Appendix A

- The `AnalystHelper` R-package documentation can be found on Github (https://github.com/SwampThingPaul/AnalystHelper). All other packages are available on CRAN.

- Data used in this assessment was retrieved from FDEP's IWR database found here: http://publicfiles.dep.state.fl.us/DEAR/IWR/.

R code used in this analysis.

```r
# Libraries
library(AnalystHelper)
library(plyr)
library(reshape2)
library(openxlsx)
library(zoo)
library(EnvStats)
library(lmtest)
library(nortest)
library(gvlma)


data.path="./iwr2020_run60_2020-10-05/"


# -----------------------------------------------------------------------
dat.qual=data.frame(QUALIFIER=c(NA,"!","A","D","E","F","I","R","T","U"
                                ,"*","?","B","H","J","K","L","M","N",
                                "O","Q","V","Y","Z"),
                    FATALYN=c("N",rep("N",9),rep("Y",14)))

# WBID3240F - Daughtrey -------------------------------------------------
# Data exported from the IWR database as an xlsx file
dat3240F=read.xlsx(paste0(data.path,"periodOfRecordData3240F.xlsx"))
dat3240F$date=date.fun(convertToDate(dat3240F$date))
dat3240F=subset(dat3240F,year%in%seq(2006,2020,1))
# site not included in original analysis
dat3240F=subset(dat3240F,station.id!="112WRD 264140081494600");

# QA/QC qualifiers
unique(dat3240F$rcode)
unique(dat3240F$xcode)
quals=as.character(unique(dat3240F$xcode))
spl=strsplit(quals,split="")
quals=data.frame(xcode=quals,
                 q1=sapply(spl,"[",1),
                 q2=sapply(spl,"[",2),
                 q3=sapply(spl,"[",3))
quals$Fatal=with(quals,
                 ifelse(q1%in%subset(dat.qual,FATALYN=="Y")$QUALIFIER|
                        q2%in%subset(dat.qual,FATALYN=="Y")$QUALIFIER|
                        q3%in%subset(dat.qual,FATALYN=="Y")$QUALIFIER,
                        "Y","N"))
dat3240F=merge(dat3240F,quals[,c("xcode","Fatal")],"xcode",all.x=T)
dat3240F=subset(dat3240F,Fatal=="N")
dat3240F$HalfMDL=with(dat3240F,
```

```r
                    ifelse(is.na(mdl)==T,result,
                          ifelse(result<=mdl,result/2,result)))

params.keep=c("CHLAC","COLOR","COND",
             "DO","DOSAT","TN","TP","TOC","TEMP")
dat3240F.xtab=dcast(subset(dat3240F,master.code%in%params.keep),
                          wbid+station.id+lat+long+date~master.code,
                          value.var="HalfMDL",mean)
dat3240F.xtab$Sal=with(dat3240F.xtab,SalinityCalc(COND,TEMP))
dat3240F.xtab$DOsat.calc=with(dat3240F.xtab,DO_PerSat(TEMP,DO,Sal))
dat3240F.xtab$DoY=as.numeric(format(dat3240F.xtab$date,"%j"))
dat3240F.xtab$CY=as.numeric(format(dat3240F.xtab$date,"%Y"))
dat3240F.xtab$month=as.numeric(format(dat3240F.xtab$date,'%m'))

## Spatially Average Data
dat3240F.xtab.mean=ddply(dat3240F.xtab,"date",summarise,
                        mean.DOSat=mean(DOsat.calc,na.rm=T))
dat3240F.xtab.mean$DoY=as.numeric(format(dat3240F.xtab.mean$date,"%j"))
dat3240F.xtab.mean$CY=as.numeric(format(dat3240F.xtab.mean$date,'%Y'))
dat3240F.xtab.mean=subset(dat3240F.xtab.mean,is.na(mean.DOSat)==F)
dat3240F.xtab.mean$dum.val=1:nrow(dat3240F.xtab.mean) #time index

