

TECHNICAL REPORT

DRAFT: Evaluation of Algal bloom potential for the Caloosahatchee River Estuary

Paul Julian^{1,2}

¹Sanibel-Captiva Conservation Foundation, PO Box 839, Sanibel, FL 33957

²Conservancy of Southwest Florida, 1495 Smith Preserve Way, Naples, FL 34102

Document History

Compiled June 3, 2021

Open Source Information

https://github.com/SwampThingPaul/LOSOM_AlgalBloom/

DOI

<https://doi.org/10.5281/zenodo.4876353>

1. Objective

The draft Northern Estuaries Algal Bloom Risk Assessment performance measure for Lake Okeechobee System Operating Manual (LOSOM) proposes to evaluate changes in discharges from Lake Okeechobee to the Caloosahatchee River Estuary (CRE) to assess the risk of algal bloom development. The premise of this performance measure is that releasing freshwater with sufficient algal biomass to the estuaries during the summer (May/June - August) favors freshwater algal blooms in the estuaries by transport of algal biomass, expansion of the freshwater habitat suitable for algal growth, and subsidizing available nutrients. This concept has been demonstrated for the St Lucie Estuary (Phlips et al. 2020) but not for the CRE. Therefore, this preliminary evaluation will explore the potential for algal bloom transport from Lake Okeechobee to the CRE and the role of discharges on algal bloom formation at the CRE headwaters.

2. Methods

2.1. Data Source

Chlorophyll-a data were retrieved from the South Florida Water Management District (SFWMD) online database (DBHYDRO; www.sfwmd.gov/dbhydro) for sites defined as Lake Okeechobee western littoral zone as defined by Walker (2020) and S-79 (Table 1) for S-77 and S-79. Only positive discharge values were considered in this analysis, therefore discharge values reported less than zero were replaced with zero. Water quality data were screened based on laboratory qualifier codes, consistent with Florida

Department of Environmental Protection’s quality assurance rule (Florida Administrative Code 2008). Any datum associated with a fatal qualifier indicating a potential data quality problem was removed from the analysis. Additional considerations in the handling of analytical data were the accuracy and sensitivity of the laboratory method used. For purposes of data analysis and summary statistics, data reported as less than the method detection limit (MDL) were assigned one-half the MDL.

Table 1.: Monitoring Locations used in this evaluation with regions defined. A map of monitoring locations is presented in Fig B1.

DataType	Site	Region
Water Quality	S79	S-79 (CRE)
	CES01	S-79 (CRE)
	S77	Littoral West (Lake Okeechobee)
	PALMOUT	Littoral West (Lake Okeechobee)
	PLN2OUT	Littoral West (Lake Okeechobee)
	TREEOUT	Littoral West (Lake Okeechobee)
Discharge	S77	From Lake (DBKEY:15635)
	S79	To Estuary (DBKEY: 00865)

2.2. Data Analysis

Daily discharge data were aggregated by month for each discharge location. Basin discharge was estimated by taking the difference from S77 and S79. If the difference between the estuary discharge (S79) and lake (S77) were less than zero then basin discharge was set to zero. Lake discharges to the estuary were determined based on the relative discharge volumes of S77 and S79. If S77 discharges were greater than S79 for any given month it was assumed that the discharge volume at S79 was from the lake (i.e. no basin flow), otherwise, the observed discharge volumes for S77 were characterized as lake water. Monthly discharge volumes were compared between high (June – August) and low (September – May) algal bloom risk periods for S77 and S79 using the Kruskal-Wallis rank-sum test. The algal bloom risk periods were based on analyses by Walker (2020), where generally chlorophyll-a concentrations and therefore algal bloom potential are greatest during May through August period. The period was adjusted for the CRE as May is typically the driest month and discharges during this period do not reflect a high risk of algal transport to the estuary. Monthly mean chlorophyll-a data were estimated for each region (i.e. CRE and Littoral West), for Lake Okeechobee Littoral West all sites were average together for any given month. Monthly mean chlorophyll-a concentrations were compared between regions using Kruskal-Wallis rank-sum test. Cross-correlation function estimation was used to evaluate potential lagged correlations between Lake Okeechobee (littoral west) and S-79 (CRE) chlorophyll concentrations where CRE chlorophyll values remained fixed and Lake Okeechobee chlorophyll values were lagged. Logistic regression was used to evaluate the potential of bloom occurrence relative to lake discharges. To evaluate

