

E2. Comparing River Nitrogen with Watershed Characteristics

Author: Thomas E. Pearson and Dave Wall, MPCA

Introduction

In-stream nitrogen (N) levels were compared against land use, climate, soils, and other watershed characteristics to determine whether this analysis showed any inconsistencies with the University of Minnesota and the Minnesota Pollution Control Agency (UMN/MPCA) source assessment findings described in Chapters D1 to D4. This analysis was conducted to determine if the relationship between watershed characteristics and stream N levels support or contradict conclusions of the N source assessment, which were derived mostly without the use of statistics or stream monitoring information.

Based on the UMN/MPCA source assessment in chapters D1 through D4, we expected to see the following types of relationships between watershed characteristics and watershed N levels:

- watersheds dominated by forests should have low river N
- watersheds with large percentages of fertilized cropland should have high river N, especially if the land is tilled or is in areas with high groundwater recharge
- river nitrogen loads should be generally independent of human population differences when evaluating rural watersheds

The evaluated watersheds included only those independent 8-digit Hydrologic Unit Code (HUC8) watersheds which: a) were not influenced by upstream main-stem rivers; and b) had two years of N yield and concentration data, obtained during years with mid-range river flows within the 2005-2009 timeframe (see Chapter B3 for more information on the selection of the watersheds meeting minimum criteria).

We analyzed the watershed characteristics and N levels in two different ways: 1) a non-statistical approach to observe the differences in land characteristics between watersheds with low, medium, high, and very high stream N levels; and 2) a statistical multiple regression analysis to identify key watershed characteristics influencing the variability in stream N levels.

This approach follows a central theme in landscape ecology, investigating relationships between spatial patterns in the landscape and ecological processes (Turner et al., 2001), and more specifically, the relationships between land use patterns in watersheds and the conditions of the streams that run through them (Allen 2004). The purpose of the watershed characteristics assessment was to gain a better understanding of similarities and differences among the watersheds with various levels of N pollution. The causes of high and low nitrate levels cannot be isolated as single variables, but are rather due to several confounding factors which involve: the presence or addition of a N source, the amount of water available to drive the N through the soil, an absence of an effective way of removing soil N (such as high density of plant roots), and a transport pathway which circumvents denitrification losses.

This analysis did not include watersheds with large metropolitan areas. This was the case because large metropolitan area watersheds water quality results were influenced by upstream main stem rivers, or we did not have two years of N yield and concentration data for these watersheds, obtained during years with mid-range river flows within the 2005-2009 timeframe.

Watershed characteristics

Methods of extracting land characteristic data

Watershed areas were delineated upstream from 79 water quality monitoring stations across Minnesota. We used ArcHydro in ArcGIS (ESRI 2012) to complete the delineations. Our ArcHydro implementation was developed using a 30-meter hydrologically conditioned digital elevation model (DEM) together with watershed walls enforced using the Minnesota Department of Natural Resources (MDNR) 16-digit catchments, and burned-in streams using the MDNR synthetic flow lines. We selected 28 watersheds that were not influenced by upstream main stem rivers and which also had two years of N yield and concentration data obtained during years with mid-range flows between 2005 and 2009. We used these 28 watersheds to extract data from a series of data layers listed in Table 1. For categorical raster layers such as the National Land Cover Data (NLCD) we calculated the area covered by specific land cover classes. For continuous raster layers such as percent soil organic material, we calculated the average percent of the material for each watershed. For vector layers such as the 2010 Census, we used a spatial overlay apportionment method and summarized the results by watershed to determine density values for each watershed. We used additional spatial overlays and raster analysis tools to determine areas where land cover characteristics overlapped, for example where row crops and shallow depth to bedrock were both present.

Table 1. List of land characteristic data layers and the associated data sources.

Forest and shrub	NLCD 2006 classes 41, 42, 43, 52
Pasture, grass and hay	NLCD 2006 classes 71, 81
Human population density (persons per acre)	U.S. Census 2010 blocks
Livestock and poultry density	MPCA Delta database for feedlots
Shallow depth to bedrock (<= 50 feet)	Preliminary Bedrock Geologic Map of Minnesota, April 2010, Minnesota Geological Survey
Sandy soil areas (>=85%)	USDA NRCS SSURGO soils data
Row crops	USDA Crop Data Layer 2009 including corn, sweet corn, soybeans, dry beans, potatoes, peas, sunflowers, sugarbeets
Small grains	USDA Crop Data Layer 2009
Wetlands	NLCD 2006 classes 90, 95
Precipitation	Minnesota State Climatology Office
Irrigation	Permitted acres from the Minnesota Department of Natural Resources
Soil organic material	USDA NRCS SSURGO soils data
Estimated area tile drained	USDA Crop Data Layer 2009, USDA NRCS SSURGO soils, USGS National Elevation Dataset 30-meter DEM
Derived data layers	
RCD	Row crops over shallow depth to bedrock
RCS	Row crops over sandy soils
RCDS	Row crops over shallow depth to bedrock or sandy soils
RC DST	Row crops over shallow depth to bedrock or sandy soils or tile drain
RCnDST	Row crops not over shallow depth to bedrock, sandy soils, or tile drain

Acronyms	
DEM	Digital elevation model
NLCD	National Land Cover Database
NRCS	Natural Resources Conservation Service
SSURGO	Soil Survey Geographic Database

Our analysis included a data layer to estimate land area with tile drainage. This layer was developed by the authors using information from scientific publications (Sugg 2007, David 2010) and interviews with technical experts working in various rural areas in the state. Our criteria included the presence of row crop agriculture from the 2009 USDA Crop Data Layer, relatively flat slopes of 3% or less from the United States Geological Survey (USGS) National Elevation Dataset 30-meter DEM, and soils that were poorly drained or very poorly drained based on Soil Survey Geographic Database (SSURGO) soils data developed by the U.S. Department of Agriculture, Natural Resources Conservation Service.

We used the data layers listed at the top of Table 1 to create additional spatial data layers to serve as explanatory variables in our analysis. These data layers are listed in Table 1 in the section titled 'Derived Data Layers.' These include row crop over shallow depth to bedrock (RCD), row crop over sandy soils (RCS), row crop over shallow depth to bedrock or sandy soils (RCDS), row crop over shallow depth to bedrock, or sandy soils, or tile drain (RCDST), and finally, row crop with no shallow depth to bedrock, no sandy soils, and no tile drain (RCnDST). The RCD, RCS, and RCDS are considered to be naturally 'leaky' agricultural systems, while the tile drain layer (TD) is considered to be an anthropogenic 'leaky' agricultural system. The RCDST is a combination of these leaky systems, and the RCnDST is a non-leaky system where nutrients are less likely to have rapid pathways to surface waters.

Data coverage for each data layer listed in Table 1 was complete for the full extent of the study area, except for the SSURGO soils layer which was not finished for all areas of Minnesota at the time of this work. However, all 28 watersheds had at least partial SSURGO data coverage, with 17 having 100% coverage; 4 having 80% to 99% coverage; 6 with 50% to 79% coverage; and 1 with less than 50% coverage. For watersheds with incomplete SSURGO data, we assumed that areas with missing data were similar to areas with data present, and we used a proportioning coefficient to reflect that assumption. SSURGO serves as source data for the sandy soil layer and the tile drainage layer, and layers derived from these two. SSURGO was also used to estimate soil organic matter content.

The only watershed with less than 50% SSURGO data coverage was the Little Fork River watershed, which had only 14% coverage. However, we do not believe the minimal SSURGO coverage in the Little Fork River watershed significantly affects the analysis. This watershed has essentially no row crop agriculture, and the only explanatory variables that are based on SSURGO are also based on the presence of row crops (RCS, RCDS, RCDST, RCnDST and TD). So with no row crop agriculture, all these variables have zero values in the Little Fork watershed regardless of the SSURGO soil patterns.

Table 2. List of watersheds with partial SSURGO data coverage

Watersheds with partial SSURGO coverage	Fraction of watershed area covered by SURGO
Otter Tail River	0.99
Crow Wing, Redeye, Long Prairie River	0.96
Wild Rice River	0.87
Snake River	0.81
Rum River	0.73
Big Fork River	0.67
Thief River	0.63
Clearwater River	0.61
Mississippi R Headwaters	0.59
Kettle River	0.54
Little Fork River	0.14

Non-statistical view of watershed characteristics compared to river nitrogen levels

The non-statistical approach we used to compare watershed characteristics with N concentrations was to categorize each watershed as a low, medium, high or very high N watershed, based on the stream N monitoring results. We then assessed the range and mean of numerous watershed characteristics for each of the four N level category watersheds.

