Nutrient Trends Along the Klamath River

https://github.com/[redacted]/Klamath_River_Nutrient_Analysis

[redacted]

Abstract

The Klamath River drains over 15,000 square miles in Southern Oregon and Northern California between the Cascade Mountains and the Basin and Range Province. The River Basin is primarily covered by forested land and agricultural or rangeland that contribute to excess nutrient loading to the river, along with nutrients from the basin's geology. This analysis looks for trends in organic carbon, nitrogen, and phosphorus concentrations along the Klamath River from 2000 through 2018. Nitrogen was found to decrease downstream of Upper Klamath Lake, though nitrogen concentrations did not change over time. Phosphorus and Organic Carbon concentrations varied with distance downstream and over time. To decrease nutrient loading more quickly in the Klamath River, it may be advised to target areas in the River's upper reaches.

Contents

List of Tables

List of Figures

<Note: set up autoreferencing for figures and tables in your document>

1 Research Question and Rationale

The Klamath River Basin covers over 15,000 square miles in Southern Oregon and Northern California between the Cascade Mountains and the Basin and Range Province. Though 75% of the basin is forested, including five national forests, another 20% of the basin is used for agriculture or as rangeland for grazing. Almost all of this agricultural land lies in the Upper Klamath River Basin. This agricultural land increases nutrient loading to the river, though previous studies have also shown that groundwater contributes significant nutrients, due to the basin's geology. California lists stretches of the Klamath River as impaired, due to mostly to high temperatures and excess nutrient loading. These high temperatures and nutrient loading, paired with a number of dams along the river's main stem, have negative impacts for the river's biotic communities. Indeed, there are two fish species listed as endangered and one fish species listed as threatened under the Federal Endangered Species Act in the Klamath River. Biotic data for the Klamath River is sparse, so nutrient data is analyzed here, due to its impact on biotic communities, including the endangered fish species in the river. Additionally, this analysis is especially relevant because the four primary Klamath River dams are under review for removal, and the removals would cost an estimated $400 million.

This analysis investigates the spatial distribution of nitrogen, phosphorus, and organic carbon along the Klamath River, beginning at two tributaries just above Upper Klamath Lake. As the majority of agricultural land is above Upper Klamath Lake and the water level of the lake is controlled at the Link River Dam, this analysis predicts that nutrient levels will decrease along the length of the river.

2 Dataset Information

The data for this analysis was downloaded from the National Water Quality Monitoring Council's Water Quality Data Portal at https://www.waterqualitydata.us/portal/. The dataset analyzed contains data on Organic Carbon, Nitrogen, and Phosphorus at 25 sampling sites along the Klamath River from January 1st, 2000 to January 1st, 2019. Not all sites collect data on all three nutrients. This nutrient dataset was paired with location data for all 159 water quality monitoring stations in the Klamath River Basin, HUC code 180102. The 25 sites for which data was downloaded were selected to provide data throughout the length of the river, and each site must have data from at least 20 sampling events.

Parameter	Summary
Number of Stations	25
Total Number of Samples	5670
Nutrients	Nitrogen, Phosphorus, Carbon
Beginning Date	January 1, 2000
End Date	January 1, 2019

