Q1: Are lake and terrestrial primary productivity coherent?

Jonathan Walter, Grace Wilkinson, Rachel Fleck, Michael Pace

4/17/2019

This document organizes for openness and reproducibility analyses of the temporal coherence of interannual variation in lake primary productivity with terrestrial primary productivity in the landscape surrounding the lake.

Data import

Data produced in 'ms1__prep.Rmd' are loaded.

```
load("~/Box Sync/NSF EAGER Synchrony/Data/RData files/ms1_analysis_inprogress1_v10873.RData")
```

```
any(sapply(analysislakes$lakedata, function(x){any(is.infinite(x))}))
```

```
## [1] FALSE
```

```
any(sapply(analysislakes$lakedata, function(x){any(is.na(x))}))
```

```
## [1] FALSE
```

```
which(sapply(analysislakes$lakedata, function(x){any(is.na(x))}))
```

```
## named integer(0)
```

```
analysislakes$lakeinfo[which(sapply(analysislakes$lakedata, function(x){any(is.na(x))})),]
```

```
##  [1] lagoslakeid       gnis_name         nhd_lat
##  [4] nhd_long          lake_area_ha      lake_perim_meters
##  [7] nhd_ftype         nhd_fcode         hu4_zoneid
## [10] hu12_zoneid       state_zoneid      elevation_m
## [13] start             end
## <0 rows> (or 0-length row.names)
```

```
# image(accndvi)
# points(lakepts.prj[which(sapply(analysislakes$lakedata, function(x){any(is.na(x))})),])

dbuff[which(sapply(analysislakes$lakedata, function(x){any(is.na(x))}))]
```

```
## numeric(0)
```

```
analysislakes$lakeinfo<-analysislakes$lakeinfo[!sapply(analysislakes$lakedata, function(x){any(is.na(x)
analysislakes$lakedata<-analysislakes$lakedata[!sapply(analysislakes$lakedata, function(x){any(is.na(x)

analysislakes$lakeinfo$tslength<-analysislakes$lakeinfo$end-analysislakes$lakeinfo$start+1
# analysislakes$lakedata<-analysislakes$lakedata[!analysislakes$lakeinfo$tslength < 20]
# analysislakes$lakeinfo<-analysislakes$lakeinfo[!analysislakes$lakeinfo$tslength < 20,]
```

```
source("~/GitHub/AquaTerrSynch/AnalysisCode/bandtest_coh.R")

tsranges<-rbind(c(2,4),c(4,Inf),c(2,Inf))
```

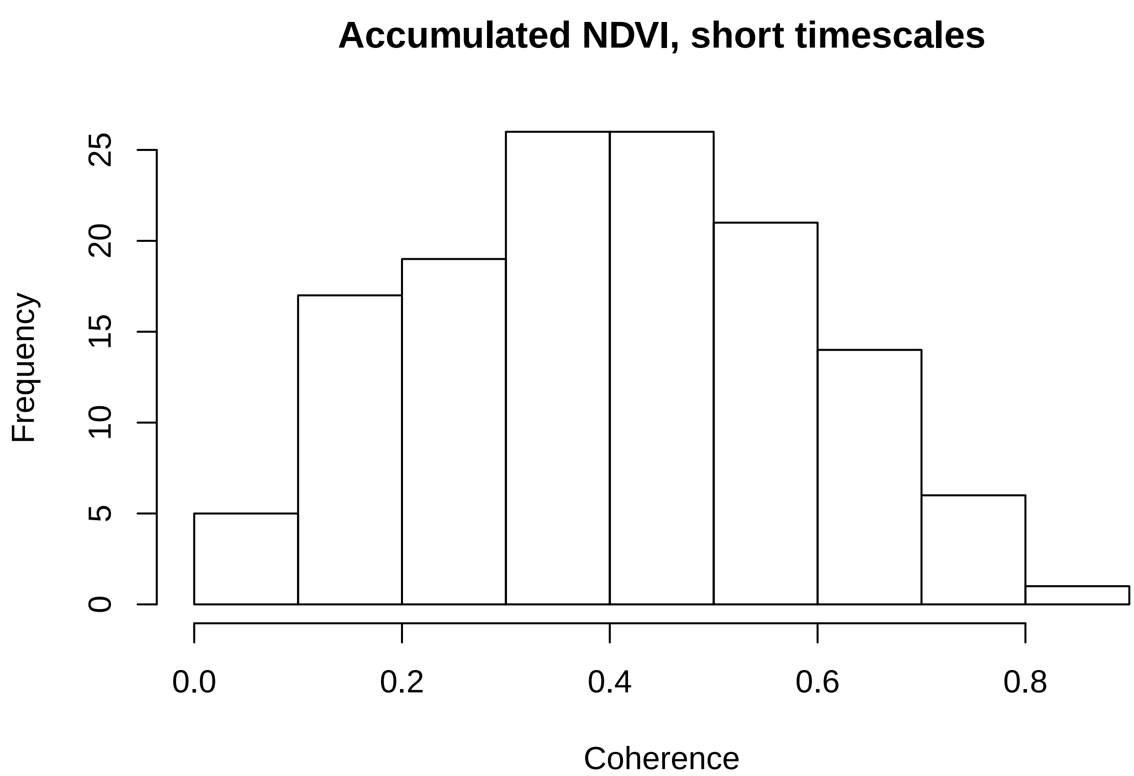
```r
coh.chlaXaccndvi<-NULL
#coh.chlaXmaxndvi<-NULL

for(lind in 1:length(analysislakes$lakedata)){
  lakedat.ii<-cleandat(analysislakes$lakedata[[lind]], as.numeric(colnames(analysislakes$lakedata[[lind
  chlaXaccndvi<-coh(lakedat.ii[1,], lakedat.ii[2,], as.numeric(colnames(analysislakes$lakedata[[lind]])
                    norm="powall", sigmethod="fast", nrand=10000)
#  chlaXmaxndvi<-coh(lakedat.ii[1,], lakedat.ii[3,], as.numeric(colnames(analysislakes$lakedata[[lind]]
#                     norm="powall", sigmethod="fast", nrand=10000)
  for(rind in 1:nrow(tsranges)){
    chlaXaccndvi<-bandtest.coh(chlaXaccndvi, tsranges[rind,])
    #chlaXmaxndvi<-bandtest.coh(chlaXmaxndvi, tsranges[rind,])
  }
  coh.chlaXaccndvi<-rbind(coh.chlaXaccndvi, c(t(as.matrix(chlaXaccndvi$bandp[,3:5]))))
#  coh.chlaXmaxndvi<-rbind(coh.chlaXmaxndvi, c(t(as.matrix(chlaXmaxndvi$bandp[,3:5]))))

}

coh.chlaXaccndvi<-as.data.frame(coh.chlaXaccndvi)
#coh.chlaXmaxndvi<-as.data.frame(coh.chlaXmaxndvi)

colnames(coh.chlaXaccndvi)<-paste0("accndvi",c("p.ts1","phi.ts1","coh.ts1","p.ts2","phi.ts2","coh.ts2",
#colnames(coh.chlaXmaxndvi)<-paste0("maxndvi",c("p.ts1","phi.ts1","coh.ts1","p.ts2","phi.ts2","coh.ts2"

coh.chlaXaccndvi$lagoslakeid<-analysislakes$lakeinfo$lagoslakeid
#coh.chlaXmaxndvi$lagoslakeid<-analysislakes$lakeinfo$lagoslakeid
```

```r
#looking for especially coherent lakes did not return easy-to-interpret examples. Proceeding with simul

tmax=50
res=0.1
tt=seq(1,tmax,res)

p1<-2
sig1<-sin(seq(0,2*pi*tmax/p1,length.out=length(tt)))
p2<-10
sig2<-sin(seq(0,2*pi*tmax/p2,length.out=length(tt)))

comb1<-sig1+0.7*sig2+3.5
comb2<-sig1+-0.7*sig2


laymat<-matrix(1,nrow=2,ncol=3)
laymat[2,]<-2:4

sig3<-sig2[tt<=20]
sig4<-sig3*0.9
sig5<-sin(seq(-pi/2,2*pi*20/p2-(pi/2),length.out=length(tt[tt<=20])))
sig6<-sig3*-1

tiff("~/Box Sync/NSF EAGER Synchrony/Manuscripts/1_CoherenceSpatialVariation/fig1_pedagogical.tif", uni
     res=300, width=6.5, height=4)
```

```r
layout(laymat)
par(mar=c(1.5,1.5,2,1.5), mgp=c(1,1,0), oma=c(2,2,0,0))

plot(NA,NA,ylim=c(-2,5.2),xlim=range(tt), xlab="", ylab="", xaxt="n",yaxt="n")
lines(tt,comb1,lwd=2)
lines(tt,comb2,lwd=2,col="red")
axis(1, at=c(0,10,20,30,40,50),labels=NA)
axis(2, at=c(-1,1.5,4), labels=NA)
mtext("Timescale specific relationship",3,line=0.25)
text(0.4,4.9,"a)",cex=1.2)

plot(NA,NA,ylim=c(-1,1),xlim=c(0,20),xaxt="n",yaxt="n",xlab="",ylab="")
axis(1, at=c(0,10,20), labels=NA)
axis(2, at=c(-1,0,1), labels=NA)
lines(tt[tt<=20],sig3,lwd=2)
lines(tt[tt<=20],sig4,lwd=2,col="red")
mtext(expression(paste(phi," = 0")))
text(1,0.9,"b)",cex=1.2)

plot(NA,NA,ylim=c(-1,1),xlim=c(0,20),xaxt="n",yaxt="n",xlab="",ylab="")
axis(1, at=c(0,10,20), labels=NA)
axis(2, at=c(-1,0,1), labels=NA)
lines(tt[tt<=20],sig3,lwd=2)
lines(tt[tt<=20],sig5,lwd=2,col="red")
mtext(expression(paste(phi," = ",pi,"/2")))
text(1,0.9,"c)",cex=1.2)

plot(NA,NA,ylim=c(-1,1),xlim=c(0,20),xaxt="n",yaxt="n",xlab="",ylab="")
axis(1, at=c(0,10,20), labels=NA)
axis(2, at=c(-1,0,1), labels=NA)
lines(tt[tt<=20],sig3,lwd=2)
lines(tt[tt<=20],sig6,lwd=2,col="red")
mtext(expression(paste(phi," = ",pi)))
text(1,0.9,"d)",cex=1.2)


mtext("Time", 1, outer=T)
mtext("Signal", 2, outer=T)

dev.off()
```

```
## pdf
##   2
```

```r
#short timescales
hist(coh.chlaXaccndvi$accndvicoh.ts1, main="Accumulated NDVI, short timescales", xlab="Coherence", ylab=
```




```
#hist(coh.chlaXmaxndvi$maxndvicoh.ts1, main="Maximum NDVI, short timescales", xlab="Coherence", ylab="F

quantile(coh.chlaXaccndvi$accndvicoh.ts1)
```

```
##          0%        25%        50%        75%       100%
## 0.03540956 0.26015941 0.40373548 0.52492077 0.81625251
```

```
#quantile(coh.chlaXmaxndvi$maxndvicoh.ts1)

alpha=0.05
sum(coh.chlaXaccndvi$accndvip.ts1<alpha)/nrow(coh.chlaXaccndvi)
```

```
## [1] 0.06666667
```

```
#sum(coh.chlaXmaxndvi$maxndvip.ts1<alpha)/nrow(coh.chlaXmaxndvi)

print(cbind(coh.chlaXaccndvi$lagoslakeid, coh.chlaXaccndvi$accndvip.ts1)[coh.chlaXaccndvi$accndvip.ts1<a
```

```
##          [,1]       [,2]
##  [1,]    5104 0.00169983
##  [2,]    5288 0.03849615
##  [3,]    6199 0.00669933
##  [4,]    6399 0.03469653
##  [5,]    6973 0.02419758
##  [6,]    7810 0.01579842
##  [7,]   79457 0.04709529
##  [8,]  136680 0.04859514
##  [9,]    5453 0.02489751
```

```
print(cbind(coh.chlaXaccndvi$lagoslakeid, coh.chlaXaccndvi$accndvip.ts2)[coh.chlaXaccndvi$accndvip.ts2<a
```

