Mixed Effects Analysis of TP Levels in Lakes Located in Six Minnesota Counties from 2015 to 2020

Carly Mahoney

April 28, 2022

Abstract

Lakes in Minnesota are a valuable resource as well places for recreation in Minnesota so it is important that we keep them healthly. One factor that can have a large impact on lake health is phorphorous. High phorphorous levels can result in algae growth and reduce nessecary oxygen levels needed by fish. Past research has suggested that climate change and specifically increase runoff due to increased rainfall is increasing phorphorous concentrations in Minnesota lakes. Using mixed effects modeling, this project looks at whether phosphorous concentrations through a summer display a pattern and whether years with greater rainfall see increased phosphorous levels.

Advisers and Key Words

- Primary Adviser: Jon Anderson
- Secondary Adviser: Jong-Min Kim
- Key word 1: Phosphorous
- Key word 2: Lakes
- Key word 3: Minnesota

Introduction

Phosphorous is the driving nutrient for aquatic plants in most Minnesota lakes meaning all other nutrients a plant needs are typically in excess (RMB, 2022). This means total phosphorous (TP) levels have a direct effect on plant growth and therefore the health of a lake. Without phosphorous, plants will die due to a lack of nutrients but when concentrations are high, there can be an algae bloom. Algae blooms lead to algae decay which reduces the levels of oxygen that are available to fish (MPC, 2009). They also make a lake unpleasant and possibly unhealthy to swim in (MPC, 2009).

The major sources through which phosphorous enters lakes are human and animal wastes, soil erosion, detergents, septic systems, and runoff from farmland or fertilized lawns (RMB, 2022). Most of these factors can only occur during months when lakes are not frozen suggesting phosphorous levels would increase throughout the summer. Phosphorous entering lakes through runoff suggests higher rainfall would result in higher phosphorous concentrations in surface water. It is also important to note that in different areas of the state, in this case, we are looking at counties, soil content and amount of farmland differ which means phosphorous levels could potentially differ between counties.

Averages for rainfall in Minnesota can be examined by division. Division four, most notably, includes Otter Tail County which contains an abundance of lakes making it a high-traffic vacation spot. The other counties it includes are Big Stone, Douglas, Grant, Pope, Stevens, Swift, Traverse, and Yellow Medicine. To keep our lakes healthy and enjoyable, it is important to understand the pattern of TP levels and the factors that have an impact on them. This project will look at total phosphorous levels in Minnesota's rainfall division four lakes using month, year, and county from the years 2015 to 2020, as descriptors.

Background

Total phosphorous concentrations in Minnesota lakes are constantly being studied and reviewed by the Minnesota Pollution Control Agency to monitor the health of our lakes. The monitoring methods include observations of the look of a lake and sample collection and laboratory testing. In 2014, MPC (2014) reported 75% of lakes statewide met quality control standards. The 25% of lakes that did not meet the standards had high phosphorous and algae content (MPC, 2014). The MPCA compares lake health by year but it does not seem to study how TP levels change throughout the summer which is part of what this project will cover.

Since TP concentrations are partially the result of runoff, it is important to note the rainfall values that occurred in the area of study. The average rainfalls for all the measurement stations in the region studied in this project from the first of May to the first of October were 18.6 inches in 2015, 19.5 inches in 2016, 21.2 inches in 2017, 19.4 inches in 2018, 22.3 inches in 2019, and 14.9 inches in 2020 (noa, 2022). These were found simply by collecting the data and averaging it. Note that rainfall from 2015 to 2020 was mostly consistent, with the exception of 2020 being low.

Jacobson *et al.* (2017) used taxa-specific ecological niche models that were developed using generalized additive models (GAMs) to determine the effects of several environmental and ecological variables on the abundance of multiple fish species in Minnesota lakes. Phosphorus was selected to capture the effects of eutrophication from 1993 through 2005 (Jacobson *et al.*, 2017). Phosphorous is significant to fish populations because some fish are more suited to nutrient-rich water than others. Jacobson *et al.* (2017) concluded that eutrophication during this time period did shift population counts for fish species, however, in more recent years populations have not changed significantly. This brings up the question is eutrophication still happening in Minnesota or are fish populations just more stable now that nutrient levels are already high?

Past research in Denmark studied the effects of rainfall and runoff, as well as other factors on TP levels in areas with greater than 50% agricultural land. Jeppesen *et al.* (2009) suggested that increased rainfall due to climate change would have profound effects on lake eutrophication which is another word for an excessive amount of nutrients such as phosphorous in a body of water. Jeppesen *et al.* (2009) used the NAM rainfall-runoff model to model water discharge. The model showed an increase in the annual mean runoff between the control and the scenario period (Jeppesen *et al.*, 2009). Jeppesen *et al.* (2009) concluded that even with dilution reducing phosphorous levels, there would be a net increase. This study used longer time periods than I will be looking at in my study. So, while this study showed increased runoff leads to a net increase in lake phosphorous concentrations in the long term, it has not been shown whether this effect can be observed in the short term. This study was also not specific to Minnesota so the results may not be completely comparable unlike the result given by Jacobson *et al.* (2017).

Jeppesen *et al.* (2009) also studied how region affected TP levels. They found that regions with differing levels of precipitation, soil composition, and agricultural practices showed significant differences in the TP composition of water bodies. The counties in this study are all adjacent to one another, however, differences between counties might still be observed. For example, one county might be heavier in summer tourism while another is heavy in agriculture causing the lakes in the two areas to have different health statuses.

Data Collection, Organization, and Observations

The data for this analysis were compiled by the Minnesota Pollution Control Agency from the years 1986 to 2020 from May through October. Each row of the data frame contains data for a single water sample. After cleaning, included in each row is the name of the lake the sample was taken from, the county the lake is in, the full date, name of the month, number of the month, day of the year, and year the sample was taken, the total phosphorous measurement in mg/L, and the sample site (stationId variable). These descriptions correspond to the variables stationName, county, sampleDate, month, monthN, year_d, year, result, and stationId respectively.

Each lake included in the analysis had multiple samples taken from it during the time period. Most of the lake samples with total phosphorous results were analyzed by RMB Environmental Laboratories. Some lakes had multiple sample sites each having a unique station ID. Samples were taken at a depth of no more than two meters and each repeated measurement with a unique station ID was taken in the same area of the lake. A few rows of data had the same station ID and sample dates. This indicates a quantity control measure was taken by duplicating a sample. The duplicate rows were averaged for this analysis.