3 Exploratory Data Analysis and Wrangling

3.1 Nutrient Data Wrangling

The raw nutrient data contains superfluous information, so all data was removed from the dataset besides sampling location information, the date and time of the sampling activity, and the measured nutrient concentrations. The raw dataset also contains information on the type of sample (total, dissolved, filtered, etc.) and the units of measurement. Since the sample measurements are not comparable across sample types, only data for the most common sample type of each nutrient were retained for analysis.

```r
## read in nutrient data
klamath <- read.csv("./Data/RAW/Klamath_River_Nutrients_Raw.csv",
                    header = T)
## read in station data
stations <- read.csv("./Data/RAW/Klamath_River_Stations_Raw.csv",
                     header = T)

## reformat date
klamath$ActivityStartDate <- as.Date(klamath$ActivityStartDate,
                                     format = "%Y-%m-%d")

## keep only useful columns, select for columns
klamath_data <- klamath %>%
  select(ActivityStartDate, ActivityStartTime.Time, MonitoringLocationIdentifier,
         CharacteristicName, ResultSampleFractionText, ResultMeasureValue,
         ResultMeasure.MeasureUnitCode,
         ResultDepthHeightMeasure.MeasureValue,
         ResultDepthHeightMeasure.MeasureUnitCode)

## rename columns
names(klamath_data) <- c("Date", "Time", "Location", "Nutrient",
                         "SampleFraction", "Concentration",
                         "ConcentrationUnits", "Depth", "DepthUnit")

## remove duplicates using distinct()
klamath_distinct <- distinct(klamath_data)
## remove one data point from each pair.

klamath_distinct <- klamath_distinct[-c(882, 884, 888, 889, 893, 896, 957,
                                        1011,879, 881, 883, 887, 946, 950,
                                        955, 959, 1008),]

## Get all data of each nutrient into the same sample fraction type and units
```

```r
## filter the nutrients separately
carbon <- filter(klamath_distinct, Nutrient == "Organic carbon")
nitrogen <- filter(klamath_distinct, Nutrient == "Nitrogen")
phosphorus <- filter(klamath_distinct, Nutrient == "Phosphorus")

## determine which sample type to keep
# summary(carbon$SampleFraction) # keep dissolved
carbon_dissolved <- filter(carbon, SampleFraction == "Dissolved")
# summary(nitrogen$SampleFraction) # keep total
nitrogen_total <- filter(nitrogen, SampleFraction == "Total")
summary(phosphorus$SampleFraction) # keep total
```

```
##                  Dissolved            Filtered, lab
##                         10                        0
## Non-Filterable (Particle)                 Suspended
##                          0                        0
##                      Total               Unfiltered
##                       2433                        6
```

```r
phosphorus_total <- filter(phosphorus, SampleFraction == "Total")
# summary(carbon_dissolved$ConcentrationUnits)

## Convert ug/l to mg/l for nitrogen and phosphorus, and make a factor
nitrogen_ug_to_mg <- nitrogen_total %>%
  filter(ConcentrationUnits == "ug/l") %>%
  mutate(Concentration = Concentration/1000) %>%
  mutate(ConcentrationUnits = as.factor("mg/l"))

phosphorus_ug_to_mg <- phosphorus_total %>%
  filter(ConcentrationUnits == "ug/l") %>%
  mutate(Concentration = Concentration/1000) %>%
  mutate(ConcentrationUnits = as.factor("mg/l"))

## Remove everything but mg/l from nitrogen_total and phosphorus_total
## bind_rows() to converted data
nitrogen_total_mg <- nitrogen_total %>%
  filter(ConcentrationUnits == "mg/l") %>%
  bind_rows(nitrogen_ug_to_mg) %>%
  mutate(ConcentrationUnits = as.factor(ConcentrationUnits))

phosphorus_total_mg <- phosphorus_total %>%
  filter(ConcentrationUnits == "mg/l") %>%
  bind_rows(phosphorus_ug_to_mg) %>%
  mutate(ConcentrationUnits = as.factor(ConcentrationUnits))
```

```
## bind_rows to re-unite all nutrient data
klamath_clean <- bind_rows(carbon_dissolved,
                           phosphorus_total_mg,
                           nitrogen_total_mg) %>%
  mutate(ConcentrationUnits = as.factor(ConcentrationUnits))

## spread data
klamath_spread <- spread(data = klamath_clean, Nutrient, Concentration)

## remove space from Organic carbon column name
names(klamath_spread)[9] <- c("Organic_Carbon")
```

The most common sample type for organic carbon was dissolved, and the most common sample type for both nitrogen and phosphorus was total. After selecting for only these types of samples, all units were converted to mg/L. The phosphorous data also contained measurements in mg/L as P and mg/kg as P. As these could not be converted to mg/L, they were removed. There were also 13 pairs of data collected at the same time and location, but with different results. As the removal of these data will will not substantially affect data trends, one data point from each pair was removed.

