generalized Damkohler number for classifying material processing in hydrological systems, Hydrol Earth Syst Sc, 17, 1133-1148, 10.5194/hess-17-1133-2013, 2013.

Omernik, J. M., and Griffith, G. E.: Ecoregions of the Conterminous United States: Evolution of a Hierarchical Spatial Framework, Environ Manage, 54, 1249-1266, 10.1007/s00267-014-0364-1, 2014.

Platt, J. R.: Strong Inference, Science, 146, 347, 10.1126/science.146.3642.347, 1964.

Qi, J. Y., Zhang, X. S., Yang, Q. C., Srinivasan, R., Arnold, J. G., Li, J., Waldholf, S. T., and Cole, J.: SWAT ungauged: Water quality modeling in the Upper Mississippi River Basin, J Hydrol, 584, ARTN 124601 10.1016/j.jhydrol.2020.124601, 2020.

Quin, A., Jaramillo, F., and Destouni, G.: Dissecting the ecosystem service of large-scale pollutant retention: The role of wetlands and other landscape features, Ambio, 44, S127-S137, 10.1007/s13280-014-0594-8, 2015.

Rabotyagov, S., Campbell, T., Jha, M., Gassman, P. W., Arnold, J., Kurkalova, L., Secchi, S., Feng, H. L., and Kling, C. L.: Least-cost control of agricultural nutrient contributions to the Gulf of Mexico hypoxic zone, Ecol Appl, 20, 1542-1555, Doi 10.1890/08-0680.1, 2010.

Rankinen, K., Granlund, K., Etheridge, R., and Seuri, P.: Valuation of nitrogen retention as an ecosystem service on a catchment scale, Hydrol Res, 45, 411-424, 10.2166/nh.2013.239, 2014.

Read, E. K., Patil, V. P., Oliver, S. K., Hetherington, A. L., Brentrup, J. A., Zwart, J. A., Winters, K. M., Corman, J. R., Nodine, E. R., Woolway, R. I., Dugan, H. A., Jaimes, A., Santoso, A. B., Hong, G. S., Winslow, L. A., Hanson, P. C., and Weathers, K. C.: The importance of lake-specific characteristics for water quality across the continental United States, Ecol Appl, 25, 943-955, 10.1890/14-0935.1, 2015.

Renard, B., Kavetski, D., Kuczera, G., Thyer, M., and Franks, S. W.: Understanding predictive uncertainty in hydrologic modeling: The challenge of identifying input and structural errors, Water Resour Res, 46, Artn W05521 10.1029/2009wr008328, 2010.

Ryther, J. H., and Dunstan, W. M.: Nitrogen, Phosphorus, and Eutrophication in the Coastal Marine Environment, Science, 171, 1008, 10.1126/science.171.3975.1008, 1971.

Saad, D. A., Schwarz, G. E., Robertson, D. M., and Booth, N. L.: A Multi-Agency Nutrient Dataset Used to Estimate Loads, Improve Monitoring Design, and Calibrate Regional Nutrient Sparrow Models, J Am Water Resour As, 47, 933-949, 10.1111/j.1752-1688.2011.00575.x, 2011.

Sanchez-Perez, J. M., Vervier, P., Garabetian, F., Sauvage, S., Loubet, M., Rols, J. L., Bariac, T., and Weng, P.: Nitrogen dynamics in the shallow groundwater of a riparian wetland zone of the Garonne, SW France: nitrate inputs, bacterial densities, organic matter supply and denitrification measurements, Hydrol Earth Syst Sc, 7, 97-107, DOI 10.5194/hess-7-97-2003, 2003.

Schilling, K. E., Chan, K. S., Liu, H., and Zhang, Y. K.: Quantifying the effect of land use land cover change on increasing discharge in the Upper Mississippi River, J Hydrol, 387, 343-345, 10.1016/j.jhydrol.2010.04.019, 2010.

Schmeier, S., and Shubber, Z.: Anchoring water diplomacy - The legal nature of international river basin organizations, J Hydrol, 567, 114-120, 10.1016/j.jhydrol.2018.09.054, 2018.

Schottler, S. P., Ulrich, J., Belmont, P., Moore, R., Lauer, J. W., Engstrom, D. R., and Almendinger, J. E.: Twentieth century agricultural drainage creates more erosive rivers, Hydrol Process, 28, 1951-1961, 10.1002/hyp.9738, 2014.

Seitzinger, S. P., Harrison, J. A., Dumont, E., Beusen, A. H. W., and Bouwman, A. F.: Sources and delivery of carbon, nitrogen, and phosphorus to the coastal zone: An overview of Global Nutrient Export from Watersheds (NEWS) models and their application, Global Biogeochemical Cycles, 19, n/a-n/a, 10.1029/2005gb002606, 2005.

Shen, C. P.: A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists, Water Resour Res, 54, 8558-8593, 10.1029/2018wr022643, 2018.

Sivapalan, M.: The secret to 'doing better hydrological science': change the question!, Hydrol Process, 23, 1391-1396, 10.1002/hyp.7242, 2009.

Soares, M. I. M.: Biological Denitrification of Groundwater, Water, Air, and Soil Pollution, 123, 183-193, 10.1023/A:1005242600186, 2000.

Sobota, D. J., Compton, J. E., McCrackin, M. L., and Singh, S.: Cost of reactive nitrogen release from human activities to the environment in the United States, Environ Res Lett, 10, Artn 025006 10.1088/1748-9326/10/2/025006, 2015.

Soller, D. R., and Reheis, M. C.: Surficial materials in the conterminous United States, 2004.

Stackpoole, S. M., Stets, E. G., and Sprague, L. A.: Variable impacts of contemporary versus legacy agricultural phosphorus on US river water quality, P Natl Acad Sci USA, 116, 20562-20567, 10.1073/pnas.1903226116, 2019.

Sutton, M. A., Oenema, O., Erisman, J. W., Leip, A., van Grinsven, H., and Winiwarter, W.: Too much of a good thing, Nature, 472, 159-161, 10.1038/472159a, 2011.

Thompson, S. E., Basu, N. B., Lascurain, J., Aubeneau, A., and Rao, P. S. C.: Relative dominance of hydrologic versus biogeochemical factors on solute export across impact gradients, Water Resour Res, 47, Artn W00j05 10.1029/2010wr009605, 2011.

Tyralis, H., Papacharalampous, G., and Langousis, A.: A Brief Review of Random Forests for Water Scientists and Practitioners and Their Recent History in Water Resources, Water, 11, 10.3390/w11050910, 2019.

Uuemaa, E., Palliser, C. C., Hughes, A. O., and Tanner, C. C.: Effectiveness of a Natural Headwater Wetland for Reducing Agricultural Nitrogen Loads, Water, 10, ARTN 287 10.3390/w10030287, 2018.

Van Meter, K. J., and Basu, N. B.: Catchment Legacies and Time Lags: A Parsimonious Watershed Model to Predict the Effects of Legacy Storage on Nitrogen Export, Plos One, 10, ARTN e0125971 10.1371/journal.pone.0125971, 2015.

Van Meter, K. J., Basu, N. B., Veenstra, J. J., and Burras, C. L.: The nitrogen legacy: emerging evidence of nitrogen accumulation in anthropogenic landscapes, Environ Res Lett, 11, Artn 035014 10.1088/1748-9326/11/3/035014, 2016.

Van Meter, K. J., Van Cappellen, P., and Basu, N. B.: Legacy nitrogen may prevent achievement of water quality goals in the Gulf of Mexico, Science, 360, 427-430, 10.1126/science.aar4462, 2018.

Van Meter, K. J., Van Cappellen, P., and Basu, N. B.: Response to Comment on "Legacy nitrogen may prevent achievement of water quality goals in the Gulf of Mexico", Science, 365, ARTN eaav3851 10.1126/science.aav3851, 2019.

