Impacts of Land Use on Water Quality in Minnesota

https://github.com/lhr12/HDA_Project

Keith Bollt, Jake Greif, Felipe Raby-Amadori, & Lindsay Roth

Contents

List of Tables

List of Figures

1 Rationale and Research Questions

Land use has a large impact on nutrient runoff into streams, lakes, and other water bodies. Understanding the causes of nutrient problems will better inform management strategies. A study looking at land use and water quality in the Chaohu Basin in China found that forest and grassland has a statistically significant negative correlation with water quality, while urbanization has a statistically significant positive correlation with water quality and farmland has a more complex relationship (Huang et al., 2013). We applied some of the basic research questions of the Huang et al. study to the United States. In the US, nutrient management has been a challenge for states trying to control harmful algal blooms and coastal dead zones in water bodies like the Gulf of Mexico (Wurtsbaugh et al., 2019).

We selected the state of Minnesota for our study because it has wide variety of land uses including a large urban center (the Twin Cities metropolitan area), natural lands in the northern part of the state, and large agricultural areas. It also has a large number of lakes with water quality sampling. Prior studies have shown that water quality in Minnesota has been adversely affected by land use changes (Christensen et al., 2012); (Maalim et al., 2013). We also want to look for seasonal relationships between water quality and land use. For example, we know that farmers apply nutrients at certain times of year and lakes naturally have higher levels of algae in summer months.

It is important that states like Minnesota have sound nutrient management regimes because they affect water quality downstream of their borders. The Gulf of Mexico has been experiencing hypoxia, or low dissolved oxygen levels, as a result of agricultural runoff in the Mississippi River basin. Minnesota is the source of the Mississippi River and a large agricultural state, so it is important that Minnesota properly manages its agricultural runoff (Minnesota Pollution Control Agency, 2019). Our project hopes to make management recommendations for Minnesota based off our findings.

In order to evaluate the relationship between land use and lacustrine water quality in the state of Minnesota, we asked the following two questions to help guide our research:

- 1. What are the predictors of nutrients based on land use in watersheds in the state of Minnesota?

    - Hypothesis 1a: The percentage of urbanized land in a lake's watershed in the state of Minnesota will have a statistically significant negative effect on water quality, as measured by higher chlorophyll a concentrations and lower secchi depth measurements.
    - Hypothesis 1b: The percentage of agricultural land in a lake's watershed in the state of Minnesota will have a statistically significant negative effect on water quality, as measured by higher chlorophyll a concentrations and lower secchi depth measurements.
    - Hypothesis 1c: The percentage of undeveloped land in a lake's watershed in the state of Minnesota will have a statistically significant positive effect on water quality, as measured by lower chlorophyll a concentrations and higher secchi depth measurements.

- 2. How do you characterize seasonal variation between the predictors of nutrients?
    - Hypothesis 2: The water quality in Minnesota lakes will demonstrate a statistically significant seasonal pattern.

By testing our hypotheses, we aim to accomplish the following goals:

- 1. Determine how land use, watershed size, and ecoregion explain variation in nutrient loading indicators.
- 2. Discern whether there a seasonal trends in nutrient loading indicators based on land use types, watershed size, and ecoregion.
- 3. Provide insight to inform decisions about nutrient management practices based on land use types, watershed size, and ecoregion.

2 Dataset Information

The data used in this analysis include data from the Lake Multi-Scaled Geospatial and Temporal Database (LAGOSNE) and the EPA ecoregion spatial datasets.

LAGOSNE is a collection of several data modules that contain information on lakes in the northern United States. The modules contain data from thousands of lakes in 17 states in the northeastern and Midwestern United States, from Missouri to Maine. The dataset includes a complete list of all lakes bigger than 4 hectares in the 17 state area, and water quality data on a large number of lakes, spanning every state.

Ecoregions are developed by the US Environmental Protection Agency and are used by planning managers to understand the type of land use that occurs in different regions of the United States. There are different levels of ecoregions. Level I has low resolution, and divides North America into 15 ecological regions, while Level IV offers fine ecological resolution for each state. This data was published by the U.S. EPA Office of Research and Development (ORD) - National Health and Environmental Effects Research Laboratory (NHEERL). Minnesota contains parts of three distinct Level II ecoregions, which we selected for our project. This level of resolution allows us to broadly capture the natural ecosystem variation in different parts of the state in our statistical models with sufficient water quality data per ecoregion, conditions that could not be met with higher resolution ecoregions.

Table 1 presents a summary of the data structure that is used.

Table 1: Data structure, including variables, descriptions, units, variable types, and source.

Column.Name	Description	Units	Variable.Type	Source
chla	Chlorophyll a	ug/L	Dependent	LAGOS-NE
secchi	Secchi depth	meters	Dependent	LAGOS-NE
IntenseUrban.pct	% med. and high intensity urban	%	Independent	LAGOS-NE
OpenUrban.pct	% developed	%	Independent	LAGOS-NE
Barren.pct	% bedrock and desert pavement	%	Independent	LAGOS-NE
Forest.pct	% deciduous	%	Independent	LAGOS-NE
GrassShrub.pct	% shrubs dominated	%	Independent	LAGOS-NE
Wetland.pct	% wetlands	%	Independent	LAGOS-NE
Pasture.pct	% areas of grasses	%	Independent	LAGOS-NE
RowCrop.pct	% cultivated crops	%	Independent	LAGOS-NE
LakeIWS.Ratio	Lake-watershed area ratio	N/A	Independent	LAGOS-NE
Season	Early/Prime/Late	N/A	Independent	-
Ecoregion	Level II Ecoregions	N/A	Independent	U.S. EPA

3 Exploratory Analysis

3.1 Overview

We came in to our exploratory analysis knowing we wanted to look at how water quality was affected by varying land use. Our exploratory analysis had four main goals: selecting a study area, selecting a study period, selecting what land use classifications (independent variables) and water quality proxies (dependent variables) to study, and preparing a dataframe to test our research questions and hypotheses.

3.2 Data Compilation

Our first step was to load the LAGOSNE data columns we thought we would need and to look at how much data each state in our study contained. We decided on Minnesota because of its high number of lakes in the LAGOSNE dataset, and its high land use variation. We observed that the latest land use classification by USGS was from 2011; the next most recent was from 2006. Therefore, we limited our study to the period from 2009 through 2014 inclusive, so we did not extrapolate too much from our 2011 land use dataset.

3.2.1 Visualize Temporal Distribution & Abundance of Data

We filtered out the variables we wanted for our study and created a "LAGOS.MN.processed" dataframe. We decided to use chlorophyll a and secchi depth as proxies of water quality after we visualized our data and confirmed that we had enough data on these two variables in Minnesota between 2009 and 2014 to make claims about land use and water quality; in fact we had over 19,000 lakes with data. Observe in Figure 1 and Table 2 that secchi depth data in Minnesota is available for a large number of lakes and for a large portion of the year, with gaps in the data every winter. Observe in Figure 2 and Table 3 a similar pattern in data availability for chlorophyll a.