Categorizing watersheds into low, medium, high and very high nitrogen levels

Twenty-eight independent watersheds with available normal flow conditions fit into one of four distinct categories based on total nitrogen (TN) and nitrite+nitrate-N (NOx) yields and concentrations. The watersheds fitting the low, medium, high, and very high categories of water N levels are shown in Table 3.

Table 3. Watershed groupings based on stream Nitrite+Nitrate-N and Total N yields and concentrations. Watersheds which did not meet selection criteria are not included in this table.

Category	Low N watersheds	Medium N watersheds	High N watersheds	Very high N watersheds
Major Watersheds	Otter Tail River	Chippewa River	Root River	Blue Earth River
	Rum River	Wild Rice River	Cannon River	Cottonwood River
	Snake River	Clearwater River	Des Moines - Headwaters	South Fork Crow
	Leech Lake River	Buffalo River	Yellow Medicine	LeSueur River
	Kettle River	Pomme de Terre	Redwood River	Watonwan River
	Mississippi R. Headwaters	Sauk River		
	Little Fork River	Sandhill River		
	Big Fork River	Marsh River		
	Thief River			
	Crow Wing + Redeye + Long Prairie Rivers			
Nitrogen ranges				
NOx FWMC (mg/l)	0.05-0.5	0.6-1.9	4.8-7.1	7.9-9.5
TN FWMC (mg/l)	0.7-1.4	1.8-3.4	5.6-8.6	9.8-11.1
NOx Yield (lbs/ac/yr)	0.07-0.53	0.37-2	3.6-8.9	9.3-18.3
TN Yield (lbs/ac/yr)	0.51-2.7	1.4-3.9	7.6-12.1	10.9-21.3

Maps of watershed nitrate and TN concentrations and yields are shown in Figures 1, 3, 5 and 7. The range and average nitrate and TN concentrations and yields for each of the four watershed categorizations in Table 3 are shown in Figures 2, 4, 6 and 8. The same watersheds remain in each of the Table 3 categories throughout all figures in this section. For example, the very high N watersheds are always represented by the Blue Earth, Cottonwood, South Fork Crow, LeSueur, and Watonwan Rivers.

As shown in Figures 1 to 8, the nitrite+nitrate-N (NOx) flow weighted mean concentrations (FWMC) and yields show four distinct ranges and means. For example, the NOx FWMC range in watersheds classified in Table 3 as having high N levels do not overlap at all with the NOx FWMC range of watersheds classified in Table 3 as having medium N levels (Figure 2). The range of TN FWMCs in the four categories of watersheds are also distinct, with no overlapping concentrations among the four categories (Figure 6). NOx yields show the same pattern of a very low range of yields to a very high range of yields in the four categories, although there is a slight overlap in ranges in a couple of the categories (Figure 4). The TN yield ranges are less distinct compared to the NOx yields, since TN includes organic N which is influenced by natural sources as well as human-induced sources (Figure 8).

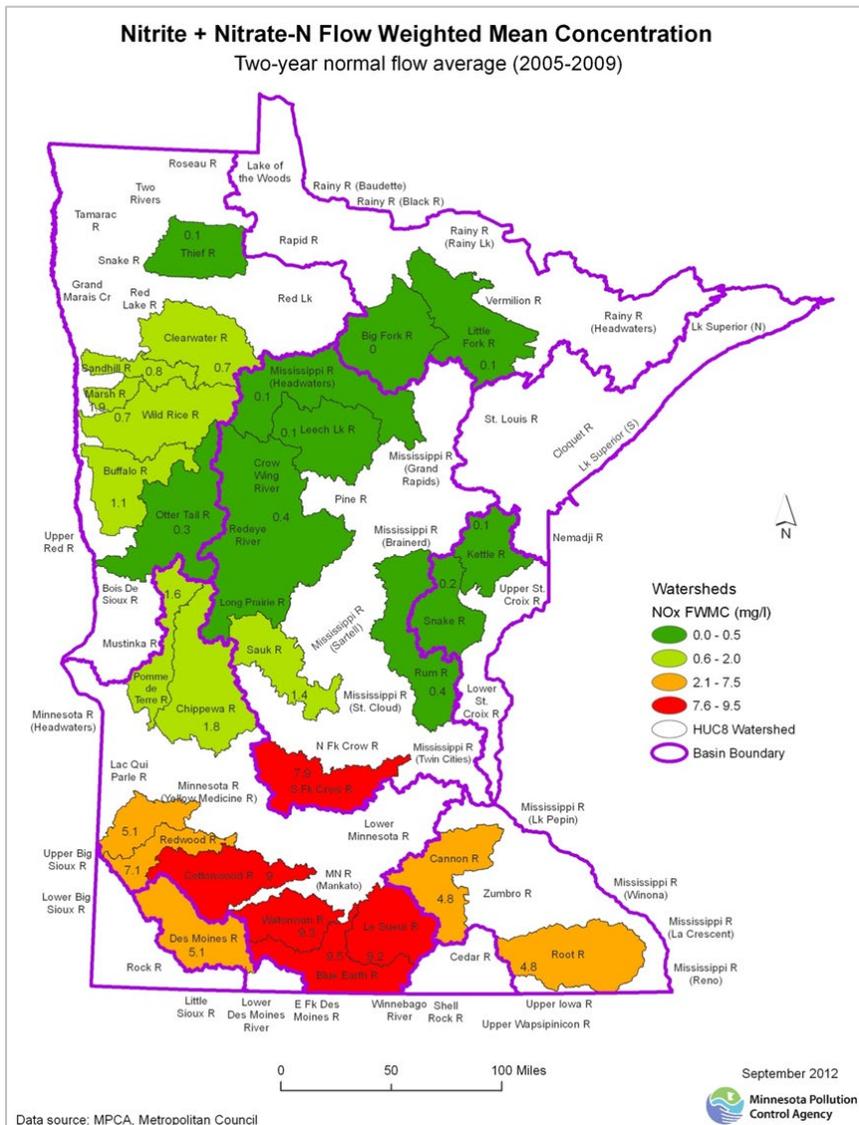


Figure 1. Nitrite+nitrate-N annual flow weighted mean concentration averages from the 28 study watersheds. Monitoring and load calculations were conducted by the MPCA and Metropolitan Council.

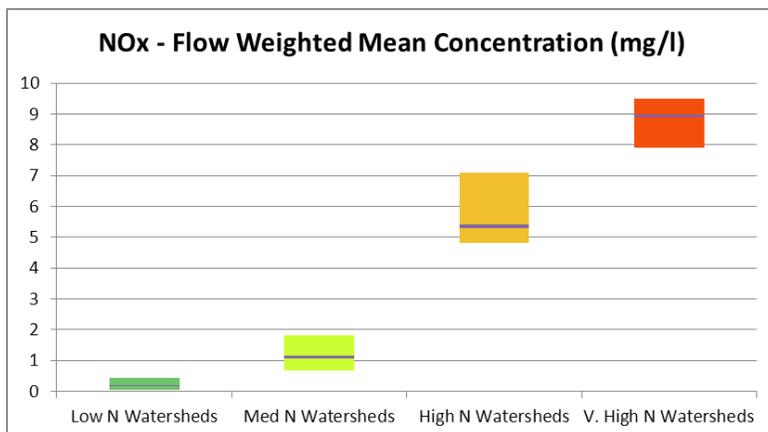


Figure 2. The range (colored bars) and mean (dark line) nitrite+nitrate-N annual flow weighted mean concentration for watersheds in each of the four river N level groupings listed in Table 3.