the complexity of algal bloom occurrence, two separate models were evaluated. The first, more simplistic logistic model compared algal bloom occurrence at S79 to lake discharges, meanwhile the second more complex model included water temperature, Lake Okeechobee discharges and C-43 basin discharges. Potential algal blooms at S79 were identified if monthly mean chlorophyll-a concentrations were greater than $20 \mu\text{g L}^{-1}$.

All plots were generated in base R and all tables were formatted using “flextable” R-package (Gohel 2021). All statistical operations were performed using R (Ver 4.0.4, R Foundation for Statistical Computing, Vienna Austria). Unless otherwise stated, all statistical operations were performed using the base R library. The critical level of significance was set at $\alpha = 0.05$. All R code for the analyses outline above can be found in **Appendix A** and source files for the analysis and this report have been posted to https://github.com/SwampThingPaul/LOSOM_AlgalBloom/.

3. Results

Monthly discharge from the S77 and S79 was significantly different between high and low-risk bloom periods (S77: $\chi^2=21.7$, $\text{df}=1$, $\rho<0.01$; S79: $\chi^2=20.97$, $\text{df}=1$, $\rho<0.01$). Generally, discharges from the S-77 during the high bloom risk period are lower than discharges observed during the high risk (Fig 1). However, discharges from the S-79 are typically higher during the high bloom risk period relative to the low bloom risk period (Fig 2). Additionally, monthly discharge volumes at S-79 have a strong seasonal signature with discharges beginning to increase in May to June and declining in September to October periods (Fig 2). Meanwhile, the seasonality of discharges at S77 is much less apparent (Fig 2).

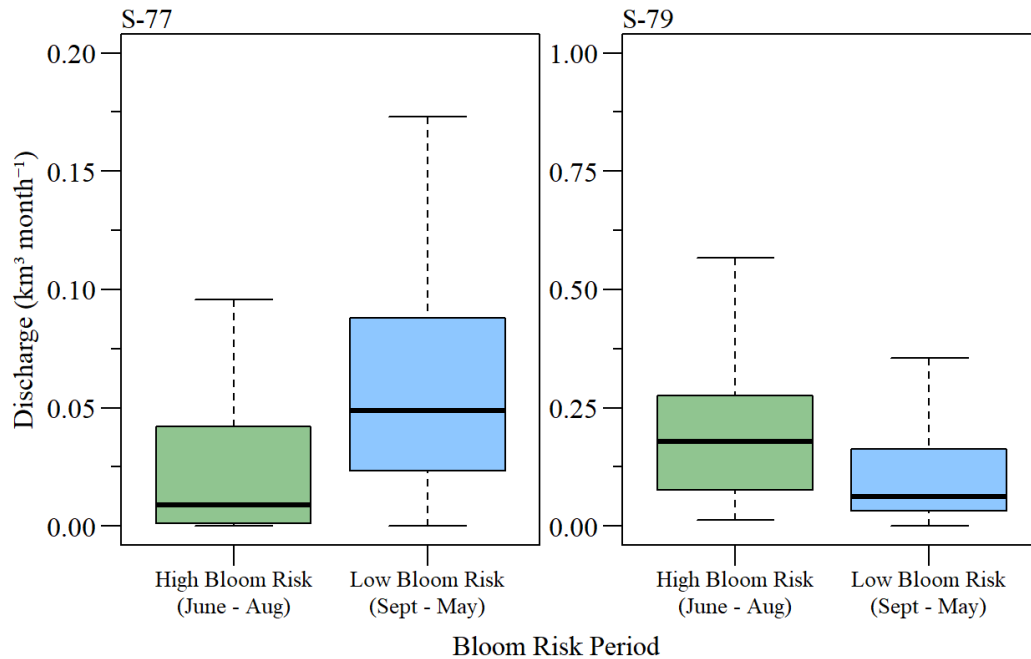


Figure 1. Comparison of monthly discharge volumes between bloom risk periods for S77 (Left) and S79 (right).