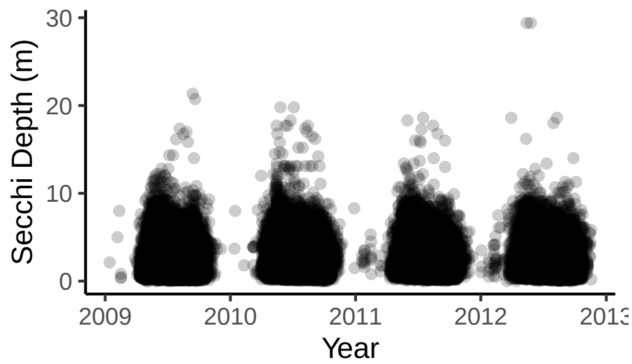

Figure 1: Secchi Depth a vs. Time. This plot provides a sense of how secchi depth measurements are distributed between and within years, as well as the general range of secchi depth measurements. Sampling occurs during the middle portion of each year, and ceases during the winter season.

Table 2: Number of secchi depth measurements in each of the years to be included in the analysis, which provides a sense of the amount of data, as well as the distribution of data.

Year	Count
2009	20849
2010	19829
2011	18025
2012	16924

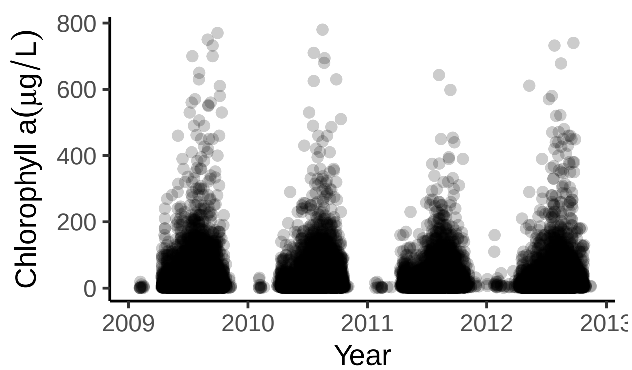

Figure 2: Chlorophyll a vs. Time. This plot provides a sense of how chlorophyll a measurements are distributed between and within years, as well as the general range of chlorophyll a measurements. Sampling occurs during the middle portion of each year, and ceases during the winter season.

Table 3: Number of chlorophyll a measurements in each of the years to be included in the analysis, which provides a sense of the amount of data, as well as the distribution of data.

Year	Count
2009	8996
2010	7576
2011	6989
2012	6206

3.3 Explore Trends in Variables by Land Use & Season

We decided which land uses we were interested in, based on our background readings and prior knowledge of water quality and land use, grouped the land uses together, and assigned them names in our dataframe. Considering the National Land Cover Database 2011 (NLCD2011) included in LAGOS-NE, we defined IntenseUrban as the sum of 23-Developed, Medium Intensity and 24-Developed, High Intensity; OpenUrban as 21-Developed, Open Space; Barren as 31-Barren Land (Rock/Sand/Clay); Forest as the sum of 41-Deciduous Forest, 42-Evergreen Forest, and 43-Mixed Forest; GrassShrub as the sum of 52-Dwarf Scrub and 71-Grassland/Herbaceous; Wetland as the sum of 90-Woody Wetlands and 95-Emergent Herbaceous Wetlands; Pasture as 81-Pasture/Hay; and finally RowCrop as 82-Cultivated Crops.

We also created a lake - interlake watersheds (IWS) ratio column, which is the lake area divided by the lake's watershed area. We reasoned a big lake in a small watershed would be affected by land use differently than a small lake in a big watershed, and we wanted this to be captured in our analysis.

Our next step was to create a growing "seasons" column in order to account for seasonality in our statistical models. We divided up the calendar year into three bins: Early (January 1 - May 15), Prime (May 16 - September 30), and Late (October 1 - December 31). These season designations are somewhat arbitrary, but are based off of our prior knowledge of when the prime biological growing season is in Minnesota. We created a dataframe for each of the three seasons.

A summary of the structure of the resulting dataset is shown in table 1.

With Figures 3, 4, and 5 we performed a preliminar exploratory visualization of our dataset. It includes a histogram of the variables, the relationship, and correlation between them for each season. It can be seen that the correlations numbers are in an acceptable range and that the histograms are showing that the variables are not normally distributed.

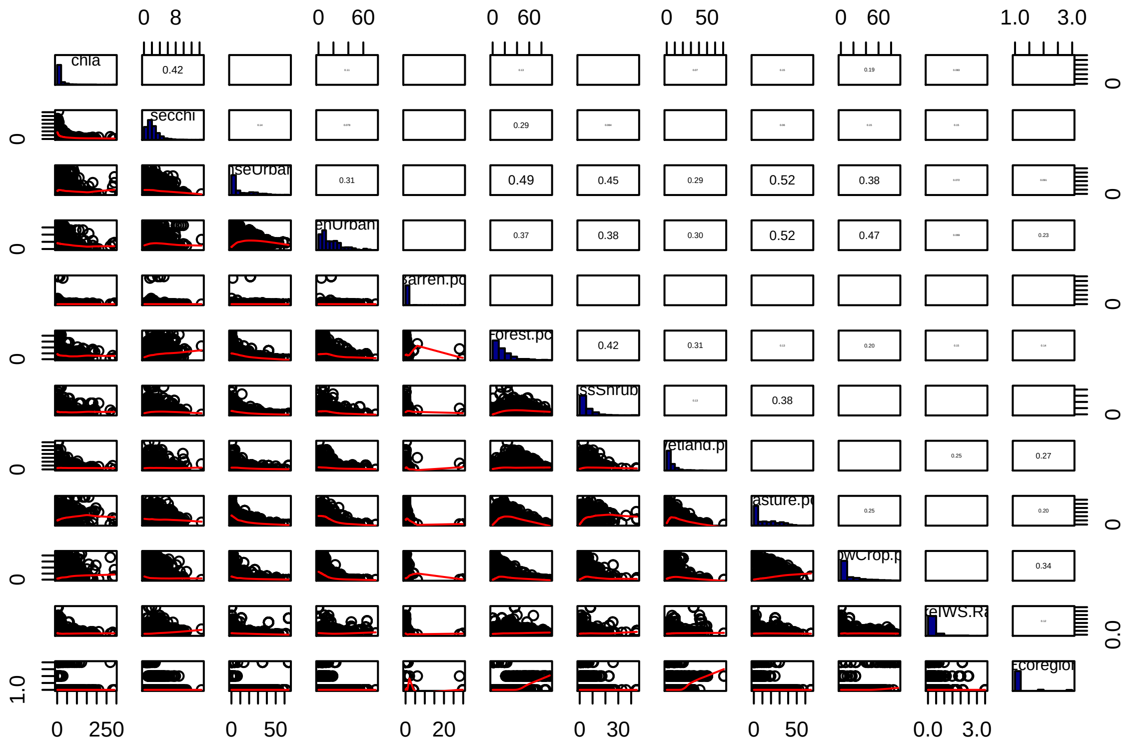

Figure 3: Plot displaying distributions of variables of interest, relationship, and correlations between them for Early Season

