Status and Trend Monitoring Networks
Trend Analysis Protocols

Division of Environmental Assessment and Restoration

Watershed Monitoring Section

Florida Department of Environmental Protection

June 2020

2600 Blair Stone Rd, MS 3560

Tallahassee, Florida 32399-2400

https://floridadep.gov/

Table of Contents

Introduction

This document describes the methods used by the Watershed Monitoring Section (WMS) for the determination of statewide and region-wide water-quality trends utilizing data generated by the WMS' Status and Trends Monitoring Networks. Complete design and operation details for these networks are provided in the Florida Watershed Monitoring Status and Trend Program Design Document. A brief summary of each network follows.

Initiated in 1998, the Trend Monitoring Network (Trend Network) was designed specifically to track water-quality trends at individual river, stream, canal, and aquifer (via wells) stations over time. To achieve this goal, fixed locations are sampled at fixed intervals (monthly or quarterly). It consists of 129 stations across the state; 78 surface water, and 51 groundwater sites. The Trend Network complements the Status Network by providing spatial and temporal information about water resources and potential changes from anthropogenic or natural influences, including extreme events (e.g., droughts and hurricanes).

The Status Monitoring Network (Status Network) was designed specifically to provide a 'snapshot' of current water quality. It was initiated in 2000 and provides an unbiased, cost-effective sampling of the state's water resources by utilizing an EPA-developed probabilistic design. It is used to provide information on regional and statewide conditions for seven water resources: large and small lakes, rivers, streams, canals and unconfined and confined groundwater (via wells). Since 2009, a statewide sample survey (cycle) for all water resources is completed annually. Historically two cycles were completed using a rotating basin approach. The first cycle took four years with sampling during 2000-2003; the second took five years with sampling during 2004-2008. Conditions from one statewide sampling cycle may be compared to another for the determination of statewide or region-wide water-quality trends.

Both of these monitoring network designs provide estimates of water-quality indicators with known statistical confidence. Data produced by the Status and Trend Networks fulfill CWA 305(b) reporting needs and complement CWA 303(d) reporting.

Trend Tests Utilized by WMS

Helsel and Hirsch (2002) categorize trend tests into those using data collected throughout a single period (monotonic trends) and those comparing data collected in two or more non-overlapping periods (step trends). DEP uses Trend Network (monotonic) data for trend determination at individual stations and for

region-wide trends. Additionally, Status Network data collected in an early and a late statewide sample cycle (step) can be evaluated for determination of region-wide and statewide trends.

The following methods are used as needed to identify water-quality changes over time (trend detection):

1. Seasonal Kendall (SK) test for individual station water-quality indicator trend detection. This analysis is run every four years for all trend network stations having sufficient data.
2. Regional Kendal (RK) test for statewide evaluation of non-flow adjusted surface water stations. The test aggregates the results of the individual SK tests into a single result for the region of interest. This analysis is run every four years and includes all trend network stations having sufficient data. It also supplements results derived from the Change Analysis described below.
3. Change Analysis (CHAN) for region-wide water-quality indicator trend detection. This is a step trend analysis comparing early cycles to later cycles. This analysis can be run periodically for flowing surface waters, lakes, and groundwater, as management needs dictate.

Data Examination / Preparation Prior to Trend Analysis

The WMS conducts a series of procedural steps on the data prior to any of the trend analyses identified above are conducted. The steps are:

1. Examine data for spurious values

Data qualifiers are used to determine the suitability of indicator values for analyses. These qualifiers provide additional information on measurement values and are referenced by codes. Data may be qualified for a variety of reasons described in **Appendix A** (from 2018 revision of Chapter 62-160.700, F.A.C.). **Note:** Prior to October 1, 2017 the qualifier definitions and codes used were those listed in the 2014 revision of Chapter 62-160.700 F.A.C. For the analyses presented in this document any data qualified with codes "**O**" and "**?**" are removed from analysis. Documenting the data analyst's use of data qualifiers in corresponding data analysis reports is required to understand the results.

2. Identify spurious data and statistically describe the analytes / indicators

WMS relies on nonparametric statistics for most water-quality data analyses. Nonparametric statistics do not rely on data belonging to a particular distribution. These techniques rank data from smallest to highest values and then use these ranks to describe and analyze the data. This process is insensitive to spurious outliers and missing values, which are common in environmental data. Nevertheless, outliers should be examined to determine if they represent

data errors. If there are errant data, a decision must be made to include or exclude them from analysis.

One method of determining outliers is the Tukey method (Agresti and Franklin 2013). This method is a nonparametric method which utilizes the first (Q1) and third (Q3) quartiles of a dataset to determine the interquartile (Q3 - Q1) range. Upper outliers are identified as values greater than [(Q3) + (1.5 * IQR)]. Lower outliers are values less than [(Q1) - (1.5 * IQR)]. Each outlier may then be considered a point in need of evaluation (PINE). These PINES need to be examined to determine if they are valid values. They are shared with the WMS's Quality Assurance Officer (QAO) and Data Manager to determine if the PINES are data entry errors. Once the PINES are examined by the QAO and DM, the Data Analyst (DA) is notified and determines whether to include or exclude the data.

The DA creates tables listing nonparametric descriptive statistics for each station's analytes (**Appendix B**). These tables allow the analyst to examine data distributions prior to analysis. In **Appendix B**, the reported minimum concentration value reflects the minimum detection level reported by the laboratory. Limitations of laboratory instruments can cause low concentrations of analytes to be undetected. This is referred to as data censorship. When the analyte level in a sample is lower than the minimum detection level of an instrument, the measurement is marked as BDL (below detection limit). The DEP laboratory marks censored / BDL measurements with the qualifiers **U** and **T** (**Appendix A**)

During analyses the DA replaces BDL values with either the highest reported minimum detection level of the instrument, a maximum likelihood estimation (MLE), or the minimum value in the period of record. If a period of record is missing a data point, the statistics (median, quartiles, ranks) are based on the number of actual values in the record. For example, if the December sample is missing from a monthly dataset, the median for the year is based only on the January-November values.

3. Determine if seasonality exists in the time series data

For time-series data, such as those obtained from the Trend Monitoring network, seasonal cycles, such as wet and dry periods, can mask trends in water-quality indicators. Water constituents commonly show seasonal cycles (seasonality); therefore, an effort should be made to collect at least one sample in each season, or four per year at a minimum. Statistically valid trends may be determined only after adjusting the data for seasonality. This adjustment is referred to as de-seasonalizing the data.

