Cornell University
Biological and Environmental Engineering

Exploring the Relationship between Tile Drainage Density and Surface Water Nitrogen Pollution in New York State

Romir Anand (ra493), Dr. Scott Steinschneider (ss3378)

August 13, 2024

1 INTRODUCTION

Nitrogen (N) in fertilizer and manure applied to croplands is a prominent cause of nutrient pollution in New York State (NYS)[1]. Excess N ends up in waterways through runoff and groundwater infiltration, which results in eutrophication of nearby water bodies and fuels the release of nitrous oxide, a greenhouse gas[2]. Because of these water quality impacts, NYS agencies charged with managing the state's water bodies have programs designed to help agricultural communities reduce the nutrient concentrations in farmland runoff (e.g., the Department of Agriculture and Market's Agricultural Non-point Source Abatement and Control Program).

Subsurface tile drainage is a prevalent agricultural management practice designed to improve crop yields in areas with high water tables or poorly drained soils[3]. This technique involves the installation of subsurface pipes (tiles) beneath the crop root zone to facilitate the removal of excess water, which enhances soil aeration and root development.

Research surveys conducted by the US. Department of Agriculture (USDA) haves shown that, in New York State, many farms in NYS use tile drainage to increase crop yields and improve soils[4]. However, tile drainage also increases baseflow in watersheds, elevates annual runoff volumes and reduces groundwater travel times[4]. Given these factors as well as the N-based-fertilizers used by farms, there is a high potential for increased surface water nutrient pollution in watersheds with a high density of tile drainage [5-6]. However, no work has established whether such a relationship exists in NYS, and the state's non-point source pollution programs do not consider tile drainage as a target for remediation.

Because tile-drained areas are invariably associated with crop cover, there is limited research that examines the impact of tile drainage on N pollution independently from crop cover[7]. To address this gap, this study evaluates the associations between tile drainage and surface water N concentrations in NYS by integrating water quality data with a novel spatial dataset of tile drainage data and additional land use information. Our objective is to isolate the specific impacts of tile drainage on surface water N concentrations from the more general impacts of agricultural land cover. Ultimately, the goal of this work is to provide decision-relevant information that can help state agencies better tailor agricultural land management policies and programs to efficiently and effectively improve water quality across NYS rivers and streams.

2 DATA AND METHODS

2.1 Nitrate Data

Most water quality data used in this study was collected from the Water Quality Portal (WQP)[8]. The WQP aggregates water quality data from a variety of federal, state, and local agencies, offering a comprehensive repository for surface water quality metrics. For this study, we extracted nitrate concentrations at all available locations in NYS, which was the most abundant and widespread nitrogen-based metric available through the WQP. The dataset included measurements of nitrate in units of milligrams per liter (mg/L) as nitrogen (N) or mg/L as nitrate (NO3), the latter of which we converted to units of mg/L as N. The data spanned from 1950 to the present, but only data from 2000 onward was used in the analysis.

The extraction process of nitrate data from the WQP was streamlined through a Python function that allows for flexible modification of inputs to retrieve the necessary data. For this study, only surface water data was required, and measurements with units in µg/L were excluded. The latitude and longitude information for the data points were not included in the initial WQP dataset. To obtain these, a separate WFS GetFeature dataset was downloaded from the WQP using the same parameters as stated which contains the location data. The location and chemical data were then combined using Python pandas to merge the datasets.

Supplementary nitrate data was obtained from the Community Science Institute (CSI), which conducts extensive water quality monitoring throughout the Finger Lakes region of New York State. This region is of particular interest due to its more extensive tile drainage systems. The CSI dataset included nitrate measurements and corresponding location data, providing a valuable complement to the WQP data[9]. CSI data was downloaded with parameters similar to those used for the WQP. Location data for CSI sampling sites were not available for download through the online portal and instead were obtained through email correspondence with the CSI. The formatting of the CSI data was adjusted in Python to align with the structure of the WQP data and was integrated with the existing data. For the purposes of this study, only data from Upstate New York was retained (sampling points north of 41°N), as nitrogen concentrations in the greater New York City area are most likely affected by urban development rather than agriculture.

After combining both WQP and CSI datasets, instances of multiple results from the same sampling location and date were averaged. This process was streamlined in Python and reduced the dataset size from 71,205 to 43,731 samples, taken from 1,913 sampling locations from 2000-present. To better understand how sampling locations and nitrate concentrations have changed over time, we show the data for each decade based on the average nitrate concentration and number of samples per location in Figure 1 and Figure 2.

