

MyLake Evaluation for Lake 227

Kateri Salk

12 February 2018

1 Summary of Progress

1. Optimizations have been run for TP, TDP, and chl. The model fits are fairly poor.
2. Initial optimization has been run for dissolved oxygen. The model fit is poor, as oxygen is not consumed quickly enough following mixing.
3. Model fits are computed for:
 - R-squared value: amount of variance explained by the relationship (ideally close to 1)
 - RMSE: mean difference between modeled value and observed value, in the units of interest (ideally close to 0)
 - Nash-Sutcliffe coefficient: how well the model explains the observations (range: -infinity to 1; ideally close to 1; 0 indicates model predicts as well as the mean observation, > 0 indicates model predicts better than the mean observation, and < 0 indicates the model predicts more poorly than the mean observation)
4. I've included some ideas for improvements to the optimization.
5. I've also included ideas for adding in a switching function for N fixers vs. non-N fixers growth.

2 Model Performance: Physics

2.1 Ice Break and Ice Freeze

Ice break and ice freeze are within 5 days of the observed dates. An important note is that the observed dates are for Lake 239, not Lake 227. It is expected that lakes of smaller size, like Lake 227, will break and freeze a few days to a week in advance of Lake 239.

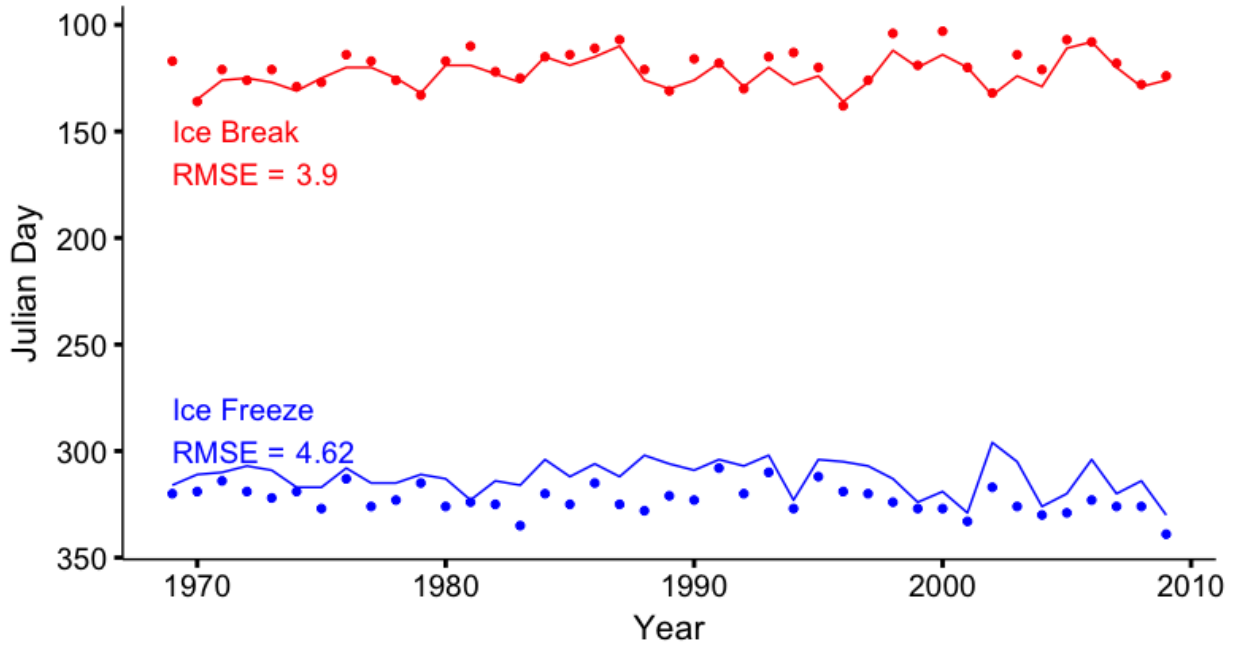


Figure 1: Date of observed (points) and modeled (lines) dates for ice break and ice freeze.

2.2 Temperature profiles

Temperatures at 1, 4, and 9 m are predicted within 1.42, 1.73, and 0.52 degrees, respectively. Nash-Sutcliffe values are all > 0 .