Van Meter, K. J., Chowdhury, S., Byrnes, D. K., and Basu, N. B.: Biogeochemical asynchrony: Ecosystem drivers of seasonal concentration regimes across the Great Lakes Basin, Limnol Oceanogr, 65, 848-862, 10.1002/lno.11353, 2020.

Vitousek, P. M., Aber, J. D., Howarth, R. W., Likens, G. E., Matson, P. A., Schindler, D. W., Schlesinger, W. H., and Tilman, D.: Human alteration of the global nitrogen cycle: Sources and consequences, Ecol. Appl., 7, 737-750, 10.2307/2269431, 1997.

Vymazal, J.: Removal of nutrients in various types of constructed wetlands, Sci Total Environ, 380, 48-65, 10.1016/j.scitotenv.2006.09.014, 2007.

Wickham, J., Stehman, S. V., Gass, L., Dewitz, J. A., Sorenson, D. G., Granneman, B. J., Poss, R. V., and Baer, L. A.: Thematic accuracy assessment of the 2011 National Land Cover Database (NLCD), Remote Sens Environ, 191, 328-341, 10.1016/j.rse.2016.12.026, 2017.

Wickham, J. D., and Wade, T. G.: Watershed level risk assessment of nitrogen and phosphorus export, Comput Electron Agr, 37, 15-24, Pii S0168-1699(02)00117-5 Doi 10.1016/S0168-1699(02)00117-5, 2002.

Wine, M. L., Cadol, D., and Makhnin, O.: In ecoregions across western USA streamflow increases during post-wildfire recovery, Environ Res Lett, 13, ARTN 014010 10.1088/1748-9326/aa9c5a, 2018a.

Wine, M. L., Makhnin, O., and Cadol, D.: Nonlinear Long-Term Large Watershed Hydrologic Response to Wildfire and Climatic Dynamics Locally Increases Water Yields, Earths Future, 6, 997-1006, 10.1029/2018ef000930, 2018b.

Wine, M. L.: Under non-stationarity securitization contributes to uncertainty and Tragedy of the Commons, J Hydrol, 568, 716-721, 10.1016/j.jhydrol.2018.11.044, 2019.

Wine, M. L.: Climatization of environmental degradation: a widespread challenge to the integrity of earth science, Hydrolog Sci J, 10.1080/02626667.2020.1720024, 2020.

Wine, M. L., and Laronne, J. B.: In Water-Limited Landscapes, an Anthropocene Exchange: Trading Lakes for Irrigated Agriculture, Earth's Future, 8, 10.1029/2019ef001274, 2020.

Wu, Y. P., and Liu, S. G.: Impacts of biofuels production alternatives on water quantity and quality in the Iowa River Basin, Biomass Bioenerg, 36, 182-191, 10.1016/j.biombioe.2011.10.030, 2012.

Yu, J., Kilgour, D. M., Hipel, K. W., and Zhao, M.: Power asymmetry in conflict resolution with application to a water pollution dispute in China, Water Resour Res, 51, 8627-8645, 10.1002/2014wr016257, 2015.

https://doi.org/10.5194/hess-2020-423
Preprint. Discussion started: 14 September 2020

Zeitoun, M., Lankford, B., Krueger, T., Forsyth, T., Carter, R., Hoekstra, A. Y., Taylor, R., Varis, O., Cleaver, F., Boelens, R., Swatuk, L., Tickner, D., Scott, C. A., Mirumachi, N., and Matthews, N.: Reductionist and integrative research approaches to complex water security policy challenges, Global Environ. Change, 39, 143-154, https://doi.org/10.1016/j.gloenvcha.2016.04.010, 2016.

Zhang, Y. K., and Schilling, K. E.: Increasing streamflow and baseflow in Mississippi River since the 1940 s: Effect of land use change, J Hydrol, 324, 412-422, 10.1016/j.jhydrol.2005.09.033, 2006.

Zuur, A. F.: Mixed effects models and extensions in ecology with R, Statistics for biology and health, Springer, New York, NY, xxii, 574 p. pp., 2009.

https://doi.org/10.5194/hess-2020-423
Preprint. Discussion started: 14 September 2020

Table 1. Importance of top-ranked predictors of TN concentrations as measured by % increase in mean square error (MSE) if a particular predictor is omitted. Wetness index (WI) of the month of [TN] measurement (WI_t) and WI of the four preceding months are considered.

Predictor	% Increase in MSE
Discharge	112.5
Year	56.1
WI_{t-1}	53.0
Day of Year	50.8
WI_{t-2}	49.7
Forest	45.6
WI_{t-3}	42.8
WI_t	40.6
WI_{t-4}	40.2
Soil nitrate (Wu and Liu, 2012)	36.1
Wetland count	22.9
Shallow subsurface flowpath from wetland to stream	22.9
Watershed area	21.5

https://doi.org/10.5194/hess-2020-423
Preprint. Discussion started: 14 September 2020

Table 2. Omitting forest area as a predictor reveals the importance of nitrogen loading. See Table 1 caption for abbreviations

Variable	% Increase in MSE
Discharge	105.8
WI_{t-1}	54.7
Total N input	54.0
Year	52.8
Day of year	48.4
WI_{t-2}	48.2
WI_{t}	42.8
WI_{t-4}	42.7
WI_{t-3}	41.4
Soil Nitrate (Wu and Liu, 2012)	40.2
Average Manning's roughness coefficient along the flowpath from wetland to stream	33.5
Average Impervious	26.4
Wetland Count	23.4

https://doi.org/10.5194/hess-2020-423
Preprint. Discussion started: 14 September 2020

Hydrology and Earth System Sciences
Discussions

Table 3. Total Nitrogen models of increasing complexity, where y_{ijt} refers to $\log_{10}$[TN] from the *t*th day of year of the *i*th year measured at the *j*th watershed. Predictors include discharge (x_0), seasonal soil nitrogen (x_1) as simulated by Wu and Liu (2012), cultivated area (x_2), wetland area (x_3), watershed area (x_4), wetness index during the preceding month (x_5), and year (α). The mixed effects model involves both fixed effects, which involve fitting scalar coefficients (β) and random effects, which fit vectors (u).