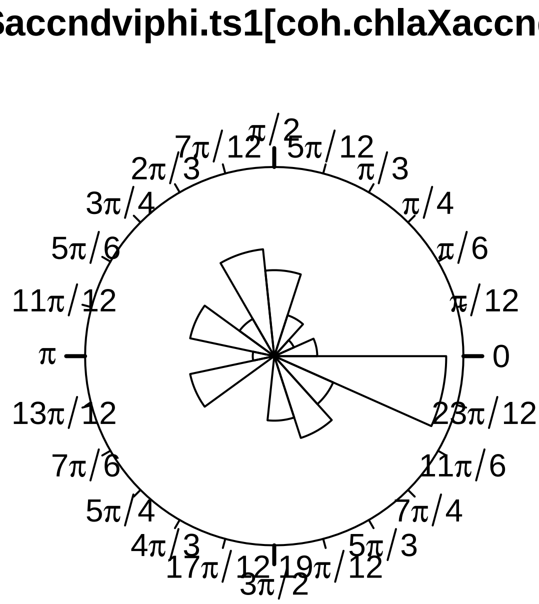
```
##        [,1]       [,2]
```

```
## [1,]    249 0.02229777
## [2,]   6301 0.02349765
## [3,]   7792 0.04729527
## [4,] 136466 0.00749925
## [5,]  14815 0.00889911
## [6,]   3280 0.03769623
## [7,]   5463 0.03249675
```

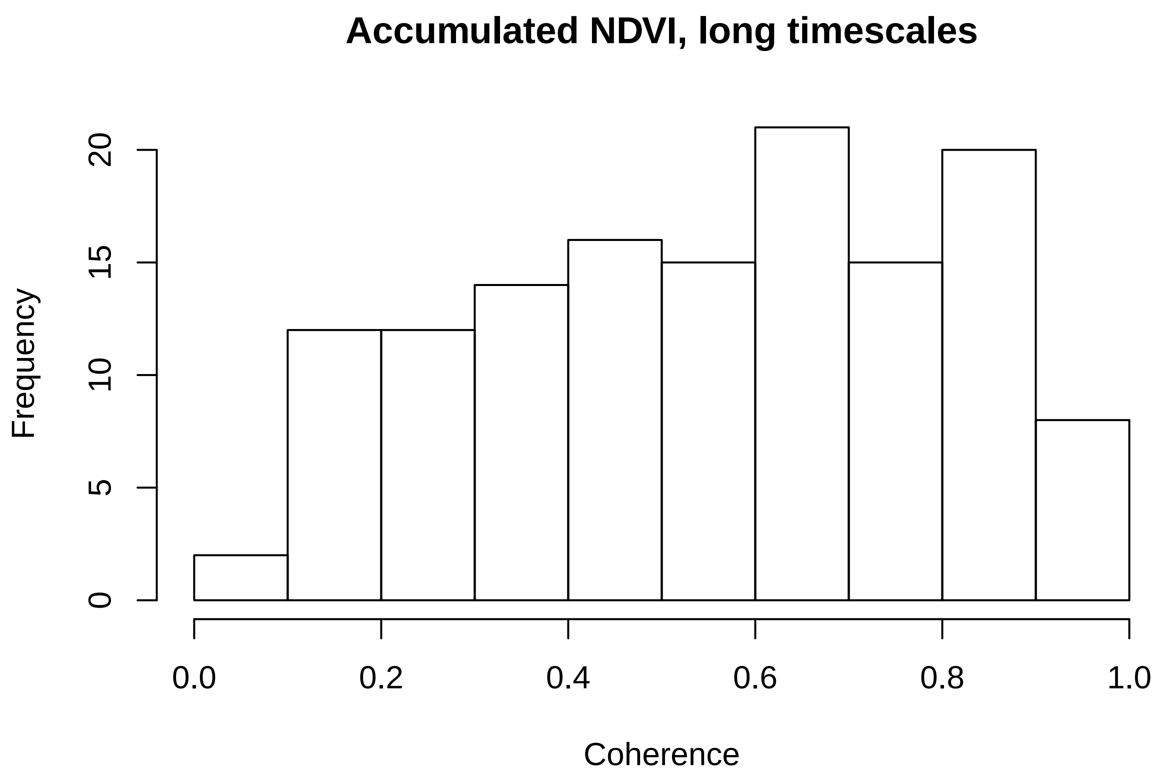
```r
cor(coh.chlaXaccndvi$accndvicoh.ts1,coh.chlaXaccndvi$accndvicoh.ts2)
```

```
## [1] -0.002969988
```

```r
# print(coh.chlaXaccndvi$accndviphi.ts1[coh.chlaXaccndvi$accndvip.ts1<alpha]/pi) #only pattern is that
# print(coh.chlaXmaxndvi$maxndviphi.ts1[coh.chlaXmaxndvi$maxndvip.ts1<alpha]/pi)

phicls<-c(-1,-.75,-0.25,0.25,0.75,1)

# hist(coh.chlaXaccndvi$accndviphi.ts1[coh.chlaXaccndvi$accndvip.ts1<0.2]/pi, main="Accumulated NDVI, s
rose(coh.chlaXaccndvi$accndviphi.ts1[coh.chlaXaccndvi$accndvip.ts1<0.3], unit="radian",
     breaks=seq(0,2*pi,length.out=16))
```




```r
#hist(coh.chlaXmaxndvi$maxndviphi.ts1[coh.chlaXmaxndvi$maxndvip.ts1<0.2]/pi, main="Maximum NDVI, short

#long timescales
hist(coh.chlaXaccndvi$accndvicoh.ts2, main="Accumulated NDVI, long timescales", xlab="Coherence", ylab=
```




```r
#hist(coh.chlaXmaxndvi$maxndvicoh.ts2, main="Maximum NDVI, long timescales", xlab="Coherence", ylab="Fr

quantile(coh.chlaXaccndvi$accndvicoh.ts2)
```

```
##         0%        25%        50%        75%       100%
## 0.06700155 0.35635453 0.56072757 0.75753276 0.96052338
```

```r
#quantile(coh.chlaXmaxndvi$maxndvicoh.ts2)

alpha=0.05
sum(coh.chlaXaccndvi$accndvip.ts2<alpha)/nrow(coh.chlaXaccndvi)
```

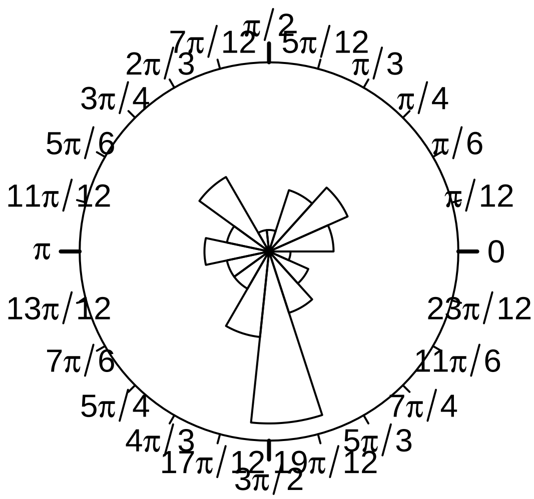
```
## [1] 0.05185185
```

```r
#sum(coh.chlaXmaxndvi$maxndvip.ts2<alpha)/nrow(coh.chlaXmaxndvi)

print(coh.chlaXaccndvi$accndviphi.ts2[coh.chlaXaccndvi$accndvip.ts2<alpha]/pi)
```

```
## [1] -0.43893809  0.25315167 -0.40935196 -0.04386325 -0.65597599  0.36382168
## [7]  0.89471121
```

```r
#print(coh.chlaXmaxndvi$maxndviphi.ts2[coh.chlaXmaxndvi$maxndvip.ts2<alpha]/pi)

# hist(coh.chlaXaccndvi$accndviphi.ts2[coh.chlaXaccndvi$accndvip.ts2<0.2]/pi, main="Accumulated NDVI, l
rose(coh.chlaXaccndvi$accndviphi.ts2[coh.chlaXaccndvi$accndvip.ts2<0.3], unit="radian",
     breaks=seq(0,2*pi,length.out=16))
```

coh.chlaXaccndvi$accndviphi.ts2[coh.chlaXaccndvi$accndvip.ts2 <



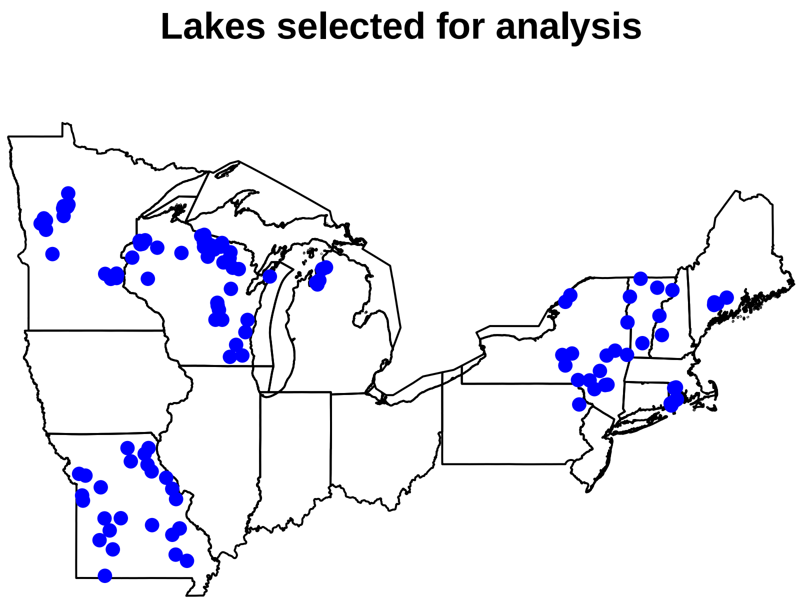

```
#hist(coh.chlaXmaxndvi$maxndviphi.ts1[coh.chlaXmaxndvi$maxndvicoh.ts2>0.6]/pi, main="Maximum NDVI, shor

tiff("~/Box Sync/NSF EAGER Synchrony/Manuscripts/1_CoherenceSpatialVariation/fig2_distributions.tif", un
     res=300, width=6.5, height=6.5)

par(mar=c(3,3,2,1),mgp=c(1.7,0.5,0),mfrow=c(2,2),cex.main=0.9)

hist(coh.chlaXaccndvi$accndvicoh.ts1, main="Short timescale coherence", xlab="Coherence", ylab="Frequen
text(par()$usr[1]+.05,0.95*par()$usr[4],"a)")
hist(coh.chlaXaccndvi$accndvicoh.ts2, main="Long timescale coherence", xlab="Coherence", ylab="Frequency
text(par()$usr[1]+.05,0.95*par()$usr[4],"b)")

par(mar=c(1,1,2,1))
rose(coh.chlaXaccndvi$accndviphi.ts1[coh.chlaXaccndvi$accndvip.ts1<0.3], unit="radian", col="lightgrey"
     breaks=c(0,pi/4,pi/2,3*pi/4,pi,5*pi/4,3*pi/2,7*pi/4,2*pi), main="Short timescale phases",
       at=c(0,pi/4,pi/2,3*pi/4,pi,-3*pi/4,-pi/2,-pi/4))
text(0.9*par()$usr[1],0.95*par()$usr[4],"c)")
rose(coh.chlaXaccndvi$accndviphi.ts2[coh.chlaXaccndvi$accndvip.ts2<0.3], unit="radian", col="lightgrey"
     breaks=c(0,pi/4,pi/2,3*pi/4,pi,5*pi/4,3*pi/2,7*pi/4,2*pi), main="Long timescale phases",
     at=c(0,pi/4,pi/2,3*pi/4,pi,-3*pi/4,-pi/2,-pi/4))
text(0.9*par()$usr[1],0.95*par()$usr[4],"d)")

dev.off()
```

```
## pdf
##   2
```

```
states<-readOGR("~/Box Sync/NSF EAGER Synchrony/Data/statesp020.shp")
```

```
## OGR data source with driver: ESRI Shapefile
## Source: "/Users/jonmacbook/Box Sync/NSF EAGER Synchrony/Data/statesp020.shp", layer: "statesp020"
## with 2895 features
## It has 9 fields
## Integer64 fields read as strings:  STATESP020 DAY_ADM YEAR_ADM
```