The data set was cut down to include only data from the years 2015 to 2020. Summary statistics by county can be seen in Table 1, summary statistics by year can be seen in Table 2, and summary statistics by month can be seen in Table 3. An important piece to note about the summary statistics is the low measurement counts for October and the low measurement counts for some of the counties. When a county has a low count such as 15, it means 15 total samples were taken from all the lakes in the county, and therefore, a county with a count that low is likely only measuring TP in two or three lakes. The low lake counts are more apparent in Figure 1. Mean functions for Grant, Big Stone, and Yellow Medicine county are more varied by year and are missing entire years because the few lakes documented did not test every year.

Table 1: Count, average, and standard deviation for TP results by county

county	avePhos	n	sdPhos
Big Stone	0.1597	81	0.1135
Douglas	0.0233	867	0.0201
Grant	0.0467	15	0.0150
Otter Tail	0.0204	2095	0.0177
Pope	0.0722	544	0.1102
Yellow Medicine	0.0322	10	0.0115

Table 2: Count, average, and standard deviation for TP results by year

year	avePhos	n	sdPhos
2015	0.0273	543	0.0393
2016	0.0296	571	0.0456
2017	0.0313	635	0.0495
2018	0.0337	641	0.0516
2019	0.0385	606	0.0761
2020	0.0321	616	0.0618

Figures 1, 2, and 3 give insight into the mean functions by county and by year. These figures tell similar stories to the ones told in Table 1 and Table 2. From the raw data, we see the general pattern arise that phosphorous levels increase throughout the summer however there appears to be an interaction between the county and the day of the summer. We can also observe the difference between counties appears to be greater than the difference between years. Only two of the county mean functions cross while the year mean functions are closer together and cross each other often. Despite being closer together, the year mean functions follow a pattern. The intercepts increase as the year increases. The only discrepancy is 2020, however, the rainfall in 2020 was much less than any of the other years, which is a possible reason why 2020 does not follow the pattern.