The Seasonal Kendall (SK) and Regional Kendall (RK) tests (see below,) remove effects of seasonal cycles. For monthly data, the conducted tests assume that each month is a season. For

quarterly data, the conducted tests assume that each quarter is a season. The quarters are: (1) December–February, (2) March–May, (3) June–August, and (4) September–November.

Analyses - Trend Network

Determine if linear time series trends exist for each indicator at each site

During the conceptual development of the Trend Monitoring Network, there was a concern that a linear trend in a time series might be found for one period, while another trend might be found for a different period at the same station. For example, suppose an upward trend was found for the first half of a time series, while a downward trend was found for the second half, or vice versa. Detectable trends depend on the period for which they are analyzed. Thus, if the linear, monotonic trend tests indicate the existence of a trend, it is only representative of the time series period of record.

Seasonal Kendall (SK) Test The statistical analyses used to evaluate data from the Trend Monitoring Network require a long period of record to be meaningful. WMS uses these sampling frequencies to determine trends: 1) monthly field data collection and laboratory analyses for rivers, streams, and canals; 2) monthly field data collection for unconfined aquifers; and 3) quarterly (once every three months) laboratory analyses for confined and unconfined aquifers, and field data collection for confined aquifers.

Gilbert (1987) stated that when testing for trends using time series data, variations added by regularly spaced cycles make it more difficult to detect trends if they exist. Regarding environmental data, Gilbert mentioned that major cycles are often referred to as seasonality. To address this issue, Hirsch and Slack (1984) developed the SK test. It removes the effect of the seasonal cycles prior to analysis. DEP uses the SK test to look for trends for each indicator at each surface water and groundwater Trend Network site. R software (R Core Team 2017) and the kendallSeasonalTrendTest function in the EnvStats R package (Millard 2013) are used to perform these analyses. An example script is provided in **Appendix C SK**.

As with seasonality, it is more difficult to detect trends in flowing surface waters with highly variable flow rates. Where available, mean daily flow rate data from nearby USGS gauging stations are associated with the surface water samples collected on the same date. With these added measurements, DEP can adjust surface water-quality data for flow before conducting the SK trend analyses. In contrast, groundwater flow rates generally are much slower, and DEP makes no flow adjustments prior to performing the SK analyses for groundwater.

If a trend exists for either flow-adjusted or nonflow-adjusted data, DEP determines the corresponding direction of the trend by using the Sen Slope (SS) estimator which measures the median difference between all observations over the time series (Gilbert 1987). The SS provides an estimate of the magnitude of change for a water-quality indicator over the period of record. Reporting a trend as increasing or decreasing indicates the direction of the slope and does not necessarily indicate impairment or improvement of the analyte being measured.

The SK test allows for missing values, while BDLs are assigned an arbitrary value as previously discussed. Each test is run as a two-sided test with the null hypothesis indicating no monotonic linear slope present in the concentration. The α-level for determining the significance of the results is 0.05.

If there are insufficient data (ISD), trend analyses are not conducted. A time series is considered to have ISD if one or more of the following reasons apply:

(1) If there are less than 10 observations in the entire time series,
(2) If a data gap in the time series represents ≥ 10% of the entire number of observations in the series,
(3) If missing values cause the margin of error to exceed 15%.

For an individual analyte, the results of a time-series analysis can result in a positive trend (numerical values increase over time), a negative trend (numerical values decrease over time), and insufficient evidence of a trend (values do not change over time) or no trend present. Insufficient evidence of a trend can result from one of three following situations:

(1) The resulting probability level of the trend analysis test is > 0.05,
(2) The resulting probability level of the test is ≤ 0.05, but the Sen-Slope estimator (see below) = 0,
(3) If the lowering of laboratory detection levels (DLs) over the period of record is believed have artificially lowered the resulting probability level of the test.

Smoothing procedures make visual inspections of time series plots easier to interpret. For each trend, WMS evaluates possible reasons for an analyte trend. If the trend is believed to be human-induced, staff discuss plausible reasons for the trend.

Determine if regional time series trends exist for water-quality indicators over the region of interest

Regional Kendall (RK) Analyses The RK test was developed by Helsel and Frans (2006) as an extension of the SK test. The RK test conducts an SK test and computes a Sen slope estimate (Sen

1968) at each site over an entire region. The test then combines the results of each SK test, determines an overall p-value of the test, and determines an overall Sen slope for the region. However, the regional Sen slope should only be used to indicate direction of the trend and not the magnitude when the RK test indicates a regional trend.

DEP uses the RK test to look for statewide trends for each indicator for the 78, non-flow adjusted, surface water trend sites. DEP uses R software (R Core Team 2017) and the rkt function in the Regional Kendall R package (Marchetto 2017) to perform these analyses. An example script is found in **Appendix C. RK**.

At each site, the RK test needs a minimum number of four values. However, if a minimum of 10 years of data are available, adjustments for spatial autocorrelation can be made. The RK test works best if the product of the number of sites times the number of years is greater than 25 (Helsel and Frans 2006). In addition to adjusting for seasonality, the "rkt" package of R has an option to adjust for the effects of spatial autocorrelation. WMS uses this option is used for each RK test.

If a regional or statewide trend exists, its direction is determined using the Sen-Slope (S-S) estimator. For each trend, WMS evaluates possible reasons for an analyte trend. If the trend is believed to be human-induced, staff discuss plausible reasons for the trend.

Analyses - Status Network

Determine if water quality indicator step trends are present

Change Analysis Change analysis (CHAN) is a step trend procedure for Status Network data which may be used to compare summarized data from one period with those from another period. Helsel et al. (2020) mention that a step trend compares two non-overlapping data sets, an early and late period of record. Changes between the periods are called step trends as values of *Y* (in this case, indicator values) step up or down from one time period to the next. The CHAN is fully described in Kincaid and Olsen (2019). DEP uses The Change Analysis function in R software's (R Core Team 2017) package spsurvey (Kincaid and Olsen 2019). The DA writes individual R scripts for each water resource analyzed.

