table lm fcr

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

#install.packages("MuMIn")  
#install.packages("rsq")  
library(MuMIn)  
library(knitr)  
library(rsq)  
library(tidyverse)

## -- Attaching packages -------------------------------------------------------------------------- tidyverse 1.2.1 --

## v ggplot2 3.0.0 v purrr 0.2.5  
## v tibble 1.4.2 v dplyr 0.7.6  
## v tidyr 0.8.1 v stringr 1.3.1  
## v readr 1.1.1 v forcats 0.3.0

## -- Conflicts ----------------------------------------------------------------------------- tidyverse\_conflicts() --  
## x dplyr::filter() masks stats::filter()  
## x dplyr::lag() masks stats::lag()

#all data  
data <- read.csv("C:/Users/wwoel/Dropbox/FCR\_TimeSeries/FCR/model\_transformed\_2013\_2016.csv")  
  
########################################################################################################################################################  
  
varall <- read.csv("C:/Users/wwoel/Dropbox/FCR\_TimeSeries/FCR/correlation matrices/2013\_2016\_selected\_correlations.csv")  
varall$Date <- NA  
varall <- varall%>%select(Date, everything())  
# get rid of the variables that have NAs for 2013 and/or 2014  
varall <- varall%>%select(-(Kd:TN\_inf\_log))  
dataall <- data[,colnames(varall)]  
# subset to get rid of NA's at beginning? this truncates dataset to after June20  
dataall$Date <- as.Date(dataall$Date)  
dataall <- dataall[dataall$Date>"2013-06-20",]  
  
  
# build a global model for all years  
modelall <- glm(Chla\_sqrt~Chla\_ARlag1\_sqrt+Turb\_NTU\_log+NH4\_inf\_log+flow\_min+Temp\_inf\_min+RelHum\_max\_log+Rain\_sum\_log+  
 WindSpeed\_max\_log+ShortWave\_mean, data = dataall, family = gaussian, na.action = "na.fail" )  
glmall <- dredge(modelall, rank = "AICc", fixed = "Chla\_ARlag1\_sqrt")

## Fixed terms are "Chla\_ARlag1\_sqrt" and "(Intercept)"

selectall <- subset(glmall, delta<2 )  
selectall <- subset(selectall, selectall$df<8)  
  
# build individual models for those within 2 AICc units  
mod1.all <- glm(Chla\_sqrt~Chla\_ARlag1\_sqrt+flow\_min+NH4\_inf\_log+ShortWave\_mean+Turb\_NTU\_log,  
 data = dataall, family = gaussian, na.action = "na.fail")  
pred.1.all <- predict(mod1.all, newdata=dataall)  
  
mod2.all <- glm(Chla\_sqrt~Chla\_ARlag1\_sqrt+flow\_min + ShortWave\_mean+Turb\_NTU\_log,   
 data = dataall, family = gaussian, na.action = "na.fail")  
pred.2.all <- predict(mod2.all, newdata=dataall)  
  
mod3.all <- glm(Chla\_sqrt~Chla\_ARlag1\_sqrt+flow\_min+RelHum\_max\_log+Turb\_NTU\_log+ShortWave\_mean, data = dataall, family = gaussian, na.action = "na.fail" )  
pred.3.all <- predict(mod3.all, newdata=dataall)  
  
mod4.all <- glm(Chla\_sqrt~Chla\_ARlag1\_sqrt+Turb\_NTU\_log+flow\_min+Temp\_inf\_min+ShortWave\_mean, data = dataall, family = gaussian, na.action = "na.fail" )  
pred.4.all <- predict(mod4.all, newdata=dataall)  
  
mod5.all <- glm(Chla\_sqrt~Chla\_ARlag1\_sqrt+Turb\_NTU\_log+flow\_min+Rain\_sum\_log+ShortWave\_mean, data = dataall, family = gaussian, na.action = "na.fail" )  
pred.5.all <- predict(mod5.all, newdata=dataall)  
  
mod6.all <- glm(Chla\_sqrt~Chla\_ARlag1\_sqrt+Turb\_NTU\_log+flow\_min+RelHum\_max\_log, data = dataall, family = gaussian, na.action = "na.fail" )  
pred.6.all <- predict(mod6.all, newdata=dataall)