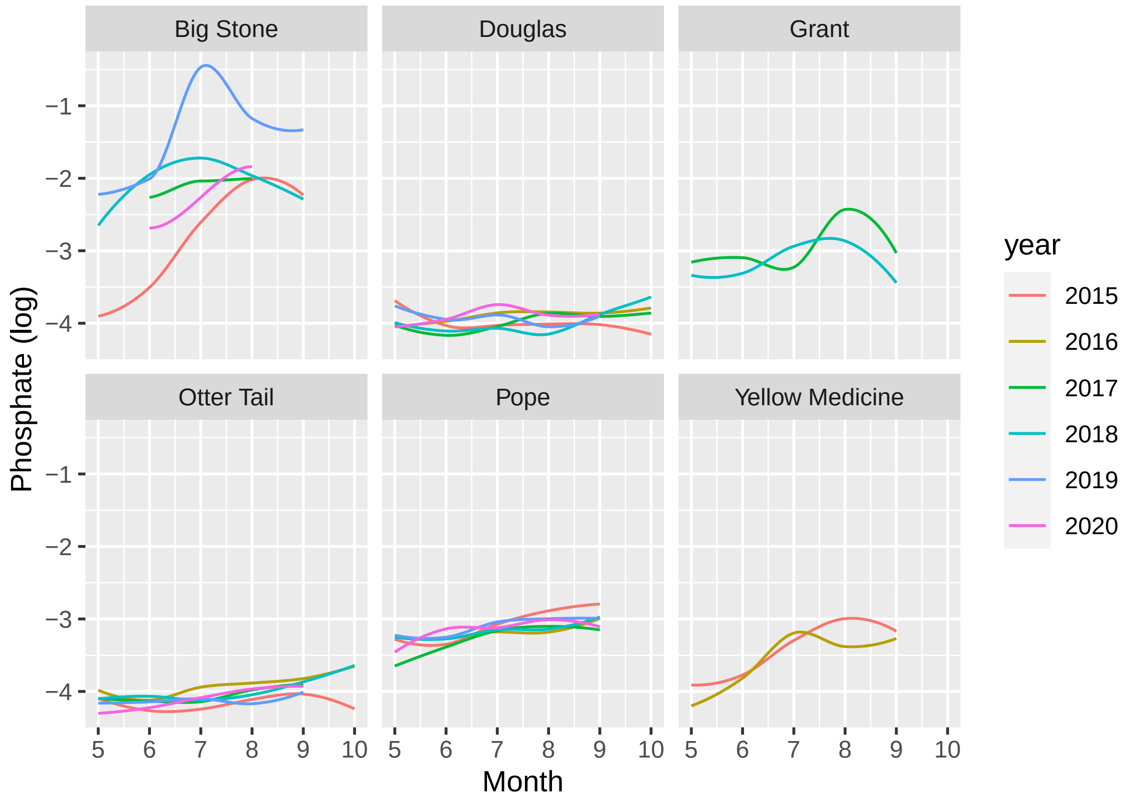


Figure 1: Mean Phosphate Measurements by Year and County

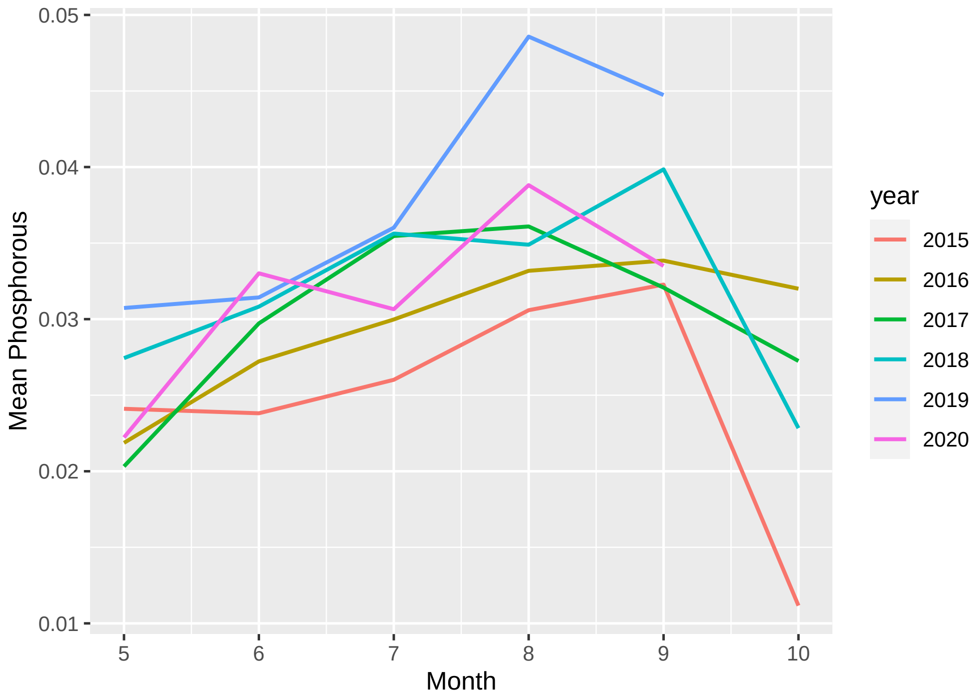


Figure 2: Mean Phosphate by Year

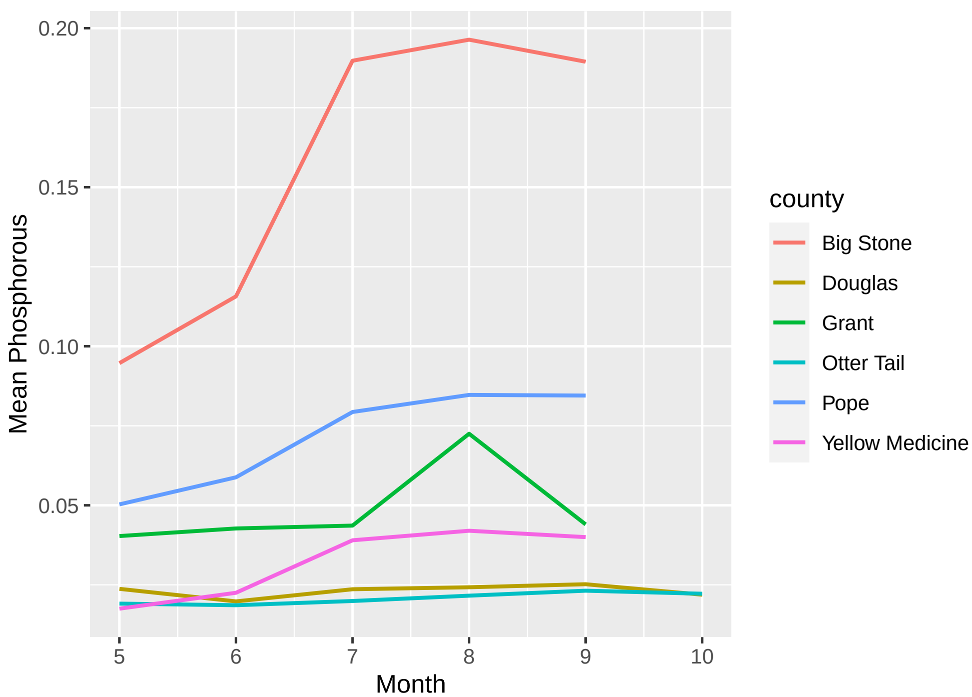


Figure 3: Mean Phosphate by County

Statistical Methods

Overview

Multiple measurements are taken from the same lake over the time period making this data appropriate for a repeated measures type analysis. Repeated measures of a single area result in correlation between the measurements, therefore violating the independence assumption of simple ANOVA (Scheiner and Gurevitch (2006)). Since simple ANOVA cannot account for the correlation within a subject, other methods must be used. The specific method used in this study is linear mixed-effects modeling. Mixed-effects modeling allows for the examining of TP results while considering variability within and across lakes simultaneously (Brown (2021)). Brown (2021) explains that mixed models allow for both fixed and random effects hence the name mixed models.

According to Brown (2021), fixed effects are population-level effects that should be consistent across experiments. For example, in this data month is a fixed effect because we expect an average relationship between it and TP concentration even if we sample a different set of lakes. Random effects, which will be represented by q in the model definition, are those that occur between dependent data points that belong to a single item or participant. In this case, a random effect would be the lake/station ID because each lake has multiple TP measurements that are dependent on one another. In the model definition, the number of lakes (144) will be represented by J and N will be the total number of TP measurements (3612).

Mixed-effects models work by using random intercepts and random slopes. Traditional linear regression creates one fixed intercept and one fixed slope which creates a single linear model for all the subjects. Mixed-effects models differ in that they allow subjects to have unique intercepts and unique slopes that vary around a fixed intercept and fixed slope. (Brown (2021)). In the TP data, intercepts may vary by lake because some lakes naturally have more or less phosphorous than others. Slopes may vary by lake because lakes may react differently to the conditions that change by month.

The linear mixed model is represented by the equation:

$$Y = X\beta + Zu + \epsilon$$

where Y is the response values for each subject represented by a $N \times 1$ vector, β is the $p \times 1$ (p is the number of predictor variables) column vector containing the fixed-effect coefficients, X is the $N \times p$ matrix (p being the predictor variables) that determines how β is assigned to individual observations within subjects, u is the $qJ \times 1$ random effects vector, Z is the $N \times qJ$ matrix which determines how u is assigned to observations, and ϵ is the error term for a particular observation and subject noa (2021). The $X\beta$, Zu, and ϵ terms are all $N \times 1$ vectors which means adding them together gives an $N \times 1$ vector, in this case y. It is important to note that u is not a direct estimation. It is assumed to have distribution $N(0, D)$ with $Var(u) = D$. If we then condition the response on this assumption, we get:

$$\begin{aligned} E(Y|u) &= X\beta + Zu \\ Var(Y|u) &= \Sigma \\ Var(Y) &= ZDZ' + \Sigma. \end{aligned}$$

The assumptions of mixed-effects modeling deal with linearity, normality of the residuals, homogeneity of residual variance, and autocorrelation Waterston (2019). All of these are considered in the model building process. Autocorrelation is the only assumption that differs from a normal linear regression model. It is the correlation that occurs between measurements in the same group. We adjust for this with the Σ matrix. When observations are independent of one another we use the matrix:

$$\Sigma_i = \sigma^2 \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

However, as we have discussed, the consequence of repeated measurements on a single group is correlations between observations making this matrix inappropriate. In the model building process to follow, multiple autocorrelation structures are tested and the best matrix for the data is described in further detail.

Model Building

The fixed effects used in the model are year_d, year, and county. Some stations sample at the beginning of a month and then near the end. To have the model account for this type of time change, I used the year_d variable. This makes a single unit of time a day instead of having a single unit of time be a month. The random variable considered is each lake's station identification because multiple measurements were taken at each station. The response variable used was result. The code for the models considered can be found in Appendix III.