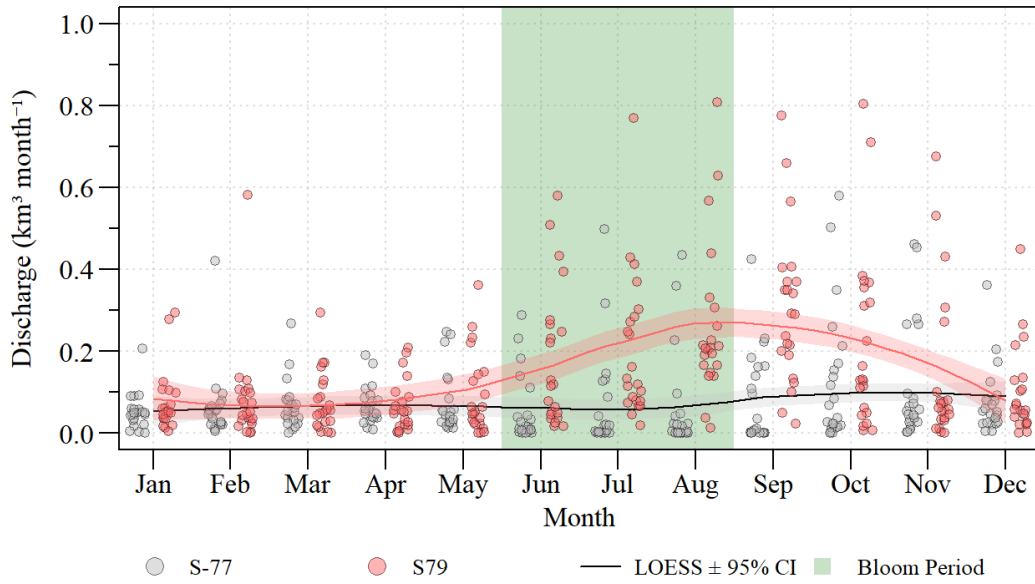


Figure 2. Monthly discharge volumes for S77 and S79 during the May 1999 to May 2021 period of record with locally estimated scatterplot smoothing (LOESS) trend and 95% confidence interval.

Monthly mean chlorophyll-a concentrations were significantly greater for the Littoral West region relative to CRE (χ^2 100, $df=1$, $\rho<0.01$; Fig 3). However, during the June to August, high bloom period chlorophyll-a concentrations are not significantly different between regions (χ^2 0.4, $df=1$, $\rho=0.54$; Fig 3). However, the cross-correlation between monthly chlorophyll-a concentration observed at CRE and that of the Littoral West are relatively low with no correlations across the 15 month lag period (Fig 4). To reduce the potential variability in the data, comparison of S79 and S77 chlorophyll concentrations were also conducted with similar results.

The logistic model to evaluate algal bloom potential at S-79 relative to lake discharges indicated a significant decrease in log-odds ratio for algal bloom (i.e. chlorophyll $>20 \mu\text{g L}^{-1}$) as discharges from the lake increases (Table 2 and Fig 5). However, the model has a low degree of fit (McFadden's R^2 0.07), therefore any algal bloom mitigation strategy developed based on this analysis should be used with caution. Meanwhile, when water temperature, lake and basin discharges are considered, model fit improves (Table 3) but not to the level to sufficiently predict algal bloom occurrence without a high level of uncertainty.