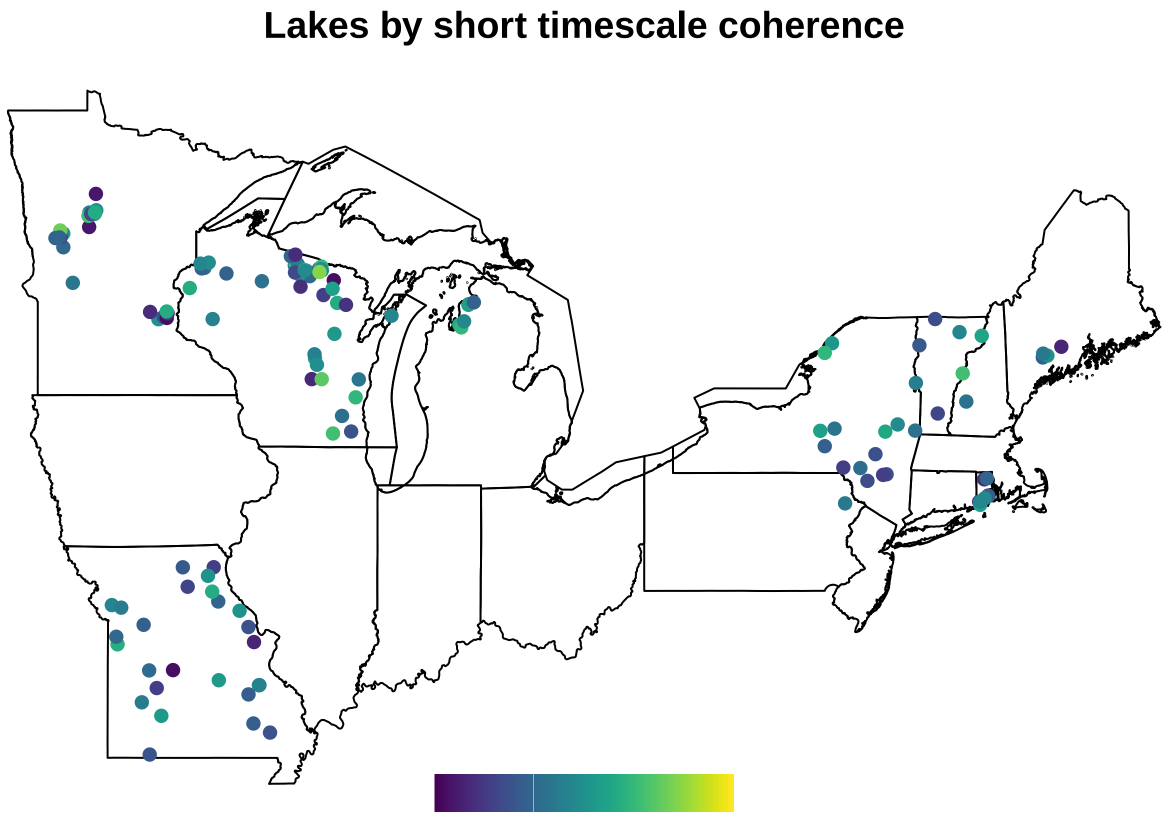
```
getstates<-c("Minnesota", "Iowa", "Wisconsin", "Illinois", "Missouri", "Michigan", "Indiana", "Ohio", "
lagosstates<-states[states@data$STATE %in% getstates,]

plot(lagosstates, main="Lakes selected for analysis")
points(analysislakes$lakeinfo$nhd_long, analysislakes$lakeinfo$nhd_lat, pch=16, cex=1, col="blue")
```



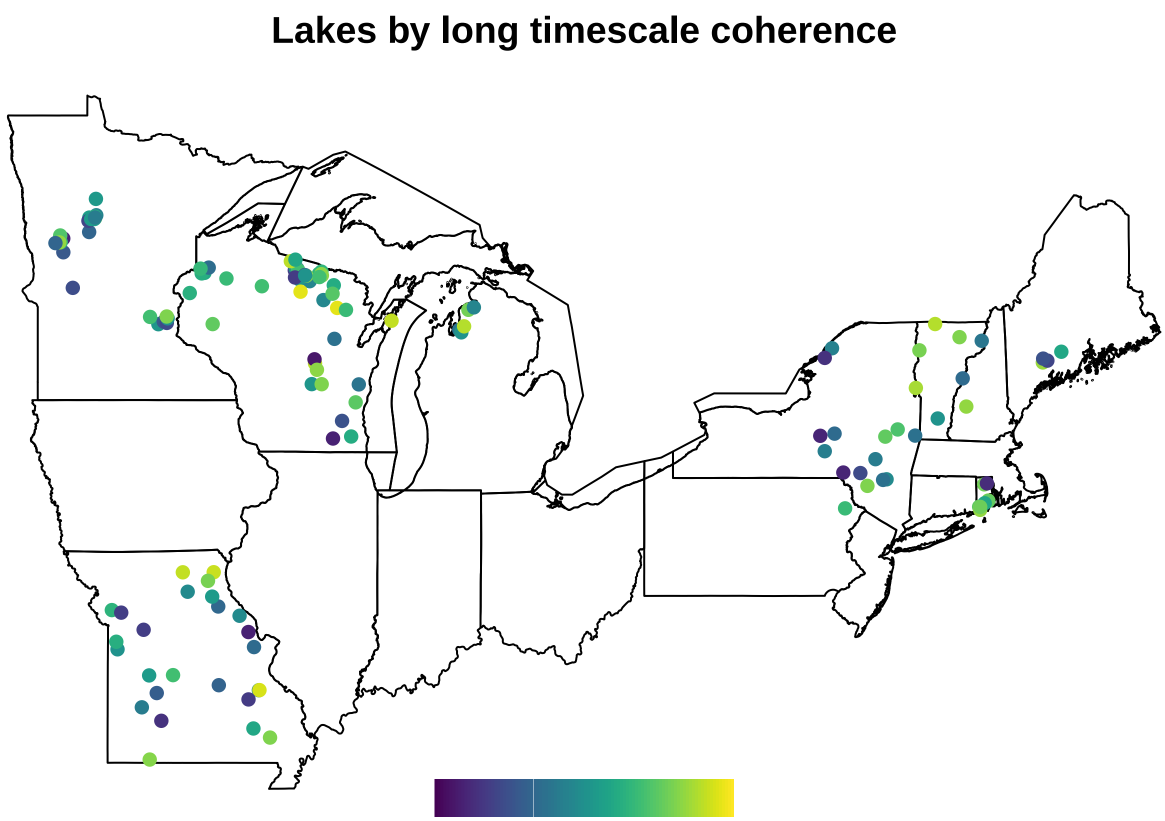

```
cohplotdata<-left_join(analysislakes$lakeinfo, coh.chlaXaccndvi, by="lagoslakeid")

pal<-viridis(100)

par(mar=c(1,0,2,0))

plot(lagosstates, main="Lakes by short timescale coherence")
par()$usr
```

```
## [1] -98.49241 -65.70056  35.45947  49.92124
```

```
points(cohplotdata$nhd_long, cohplotdata$nhd_lat, pch=16, cex=1, col=pal[round(cohplotdata$accndvicoh.t
colorbar.plot(x=mean(par("usr")[1:2]),y=par("usr")[3],strip=1:100,col=pal,horizontal = T)
```




```
plot(lagosstates, main="Lakes by long timescale coherence")
points(cohplotdata$nhd_long, cohplotdata$nhd_lat, pch=16, cex=1, col=pal[round(cohplotdata$accndvicoh.t
colorbar.plot(x=mean(par("usr")[1:2]),y=par("usr")[3],strip=1:100,col=pal,horizontal = T)
```




```
laymat=matrix(1,nrow=2,ncol=13)
laymat[2,]<-2
laymat[,13]<-3

tiff("~/Box Sync/NSF EAGER Synchrony/Manuscripts/1_CoherenceSpatialVariation/fig3_coherencemap.tif", uni
     res=300, width=6.5, height=7.5)

layout(laymat)
par(mar=c(0,0,1.5,0))

plot(lagosstates, main="Short timescale coherence")
points(cohplotdata$nhd_long, cohplotdata$nhd_lat, pch=16, cex=1.5, col=pal[round(cohplotdata$accndvicoh
text(0.99*par()$usr[1],0.99*par()$usr[4],"a)",cex=1.5)

plot(lagosstates, main="Long timescale coherence")
points(cohplotdata$nhd_long, cohplotdata$nhd_lat, pch=16, cex=1.5, col=pal[round(cohplotdata$accndvicoh
text(0.99*par()$usr[1],0.99*par()$usr[4],"b)",cex=1.5)

par(mar=c(5,1,5,1))
image(matrix(1:100,nrow=1),col=pal,xaxt="n",yaxt="n")
axis(2,at=seq(0,1,0.2))

dev.off()
```

```
## pdf
##   2
```

```
dt<-lagosne_load("1.087.3")
```

```
dt.conn<-dt$buffer500m.conn
dt.conn<-dt.conn[,!grepl("sum_lengthm",colnames(dt.conn))]
dt.conn<-dt.conn[,colnames(dt.conn)!="buffer500m_nhdid"]

dt.chag<-dt$hu12.chag
dt.chag<-dt.chag[,!grepl("_min",colnames(dt.chag))]
dt.chag<-dt.chag[,!grepl("_max",colnames(dt.chag))]
dt.chag<-dt.chag[,!grepl("_ha",colnames(dt.chag))]
dt.chag<-dt.chag[,!colnames(dt.chag)=="borderhu12s"]
dt.chag$hu12_dep_no3_tavg_mean<-rowMeans(dt.chag[,grepl("hu12_dep_no3",colnames(dt.chag)) &
                                                    grepl("_mean",colnames(dt.chag))])
dt.chag$hu12_dep_no3_tavg_std<-rowMeans(dt.chag[,grepl("hu12_dep_no3",colnames(dt.chag)) &
                                                    grepl("_std",colnames(dt.chag))])
dt.chag$hu12_dep_so4_tavg_mean<-rowMeans(dt.chag[,grepl("hu12_dep_so4",colnames(dt.chag)) &
                                                     grepl("_mean", colnames(dt.chag))])
dt.chag$hu12_dep_so4_tavg_std<-rowMeans(dt.chag[,grepl("hu12_dep_so4",colnames(dt.chag)) &
                                                     grepl("_std", colnames(dt.chag))])
dt.chag$hu12_dep_totaln_tavg_mean<-rowMeans(dt.chag[,grepl("hu12_dep_totaln",colnames(dt.chag)) &
                                                       grepl("_mean", colnames(dt.chag))])
dt.chag$hu12_dep_totaln_tavg_std<-rowMeans(dt.chag[,grepl("hu12_dep_totaln",colnames(dt.chag)) &
                                                       grepl("_std", colnames(dt.chag))])
dt.chag<-dt.chag[,!(grepl("hu12_dep",colnames(dt.chag)) & grepl("_19",colnames(dt.chag)))]
dt.chag<-dt.chag[,!(grepl("hu12_dep",colnames(dt.chag)) & grepl("_20",colnames(dt.chag)))]


dt.geo<-dt$lakes.geo
dt.geo<-dt$lakes.geo[,!colnames(dt.geo) %in% c("state_zoneid","iws_zoneid","edu_zoneid")]

dt.lulc<-dt$hu12.lulc
dt.lulc<-dt.lulc[,!grepl("_ha_",colnames(dt.lulc))]
dt.lulc<-dt.lulc[,!grepl("_nlcd1992_",colnames(dt.lulc))]
dt.lulc<-dt.lulc[,!grepl("_nlcd2006_",colnames(dt.lulc))]
dt.lulc<-dt.lulc[,!grepl("_nlcd2001_",colnames(dt.lulc))]
dt.lulc<-dt.lulc[,colnames(dt.lulc)!="hu12_damdensity_pointsperha"]
dt.lulc<-dt.lulc[,colnames(dt.lulc)!="hu12_damdensity_pointcount"]
dt.lulc<-dt.lulc[,colnames(dt.lulc)!="hu12_roaddensity_sum_lengthm"]
dt.lulc<-dt.lulc[,!grepl("_min",colnames(dt.lulc))]
dt.lulc<-dt.lulc[,!grepl("_max",colnames(dt.lulc))]

#depth
depth<-lagosne_select(table="lakes_limno", vars=c("lagoslakeid","maxdepth"))
depth<-depth[depth$lagoslakeid %in% analysislakes$lakeinfo$lagoslakeid,] #use max depth because it's mo

#growing season Chlorophyll-a
chla<-lagosne_select(table="epi_nutr", vars=c("lagoslakeid","samplemonth","chla"))
chla<-chla[chla$lagoslakeid %in% analysislakes$lakeinfo$lagoslakeid,]
gs.chla<-chla[chla$samplemonth %in% 5:9,]
avg.chla<-aggregate(chla ~ lagoslakeid, data=gs.chla, FUN=mean, na.rm=T)


#Chlorophyll-a TSI class
#TSI(CHL) = 9.81 ln(CHL) + 30.6
tsi.chl<-data.frame(lagoslakeid=avg.chla$lagoslakeid, tsi=9.81 * log(avg.chla$chla) + 30.6)
```

```r
tsi.chl$tsi.cat<-rep("lake",nrow(tsi.chl))

tsi.chl$tsi.cat[tsi.chl$tsi < 40]<-"oligotrophic"
tsi.chl$tsi.cat[tsi.chl$tsi >=40 & tsi.chl$tsi < 50]<-"mesotrophic"
tsi.chl$tsi.cat[tsi.chl$tsi >=50 & tsi.chl$tsi < 70]<-"eutrophic"
tsi.chl$tsi.cat[tsi.chl$tsi >= 70] <-"hypereutrophic"

#CV of terrestrial NDVI
cv.accndvi<-NULL
for(lake in 1:length(analysislakes$lakedata)){
  tmp<-analysislakes$lakedata[[lake]][rownames(analysislakes$lakedata[[lake]])=="accndvi",]
  cv.accndvi<-c(cv.accndvi, sd(tmp)/mean(tmp))
 # rm(tmp)
}
cv.accndvi<-data.frame(lagoslakeid=as.numeric(names(analysislakes$lakedata)), cv.accndvi=cv.accndvi)

#shoreline development ratio
sdev<-analysislakes$lakeinfo$lake_perim_meters/(2*sqrt(pi*analysislakes$lakeinfo$lake_area_ha*10000))
shoredev<-data.frame(lagoslakeid=analysislakes$lakeinfo$lagoslakeid,shoredev=sdev)

preds<-analysislakes$lakeinfo[,colnames(analysislakes$lakeinfo) %in% c("lagoslakeid","end","start")]
preds$tslength<-preds$end-preds$start + 1
preds<-left_join(preds, dt.geo, by="lagoslakeid")
preds<-left_join(preds, dt.conn, by="lagoslakeid")
preds<-left_join(preds, dt.chag, by="hu12_zoneid")
preds<-left_join(preds, dt.lulc, by="hu12_zoneid")
preds<-left_join(preds, avg.chla, by="lagoslakeid")
preds<-left_join(preds, cv.accndvi, by="lagoslakeid")
preds<-left_join(preds, depth, by="lagoslakeid")

# modvars.conn<-left_join(pred.conn, coh.chlaXaccndvi, by="lagoslakeid")
# modvars.chag<-left_join(pred.chag, coh.chlaXaccndvi, by="lagoslakeid")

#huc2 and huc4 watershed codes
huc_codes<-read.csv("~/GitHub/AquaTerrSynch/AnalysisCode/match_huc_codes.csv", colClasses = 'character')

#state info
states<-lagosne_select(table="state", vars=c("state_zoneid","state_name"))

for(nn in 1:ncol(preds)){

  if(is.factor(preds[,nn])){
    preds[,nn]<-factor(preds[,nn])
  }

}
```

```r
rfdat.cohst<-left_join(coh.chlaXaccndvi[,c(10,3)], preds)
```

```
## Joining, by = "lagoslakeid"
```

```r
rfdat.cohst<-rfdat.cohst[,!colnames(rfdat.cohst) %in% c("lagoslakeid","start","end","lakes_nhdid","hu12_
rfdat.cohst<-rfdat.cohst[,!grepl("borderhu12s",colnames(rfdat.cohst))]
```

```r
for(nn in 1:ncol(rfdat.cohst)){
  if(is.character(rfdat.cohst[,nn])){
    rfdat.cohst[,nn]<-as.factor(rfdat.cohst[,nn])
  }
}

cf.cohst<-party::cforest(accndvicoh.ts1 ~ ., data=rfdat.cohst, controls=cforest_control(ntree=50000,min

varimp.coh.st<-varimp(cf.cohst)
print(varimp.coh.st[order(varimp.coh.st, decreasing=T)][1:20])
```