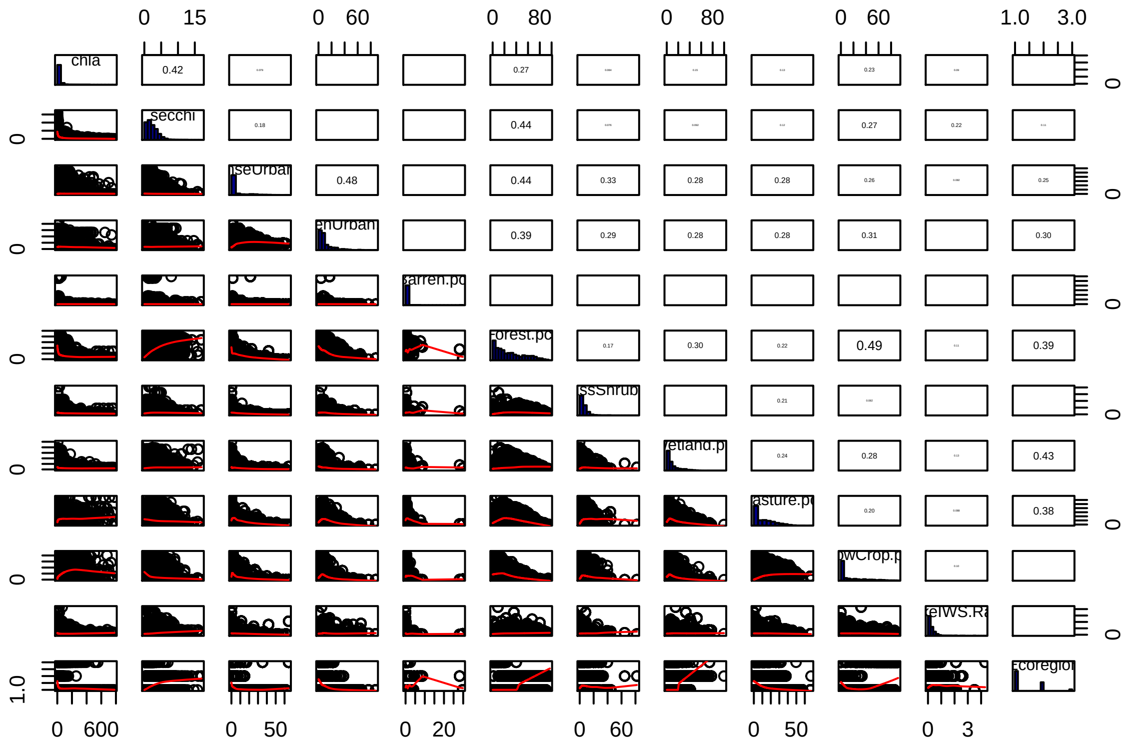

Figure 4: Plot displaying distributions of variables of interest, relationship, and correlations between them for Prime Season

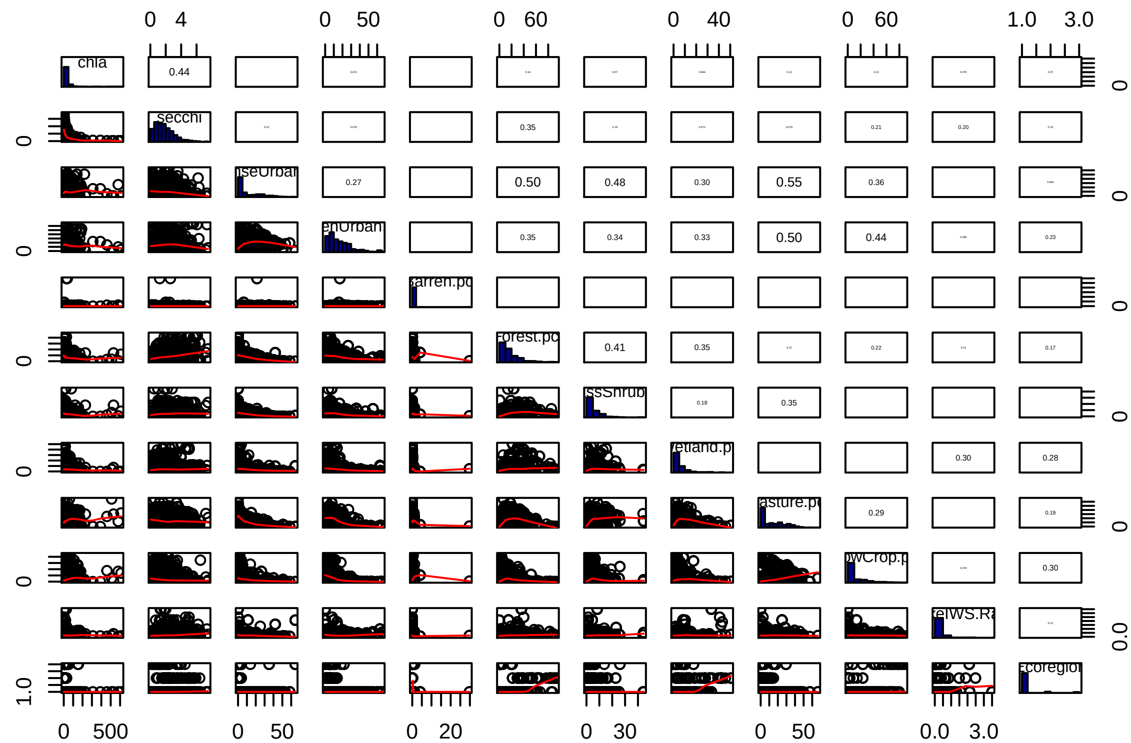

Figure 5: Plot displaying distributions of variables of interest, relationship, and correlations between them for Late Season

3.3.1 Visualize Trends in Variables by Land Use

Next, we visualized land use percentage versus secchi depth and versus chlorophyll a for eight representative land uses for the Late season. We aimed to visually confirm that we might find a relationship between land use and our two response variables when we ran our statistical tests in our analysis section. Specifically, we expected to see an increase in secchi depth and a decrease in chlorophyll a with an increasing percentage of undeveloped land. Figures 6 and 7 explore the trends in secchi depth and chlorophyll a by the percentage of each land use type in the late season. The late season was chosen for exploration because it had to lowest and highest mean secchi depth and mean chlorophyll a, respectively. Observe in Figure 6 (secchi depth) and Figure 7 (chlorophyll a) that some ggplots showed a possible relationship between some land uses and our response variables in the directions that we predicted in Hypotheses 1a, 1b, and 1c, whereas some ggplots did not appear to show a relationship.