Prior to running CHAN on Status Network data, several data preparatory steps are necessary, as follows:

1. ESRI's ArcGIS is used to determine site locations from the early period's sampled sites that do not fall onto the extent of water resource coverages used for the late time period. Sites that do not fall on the more current resource coverage are removed from the sampled sites list for the early period. The target populations of flowing waters and lakes for statewide sample surveys conducted during 2000-2008 (cycles 1 &2) differ from those for cycles beginning in 2009. For comparison the three current flowing water resources and two lake resources must be combined into two datasets: flowing waters and lakes. This also increases the sample sizes of the respective datasets, as a relatively large number of sites are removed from the early period because they were in locations excluded from more recent coverages. This is unnecessary for confined and unconfined aquifer wells, as their target population definitions have remained stable since the inception of the Status Network
2. R software projects are created for each resource for which CHAN is to be run and data from the remaining early period sites that were not excluded by Step 1 are then imported from the production Oracle database into R dataframes.
3. Data from the more current time period for each water resources are then imported into each of the dataframes created in Step 2. A factor column defining the time period is added to each of the dataframes and populated.
4. Analyte/indicator value distributions are then examined using the methods described above in **Data Examination / Preparation Prior to Trend Analysis** by using the R package EnvStats (Millard 2013). Unusual values are tagged as "points in need of examination" and provided to the data manager and QA officer for review.
5. R scripts are written for each of the water resources to provide data distributions. These scripts call and, therefore, run the change analyses utilizing the R package spsurvey's CHAN function, and export the analysis results to provide graphics by utilizing the R package ggplot2 (Wickham 2016).

Next, the CHAN test is used to generate spatial-weight matrices for each water resource by using the extent of each water resource in each of the six Status Network reporting units. Once done, the change analysis function then generates statewide spatially weighted mean and median values, plus upper and lower confidence bounds (CBs) for water quality indicators from both the early period and the late period. An example script is located in **Appendix C. CHAN**.

For surface water, the minimum subset of indicators to be examined included Chl-*a*, pH, DO, SC, Temp, NO_3-NO_2, TKN, TN, TP, and TOC. Except for Chl-*a*, the minimum subset of indicators for confined and unconfined aquifers is the same as those for surface water.

The CHAN test calculates the difference in spatially weighted means for each indicator between the early (E) and late (L) periods and a 95 % confidence interval for the difference. These values are

referred to as difference estimates. If the CBs for the difference estimates between the two periods (L minus E) do not include 0, there is statistical evidence that the values differ. Note that CHAN does not calculate p-values. However, if 0 does not lie within the 95 % CBs for the difference estimate, the p-value is known to be < 0.05.

Summarize Major Water Quality Issues

Using appropriate graphics and the results from the trend analyses tests, WMS summarizes the major water quality issues and reports them annually to the department and biennially to the US EPA. These summaries can lead to recommendations on future sampling of both surface waters and groundwaters.

References

Agresti, A. and .C. Franklin. 2013. Statistics: The Art and Science of Learning from Data. 3rd Edition. Pearson Education, Inc. ISBN 0-321-75594-4.

Gilbert, R.O. 1987. *Statistical methods for environmental pollution monitoring*. New York: John Wiley and Sons, Inc.

Helsel, D.R., and R.M. Hirsch. 2002. *Statistical methods in water resources*, Chapter A3. In: Techniques of water-resources investigations of the United States Geological Survey, Book 4, Hydrologic analysis and interpretation. https://pubs.usgs.gov/twri/twri4a3/twri4a3.pdf, accessed March 4, 2020.

Helsel, D.R., and L. M Frans. 2006. Regional Kendall test for trend. *Environ Sci Technol*. 2006;40(13):4066-4073. doi:10.1021/es051650b.

Helsel, D.R., Hirsch, R.M., Ryberg, K.R., Archfield, S.A., and Gilroy, E.J. 2020. Statistical methods in water resources: U.S. Geological Survey Techniques and Methods, book 4, chapter A3, 458 p., https://doi.org/10.3133/tm4a3. [Supersedes USGS Techniques of Water-Resources Investigations, book 4, chapter A3, version 1.1.]

Hirsch, R.M., and J.R. Slack. 1984. Nonparametric trend test for seasonal data with serial dependence. *Water Resources Research* 20(6): 727–732. https://www.researchgate.net/publication/224839899_Non-Parametric_Trend_Test_for_Seasonal_Data_With_Serial_Dependence.

Kincaid, T.M., and A.R. Olsen. 2019. *Spsurvey: Spatial survey design and analysis.* R package version 3.3. https://cran.r-project.org/web/packages/spsurvey/spsurvey.pdf.

Marchetto, Aldo. 2017. rkt: Mann-Kendall Test, Seasonal and Regional Kendall Tests. R package version 1.5. https://CRAN.R-project.org/package=rkt.

Millard, S.P. 2013. *EnvStats: An R package for environmental statistics*. New York: Springer. ISBN 978-1-4614-8455-4. http://www.springer.com.

R Core Team. 2017. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing Report 3-900051-07-0, Vienna, Austria. http://www.R-project.org/.

Sen, P.K. 1968. Estimates of the Regression Coefficient based on Kendall's Tau. Journal of the American Statistical Association, 63, 1379-1389. http://dx.doi.org/10.1080/01621459.1968.10480934.

Wickham, H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

Appendices

Appendix A. Data Qualifiers

Data qualifiers used by the department beginning October 1, 2017

This is a two-column table. Column 1 lists the qualifier code and Column 2 lists a description.

Code	Definition
A	Value reported is the arithmetic mean (average) of two or more determinations.
B	Results based upon colony counts outside the acceptable range. This code applies to microbiological tests and specifically to membrane filter colony counts. The code is to be used if the colony count is generated from a plate in which the total number of coliform colonies is outside the method indicated ideal range.
G	Indicates that the analyte was detected at or above the method detection limit in both the sample and the associated field collected blank, and the value of the blank is greater than 10% of the associated sample value.
I	The reported value is greater than or equal to the laboratory method detection limit but less than the laboratory practical quantification limit.
J	Estimate value. **Shall be accompanied by a detailed explanation** to justify the reason(s) for designating the value as estimated. Examples of situations in which a "J" code must be reported include: instances where a quality control item associated with the reported value failed to meet the established quality control criteria (the specific failure must be identified); instances when the sample matrix interfered with the ability to make any accurate determination; instances when data are questionable because of improper laboratory or field protocols (e.g., composite sample was collected instead of a grab sample); instances when the analyte was detected at or above the method detection limit in an analytical laboratory blank other than the method blank (such as calibration blank) and the value the blank is greater than 10% of the associated sample value; or instances when the field or laboratory calibrations or calibration verifications did not meet calibration acceptance criteria.
K	Off-scale low. The actual value is known to be less than the value given. (Only used for lab analyses.)
L	Off-scale high. The actual value is known to be greater than the value given. (Only used for lab analyses.)
N	Presumptive evidence of presence of material; component tentatively identified based on mass spectral library search or there is an indication that the analyte is present, but quality control requirements for the confirmation were not met.
O	Sampled but analysis lost or not performed.
Q	Sample held beyond the accepted holding time. Value is derived from a sample that was prepared or