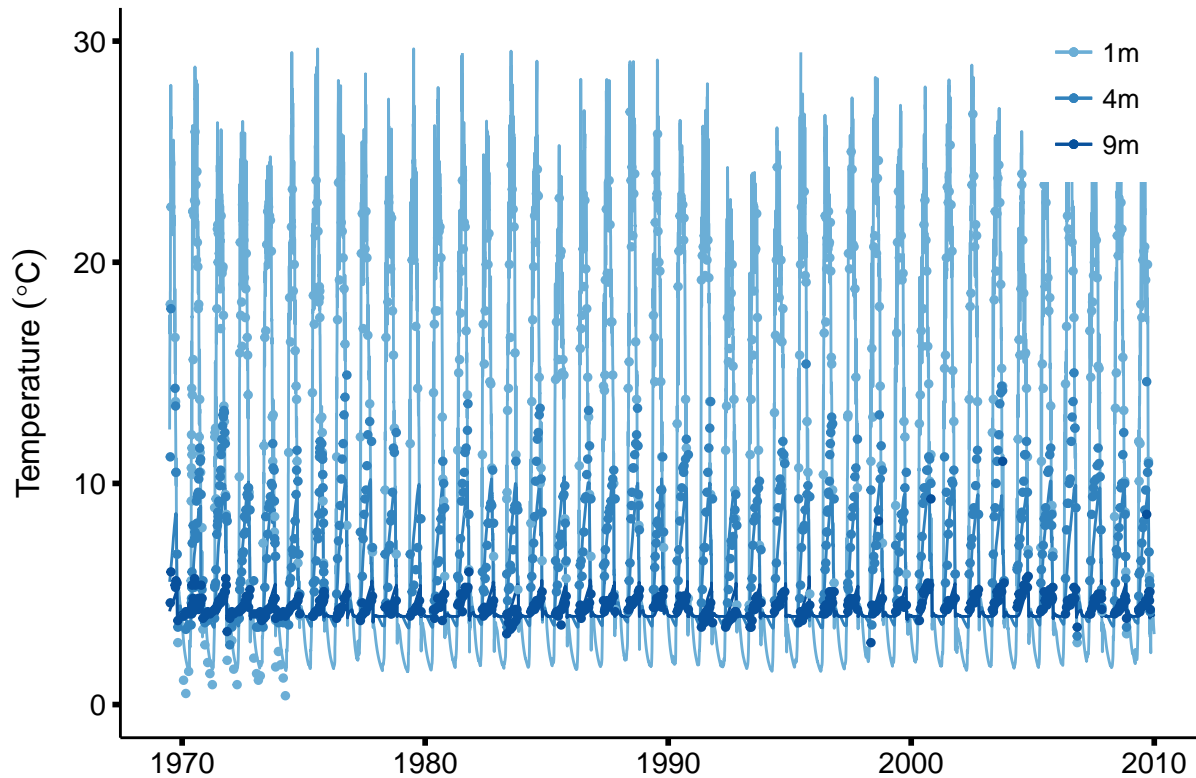


Figure 2: Temperatures at 1, 4, and 9 m depth for observed (points) and modeled (lines) dates.

```
temp1m.regression <- lm(mod2.match$obs.Temp1m ~ mod2.match$mod.Temp1m)
summary(temp1m.regression)$adj.r.squared

## [1] 0.9464716
mse.temp1m <- mean(residuals(temp1m.regression)^2); rmse.temp1m <- sqrt(mse.temp1m); rmse.temp1m

## [1] 1.421932
NashSutcliffe.temp1m <- NSE(mod2.match$mod.Temp1m, mod2.match$obs.Temp1m); NashSutcliffe.temp1m

## [1] 0.8733446
temp4m.regression <- lm(mod2.match$obs.Temp4m ~ mod2.match$mod.Temp4m)
summary(temp4m.regression)$adj.r.squared

## [1] 0.5513304
mse.temp4m <- mean(residuals(temp4m.regression)^2); rmse.temp4m <- sqrt(mse.temp4m); rmse.temp4m

## [1] 1.734417
```

```

NashSutcliffe.temp4m <- NSE(mod2.match$mod.Temp4m, mod2.match$obs.Temp4m); NashSutcliffe.temp4m
## [1] 0.372049
temp9m.regression <- lm(mod2.match$obs.Temp9m ~ mod2.match$mod.Temp9m)
summary(temp9m.regression)$adj.r.squared
## [1] 0.1872678
mse.temp9m <- mean(residuals(temp9m.regression)^2); rmse.temp9m <- sqrt(mse.temp9m); rmse.temp9m
## [1] 0.5190508
NashSutcliffe.temp9m <- NSE(mod2.match$mod.Temp9m, mod2.match$obs.Temp9m); NashSutcliffe.temp9m
## [1] 0.05675739

```

3 Model Performance: Phosphorus

The model fits are poor for phosphorus-related variables. This parameterization is optimized for 1979-1989, but the fit for that period is still not good.

Parameters for optimization period 1979-1989 (fit for this model is shown in this report):