Model	Model Form	New term	AIC	BIC
1	$y_{ijt} = \beta_0 + u_{0_j} + \beta_1 \bullet \sin\left(\frac{2\pi t}{365}\right) + \beta_2 \bullet \cos\left(\frac{2\pi t}{365}\right) + e_{ijt}$		11151.78	11185.97
2	$y_{ijt} = \beta_0 + u_{0_j} + \beta_1 \bullet \sin\left(\frac{2\pi t}{365}\right) + \beta_2 \bullet \cos\left(\frac{2\pi t}{365}\right) + \beta_3 \bullet \sin\left(\frac{4\pi t}{365}\right) + \beta_4 \bullet \cos\left(\frac{4\pi t}{365}\right) + e_{ijt}$	2nd harmonic	10886.61	10934.48
3	$y_{ijt} = \beta_0 + u_{0_j} + u_{1_j} \bullet \sin\left(\frac{2\pi t}{365}\right) + u_{2_j} \bullet \cos\left(\frac{2\pi t}{365}\right) + \beta_1 \bullet \sin\left(\frac{4\pi t}{365}\right) + \beta_2 \bullet \cos\left(\frac{4\pi t}{365}\right) + e_{ijt}$	1st harmonic, random effect	9479.995	9562.049
4	$y_{ijt} = \beta_0 + u_{0_j} + u_{1_j} \bullet \sin\left(\frac{2\pi t}{365}\right) + u_{2_j} \bullet \cos\left(\frac{2\pi t}{365}\right) + u_{3_j} \bullet \sin\left(\frac{4\pi t}{365}\right) + u_{4_j} \bullet \cos\left(\frac{4\pi t}{365}\right) + e_{ijt}$	2nd harmonic, random effect	9053.358	9196.952
5	$y_{ijt} = \beta_0 + u_{0_j} + u_{1_j} \bullet \sin\left(\frac{2\pi t}{365}\right) + u_{2_j} \bullet \cos\left(\frac{2\pi t}{365}\right) + u_{3_j} \bullet \sin\left(\frac{4\pi t}{365}\right) + u_{4_j} \bullet \cos\left(\frac{4\pi t}{365}\right) + \beta_1 \bullet x_{0_{ijt}} + e_{ijt}$	Discharge, fixed effect	7303.277	7453.706
6	$y_{ijt} = \beta_0 + u_{0_j} + u_{1_j} \bullet \sin\left(\frac{2\pi t}{365}\right) + u_{2_j} \bullet \cos\left(\frac{2\pi t}{365}\right) + u_{3_j} \bullet \sin\left(\frac{4\pi t}{365}\right) + u_{4_j} \bullet \cos\left(\frac{4\pi t}{365}\right) + u_{5_j} \bullet x_{0_{ijt}} + e_{ijt}$	Discharge, random effect	6193.4	6384.855
7	$y_{ijt} = \beta_0 + u_{0_j} + u_{1_j} \bullet \sin\left(\frac{2\pi t}{365}\right) + u_{2_j} \bullet \cos\left(\frac{2\pi t}{365}\right) + u_{3_j} \bullet \sin\left(\frac{4\pi t}{365}\right) + u_{4_j} \bullet \cos\left(\frac{4\pi t}{365}\right) + u_{5_j} \bullet x_{0_{ijt}} + \alpha_{(i)} + e_{ijt}$	Year, fixed effect	6131.496	6404.933
8	$y_{ijt} = \beta_0 + u_{0_j} + u_{1_j} \bullet \sin\left(\frac{2\pi t}{365}\right) + u_{2_j} \bullet \cos\left(\frac{2\pi t}{365}\right) + u_{3_j} \bullet \sin\left(\frac{4\pi t}{365}\right) + u_{4_j} \bullet \cos\left(\frac{4\pi t}{365}\right) + u_{5_j} \bullet x_{0_{ijt}} + \alpha_{(i)} + \beta_1 \bullet x_{1_t} + e_{ijt}$	Soil nitrogen, fixed effect	6113.457	6393.724
9	$y_{ijt} = \beta_0 + u_{0_j} + u_{1_j} \bullet \sin\left(\frac{2\pi t}{365}\right) + u_{2_j} \bullet \cos\left(\frac{2\pi t}{365}\right) + u_{3_j} \bullet \sin\left(\frac{4\pi t}{365}\right) + u_{4_j} \bullet \cos\left(\frac{4\pi t}{365}\right) + u_{5_j} \bullet x_{0_{ijt}} + \alpha_{(i)} + \beta_1 \bullet x_{1_t} + \beta_2 \bullet x_{2_j} + e_{ijt}$	Cultivated area, fixed effect	6059.893	6346.99
10	$y_{ijt} = \beta_0 + u_{0_j} + u_{1_j} \bullet \sin\left(\frac{2\pi t}{365}\right) + u_{2_j} \bullet \cos\left(\frac{2\pi t}{365}\right) + u_{3_j} \bullet \sin\left(\frac{4\pi t}{365}\right) + u_{4_j} \bullet \cos\left(\frac{4\pi t}{365}\right) + u_{5_j} \bullet x_{0_{ijt}} + \alpha_{(i)} + \beta_1 \bullet x_{1_t} + \beta_2 \bullet x_{2_j} + \beta_3 \bullet x_{3_j} + e_{ijt}$	Wetland area, fixed effect	6057.12	6351.047
11	$y_{ijt} = \beta_0 + u_{0_j} + u_{1_j} \bullet \sin\left(\frac{2\pi t}{365}\right) + u_{2_j} \bullet \cos\left(\frac{2\pi t}{365}\right) + u_{3_j} \bullet \sin\left(\frac{4\pi t}{365}\right) + u_{4_j} \bullet \cos\left(\frac{4\pi t}{365}\right) + u_{5_j} \bullet x_{0_{ijt}} + \alpha_{(i)} + \beta_1 \bullet x_{1_t} + \beta_2 \bullet x_{2_j} + \beta_3 \bullet x_{3_j} + \beta_4 \bullet x_{4_j} + e_{ijt}$	Watershed area, fixed effect	6044.125	6344.88

https://doi.org/10.5194/hess-2020-423
Preprint. Discussion started: 14 September 2020

Hydrology and Earth System Sciences
Discussions

12	$y_{ijt} = \beta_0 + u_{0_j} + u_{1_j} \bullet \sin\left(\frac{2\pi t}{365}\right) + u_{2_j} \bullet \cos\left(\frac{2\pi t}{365}\right) + u_{3_j} \bullet \sin\left(\frac{4\pi t}{365}\right) + u_{4_j} \bullet \cos\left(\frac{4\pi t}{365}\right) + u_{5_j} \bullet x_{0_{ijt}} + \alpha_{(i)} + \beta_1 \bullet x_{1_t} + \beta_2 \bullet x_{2_j} + \beta_3 \bullet x_{3_j} + \beta_4 \bullet x_{4_j} + \beta_5 \bullet x_{5_{ijt}} + e_{ijt}$	Wetness index, fixed effect	6030.475	6338.06
13	$y_{ijt} = \beta_0 + u_{0_j} + u_{1_j} \bullet \sin\left(\frac{2\pi t}{365}\right) + u_{2_j} \bullet \cos\left(\frac{2\pi t}{365}\right) + u_{3_j} \bullet \sin\left(\frac{4\pi t}{365}\right) + u_{4_j} \bullet \cos\left(\frac{4\pi t}{365}\right) + u_{5_j} \bullet x_{0_{ijt}} + \alpha_{(i)} + \beta_1 \bullet x_{1_t} + \beta_2 \bullet x_{2_j} + \beta_3 \bullet x_{3_j} + \beta_4 \bullet x_{4_j} + \beta_5 \bullet x_{5_{ijt}} + \beta_6 \bullet x_{5_{ijt}} \bullet x_{1_t} + e_{ijt}$	Interaction of wetness index, soil NO_3-N	6022.122	6336.535
14	$y_{ijt} = \beta_0 + u_{0_j} + u_{1_j} \bullet \sin\left(\frac{2\pi t}{365}\right) + u_{2_j} \bullet \cos\left(\frac{2\pi t}{365}\right) + u_{3_j} \bullet \sin\left(\frac{4\pi T}{365}\right) + u_{4_j} \bullet \cos\left(\frac{4\pi t}{365}\right) + u_{5_j} \bullet x_{0_{ijt}} + \alpha_{(i)} + \beta_1 \bullet x_{1_t} + \beta_2 \bullet x_{2_j} + \beta_3 \bullet x_{3_j} + \beta_4 \bullet x_{4_j} + \beta_5 \bullet x_{5_{ijt}} + \beta_6 \bullet x_{5_{ijt}} \bullet x_{1_t} + \beta_7 \bullet x_{5_{ijt}} \bullet x_{2_j} + e_{ijt}$	Interaction of wetness index, cultivated area	6018.495	6339.736

https://doi.org/10.5194/hess-2020-423
Preprint. Discussion started: 14 September 2020

Hydrology and Earth System Sciences
Open Access
Discussions

Table 4. Spearman correlation between the first harmonic [TN] amplitude associated with each watershed and watershed metrics.