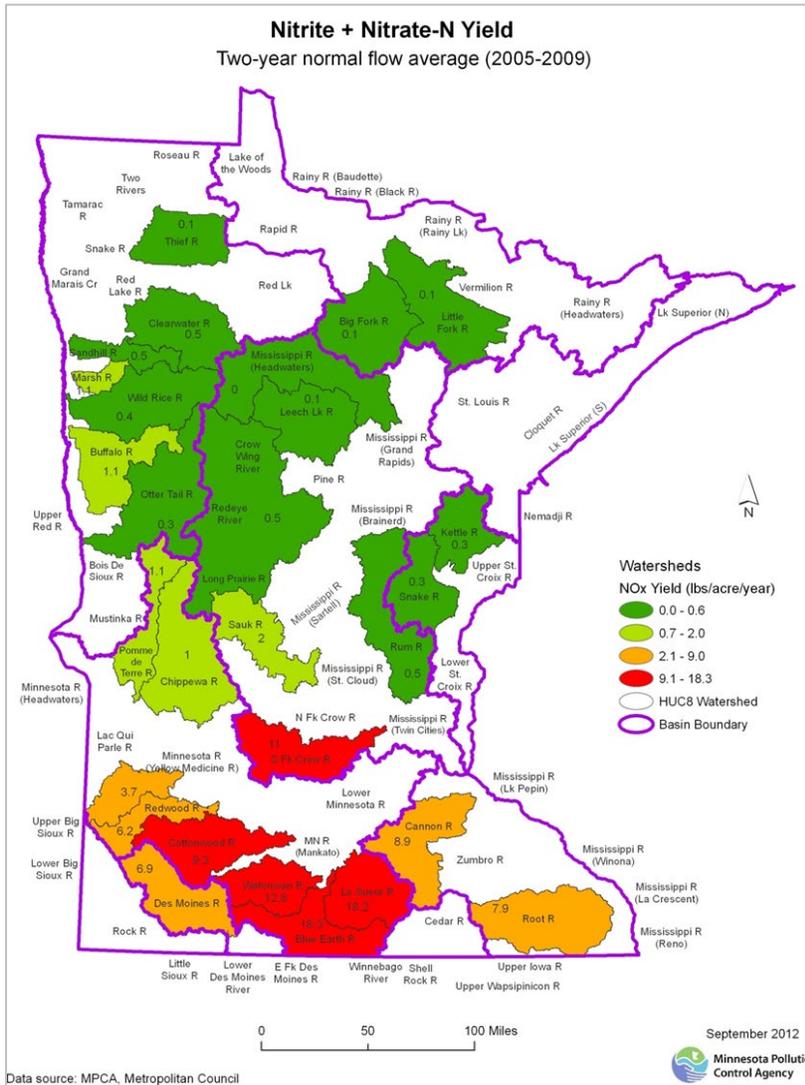


Figure 3. Nitrite+nitrate-N annual yield averages from the 28 study watersheds. Monitoring and yield calculations were conducted by the MPCA and Metropolitan Council.

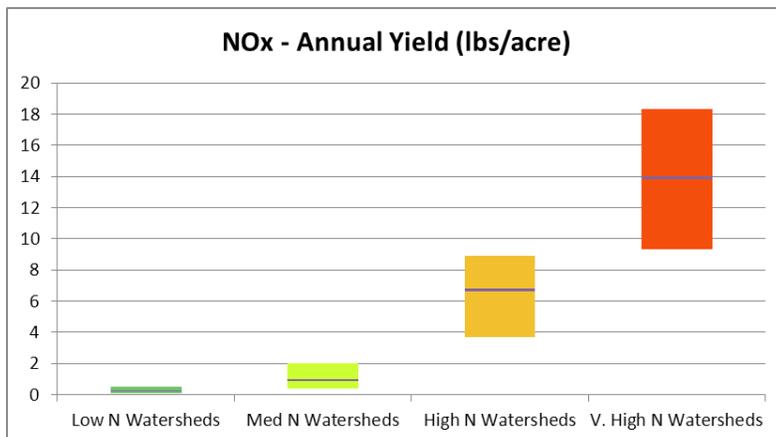


Figure 4. The range (colored bars) and mean (dark line) nitrite+nitrate-N annual yield for watersheds in each of the four river N level groupings listed in Table 3.

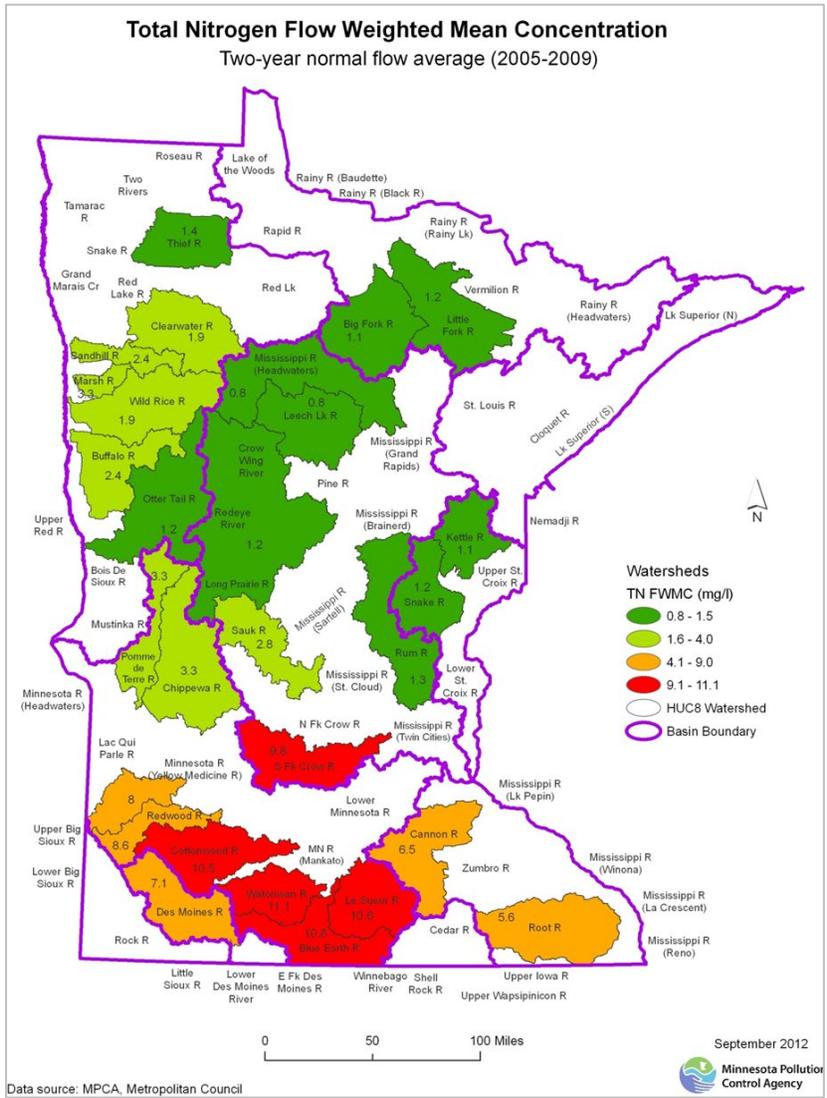


Figure 5. Total nitrogen annual flow weighted mean concentration averages from the 28 study watersheds. Monitoring and load calculations conducted by the MPCA and Metropolitan Council.

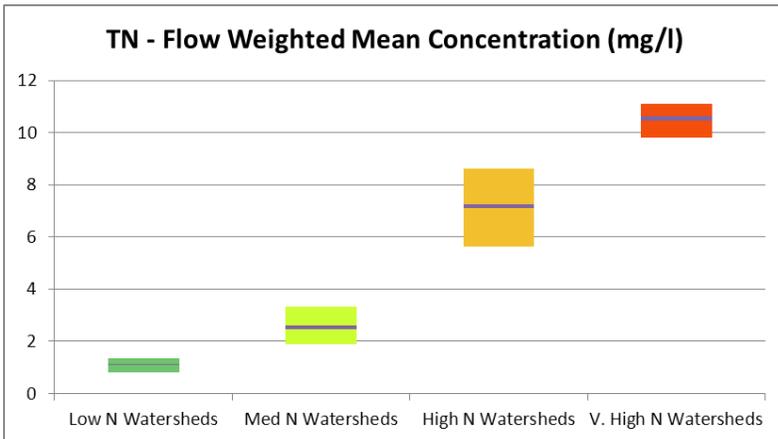


Figure 6. The range (colored bars) and mean (dark line) TN annual flow weighted mean concentration for watersheds in each of the four river N level groupings listed in Table 3.