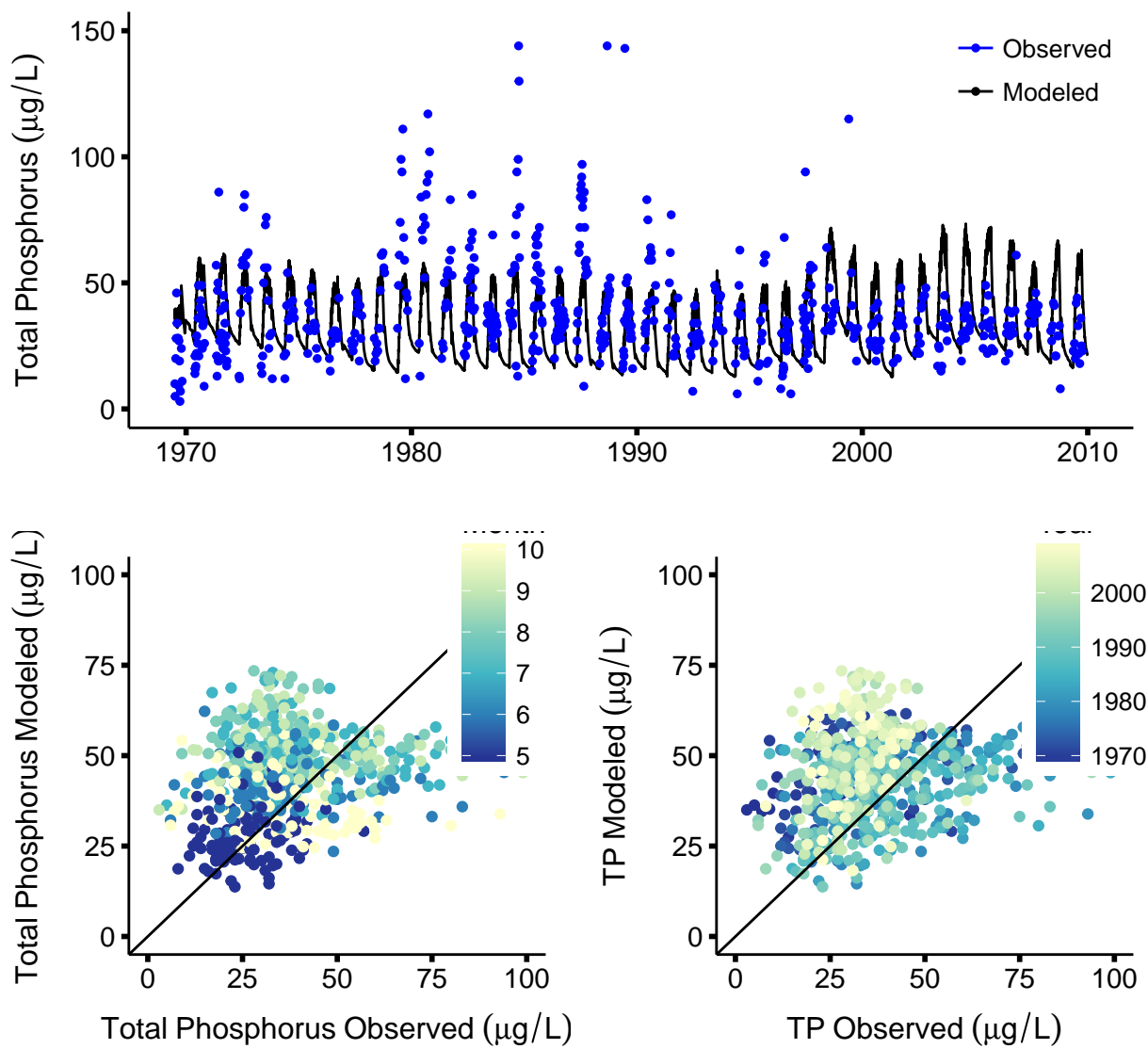
- Par_sat: 0.0002421
- w_chl: 0.1103
- m_twty: 0.02
- g_twty: 0.5987
- P_half: 0.7408

Parameters for optimization period 1969-1979 (not shown):

- Par_sat: 0.000238
- w_chl: 0.114
- m_twty: 0.02
- g_twty: 1.12
- P_half: 1.68

3.1 Total Phosphorus

Model predictions tend to be underestimated in the 1980s but overestimated in the 2000s.



```
TPregression <- lm(mod.match$obs.TP ~ mod.match$mod.TP)
summary(TPregression)$adj.r.squared
```

```
## [1] 0.01953384
```

```
mse.TP <- mean(residuals(TPregression)^2); rmse.TP <- sqrt(mse.TP); rmse.TP
```