Variable	Correlation
Watershed area covered by grassland	-0.393
Watershed area covered by barren land	-0.278
Number of wetlands in a watershed	-0.259
Average Manning's roughness coefficient along the flowpath from wetland to stream	-0.246
Watershed area covered by pasture	-0.237
Average proportion of wetland drainage areas with impervious surfaces	0.236
Number of wetlands per unit watershed area	-0.209
Area of a watershed	-0.207
Total wetland area in a watershed	-0.202

https://doi.org/10.5194/hess-2020-423
Preprint. Discussion started: 14 September 2020

Hydrology and Earth System Sciences
Open Access
Discussions

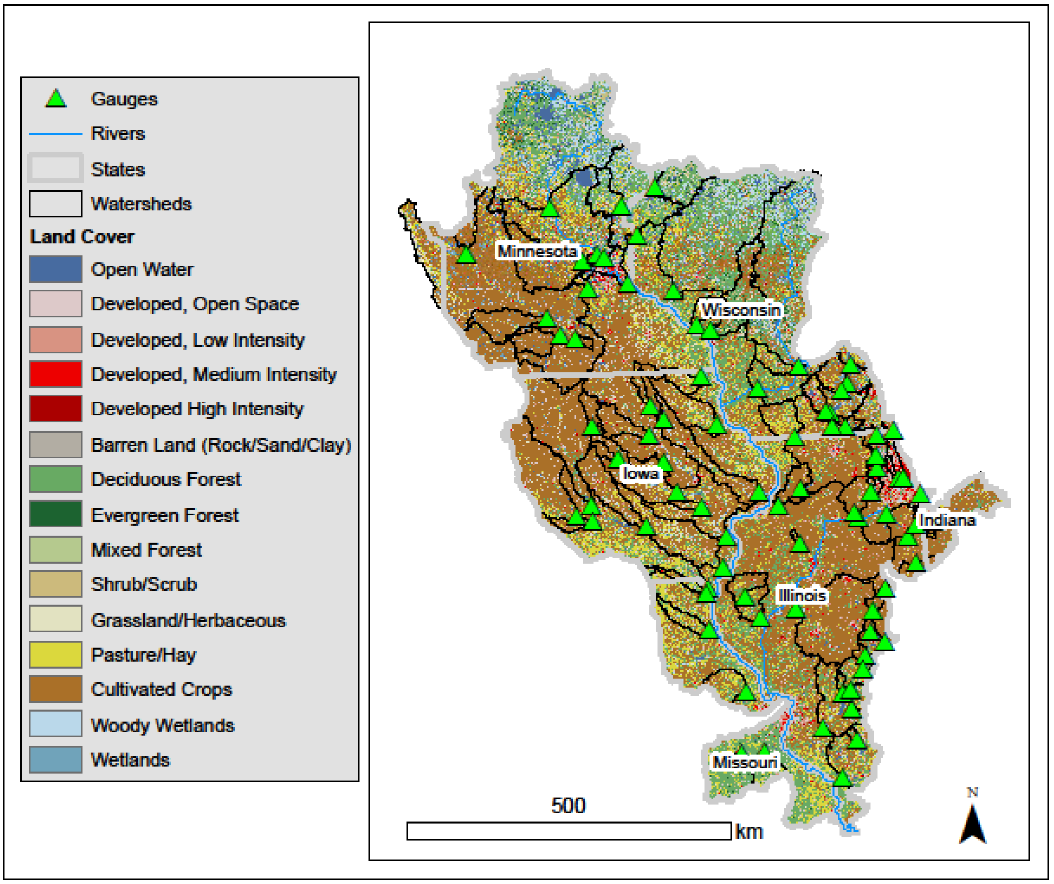

Figure 1: Land cover, study basins, and stream gauges in the Upper Mississippi River basin, USA.

https://doi.org/10.5194/hess-2020-423
Preprint. Discussion started: 14 September 2020

Hydrology and Earth System Sciences
Open Access
Discussions

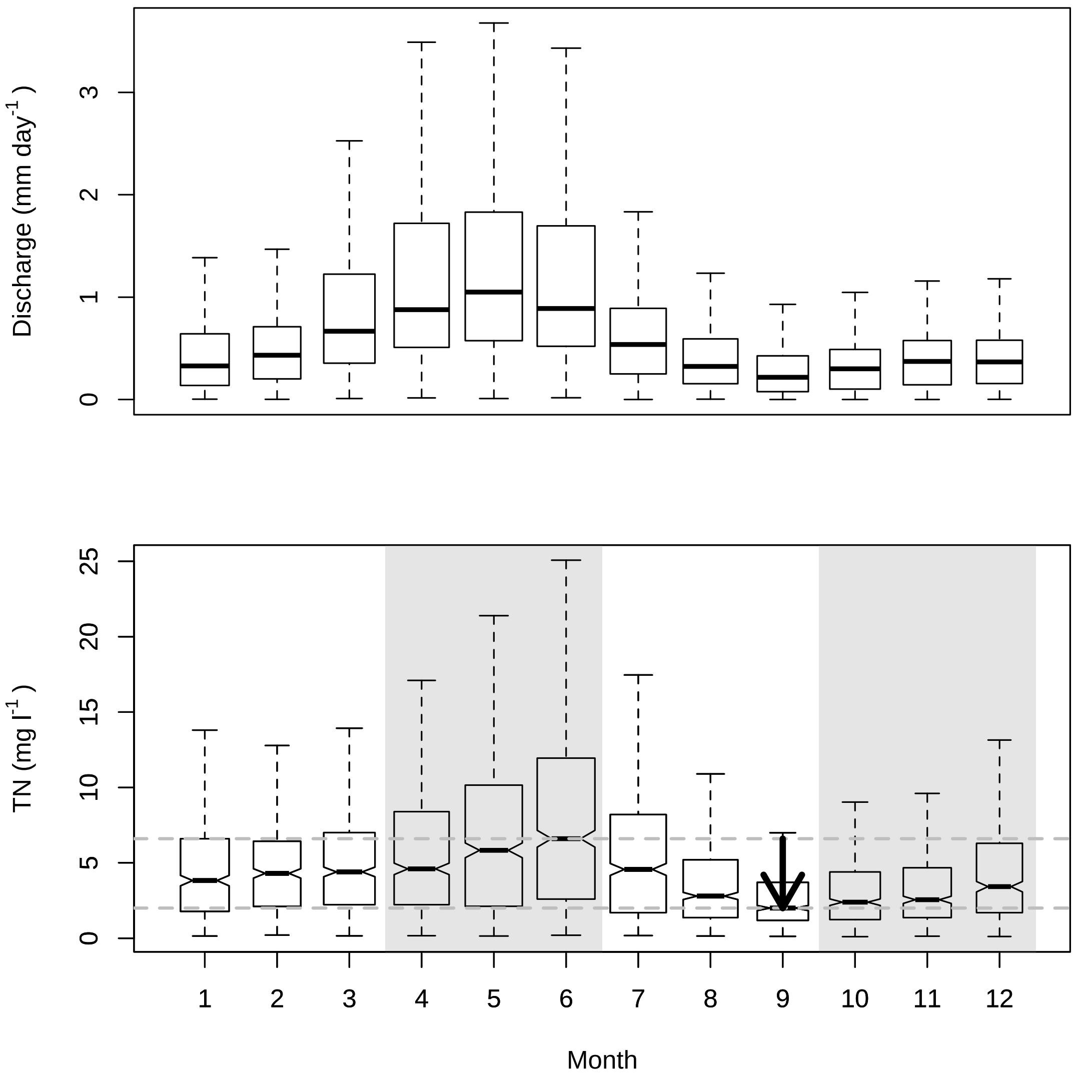

Figure 2: Stream total nitrogen (TN) concentrations peak in June following spring fertilizer loading. By September concentrations have decreased by 70%, consistent with contemporary nutrient loading as the primary cause of observed water quality impoverishment. Data are from 82 watersheds in the Upper Mississippi River basin (1995-2007). Recommended timing for fertilizer application (i.e., spring and fall) are shaded gray. (The gray dashed lines respectively refer to the highest and lowest monthly median [TN].)

https://doi.org/10.5194/hess-2020-423
Preprint. Discussion started: 14 September 2020