3.2 Station Location Data Wrangling

The locations of the stations were reported using three different datums; NAD27, NAD83, and WGS84. However, all three datums were treated as NAD83 because the difference in computed distance is small compared to the scale that separates monitoring stations. Also, rather than following the path of the river to determine station separation distance, the distances were computed as straight line distance from the entrance to Upper Klamath Lake, station ID KLAMATHTRIBES_WQX-WR6000. This distance was used an estimated proxy for distance downstream.

```
## select location columns
stations_loc <- stations %>%
  select(MonitoringLocationIdentifier, MonitoringLocationName,
         LatitudeMeasure,LongitudeMeasure,
         HorizontalCoordinateReferenceSystemDatumName)
## rename location column to match klamath spread column name for joining
names(stations_loc)[1] <- c("Location")
# make into an sf object
stations_sf <- st_as_sf(stations_loc,
                        coords = c('LongitudeMeasure',
                                   'LatitudeMeasure'),crs=4269)
## change coordinate system to UTM Zone 10N, EPSG: 26910, for distance calculations
stations_UTM <- st_transform(stations_sf, crs=26910)
```

```r
## calculate distance from site KLAMATHTRIBES_WQX-WR6000 to all other sites
Distance_to_MODOC <- stations_UTM %>%
  filter(Location == 'KLAMATHTRIBES_WQX-WR6000') %>%
  st_distance(stations_UTM) %>%   #Compute distances to all other sites
  data.frame() %>%  t             #Transpose the result
## add the distance vector to the stations_loc dataframe,
## converting from meters to kilometers
stations_UTM$distance <- Distance_to_MODOC/1000

## change two upstream sites to have negative distances
## KLAMATHTRIBES_WQX-SR0080 and KLAMATHTRIBES_WQX-WR2000
stations_UTM$distance[125] <- stations_UTM$distance[125]*-1
stations_UTM$distance[130] <- stations_UTM$distance[130]*-1

## join nutrient data with location data
klamath_spread_loc <- left_join(klamath_spread, stations_UTM, by = "Location")
klamath_spread_sf <- st_as_sf(klamath_spread_loc,sf_column_name = 'geometry')

## create a separate sf object that is only the origin station
station_origin <- filter(klamath_spread_sf,
                         Location == 'KLAMATHTRIBES_WQX-WR6000')

## create a separate data frame without the three stations upstream of
## Link River Dam
klamath_trend_sf <- klamath_spread_sf %>%
  filter(Location != 'KLAMATHTRIBES_WQX-WR6000',
         Location != 'KLAMATHTRIBES_WQX-SR0080',
         Location != 'KLAMATHTRIBES_WQX-WR2000')
```

```r
## data summary
summary(klamath_data)
```

```
##       Date                  Time                             Location
##  Min.   :2000-05-01           : 414   KLAMATHTRIBES_WQX-WR6000: 714
##  1st Qu.:2008-04-08   00:00:00: 248   KLAMATHTRIBES_WQX-SR0080: 696
##  Median :2010-10-19   13:00:00:  97   KLAMATHTRIBES_WQX-WR2000: 636
##  Mean   :2010-05-24   10:00:00:  87   YUROKTEP_WQX-TC         : 298
##  3rd Qu.:2013-08-22   09:30:00:  86   CEDEN-KR12850_KHSA      : 288
##  Max.   :2018-10-24   11:00:00:  77   CEDEN-KR15626_KHSA      : 288
##                       (Other) :4661   (Other)                 :2750
##            Nutrient                      SampleFraction Concentration
##  Nitrogen      :2033   Dissolved                 : 980   Min.   :  0.0075
##  Organic carbon:1179   Filtered, lab             :   6   1st Qu.:  0.2200
##  Phosphorus    :2458   Non-Filterable (Particle):   5   Median :  2.4300
##                        Suspended                 : 188   Mean   : 52.3953
```

```
##                        Total                    :4485   3rd Qu.: 56.0000
##                        Unfiltered               :   6   Max.   :873.0000
##                                                         NA's   :6
##   ConcentrationUnits     Depth       DepthUnit
##             :   3     Min.   :0.100    :3342
##  mg/kg as P:  24      1st Qu.:0.500   m:2328
##  mg/l       :3161     Median :0.500
##  mg/l as P : 436      Mean   :0.473
##  ug/l       :2046     3rd Qu.:0.500
##                       Max.   :0.500
##                       NA's   :3342
```

```
## summary table
klamath_summary_UTM <- klamath_spread_sf %>%
  group_by(Location) %>%
  summarize(meanN = mean(Nitrogen, na.rm = T),
            meanP = mean(Phosphorus, na.rm = T),
            meanC = mean(Organic_Carbon, na.rm = T))

klamath_summary_UTM
```

```
## Simple feature collection with 24 features and 4 fields
## geometry type:  POINT
## dimension:      XY
## bbox:           xmin: 416473.8 ymin: 4559613 xmax: 613495.1 ymax: 4735530
## epsg (SRID):    26910
## proj4string:    +proj=utm +zone=10 +ellps=GRS80 +towgs84=0,0,0,0,0,0,0 +units=m +no_d
## # A tibble: 24 x 5
##    Location             meanN  meanP  meanC           geometry
##    <chr>                <dbl>  <dbl>  <dbl>        <POINT [m]>
##  1 CEDEN-105KL0330      0.186 0.0561   1.85 (416557.3 4596523)
##  2 CEDEN-105KL2057      0.252 0.0516   3.97 (440707.2 4559613)
##  3 CEDEN-105KL9073      1.01  0.119    1.36 (546098.7 4641909)
##  4 CEDEN-KR00600_KHSA   0.245 0.0408   1.42 (416473.8 4596524)
##  5 CEDEN-KR03850_KHSA   0.242 0.0470   1.65 (435296.2 4564200)
##  6 CEDEN-KR04350_KHSA   0.314 0.0550   1.99 (440539.5 4559614)
##  7 CEDEN-KR05910      NaN     0.134  NaN    (455382.2 4572930)
##  8 CEDEN-KR10080      NaN     0.150  NaN    (464737.1 4619853)
##  9 CEDEN-KR12850_KHSA   0.620 0.107    3.46 (481735.2 4632291)
## 10 CEDEN-KR15626_KHSA   0.810 0.130    4.06 (511208.8 4631732)
## # ... with 14 more rows
```

```
## dataframe of only sampling events with all three nutrients
klamath_complete <- klamath_spread %>%
  filter(!is.na(Nitrogen) & !is.na(Phosphorus) & !is.na(Organic_Carbon))
## there are no sampling events where data was collected for all nutrients
```

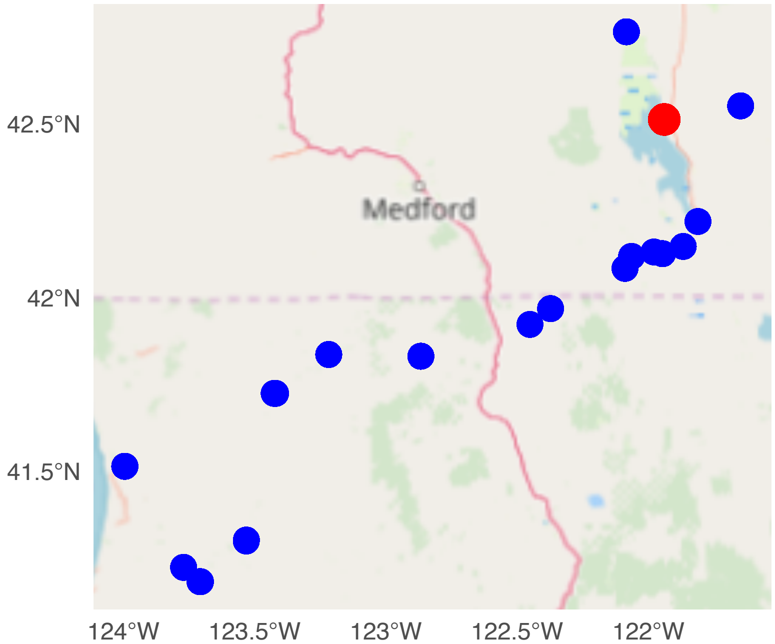

Figure 1: Locations of the 25 stations for which data was analyzed along the Klamath River. The station at the entrance of Upper Klamath Lake is shown in red and was used as the upstream origin for further analysis.