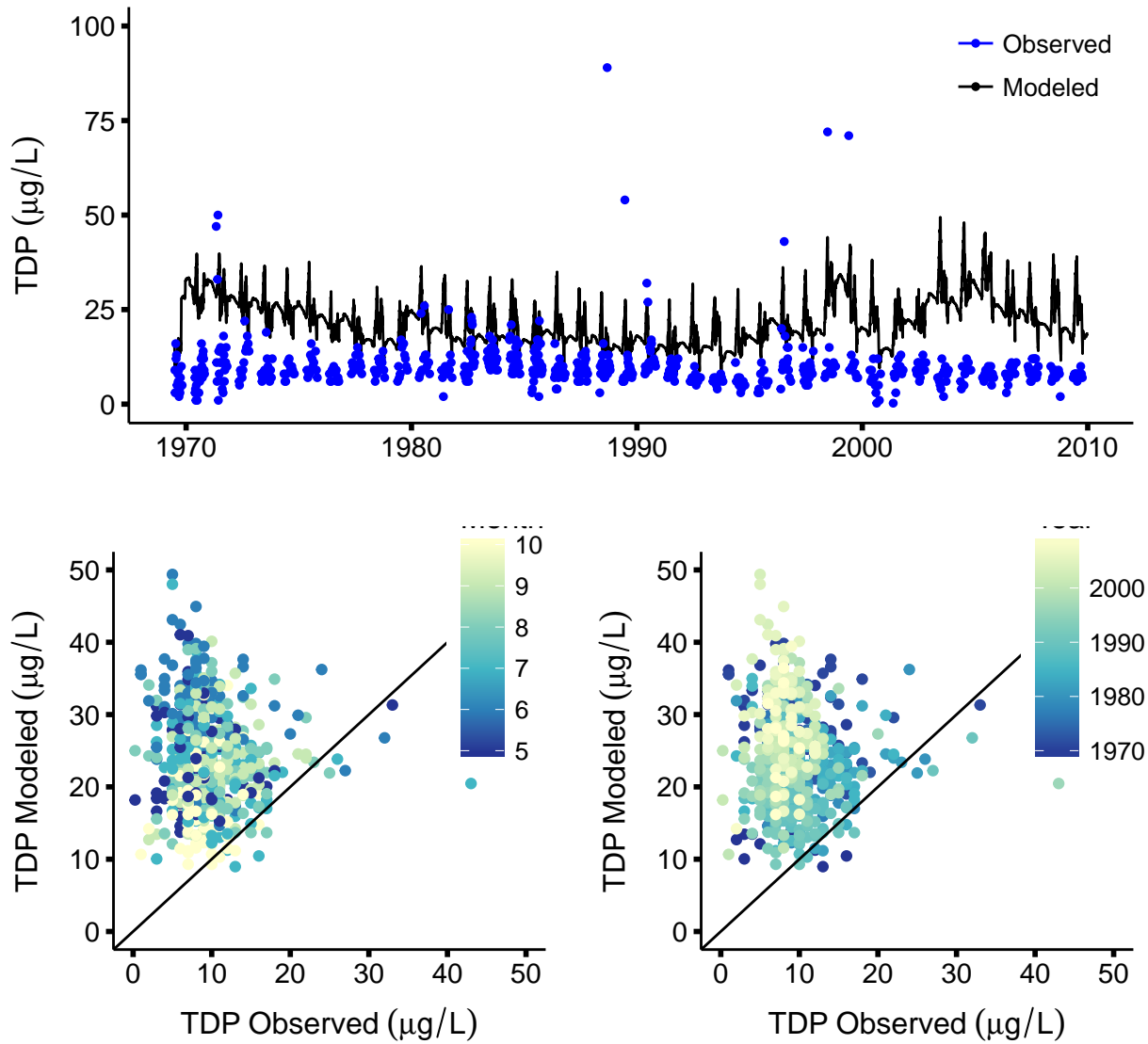
```
## [1] 38.07992
```

```
NashSutcliffe.TP <-NSE(mod.match$mod.TP, mod.match$obs.TP); NashSutcliffe.TP
```

```
## [1] -0.009707243
```

3.2 Total Dissolved Phosphorus

TDP is nearly always overestimated (too much P in the dissolved pool, not enough in the particulate pool).



```
TDPregression <- lm(mod.match$obs.TDP ~ mod.match$mod.TDP)
summary(TDPregression)$adj.r.squared

## [1] 0.001448681

mse.TDP <- mean(residuals(TDPregression)^2); rmse.TDP <- sqrt(mse.TDP); rmse.TDP

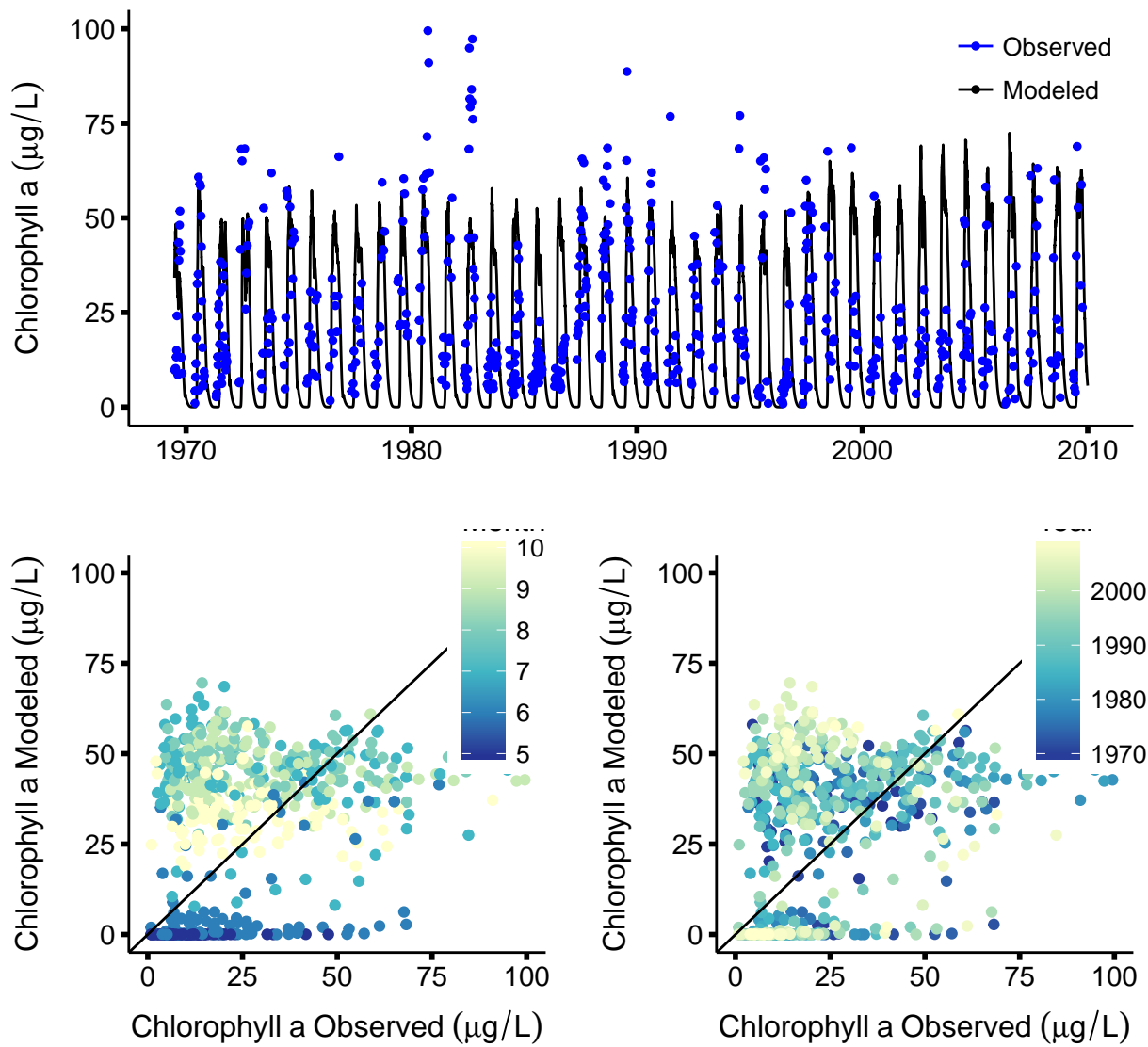
## [1] 26.4752

NashSutcliffe.TDP <- NSE(mod.match$mod.TDP, mod.match$obs.TDP); NashSutcliffe.TDP

## [1] -0.2391471
```