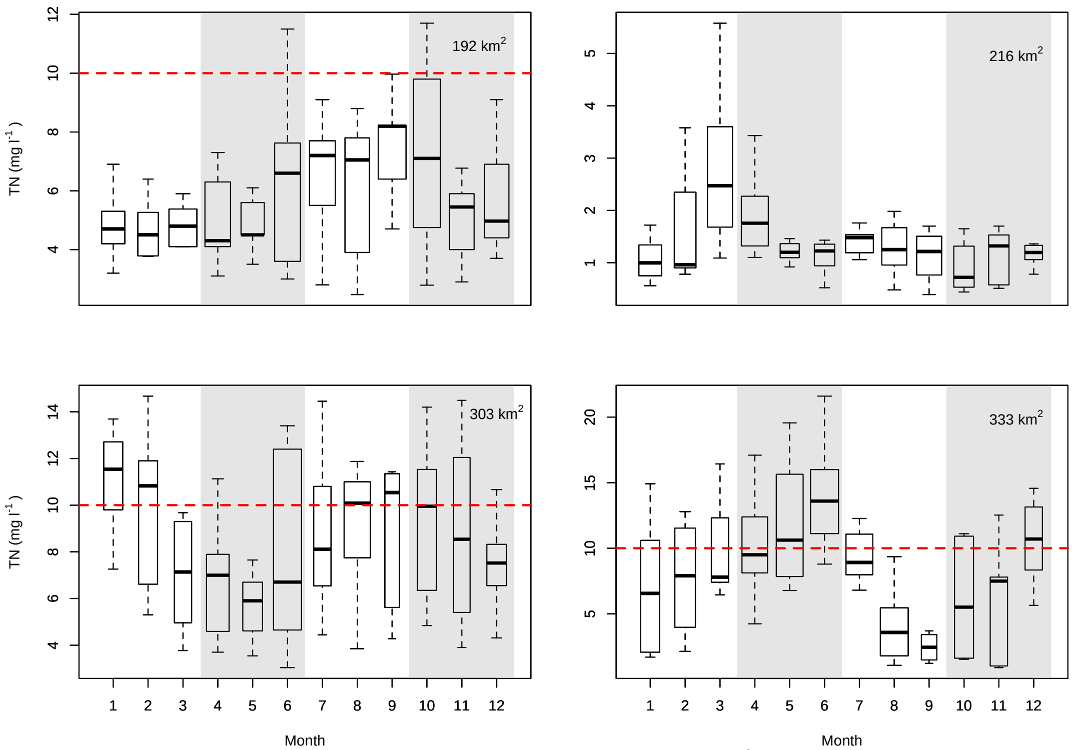

Figure 3: A range of conditions are observed in smaller UMRB watersheds (<350 km^2), e.g., here TN concentrations peak in September, March, January, and June, respectively. (The red dashed line references the current maximum contaminant level for nitrate.)

Hydrology and Earth System Sciences
Open Access
Discussions

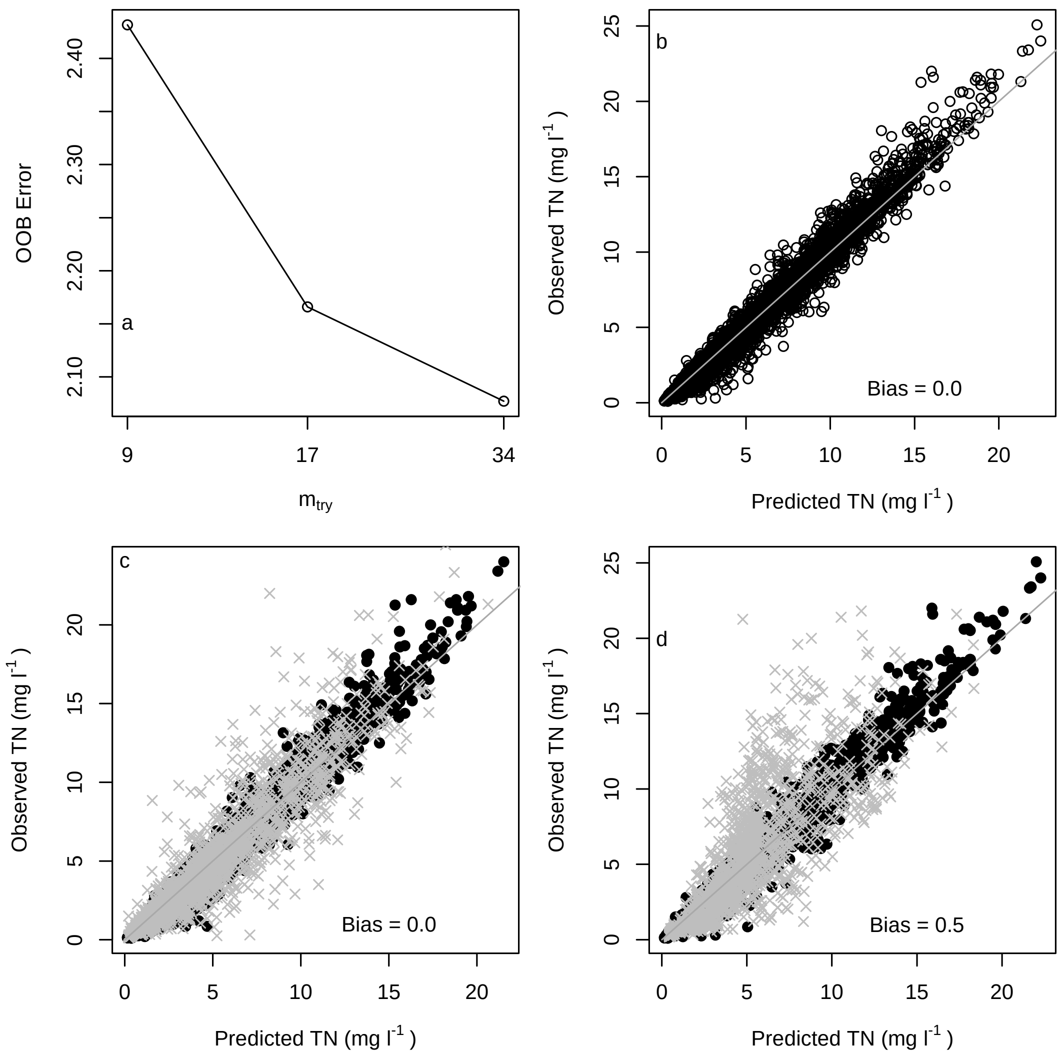

Figure 4: (a)The m_{try} parameter in random forest is optimized by minimizing out of bag (OOB) error. (b) Random forest predictions in which all observations are made available for training. (c) Random forest predictions on the training (black) and verification (gray) datasets, where 70% of samples from each site were selected. (d) Random forest predictions when training was performed on all observations from each of 70% of all sites.

https://doi.org/10.5194/hess-2020-423
Preprint. Discussion started: 14 September 2020