#2013-2016 dataset model diagnostics table  
# make a table of model diagnostics and descriptions  
table1316 = array(NA,dim=c(6,4))  
row.names(table1316) = c('Mod 1','Mod 2','Mod 3', 'Mod 4', 'Mod 5', 'Mod 6')  
colnames(table1316) = c("Model Equation", "AICc", "R2", 'df')  
table1316[1,1] = "0.31Chla(t-1) - 5.01flow\_min - 0.002ShortWave\_mean + 0.15Turbidity\_log + 0.14NH4\_inf\_log + 1.35"  
table1316[1,2] = round(selectall[1,13], digits = 2)  
table1316[1,3] = round((rsq(mod1.all, type = 'sse')), digits = 3)  
table1316[1,4] = selectall[1,11]  
table1316[2,1] = "0.30Chla(t-1) - 4.55flow\_min - 0.002ShortWave\_mean + 0.13Turbidity\_log + 1.60"  
table1316[2,2] = round(selectall[2,13], digits = 2)  
table1316[2,3] = round((rsq(mod2.all, type = 'sse')), digits = 3)  
table1316[2,4] = selectall[2,11]  
table1316[3,1] = "0.32Chla(t-1) - 4.71flow\_min - 0.001ShortWave\_mean + 0.12Turbidity\_log + 0.70RelHumimidty\_max\_log - 1.67"  
table1316[3,2] = round(selectall[3,13], digits = 2)  
table1316[3,3] = round((rsq(mod3.all, type = 'sse')), digits = 3)  
table1316[3,4] = selectall[3,11]  
table1316[4,1] = "0.32Chla(t-1) - 4.33flow\_min - 0.002ShortWave\_mean + 0.14Turbidity\_log + 0.019Temp\_inflow\_min + 1.3"  
table1316[4,2] = round(selectall[4,13], digits = 2)  
table1316[4,3] = round((rsq(mod4.all, type = 'sse')), digits = 3)  
table1316[4,4] = selectall[4,11]  
table1316[5,1] = "0.32Chla(t-1) - 4.49flow\_min - 0.002ShortWave\_mean + 0.13Turbidity\_log + 0.014Rain\_sum\_log + 1.62"  
table1316[5,2] = round(selectall[5,13], digits = 2)  
table1316[5,3] = round((rsq(mod5.all, type = 'sse')), digits = 3)  
table1316[5,4] = selectall[5,11]  
table1316[6,1] = "0.35Chla(t-1) - 5.26flow\_min + 0.12Turbidity\_log + 1.09RelHumidity\_max\_log - 3.74"  
table1316[6,2] = round(selectall[6,13], digits = 2)  
table1316[6,3] = round((rsq(mod6.all, type = 'sse')), digits = 3)  
table1316[6,4] = selectall[6,11]  
  
kable(table1316)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Model Equation | AICc | R2 | df |
| Mod 1 | 0.31Chla(t-1) - 5.01flow\_min - 0.002ShortWave\_mean + 0.15Turbidity\_log + 0.14NH4\_inf\_log + 1.35 | 127.38 | 0.447 | 7 |
| Mod 2 | 0.30Chla(t-1) - 4.55flow\_min - 0.002ShortWave\_mean + 0.13Turbidity\_log + 1.60 | 127.97 | 0.431 | 6 |
| Mod 3 | 0.32Chla(t-1) - 4.71flow\_min - 0.001ShortWave\_mean + 0.12Turbidity\_log + 0.70RelHumimidty\_max\_log - 1.67 | 128.18 | 0.442 | 7 |
| Mod 4 | 0.32Chla(t-1) - 4.33flow\_min - 0.002ShortWave\_mean + 0.14Turbidity\_log + 0.019Temp\_inflow\_min + 1.3 | 128.7 | 0.439 | 7 |
| Mod 5 | 0.32Chla(t-1) - 4.49flow\_min - 0.002ShortWave\_mean + 0.13Turbidity\_log + 0.014Rain\_sum\_log + 1.62 | 129 | 0.438 | 7 |
| Mod 6 | 0.35Chla(t-1) - 5.26flow\_min + 0.12Turbidity\_log + 1.09RelHumidity\_max\_log - 3.74 | 129.21 | 0.424 | 6 |

################################# 2013 ########################################################################################################33  
  
var13 <- read.csv("C:/Users/wwoel/Dropbox/FCR\_TimeSeries/FCR/correlation matrices/2013\_selected\_correlations.csv")  
var13$Date <- NA  
var13 <- var13%>%select(Date, everything())  
data13 <- data[,colnames(var13)]  
data13$Date <- as.Date(data13$Date)  
# subset to get rid of NA's at beginning?  
data13 <- data13[data13$Date>"2013-06-20" & data13$Date<"2014-01-01",]  
  
