**Analysis**:

Before creating our models, we tested for normality of the distribution of our variables using qqplots and Shaprio-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.

**Results:**

Early Season Chla:

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).

|  |  |
| --- | --- |
| Predictor | Coefficient |
| Open Urban | -0.0107689 |
| Forest | -0.0152622 |
| Pasture | 0.0084965 |
| Row Crop | 0.00320106 |
| Ratio | -0.3086167 |

Early Season Secchi:

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 (check this phrase).

|  |  |
| --- | --- |
| Predictor | Coefficient |
| Open Urban | 0.01236809 |
| Barren | 0.01956408 |
| Forest | 0.01228808 |
| Grass Shrub | 0.01024311 |
| Ratio | 0.2144161 |

Prime Season Chla:

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.

|  |  |
| --- | --- |
| Predictor | Coefficient |
| Intense Urban | -0.0100841 |
| Open Urban | -0.0252362 |
| Forest | -0.024479 |
| Grass Shrub | -0.0252642 |
| Wetland | -0.0125302 |
| Row Crop | -0.0069355 |
| Ratio | -0.4174341 |

Prime Season Secchi:

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

|  |  |
| --- | --- |
| Predictor | Coefficient |
| Intense Urban | 0.00423021 |
| Open Urban | 0.01558668 |
| Barren | 0.00759661 |
| Forest | 0.01596932 |
| Grass Shrub | 0.01445791 |
| Pasture | 0.00332981 |
| Row Crop | 0.00463991 |
| Ratio | 0.3943152 |

Late Season Chla:

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.

|  |  |
| --- | --- |
| Predictor | Coefficient |
| Intense Urban | -0.0133056 |
| Open urban | -0.0237716 |
| Forest | -0.02487 |
| Grass Shrub | -0.0281489 |
| Wetland | -0.0112843 |
| Ratio | -0.2633484 |

Late Season Secchi:

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, coefficient = ), 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.

|  |  |
| --- | --- |
| Predictor | Coefficient |
| Intense Urban | 0.00386273 |
| Open Urban | 0.01453631 |
| Forest | 0.01619299 |
| Grass Shrub | 0.01761664 |
| Ratio | 0.2467001 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Chlorophyll a** | |  | **Secchi Depth** | |  |
| **Predictor** | **Early** | **Prime** | **Late** | **Early** | **Prime** | **Late** |
| Intense Urban |  | -0.0101 | -0.0133 |  | 0.0042 | 0.0039 |
| Open Urban | -0.0108 | -0.0252 | -0.0238 | 0.0124 | 0.0156 | 0.0145 |
| Barren |  |  |  | 0.0196 | 0.0076 |  |
| Forest | -0.0153 | -0.0245 | -0.0249 | 0.0123 | 0.016 | 0.0162 |
| Grass/Shrub |  | -0.0253 | -0.0281 | 0.0102 | 0.0145 | 0.0176 |
| Wetland |  | -0.0125 | -0.0113 |  |  |  |
| Pasture | 0.0085 |  |  |  | 0.0033 |  |
| Row Crop | 0.0032 | -0.0069 |  |  | 0.0046 |  |
| Lake IWS Ratio | -0.3086 | -0.4174 | -0.2633 | 0.2144 | 0.3943 | 0.2467 |

Conclusions:

* Difference between seasons
* Land use has significant impact on both response variables
* Prime has highest number of significant variables
* Row crop increasing chla in early, decreasing in prime, and not significant in late
* Initial assumption that ag and urbanized land would negatively impact water quality, and forest and natural lands have positive impact on water quality
* Similarities to hypotheses:
  + Pasture and row crop degrades wq in early season
  + Forest improves wq all seasons
  + Wetland improves wq prime and late (chla only)
  + Grass shrub improving wq
* Differences from hypotheses:
  + Intense urban improves water quality in prime and late season
  + Open urban improves in all seasons
  + Row crop improve wq in prime season
* Policy recommendations
  + Urban is not the issue in MN, efforts should be focused on other land uses
  + Understand that non point sources are hard to regulate
  + Land use should be considered but should not be the only factor considered
* Limitation of study:
  + Not considering all of the variables contributing to water quality
    - Low R-sqared
    - Need to consider this when creating policy
  + Uneven sampling between ecoregions
    - Ecoregions with less lakes are not as thoroughly accounted for by this study
    - Other water quality issues related to groundwater not captured by this dataset
* Future studies
  + More up to date land cover
  + Changes in land cover
  + More official and academically defensible “seasons”
  + Groundwater quality should be included in studies

The aim of this project was to examine the impacts that land use had on water quality in Minnesota and if there was a seasonal characterization of these effects. Our findings support some of our hypotheses and does not support others. An increase in percent forest land use