Hydrology and Earth System Sciences
Discussions

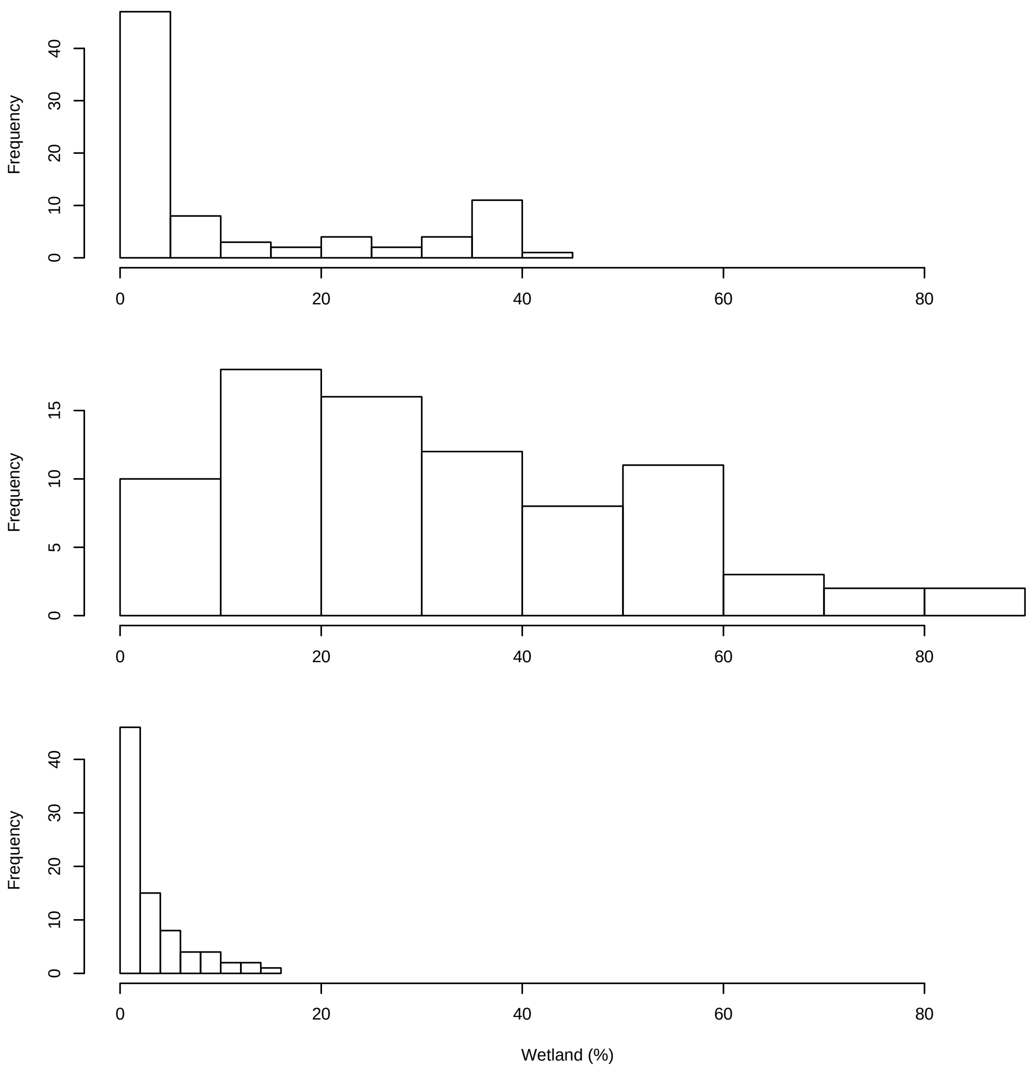

Figure 5: Proportion of wetlands within each of the 82 study watersheds in the Upper Mississippi River. Historic distribution of wetlands from the (Top) Global Lakes and Wetlands Database (Lehner and Doll, 2004) and (Middle) Horvath et al. (2017). (Bottom) Contemporary distribution of wetlands from the National Wetland Inventory.

https://doi.org/10.5194/hess-2020-423
Preprint. Discussion started: 14 September 2020

Hydrology and Earth System Sciences
Open Access
EGU
Discussions

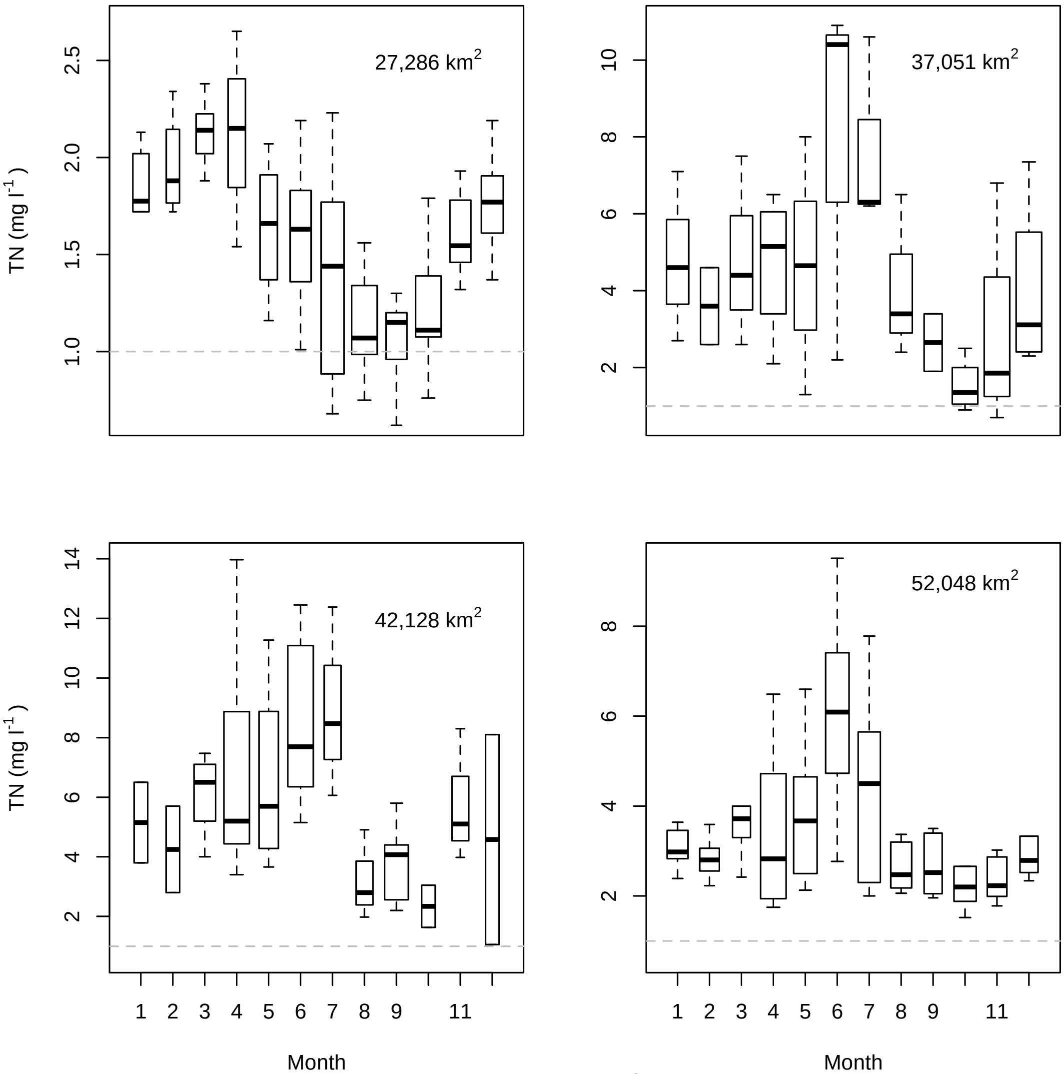

Figure 6: Four examples of larger watersheds in the UMRB (>27,000 km^2) showing Total Nitrogen [TN] concentrations tend to peak in June or July and reach a trough in September or October. (The grey dashed line references 1 mg l^{-1}.)

https://doi.org/10.5194/hess-2020-423
Preprint. Discussion started: 14 September 2020