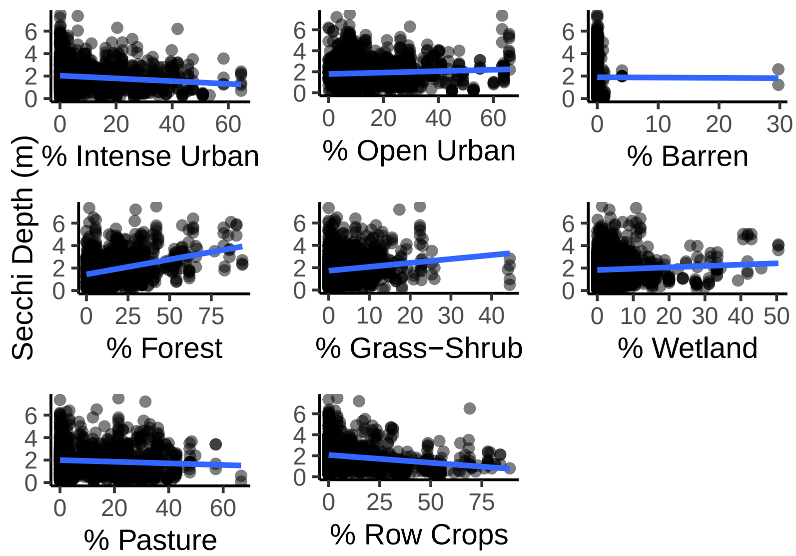

Figure 6: Plots displaying trends in secchi depth relative to the given land uses during the 'Late Season'. The presence of trends supports the inclusion of the above land uses in the analysis. Note that the x-axes are on different scales.

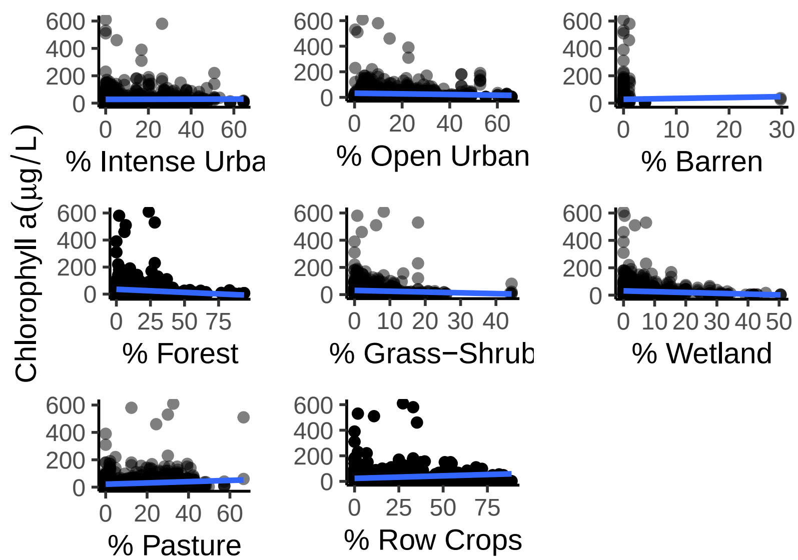

Figure 7: Plots displaying trends in chlorophyll a concentrations relative to the given land uses during the 'Late Season'. The presence of trends supports the inclusion of the above land uses in the analysis. Note that the x-axes are on different scales.

Because of the presence of some relationships in Figures 6 and 7, we determined that we had, in fact selected useful land use classifications and useful seasonal demarcations. We performed a summary function, and gave all of our columns useful names and removed NAs. Our statistical tests would not all run if our dataset contained NAs. The next step we performed was to convert our dataframes to sf dataframes so we could add Level II ecoregion column.

3.3.2 Visualize Spatial Distribution of Data in Each Season

Our final exploratory analysis test was to see if we had enough lake data for each Level II ecoregion for each of our three seasons. We created three tables, one for each season, with a lake sample count for each ecoregion. See Tables 4, 5, and 6. From these tables, we learned that we had enough data to proceed with our statistical analysis, and therefore Level II ecoregions were the correct resolution for our study. We also displayed this information spatially in Figures 8, 9, and 10; each figure corresponds to a different season. We discovered that there are a large number of lake samples around the Twin Cities area, and that there is enough spatial variation in our data to capture the three different Level II ecoregions for all three of our seasons. We concluded our data wrangling by exporting our three non-sf season processed dataframes as csv files to our processed data folder.

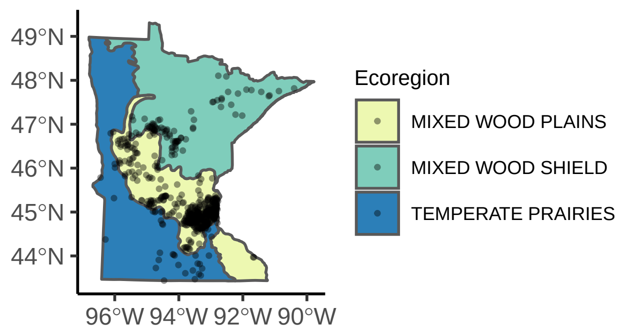

Figure 8: Visual representation of the available secchi depth and/or chlorophyll a data in each Ecoregion during the 'Early Season', which confirms that there are a suitable number of lakes in each Ecoregion.

Table 4: Number of lakes with available secchi depth and/or chlorophyll a data in each Ecoregion during the 'Early Season.'

Ecoregion	Count
MIXED WOOD PLAINS	454
MIXED WOOD SHIELD	82
TEMPERATE PRAIRIES	37

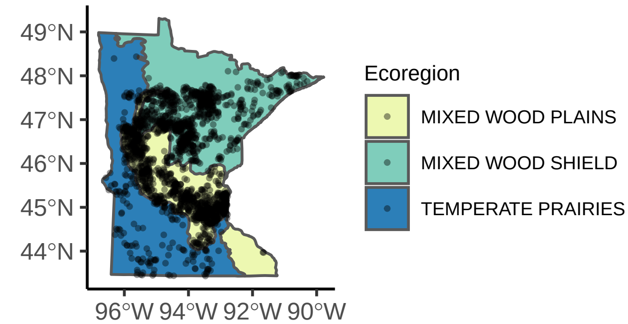

Figure 9: Visual representation of the available secchi depth and/or chlorophyll a data in each Ecoregion during the 'Prime Season', which confirms that there are a suitable number of lakes in each Ecoregion.