Code	Definition
	analyzed after the approved holding time restrictions for sample preparation or analysis.
R	Significant rain (typically in excess of ½ inch) in the past 48 hours, which might contribute to a lower or higher than normal value.
S	Secchi disk visible to bottom of waterbody. The value reported is the depth of the waterbody at the location of the Secchi disk measurement.
T	Value reported is less than the laboratory method detection limit. Value reported for informational purposes only and shall not be used in statistical analysis.
U	Indicates that the compound was analyzed for but not detected. The reported value shall be the method detection limit.
V	Indicates that the analyte was detected at or above the method detection limit in both the sample and the associated method blank and the blank value was greater than 10% of the associated sample value.
X	Indicates, when reporting results from a Stream Condition Index Analysis (LT 7200 and FS 7420), that insufficient individuals were present in the sample to achieve a minimum of 280 organisms for identification (the method calls for two aliquots of 140-160 organisms), suggesting either extreme environmental stress or a sampling error.
Y	The laboratory analysis was from an unpreserved or improperly preserved sample. The data may not be accurate.
Z	Too many colonies were present for accurate counting. Historically, this condition has been reported as “too numerous to count” (TNTC). The “Z” qualifier code shall be reported when the total number of colonies of all types is more than 200 in all dilutions of the sample tested using a membrane filter technique. When applicable to the observed test results, a numeric value for the colony count for the microorganism tested may be estimated from the highest dilution factor (smallest sample volume) used for the test and reported with the qualifier code.
!	Indicates that the reported value deviates from historically established concentration ranges.
?	Data are rejected and should not be used. Some or all of the quality control data for the analyte were outside criteria, and the presence or absence of the analyte cannot be determined from the data.

Appendix B. Descriptive Statistics Example

Example of descriptive statistics table for the distribution of total nitrate (TN) and TP) values from a Status Network sample survey.

This *is a nine-column table. Column 1 lists the analyte, Column 2 lists the measurement unit, Column 3 lists the number of samples, Column 4 lists the number of values Below Detection Limits (BDLs), Column 5 lists the minimum value, Column 6 lists the first quartile value, Column 7 lists the median value, Column 8 lists the third quartile value, and Column 9 lists the maximum value.*

Analyte	***Measurement Unit***	***Number of Samples***	***Number of BDLs***	***Minimum Value***	***25th Percentile Value***	***Median Value***	***75th Percentile Value***	***Maximum Value***
TN	mg/L	30	7	0.05	0.09	1.00	2.00	10.30
TP	mg/L	29	4	0.05	0.10	0.10	0.15	1.30