The first step taken to build the model was determining the fixed effects structure. All three of the fixed effect variables were used to build the model. The test used to compare the model without the year variable (mod2) to the model with the year variable (mod3) showed no evidence year helped explain the TP result after adjusting for the other variables in the model. I was interested in looking at the year variable and the p-value for the ANOVA test was just over 0.05 so I chose to keep the year variable. The Akaike Information Criteria (AIC) was used to assess the rest of the models with results shown in Table 3. The best fixed effects model was mod4 which included the year_d effect as a second degree polynomial as well as an interaction between year_d and county. I also created a model using splines (mod6) and found that it performed almost identical to the polynomial model (mod4). I chose the polynomial model due to the slightly lower AIC value.

Table 3: AIC values for the fixed effects determing models

	df	AIC
initialmod	14	-14452.11
mod2	19	-14554.60
mod3	14	-14554.37
mod4	25	-14611.30
mod5	15	-14450.58
mod6	25	-14610.94

Using mod4 as the base, the second step was to determine the random effect structure. AIC was used again to determine the best model (mod4b) which allowed a random adjustment to be made for each station ID each day of the year. The code that represents this structure is random = ~ year_d | stationId.

Table 4: AIC values for the random effect determing models

	df	AIC
mod4a	25	-14258.36
mod4b	27	-14682.05
mod4c	27	-14645.23

Finally, mod4b was used as a base to determine the error correlation structure. This is the step that adjusts for autocorrelation. After examining AIC values (Table 5) for multiple covariance matrix structures, the best one is the first-order autoregressive or AR(1):

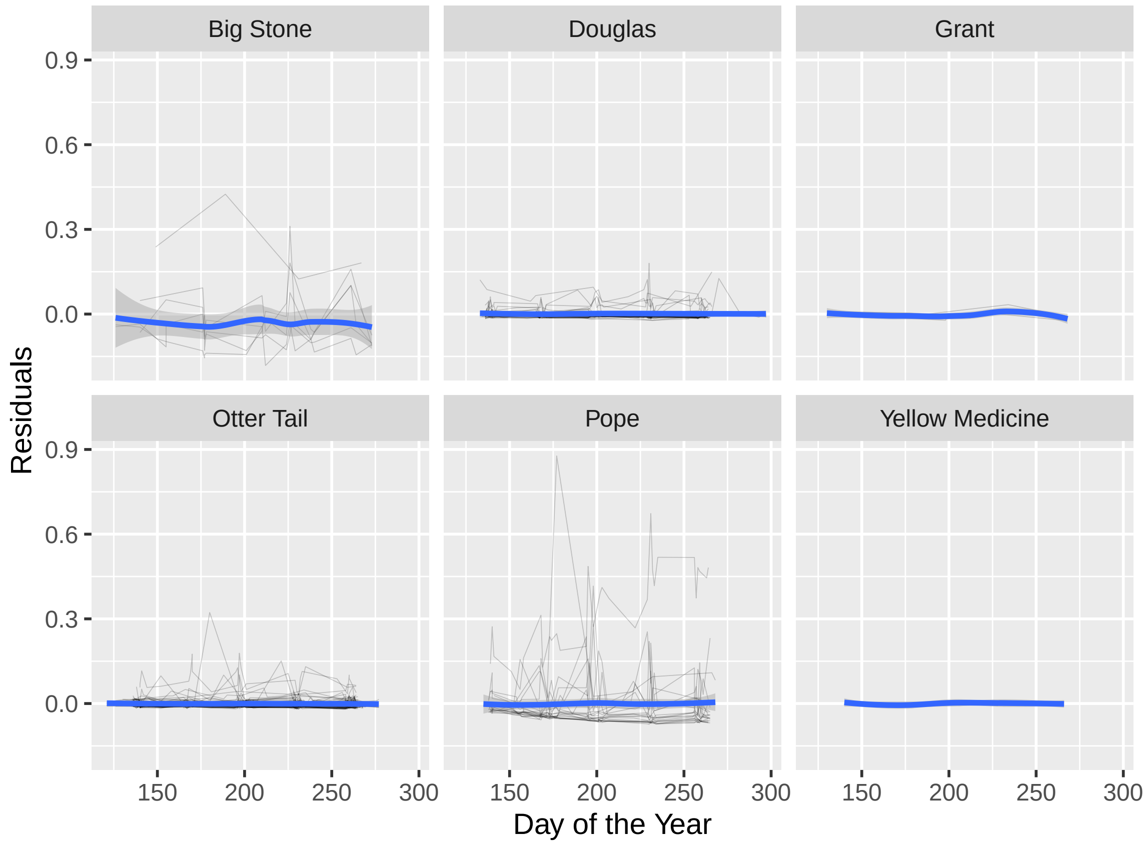
$$\Sigma_i = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 \\ \rho & 1 & \rho \\ \rho^2 & \rho & 1 \end{pmatrix}.$$

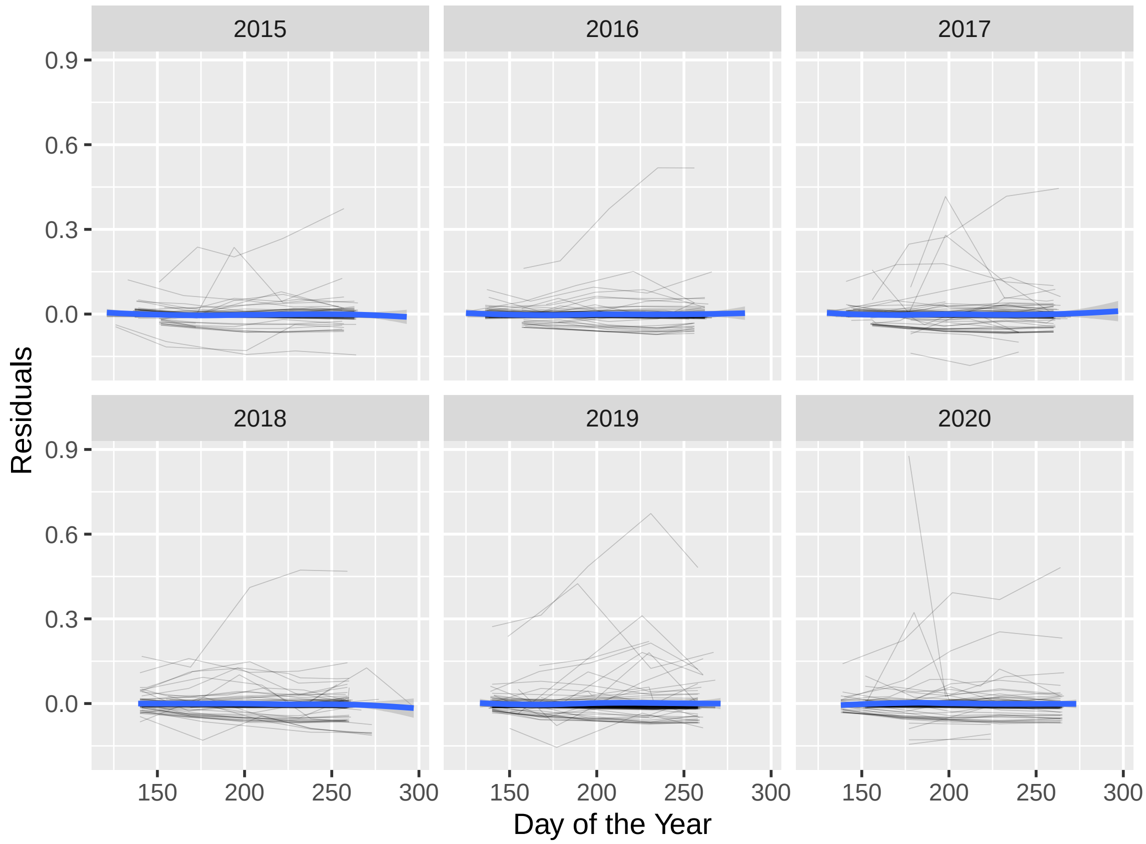
The main diagonal values are all equal to 1 which means the variance is equal at all time points with common elements. However, AR(1) does allow for correlation between time points to differ. In the context of the phosphorous data, this means the correlation that occurs between a lake's June and July samples does not have to equal the correlation that occurs between the lake's July and August samples. This autocorrelation matrix increases by a power unit across the columns and down the rows. This means as two sample dates get further away from each other, the less correlation there is between the two samples.