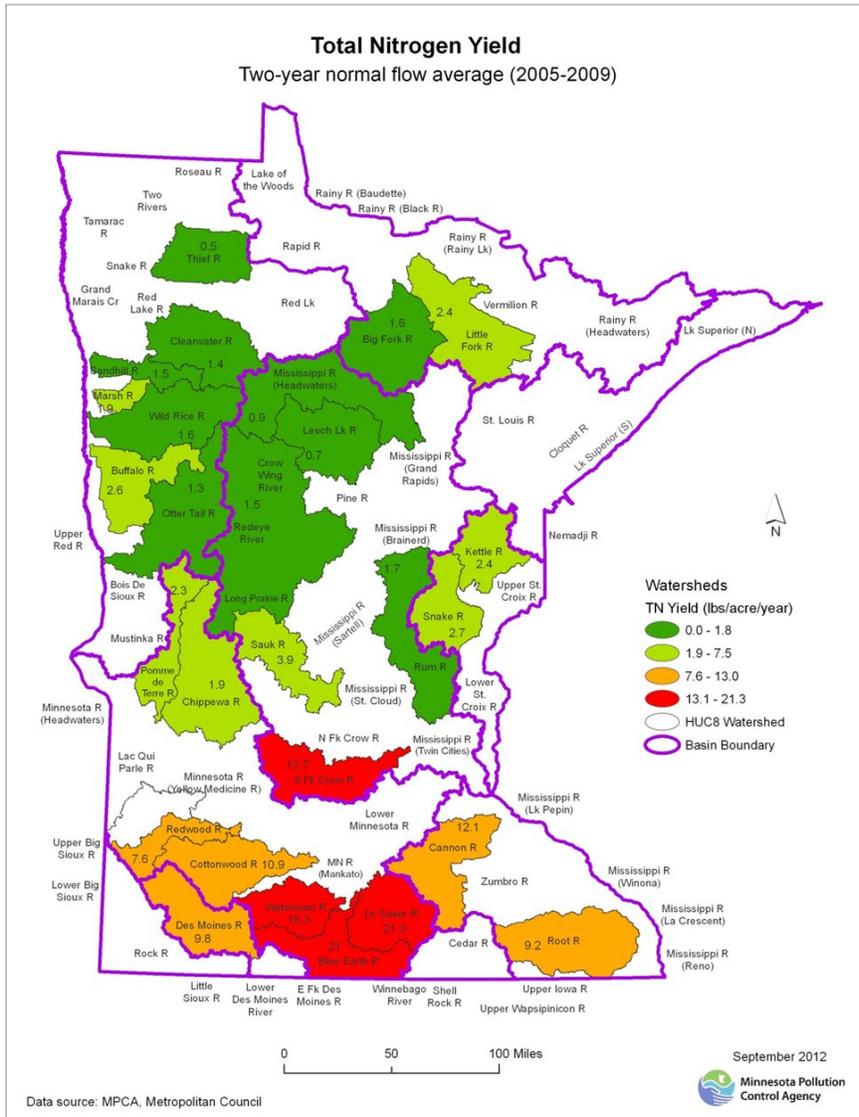


Figure 7. Total nitrogen annual yield from the 28 study watersheds. Monitoring and yield calculations conducted by the MPCA and Metropolitan Council.

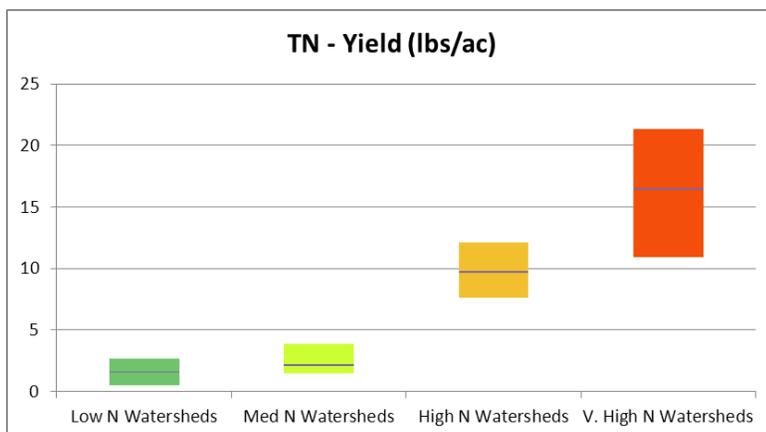


Figure 8. The range (colored bars) and mean (dark line) TN yield for watersheds in each of the four river N level groupings listed in Table 3.

Graphical depictions of watershed land characteristics with nitrogen levels

The range of watershed characteristics for each of the four stream N level categorizations (as listed in Table 3) are shown in Figures 9 to 21. Each bar in these figures represents the range in land use for watersheds assigned to that stream N level category, and the dark line in the middle of the colored bars represents the average of watersheds grouped in each category.

Note: The following results were not used in any way for estimating N source contributions in Section D of this report. The N source assessment uses a completely different approach which does not include statistical relationships between land characteristics and monitoring results.

Forest and grasses

The average percent of watershed land area in forest and grasses is inversely related to the watershed N level, yet there is overlap in the ranges of land percentages in forest and grass among the four categories (Figure 9). The low N watersheds have between 15% and 71% of land in forest and grassland, with a mean of 53%. In contrast, the very high N watersheds have 3% to 15% of their land in forest and grasses, with a mean of 7%.

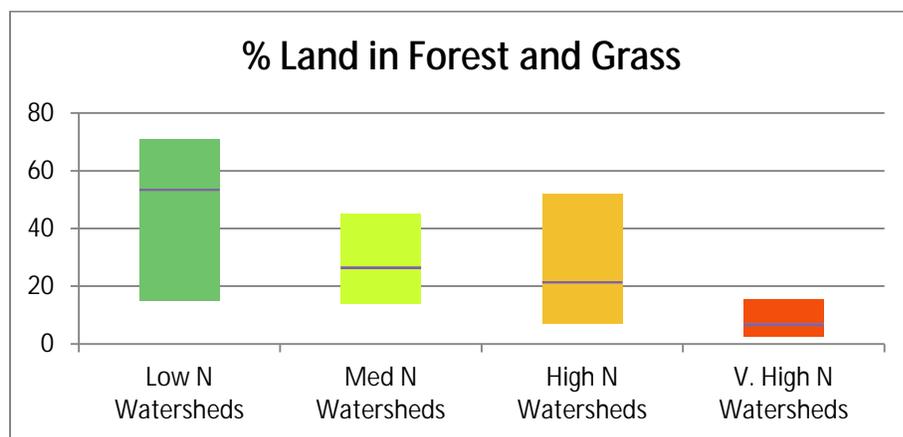


Figure 9. The range (colored bars) and mean (dark line) percent of land in forest and grasses for watersheds classified under each of the four river N level groupings (as listed in Table 3).

Human population

The range in human population densities among the four categories of N level watersheds does not show any definitive patterns (Figure 10), suggesting that differences in human population among the studied watersheds is not a major factor influencing water N ranges in the studied watersheds. Note, however, that the watersheds with major urban centers, such as the Twin Cities, Rochester, or Duluth, did not meet the watershed selection criteria and are not included among the watersheds assessed within this chapter. It is possible that if the evaluated watersheds had included larger urban areas that an effect from high human population centers would be observed.

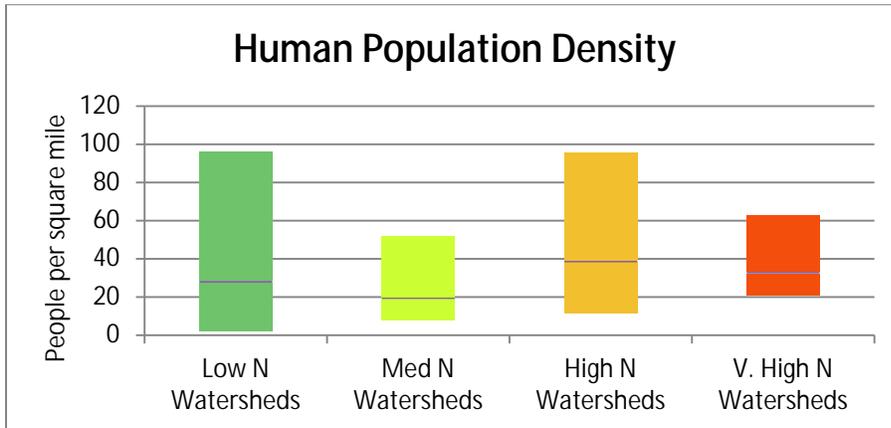


Figure 10. The range (colored bars) and mean (dark line) human population density for watersheds classified in each of the four river N level groupings (as listed in Table 3).