# WQS Eval
dat3240F$time=with(dat3240F,ifelse(nchar(time)==3,
                              paste0(0,time),time))
dat3240F$datetime=with(dat3240F,
                    date.fun(paste(date,time),form="%F %H%M"))
dat3240F.xtab2=dcast(subset(dat3240F,
                        master.code%in%params.keep&
                          station.id!="112WRD  264140081494600"),
                    wbid+station.id+lat+long+datetime~master.code,
                    value.var="HalfMDL",mean)
dat3240F.xtab2$CY=as.numeric(format(dat3240F.xtab2$date,"%Y"))
dat3240F.xtab2$Sal=with(dat3240F.xtab2,SalinityCalc(COND,TEMP))
dat3240F.xtab2$DOsat.calc=with(dat3240F.xtab2,DO_PerSat(TEMP,DO,Sal))
dat3240F.xtab2$DO.TOD.WQS=with(dat3240F.xtab2,
                              DO.TOD.WQS.stream(datetime))
dat3240F.xtab2$exceed=with(dat3240F.xtab2,
                        ifelse(DOsat.calc<DO.TOD.WQS,1,0))

rslt.3240F=ddply(dat3240F.xtab2,c("station.id","CY"),
              summarise,N.exceed=sum(exceed,na.rm=T),
              N.val=N.obs(DOsat.calc))
rslt.3240F$PerExceed=with(rslt.3240F,N.exceed/N.val)*100
rslt.3240F$status=with(rslt.3240F,ifelse(PerExceed>10,1,0))
rslt.3240F=ddply(rslt.3240F,"station.id",summarise,
              sum.status=sum(status),n.val=N.obs(status))

# Kendall Trend
ken.rslt=with(dat3240F.xtab.mean,
            cor.test(mean.DOSat,as.numeric(date),method="kendall"))

# Seasonal Kendall
```

```r
dat3240F.xtab.mean.sea=ddply(dat3240F.xtab,c("month","CY"),summarise,
                             mean.DOSat=mean(DOsat.calc,na.rm=T))
sea.rslt=kendallSeasonalTrendTest(mean.DOSat~month+CY,
                                  data=dat3240F.xtab.mean.sea)
print(sea.rslt)

# Individual station Kendall Trend
samp.screen=ddply(dat3240F.xtab,"station.id",summarise,
                  N.val=N.obs(DOsat.calc))
samp.screen=subset(samp.screen,N.val>20)
dat3240F.trend=subset(dat3240F.xtab,
                      station.id%in%samp.screen$station.id)
dat3240F.trend$date.num=as.numeric(dat3240F.trend$date)
site.trend=ddply(dat3240F.trend,"station.id",summarise,
                 N.val=N.obs(DOsat.calc,"NaN"),
                 est=cor.test(DOsat.calc,date.num,
                              method="kendall")$estimate,
                 pval=cor.test(DOsat.calc,date.num,
                               method="kendall")$p.value)

# lm trend
dep.trend=lm(mean.DOSat~date,dat3240F.xtab.mean)
summary(dep.trend)
gvlma(dep.trend)
layout(matrix(1:4,2,2));plot(dep.trend)

bgtest(dep.trend,order=3)
shapiro.test(residuals(dep.trend))

## Data distribution
tmp=subset(dat3240F.xtab,is.na(DOsat.calc)==F)$DOsat.calc
shapiro.test(tmp)
ad.test(tmp)

plotdist(tmp,"norm",para=list(mean=mean(tmp),sd=sd(tmp)))

# WBID3240Q - Popash  ----------------------------------------------
dat3240Q=read.xlsx(paste0(data.path,"periodOfRecordData3240Q.xlsx"))
dat3240Q$date=date.fun(convertToDate(dat3240Q$date))
dat3240Q=subset(dat3240Q,year%in%seq(2006,2020,1))

# QA/QC qualifiers
unique(dat3240Q$rcode)
unique(dat3240Q$xcode)
quals=as.character(unique(dat3240Q$xcode))
spl=strsplit(quals,split="")
quals=data.frame(xcode=quals,
                 q1=sapply(spl,"[",1),