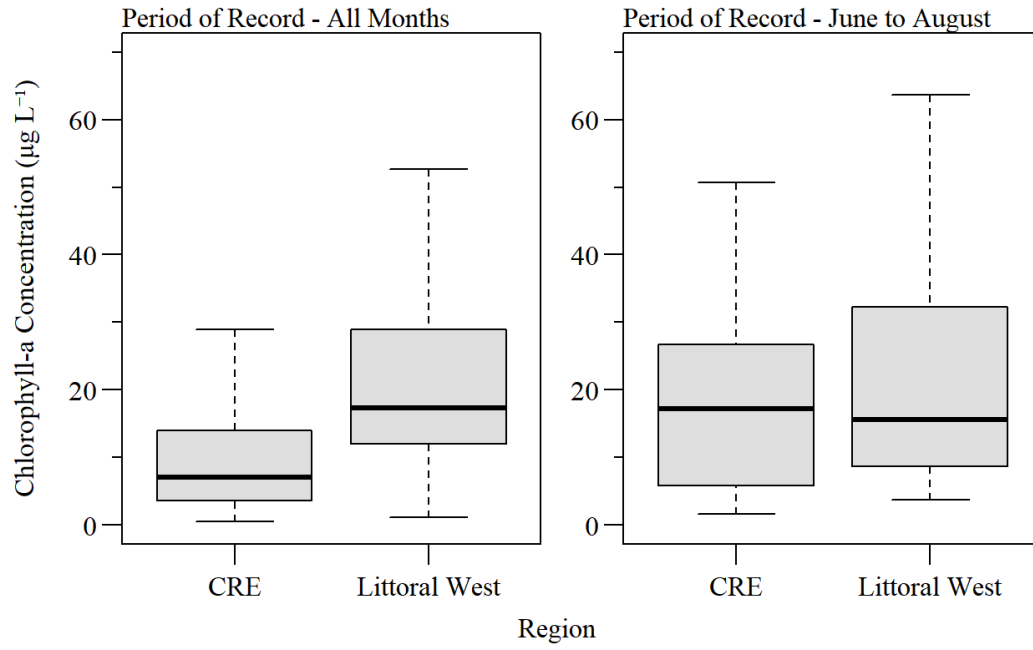


Figure 3. Monthly mean chlorophyll-a concentration between regions for all months (left) and algal bloom period (right).

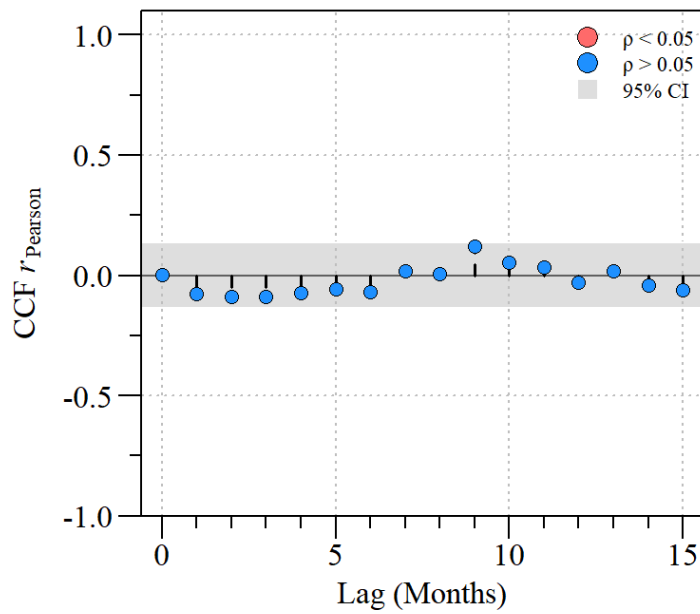


Figure 4. Cross-correlation function results comparing monthly mean chlorophyll-a concentration between CRE and lagged Littoral West.

Table 2.: Simple logistic model summary comparing algal bloom potential to lake discharges.

	Estimate	Standard Error	z-value	p-value
(Intercept)	-1.038	0.246	-4.216	0.000
Lake Discharge	-12.165	4.860	-2.503	0.012

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 183.1 on 201 degrees of freedom