Irrigated agriculture

Differences in stream N levels did not appear to be closely associated with low or high percentages of the watershed under irrigation. The highest average percentage of land under irrigation was in the medium N watershed category (Figure 11). While irrigated fields could contribute N to localized surface waters, the total amount of irrigated acreage was less than 9% in all watersheds and was, therefore, not a dominant land use in any of the studied watersheds. Irrigation does not appear to be an important factor affecting the very high N level watersheds, as these five watersheds each had less than 1% of the land in irrigated agriculture.

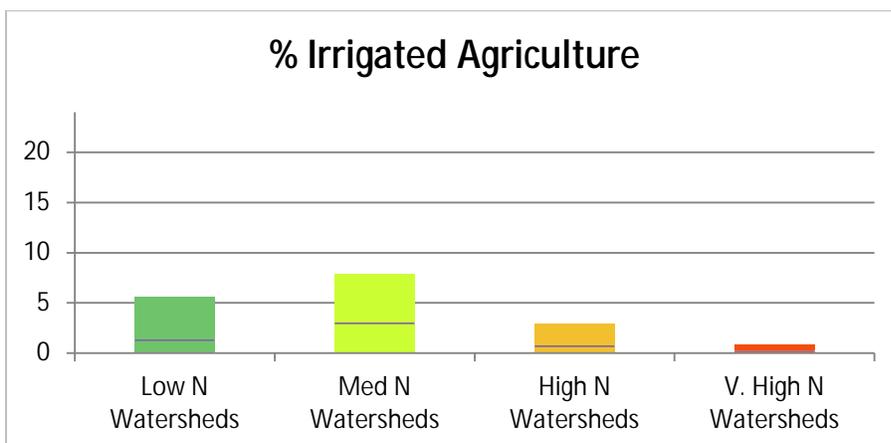


Figure 11. The range (colored bars) and mean (dark line) percent of land under irrigated agricultural production for watersheds in each of the four river N level groupings (as listed in Table 3).

Soil organic matter

Soil organic matter ranges and means were highest in the watersheds with the lowest surface water N levels, followed by the medium N watersheds (Figure 12). The high soil organic matter in the low N watersheds is likely attributable to the abundance of wetland and peat soils common in the northern part of the state where river N levels are low. The high and very high N watersheds had the lowest percent soil organic matter. Soil organic matter is one source of N to waters, but is transported to waters most readily when converted to mobile N forms through a mineralization process affected by temperature, soil moisture, and soil oxygen.

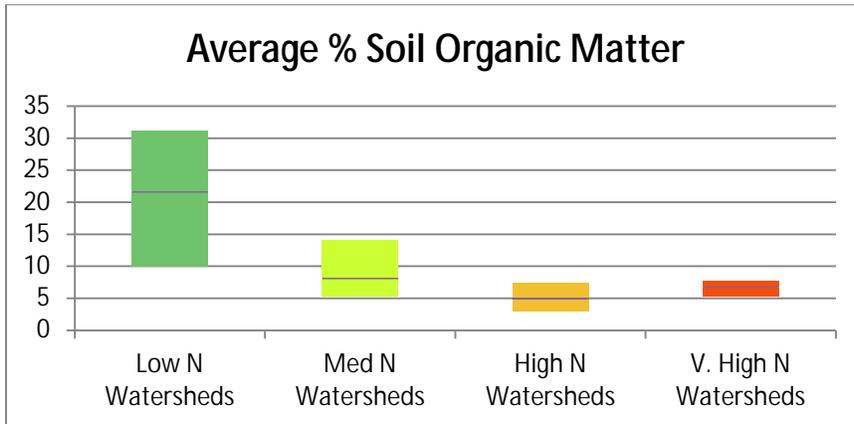


Figure 12. The range (colored bars) and mean (dark line) of the spatial average soil organic matter (%), in watersheds classified under each of four river N level groupings (as listed in Table 3).

Wetlands

The average percent of watershed land in wetlands is inversely related to river N levels (Figure 13). The high and very high river N watersheds have an average of about 3% of the watershed area in wetlands. The mean percent of land with wetlands increases to 8% and 29% in the medium and low N watershed categories, respectively. Wetlands remove considerable amounts of nitrate. However, the low N in watersheds with more wetlands is not necessarily attributable to the wetlands, since these same watersheds also have different land use, soils, and land cover as compared to the higher N loading watersheds.

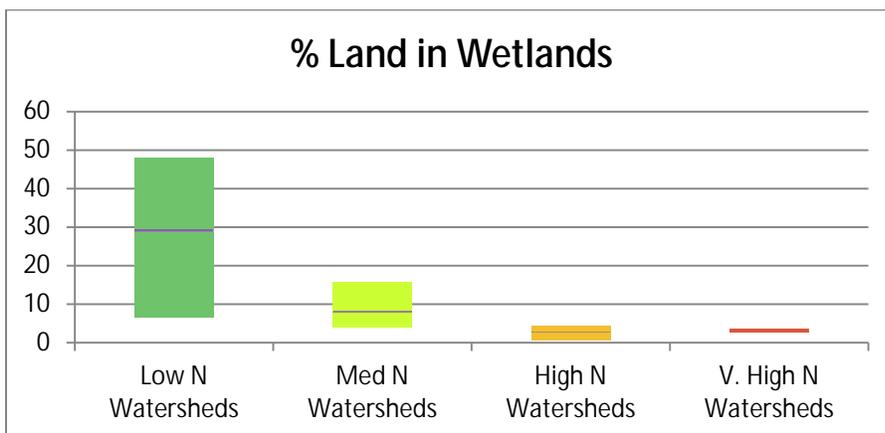


Figure 13. The range (colored bars) and mean (dark line) of the percentage of land with wetlands, in watersheds classified under each of four river N level groupings (as listed in Table 3).

Small grains

The watersheds with the most land in small grain production had low to medium N levels (Figure 14). The small grains are often grown in areas where soils and climate are less suitable for row crop production and, therefore, we cannot directly attribute small grains as a cause of high or low nitrate. Rather, we can only note that our high N watersheds are those with relatively low percentages of land planted to small grains.

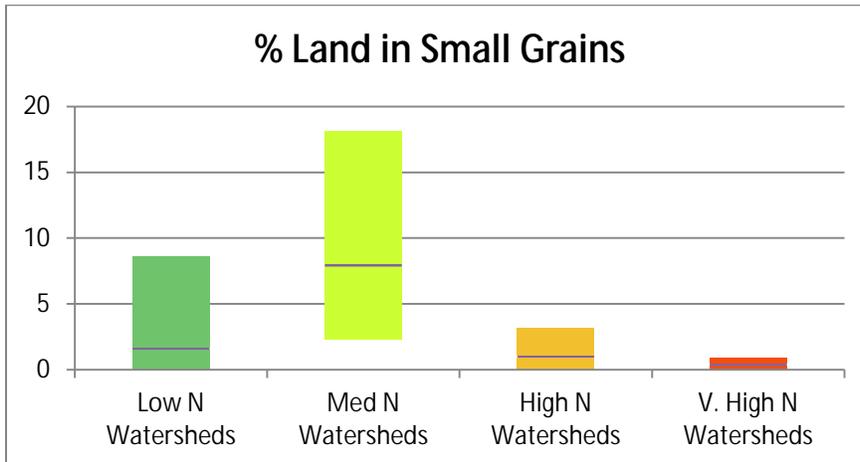


Figure 14. The range (colored bars) and mean (dark line) of the percentage of land in small grain production in watersheds classified under each of four river N level groupings (as listed in Table 3).

Precipitation

Average annual precipitation was slightly lower in the low and medium N category watersheds as compared to the high and very high N watersheds (Figure 15). However, there is considerable overlap in precipitation levels among the four N categories.