Table 5: Number of lakes with available secchi depth and/or chlorophyll a data in each Ecoregion during the 'Prime Season.'

Ecoregion	Count
MIXED WOOD PLAINS	1055
MIXED WOOD SHIELD	860
TEMPERATE PRAIRIES	180

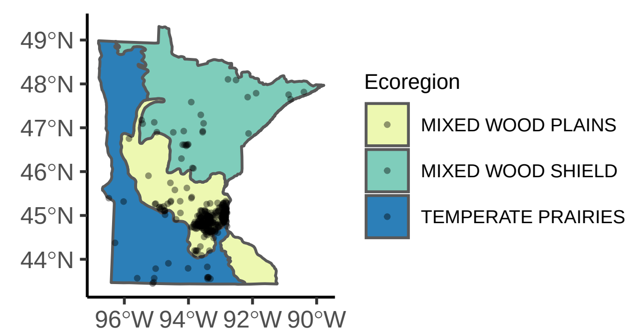

Figure 10: Visual representation of the amount of lakes with available secchi depth and/or chlorophyll a data in each Ecoregion during the 'Late Season', which confirms that there are a suitable number of lakes in each Ecoregion.

Table 6: Number of lakes with available secchi depth and/or chlorophyll a data in each Ecoregion during the 'Late Season.'

Ecoregion	Count
MIXED WOOD PLAINS	278
MIXED WOOD SHIELD	28
TEMPERATE PRAIRIES	19

4 Analysis

Before creating our models, we tested for correlation between all of our variables to make sure none of them needed to be excluded. Then, we tested for normality of the distribution of our variables using qqplots and Shapiro-Wilkes tests. We used the function BestNormalize to determine what the best transformation for each variable would be if it were needed and used that transformation accordingly. We log-transformed chlorophyll a and secchi depth. We did not find an effective transformation for our land use variables or the lake IWS ratio, which we expected because they are proportion and ratio values. Therefore, we used them in our models with their original distributions.

For our analysis, we ran mixed effect linear models with land use as fixed effects and ecoregion as random effects in order to account for variability between ecoregions without considering each ecoregion as its own factor. We created models for both Chlorophyll a and Secchi depth as response variables for each of the three seasons making a total of six models. To determine the most parsimonious models, we eliminated non-significant variables with the highest p-values one by one until all remaining variables were significant. To check that this model was the best fit for the data, we ran an ANOVA on all of the models together to determine which model had the lowest AIC. If our simplest model had the lowest AIC or it's AIC was not more than 3 points away from the lowest score, we chose that model as our best fit.

4.1 Research Questions

1. What are the predictors of nutrients based on land use in watersheds in the state of Minnesota?
2. How do you characterize seasonal variation between the predictors of nutrients?

4.2 Results

Instead of splitting our results into two subsections, one for each research question, we explore both questions simultaneously. We formatted our results section in this way because our linear models test both questions at the same time.

4.2.1 Early Season Chlorophyll a

The significant predictors of early season chlorophyll a are Open Urban Percent ($p = 2.453 * 10^{-5}$), Forest Percent ($p < 2.2 * 10^{-16}$), Pasture Percent ($p = 2.853 * 10^{-5}$), Row Crop Percent ($p = 0.04671$), and Lake IWS Ratio ($p = 2.737 * 10^{-6}$). The marginal $R^2 = 0.1058$ (represents the variance explained by the fixed effects), and the conditional $R^2 = 0.1058$ (interpreted as a variance explained by the entire model, including both fixed and random effects). Table 7 shows the coefficients for each significant explanatory variable.

Table 7: Model Coefficients. Early Season Chlorophyll a.

Predictor	Coefficient
Open Urban	-0.01077
Forest	-0.01526
Pasture	0.00850
Row Crop	0.00320
Ratio	-0.30862

Since the outcome variable of the model is log transformed, the coefficients must be exponentiated to be interpreted correctly. Table 8 presents the percentage increase or decrease that a 1% increase of the explanatory variables have on early season chlorophyll a.

Table 8: Transformed Model Coefficients. Early Season Chlorophyll a.

Predictor	Coefficient
OpenUrban	-2.50%
Forest	-3.50%
Pasture	2.00%
RowCrop	0.70%
LakeIWS.Ratio	-50.90%

Figure 11 shows Early Season log Chlorophyll a vs. Lake IWS Ratio together with the Mixed Effects Model that we determined.

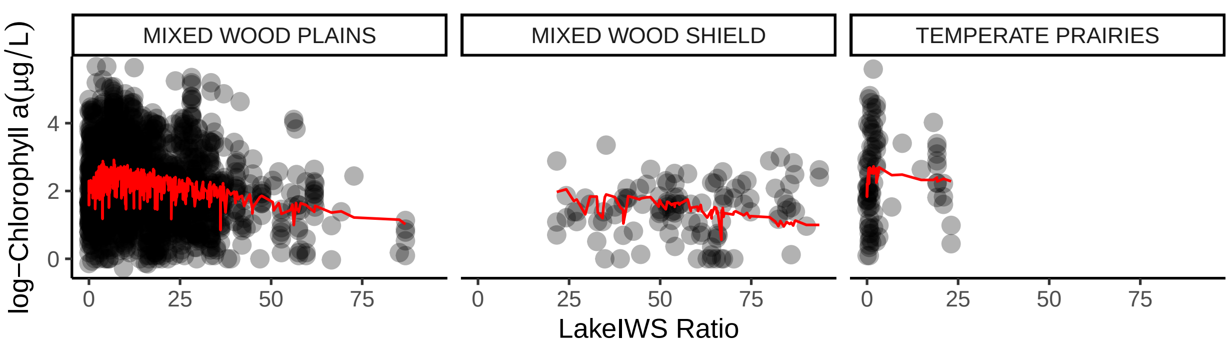

Figure 11: Early Season log Chlorophyll a vs. Lake IWS Ratio and Mixed Effects Model

4.2.2 Early Season Secchi Depth

The significant predictors of early season secchi depth are Open Urban Percent (p < 2.2 * 10^-16), Barren Percent (p = 0.0316103), Forest Percent (p = < 2.2 * 10^-16), Grass Shrub Percent (p = 0.0007119), and Lake IWS Ratio (p = 1.856 * 10^-7). The marginal R^2 = 0.1081, and the conditional R^2 = 0.2648. This tells us that 16% of the variability in secchi depth in the early season can be explained by the variance of ecoregion. Table 9 shows the coefficients for each significant explanatory variable.

Table 9: Model Coefficients. Early Season Secchi Depth.