There were no sampling events where phosphorus, nitrogen, and organic carbon were all sampled at the same time.

As can be seen in figures 2, 3, and 4, the concentration of each nutrient appears to decrease downstream in the river. However, this trend does not appear to hold for the origin site and the two upstream sites

Figures 5, 6, and 7 show similarly that there may be a slight decreasing trend over time .

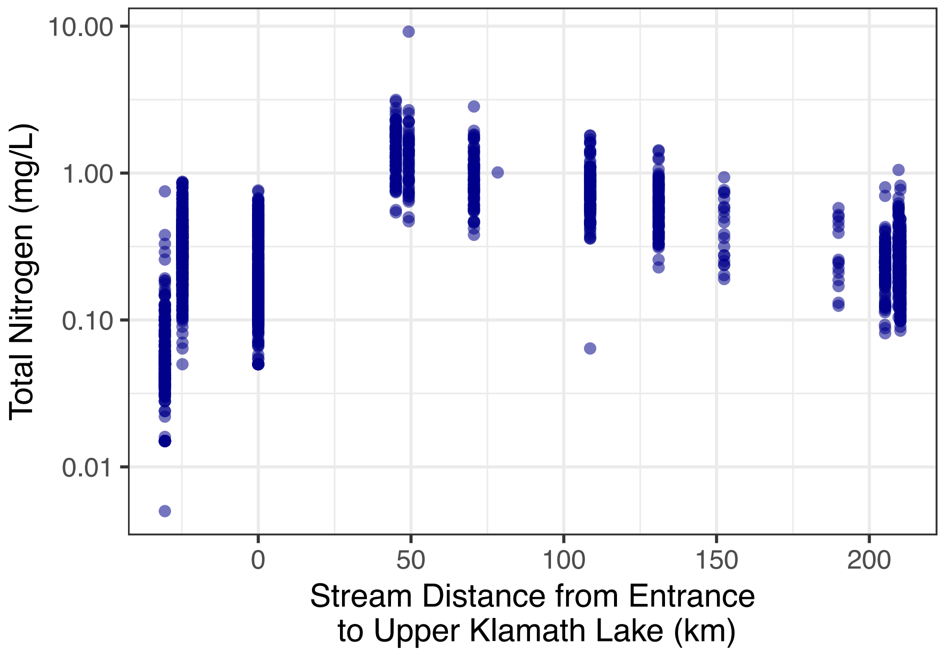

Figure 2: Nitrogen concentration versus distance downstream.

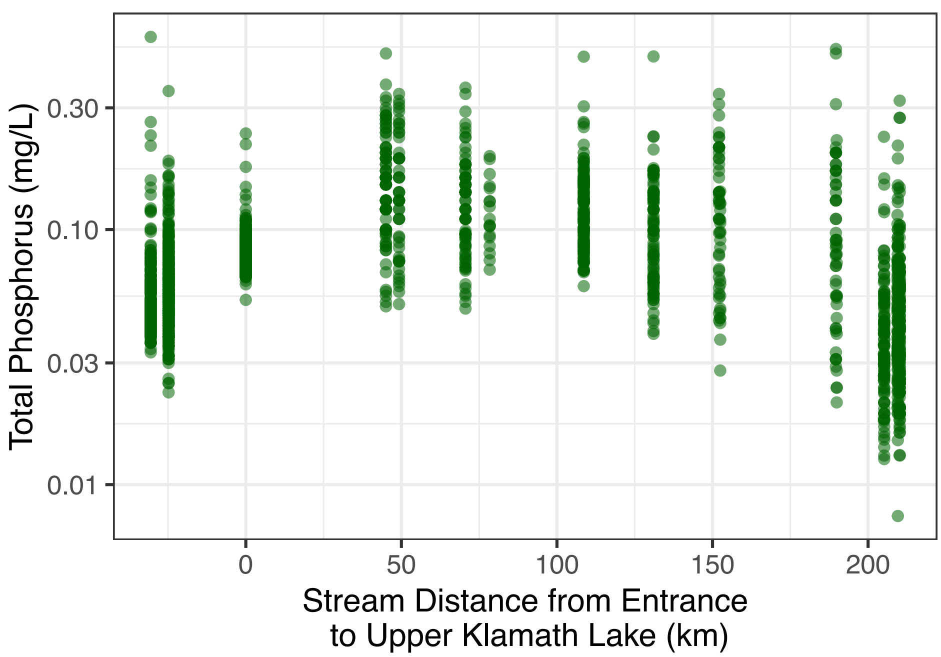

Figure 3: Phosphorus concentration versus distance downstream.