3.3 Chlorophyll

Chlorophyll tends to be underestimated in the spring (missing the diatom bloom?) and overestimated in the summer and fall. There isn't an apparent effect of year for chlorophyll.



```
chlaregression <- lm(mod.match$obs.chla ~ mod.match$mod.chla)
summary(chlaregression)$adj.r.squared
```

```
## [1] 0.08001171
```

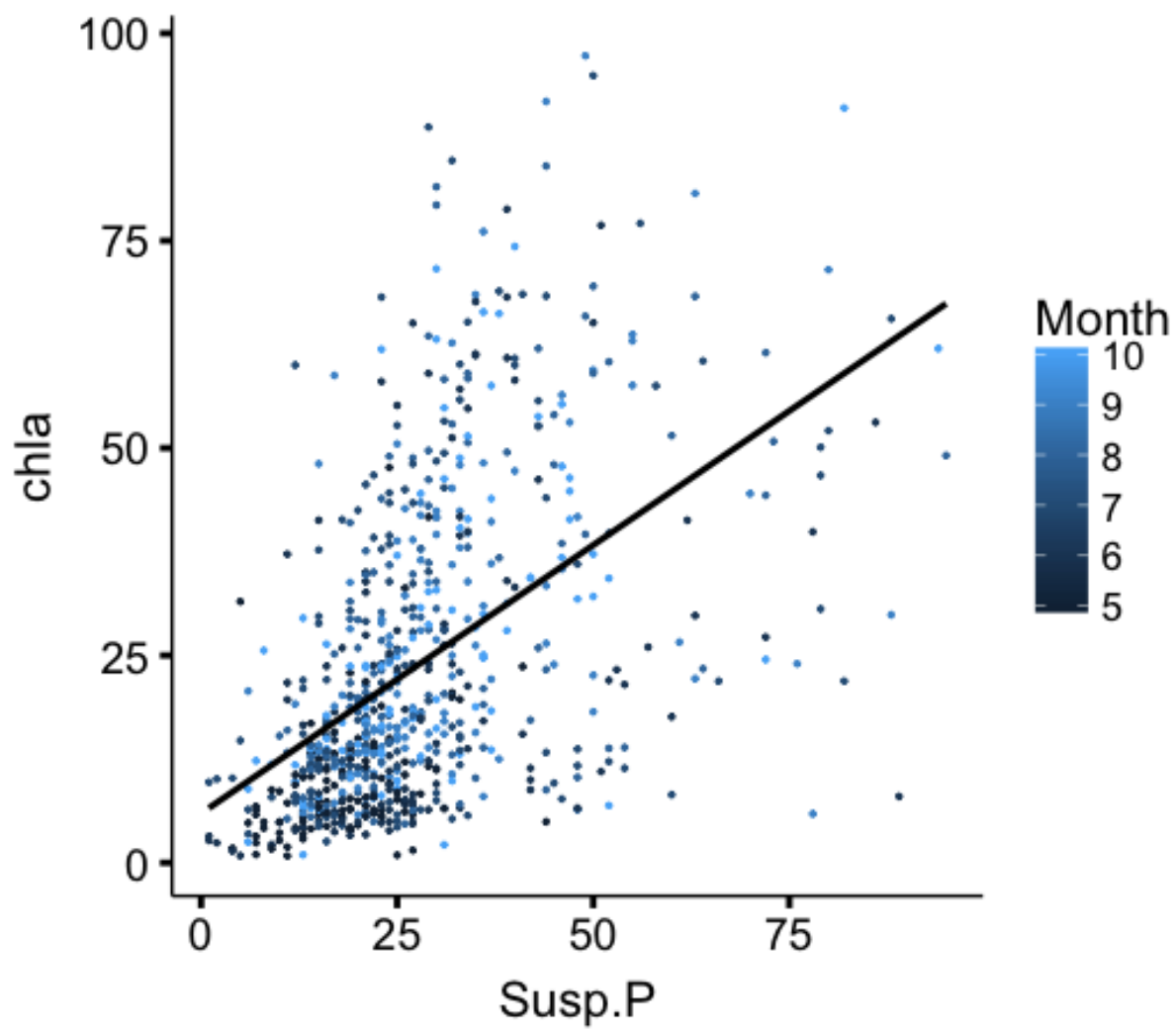
```
mse.chl <- mean(residuals(chlaregression)^2); rmse.chl <- sqrt(mse.chl); rmse.chl
```

```
## [1] 21.91558
```

```
NashSutcliffe.chl <- NSE(mod.match$mod.chla, mod.match$obs.chla); NashSutcliffe.chl
```

```
## [1] -0.3402085
```