Predictor	Coefficient
OpenUrban	0.01237
Barren	0.01956
Forest	0.01229
GrassShrub	0.01024
LakeIWS.Ratio	0.21442

Since the outcome variable of the model is log transformed, the coefficients must be exponentiated to be interpreted. Table 10 presents the percentage increase or decrease that a 1% increase of the explanatory variables have on early season secchi depth.

Table 10: Transformed Model Coefficients. Early Season Secchi Depth.

Predictor	Coefficient
OpenUrban	2.90%
Barren	4.60%
Forest	2.90%
GrassShrub	2.40%
LakeIWS.Ratio	63.80%

Figure 12 shows Early Season log Secchi Depth vs. Lake IWS Ratio together with the Mixed Effects Model that we determined.

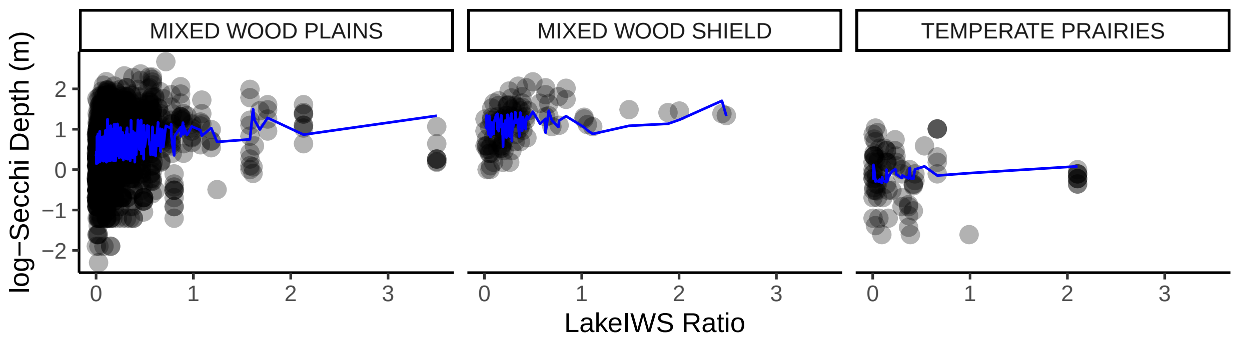

Figure 12: Early Season log Secchi Depth vs. Lake IWS Ratio and Mixed Effects Model

4.2.3 Prime Season Chlorophyll a

The significant predictors of prime season chlorophyll a are Intense Urban Percent ($p < 2.2 * 10^{-16}$), Open Urban Percent ($p < 2.2 * 10^{-16}$), Forest Percent ($p < 2.2 * 10^{-16}$), Grass Shrub Percent ($p < 2.2 * 10^{-16}$), Wetland Percent ($p < 2.2 * 10^{-16}$), Row Crop Percent ($p < 2.2 * 10^{-16}$), and Lake IWS Ratio ($p < 2.2 * 10^{-16}$). The marginal $R^2 = 0.1665$ and the conditional $R^2 = 0.2513$. This tells us that 9% of the variability in prime season chlorophyll a can be explained by the variance of ecoregion. Table 11 shows the coefficients for each significant explanatory variable.

Table 11: Model Coefficients. Prime Season Chlorophyll a.

Predictor	Coefficient
IntenseUrban	-0.01008
OpenUrban	-0.02524
Forest	-0.02448
GrassShrub	-0.02526
Wetland	-0.01253
RowCrop	-0.00694
LakeIWS.Ratio	-0.41743

Since the outcome variable of the model is log transformed, the coefficients must be exponentiated to be interpreted. Table 12 presents the percentage increase or decrease that a 1% increase of the explanatory variables have on prime season chlorophyll a.

Table 12: Transformed Model Coefficients. Prime Season Chlorophyll a.

Predictor	Coefficient
IntenseUrban	-2.30%
OpenUrban	-5.60%
Forest	-5.50%
GrassShrub	-5.70%
Wetland	-2.80%
RowCrop	-1.60%
LakeIWS.Ratio	-61.80%

Figure 13 displays Prime Season log Chlorophyll a vs. Lake IWS Ratio together with the Mixed Effects Model that we determined.

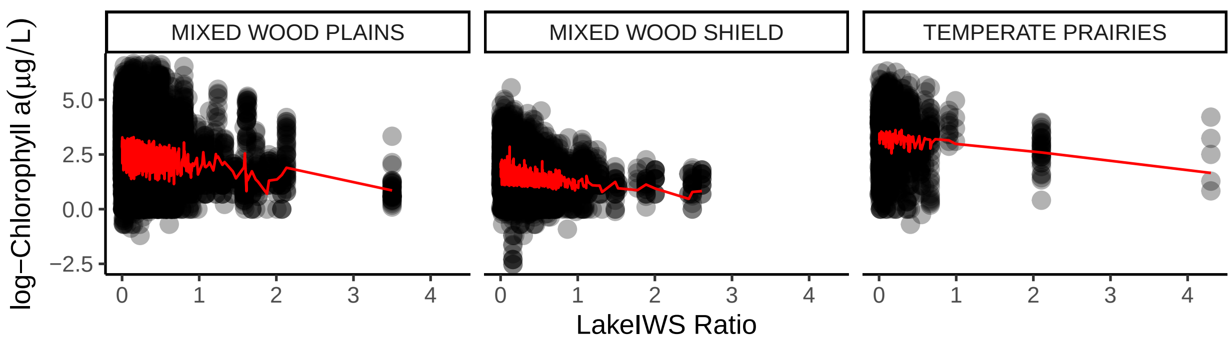

Figure 13: Prime Season log Chlorophyll a vs. Lake IWS Ratio and Mixed Effects Model

4.2.4 Prime Season Secchi Depth

The significant predictors of prime season secchi depth are Intense Urban Percent (p = 3.028 * 10^-7), Open Urban Percent (p < 2.2 * 10^-16), Barren Percent (p = 0.04061), Forest Percent (p < 2.2 10^-16), Grass Shrub Percent (p < 2.2 * 10^-16), Pasture Percent (p = 1.234 * 10^-8), Row Crop Percent (p < 2.2 * 10^-16), and Lake IWS Ratio (p < 2.2 *10^-16). The marginal R^2 = 0.1241 and the conditional R^2 = 0.4032. This tells us that 28% of the variability in prime season secchi depth can be explained by the variance of ecoregion. Table 13 shows the coefficients for each significant explanatory variable.

Table 13: Model Coefficients. Prime Season Secchi Depth.

Predictor	Coefficient
IntenseUrban	0.00423
OpenUrban	0.01559
Barren	0.00760
Forest	0.01597
GrassShrub	0.01446
Pasture	0.00333
RowCrop	0.00464
LakeIWS.Ratio	0.39432

Since the outcome variable of the model is log transformed, the coefficients must be exponentiated to be interpreted. Table 14 presents the percentage increase or decrease that a 1% increase of the explanatory variables have on prime season secchi depth.