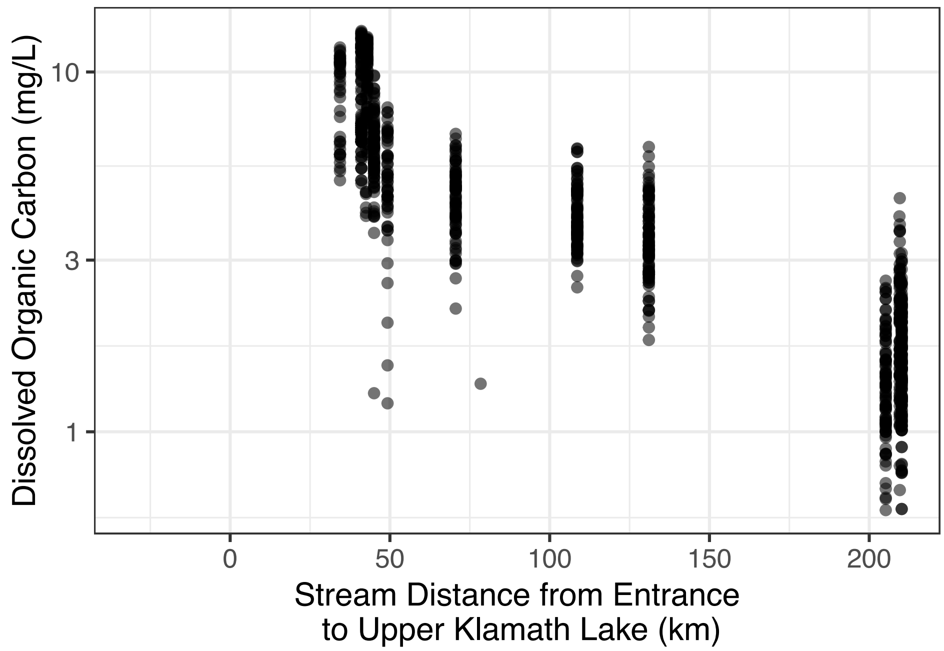

Figure 4: Organic Carbon concentration versus distance downstream.

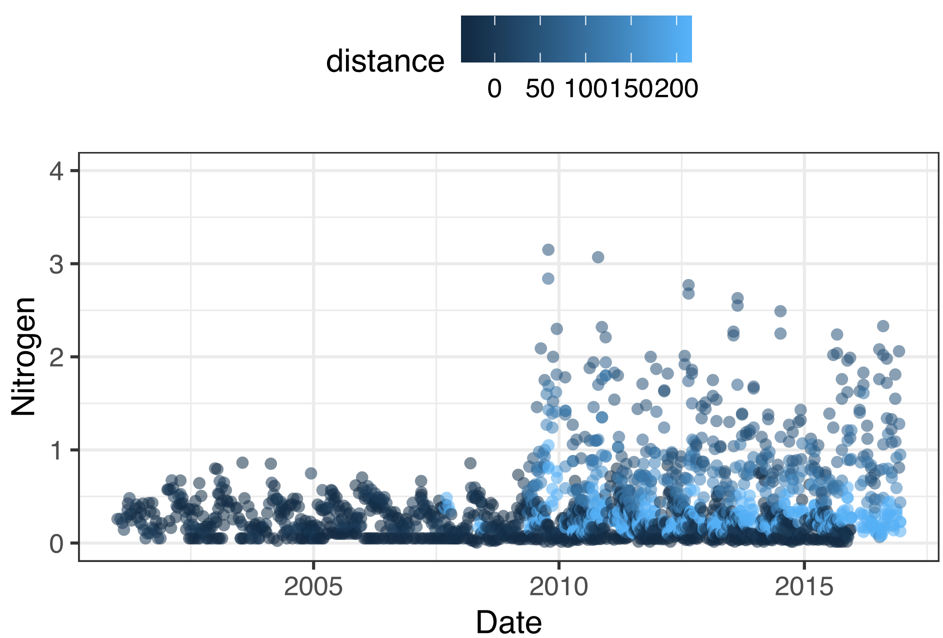

Figure 5: Nitrogen Concentration from 2000 through 2018, shaded by distance from the entrance to Upper Klamath Lake.

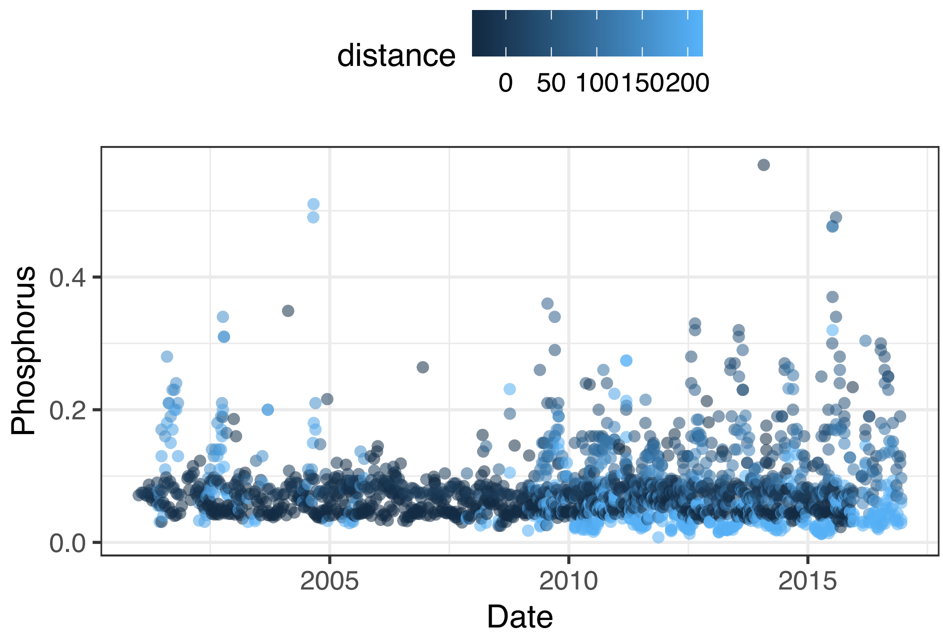

Figure 6: Phosphorus Concentration from 2000 through 2018, shaded by distance from the entrance to Upper Klamath Lake.