# build a global model for 2013  
model13 <- glm(Chla\_sqrt~Chla\_ARlag1\_sqrt +Turb\_NTU\_log+NO3NO2\_log+SRP\_log+mean\_flow+Temp\_inf\_max+Rain\_sum\_log+WindSpeed\_mean\_log  
 +ShortWave\_mean, data = data13, family = gaussian, na.action = "na.fail" )  
glm13 <- dredge(model13, rank = "AICc", fixed = "Chla\_ARlag1\_sqrt")

## Fixed terms are "Chla\_ARlag1\_sqrt" and "(Intercept)"

select13 <- subset(glm13, delta<2 )  
  
# models <2 units different AICc values have been selected  
# now build those models individually and run summary statistics  
mod1 <- glm(Chla\_sqrt~Chla\_ARlag1\_sqrt+mean\_flow+ShortWave\_mean+Turb\_NTU\_log, data = data13, family = gaussian, na.action = na.fail)  
pred.1 <- predict(mod1, newdata=data13)  
mod2 <- glm(Chla\_sqrt~Chla\_ARlag1\_sqrt+mean\_flow+NO3NO2\_log+ShortWave\_mean+Turb\_NTU\_log, data = data13,   
 family = gaussian, na.action = "na.fail")  
pred.2 <- predict(mod2, newdata=data13)  
mod3 <- glm(Chla\_sqrt~Chla\_ARlag1\_sqrt+mean\_flow+Rain\_sum\_log+ShortWave\_mean+Turb\_NTU\_log,data = data13, family = gaussian, na.action = na.fail)  
pred.3 <- predict(mod3, newdata=data13)  
mod4 <- glm(Chla\_sqrt~Chla\_ARlag1\_sqrt+mean\_flow+ShortWave\_mean+Temp\_inf\_max+Turb\_NTU\_log, data = data13, family = gaussian, na.action = "na.fail")  
pred.4 <- predict(mod4, newdata=data13)  
mod5 <- glm(Chla\_sqrt~Chla\_ARlag1\_sqrt+mean\_flow+NO3NO2\_log+Rain\_sum\_log+ShortWave\_mean+Turb\_NTU\_log, data = data13, family = gaussian, na.action = "na.fail")  
pred.5 <- predict(mod5, newdata=data13)  
  
  
# now use the 2013 model for the entire dataset  
data13.all <- data[,colnames(var13)]  
data13.all$Date <- as.Date(data13.all$Date)  
# subset to get rid of NA's at beginning?  
data13.all <- data13.all[data13.all$Date>"2013-06-20",]  
  
model13.all <- glm(Chla\_sqrt~Chla\_ARlag1\_sqrt +Turb\_NTU\_log+NO3NO2\_log+SRP\_log+mean\_flow+Temp\_inf\_max+Rain\_sum\_log+WindSpeed\_mean\_log  
 +ShortWave\_mean, data = data13.all, family = gaussian, na.action = "na.fail" )  
glm13.all <- dredge(model13.all, rank = "AICc", fixed = "Chla\_ARlag1\_sqrt")

## Fixed terms are "Chla\_ARlag1\_sqrt" and "(Intercept)"

select13.all <- subset(glm13.all, delta<2 )  
  
  
# models <2 units different AICc values have been selected  
# now build those models individually and run summary statistics  
mod1.13all <- glm(Chla\_sqrt~Chla\_ARlag1\_sqrt+mean\_flow+ShortWave\_mean+Turb\_NTU\_log, data = data13.all, family = gaussian, na.action = na.fail)  
pred.1.13all <- predict(mod1.13all, newdata=data13.all)  
mod2.13all <- glm(Chla\_sqrt~Chla\_ARlag1\_sqrt+mean\_flow+NO3NO2\_log+ShortWave\_mean+Turb\_NTU\_log, data = data13.all,   
 family = gaussian, na.action = "na.fail")  
pred.2.13all <- predict(mod2.13all, newdata=data13.all)  
mod3.13all <- glm(Chla\_sqrt~Chla\_ARlag1\_sqrt+mean\_flow+Rain\_sum\_log+ShortWave\_mean+Turb\_NTU\_log,data = data13.all, family = gaussian, na.action = na.fail)  
pred.3.13all <- predict(mod3.13all, newdata=data13.all)  
mod4.13all <- glm(Chla\_sqrt~Chla\_ARlag1\_sqrt+mean\_flow+ShortWave\_mean+Temp\_inf\_max+Turb\_NTU\_log, data = data13.all, family = gaussian, na.action = "na.fail")  
pred.4.13all <- predict(mod4.13all, newdata=data13.all)  
mod5.13all <- glm(Chla\_sqrt~Chla\_ARlag1\_sqrt+mean\_flow+NO3NO2\_log+Rain\_sum\_log+ShortWave\_mean+Turb\_NTU\_log, data = data13.all, family = gaussian, na.action = "na.fail")  
pred.5.13all <- predict(mod5.13all, newdata=data13.all)