Right now, we are assuming a phytoplankton P:chl ratio of 0.42 (Hecky et al. 1993). Does this ratio always hold? Here is a plot of particulate P (phytoplankton + other particulates) vs. chl concentrations. The slope of the line is 0.65, but the R-squared value is only 0.26. This is significantly dependent on month. I would conclude that assuming a consistent P:chl ratio is not appropriate, and perhaps we should use a different metric for predicting phytoplankton biomass. I suggest using particulate P as a phytoplankton metric would be appropriate.

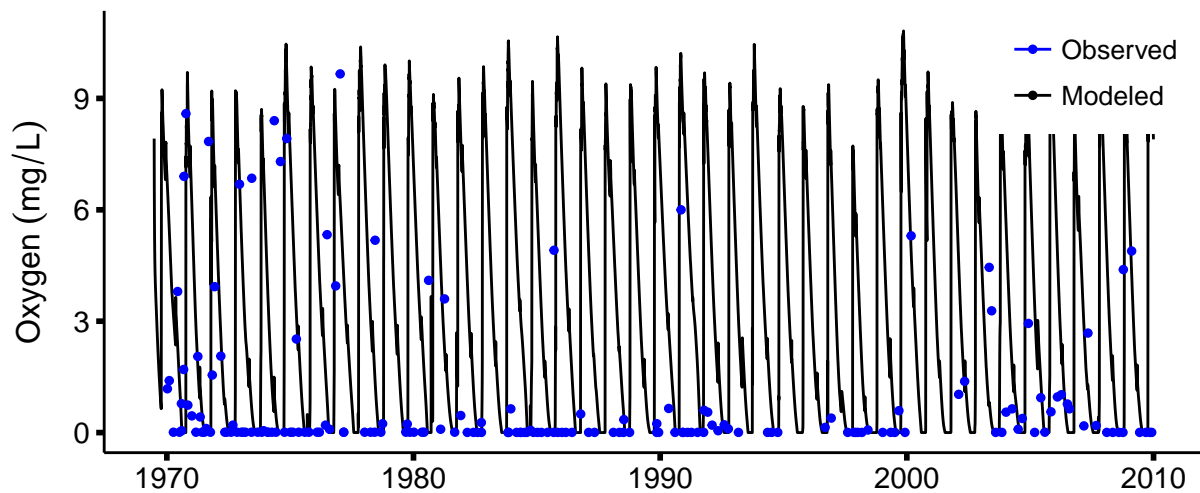
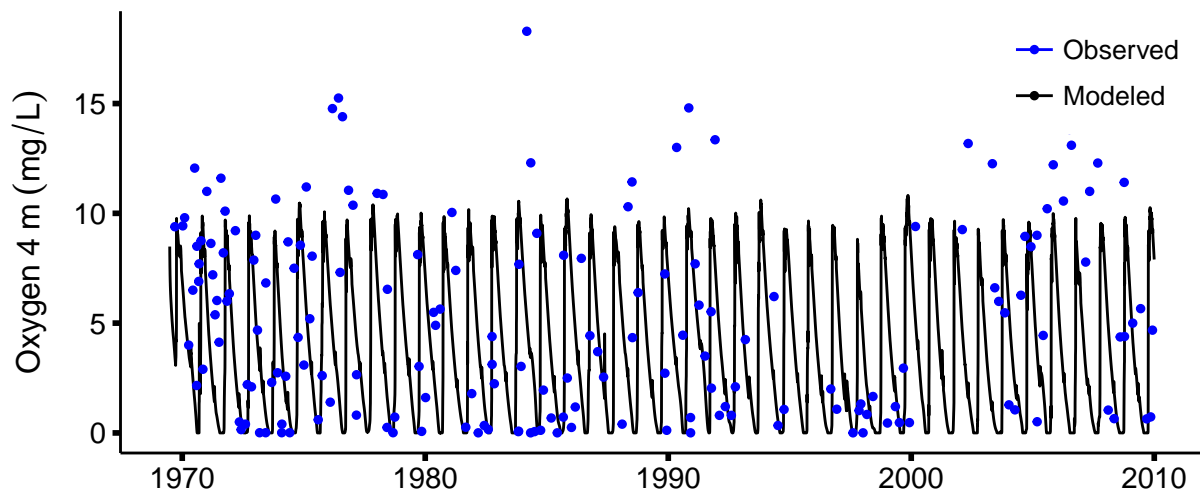