                 q2=sapply(spl,"[",2),
                 q3=sapply(spl,"[",3))
quals$Fatal=with(quals,
                 ifelse(q1%in%subset(dat.qual,FATALYN=="Y")$QUALIFIER|
                        q2%in%subset(dat.qual,FATALYN=="Y")$QUALIFIER|
```

```r
                              q3%in%subset(dat.qual,FATALYN=="Y")$QUALIFIER,
                              "Y","N"))
dat3240Q=merge(dat3240Q,quals[,c("xcode","Fatal")],"xcode",all.x=T)
dat3240Q=subset(dat3240Q,Fatal=="N")
dat3240Q$HalfMDL=with(dat3240Q,
                      ifelse(is.na(mdl)==T,result,
                              ifelse(result<=mdl,result/2,result)))

params.keep=c("CHLAC","COLOR","COND",
               "DO","DOSAT","TN","TP","TOC","TEMP")
dat3240Q.xtab=dcast(subset(dat3240Q,master.code%in%params.keep),
                              wbid+station.id+lat+long+date~master.code,
                              value.var="HalfMDL",mean)
dat3240Q.xtab$Sal=with(dat3240Q.xtab,SalinityCalc(COND,TEMP))
dat3240Q.xtab$DOsat.calc=with(dat3240Q.xtab,DO_PerSat(TEMP,DO,Sal))
dat3240Q.xtab$DoY=as.numeric(format(dat3240Q.xtab$date,"%j"))
dat3240Q.xtab$CY=as.numeric(format(dat3240Q.xtab$date,"%Y"))
dat3240Q.xtab$month=as.numeric(format(dat3240Q.xtab$date,'%m'))

## Spatially Average Data
dat3240Q.xtab.mean=ddply(dat3240Q.xtab,"date",summarise,
                              mean.DOSat=mean(DOsat.calc,na.rm=T))
dat3240Q.xtab.mean$DoY=as.numeric(format(dat3240Q.xtab.mean$date,"%j"))
dat3240Q.xtab.mean$CY=as.numeric(format(dat3240Q.xtab.mean$date,'%Y'))
dat3240Q.xtab.mean=subset(dat3240Q.xtab.mean,is.na(mean.DOSat)==F)
dat3240Q.xtab.mean$dum.val=1:nrow(dat3240Q.xtab.mean) #time index

# WQS Eval
dat3240Q$time=with(dat3240Q,ifelse(nchar(time)==3,
                                    paste0(0,time),time))
dat3240Q$datetime=with(dat3240Q,
                       date.fun(paste(date,time),form="%F %H%M"))
dat3240Q.xtab2=dcast(subset(dat3240Q,
                              master.code%in%params.keep),
                     wbid+station.id+lat+long+datetime~master.code,
                     value.var="HalfMDL",mean)
dat3240Q.xtab2$CY=as.numeric(format(dat3240Q.xtab2$date,"%Y"))
dat3240Q.xtab2$Sal=with(dat3240Q.xtab2,SalinityCalc(COND,TEMP))
dat3240Q.xtab2$DOsat.calc=with(dat3240Q.xtab2,DO_PerSat(TEMP,DO,Sal))
dat3240Q.xtab2$DO.TOD.WQS=with(dat3240Q.xtab2,
                                DO.TOD.WQS.stream(datetime))
dat3240Q.xtab2$exceed=with(dat3240Q.xtab2,
                            ifelse(DOsat.calc<DO.TOD.WQS,1,0))

rslt.3240Q=ddply(dat3240Q.xtab2,c("station.id","CY"),
                  summarise,N.exceed=sum(exceed,na.rm=T),
                  N.val=N.obs(DOsat.calc))
rslt.3240Q$PerExceed=with(rslt.3240Q,N.exceed/N.val)*100
rslt.3240Q$status=with(rslt.3240Q,ifelse(PerExceed>10,1,0))
rslt.3240Q=ddply(rslt.3240Q,"station.id",summarise,
                  sum.status=sum(status),n.val=N.obs(status))

# Kendall Trend
```

```r
ken.rslt=with(dat3240Q.xtab.mean,
             cor.test(mean.DOSat,as.numeric(date),method="kendall"))

# Seasonal Kendall
dat3240Q.xtab.mean.sea=ddply(dat3240Q.xtab,c("month","CY"),summarise,