#2013 data only model diagnostic table  
table2013 = array(NA,dim=c(5,5))  
row.names(table2013) = c('Mod 1','Mod 2','Mod 3', 'Mod 4', 'Mod 5')  
colnames(table2013) = c("Model Equation", "AICc", "R2 Training", "R2 Full", 'df')  
table2013[1,1] = "0.32Chla(t-1) - 3.6mean\_flow - 0.002ShortWave\_mean + 0.14Turbidity\_log + 1.58"  
table2013[1,2] = round(select13[1,13], digits = 2)  
table2013[1,3] = round((rsq(mod1, type = 'sse')), digits = 3)  
table2013[1,4] = round((rsq(mod1.13all, type = 'sse')), digits = 3)  
table2013[1,5] = select13[1,11]  
table2013[2,1] = "0.30Chla(t-1) - 3.9mean\_flow - 0.002ShortWave\_mean + 0.15Turbidity\_log - 0.09NO3NO2\_log + 1.7"  
table2013[2,2] = round(select13[2,13], digits = 2)  
table2013[2,3] = round((rsq(mod2, type = 'sse')), digits = 3)  
table2013[2,4] = round((rsq(mod2.13all, type = 'sse')), digits = 3)  
table2013[2,5] = select13[2,11]  
table2013[3,1] = "0.33Chla(t-1) - 3.7mean\_flow - 0.002ShortWave\_mean + 0.13Turbidity\_log + 0.017Rain\_sum\_log + 1.6"  
table2013[3,2] = round(select13[3,13], digits = 2)  
table2013[3,3] = round((rsq(mod3, type = 'sse')), digits = 3)  
table2013[3,4] = round((rsq(mod3.13all, type = 'sse')), digits = 3)  
table2013[3,5] = select13[3,11]  
table2013[4,1] = "0.33Chla(t-1) - 3.4mean\_flow - 0.003ShortWave\_mean + 0.15Turbidity\_log + 0.019Temp\_inflow\_max + 1.3"  
table2013[4,2] = round(select13[4,13], digits = 2)  
table2013[4,3] = round((rsq(mod4, type = 'sse')), digits = 3)  
table2013[4,4] = round((rsq(mod4.13all, type = 'sse')), digits = 3)  
table2013[4,5] = select13[4,11]  
table2013[5,1] = "0.32Chla(t-1) - 4.0mean\_flow - 0.002ShortWave\_mean + 0.15Turbidity\_log - 0.09NO3NO2\_log + 0.016Rain\_sum\_log + 1.7"  
table2013[5,2] = round(select13[5,13], digits = 2)  
table2013[5,3] = round((rsq(mod5, type = 'sse')), digits = 3)  
table2013[5,4] = round((rsq(mod5.13all, type = 'sse')), digits = 3)  
table2013[5,5] = select13[5,11]  
  
kable(table2013)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Model Equation | AICc | R2 Training | R2 Full | df |
| Mod 1 | 0.32Chla(t-1) - 3.6mean\_flow - 0.002ShortWave\_mean + 0.14Turbidity\_log + 1.58 | 14.81 | 0.764 | 0.419 | 6 |
| Mod 2 | 0.30Chla(t-1) - 3.9mean\_flow - 0.002ShortWave\_mean + 0.15Turbidity\_log - 0.09NO3NO2\_log + 1.7 | 14.85 | 0.764 | 0.429 | 5 |
| Mod 3 | 0.33Chla(t-1) - 3.7mean\_flow - 0.002ShortWave\_mean + 0.13Turbidity\_log + 0.017Rain\_sum\_log + 1.6 | 15.26 | 0.766 | 0.429 | 6 |
| Mod 4 | 0.33Chla(t-1) - 3.4mean\_flow - 0.003ShortWave\_mean + 0.15Turbidity\_log + 0.019Temp\_inflow\_max + 1.3 | 15.89 | 0.765 | 0.427 | 5 |
| Mod 5 | 0.32Chla(t-1) - 4.0mean\_flow - 0.002ShortWave\_mean + 0.15Turbidity\_log - 0.09NO3NO2\_log + 0.016Rain\_sum\_log + 1.7 | 16.25 | 0.767 | 0.438 | 6 |