Hydrology and Earth System Sciences
Discussions

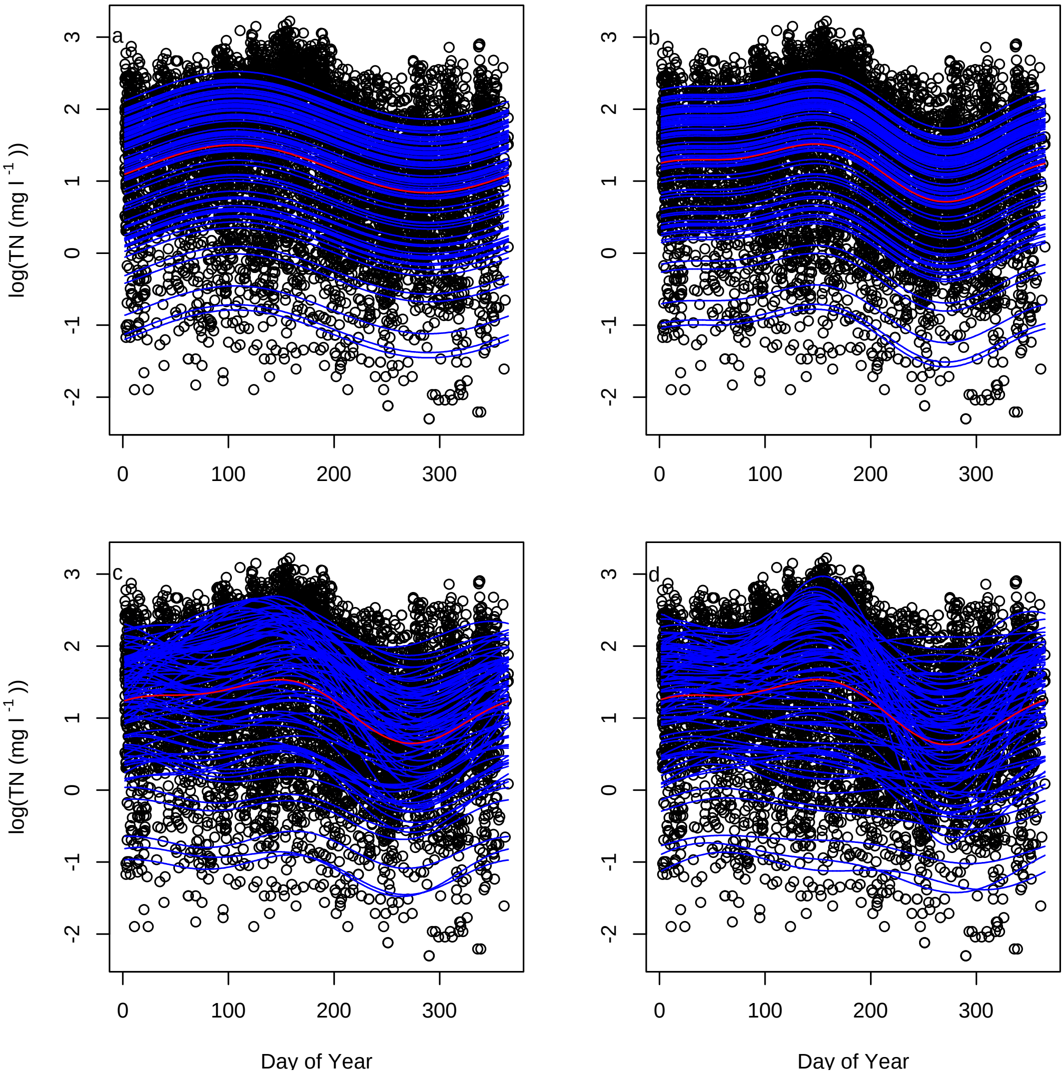

Figure 7: Linear Mixed Effect Modeling [TN] with a random intercept and first harmonic (a), adding a second harmonic (b), assigning the first harmonic a random effect (c), and assigning both harmonics as random effects (d). These model iterations correspond to equations 1-4 of Table 3. The red line represents the fixed effects and the blue lines indicate the random effects (by watershed).

https://doi.org/10.5194/hess-2020-423
Preprint. Discussion started: 14 September 2020

Hydrology and Earth System Sciences
Discussions

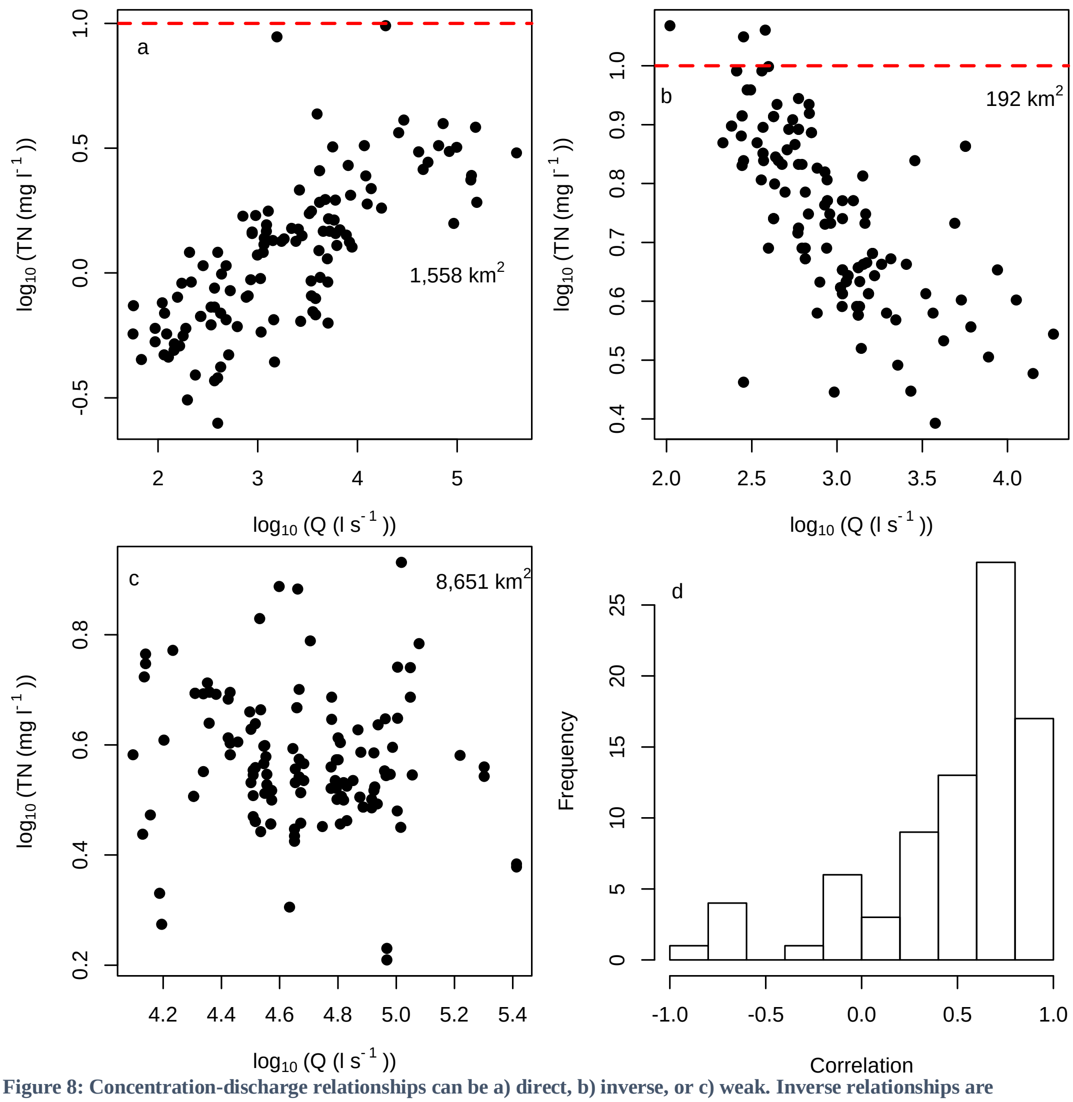

Figure 8: Concentration-discharge relationships can be a) direct, b) inverse, or c) weak. Inverse relationships are observed in watersheds in which 50% or more of the area is drained artificially by tiles. d) Typically, the Spearman correlation between concentration and discharge is direct, whereas fewer watersheds exhibit inverse relationships. The red stippled line in a) and b) corresponds to the maximum contaminant level for nitrate in the US.

https://doi.org/10.5194/hess-2020-423
Preprint. Discussion started: 14 September 2020