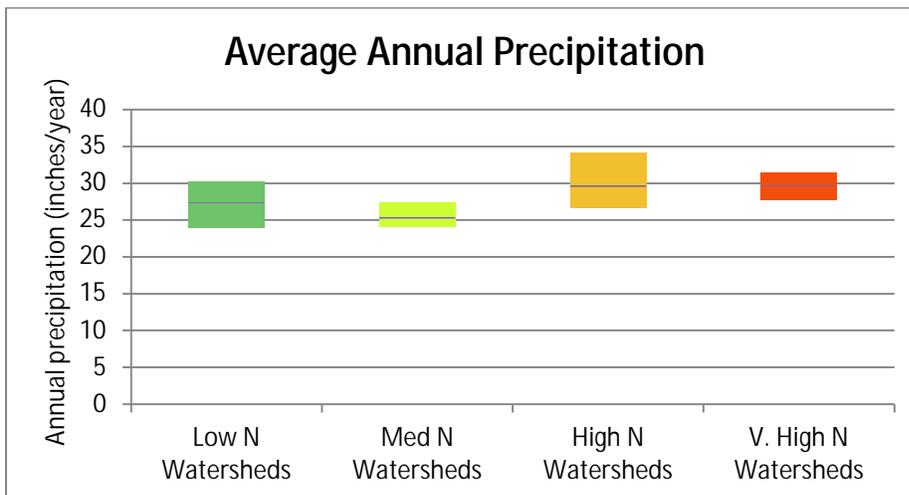


Figure 15. The range (colored bars) and mean (dark line) of the 30 year annual precipitation in watersheds classified under each of four river N level groupings (as listed in Table 3).

Land in row crops over sandy soils

The medium, high, and very high N watersheds each had similar percentages of land in row crop over sandy soils (Figure 16). The “low” river N watersheds had a lower fraction of land in row crop in general, and similarly had a lower percentage of row crops over sands as compared to the other watershed categories.

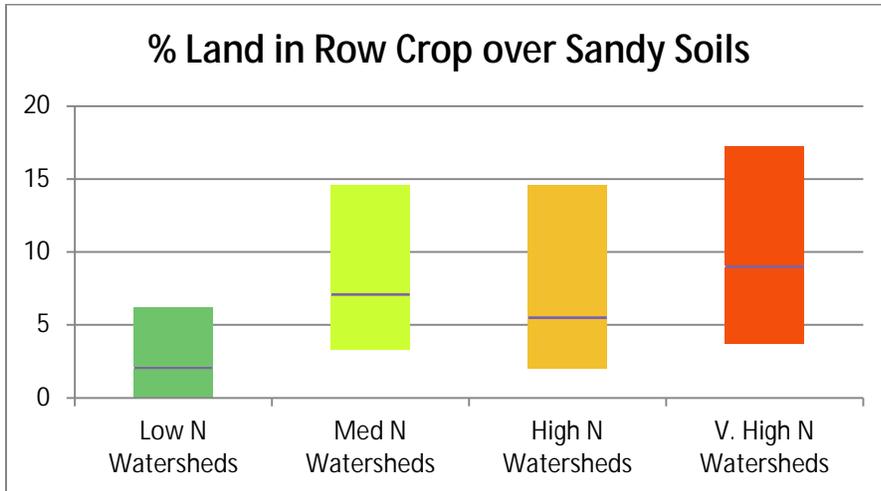


Figure 16. The range (colored bars) and mean (dark line) of the percentage of land in row crops over sandy subsoils, in watersheds classified under each of four river N level groupings (as listed in Table 3).

Land in row crops over shallow bedrock soils

The high and very high river N watersheds each had a couple of watersheds in regions with over 5% of the land having shallow depth to bedrock combined with row crop production. The low and medium N level categories did not have appreciable land with row crop over shallow depth to bedrock (Figure 17).

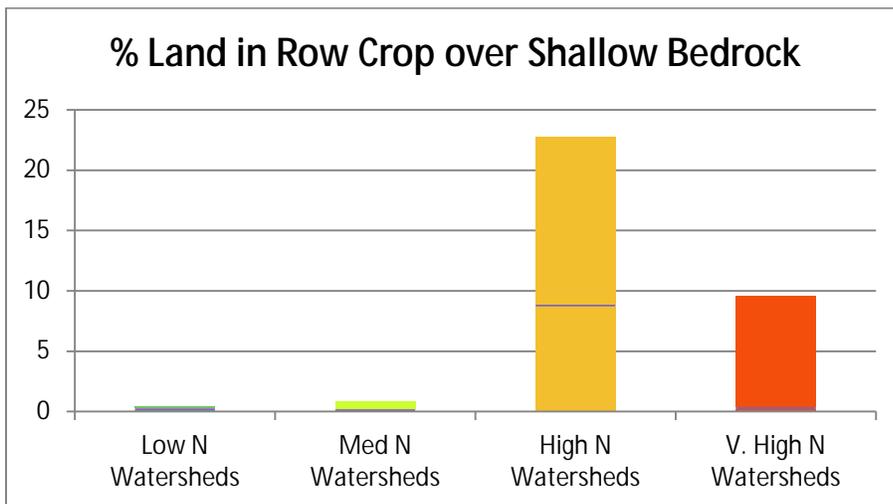


Figure 17. The range (colored bars) and mean (dark line) of the percentage of land in row crops over shallow depth to bedrock soils, in watersheds classified under each of four river N level groupings (as listed in Table 3).

Animal density

The mean watershed livestock density increases from 20 animal units (AU) per square mile in low N watersheds to 225 AUs per square mile in very high N watersheds (Figure 18). An AU is a measure used in feedlot regulations to approximate manure from a 1,000 pound beef cow. One AU represents 56 turkeys, or 0.7 dairy cows, or 3.3 finishing swine. The pattern in Figure 14 does not necessarily mean that livestock is a significant source of N in surface waters since livestock are concentrated in areas where other N sources, such as fertilizer, are also added to soil.

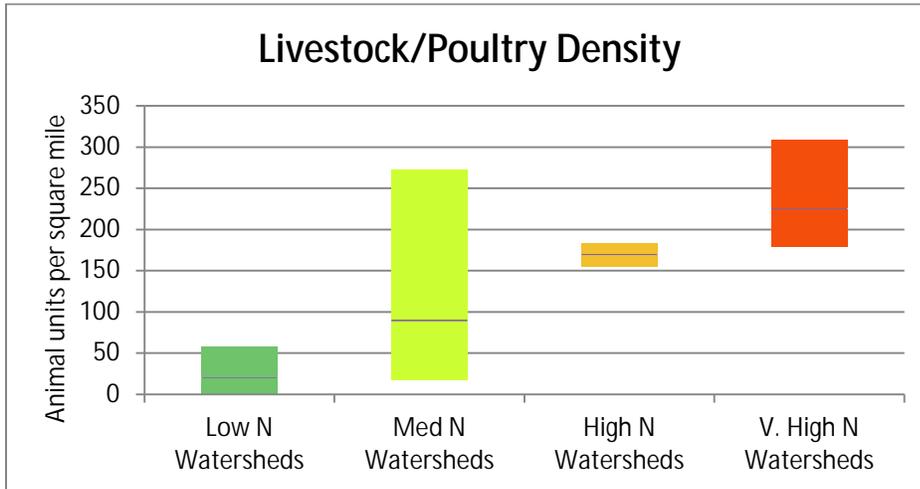


Figure 18. The range (colored bars) and mean (dark line) livestock and poultry AU density in watersheds classified under each of four river N level groupings (as listed in Table 3).

Row crops

The mean percent of watersheds in row crop production increases from about 5% in low N watersheds, to 39% in medium N watersheds, to 60% in high N watersheds, and 76% in very high N watersheds. Row crops are often located in areas that also have tile drainage and animal agriculture production. Therefore, we cannot conclude from this assessment that row crops are the key explanatory variable for stream N levels; rather it appears that row crops directly correlate with N levels in the watersheds used for this analysis.

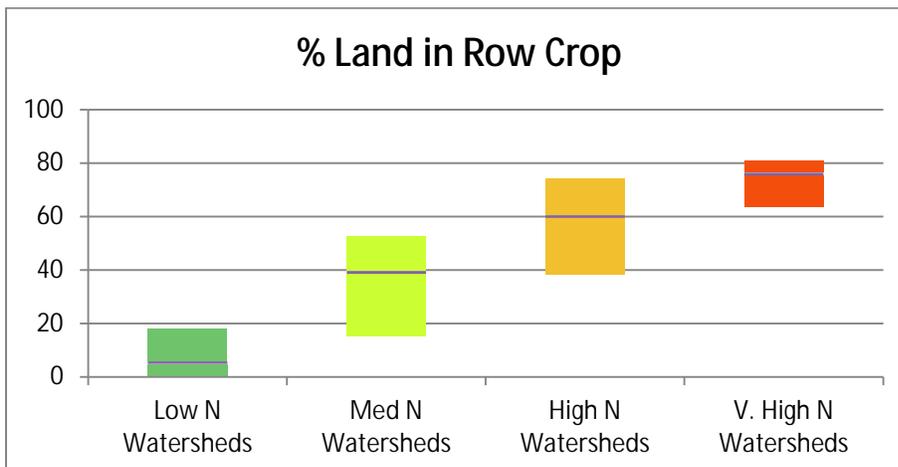


Figure 19. The range (colored bars) and mean (dark line) percent of land in row crop production for watersheds classified under each of the four river N level groupings (as listed in Table 3).