Appendix C. General examples of SK, RK, and CHAN scripts

SK script for flow-adjusted loop for total phosphorous

```
################################################################################
#### 9/12/2019 Seasonal Kendall flow adjusted code for total phosphorus
####  analyses written by Jay Silvanima.
####
#### This script loads all Phosphorus, Total (as P)  and Stream Discharge data for all surface
####      water trend stations having stream discharge data.
#### It then loops through a station list of those stations  having discharge data and does the
####      following for each station.
####  1) Runs a lowess regression and stores  the residuals of Total Phosporus vs Flow.
####  3) Runs Seasonal Kenadall Tests for the residuals and Total Phosporus
####      (flow adjusted and non-flow adjusted).
####  4) Exports scatter plots of the residuals, and Total Phosporus by date.
####  5) Exports boxplots of the residual and Total Phosporus data by month
####  6) Exports text files of the SKT output for residuals and Total Phosporus.
####  7) For each station it stores the basic SKT flow adjusted output as a row in
####      dataframe named SKTstats. The cells of the table are station, indicator,
####       tau, slope, intercept, p-value(Het), p-value(Trend).
#### Once done six files are created in the below directory for each station
####  along with an excel spreadsheet containing the statistics for each station.
################################################################################

setwd ('Z:/data analysis/SW Trend Analysis 1999-2018/FLOW ADJUSTED/TP')

### Load station list for stations having daily discharge data.
### Note that if a station is loaded which does not have discharge data
###  the program will stop and an error will be provided that looks like
###  "Error in xj[i] : invalid subscript type 'list'".
```

```
swtrendstations<-read.csv('Z:/data analysis/SW Trend Analysis 1999-2018/swtrendstations_flow.csv')

stnList<-swtrendstations$PK.STATION
nStns<-length(stnList)
trenddata1<-allData[sapply(allData[,"FK_PARAM_CODE"],function(x)any(x==c(665,60))),]

##Load EnvStats and chron libraries
library("EnvStats", lib.loc="C:/R/R 3.1.0/R-3.1.0/library")
library("chron", lib.loc="C:/R/R 3.1.0/R-3.1.0/library")

##create dataframe to hold out put from Seasonal Kendall test (tau, slope, p-value)
SKTstats <- data.frame(matrix(ncol=7, nrow=1))
names(SKTstats) <- c("Station", "Indicator","Tau","Slope", "Intercept", "P-Value(Het)", "P-Value(Trend)")

##loop to produce graphics and SKT analysis output to files

for (istn in 1:nStns){

  ##Create the subtitle (q) and y-axis label (u) for graphs from the station number and indicator column
  q<-c(swtrendstations$PK.STATION[istn])
##  u<-(trenddata2[1,9])
  q
##  u

##Select station to run trend analysis on
##Create dataframe trendData2 with data coming the selected station

  trenddata2<-trenddata1[sapply(trenddata1[,"FK_STATION"],function(x)any(x==c(q))),]

##  Select only values for parcodes desired-i.e. TP (665) and Flow (60)
```

```
##  Need to have flow and values on same spreadsheet to make stats function!!!!!

 trenddata3<-trenddata2[sapply(trenddata2[,"FK_PARAM_CODE"],function(x)any(x==c(665))),]
 trenddata4<-trenddata2[sapply(trenddata2[,"FK_PARAM_CODE"],function(x)any(x==c(60))),]

##  Change the names of the columns for the flow data
 names(trenddata4)[12]<-'FLOW'
 names(trenddata4)[13]<-'FLOW.QUALIFIER'
 names(trenddata4)[14]<-'FLOW.UNITS'

##  Change the names of the columns for the data
 names(trenddata3)[12]<-'TP'
 names(trenddata3)[13]<-'TP.QUALIFIER'
 names(trenddata3)[14]<-'TP.UNITS'

##  Create dataframe by merging TP data with Flow data
 trenddata6<-merge.data.frame(trenddata3,trenddata4, by=1:7)


##  and order with respect to date and location
 trenddata6<-trenddata6[order(trenddata6[,6]),]
 trenddata6<-trenddata6[order(trenddata6[,1]),]

## Date conversions for SKT
## need to create column with month of year only
 trenddata6$year<-as.numeric(trenddata6$YEAR.x)
 trenddata6$COLLECTION_DATE<-as.Date(trenddata6$COLLECTION_DATE, "%m/%d/%Y")
 trenddata6$moy<-months(trenddata6$COLLECTION_DATE)

## Create dataframe for the combined TP and flow as per above data prep.
 trenddata7<-trenddata6[sapply(trenddata6[,"FK_STATION"],function(x)any(x==c(stnList[istn]))),]
```

```
## **Start of statistics**
## calculate loess residuals for pth variable

  tdata<-data.frame(x)
  tdata<-loess(trenddata7$TP ~ trenddata7$FLOW)
    trenddata7$lresids<-residuals(tdata)

##Scatter Plot of values
  jpeg(paste(q,'Residuals_Plot_TP.jpg'))

  plot(trenddata7$COLLECTION_DATE, trenddata7$lresids, main=(paste("RESIDUALS TP/STREAM FLOW
PK_STATION = ",q)),  xlab="Date", ylab="Residuals, TP mgL/FLOW CFS",
       xlim=c(min(trenddata7$COLLECTION_DATE),max(trenddata7$COLLECTION_DATE)), )

  dev.off()

##Boxpolt of value data vs sorted months
## Prepping data for box plot of seasonal data
  trenddata7$month_fac = factor(trenddata7$moy, levels = month.name)
  sort(trenddata7$month_fac)

  jpeg(paste(q,"FLOWRESIDS_BoxPlot_TP.jpg"))

  boxplot(trenddata7$lresids~sort(trenddata7$month_fac), data=trenddata7, main=(paste("PK_STATION = ",q)),
xlab="Date", ylab="Residuals, mgL/CFS")

  dev.off()

##run skt and store output to dataframe and txt file
```

```
  sktstats<-
kendallSeasonalTrendTest(trenddata7$lresids,trenddata7$MONTH,trenddata7$year,alternative="two.sided",correct =
TRUE,ci.slope = TRUE,conf.level = 0.95,
                         independent.obs=TRUE,season_name= NULL, year_name = null)


  SKTstats[istn,] <- c(q, "Flow Resids Phosphorus, Total (as P)", sktstats[[4]], sktstats[[3]])


  sink(paste(q,"_TP_CFS_SKT.txt"))
  print(kendallSeasonalTrendTest(trenddata7$lresids,trenddata7$MONTH,trenddata7$year,alternative="two.sided",correct =
TRUE,ci.slope = TRUE,conf.level = 0.95,
                       independent.obs=TRUE,season_name= NULL, year_name = null))
  sink()


##Scatter Plot of TP values
  jpeg(paste(q,"TP_Scatter_Plot.jpg"))


  plot(trenddata7$COLLECTION_DATE, trenddata7$TP, main=(paste("Total Phosphorus",q)),  xlab="Date", ylab="TP
mg/L",
     xlim=c(min(trenddata7$COLLECTION_DATE),max(trenddata7$COLLECTION_DATE)), )


  dev.off()


##Boxpolt of value data vs sorted months


  jpeg(paste(q,'BoxPlot_TP.jpg'))


  boxplot(trenddata7$TP~sort(trenddata7$month_fac), data=trenddata4, main=(paste("Total Phosphorus\nBoxplots per
month\n1998 to present  PK_STATION = ",q)), xlab="Date", ylab="TP, mg/L")


  dev.off()


##Run SKT for non flow adjusted data.
```

```
  sink(paste(q,'TP_SKT.txt'))

  print(kendallSeasonalTrendTest(trenddata7$TP,trenddata7$MONTH,trenddata7$year,alternative="two.sided",correct =
TRUE,ci.slope = TRUE,conf.level = 0.95,

                    independent.obs=TRUE,season_name= NULL, year_name = null))

  sink()


}


##Write out table with statistics

write.csv(SKTstats,file='TN_CFS_Stats.csv')
```