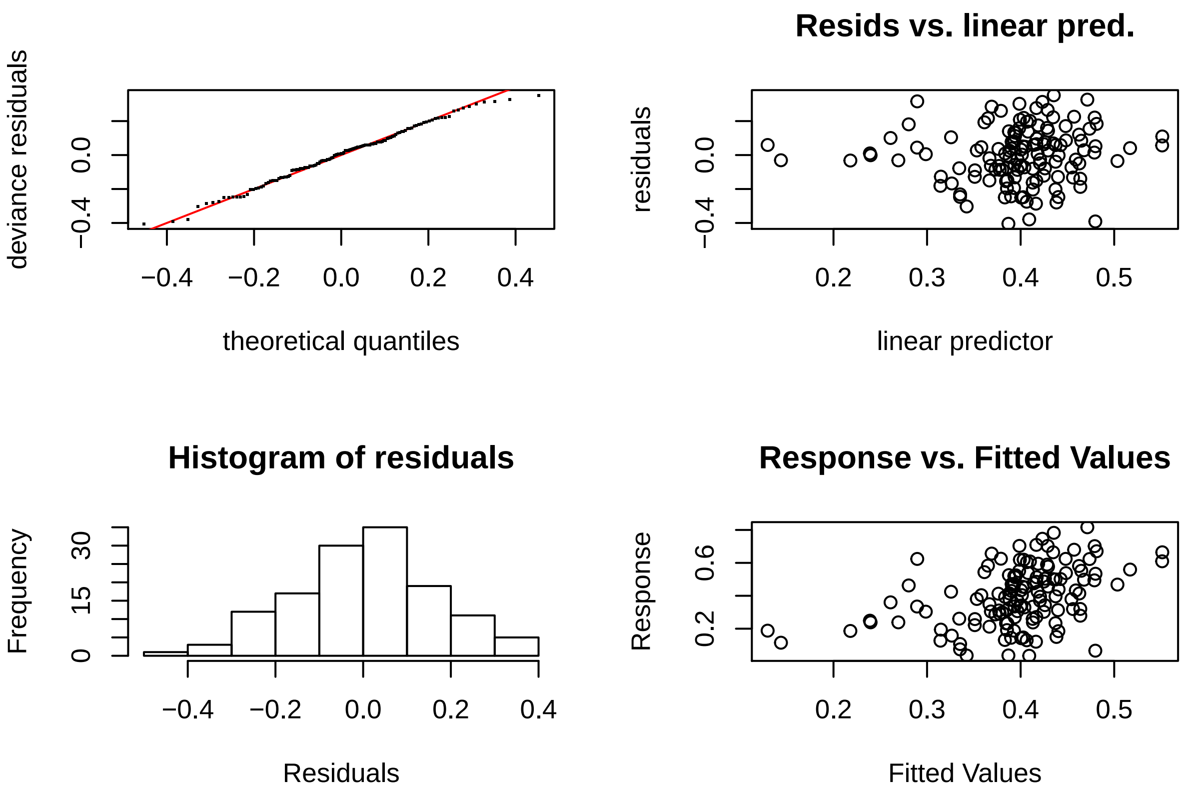
```
## buffer500m_streamdensity_midreaches_density_mperha
##                                        3.830673e-04
##     buffer500m_streamdensity_streams_density_mperha
##                                        2.675314e-04
##                                 hu12_nlcd2011_pct_31
##                                        2.461680e-04
##                              hu12_baseflowindex_mean
##                                        2.325104e-04
##       hu12_prism_ppt_30yr_normal_800mm2_annual_mean
##                                        1.795815e-04
##                             upstream_lakes_4ha_count
##                                        1.680815e-04
##                                 hu12_nlcd2011_pct_82
##                                        1.168098e-04
##                                 hu12_nlcd2011_pct_23
##                                        1.132611e-04
## buffer500m_streamdensity_headwaters_density_mperha
##                                        1.112351e-04
##                                        hu12_tri_mean
##                                        1.076807e-04
##                  hu12_surficialgeology_lac_clay_pct
##                                        1.053586e-04
##      hu12_prism_tmax_30yr_normal_800mm2_annual_mean
##                                        1.034309e-04
##                                      hu12_slope_mean
##                                        1.020266e-04
##                                       hu12_slope_std
##                                        1.004384e-04
##                                         hu12_tri_std
##                                        9.254496e-05
##                                 hu12_nlcd2011_pct_22
##                                        8.243305e-05
##                                 hu12_nlcd2011_pct_52
##                                        7.301085e-05
##                          upstream_lakes_4ha_area_ha
##                                        6.843541e-05
##                                 hu12_nlcd2011_pct_24
##                                        6.810628e-05
##                            upstream_lakes_10ha_count
##                                        5.973624e-05
```

```r
#hist(predcoh.st)
#hist(modvars.accndvi$accndvicoh.ts1)
```

```r
predcoh.st<-predict(cf.cohst, newdata=rfdat.cohst,type="response")
# plot(predcoh.st, rfdat.cohst$accndvicoh.ts1, xlab="predicted", ylab="empirical", main="Coherence, sho
#       xlim=c(0,1), ylim=c(0,1))
# abline(a=0,b=1)
cor.test(predcoh.st,rfdat.cohst$accndvicoh.ts1)
```

```
##
##  Pearson's product-moment correlation
##
## data:  predcoh.st and rfdat.cohst$accndvicoh.ts1
## t = 14.857, df = 133, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.7166293 0.8459872
## sample estimates:
##       cor
## 0.7899407
```

```r
lwgt<-preds$tslength/mean(preds$tslength)

gam.cohst<-gam(accndvicoh.ts1 ~ s(buffer500m_streamdensity_midreaches_density_mperha) + s(hu12_nlcd2011_
               s(buffer500m_streamdensity_streams_density_mperha) + s(hu12_baseflowindex_mean) +
                 s(hu12_prism_ppt_30yr_normal_800mm2_annual_mean), data=rfdat.cohst, gamma=1, weights=lw
gam.check(gam.cohst)
```




```
##
## Method: GCV   Optimizer: magic
## Smoothing parameter selection converged after 12 iterations.
## The RMS GCV score gradient at convergence was 3.606828e-08 .
```

```
## The Hessian was positive definite.
## Model rank =  46 / 46
##
## Basis dimension (k) checking results. Low p-value (k-index<1) may
## indicate that k is too low, especially if edf is close to k'.
##
##                                                         k'  edf k-index
## s(buffer500m_streamdensity_midreaches_density_mperha) 9.00 1.00    0.97
## s(hu12_nlcd2011_pct_31)                               9.00 1.00    0.99
## s(buffer500m_streamdensity_streams_density_mperha)    9.00 4.87    1.13
## s(hu12_baseflowindex_mean)                            9.00 2.31    1.07
## s(hu12_prism_ppt_30yr_normal_800mm2_annual_mean)      9.00 1.39    1.08
##                                                       p-value
## s(buffer500m_streamdensity_midreaches_density_mperha)    0.33
## s(hu12_nlcd2011_pct_31)                                  0.39
## s(buffer500m_streamdensity_streams_density_mperha)       0.93
## s(hu12_baseflowindex_mean)                               0.86
## s(hu12_prism_ppt_30yr_normal_800mm2_annual_mean)         0.74
```

```r
concurvity(gam.cohst)
```