Table 5: AIC values for the possible autocorrelation adjustments

	df	AIC
mod4b.car1	28	-14680.05
mod4b.ar0ma1	28	-14685.00
mod4b.cs	28	-14680.05
mod4b.exp	28	-14680.05
mod4b.gaus	28	-14680.05
mod4b.Lin	28	-14680.05

Here we examine residuals to see if the model has any notable problems.


Big Stone
Douglas
Grant
Otter Tail
Pope
Yellow Medicine
Residuals
0.9
0.6
0.3
0.0
150
200
250
300
Day of the Year

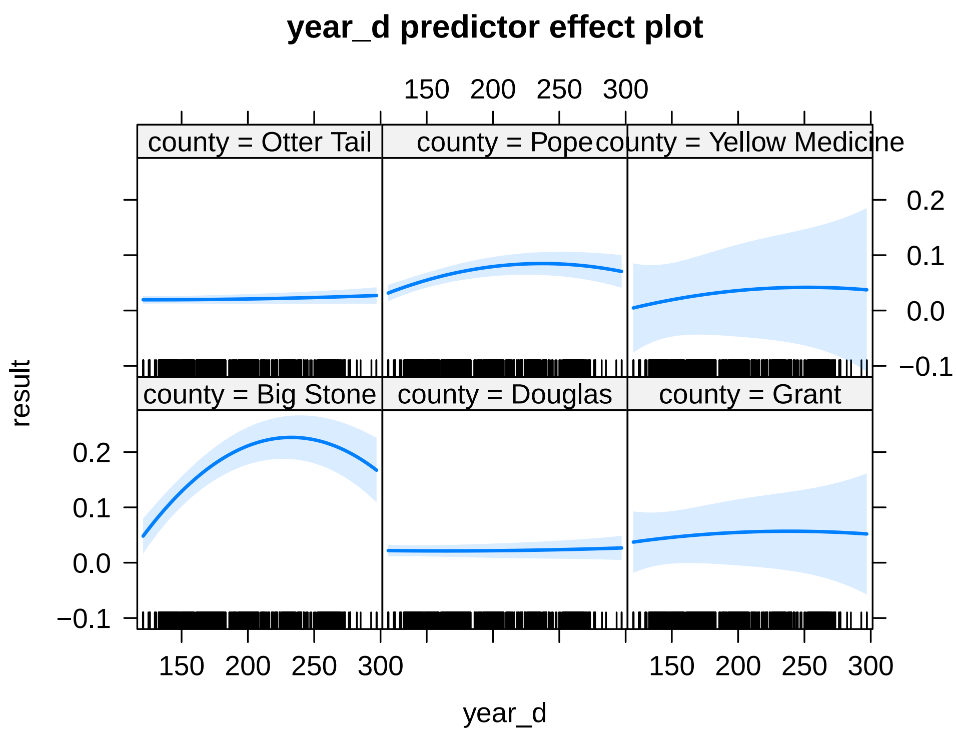


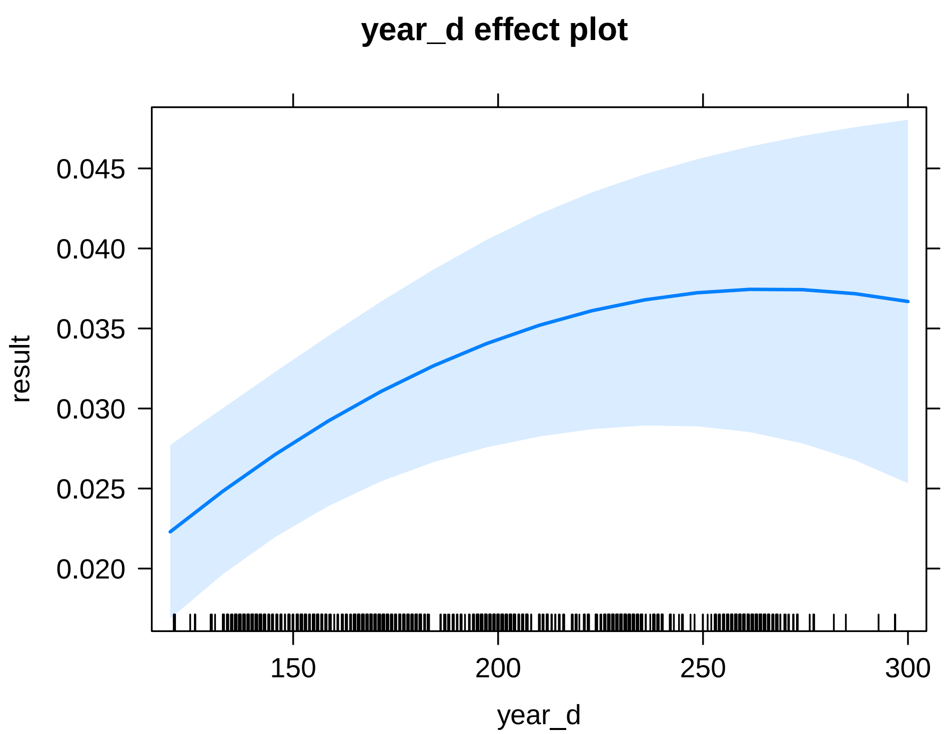
There are some apparent issues with the developed model. Outliers are present almost every year and non-constant variance is apparent in Pope County. The model was not changed in any way due to these issues, but it is important to acknowledge they do exist.