RK script for total phosphorous

```
################################################################################
####
####  Regional Kendall Tau Tests for flow adjusted data coming from the
####      46 surface water trend stations co-located with USGS Gage Stations.
####  Period of record January 1 1998 through present.
####  Utilizes dataframe (allData2) which contains the merged data for stations
####..    Charlie Creek and Aerojet Canal because these stations had to be moved
####..    due to bridge construction.
################################################################################


###Script written by Jay Silvanima 11/15/2019.


setwd ('Z:/data analysis/SW Regional Trend Analysis/1998-2019/NO Flow Adjustment')


##  Load Regional Kendall Tau (rkt) package and lubridate package to deal with dates.


library(rkt)
library(lubridate)


##Date conversions
allData2$month<-as.character(allData2$MONTH)
allData2$year<-as.numeric(allData2$YEAR)
allData2$COLLECTION_DATE<-as.Date(allData2$COLLECTION_DATE, "%m/%d/%Y")


allData2$month[allData2$MONTH == 1] <- 'January'
allData2$month[allData2$MONTH == 2] <- 'February'
allData2$month[allData2$MONTH == 3] <- 'March'
allData2$month[allData2$MONTH == 4] <- 'April'
allData2$month[allData2$MONTH == 5] <- 'May'
allData2$month[allData2$MONTH == 6] <- 'June'
```

```
allData2$month[allData2$MONTH == 7] <- 'July'
allData2$month[allData2$MONTH == 8] <- 'August'
allData2$month[allData2$MONTH == 9] <- 'September'
allData2$month[allData2$MONTH == 10] <- 'October'
allData2$month[allData2$MONTH == 11] <- 'November'
allData2$month[allData2$MONTH == 12] <- 'December'

## Prepping data for box plot of seasonal data
allData2$month_fac = factor(allData2$month, levels = month.name)
sort(allData2$month_fac)

#################################################################################
##  Select only values for parcodes desired-i.e. Phosphorus, Total (665),
##  Need to have flow and values on same spreadsheet to make stats function!!!!!
#########  Combine TKN and NOx Total results

##  Select only values for parcode desired-i.e. Phosphorus, Total (665)

trenddata3<-allData2[sapply(allData2[,"FK_PARAM_CODE"],function(x)any(x==c(665))),]

##  Change the names of the columns for the flow data
names(trenddata3)[12]<-'TP'
names(trenddata3)[13]<-'TP.QUALIFIER'
names(trenddata3)[14]<-'TP.UNITS'

##  and order with respect to date and location
trenddata3<-trenddata3[order(trenddata3[,6]),]
trenddata3<-trenddata3[order(trenddata3[,1]),]

## Date conversions for SKT
## need to create column with month of year only
```

```
trenddata3$year<-as.numeric(trenddata3$YEAR)
trenddata3$COLLECTION_DATE<-as.Date(trenddata3$COLLECTION_DATE, "%m/%d/%Y")
trenddata3$moy<-months(trenddata3$COLLECTION_DATE)

##To export dataframe to file see below....

write.csv(trenddata3, file = 'TP_All.csv')

###
### Remove data qualified with fatal qualiers (?,O,G,Y,V) manually
### then import corrected file
###

trenddata6<-TP_All_Corrected

## Date conversions for Plotting
## need to create column with month of year only
trenddata6$year<-as.numeric(trenddata6$YEAR)
trenddata6$COLLECTION_DATE<-as.Date(trenddata6$COLLECTION_DATE, "%m/%d/%Y")
trenddata6$moy<-months(trenddata6$COLLECTION_DATE)

trenddata6$month_fac = factor(trenddata6$moy, levels = month.name)
sort(trenddata6$month_fac)

hist(trenddata6$MONTH)
boxplot(trenddata6$TP~trenddata6$FK_STATION)
summary(trenddata6)

trenddata6$block <- (trenddata6$FK_STATION * 12) + trenddata6$MONTH

boxplot(trenddata6$TP~trenddata6$FK_STATION)
```

```
##Boxpolt of value data vs sorted months

jpeg(paste("TP by month.jpg"))

boxplot(trenddata6$TP~sort(trenddata6$month_fac), data=trenddata6, main=('SW Trend Stations 1998-2019'), xlab="Date",
ylab="TP mg/L")

dev.off()

#Scatter Plot of TOC

jpeg(paste("TP_Scatter_Plot.jpg"))

plot(trenddata6$COLLECTION_DATE, trenddata6$TP, main=(paste("SW Trend Stations 1998-2019")),  xlab="Date",
ylab="TP mg/L",
    xlim=c(min(trenddata6$COLLECTION_DATE),max(trenddata6$COLLECTION_DATE)), )

dev.off()

##  Run rkt using YEAR, VALUE, FK_STATION+block (MONTH) using correction for
##    spatial autocorrellation and using a median for values sharing the same
##    date.

rkt(trenddata6$YEAR,trenddata6$TP, trenddata6$block, , correct = T, rep = "m")

sink(paste("TP_RKT_1998.txt"))

print(rkt(trenddata6$YEAR,trenddata6$TP, trenddata6$block, , correct = T, rep = "m"))

sink()
```