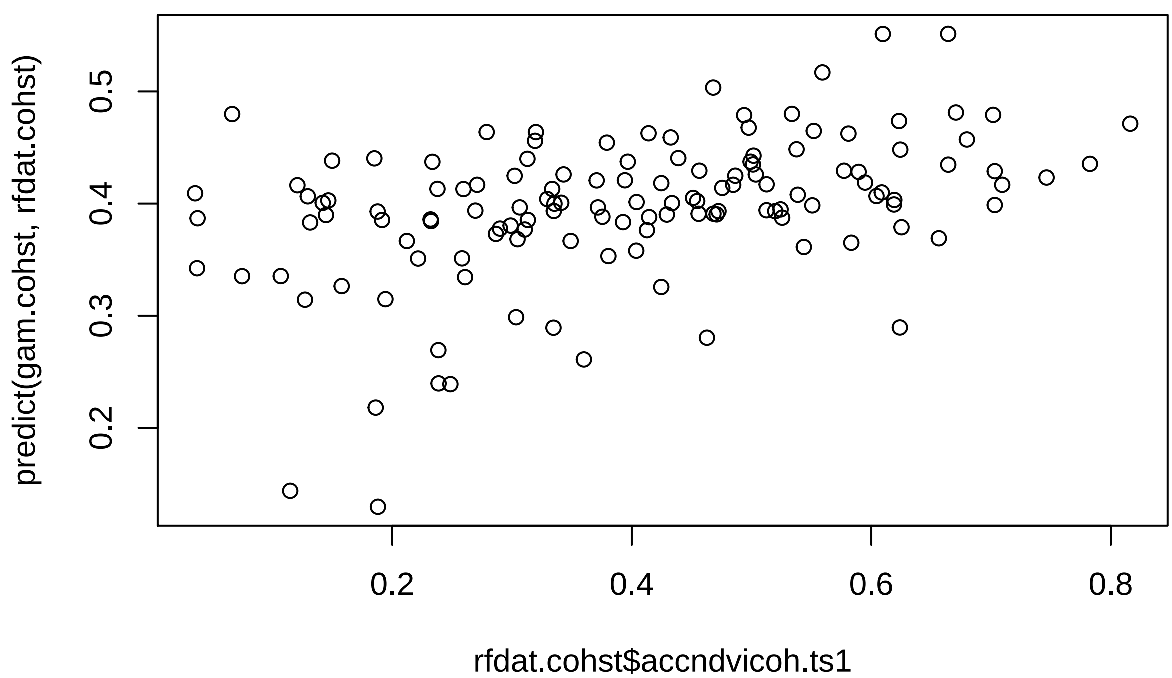
```
##                  para
## worst    3.336985e-22
## observed 3.336985e-22
## estimate 3.336985e-22
##          s(buffer500m_streamdensity_midreaches_density_mperha)
## worst                                                0.8399928
## observed                                             0.5620454
## estimate                                             0.5751856
##          s(hu12_nlcd2011_pct_31)
## worst                  0.9837860
## observed               0.5340242
## estimate               0.4327949
##          s(buffer500m_streamdensity_streams_density_mperha)
## worst                                             0.7594393
## observed                                          0.3213433
## estimate                                          0.5561478
##          s(hu12_baseflowindex_mean)
## worst                     0.8511012
## observed                  0.5587086
## estimate                  0.6868655
##          s(hu12_prism_ppt_30yr_normal_800mm2_annual_mean)
## worst                                           0.9835253
## observed                                        0.6313437
## estimate                                        0.6924323
```

```r
summary(gam.cohst)
```

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## accndvicoh.ts1 ~ s(buffer500m_streamdensity_midreaches_density_mperha) +
##     s(hu12_nlcd2011_pct_31) + s(buffer500m_streamdensity_streams_density_mperha) +
```

```
##     s(hu12_baseflowindex_mean) + s(hu12_prism_ppt_30yr_normal_800mm2_annual_mean)
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   0.3972     0.0147   27.01   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                                                      edf Ref.df     F
## s(buffer500m_streamdensity_midreaches_density_mperha) 1.000  1.000 2.325
## s(hu12_nlcd2011_pct_31)                               1.000  1.000 3.237
## s(buffer500m_streamdensity_streams_density_mperha)    4.866  5.861 0.590
## s(hu12_baseflowindex_mean)                            2.313  2.917 1.486
## s(hu12_prism_ppt_30yr_normal_800mm2_annual_mean)      1.386  1.681 0.207
##                                                      p-value
## s(buffer500m_streamdensity_midreaches_density_mperha)  0.1298
## s(hu12_nlcd2011_pct_31)                                0.0744 .
## s(buffer500m_streamdensity_streams_density_mperha)     0.6861
## s(hu12_baseflowindex_mean)                             0.2744
## s(hu12_prism_ppt_30yr_normal_800mm2_annual_mean)       0.7568
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.106   Deviance explained = 17.8%
## GCV = 0.03142  Scale est. = 0.028687  n = 133
```

```
plot(rfdat.cohst$accndvicoh.ts1, predict(gam.cohst, rfdat.cohst))
```



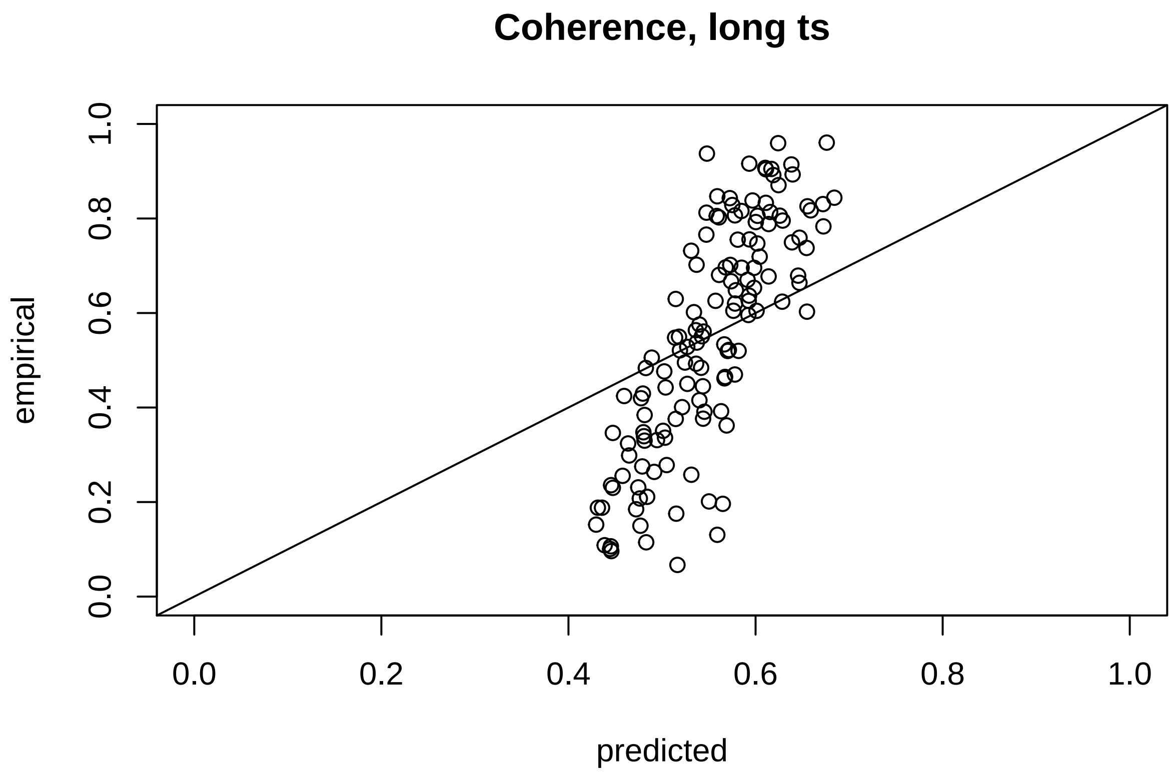

```
rfdat.cohlt<-left_join(coh.chlaXaccndvi[,c(10,6)], preds)
```

```
## Joining, by = "lagoslakeid"
```

```r
rfdat.cohlt<-rfdat.cohlt[,!colnames(rfdat.cohlt) %in% c("lagoslakeid","start","end","lakes_nhdid","hu12_
rfdat.cohlt<-rfdat.cohlt[,!grepl("borderhu12s",rfdat.cohlt)]

for(nn in 1:ncol(rfdat.cohlt)){
  if(is.character(rfdat.cohlt[,nn])){
    rfdat.cohlt[,nn]<-as.factor(rfdat.cohlt[,nn])
  }
}

cf.cohlt<-party::cforest(accndvicoh.ts2 ~ ., data=rfdat.cohlt, controls=cforest_control(ntree=50000,minc

varimp.coh.st<-varimp(cf.cohlt)
print(varimp.coh.st[order(varimp.coh.st, decreasing=T)][1:20])
```