                            mean.DOSat=mean(DOsat.calc,na.rm=T))
sea.rslt=kendallSeasonalTrendTest(mean.DOSat~month+CY,
                                 data=dat3240Q.xtab.mean.sea)
print(sea.rslt)

# Individual station Kendall Trend
samp.screen=ddply(dat3240Q.xtab,"station.id",summarise,
                 N.val=N.obs(DOsat.calc))
samp.screen=subset(samp.screen,N.val>20)
dat3240Q.trend=subset(dat3240Q.xtab,
                     station.id%in%samp.screen$station.id)
dat3240Q.trend$date.num=as.numeric(dat3240Q.trend$date)
site.trend=ddply(dat3240Q.trend,"station.id",summarise,
                N.val=N.obs(DOsat.calc,"NaN"),
                est=cor.test(DOsat.calc,date.num,
                            method="kendall")$estimate,
                pval=cor.test(DOsat.calc,date.num,
                             method="kendall")$p.value)


# lm trend
dep.trend=lm(mean.DOSat~date,dat3240Q.xtab.mean)
summary(dep.trend)
gvlma(dep.trend)
layout(matrix(1:4,2,2));plot(dep.trend)

bgtest(dep.trend,order=3)
shapiro.test(residuals(dep.trend))

## Data distribution
tmp=subset(dat3240Q.xtab,is.na(DOsat.calc)==F)$DOsat.calc
shapiro.test(tmp)
ad.test(tmp)

plotdist(tmp,"norm",para=list(mean=mean(tmp),sd=sd(tmp)))

# WBID3235C - Cypress ---------------------------------------------
dat3235C=read.xlsx(paste0(data.path,"periodOfRecordData3235C.xlsx"))
dat3235C$date=date.fun(convertToDate(dat3235C$date))
dat3235C=subset(dat3235C,year%in%seq(2006,2020,1))

# Sites in IWR databased within Cypress Creek
cypresscreeksites=c("21FLBABRCYPRESS_HEAD",
                   "21FLBABRCYPRESS_OUTFLOW",
                   "21FLFTM 28020237",
                   "21FLGW  56335",
                   "21FLFTM G3SD0084")
# data discrepancies
# dat3235C=subset(dat3235C,station.id%in%cypresscreeksites)
```

```r
# QA/QC qualifiers
unique(dat3235C$rcode)
unique(dat3235C$xcode)
quals=as.character(unique(dat3235C$xcode))
spl=strsplit(quals,split="")
quals=data.frame(xcode=quals,
                 q1=sapply(spl,"[",1),
                 q2=sapply(spl,"[",2),
                 q3=sapply(spl,"[",3))
quals$Fatal=with(quals,
                 ifelse(q1%in%subset(dat.qual,FATALYN=="Y")$QUALIFIER|
                        q2%in%subset(dat.qual,FATALYN=="Y")$QUALIFIER|
                        q3%in%subset(dat.qual,FATALYN=="Y")$QUALIFIER,
                        "Y","N"))
dat3235C=merge(dat3235C,quals[,c("xcode","Fatal")],"xcode",all.x=T)
dat3235C=subset(dat3235C,Fatal=="N")
dat3235C$HalfMDL=with(dat3235C,
                      ifelse(is.na(mdl)==T,result,
                             ifelse(result<=mdl,result/2,result)))


params.keep=c("CHLAC","COLOR","COND",
              "DO","DOSAT","TN","TP","TOC","TEMP")
dat3235C.xtab=dcast(subset(dat3235C,master.code%in%params.keep),
                    wbid+station.id+lat+long+date~master.code,
                    value.var="HalfMDL",mean)
dat3235C.xtab$Sal=with(dat3235C.xtab,SalinityCalc(COND,TEMP))
dat3235C.xtab$DOsat.calc=with(dat3235C.xtab,DO_PerSat(TEMP,DO,Sal))
dat3235C.xtab$DoY=as.numeric(format(dat3235C.xtab$date,"%j"))
dat3235C.xtab$CY=as.numeric(format(dat3235C.xtab$date,"%Y"))