CHAN for flowing waters from 2000-2017

```
# This script is written to utilize the change.analysis section of the spsurvey package.
# by Jay Silvanima
# 02/11/2019
# Tony Olsen incorporated additional code during review
# 03/01/ - 03/06/2019
# Load psurvey.analysis library
library("spsurvey")

# import the combined Exclusion and Results file that will be used for the analysis

# This is the Flowing Waters file for 2000-2017

setwd ('Z:/data analysis/Change Analysis/Cycle1 FW 2000-2017')

##Descriptive Stats tmdl basins
##Load combined file containing information from both time periods.
##File created independently from this script.

comb <- read.csv('C1C32017COMB.csv')

##comb <- read.csv('Original_Data/C1C32017COMB.csv')

#####################################################################
###NNC and DO Calculations
#####################################################################

names(comb)
```

```
comb$TN<-(comb$Kjeldahl.Nitrogen..Total..as.N+comb$Nitrate.Nitrite..Total..as.N)



### Pass=1 AND Fail=0
## Using the NNC maximum values

comb$TNcat<-ifelse((comb$TN_NNC >= comb$TN),1,0)

comb$TPcat<-ifelse((comb$TP_NNC >= comb$Phosphorus..Total..as.P),1,0)

comb$ChlaCat<- ifelse((comb$CHLA_NNC >= comb$Chlorophyll.A..Monochromatic) ,0,1)

#load lubridate
#need the date in the yyyy-mm-dd format

library(lubridate)

# Generate the files used in change.analysis

# do equal area projection for variance estimation
tmp <- marinus(comb$LAT_DEC, comb$LONG_DEC)
comb$xmarinus <- tmp[,'x']
comb$ymarinus <- tmp[,'y']

nr<-nrow(comb)

## The following is the targetsize in miles for the flowing waters in
## Florida's 29  drainage basins. Excludes canals within the Northwest Florida
## and Suwannee River Water Management Districts as these canals were never
## included in our statistical surveys due to their limited extent.
```

```
## The resource extents for three basins are exluded because no sites were
## selected during cycle one from these basins.

targetsizeGroup <- c("Apalachicola - Chipola" = 1814.74,
                     "Caloosahatchee" = 249.40,
                     ##"Charlotte Harbor" = 52.45,
                     "Choctawhatchee - St. Andrew" = 2681.42,
                     "Everglades" = 456.09,
                     "Everglades West Coast" = 326.23,
                     "Fisheating Creek" = 248.35,
                     ## "Florida Keys" =  0.09,
                     "Indian River Lagoon" = 117.09,
                     "Kissimmee River" = 736.68,
                     "Lake Okeechobee" = 157.98,
                     "Lake Worth Lagoon - Palm Beach Coast" = 215.41,
                     "Lower St. Johns" = 1268.08,
                     "Middle St. Johns" = 616.44,
                     "Nassau - St. Marys" = 520.24,
                     "Ochlockonee - St. Marks" = 1474.90,
                     "Ocklawaha" = 380.42,
                     "Pensacola" = 2325.68,
                     "Perdido" = 272.02,
                     "Sarasota Bay - Peace - Myakka" = 1637.37,
                     "Southeast Coast - Biscayne Bay" = 317.48,
                     "Springs Coast" = 147.83,
                     "St. Lucie - Loxahatchee" = 216.34,
                     "Suwannee" = 1969.26,
                     "Tampa Bay" = 146.52,
                     "Tampa Bay Tributaries" = 1011.69,
                     ##"Upper East Coast" = 121.93,
                     "Upper St. Johns" = 537.52,
```

```
            "Withlacoochee" = 436.31)


## The following is the targetsize in miles for the flowing waters in
## Florida's 6 reporting units per 2017 coverage. Excludes canals within the Northwest Florida
## and Suwannee River Water Management Districts as these canals were never
## included in our statistical surveys due to their limited extent.

##targetsizeZone <- c("ZONE 1"=8629.32 ,"ZONE 2"=1901.23,
##                "ZONE 3"=3615.71,"ZONE 4"=3547.18,
##                "ZONE 5"=1455.62,"ZONE 6"=1384.50)

## The following is the targetsize in miles for the flowing waters in
## Florida's 6 reporting units. Excludes canals within the Northwest Florida
## and Suwannee River Water Management Districts as these canals were never
## included in our statistical surveys due to their limited extent.
## The resource extents for three TMDL basins (Charlotte Harbor, Florida Keys,
## and Upper East Coast are exluded because no sites were
## selected during cycle one from these basins.

targetsizeZone <- c("ZONE 1"=8629.32 ,"ZONE 2"=1901.23,
                    "ZONE 3"=3497.13,"ZONE 4"=3509.88,
                    "ZONE 5"=1438.75,"ZONE 6"=1384.50)

# Look at Cycle, Zone and Group
addmargins(table(comb$GROUP_NAME, comb$C3_REPORTING_UNIT, comb$CYCLE,
                 useNA = 'ifany'))
# have 768 observations from Cycle 1 and 711 from Cycle 2. No missing values.
# All sites are target and sampled.  Assume this is all sites evaluated.
# Do not know if this is true. Makes assumption that sample frame matches the
# target population and that all sites could be sampled.
```

```
comb$ZoneGroup <- factor(paste(comb$C3_REPORTING_UNIT, comb$GROUP_NAME, sep = "_"))
levels(comb$ZoneGroup)
# note that Suwannee Group is split between Zone 1 and Zone 2. Is this an issue?

##### Weight adjustment

# Cycle 1: Adjust weights to 2017 framesize with 29 basins used in the design
comb$wgt_1 <- adjwgt(comb$CYCLE == 1, comb$WGT, comb$GROUP_NAME,
                     framesize = targetsizeGroup)

# Cycle 3: Adjust weights to 2017 framesize with 6 zones used in the design
comb$wgt_3 <- adjwgt(comb$CYCLE == 3, comb$WGT, comb$C3_REPORTING_UNIT,
                     framesize = targetsizeZone)

addmargins(table(comb$wgt_1 == 0, comb$wgt_3 == 0))

comb$wgt <- comb$wgt_1 + comb$wgt_3

# look at weight sums for Cycle 1
tmp <- tapply(comb$wgt_1, list(comb$GROUP_NAME, comb$C3_REPORTING_UNIT), sum)
tmp[is.na(tmp)] <- 0
round(addmargins(tmp), 1)
# Note that Charlotte Harbor and Upper East Coast do not have sites.
# Also Group length totals 20281.5

# look at weight sums for Cycle 3
tmp <- tapply(comb$wgt_3, list(comb$GROUP_NAME, comb$C3_REPORTING_UNIT), sum)
tmp[is.na(tmp)] <- 0
round(addmargins(tmp), 1)
# Zone Length totals 20360.8 which is close to length for Cycle 1.
```

```
# Note differences in GIS layer scale and how extracted account for differences.
# Not unusual for this to happen.

## write.csv(comb,file='comb.csv')

#Now to get the dataframe for the sites

mysites <- data.frame(siteID=comb$PK_RANDOM,
                      Survey1=ifelse(comb$CYCLE=="1",TRUE,FALSE),
                      Survey2=ifelse(comb$CYCLE=="3", TRUE,FALSE))

#########################################################################
### Define subpopulations (reporting units) and create design for
### Florida's TMDL basins
#########################################################################

mysubpop <- data.frame(siteID=comb$PK_RANDOM,
                       Combined=rep("Basins Combined", nr),
                       Zone=comb$C3_REPORTING_UNIT, # do Zone estimates as well
                       Basin=comb$GROUP_NAME)

## mydsgn <- data.frame(siteID=comb$PK_RANDOM,
##                      wgt=comb$wgt,
##                      xcoord=comb$xmarinus ,
##                      ycoord=comb$ymarinus ,
##                      stratum=comb$GROUP_NAME)
# since use local nhbd variance estimator (in most cases) likely
# okay to ignore stratum so that nbhd uses nearby sites regardless if they
# are in neighboring GROUP or ZONE. have very small sample sizes otherwise.
mydsgn <- data.frame(siteID=comb$PK_RANDOM,
                     wgt=comb$wgt,
```

```
                xcoord=comb$xmarinus ,
                ycoord=comb$ymarinus)

##############################################################################
### Create categorical and continuous distribution of specific indicator
### to run change analysis on
##############################################################################

### Total Nitrogen

mydata.cat <- data.frame(siteID=comb$PK_RANDOM,
                    TNcategory=comb$TNcat)

mydata.cont <- data.frame(siteID=comb$PK_RANDOM,
                     TN=comb$TN)

#### Total Nitrogen change.analysis

CategoryTNCAT <- change.analysis(sites=mysites, subpop=mysubpop, design=mydsgn,