```
##                           hu12_nlcd2011_pct_90
##                                   3.481914e-04
##                          hu12_canopy2001_mean
##                                   3.130793e-04
##                                    hu6_zoneid
##                                   2.371690e-04
##                                     cv.accndvi
##                                   2.106509e-04
##                                    hu8_zoneid
##                                   1.762389e-04
##                 hu12_groundwaterrecharge_mean
##                                   1.123435e-04
##                     hu12_dep_totaln_tavg_mean
##                                   7.653532e-05
##                          hu12_nlcd2011_pct_81
##                                   6.590974e-05
##                        hu12_dep_so4_tavg_mean
##                                   5.001752e-05
## wlconnections_allwetlands_contributing_area_ha
##                                   4.885977e-05
##                              hu12_runoff_mean
##                                   4.846892e-05
##                        hu12_dep_no3_tavg_mean
##                                   4.741097e-05
##                                    hu4_zoneid
##                                   4.692585e-05
##                          hu12_nlcd2011_pct_82
##                                   4.512522e-05
##               hu12_surficialgeology_solut_pct
##                                   4.220645e-05
##                         hu12_dep_so4_tavg_std
##                                   3.939962e-05
##                      hu12_dep_totaln_tavg_std
##                                   3.878235e-05
##  hu12_prism_tmax_30yr_normal_800mm2_annual_std
##                                   3.811235e-05
##                          hu12_nlcd2011_pct_21
##                                   3.656769e-05
##           hu12_surficialgeology_till_sand_pct
##                                   3.634218e-05
```

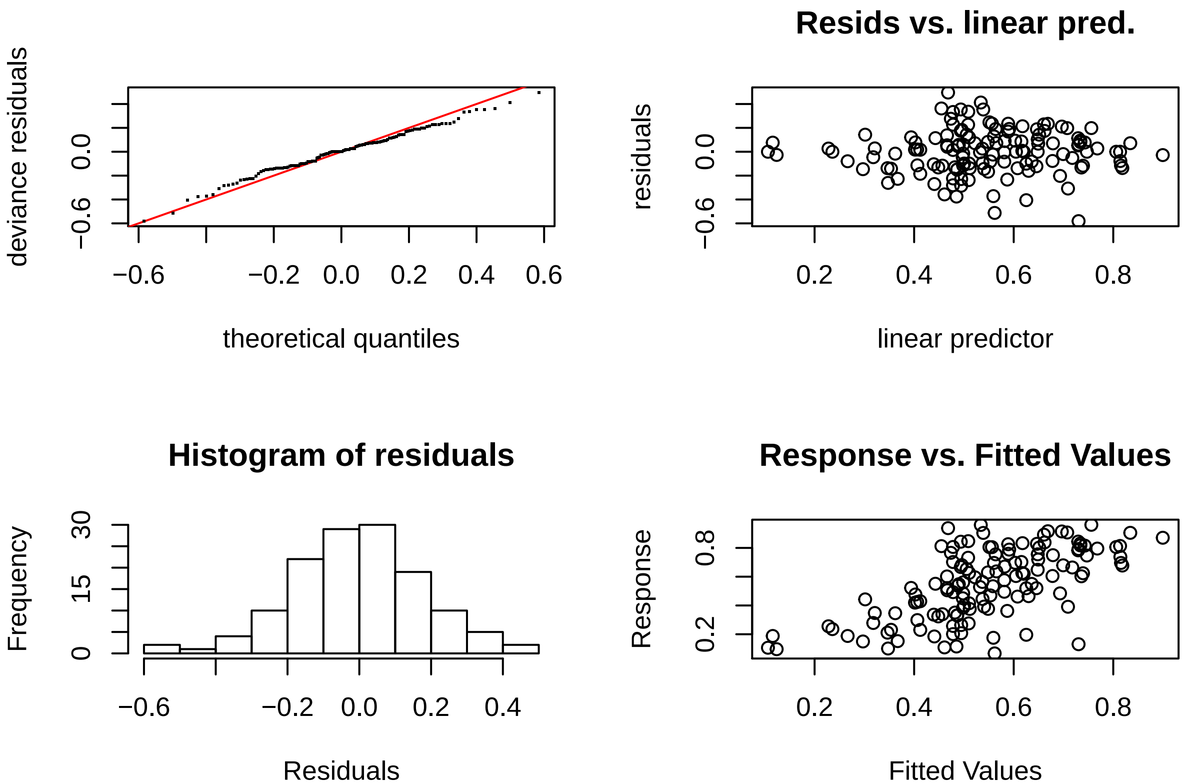
```r
#hist(predcoh.st)
#hist(modvars.accndvi$accndvicoh.ts1)
predcoh.lt<-predict(cf.cohlt, newdata=rfdat.cohlt,type="response")
plot(predcoh.lt, rfdat.cohlt$accndvicoh.ts2, xlab="predicted", ylab="empirical", main="Coherence, long t
     xlim=c(0,1), ylim=c(0,1))
abline(a=0,b=1)
```




```r
cor.test(predcoh.lt,rfdat.cohlt$accndvicoh.ts2)
```

```
##
##  Pearson's product-moment correlation
##
## data:  predcoh.lt and rfdat.cohlt$accndvicoh.ts2
## t = 15.72, df = 133, p-value < 2.2e-16
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.7378433 0.8583164
## sample estimates:
##       cor
## 0.8062857
```

```r
lwgt<-preds$tslength/mean(preds$tslength)

gam.cohlt<-gam(accndvicoh.ts2 ~ s(hu12_nlcd2011_pct_90 ) + s(hu12_canopy2001_mean) + s(cv.accndvi) +
                 s(hu12_dep_totaln_tavg_mean) + hu6_zoneid, data=rfdat.cohlt, gamma=1, weights=lwgt)
gam.check(gam.cohlt)
```




```
##
## Method: GCV   Optimizer: magic
## Smoothing parameter selection converged after 10 iterations.
## The RMS GCV score gradient at convergence was 4.275192e-08 .
## The Hessian was positive definite.
## Model rank =  68 / 68
##
## Basis dimension (k) checking results. Low p-value (k-index<1) may
## indicate that k is too low, especially if edf is close to k'.
##
##                                 k'  edf k-index p-value
## s(hu12_nlcd2011_pct_90)       9.00 1.69    0.94    0.21
## s(hu12_canopy2001_mean)       9.00 3.22    0.98    0.33
## s(cv.accndvi)                 9.00 1.00    0.88    0.10
## s(hu12_dep_totaln_tavg_mean) 9.00 1.00    1.05    0.71
```

```r
concurvity(gam.cohlt)
```

```
##               para s(hu12_nlcd2011_pct_90) s(hu12_canopy2001_mean)
## worst    0.9893564               0.9300358               0.8699693
## observed 0.9893564               0.9057199               0.6241289
## estimate 0.9893564               0.7691194               0.8104988
##          s(cv.accndvi) s(hu12_dep_totaln_tavg_mean)
## worst        0.9294917                    0.9889380
## observed     0.8750000                    0.9046704
## estimate     0.8592849                    0.8655129
```

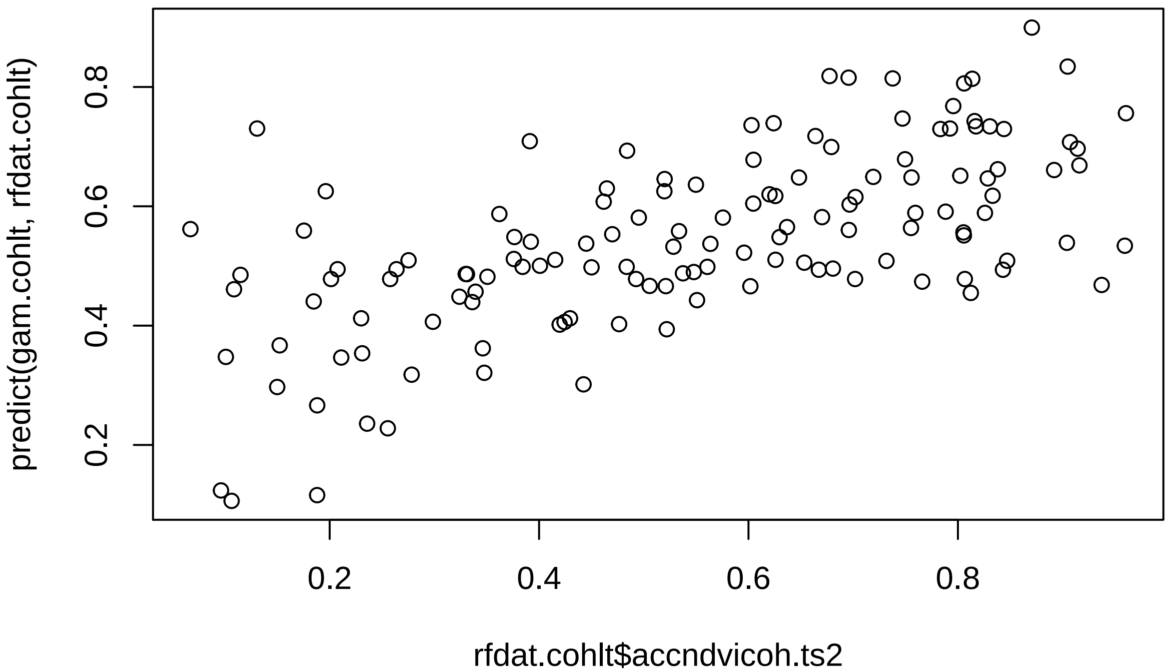
```r
summary(gam.cohlt)
```

```
##
```

```
## Family: gaussian
## Link function: identity
##
## Formula:
## accndvicoh.ts2 ~ s(hu12_nlcd2011_pct_90) + s(hu12_canopy2001_mean) +
##     s(cv.accndvi) + s(hu12_dep_totaln_tavg_mean) + hu6_zoneid
##
## Parametric coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)        0.84626    0.10090   8.387 4.55e-13 ***
## hu6_zoneidHU6_14  0.24295    0.19365   1.255 0.212721
## hu6_zoneidHU6_15 -0.18377    0.13791  -1.333 0.185875
## hu6_zoneidHU6_19 -0.61342    0.27494  -2.231 0.028027 *
## hu6_zoneidHU6_21 -0.45554    0.18307  -2.488 0.014574 *
## hu6_zoneidHU6_22 -0.31872    0.19824  -1.608 0.111215
## hu6_zoneidHU6_23 -0.15857    0.17048  -0.930 0.354665
## hu6_zoneidHU6_35 -0.45995    0.19173  -2.399 0.018389 *
## hu6_zoneidHU6_37 -0.48125    0.26436  -1.820 0.071831 .
## hu6_zoneidHU6_38 -0.26506    0.11835  -2.240 0.027449 *
## hu6_zoneidHU6_4  -0.21511    0.16368  -1.314 0.191935
## hu6_zoneidHU6_40 -0.44925    0.17773  -2.528 0.013127 *
## hu6_zoneidHU6_41 -0.15293    0.13589  -1.125 0.263241
## hu6_zoneidHU6_44 -0.48265    0.15640  -3.086 0.002658 **
## hu6_zoneidHU6_45 -0.27283    0.14926  -1.828 0.070702 .
## hu6_zoneidHU6_46 -0.83866    0.21482  -3.904 0.000177 ***
## hu6_zoneidHU6_47 -0.46276    0.15207  -3.043 0.003028 **
## hu6_zoneidHU6_48 -0.25114    0.26150  -0.960 0.339283
## hu6_zoneidHU6_49 -0.37360    0.26501  -1.410 0.161877
## hu6_zoneidHU6_7  -0.09754    0.16381  -0.595 0.552944
## hu6_zoneidHU6_70 -0.24197    0.15977  -1.514 0.133221
## hu6_zoneidHU6_73 -0.39942    0.15829  -2.523 0.013282 *
## hu6_zoneidHU6_75 -0.77860    0.28479  -2.734 0.007465 **
## hu6_zoneidHU6_76 -0.38087    0.19088  -1.995 0.048870 *
## hu6_zoneidHU6_8   0.18237    0.24874   0.733 0.465243
## hu6_zoneidHU6_83 -0.35948    0.14265  -2.520 0.013401 *
## hu6_zoneidHU6_84 -0.22060    0.20676  -1.067 0.288695
## hu6_zoneidHU6_86 -0.10141    0.24847  -0.408 0.684086
## hu6_zoneidHU6_89 -0.58406    0.20112  -2.904 0.004581 **
## hu6_zoneidHU6_90  0.13563    0.24620   0.551 0.582981
## hu6_zoneidHU6_91  0.04376    0.19560   0.224 0.823438
## hu6_zoneidHU6_93 -0.27056    0.13784  -1.963 0.052585 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Approximate significance of smooth terms:
##                                edf Ref.df      F p-value
## s(hu12_nlcd2011_pct_90)      1.694  2.142  0.762 0.48034
## s(hu12_canopy2001_mean)      3.224  4.073  0.788 0.54444
## s(cv.accndvi)                1.000  1.000 10.312 0.00179 **
## s(hu12_dep_totaln_tavg_mean) 1.000  1.000  0.549 0.46040
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.179   Deviance explained = 41.3%
```