Results

There were 3612 observations spread across 144 lakes that were used to build the model. Three explanatory variables were used with some significant findings found for each of them.

The day of the year effect showed that overall, phosphorous levels increased throughout the summers as shown in the year_d effect plot. However, an interaction was present between the county and the day of the year the sample was taken. The year_d predictor effect plot shows the average change that occurs in phosphorous levels over the summer by county. The phosphorous levels in each county follow the overall increasing pattern but the increase is more apparent in Pope and Big Stone County.





There is evidence TP concentrations in some counties differ when compared to other counties. Two notable counties are Pope County and Big Stone County. Every county's mean when compared to Big Stone County's is significantly different after adjusting for the day of the year the sample was taken and the year. The only mean TP concentration that shows no evidence of being different from Pope County's is Grant County. There is no evidence means from other counties differ from one another. All mean comparisons are shown in Table 6. Using Pope Country and Ottertail County as examples, the Table 7 estimates can be interpreted as such: It is estimated the mean phosphorus concentration in Pope County lakes is 0.0589 mg/L higher than it is in Ottertail County lakes after adjusting for the day of the year the sample was gathered and the year.

Table 6: Comparing mean differences for counties

	contrast	estimate	SE	df	t.ratio	p.value	.group
13	Otter Tail - Pope	-0.0589	0.0100	149	-5.8843	0.0000	1234
8	Douglas - Pope	-0.0581	0.0111	149	-5.2316	0.0000	1234
6	Douglas - Grant	-0.0332	0.0314	149	-1.0586	0.8968	12345678
11	Grant - Pope	-0.0249	0.0319	149	-0.7790	0.9707	12 56 90
14	Otter Tail - Yellow Medicine	-0.0155	0.0430	149	-0.3591	0.9992	1 3 5 7 9 A
9	Douglas - Yellow Medicine	-0.0147	0.0433	149	-0.3385	0.9994	1 3 5 7 9 A
7	Douglas - Otter Tail	0.0008	0.0081	149	0.0991	1.0000	5678
12	Grant - Yellow Medicine	0.0186	0.0527	149	0.3525	0.9993	1234567890ABC
10	Grant - Otter Tail	0.0340	0.0310	149	1.0970	0.8819	34 78 AB D
15	Pope - Yellow Medicine	0.0434	0.0437	149	0.9938	0.9193	2 4 6 8 0 BC
4	Big Stone - Pope	0.1327	0.0195	149	6.8183	0.0000	AB D
2	Big Stone - Grant	0.1576	0.0352	149	4.4748	0.0002	90ABCDE
5	Big Stone - Yellow Medicine	0.1761	0.0462	149	3.8153	0.0027	DE
1	Big Stone - Douglas	0.1908	0.0185	149	10.2961	0.0000	C E
3	Big Stone - Otter Tail	0.1916	0.0179	149	10.7065	0.0000	C E

Table 7 shows comparisons for mean phosphorous levels by year. The only two-year pairs that have significant differences in TP levels are 2015 and 2019 and 2017 and 2019. The year effect plot shows no consistent pattern in changing TP concentrations by year. It is estimated the mean phosphorus concentration in 2019 was 0.005 mg/L higher than it was in 2015 and 2017 after adjusting for the day of year the sample was gathered and the county the lake was in.