Table 14: Transformed Model Coefficients. Prime Season Secchi Depth.

Predictor	Coefficient
IntenseUrban	1.00%
OpenUrban	3.70%
Barren	1.80%
Forest	3.80%
GrassShrub	3.40%
Pasture	0.80%
RowCrop	1.10%
LakeIWS.Ratio	147.90%

Figure 14 shows Prime Season log Secchi Depth vs. Lake IWS Ratio together with the Mixed Effects Model that we determined.

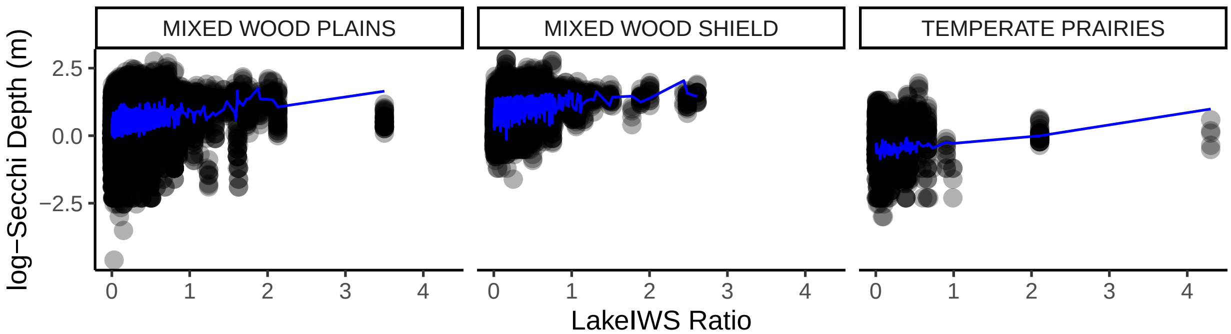

Figure 14: Prime Season log Secchi Depth vs. Lake IWS Ratio and Mixed Effects Model

4.2.5 Late Season Chlorophyll a

The significant predictors of late season chlorophyll a are Intense Urban Percent (p = 3.663* 10^-6), Open Urban Percent (p < 2.2 * 10^-16), Forest Percent (p < 2.2 * 10^-16), Grass Shrub Percent (p = 0.3.231 * 10^-5, coefficient = -0.02814891), Wetland Percent (p = 0.035393), and Lake IWS Ratio (p = 0.001653). The marginal R^2 = 0.1515 and the conditional R^2 = 0.2314. This tells us that 8% of the variability in late season chlorophyll a can be explained by the variance of ecoregion. Table 15 shows the coefficients for each significant explanatory variable.

Table 15: Model Coefficients. Late Season Chlorophyll a.

Predictor	Coefficient
IntenseUrban	-0.01331
OpenUrban	-0.02377
Forest	-0.02487
GrassShrub	-0.02815
Wetland	-0.01128
LakeIWS.Ratio	-0.26335

Since the outcome variable of the model is log transformed, the coefficients must be exponentiated to be interpreted. Table 16 displays the percentage increase or decrease that a 1% increase of the explanatory variables have on late season chlorophyll a.

Table 16: Transformed Model Coefficients. Late Season Chlorophyll a.

Predictor	Coefficient
IntenseUrban	-3.00%
OpenUrban	-5.30%
Forest	-5.60%
GrassShrub	-6.30%
Wetland	-2.60%
LakeIWS.Ratio	-45.50%

Figure 15 shows Late Season log Chlorophyll a vs. Lake IWS Ratio together with the Mixed Effects Model that we determined.

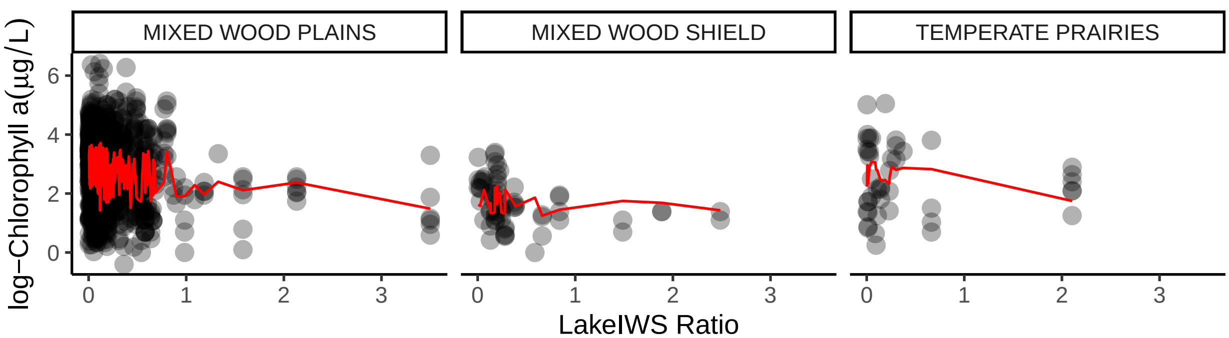

Figure 15: Late Season log Chlorophyll a vs. Lake IWS Ratio and Mixed Effects Model

4.2.6 Late Season Secchi Depth

The significant predictors of late season chlorophyll a are Intense Urban Percent (p = 0.03779), Open Urban Percent (p < 2.2 * 10^-16), Forest Percent (p < 2.2 * 10^-16), Grass Shrub Percent (p = 8.468 * 10^-5), and Lake IWS ratio (p = 3.170 * 10^-6). The marginal R^2 = 0.1595 and the conditional R^2 = 0.1889. This tells us that 3% of the variability in late season secchi depth can be explained by the variance of ecoregion. Table 17 shows the coefficients for each significant explanatory variable.

Table 17: Model Coefficients. Late Season Secchi Depth.

Predictor	Coefficient
IntenseUrban	0.00386
OpenUrban	0.01454
Forest	0.01619
GrassShrub	0.01762
LakeIWS.Ratio	0.24670

Since the outcome variable of the model is log transformed, the coefficients must be exponentiated to be interpreted. Table 18 presents the percentage increase or decrease that a 1% increase of the explanatory variables have on late season secchi.

Table 18: Transformed Model Coefficients. Late Season Secchi Depth.

Predictor	Coefficient
IntenseUrban	0.90%
OpenUrban	3.40%
Forest	3.80%
GrassShrub	4.10%
LakeIWS.Ratio	76.50%

Figure 16 shows Late Season log Secchi Depth vs. Lake IWS Ratio together with the Mixed Effects Model that we determined.