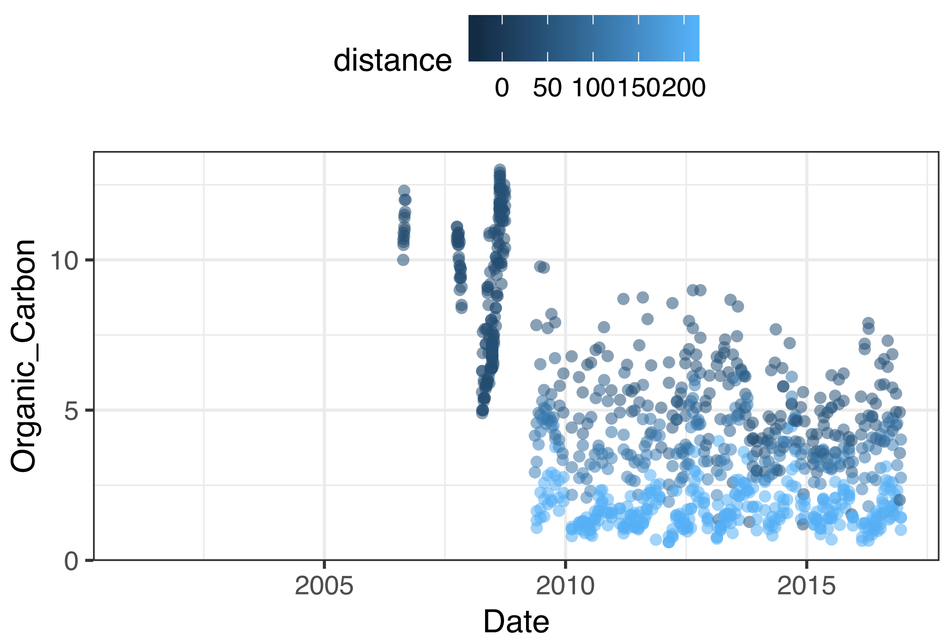

Figure 7: Organic Carbon Concentration from 2000 through 2018, shaded by distance from the entrance to Upper Klamath Lake.

4 Analysis

A fixed effects model was created for each nutrient, using distance downstream and sample date as independent variables with an interaction. This analysis was only performed using data from the sites downstream of Upper Klamath Falls, where there appears to be a trend in Figures 2, 3, and 4.

```r
## concentration as determined by date and distance downstream
lm_nitrogen <- lm(data = klamath_trend_sf, Nitrogen ~ distance*Date)
summary(lm_nitrogen)
```

```
##
## Call:
## lm(formula = Nitrogen ~ distance * Date, data = klamath_trend_sf)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.8149 -0.1745 -0.0221  0.1071  7.8950
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    2.021e+00  7.263e-01   2.783  0.00551 **
## distance      -8.408e-03  4.555e-03  -1.846  0.06524 .
## Date          -2.444e-05  4.583e-05  -0.533  0.59403
## distance:Date  1.009e-07  2.877e-07   0.351  0.72586
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4284 on 828 degrees of freedom
##   (1104 observations deleted due to missingness)
## Multiple R-squared:  0.5102, Adjusted R-squared:  0.5085
## F-statistic: 287.6 on 3 and 828 DF,  p-value: < 2.2e-16
```

```r
lm_nitrogen2 <- lm(data = klamath_trend_sf, Nitrogen ~ distance + Date)
summary(lm_nitrogen2)
```

```
##
## Call:
## lm(formula = Nitrogen ~ distance + Date, data = klamath_trend_sf)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.8216 -0.1736 -0.0221  0.1093  7.8881
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)  1.788e+00  2.908e-01   6.147 1.23e-09 ***
## distance    -6.812e-03  2.321e-04 -29.356  < 2e-16 ***
## Date        -9.690e-06  1.824e-05  -0.531    0.595
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4281 on 829 degrees of freedom
##   (1104 observations deleted due to missingness)
## Multiple R-squared:  0.5102, Adjusted R-squared:  0.509
## F-statistic: 431.7 on 2 and 829 DF,  p-value: < 2.2e-16
```

```
lm_nitrogen3 <- lm(data = klamath_trend_sf, Nitrogen ~ distance)
summary(lm_nitrogen3)
```

```
##
## Call:
## lm(formula = Nitrogen ~ distance, data = klamath_trend_sf)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.8312 -0.1745 -0.0236  0.1075  7.8808
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  1.6343816  0.0363723   44.94   <2e-16 ***
## distance    -0.0068048  0.0002315  -29.39   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.428 on 830 degrees of freedom
##   (1104 observations deleted due to missingness)
## Multiple R-squared:   0.51,  Adjusted R-squared:  0.5094
## F-statistic: 863.9 on 1 and 830 DF,  p-value: < 2.2e-16
```