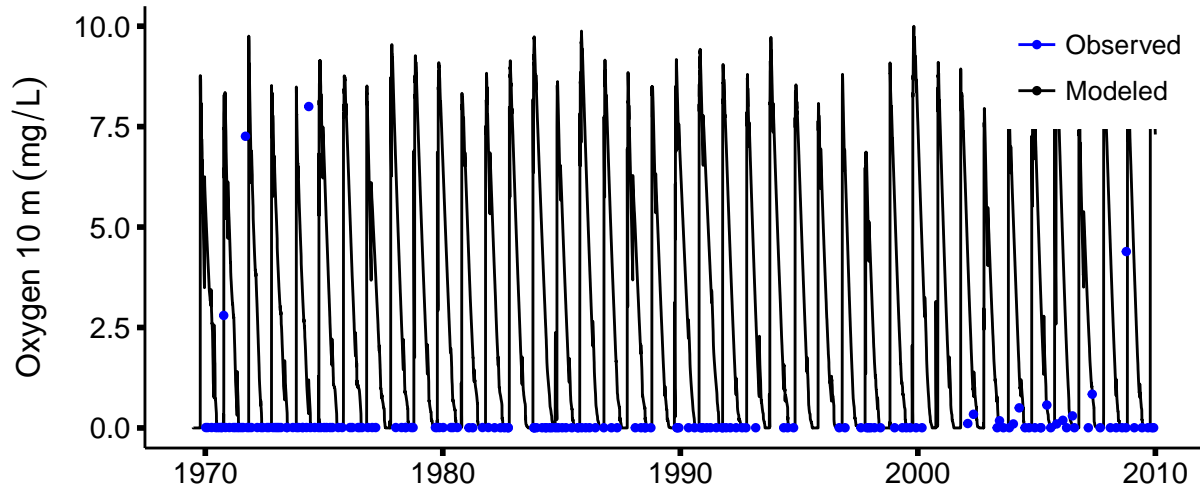
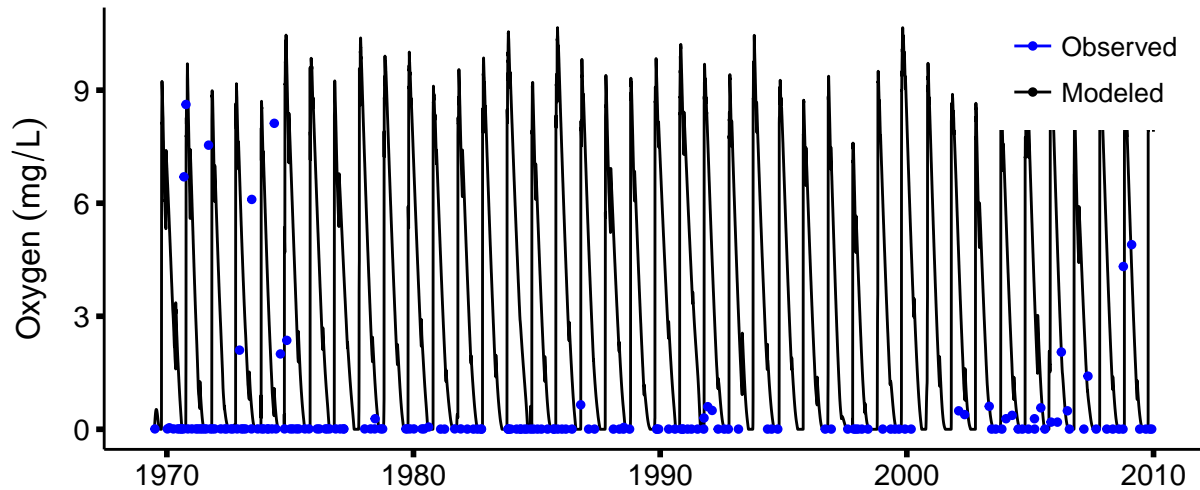
4 Model Performance: Oxygen

The model fits for oxygen are poor. When water column mixing occurs, the oxygen concentrations don't decrease quickly enough.

Oxygen concentrations were optimized separately from phosphorus. Parameters for optimization period 1979-1989 (fit for this model is shown in this report):