Tile drainage estimates

The relationship between watershed N level categories and percent of estimated tile-drained land (Figure 20) has a similar pattern as percent under row crop production. The mean percent of watershed with estimated tile-drained land is 0.2% in low N watersheds, 5% in medium N watersheds, 22% in high N watersheds, and 42% in very high N watersheds. The similarity between the row crop and tile drain variables is not unexpected because the criteria used to estimate tilled lands includes row crop production together with certain slope and soil conditions; thus, these variables are not independent of each other.

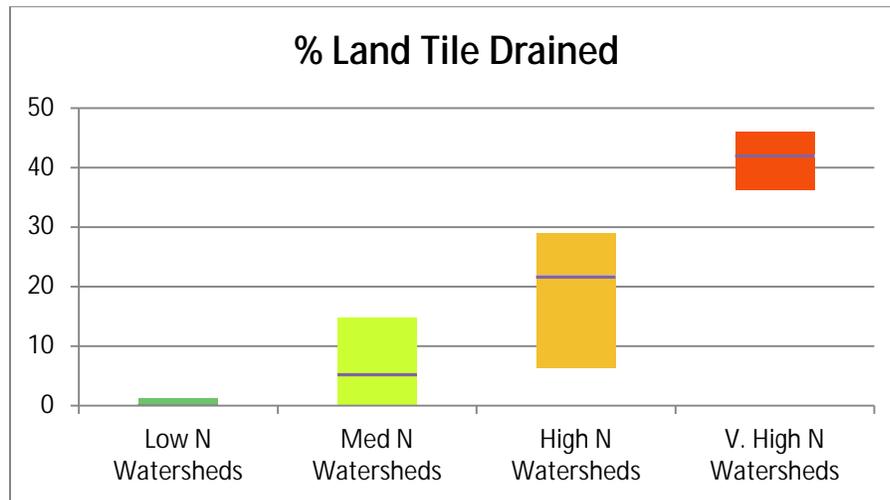


Figure 20. The range (colored bars) and mean (dark line) percent of land estimated to be tile-drained in watersheds classified under each of four river N level groupings (as listed in Table 3).

Row crops over leaky soils

The most distinct pattern observed between watershed N levels and land characteristics was with percent of row crop land in the watershed over leaky soils. “Leaky soils” included estimated tile-drained lands, sandy soils/subsoils, and shallow depth to bedrock (Figure 21). The four watershed N level categories each had a distinct and narrow range of percent row crop over leaky soils.

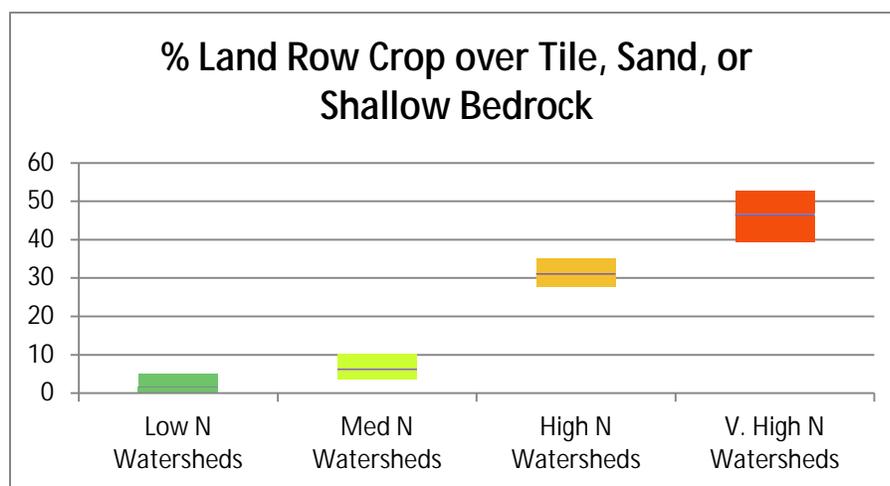


Figure 21. The range (colored bars) and mean (dark line) percent of land in row crops underlain by either tile-lines, shallow bedrock or sandy subsoils, in watersheds classified under each of four river N level groupings as listed in Table 3.

Patterns from graphs

The patterns of watershed characteristics associated with the low to very high river N levels do not show any inconsistencies with the UMN/MPCA source assessment described in chapters D1 to D4, and instead show several relationships which are generally consistent with the findings of the UMN/MPCA source assessment. A statistical analysis using this information is presented in the following section.

The Low N watersheds are characterized by having relatively high wetlands, high soil organic matter, high forest and grass-lands, and low row crop, low tile drainage, and low animal density. The very high N watersheds are characterized by having relatively low wetlands, low forest and grass, low small grain crops, and high row crop, high tile drainage, and high animal density.

Statistical assessment of watershed characteristics and river nitrogen

Methods

We used ordinary least squares (OLS) multiple linear regression analysis to examine relationships between four dependent variables and a set of 18 explanatory variables. Our four dependent variables were nitrite+nitrate (nitrate) flow weighted mean concentration (NO_x FWMC), TN FWMC, nitrate yield (NO_x Yield) and TN yield (TN Yield). Our 18 explanatory variables and their data sources are listed in Table 1. We considered many combinations of explanatory variables in an attempt to find the strongest regression models for each dependent variable. Scatter plots were examined using all combinations of dependent and explanatory variables. In cases where we found linear relationships (e.g., row crops and NO_x FWMC), the explanatory variables were included in preliminary regression models. Where relationships were non-linear, we used logarithmic and exponential transformations with the explanatory variables, and included the transformed variables in the preliminary models (e.g., forest/shrub and NO_x FWMC). Explanatory variables that had strong correlations with dependent variables were considered to be the best candidates for the preliminary regression models. We used tests of statistical significance for explanatory variables, statistical significance of overall models, distribution of model residuals, variable inflation scores (VIF) that measure variable collinearity, Akaike's Information Criterion (AIC) scores that measure overall model fit, and R-squared values to evaluate the strength of each preliminary regression model (Quinn and Keough 2002).

A number of the explanatory variables that we initially considered to be good candidates for inclusion in the final models were not statistically significant in the regression analysis. These included percent of watershed with forest and shrub; pasture, grass and hay; wetlands; human population density; livestock and poultry density; small grain cultivation; irrigated agriculture; and soil organic matter. Other explanatory variables were statistically significant in the analysis but were highly correlated with other explanatory variables, as indicated by high VIF scores. These included row crops, row crops over shallow depth to bedrock (RCD), row crops over sandy soils (RCS), row crops over shallow depth to bedrock or sandy soils (RCDS), row crops over shallow depth to bedrock or sand soils or tile drains (RCDST), row crops not over shallow depth to bedrock, sandy soils or tile drain (RCnDST), and tile drained areas. After completing this exploratory analysis we selected the strongest statistically significant models for each dependent variable.

Results

Equations for the final models are listed in Table 4 and results from the statistical tests are included in Table 5. The final models for each of the four dependent variables were statistically significant at the $p < 0.01$ level. All explanatory variables were statistically significant at the $p < 0.01$ level. All four models had high R-squared values, each over 0.9. And each model had a comparatively low AIC score; a lower AIC score indicates a stronger model fit. VIF for each model were below an acceptable threshold of 7.5 indicating that collinearity among explanatory variables was not significant. Jarque-Bera tests indicated that model residuals were normally distributed for all four models (ESRI 2012). This result suggests that the models are unbiased and that they capture the critical explanatory variables. Koenker tests for each model indicated that the model relationships exhibited stationarity or consistency across geographic space. Global Moran's Index tests confirmed a random spatial distribution of model residual for the NO_x and TN FWMC models; however, the Global Moran's Index tests for the NO_x and TN yield models showed statistically significant spatial autocorrelation in the residuals. This result is in contrast to the results from the Jarque-Bera and Koenker tests cited above. Spatial autocorrelation in model residuals indicates spatial clustering of high and low values, and suggests that the model is predicting well in some parts of the study area and not as well in others; this is usually caused by important explanatory variables being absent from the model, or non-stationarity in the model (ESRI 2012). We felt that the models did include the important explanatory variables, so we used geographically weighted regression (GWR), a method that specifically addresses non-stationarity, to determine whether non-stationarity was the cause of the spatial autocorrelation.