```
## GCV = 0.067188  Scale est. = 0.047675  n = 134
```

```
plot(rfdat.cohlt$accndvicoh.ts2, predict(gam.cohlt, rfdat.cohlt))
```




```
rfdat.phist<-left_join(coh.chlaXaccndvi[,c(10,2)], preds)
```

```
## Joining, by = "lagoslakeid"
```

```
rfdat.phist<-rfdat.phist[,!colnames(rfdat.phist) %in%
                          c("lagoslakeid","start","end","lakes_nhdid","hu12_zoneid","tslength","county_
rfdat.phist<-rfdat.phist[,!grepl("borderhu12s",colnames(rfdat.phist))]

rfdat.phist<-rfdat.phist[coh.chlaXaccndvi$accndvip.ts1<0.3,]

for(nn in 1:ncol(rfdat.phist)){
  if(is.character(rfdat.phist[,nn])){
    rfdat.phist[,nn]<-as.factor(rfdat.phist[,nn])
  }
}

cf.phist<-party::cforest(cos(accndviphi.ts1) ~ ., data=rfdat.phist,
                         controls=cforest_control(ntree=50000,mincriterion = 0.9,mtry=3))

varimp.phi.st<-varimp(cf.phist)
print(varimp.phi.st[order(varimp.phi.st, decreasing=T)][1:20])
```

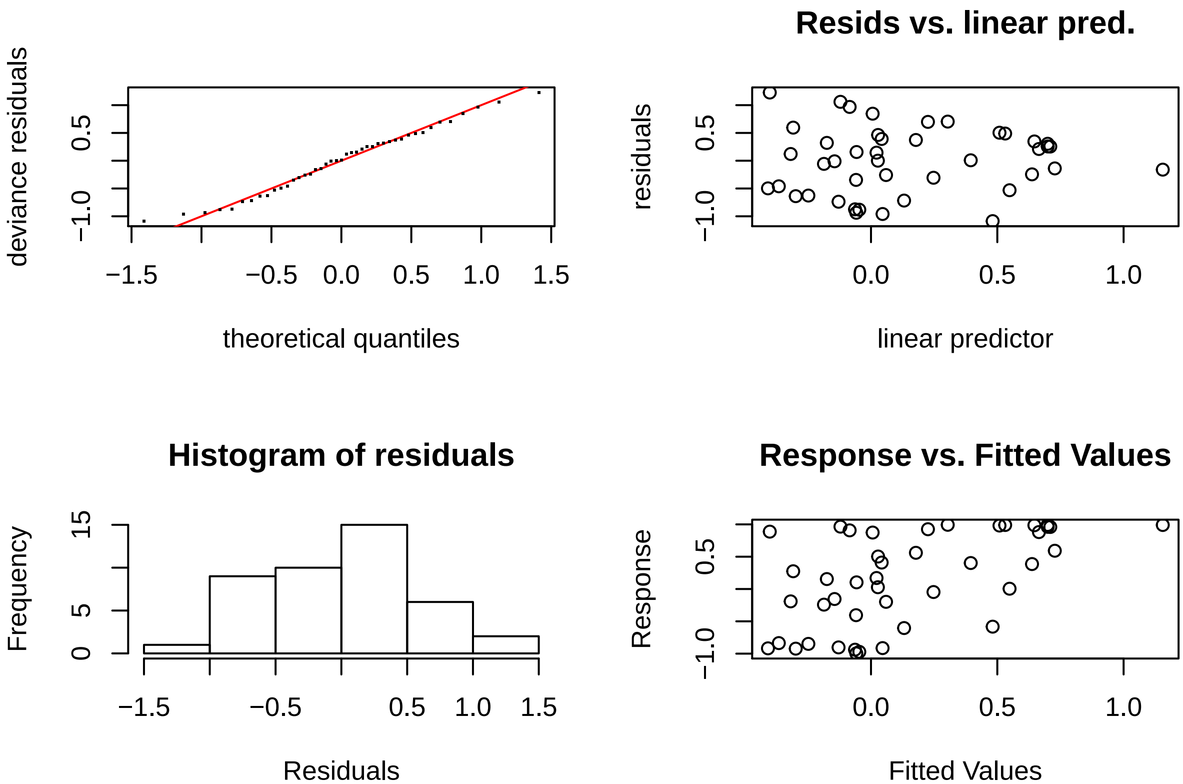
```
##                                          maxdepth
##                                      0.0072861631
##                             upstream_lakes_4ha_count
##                                      0.0054034420
## wlconnections_openwaterwetlands_contributing_area_
##                                      0.0044939538
##                               upstream_lakes_4ha_area_ha
##                                      0.0040036075
```

```
##       wlconnections_openwaterwetlands_shoreline_km
##                                       0.0038026868
##                                     lakeconnection
##                                       0.0037334924
##                          upstream_lakes_10ha_count
##                                       0.0034657661
##                        upstream_lakes_10ha_area_ha
##                                       0.0028686902
##                    hu12_surficialgeology_ice_pct
##                                       0.0014542935
## buffer500m_streamdensity_midreaches_density_mperha
##                                       0.0013807573
##              wlconnections_openwaterwetlands_count
##                                       0.0012124286
##                              hu12_dep_so4_tavg_std
##                                       0.0010565536
##                               hu12_nlcd2011_pct_90
##                                       0.0010192021
##                     wlconnections_allwetlands_count
##                                       0.0009533889
##                             hu12_dep_so4_tavg_mean
##                                       0.0008328681
##             wlconnections_allwetlands_shoreline_km
##                                       0.0007595655
##                              hu12_dep_no3_tavg_std
##                                       0.0007353636
##                               hu12_nlcd2011_pct_71
##                                       0.0007014931
##     wlconnections_allwetlands_contributing_area_ha
##                                       0.0004512432
##                             hu12_dep_no3_tavg_mean
##                                       0.0004267939
```

```
predphi.st<-predict(cf.phist, newdata=rfdat.phist,type="response")
cor.test(predphi.st,cos(rfdat.phist$accndviphi.ts1))
```

```
##
##  Pearson's product-moment correlation
##
## data:  predphi.st and cos(rfdat.phist$accndviphi.ts1)
## t = 9.924, df = 41, p-value = 1.835e-12
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.7221922 0.9107674
## sample estimates:
##       cor
## 0.8402753
```

```
lwgt<-preds$tslength[coh.chlaXaccndvi$accndvip.ts1<0.3]/mean(preds$tslength[coh.chlaXaccndvi$accndvip.ts

gam.phist<-gam(cos(accndviphi.ts1) ~ s(maxdepth) + s(upstream_lakes_4ha_count) +
                 s(wlconnections_openwaterwetlands_contributing_area_),
               data=rfdat.phist, gamma=1, weights=lwgt)
gam.check(gam.phist)
```




```
##
## Method: GCV   Optimizer: magic
## Smoothing parameter selection converged after 13 iterations.
## The RMS GCV score gradient at convergence was 5.649181e-08 .
## The Hessian was positive definite.
## Model rank =  28 / 28
##
## Basis dimension (k) checking results. Low p-value (k-index<1) may
## indicate that k is too low, especially if edf is close to k'.
##
##                                                        k'  edf k-index
## s(maxdepth)                                          9.00 2.04    1.14
## s(upstream_lakes_4ha_count)                          9.00 1.00    1.18
## s(wlconnections_openwaterwetlands_contributing_area_) 9.00 1.00    0.92
##                                                      p-value
## s(maxdepth)                                             0.73
## s(upstream_lakes_4ha_count)                             0.85
## s(wlconnections_openwaterwetlands_contributing_area_)    0.17
```

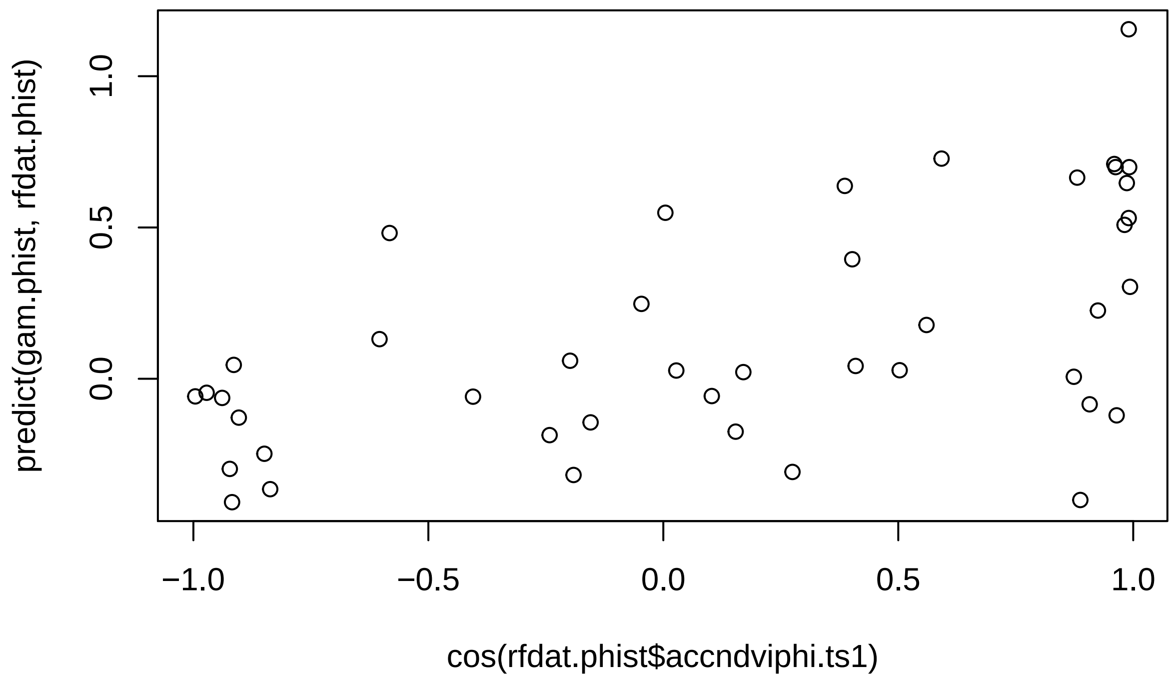
```
concurvity(gam.phist)
```

```
##                  para s(maxdepth) s(upstream_lakes_4ha_count)
## worst    4.979222e-11   0.9967232                   1.0000000
## observed 4.979222e-11   0.6967102                   0.7686527
## estimate 4.979222e-11   0.6517048                   0.7828092
##          s(wlconnections_openwaterwetlands_contributing_area_)
## worst                                                1.0000000
## observed                                             0.9144358
## estimate                                             0.9103020
```