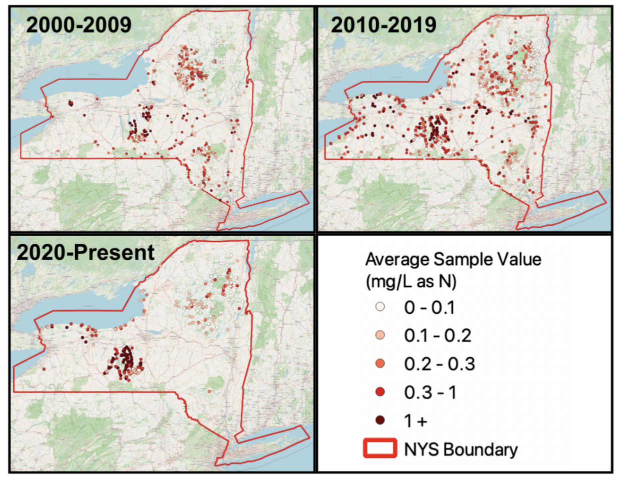

Figure 1: Nitrate concentrations in New York State per sample location by decade.

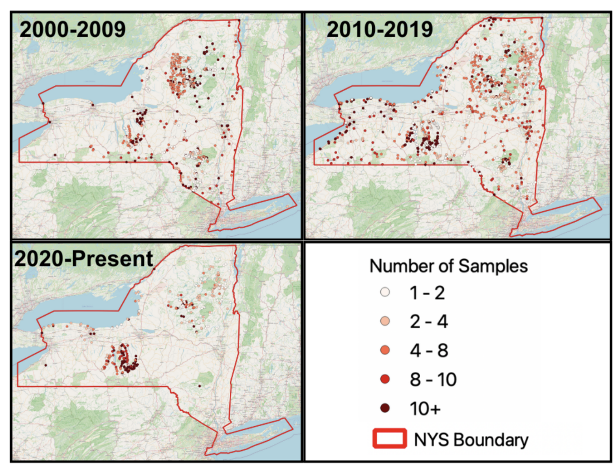

Figure 2: Nitrate sample counts in New York State per sample location by decade.

2.2 Crop Coverage and Tile Drainage Data

Land use data from the National Land Cover Database (NLCD) for the year 2016 was utilized to identify areas of cultivated crops. This dataset is a 30-meter raster-based representation of land cover types across the United States and was specifically used to isolate areas of crop coverage. The NLCD data is essential for assessing the impact of crop coverage on surface water quality, so that it can be compared against that of tile drainage[10].

Because tile drainage is privately installed on a farm-by-farm basis, a precise, parcel-level database of tile drained cropland across the country is not available. Instead, we source tile drainage data from Valayamkunnath et al. (2020), who estimated the presence of tile drainage at 30-meter resolution based on USDA Census data, land use data, and SRTM-DEM slope11. Figure 3 shows both crop coverage and estimated tile drainage across NYS.

Watersheds for each water quality sampling location were delineated using the USGS StreamStats batch processor[12]. The result was a geodataframe with polygons corresponding to each sampling point entered. Using QGIS, the NLCD and tile drainage rasters were coarsened from 30m to 100m and then were aggregated (summed) within each delineated watershed. This ultimately led to estimates of crop coverage and tile drainage area (and percent area) for the watershed of each nitrate sampling location. These data were then merged with the final nitrate data in Python.

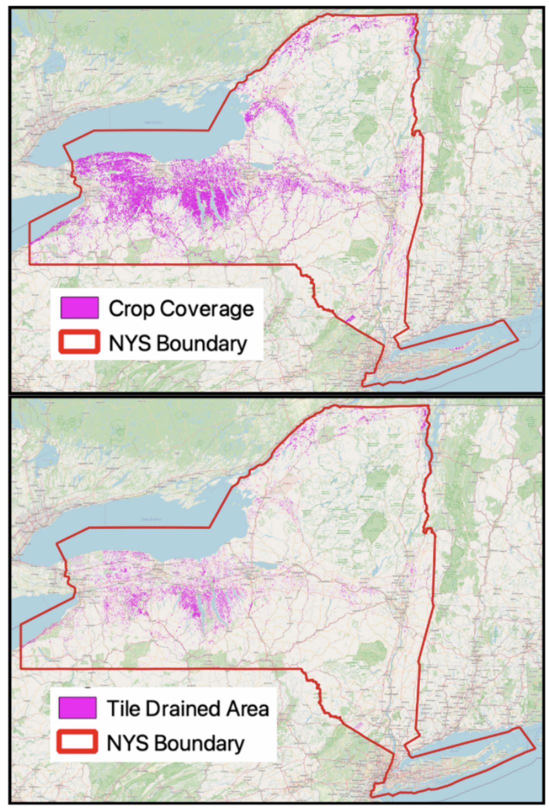

Figure 3: Crop Coverage and Tile Drained Area in New York State.