Hydrology and Earth System Sciences
Open Access
Discussions

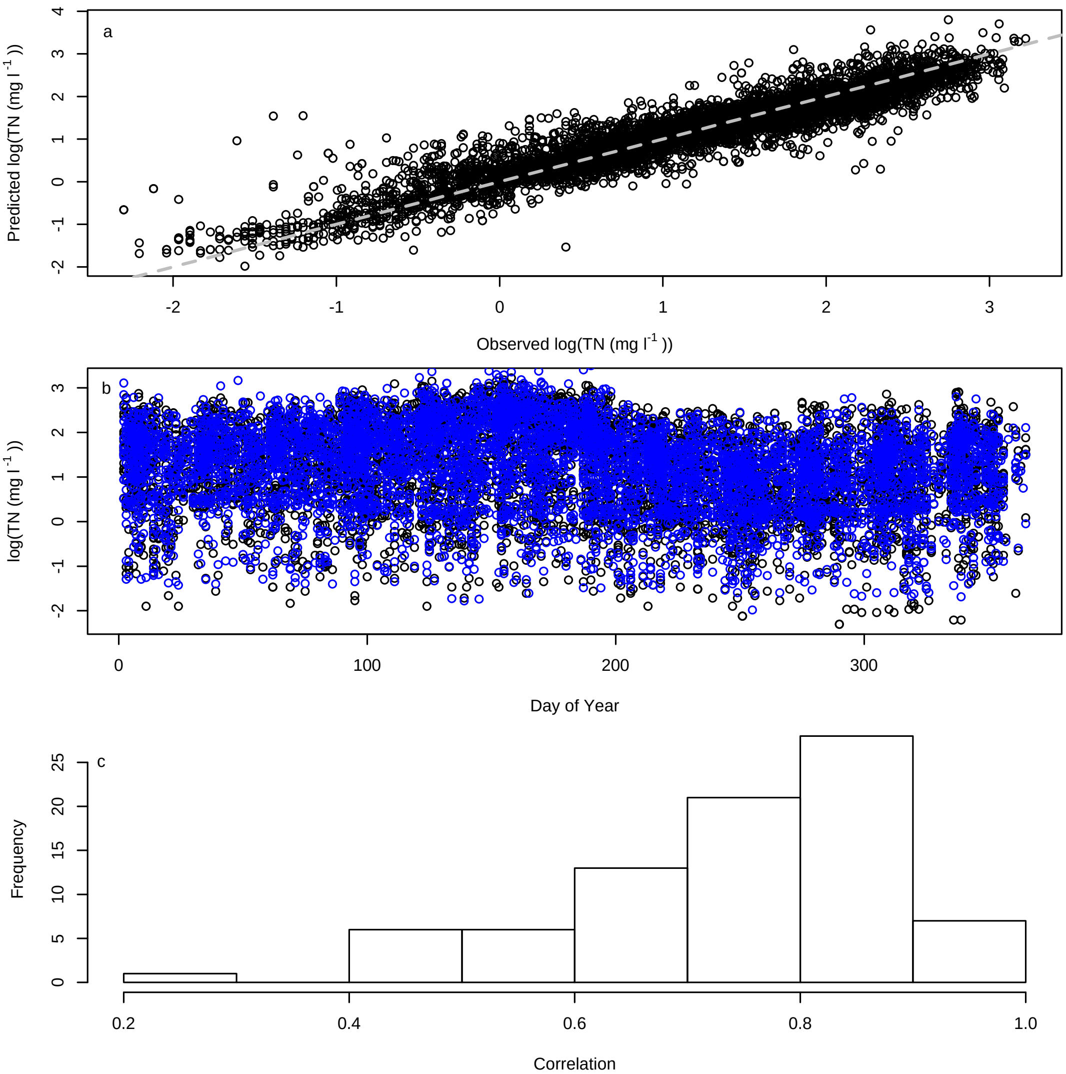

Figure 9: Mixed effects modeling described TN variability (top), including correspondence between the seasonal distributions of measurements (black) and model predictions (blue). Assessment of the Spearman correlation between measured and modeled values indicates strong (>0.6) to very strong (>0.8) agreement in most watersheds.

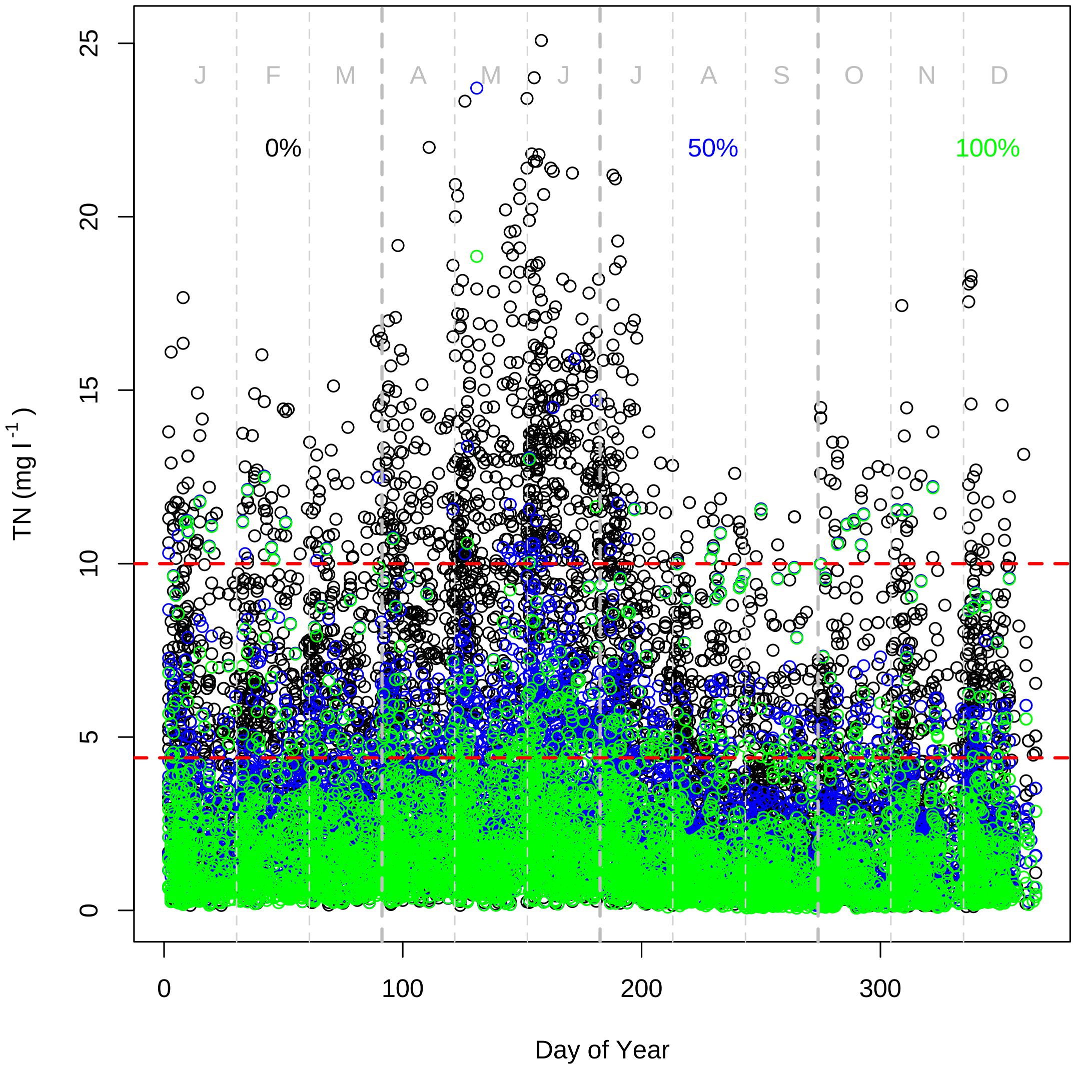

Figure 10: Observed (black) and estimated [TN] for 50% (blue) and 100% (green) restoration of the historic wetland extent using our final model forced by Horvath et al. (2017). Red lines, provided as a reference, indicate the maximum contaminant levels for nitrate in the US 10 mg l^{-1} and Germany 4.4 mg l^{-1}.