Residual deviance: 171 on 200 degrees of freedom

AIC: 175; BIC: 181.6

McFadden's Pseudo R^2 : 0.07

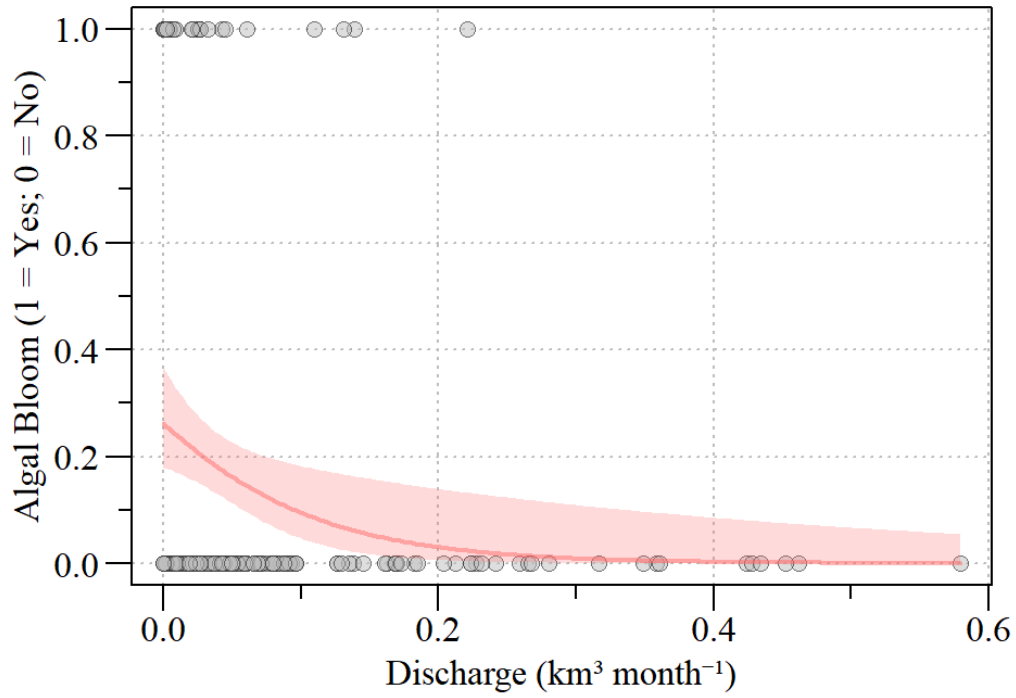


Figure 5. Algal bloom occurrence at S-79 compared to lake discharges with fitted logistic model with 95% confidence interval.

Table 3.: Logistic model summary comparing algal bloom potential to water temperature, C-43 basin and lake discharges.

	Estimate	Standard Error	z-value	p-value
(Intercept)	-9.625	2.175	-4.424	0.000
Water Temp	0.326	0.080	4.088	0.000
Lake Discharge	-7.384	3.884	-1.901	0.057
Basin Discharge	-4.941	2.681	-1.843	0.065

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 183.1 on 201 degrees of freedom