                            data.cat=mydata.cat,data.cont=mydata.cont, test=c("mean","median"))

##############################################################################
##### write out total nitrogen change analysis results
##############################################################################

CategoryTNCAT

write.csv(CategoryTNCAT$catsum,file='TNcatsumC1C3.csv')
write.csv(CategoryTNCAT$contsum_mean,file='TNcontsummeanC1C3.csv')
write.csv(CategoryTNCAT$contsum_median,file='TNcontsummedianC1C3.csv')
```

```
### Nitrate + Nitrite as N (NOx)


mydata.cont <- data.frame(siteID=comb$PK_RANDOM,
                          NOx=comb$Nitrate.Nitrite..Total..as.N)

#### NOx change.analysis

CategoryNOx <- change.analysis(sites=mysites, subpop=mysubpop, design=mydsgn,
                               data.cont=mydata.cont, test=c("mean","median"))

################################################################################
##### write out NOx change analysis results
################################################################################

CategoryNOx

write.csv(CategoryNOx$contsum_mean,file='NOxcontsummeanC1C3.csv')
write.csv(CategoryNOx$contsum_median,file='NOxcontsummedianC1C3.csv')


### TKN


mydata.cont <- data.frame(siteID=comb$PK_RANDOM,
                          TKN=comb$Kjeldahl.Nitrogen..Total..as.N)

#### TKN change.analysis

CategoryTKN <- change.analysis(sites=mysites, subpop=mysubpop, design=mydsgn,
                               data.cont=mydata.cont, test=c("mean","median"))
```

```
##############################################################################
##### write out TKN change analysis results
##############################################################################

CategoryTKN

write.csv(CategoryTKN$contsum_mean,file='TKNcontsummeanC1C3.csv')
write.csv(CategoryTKN$contsum_median,file='TKNcontsummedianC1C3.csv')



#### Total Phosphorus

mydata.cat <- data.frame(siteID=comb$PK_RANDOM,
                         TPcategory=comb$TPcat)

mydata.cont <- data.frame(siteID=comb$PK_RANDOM,
                          TP=comb$Phosphorus..Total..as.P)

#### Total Phosphorus change.analysis

CategoryTPCAT <- change.analysis(sites=mysites, subpop=mysubpop, design=mydsgn,
                                 data.cat=mydata.cat,data.cont=mydata.cont, test=c("mean","median"))



##############################################################################
##### write out total phosphorus change analysis results
##############################################################################

CategoryTPCAT
```

```
write.csv(CategoryTPCAT$catsum,file='TPcatsumC1C3.csv')
write.csv(CategoryTPCAT$contsum_mean,file='TPcontsummeanC1C3.csv')
write.csv(CategoryTPCAT$contsum_median,file='TPcontsummedianC1C3.csv')

#### Chlorophyll A

mydata.cat <- data.frame(siteID=comb$PK_RANDOM,
                         Chlacategory=comb$ChlaCat)

mydata.cont <- data.frame(siteID=comb$PK_RANDOM,
                          ChlA=comb$Chlorophyll.A..Monochromatic)

#### Chlorophyll A corrected change.analysis


CategoryChlA <- change.analysis(sites=mysites, subpop=mysubpop, design=mydsgn,
                                data.cat=mydata.cat, data.cont=mydata.cont, test=c("mean","median"))

################################################################################
##### write out chlorophyll a corrected change analysis results
################################################################################

CategoryChlA

write.csv(CategoryChlA$catsum,file='ChlAcatsumC1C3.csv')
write.csv(CategoryChlA$contsum_mean,file='ChlAcontsummeanC1C3.csv')
write.csv(CategoryChlA$contsum_median,file='ChlAcontsummedianC1C3.csv')

#### Total Chloride

mydata.cont <- data.frame(siteID=comb$PK_RANDOM,
```

```
                CL=comb$Chloride..Total)

#### Total Chloride change.analysis

CategoryCL <- change.analysis(sites=mysites, subpop=mysubpop, design=mydsgn,
                    data.cont=mydata.cont, test=c("mean","median"))

###############################################################################
##### write out Chloride corrected change analysis results
###############################################################################

CategoryCL

write.csv(CategoryCL$contsum_mean,file='CLcontsummeanC1C3.csv')
write.csv(CategoryCL$contsum_median,file='CLcontsummedianC1C3.csv')

# Specific conductance field
# Create Specific Conductance categories of 0, 1000, 10000 microsiemens

SCCat <- cut(comb$Specific.Conductance..Field, breaks=c(0,1000,10000), include.lowest=TRUE)
comb$SCCat <- SCCat
comb$SCCat <- as.factor(comb$SCCat)

#### Specific conductance change.analysis

mydata.cat <- data.frame(siteID=comb$PK_RANDOM,
                speccond=comb$SCCat)

mydata.cont <- data.frame(siteID=comb$PK_RANDOM,
                 SCField=comb$Specific.Conductance..Field)
```

```
CategorySpCond <- change.analysis(sites=mysites, subpop=mysubpop, design=mydsgn,
                        data.cat=mydata.cat,data.cont=mydata.cont, test=c("mean","median"))


##############################################################################
##### write out Specific Conductance change analysis results
##############################################################################

CategorySpCond

write.csv(CategorySpCond$catsum,file='SpCondcatsumC1C3.csv')
write.csv(CategorySpCond$contsum_mean,file='SpCondMeanC1C3.csv')
write.csv(CategorySpCond$contsum_median,file='SpCondMedianC1C3.csv')

# pH field

#### pH change.analysis

mydata.cont <- data.frame(siteID=comb$PK_RANDOM,
                    pHField=comb$pH..Field)

CategorypH <- change.analysis(sites=mysites, subpop=mysubpop, design=mydsgn,
                        data.cont=mydata.cont, test=c("mean","median"))

##############################################################################
##### write out pH change analysis results
##############################################################################

CategorypH

write.csv(CategorypH$contsum_mean,file='pHCondMeanC1C3.csv')
```

```
write.csv(CategorypH$contsum_median,file='pHCondMedianC1C3.csv')

# Temperature

#### Temperature change.analysis

mydata.cont <- data.frame(siteID=comb$PK_RANDOM,
                          TempField=comb$Water.Temperature)

CategoryTemp <- change.analysis(sites=mysites, subpop=mysubpop, design=mydsgn,
                                data.cont=mydata.cont, test=c("mean","median"))

############################################################################
##### write out temperature change analysis results
############################################################################

CategoryTemp

write.csv(CategoryTemp$contsum_mean,file='TempCondMeanC1C3.csv')
write.csv(CategoryTemp$contsum_median,file='TempCondMedianC1C3.csv')

# Dissolved Oxygen

#### Dissolved Oxygen change.analysis

mydata.cont <- data.frame(siteID=comb$PK_RANDOM,
                          DO=comb$Oxygen..Dissolved..Field)

CategoryDO <- change.analysis(sites=mysites, subpop=mysubpop, design=mydsgn,
                              data.cont=mydata.cont, test=c("mean","median"))
```

```
################################################################################
##### write out Dissolved Oxygen analysis results
################################################################################

CategoryDO

write.csv(CategoryDO$contsum_mean,file='DOCondMeanC1C3.csv')
write.csv(CategoryDO$contsum_median,file='DOCondMedianC1C3.csv')

# Total Organic Carbon

#### Total Organic Carbon change.analysis

mydata.cont <- data.frame(siteID=comb$PK_RANDOM,
                          TOC=comb$Organic.Carbon..Total)

CategoryTOC <- change.analysis(sites=mysites, subpop=mysubpop, design=mydsgn,
                               data.cont=mydata.cont, test=c("mean","median"))

################################################################################
##### write out Dissolved Oxygen analysis results
################################################################################

CategoryTOC

write.csv(CategoryTOC$contsum_mean,file='TOCCondMeanC1C3.csv')
write.csv(CategoryTOC$contsum_median,file='TOCCondMedianC1C3.csv')
```