dat3235C.xtab$month=as.numeric(format(dat3235C.xtab$date,'%m'))


## Spatially Average Data
dat3235C.xtab.mean=ddply(dat3235C.xtab,"date",summarise,
                         mean.DOSat=mean(DOsat.calc,na.rm=T))
dat3235C.xtab.mean$DoY=as.numeric(format(dat3235C.xtab.mean$date,"%j"))
dat3235C.xtab.mean$CY=as.numeric(format(dat3235C.xtab.mean$date,'%Y'))
dat3235C.xtab.mean=subset(dat3235C.xtab.mean,is.na(mean.DOSat)==F)
dat3235C.xtab.mean$dum.val=1:nrow(dat3235C.xtab.mean) #time index


# WQS Eval
dat3235C$time=with(dat3235C,ifelse(nchar(time)==3,
                                   paste0(0,time),time))
dat3235C$datetime=with(dat3235C,
                       date.fun(paste(date,time),form="%F %H%M"))
dat3235C.xtab2=dcast(subset(dat3235C,
                            master.code%in%params.keep),
                     wbid+station.id+lat+long+datetime~master.code,
                     value.var="HalfMDL",mean)
dat3235C.xtab2$CY=as.numeric(format(dat3235C.xtab2$date,"%Y"))
dat3235C.xtab2$Sal=with(dat3235C.xtab2,SalinityCalc(COND,TEMP))
dat3235C.xtab2$DOsat.calc=with(dat3235C.xtab2,DO_PerSat(TEMP,DO,Sal))
dat3235C.xtab2$DO.TOD.WQS=with(dat3235C.xtab2,
                              DO.TOD.WQS.stream(datetime))
```

```r
dat3235C.xtab2$exceed=with(dat3235C.xtab2,
                            ifelse(DOsat.calc<DO.TOD.WQS,1,0))

rslt.3235C=ddply(dat3235C.xtab2,c("station.id","CY"),
                summarise,N.exceed=sum(exceed,na.rm=T),
                N.val=N.obs(DOsat.calc))
rslt.3235C$PerExceed=with(rslt.3235C,N.exceed/N.val)*100
rslt.3235C$status=with(rslt.3235C,ifelse(PerExceed>10,1,0))
rslt.3235C=ddply(rslt.3235C,"station.id",summarise,
                sum.status=sum(status),n.val=N.obs(status))

# Kendall Trend
ken.rslt=with(dat3235C.xtab.mean,
              cor.test(mean.DOSat,as.numeric(date),method="kendall"))

# Seasonal Kendall
dat3235C.xtab.mean.sea=ddply(dat3235C.xtab,c("month","CY"),summarise,
                              mean.DOSat=mean(DOsat.calc,na.rm=T))
sea.rslt=kendallSeasonalTrendTest(mean.DOSat~month+CY,
                                  data=dat3235C.xtab.mean.sea)
print(sea.rslt)

# Individual station Kendall Trend
samp.screen=ddply(dat3235C.xtab,"station.id",summarise,
                  N.val=N.obs(DOsat.calc))
samp.screen=subset(samp.screen,N.val>20)
dat3235C.trend=subset(dat3235C.xtab,
                      station.id%in%samp.screen$station.id)
dat3235C.trend$date.num=as.numeric(dat3235C.trend$date)
site.trend=ddply(dat3235C.trend,"station.id",summarise,
                N.val=N.obs(DOsat.calc,"NaN"),
                est=cor.test(DOsat.calc,date.num,
                              method="kendall")$estimate,
                pval=cor.test(DOsat.calc,date.num,
                              method="kendall")$p.value)

# lm trend
dep.trend=lm(mean.DOSat~date,dat3235C.xtab.mean)
summary(dep.trend)
gvlma(dep.trend)
layout(matrix(1:4,2,2));plot(dep.trend)

bgtest(dep.trend,order=3)
shapiro.test(residuals(dep.trend))

## Data distribution
tmp=subset(dat3235C.xtab,is.na(DOsat.calc)==F)$DOsat.calc
shapiro.test(tmp)
ad.test(tmp)

plotdist(tmp,"norm",para=list(mean=mean(tmp),sd=sd(tmp)))
```