2.3 Data Analysis

The combined dataset of sampling locations, average nitrate concentrations, sample counts, and corresponding crop coverage and tile-drained areas was analyzed to investigate the impact of tile drainage on nitrate concentrations in surface water. Two multiple linear regression models were developed to determine the significance of the impact of tile drainage on surface water nitrate concentrations in NYS. Both models included the percent area of crop cover (*crop*) as one predictor. The first model used the percent area of tile drainage ($tile_{area}$) as the second predictor, while the second model used the proportion of crop cover that is tile drained ($tile_{proportion} = tile_{area}/crop$) as the second predictor. These two models were used to ensure model results were not overly sensitive to the form of the predictor used to represent tile drainage. In addition, $tile_{proportion}$ used in Model 2 is less correlated to crop cover than $tile_{area}$ used in Model 1 (correlation between $tile_{area}$ and $crop$ is 0.94), which could help isolate the effect of tile drainage on nitrate concentrations and mitigate the confounding influence of crop cover.

Model 1
$\ln(\hat{\bar{N}}) = \beta_0 + \beta_1 crop + \beta_2 tile_{area}$
Model 2
$\ln(\hat{\bar{N}}) = \beta_0 + \beta_1 crop + \beta_2 tile_{proportion}$

Table 1: Equations for Models 1 and 2

For both models, we observed that the relationship between average nitrogen concentrations (denoted $\bar{N}$) and the predictors (crop cover and tile drainage) exhibited a non-linear, exponential relationship when modeled using a standard multiple linear regression. To address this non-linearity and satisfy the assumptions of linear regression, we applied a natural logarithm transformation to the nitrogen concentrations.

For both Model 1 and Model 2, we used two estimation techniques, Ordinary Least Squares (OLS) and Weighted Least Squares (WLS), both fit using the `Statsmodels` library in Python. The weights used in the WLS models were the number of nitrate samples taken for each location (shown in Figure 2). This approach gives greater weight in the regression to those sampling locations with mean nitrate concentration ($\bar{N}$) based on more samples, which helps avoid noisy estimates of mean nitrate from exerting undue influence on the final regression coefficient estimates. Ultimately, we report on four separate model outcomes, the OLS and WLS estimates for both models derived for Model 1 and Model 2. Consistency in the estimated effects of tile drainage across all four model versions will help increase confidence in the robustness of the results.

3 RESULTS

Estimated coefficients from the multiple linear regression models are presented in Table 2. Two main insights emerge from Table 2. First, for both models and both estimation techniques, we find a significant, positive relationship between average nitrate concentration and crop coverage area ($\beta_1 > 0$; p-value < 0.0015 where $H_0 : \beta_n = 0$). This relationship is expected and indicates that nitrate concentrations are higher downstream of areas with a greater density of agricultural land.

The second insight that emerges from this analysis is that tile drainage exhibits a significant, negative effect on average nitrate concentration ($\beta_2 < 0$). All of the coefficients were statistically significant ($p < 0.05$ where $H_0 : \beta_n = 0$).

Model 1	OLS	WLS
β_0 (p)	$-1.321(0.000)$	$-1.188(0.000)$
β_1 (p)	0.045(0.000)	0.049(0.000)
β_2 (p)	$-0.028(0.001)$	$-0.026(0.000)$
r^2	0.386	0.554
Model 2	OLS	WLS
β_0 (p)	$-1.146(0.000)$	$-1.055(0.000)$
β_1 (p)	0.033(0.000)	0.037(0.000)
β_2 (p)	$-0.495(0.013)$	$-0.335(0.035)$
r^2	0.379	0.546