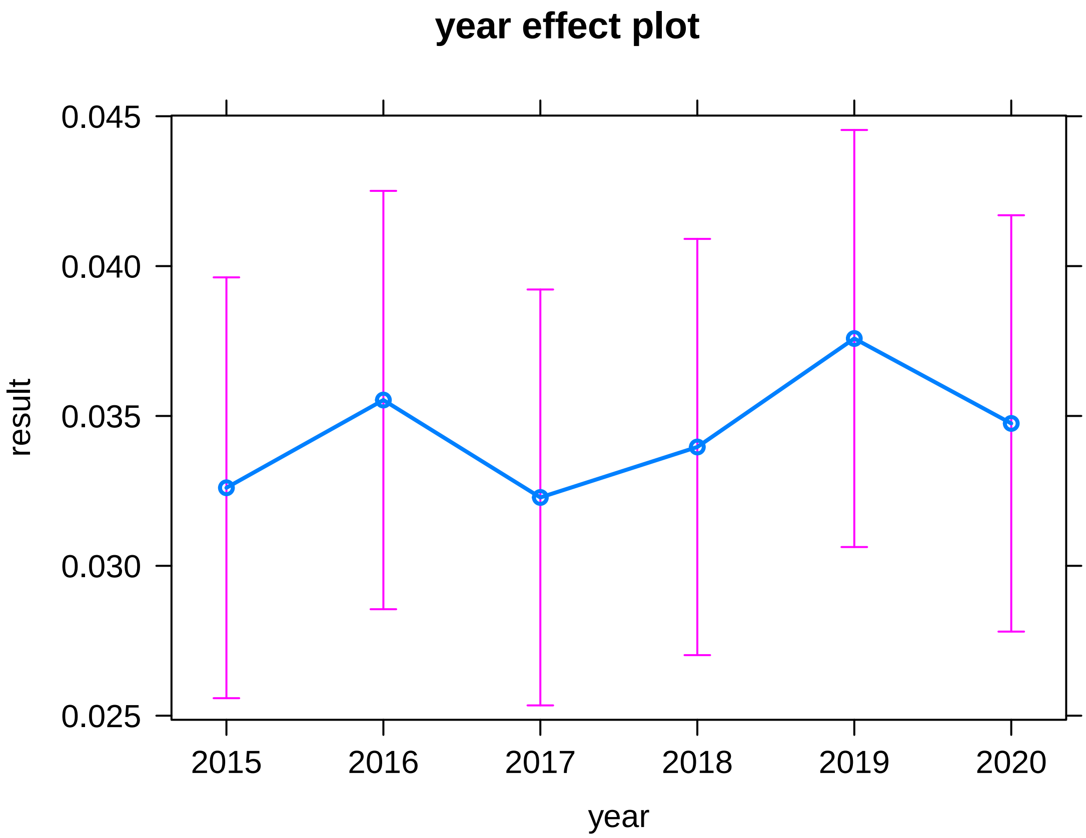


Table 7: Comparing mean differences for years

	contrast	estimate	SE	df	t.ratio	p.value	.group
11	2017 - 2019	-0.0053	0.0016	3440	-3.3433	0.0108	12
4	2015 - 2019	-0.0050	0.0017	3440	-3.0086	0.0316	1
13	2018 - 2019	-0.0036	0.0016	3440	-2.2546	0.2133	12
1	2015 - 2016	-0.0029	0.0017	3440	-1.7501	0.4985	12
12	2017 - 2020	-0.0025	0.0016	3440	-1.5724	0.6169	12
5	2015 - 2020	-0.0021	0.0017	3440	-1.2899	0.7907	12
8	2016 - 2019	-0.0021	0.0016	3440	-1.2606	0.8064	12
10	2017 - 2018	-0.0017	0.0016	3440	-1.0731	0.8922	12
3	2015 - 2018	-0.0014	0.0016	3440	-0.8352	0.9610	12
14	2018 - 2020	-0.0008	0.0016	3440	-0.5019	0.9961	12
2	2015 - 2017	0.0003	0.0016	3440	0.1988	1.0000	12
9	2016 - 2020	0.0008	0.0016	3440	0.4820	0.9968	12
7	2016 - 2018	0.0016	0.0016	3440	0.9757	0.9257	12
15	2019 - 2020	0.0028	0.0016	3440	1.7690	0.4861	12
6	2016 - 2017	0.0033	0.0016	3440	2.0009	0.3419	2

Discussion

Overall, this study gave a general overview of the total phosphorous situation for lakes in six Minnesota counties. When we consider years, there is not a consistent pattern to how phosphorous concentrations in lakes are changing. We suggested that years with more rain would have higher phosphorous levels but we do not have solid evidence from our results to be able to confirm this. The two year pairs that were significantly different from one another (2019 and 2015 and 2019 and 2017) follow the suggested rain and phosphorous pattern but 2020 had the least amount of rain out of all the years, and although the not significant, it had higher phosphorous levels than three of the years. This studied only considered six years so although we do not have sufficient evidence to say lake phosphorous concentrations are increasing in the short term, it is possible in ten years we will see every summer having significantly higher concentrations than say 2015.

In just the six counties we looked at, we saw significant differences between some of them. It would be beneficial to get data on more lakes in Yellow Stone, Grant, and Big Stone because the results for these counties may be misleading. For example, Pope TP concentrations were significantly different than all the other counties except Grant, however, only two lakes were observed in Grant. It is possible, if the samplers for these two lakes suspected the lakes had higher phosphorous concentrations then they would have been more likely to test these lakes over other lakes. In reality, Grant's lake phosphorous concentrations could on average be closer to Ottertail County's but the two lakes sampled are not sufficient enough to catch it.

There is evidence that Pope has higher lake TP concentrations than Ottertail and Douglas based on graphical evidence. This may be due to differences in lake and land usage in the counties. A future study could be done using a larger variety of Minnesota counties with varied land usage or one that considers the amount of tourism lakes acquire during the summer.

We have weak, graphical evidence to believe that phosphorous levels increase through the summer. All counties showed an increase but from the year_d effect plot, we can see the effects are not significant within error for most of the counties. Big Stone County shows a significant increase at least at the beginning of the summer and Pope County shows an increase that is either significant or close to being significant.

The increase from May to June was slightly unexpected. I hypothesized that May phosphorous levels would be elevated due to the runoff from melting snow in the spring. Since this was not the case, we could speculate either runoff is not the main cause of increasing TP levels, or the time between the first sample and the end of melting snow is a sufficient amount of time for the phosphorous cycle to reduce the TP concentrations. Future work could include a more controlled study that measures phosphorous right after it rains and then continues to monitor it until the next time it rains.

Appendix I: Data Preparation

```
path = "/home/mahon373/Git/CarlyMSenSem/Lake Data"

multmerge = function(mypath){
  filenames=list.files(path=mypath, full.names=TRUE)
  datalist = lapply(filenames, function(x){read.csv(file=x,header=T)})
  Reduce(function(x,y) {merge(x,y,all = TRUE)}, datalist)
}

# Used the function I created to compile the csv files from each county

full_data = multmerge(path)
```

```
PLakes=full_data[full_data$parameter=="Phosphorus as P",]
PLakes = PLakes[,-c(8:25)]
PLakes$sampleDate = as.Date(PLakes$sampleDate)
PLakes$result=as.numeric(PLakes$result)
LakesF = PLakes[PLakes$sampleDate > "2015-01-01",]

#added a year variable and a month variable

LakesF = LakesF %>% mutate(year = year(sampleDate))
LakesF = LakesF %>% mutate(year = as.character(year))
LakesF = LakesF %>% mutate(month = month(sampleDate, label=TRUE))
LakesF = LakesF %>% mutate(month = as.character(month))

#averaged duplicate measurements from the same day a created one row of data

LakesF = aggregate(LakesF$result,
      by=list(stationId=LakesF$stationId,sampleDate=LakesF$sampleDate,
             county=LakesF$county,year=LakesF$year,month=LakesF$month,
             stationName=LakesF$stationName),data=LakesF,FUN=mean)

#renamed column x to result

average = LakesF[,7]
LakesF = LakesF[,-7]
LakesF["result"] <- average
LakesF = LakesF %>% mutate(monthN = as.numeric(month))
LakesF = LakesF %>% mutate(year_d = yday(sampleDate))
```

Appendix II: Summary and Descriptive Statistics

```
dplyr::glimpse(PLakes)

summary(PLakes$result)

summary(PLakes)
```

```
head(LakesF)
```

```
glimpse(LakesF)

summary(LakesF$result)

summary(LakesF)
```

```
kable(LakesF %>% group_by(year) %>%
        summarise(avePhos=mean(result),n=n(),sdPhos=sd(result)))

kable(LakesF %>% group_by(county) %>%
        summarise(avePhos=mean(result),n=n(),sdPhos=sd(result)))

kable(LakesF %>% group_by(monthN) %>%
        summarise(avePhos=mean(result),n=n(),sdPhos=sd(result)))
```

```
#This code allows you to look at individual plots by year instead of all six years at once.

PlotYear = function(x,y,z){
  ggplot(data = LakesF, aes(x = sampleDate, y = log(result) , group=stationId)) +
  geom_line(size=.15, alpha=.20) +
  geom_smooth(aes(group = year),method="loess",span=.8,se=FALSE) +
  xlab("Date") + ylab("Phosphate") +  ggtitle(z) +
  scale_x_date(limit=c(as.Date(x),as.Date(y)))
}

PlotYear("2020-05-01","2020-10-11","2020")
PlotYear("2019-05-01","2019-10-11","2019")
PlotYear("2018-05-01","2018-10-11","2018")
PlotYear("2017-05-01","2017-10-11","2017")
PlotYear("2016-05-01","2016-10-11","2016")
PlotYear("2015-05-01","2015-10-11","2015")
```