Downstream distance has a statistically significant impact on nitrogen concentration in the Klamath River, but there has been no statistically significant time-related trend on nitrogen. ($F_{1,830}$=863.9, $p<0.001$, R^2=0.51). The model predicts a nitrogen concentration of 1.63 mg/L at Klamath Lake that decreases by 0.068 mg/L with every 10 kilometers traveled downstream.

```
## concentration as determined by date and distance downstream
lm_phosphorus <- lm(data = klamath_trend_sf, Phosphorus ~ distance*Date)
summary(lm_phosphorus)
```

```
##
## Call:
## lm(formula = Phosphorus ~ distance * Date, data = klamath_trend_sf)
##
```

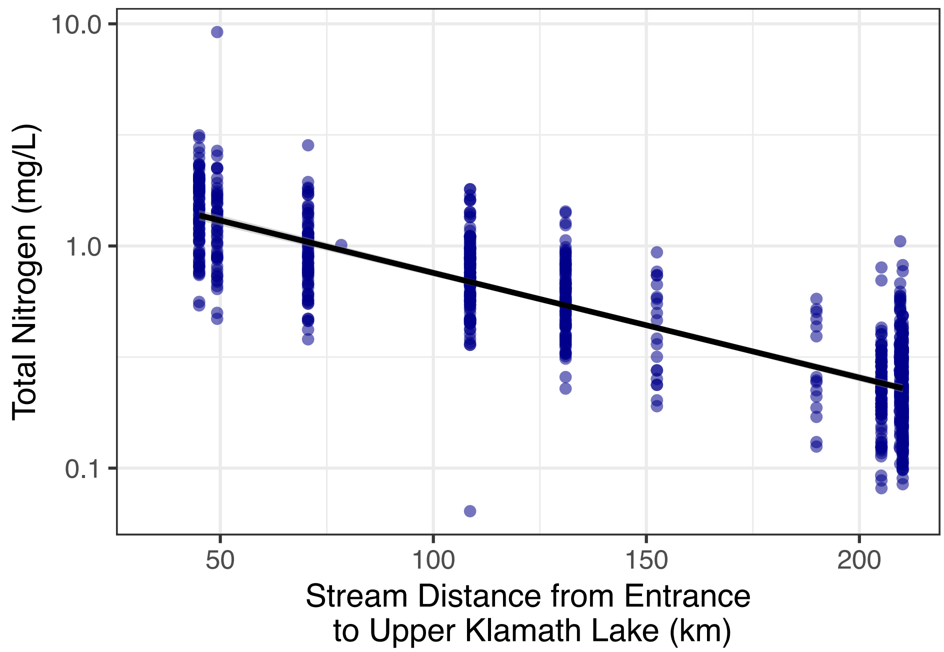

Figure 8: Nitrogen concentrations along the Klamath River with a trend line for the linear model shown in black.

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.11649 -0.03443 -0.01078  0.02081  0.40942
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    6.861e-02  7.802e-02   0.879  0.37938
## distance       9.750e-04  4.633e-04   2.105  0.03559 *
## Date           7.986e-06  4.966e-06   1.608  0.10820
## distance:Date -1.058e-07  2.955e-08  -3.581  0.00036 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05895 on 942 degrees of freedom
##   (990 observations deleted due to missingness)
## Multiple R-squared:  0.3472, Adjusted R-squared:  0.3451
## F-statistic:   167 on 3 and 942 DF,  p-value: < 2.2e-16
```

Unlike nitrogen, phosphorous concentration is affected by both distance downstream and time, and there is a significant interaction between time and distance downstream ($F_{3,942}$=167, $p<0.001$, R^2=0.35). However, the effect of date alone does not have a statistically significant impact on phosphorus concentration. This model predicts that phosphorous concentration will vary according to the following equation, beginning at Klamath Lake on January 1, 2000:

$$P = 0.0686 + 9.75 * 10^{-4} * distance + 7.99 * 10^{-6} * days - 1.06 * 10^{-7} * distance * days$$

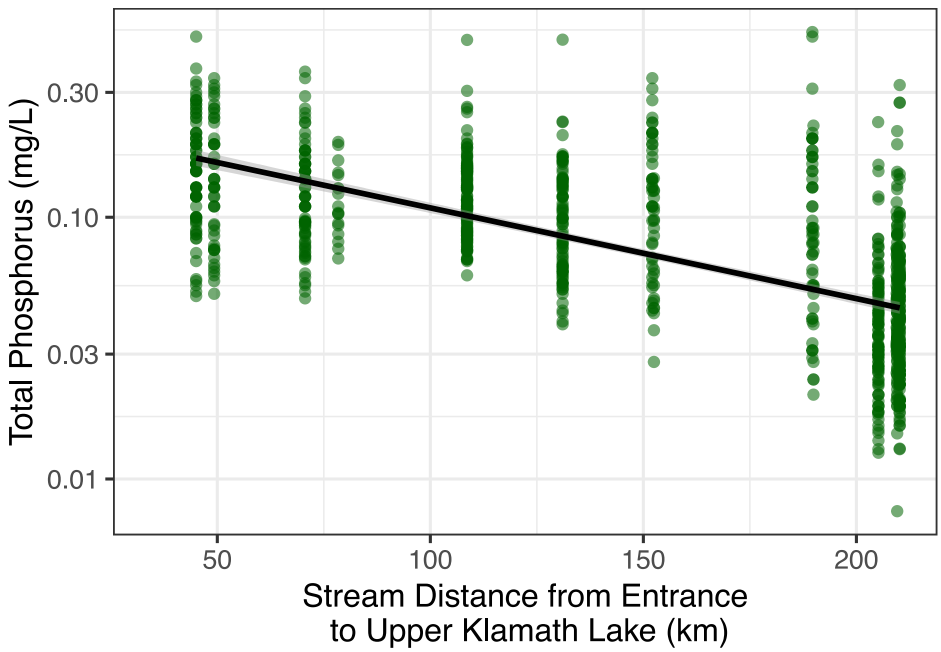

Figure 9: Phosphorus concentrations along the Klamath River with a trend line for the linear model shown in black.