Table 2: OLS and WLS coefficients and r^2 results for Models 1 and 2

4 DISCUSSION AND CONCLUSION

The strong negative relationship between nitrate concentrations and tile drainage variables observed in this study is unexpected, as previous work has shown that tile drainage tends to increase nitrogen concentrations in surface water[5,6]. One possible explanation for this result is that farms in New York State using more tile drainage, which tend to cluster around the Finger Lakes region, may also be implementing USDA Best Management Practices (BMPs), such as wood chip bioreactors on tile drainage outflows, cover crops, no-till farming, or other conservation practices[13,14]. While the clustering of tile drainage in the Finger Lakes region is likely driven by necessity (i.e., the provenance of shallow, poorly draining soils in this region), it's possible that this region also sees a higher provenance of conservation practices due to cultural norms, more active water conservation groups, or greater attention given by county and state water quality agencies. To more robustly establish the impact of tile drainage on nitrogen concentrations in New York's waterways, further research should incorporate additional variables to account for the adoption of BMPs by farms across the state. This would help to provide a more definitive answer regarding the effects of tile drainage on New York's surface water quality.

5 OPEN SCIENCE

All data and code used in this study can be found at this github link.

6 REFERENCES

[1]Vitousek, P. M.; Aber, J. D.; Howarth, R. W.; Likens, G. E.; Matson, P. A.; Schindler, D. W.; Schlesinger, W. H.; Tilman, D. G., Human alteration of the global nitrogen cycle: sources and consequences. *Ecological Applications* **1997**, *7* (3), 737-750.

[2]Reay, D. S.; Davidson, E. A.; Smith, K. A.; Smith, P.; Melillo, J. M.; Dentener, F.; Crutzen, P. J., Global agriculture and nitrous oxide emissions. *Nature Climate Change* **2012**, *2* (6), 410-416.

[3]Blann, K. L., Anderson, J. L., Sands, G. R. & Vondracek, B. Effects of agricultural drainage on aquatic ecosystems: a review. *Critical Reviews in Environmental Science and Technology* **2009**, *39*, 909–1001.

[4]Randall, G. W. and Iragavarapu, T. K., Impact of Long-Term Tillage Systems for Continuous Corn on Nitrate Leaching to Tile Drainage. *Journal of Environmental Quality* **1995**, *24*, 360-366. https://doi.org/10.2134/jeq1995.00472425002400020020x

[5]Skaggs, R. W., Breve, M. A. & Gilliam, J. W. Hydrologic and water quality impacts of agricultural drainage. *Critical Reviews in Environmental Science and Technology* **1994**, *24*, 1–32.

[6]Van den Eertwegh, G. A. P. H., Nieber, J. L., De Louw, P. G. B., Van Hardeveld, H. A. & Bakkum, R. Impacts of drainage activities for clay soils on hydrology and solute loads to surface water. *Irrigation and Drainage* **2006**, *55*, 235–245.

[7]Drury, C. F., Tan, C. S., Welacky, T. W., Reynolds, W. D., Zhang, T. Q., Oloya, T. O., McLaughlin, N. B. & Gaynor, J. D., Reducing Nitrate Loss in Tile Drainage Water with Cover Crops and Water-Table Management Systems. *Journal of Environmental Quality* **2014**, *43*, 587-598. https://doi.org/10.2134/jeq2012.0495

[8]Environmental Protection Agency. (2024). Water Quality Data Portal. Retrieved from https://www.waterqualitydata.us/

[9]Community Science Institute. (2024). Streams and Lakes Data. Retrieved from http://www.database.communityscience.org/queries

[10]Multi-Resolution Land Characteristics (MRLC) Consortium. (2016). NLCD 2016 Land Cover Conterminous United States. U.S. Geological Survey. Retrieved from https://www.mrlc.gov/data/nlcd-2016-land-cover-conus

[11]Valayamkunnath, P., Barlage, M., Chen, F. et al. Mapping of 30-meter resolution tile-drained croplands using a geospatial modeling approach. *Scientific Data* **2020**, *7*, 257. https://doi.org/10.1038/s41597-020-00596-x

[12]United States Geological Survey (2024). StreamStats Batch Processor v1.3.4. Retrieved from

https://streamstats.usgs.gov/ss/

[13]U.S. Forest Service. (n.d.). Best management practices. U.S. Department of Agriculture. Retrieved from https://www.fs.usda.gov/naturalresources/watershed/bmp.shtml

[14]Natural Resources Conservation Service. (2020). Denitrifying bioreactor (Conservation Practice Standard 605). U.S. Department of Agriculture. Retrieved from https://www.nrcs.usda.gov/sites/default/files/2022-09/Denitrifying$_{B}ioreactor_{6}05_{N}HCP_{C}PS_{2020.pdf}$