Geographically weighted regression calculates explanatory variable coefficients for each feature in the model, based on a set of neighboring features within a specified search radius, rather than the full dataset as in OLS, and thus allows model relationships to vary across space. We used a fixed distance search radius calculated by ArcGIS to be the optimal distance for model development based on model AIC scores (ESRI 2012). The calculated search distance was 91.18 miles. We ran GWR with the NO_x and TN yield models, and then ran Global Moran's Index tests on the GWR results. The Moran's Index for these models showed random spatial distribution of residuals, results that suggest non-stationarity was the issue with our original OLS models and the issue was resolved by using GWR. These results also indicate that the GWR models predict well across the study area and that the models are well specified and include the important explanatory variables. The GWR also gave lower AIC scores and higher R-squared when compared with the OLS models, suggesting a better fit with GWR for the NO_x yield and TN yield models. We also tested the GWR with the NO_x FWMC and TN FWMC but in both cases our AIC scores increased and R-squared values decreased compared to our original OLS results, suggesting that for the FWMC models, the GWR does not represent an improvement over the OLS method.

Table 4. Multiple regression equations for nitrite+nitrate-N flow weighted mean concentrations in mg/l (NOx FWMC); nitrite+nitrate-N yield in lbs/acre (NOx Yield); TN flow weighted mean concentration in mg/l (TN FWMC); and TN yield in lbs/acre (TN Yield). Explanatory variables include estimated percent of land with tile drain in the watershed (TD), percent row crop with shallow depth to bedrock or sandy soils (RCDS), and 30-year average precipitation. Explanatory variables were scaled to have a mean of 0 and standard deviation of 1 (method: ((value - mean) / standard deviation) and, therefore, these equations cannot be used for prediction with data not included in the original dataset.

Regression equations	Model
(0.13) (0.14) (0.14) NOx FWMC = 2.98 + 2.98 TD + 0.66 RCDS	Standard Errors OLS
(0.29) (0.33) (0.29) NOx Yield = 2.41 + 3.93 TD + 1.42 Precipitation 30 year average	Standard Errors (mean values) GWR (mean values)
(0.13) (0.14) (0.14) TN FWMC = 4.33 + 3.24 TD + 0.66 RCDS	Standard Errors OLS
(0.39) (0.44) (0.39) TN Yield = 4.08 + 4.22 TD + 1.76 Precipitation 30 year average	Standard Errors (mean values) GWR (mean values)
Acronyms	
Tile Drainage	TD
Row crops over shallow depth to bedrock or sandy soils	RCDS
Ordinary Least Squares Regression	OLS
Geographically Weighted Regression	GWR

Table 5. Model parameters and test results

	NOx FWMC	NOx yield	TN FWMC	TN yield
Sample size	28	28	28	27
Adjusted R-squared	0.96	0.98	0.97	0.98
AIC	63.48	73.12	65.15	85.44
Model p-value	< 0.01	< 0.01	< 0.01	< 0.01
VIF	1.22	1.26	1.22	1.29
Model	OLS	GWR	OLS	GWR
Moran's Index score	0.12	0.18	0.06	0.12
Moran's Index z-score	1.15	1.56	0.70	1.20
Moran's Index p-value	0.25	0.12	0.49	0.23
GWR Search Radius	NA	91.18 miles	NA	91.18 miles

Discussion

The N concentration models (NOx FWMC and TN FWMC) suggest that row crop practices using tile drainage and row crop practices on naturally sensitive lands with high groundwater recharge explain much of the nitrate concentration variability in the 28 Minnesota rivers. Sensitive lands in this context are defined as areas that have a depth to bedrock of less than 50 feet, or sand content in the subsoil greater than 85%, or both. The N yield models (NOx Yield and TN Yield) suggest that N yields in the 28 watersheds are influenced largely by row crop practices using tile drainage and by precipitation. That

precipitation is a significant explanatory variable in the yield models is not surprising since yield (pounds/acre/year) for any chemical parameter is affected by river flow, which, in turn, is largely influenced by precipitation.

We scaled the explanatory variable data to have a mean of zero and a standard deviation of one, so that a comparison of the relative strength of the variable coefficients in influencing N level variability would be possible. As shown in the concentration equations in Table 4, the influence of the estimated tile drain variable on nitrate has four and half times the magnitude of the influence of the RCDS variable, and for TN tile drainage has almost five times the magnitude of influence as RCDS. In the yield equation for nitrate, the estimated tile drain variable has nearly three times the influence of the precipitation variable, and for TN yield it has more than two times the influence. These coefficient values suggest that the amount of watershed land in tile drainage is the leading predictor of river nitrate and TN concentrations and yields.

In addition, the GWR analysis for the N yield models showed specific spatial trends in the model relationships, as indicated by the variance in the explanatory variable coefficients. Specifically, the model coefficients for estimated percent of watershed with tile drainage, and mean annual precipitation are higher in southern Minnesota than in the northern half of the state (Figures 1-4). This result suggests that with higher amounts of tile drainage and precipitation in the study watersheds, these explanatory variables have increased influence on levels of nitrate and TN yield in corresponding rivers.

Maps showing the spatial pattern of explanatory variables in the regression models are included in Appendix E2-1. Maps showing the GWR coefficients for explanatory variables in the NOx and TN yield models are included in Appendix E2-2. And scatter plots showing relationships between dependent variables and the explanatory variables in the regression models are included in Appendix E2-3.

Summary

The strong correlation between estimated tile drained lands and high nitrate and TN yields and FVMCs is generally consistent with the UMN/MPCA source assessment findings (Chapters D1-D4) showing tile drained cropland as the largest contributor to N loads in the state. The source assessment showed that cropland groundwater was the second highest N source/pathway. This is somewhat consistent with the statistical modeling results showing that cropland over potentially high groundwater recharge lands (shallow bedrock and sandy soils) was another important variable correlated with nitrate and TN FVMCs. The cropland over shallow bedrock and sands variable was not, however, found to be a key explanatory variable affecting nitrate or TN yield in the best statistical models.

The UMN/MPCA N source assessment also showed that loads/yields are highly dependent on precipitation. This is generally consistent with the best statistical models for N yield, which showed that average annual precipitation in the watershed was the second most important variable after tile drainage affecting variability in watershed nitrate and TN yields. Future analyses should assess whether groundwater recharge, integrating precipitation and geologic sensitivity, over cropland is correlated with nitrate and TN yield.

As noted earlier, the statistical analyses do not show causes, but relationships. The multiple regression analyses, along with the single variable graphs and scatter plots, did not show results that are inconsistent with the source assessment findings, and there were several relationships which supported the source assessment findings.

References

- Allan DJ. 2004. Landscapes and Riverscapes: The Influence of Land Use on Stream Ecosystems. *Annual Review of Ecology, Evolution, and Systematics*. Vol. 35, pp. 257-284.
- David, Mark B. 2010. Sources of Nitrate Yields in the Mississippi River Basin. *Journal of Environmental Quality* 39: 1657-1667.
- ESRI. 2012. ArcGIS 10.1. Environmental Systems Research Institute. Redlands, California.
- Mitchell, A. 2005. *GIS Analysis: Spatial Measurements and Statistics*. ERSI Press. Redlands, California.
- Quinn GP and MJ Keough. 2002. *Experimental Design and Data Analysis for Biologists*. Cambridge University Press. Cambridge.
- Sugg, Zachary. 2007. *Assessing U.S. Farm Drainage: Can GIS Lead to Better Estimates of Subsurface Drainage Extent*. World Resources Institute. Washington, DC.
- Turner MG, Gardner RH, O'Neill RV. 2001. *Landscape Ecology in Theory and Practice: Pattern and Process*. Springer. New York.