```r
summary(gam.phist)
```

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## cos(accndviphi.ts1) ~ s(maxdepth) + s(upstream_lakes_4ha_count) +
##     s(wlconnections_openwaterwetlands_contributing_area_)
##
## Parametric coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  0.14545    0.09486   1.533    0.134
##
## Approximate significance of smooth terms:
##                                                         edf Ref.df     F
## s(maxdepth)                                           2.038  2.535 4.821
## s(upstream_lakes_4ha_count)                           1.000  1.000 0.479
## s(wlconnections_openwaterwetlands_contributing_area_) 1.000  1.000 0.722
##                                                       p-value
## s(maxdepth)                                           0.00774 **
## s(upstream_lakes_4ha_count)                           0.49303
## s(wlconnections_openwaterwetlands_contributing_area_) 0.40087
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## R-sq.(adj) =  0.256   Deviance explained = 32.7%
## GCV = 0.43804  Scale est. = 0.38671   n = 43
```

```r
plot(cos(rfdat.phist$accndviphi.ts1), predict(gam.phist, rfdat.phist))
```




```r
rfdat.philt<-left_join(coh.chlaXaccndvi[,c(10,5)], preds)
```

```
## Joining, by = "lagoslakeid"
```

```r
rfdat.philt<-rfdat.philt[,!colnames(rfdat.philt) %in%
                           c("lagoslakeid","start","end","lakes_nhdid","hu12_zoneid","tslength","county_
rfdat.philt<-rfdat.philt[,!grepl("borderhu12s",colnames(rfdat.philt))]

rfdat.philt<-rfdat.philt[coh.chlaXaccndvi$accndvip.ts2<0.3,]

for(nn in 1:ncol(rfdat.philt)){
  if(is.character(rfdat.philt[,nn])){
    rfdat.philt[,nn]<-as.factor(rfdat.philt[,nn])
  }
}

cf.philt<-party::cforest(cos(accndviphi.ts2) ~ ., data=rfdat.philt,
                         controls=cforest_control(ntree=50000,mincriterion = 0.9,mtry=3))

varimp.phi.lt<-varimp(cf.philt)
print(varimp.phi.lt[order(varimp.phi.lt, decreasing=T)][1:20])
```

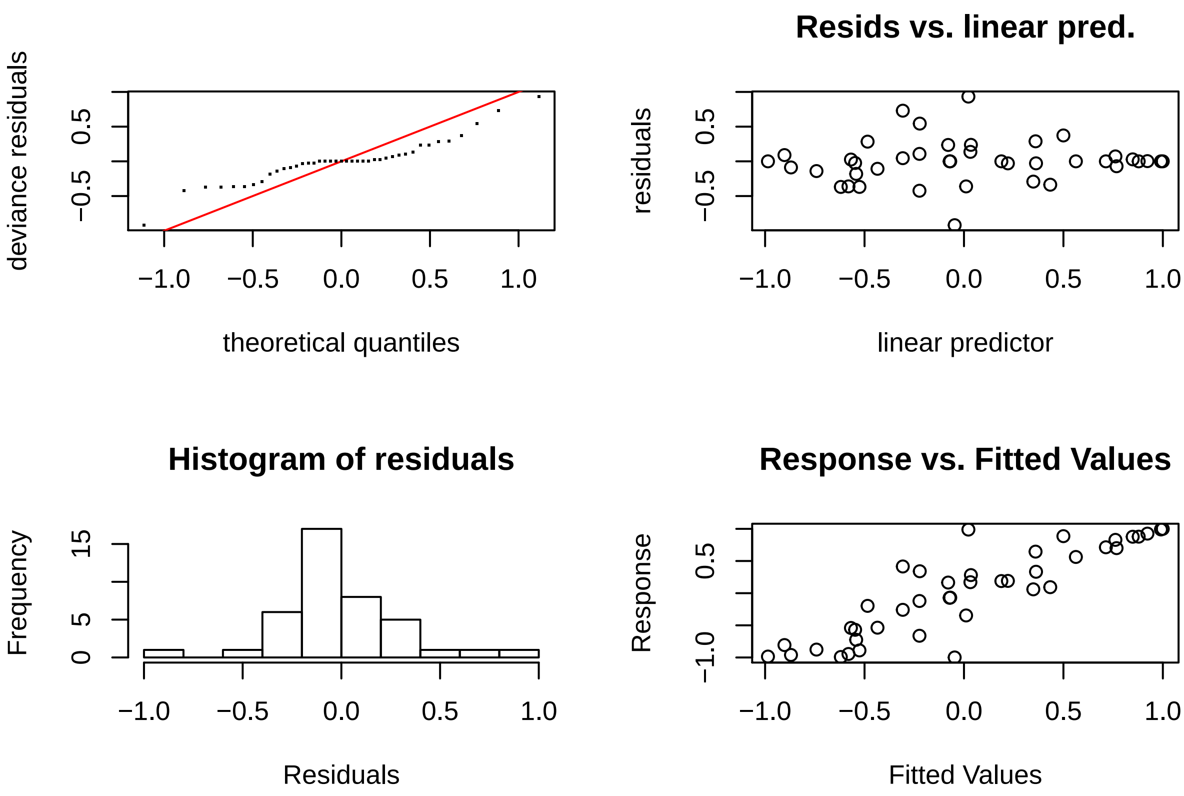
```
##                                         hu4_zoneid
##                                       0.0080082662
##                                         hu6_zoneid
##                                       0.0055070889
##                                         hu8_zoneid
##                                       0.0042286575
##                                      hu12_tri_mean
##                                       0.0024504229
##                                    hu12_slope_mean
##                                       0.0024487386
## buffer500m_streamdensity_headwaters_density_mperha
##                                       0.0021148623
##                                     lakeconnection
##                                       0.0019690957
##    buffer500m_streamdensity_streams_density_mperha
##                                       0.0019106856
##                                 hu12_nlcd2011_pct_71
##                                       0.0016058935
##      hu12_prism_tmin_30yr_normal_800mm2_annual_std
##                                       0.0015328718
##                                    hu12_runoff_std
##                                       0.0013653089
##    hu12_prism_tmean_30yr_normal_800mm2_annual_mean
##                                       0.0011474586
##                                 hu12_nlcd2011_pct_81
##                                       0.0010799255
##     hu12_prism_tmin_30yr_normal_800mm2_annual_mean
##                                       0.0010716364
## buffer500m_streamdensity_midreaches_density_mperha
##                                       0.0010514317
##      hu12_prism_tmax_30yr_normal_800mm2_annual_std
##                                       0.0010059379
##     hu12_prism_tmean_30yr_normal_800mm2_annual_std
##                                       0.0008226658
##                              hu12_dep_so4_tavg_mean
```

```
##                                                 0.0007866420
##                                        hu12_nlcd2011_pct_41
##                                                 0.0007640883
##                                        hu12_nlcd2011_pct_90
##                                                 0.0007429359
```

```
predphi.lt<-predict(cf.philt, newdata=rfdat.philt,type="response")
cor.test(predphi.lt,cos(rfdat.philt$accndviphi.ts2))
```

```
##
##  Pearson's product-moment correlation
##
## data:  predphi.lt and cos(rfdat.philt$accndviphi.ts2)
## t = 8.0713, df = 41, p-value = 5.294e-10
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.6317088 0.8773290
## sample estimates:
##       cor
## 0.7834154
```

```
lwgt<-preds$tslength[coh.chlaXaccndvi$accndvip.ts2<0.3]/mean(preds$tslength[coh.chlaXaccndvi$accndvip.t

gam.philt<-gam(cos(accndviphi.ts2) ~ hu4_zoneid + s(hu12_tri_mean) +
                 s(buffer500m_streamdensity_headwaters_density_mperha),
               data=rfdat.philt, gamma=1, weights=lwgt)
gam.check(gam.philt)
```




```
##
## Method: GCV   Optimizer: magic
```

```
## Smoothing parameter selection converged after 5 iterations.
## The RMS GCV score gradient at convergence was 8.419227e-07 .
## The Hessian was positive definite.
## Model rank =  38 / 38
##
## Basis dimension (k) checking results. Low p-value (k-index<1) may
## indicate that k is too low, especially if edf is close to k'.
##
##                                                        k'  edf k-index
## s(hu12_tri_mean)                                     9.00 1.61    0.89
## s(buffer500m_streamdensity_headwaters_density_mperha) 9.00 3.08    1.25
##                                                      p-value
## s(hu12_tri_mean)                                        0.18
## s(buffer500m_streamdensity_headwaters_density_mperha)    0.94
```

```r
concurvity(gam.philt)
```

```
##          para s(hu12_tri_mean)
## worst       1        1.0000000
## observed    1        0.9893772
## estimate    1        0.9685533
##          s(buffer500m_streamdensity_headwaters_density_mperha)
## worst                                                 1.0000000
## observed                                              0.8470862
## estimate                                              0.9666434
```

```r
summary(gam.philt)
```

```
##
## Family: gaussian
## Link function: identity
##
## Formula:
## cos(accndviphi.ts2) ~ hu4_zoneid + s(hu12_tri_mean) + s(buffer500m_streamdensity_headwaters_density_m
##
## Parametric coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      -0.01039    0.41312  -0.025    0.980
## hu4_zoneidHU4_12  0.23266    1.56218   0.149    0.883
## hu4_zoneidHU4_16  0.17390    1.06205   0.164    0.872
## hu4_zoneidHU4_24 -0.68575    0.60697  -1.130    0.275
## hu4_zoneidHU4_25 -0.41187    0.50616  -0.814    0.428
## hu4_zoneidHU4_27 -0.06844    0.56008  -0.122    0.904
## hu4_zoneidHU4_29  0.92773    0.61623   1.505    0.151
## hu4_zoneidHU4_30  0.25559    0.49004   0.522    0.609
## hu4_zoneidHU4_32 -0.11949    0.50093  -0.239    0.814
## hu4_zoneidHU4_33 -0.45347    0.64351  -0.705    0.491
## hu4_zoneidHU4_34  0.69963    0.73670   0.950    0.356
## hu4_zoneidHU4_4   1.72457    1.36129   1.267    0.223
## hu4_zoneidHU4_5  -0.30135    1.66814  -0.181    0.859
## hu4_zoneidHU4_51  0.52965    0.68886   0.769    0.453
## hu4_zoneidHU4_54  0.77799    0.56567   1.375    0.188
## hu4_zoneidHU4_60 -0.28665    0.53658  -0.534    0.600
## hu4_zoneidHU4_63  0.36168    0.59867   0.604    0.554
## hu4_zoneidHU4_64  0.45929    0.68133   0.674    0.510
```

```
## hu4_zoneidHU4_65 -0.46038    0.49235  -0.935    0.363
## hu4_zoneidHU4_67 -0.46011    1.89912  -0.242    0.812
##
## Approximate significance of smooth terms:
##                                                       edf Ref.df     F
## s(hu12_tri_mean)                                     1.61  2.014 0.397
## s(buffer500m_streamdensity_headwaters_density_mperha) 3.08  3.664 0.689
##                                                      p-value
## s(hu12_tri_mean)                                       0.696
## s(buffer500m_streamdensity_headwaters_density_mperha)   0.436
##
## R-sq.(adj) =   0.46   Deviance explained =   78%
## GCV = 0.61573  Scale est. = 0.24495   n = 41
```

```r
plot(cos(rfdat.philt$accndviphi.ts2), predict(gam.philt, rfdat.philt))
```