```
## concentration as determined by date and distance downstream
lm_carbon <- lm(data = klamath_trend_sf, Organic_Carbon ~ distance*Date)
summary(lm_carbon)
```

```
##
## Call:
## lm(formula = Organic_Carbon ~ distance * Date, data = klamath_trend_sf)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -4.3890 -0.7198 -0.0842  0.7148  4.3099
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)    3.842e+01  1.200e+00   32.00   <2e-16 ***
## distance      -1.921e-01  1.026e-02  -18.72   <2e-16 ***
## Date          -1.977e-03  7.948e-05  -24.87   <2e-16 ***
## distance:Date  1.045e-05  6.610e-07   15.82   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.397 on 966 degrees of freedom
##   (966 observations deleted due to missingness)
## Multiple R-squared:  0.8016, Adjusted R-squared:  0.801
```

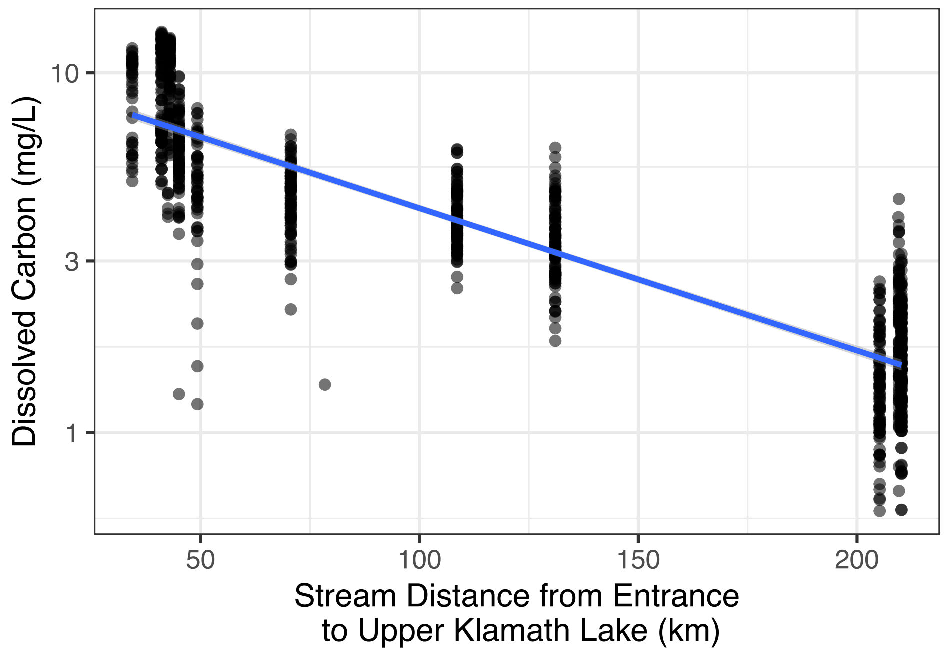

Figure 10: Phosphorus concentrations along the Klamath River with a trend line for the linear model shown in blue

```
## F-statistic:  1301 on 3 and 966 DF,  p-value: < 2.2e-16
```

Carbon concentration is significantly affected by distance downstream, data, and the interaction between distance and date ($F_{3,966}$=1301, p<0.001, R^2=0.80). The linear model predicts carbon will vary according to the following formula:

$$C = 38.4 - 0.192 * distance - 1.98 * 10^{-3} * days + 1.05 * 10^{-5} * distance * days$$

While none of the three sets of nutrient data precisely follow a normal distribution, they are quite consistent for environmental data. There is also an increase in variance for each of the three nutrients over distance and time, but this increase in variance was not accounted for.

5 Summary and Conclusions

While the statistical analyses performed indicated that the concentrations of all three nutrients vary with distance downstream and over time (with the exception of nitrogen), there are obviously other factors at play. There is a great variation in the concentrations of all three nutrients, and the models explain only a small part of this variation for phosphorus and nitrogen, according to their R^2 values. While it was hypothesized that nutrient concentrations would decrease downstream, nutrient levels were surprisingly low at and above Upper Klamath Lake, despite agricultural activity in the area. This may offer some corroborating evidence to the idea that nutrients are coming from both land use activities and the natural geology of the basin.

Phosphorus concentrations do not follow the same spatial or temporal trends as organic carbon or nitrogen, but this reason is unknown. The temporal trend in nutrient levels is also of low magnitude. While the slowly decreasing trend is positive for the river's water quality, this may indicate that regulations may be needed to more quickly decrease nutrient loading to the river.