Appendix III: Methods

Model Building

Determining the fixed effects structure:

```
initialmod <- lme(result ~ year + year_d + county, data=LakesF,
                  random=~ 1 | stationId, method="ML")

mod2 <- lme(result ~ year + year_d*county, data=LakesF,
            random=~ 1 | stationId, method="ML" )

mod3 <- lme(result ~ year_d*county, data=LakesF,
            random=~ 1 | stationId, method="ML" )

mod4 <- lme(result ~ poly(year_d,deg=2,raw=TRUE)*county + year, data=LakesF,
            random=~ 1 | stationId, method="ML" )

mod5 <- lme(result ~ poly(year_d,deg=2,raw=TRUE) + county + year, data=LakesF,
            random=~ 1 | stationId, method="ML" )

mod6 <- lme(result ~ county*(splines::ns(year_d,df=2)) + year , data=LakesF,
            random=~1 | stationId, method = "ML")

kable(AIC(initialmod,mod2,mod3,mod4,mod5,mod6))
anova(mod2,mod3)
```

Determining the best structure for the random effects:

```
mod4a <- lme(result ~ poly(year_d,deg=2,raw=TRUE)*county + year, data=LakesF,
             random=~ 1 | stationId, method="REML" )

mod4b <- lme(result ~ poly(year_d,deg=2,raw=TRUE)*county + year, data=LakesF,
             random=~ year_d | stationId, method="REML" )

mod4c <- lme(result ~ poly(year_d,deg=2,raw=TRUE)*county + year, data=LakesF,

kable(AIC(mod4a,mod4b,mod4c))
```

Adjusting for autocorrelation:

```
mod4b.car1 <- update(mod4b, correlation = corCAR1(form=~1))
mod4b.ar0ma1 <- update(mod4b, correlation = corARMA(form=~1,p=1,q=0))
mod4b.cs <- update(mod4b, correlation = corCompSymm(0.1, form=~1))
mod4b.Lin <- update(mod4b, correlation = corLin(form=~ 1))
mod4b.exp <- update(mod4b, correlation = corExp(form=~ 1 ))
mod4b.gaus <- update(mod4b, correlation = corGaus(form=~ 1))

kable(AIC(mod4b.car1,mod4b.ar0ma1,mod4b.cs,mod4b.exp,mod4b.gaus,mod4b.Lin))
```

Checking Assumptions

```
resids<-residuals(mod4b.ar0ma1,level=0)
fittedvals<-predict(mod4b.ar0ma1, level=0)
lakedframe<-data.frame(LakesF,resids,fittedvals)
```

```
ggplot(data = lakedframe, aes(x = year_d, y = resids, group=stationId)) +
geom_line(size=.15, alpha=.20) +
geom_smooth(aes(group = county),method="loess",span=.8) +
facet_wrap(~ county) + xlab("Day of the Year") + ylab("Residuals")

ggplot(data = lakedframe, aes(x = year_d, y = resids, group=stationId)) +
geom_line(size=.15, alpha=.20) +
geom_smooth(aes(group = year),method="loess",span=0.8) +
facet_wrap(~ year) + xlab("Day of the Year") + ylab("Residuals")
```

Results

```
summary(mod4b.ar0ma1)
```

```
plot(predictorEffects(mod4b.ar0ma1, ~ year_d))
plot(effect(term="year_d",mod=mod4b.ar0ma1),multiline=TRUE)
```

```
kable(multcomp::cld(pairs(emmeans(mod4b.ar0ma1,"county"))),digits=4,
      caption="Comparing mean differences for counties")
```

```
plot(effect(term="year",mod=mod4b.ar0ma1))
kable(multcomp::cld(pairs(emmeans(mod4b.ar0ma1,"year"))),digits=4,
      caption="Comparing mean differences for years")
```

References

(2009). "Phosphorus." URL https://www.pca.state.mn.us/water/phosphorus.

(2014). "The state of lakes." URL https://www.pca.state.mn.us/water/state-lakes.

(2021). "Introduction to Linear Mixed Models." URL https://stats.oarc.ucla.edu/other/mult-pkg/introduction-to-linear-mixed-models/.

(2022). "Create a Minnesota Climate Summary Table." URL https://www.dnr.state.mn.us/climate/historical/summary.html.

(2022). "Total Phosphorus." URL https://www.rmbel.info/primer/total-phosphorus/.

Brown VA (2021). "An Introduction to Linear Mixed-Effects Modeling in R." *Advances in Methods and Practices in Psychological Science*, **4**(1), 2515245920960351. ISSN 2515-2459. doi:10.1177/2515245920960351. URL https://doi.org/10.1177/2515245920960351.

Jacobson PC, Hansen GJA, Bethke BJ, Cross TK (2017). "Disentangling the effects of a century of eutrophication and climate warming on freshwater lake fish assemblages." *PLOS ONE*, **12**(8), e0182667. ISSN 1932-6203. doi:10.1371/journal.pone.0182667. URL https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0182667.

Jeppesen E, Kronvang B, Meerhoff M, Søndergaard M, Hansen KM, Andersen HE, Lauridsen TL, Liboriussen L, Beklioglu M, Özen A, Olesen JE (2009). "Climate Change Effects on Runoff, Catchment Phosphorus Loading and Lake Ecological State, and Potential Adaptations." *Journal of Environmental Quality*, **38**(5), 1930–1941. ISSN 00472425. doi:10.2134/jeq2008.0113. URL http://doi.wiley.com/10.2134/jeq2008.0113.

Scheiner SM, Gurevitch J (2006). *Design and Analysis of Ecological Experiments.* Oxford University Press, New York. ISBN 9780198030225. OCLC: 1058736933, URL http://www.SLQ.eblib.com.au/patron/FullRecord.aspx?p=271300.

Waterston AUaE (2019). *Chapter 21 Testing the assumptions | PPLS PhD Training Workshop: Statistics and R*. URL https://bookdown.org/animestina/phd_july_19/testing-the-assumptions.html.