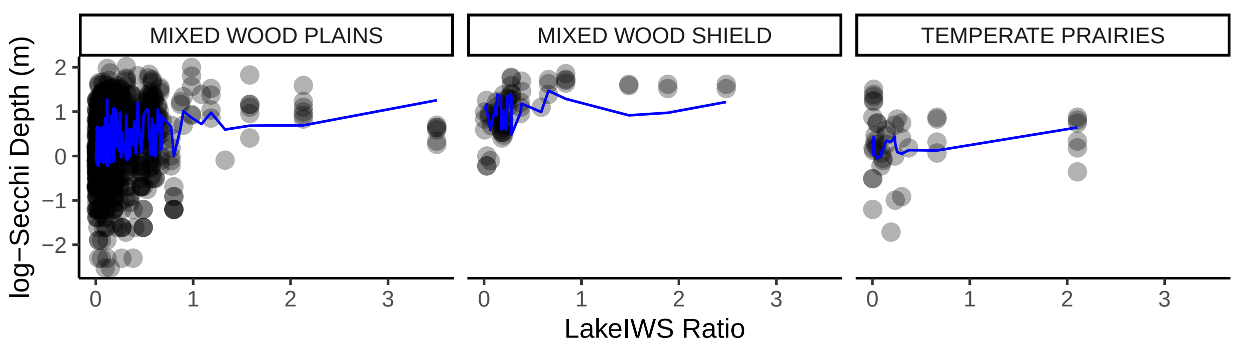

Figure 16: Late Season log Secchi Depth vs. Lake IWS Ratio and Mixed Effects Model

5 Summary and Conclusions

The aim of this project was to examine the impacts that land use has on water quality in Minnesota and if there was a seasonal characterization of these effects. Our findings support some of our hypotheses and do not support others. Our initial assumption as reflected in hypotheses 1a, 1b, and 1c was that urbanized and agricultural land would negatively impact water quality, and forest and natural lands have positive impact on water quality. We also predicted in hypothesis 2 that we would see seasonal variation in our results. Our results are less clear cut than our hypotheses predicted they would be.

Our second hypothesis is largely supported by our statistical tests. We found a difference between seasons in our dataset. Observe in Table 7 that no two seasons for the two variables we looked at found the exact same variance coefficient. In many cases, the difference in coefficients was quite large, while in other cases the difference was close to negligible and potentially the result of unaccounted for variation or statistical noise. Observe our findings for Forest land cover. Prime and late season have similar results for both chlorophyll a and secchi depth, but these two values are quite different than the value for early season. In many cases, we found different significant land uses for the same dependent variable for different seasons. For example, we found that Intense Urban only has statistically significant coefficients for the prime and late seasons. We also found that prime has highest number of significant variables.

Our first hypothesis is partially supported by our statistical tests. Observe from Table 7 that our findings do not support our hypotheses 1a, 1b, and 1c for all the land uses that we looked at. Our initial assumption was that agricultural and urbanized land would negatively impact water quality, and forest and natural lands would have a positive impact on water quality. We found that land use has significant impact on both response variables and that an increase in chlorophyll a was paired with a decrease in secchi depth in all but one case (row crop, early). These impacts were sometimes but not always in the direction our hypotheses predicted.

Row crop showed increasing chlorophyll a in the early season, decreasing in prime, and was not significant in late. If hypothesis 1b, which dealt with agricultural land use, was to be true, we would have predicted to see increasing chlorophyll a for all three seasons. It is worth noting that the China study by Huang et al. (2013) found a similarly complex result from agricultural land use. We found several other noteworthy similarities and differences between our hypotheses 1a, 1b, and 1c:

Similarities to hypotheses:

- Pasture and row crop degrades water quality in early season
- Forest improves water quality all seasons
- Wetland improves water quality prime and late (chlorophyll a only)
- Grass shrub improves water quality

Differences from hypotheses:

- Intense urban improves water quality in prime and late season
- Open urban improves in all seasons

- Row crop improves water quality in prime season

A major limitation of this study is that it does not consider all of the factors contributing to water quality. This can be seen in our relatively low R^2 values for all of our models. Another limitation is the uneven distribution of lakes across the three ecoregions. Ecoregions with less lakes will not be as thoroughly accounted for by this study, and therefore other water quality issues, such as those related to groundwater, are not captured by this dataset. These limitations should be considered when developing further studies and governing policies. Future studies could use more up to date land cover or alternatively examine the effects of changes in land cover over time.

Our findings offer potential insight for watershed managers in Minnesota, as well as in the Mississippi River basin. We were able to show in some cases that increasing the amount of developed land in a watershed decreases its water quality. In addition, our analysis found some seasonal variation in water quality. Taken together, these two findings suggest that land use management decisions in Minnesota potentially affect water quality far downstream of state lines. We recommend that managers consider increasing or slowing the decrease of undeveloped land in Minnesota. Managers also ought to be conscious of seasonal patterns in water quality impacts of land use. Potentially, states and Canadian provinces downstream of Minnesota could work together to improve land use practices at a watershed scale. The Mississippi River Basin Healthy Watershed Initiative was founded in 2009 and is a 13-state collaborative effort to pool resources from Farm Bill programs to improve water quality through nutrient management in the Mississippi River basin (Natural Resources Conservation Service, 2019). The Initiative is a strong step towards watershed-scale nutrient management, but it could benefit from an improved understanding of the types of effects different spatial and temporal land use management decisions have on downstream water quality.

6 References

Christensen, Victoria G., et al. “Relations between Retired Agricultural Land, Water Quality, and Aquatic-Community Health, Minnesota River Basin.” Journal of Environmental Quality, vol. 41, no. 5, 2012, pp. 1459-72. ProQuest, https://login.proxy.lib.duke.edu/login?url=https://search.proquest.com/docview/1 95114913?accountid=10598.

Huang, J., et al. “Evaluation of the Impacts of Land Use on Water Quality: A Case Study in The Chaohu Lake Basin,” The Scientific World Journal, vol. 2013, Article ID 329187, 7 pages, 2013. https://doi.org/10.1155/2013/329187.

Maalim, Fukhrudin K., et al. “Modeling the impact of land use changes on runoff and sediment yield in the Le Sueur watershed, Minnesota using GeoWEPP.” CATENA Volume 107, 2013, Pages 35-45, https://doi.org/10.1016/j.catena.2013.03.004.

“Mississippi River Basin Healthy Watersheds Initiative-Natural Resources Conservation Service.” Natural Resources Conservation Service, www.nrcs.usda.gov/wps/portal/nrcs/detailfull /national/programs/initiatives/?cid=stelprdb1048200.

“Nutrient Reduction Strategy.” Minnesota Pollution Control Agency, 6 Mar. 2019, www.pca.state.mn.us/water/nutrient-reduction-strategy#resources-b7ca91dc.