Residual deviance: 148.2 on 198 degrees of freedom

AIC: 156.2; BIC: 169.5

McFadden's Pseudo R^2 : 0.19

Appendix A. Reproducible R-Code

```
## Libraries
#devtools::install_github("SwampThingPaul/AnalystHelper")
library(AnalystHelper);

library(plyr)
library(reshape2)
library(zoo)

# -----
dates=date.fun(c("1999-05-01","2021-04-30"))

# Discharge -----
q.dbkeys=data.frame(SITE=c("S79","S78","S77"),
                    DBKEY=c("00865","DJ236","15635"))
q.cre.dat=data.frame()
for(i in 1:nrow(q.dbkeys)){
  tmp=DBHYDRO_daily(dates[1],dates[2],q.dbkeys$DBKEY[i])
  tmp$DBKEY=as.character(q.dbkeys$DBKEY[i])
  q.cre.dat=rbind(q.cre.dat,tmp)
  print(i)
}
q.cre.dat=merge(q.cre.dat,q.dbkeys,"DBKEY")
q.cre.dat$Date.EST=date.fun(q.cre.dat$Date)

q.cre.dat.xtab=dcast(q.cre.dat,Date.EST~SITE,
                    value.var="Data.Value",mean)
q.cre.dat.xtab$month=as.numeric(format(q.cre.dat.xtab$Date,"%m"))
q.cre.dat.xtab$CY=as.numeric(format(q.cre.dat.xtab$Date,"%Y"))
q.cre.dat.xtab$monCY=with(q.cre.dat.xtab,
                        date.fun(paste(CY,month,"01",sep="-")))
q.cre.dat.xtab$WY=WY(q.cre.dat.xtab$Date.EST)

#Positive discharge only
q.cre.dat.xtab$S77=with(q.cre.dat.xtab,ifelse(S77<0,0,S77))
q.cre.dat.xtab$S79=with(q.cre.dat.xtab,ifelse(S79<0,0,S79))
q.cre.dat.xtab$C43=with(q.cre.dat.xtab,ifelse(S79<S77,0,S79-S77))
q.cre.dat.xtab$Lake=with(q.cre.dat.xtab,ifelse(S77<S79,0,S77-S79))

q.cre.dat.xtab.mon=ddply(q.cre.dat.xtab,c("monCY","month",'CY'),
                        summarise,
                        S77=sum(cfs.to.km3d(S77),na.rm=T),
                        S79=sum(cfs.to.km3d(S79),na.rm=T))
q.cre.dat.xtab.mon$C43=with(q.cre.dat.xtab.mon,
                        ifelse(S79<S77,0,S79-S77))
q.cre.dat.xtab.mon$Lake=with(q.cre.dat.xtab.mon,
                        ifelse(S77>S79,S79,S77))
q.cre.dat.xtab.mon$Period=with(q.cre.dat.xtab.mon,
```



```

        ifelse(month%in%c(6:8),
              "Bloom","Non-Bloom"))

## Comparison of monthly discharge
## between bloom/non-bloom periods

boxplot(S77~Period,q.cre.dat.xtab.mon)
kruskal.test(S77~Period,q.cre.dat.xtab.mon)

boxplot(S79~Period,q.cre.dat.xtab.mon)
kruskal.test(S79~Period,q.cre.dat.xtab.mon)

# WQ Data -----

params=data.frame(Test.Number=c(61,179,7),
                  param=c("Chla","Chla","Temp"))

wq.sites=data.frame(Station.ID=
                    c("S79","CES01","S77",
                      "PALMOUT","PLN2OUT","TREEOUT"),
                    Region=c(rep("CRE",2),rep("LitWest",4)))

wq.dat=DBHYDRO_WQ(dates[1],dates[2],
                  wq.sites$Station.ID,params$Test.Number)
wq.dat=merge(wq.dat,params,"Test.Number")
unique(wq.dat$Collection.Method)
wq.dat=subset(wq.dat,Collection.Method=="G")

wq.dat.xtab=dcast(wq.dat,Station.ID+Date.EST~param,
                 value.var="HalfMDL",mean)
wq.dat.xtab$WY=WY(wq.dat.xtab$Date.EST)
wq.dat.xtab$month=as.numeric(format(wq.dat.xtab$Date.EST,"%m"))
wq.dat.xtab$CY=as.numeric(format(wq.dat.xtab$Date.EST,"%Y"))
wq.dat.xtab$monCY=with(wq.dat.xtab,
                      date.fun(paste(CY,month,"01",sep="-")))
wq.dat.xtab=merge(wq.dat.xtab,wq.sites,"Station.ID")

## Regional Comparison
Chla.region=ddply(wq.dat.xtab,c("monCY","month","Region"),
                 summarise(mean.val=mean(Chla,na.rm=T))
Chla.region$bloom=ifelse(Chla.region$mean.val>20,1,0)

kruskal.test(mean.val~Region,Chla.region)
kruskal.test(mean.val~Region,subset(Chla.region,month%in%c(6:8)))
plot(mean.val~month,subset(Chla.region,Region=="CRE"))
plot(mean.val~month,subset(Chla.region,Region=="LitWest"))

## Cross Correlation
Chla.region.xtab=dcast(wq.dat.xtab,monCY~Region,