- k_{chl} : 1.3665
- k_{POP} : 4.7616
- k_{POC} : 0.2427
- k_{DOP} : 3.3683
- k_{DOC} : 0.02
- Km_{O2} : 0.0226
- Kin_{O2} : 1.9823





```
O2regression4m <- lm(mod3.match$obs.O2.4m ~ mod3.match$mod.Oxy4m)
summary(O2regression4m)$adj.r.squared
```

```
## [1] -0.002908553
```

```
mse.024m <- mean(residuals(O2regression4m)^2); rmse.024m <- sqrt(mse.024m); rmse.024m
```

```
## [1] 4.420057
```

```
NashSutcliffe.024m <- NSE(mod3.match$mod.Oxy4m, mod3.match$obs.O2.4m); NashSutcliffe.024m
```

```
## [1] -0.7277797
```

```
O2regression6m <- lm(mod3.match$obs.O2.6m ~ mod3.match$mod.Oxy6m)
summary(O2regression6m)$adj.r.squared
```

```
## [1] -0.005525946
```

```
mse.026m <- mean(residuals(O2regression6m)^2); rmse.026m <- sqrt(mse.026m); rmse.026m
```

```
## [1] 2.112079
NashSutcliffe.026m <- NSE(mod3.match$mod.Oxy6m, mod3.match$obs.02.6m); NashSutcliffe.026m

## [1] -3.753451
O2regression8m <- lm(mod3.match$obs.02.8m ~ mod3.match$mod.Oxy8m)
summary(O2regression8m)$adj.r.squared

## [1] 0.007287958
mse.028m <- mean(residuals(O2regression8m)^2); rmse.028m <- sqrt(mse.028m); rmse.028m

## [1] 1.329809
NashSutcliffe.028m <- NSE(mod3.match$mod.Oxy8m, mod3.match$obs.02.8m); NashSutcliffe.028m

## [1] -10.81276
O2regression10m <- lm(mod3.match$obs.02.10m ~ mod3.match$mod.Oxy10m)
summary(O2regression10m)$adj.r.squared

## [1] 0.009458649
mse.0210m <- mean(residuals(O2regression10m)^2); rmse.0210m <- sqrt(mse.0210m); rmse.0210m

## [1] 0.8868525
NashSutcliffe.0210m <- NSE(mod3.match$mod.Oxy10m, mod3.match$obs.02.10m); NashSutcliffe.0210m

## [1] -21.39327
```

5 Stoichiometric indicators of N fixer shift in L227

We need to come up with a stoichiometric indicator to act as a switch for N fixer vs. non-N fixer growth in Lake 227. The key challenge here is that N fixers show up after 1975, but they don't dominate the entire season. Is there stoichiometric indicator that can capture intra-annual changes in the relative availability of N for phytoplankton?

I suggest using either

1. Combined fertilizer and inflow N:P ratios (low on fertilizer days, higher on non-fertilizer days)
2. DIN concentration

We would need to choose a threshold value, in either case, to serve as the switch point of a sigmoid function that dictates the proportion of phytoplankton growth allocated to N fixers vs. non-N fixers.

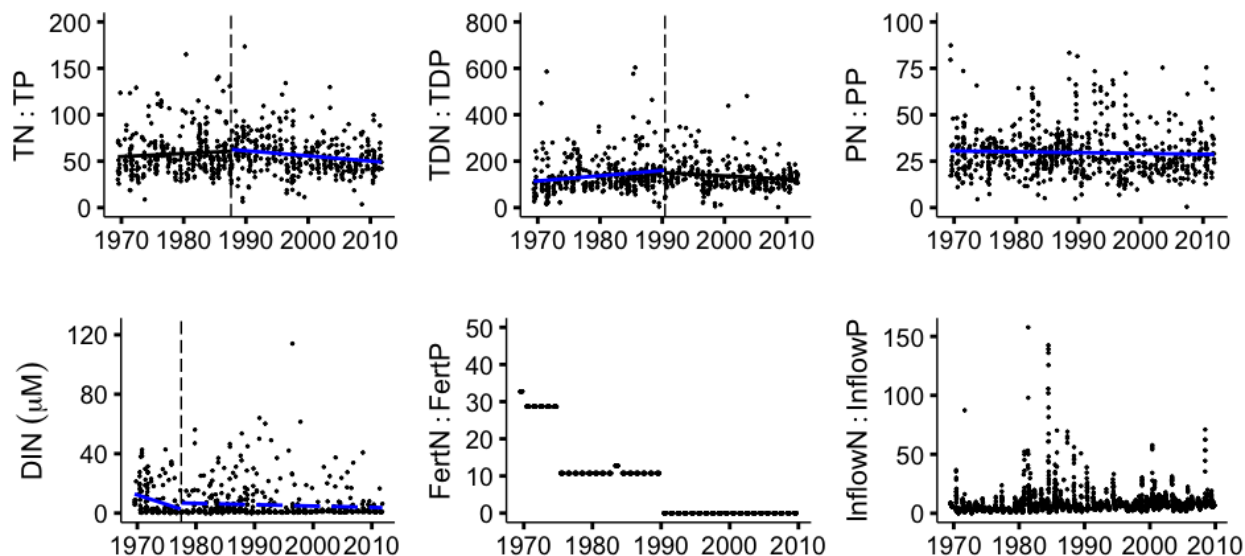


Figure 3: N:P ratios in various pools in Lake 227. Vertical lines represent calculated breakpoints for linear regressions. Linear regressions with a slope significantly different than zero are represented with a blue line, and those with a slope not significantly different from zero are represented with a black line.

6 Next steps and follow-up questions

1. I propose using TPP and TDP as the optimization variables for phosphorus, rather than TP, TDP, and chl. I would have liked to optimize with only phytoplankton-associated PP, but the observation data from L227 contain only aggregate PP.
2. We need to determine ways to make the parameterization for both P and oxygen to be more successful. Any ideas for parameters we have not tried?
3. The next step after parameterization is to add a sigmoid function to parse phytoplankton growth between the two pools of phytoplankton. I will need help with the code on this. I suggest 16 for N:P and 1 μM for DIN concentration as a starting point for the switch point.
4. I will need to generate several years of data for simulated weather conditions plus changes to fertilization regime. Is there an algorithm of some sort that can generate simulated data?