```

```

        value.var = "Chla",
        fun.aggregate=function(x) mean(x,na.rm=T))
plot(CRE~LitWest,Chla.region.xtab)

Chla.region.xtab=merge(Chla.region.xtab,q.cre.dat.xtab.mon,"monCY")

layout(matrix(1:4,1,4,byrow=T))
for(h in 0:3){
  tmp.dat=data.frame(CRE.chla=lag(as.zoo(Chla.region.xtab$CRE),
                                0,na.pad=T),
                    Lake.chla=lag(as.zoo(Chla.region.xtab$LitWest),
                                -h,na.pad=T),
                    Lake=Chla.region.xtab$Lake)
  plot(CRE.chla~Lake.chla,tmp.dat,pch=21,"grey")
}

# ccf function in R
with(subset(Chla.region.xtab,is.na(CRE)==F&is.na(LitWest)==F),
     ccf(CRE,LitWest,type="correlation"))

ccf.chla.rslt=data.frame()
for(h in 0:15){
  lagged=lag(as.zoo(Chla.region.xtab$LitWest),-h,na.pad=T)
  tmp.dat=as.zoo(Chla.region.xtab$CRE)
  stat=with(data.frame(lag=lagged,dat=tmp.dat),
            cor.test(lag,dat,method="pearson"))
  ccf.chla.rslt=rbind(ccf.chla.rslt,
                    data.frame(lag=h,
                                estimate=as.numeric(stat$estimate),
                                pval=stat$p.value))
}
ccf.chla.rslt

## Logistic Regression
# Simple Model
# data prep
bloom.region.xtab=dcast(Chla.region,monCY~Region,
                        value.var = "bloom",
                        fun.aggregate=function(x) mean(x,na.rm=T))
bloom.region.xtab$bothbloom=with(bloom.region.xtab,
                                ifelse(CRE==1&LitWest==1,1,0))
bloom.region.xtab2=merge(bloom.region.xtab,
                        q.cre.dat.xtab.mon,
                        "monCY")

# Data Explore
plot(CRE~Lake,bloom.region.xtab2)
plot(CRE~S77,bloom.region.xtab2)

```

```

plot(bothbloom~Lake,bloom.region.xtab2)
plot(bothbloom~S77,bloom.region.xtab2)

mod=glm(CRE~Lake,bloom.region.xtab2,family="binomial")
summary(mod)
confint(mod)

# diagnostics
plot(1:length(rstandard(mod)),rstandard(mod))
plot(mod,which=4,id.n=5);#identify the top five largest values

# McFadden's r2
glm.null.loglike=logLik(glm(CRE~1,
                             bloom.region.xtab2,
                             family="binomial"))[1]
mcR2=1-logLik(mod)[1]/glm.null.loglike
mcR2

#classification rate
prop.table(table(bloom.region.xtab2$CRE,
                  predict(mod,
                           data.frame(Lake=bloom.region.xtab2$Lake),
                           type="response")>0.5))

# Multi-parameter Model
# Data prep
Chla.region2=ddply(wq.dat.xtab,c("monCY","month","Region"),
                   summarise,
                   Chla=mean(Chla,na.rm=T),
                   Temp=mean(Temp,na.rm=T))
Chla.region2$bloom=ifelse(Chla.region2$Chla>20,1,0)

bloom.region.xtab2=merge(subset(Chla.region2,Region=="CRE"),
                          q.cre.dat.xtab.mon,c("monCY","month"))

# Data Explore
plot(bloom~Temp,bloom.region.xtab2)
plot(bloom~Lake,bloom.region.xtab2)
plot(bloom~C43,bloom.region.xtab2)

mod2=glm(bloom~Temp+Lake+C43,bloom.region.xtab2,family="binomial")
summary(mod2)
confint(mod2)

# diagnostics
plot(1:length(rstandard(mod2)),rstandard(mod2))
plot(mod2,which=4,id.n=5);#identify the top five largest values

# McFadden's r2

```

```
glm.null.loglike=logLik(glm(bloom~1,  
                           bloom.region.xtab2,  
                           family="binomial"))[1]  
mcR2=1-logLik(mod2)[1]/glm.null.loglike  
mcR2
```

Appendix B. Sample Map



Figure B1. Map of monitoring locations presented in Table 1.

References

- Florida Administrative Code. 2008. “Chapter 62-160 Quality Assurance.” .
- Gohel, D. 2021. “flextable: Functions for Tabular Reporting.” CRAN R-Project. <http://CRAN.R-project.org/package=flextable>.
- Phlips, Edward J., Susan Badylak, Natalie G. Nelson, and Karl E. Havens. 2020. “Hurricanes, El Niño and harmful algal blooms in two sub-tropical Florida estuaries: Direct and indirect impacts.” *Scientific Reports* 10 (1): 1910. <http://www.nature.com/articles/s41598-020-58771-4>.
- Walker, W. 2020. *DRAFT Chlorophyll-a Models for LOSOM Applications*. Technical Report. US Department of the Interior.