Introduction

The goal of this repository is to analyze trends within MDNet data, as well as plot values for various sites at different time periods.

File List

* Seasonal Kendall Trend Test
  + Sktt.m
  + Taub.m
* Data Files
* Files to run and sort data
  + Cleandata.m
  + GreatLakesMDN.m
  + MDNmonth.m
  + Sort2005.m
  + SortallMDN.m
* Files of Trend Summaries
* Workspace for Great Lakes variables
  + GLworkspace.mat

Description of Files

# **sktt.m**

This file is run in conjunction with ktaub.m. Calculates a non-parameteric trend test of monotonic trends using seasons. Function outputs and inputs are: function [taubsea tausea Sens h sig sigAdj Zs Zmod Ss Sigmas CIlower CIupper] = sktt(datain, alpha, wantplot,StartSeason). Datain should be in the following form: 3 columns, the first listing year, the second listing month (season), and the third listing the actual data. This function can take into account multiple observations per season, but my data has been pre-analyzed, computing the monthly averages for each year.

# **ktaub.m**

This file is run automatically when sktt.m is run. The data is formatted and subsequently run by sktt.

# **Data Files**

This is a folder of Excel files of MDN data. The format of the columns is: SiteID, Date On, Date Off, RGPPT, SVOL, SUBPPT, HgConc, HgDep. These must be downloaded before analysis can occur.

# **Cleandata.m**

The purpose of this function is to remove outliers from HgConc data. The input is an imported MDN file, and it returns a cleaned matrix of HgConc data for all time. It analyzes the data based upon length and identifies and removes any outliers using Rosner’s test for identifying *k* outliers as described in *Statistical Methods for Environmental Pollution Monitoring*, Gilbert 1987. An r-value is calculated from:

Ri+1 = |x(i)-x(i)|/s(i)

Where x(i) is the largest value remaining, x(i) is the mean after the largest value is removed, and s(i) is the standard deviation after the largest value is removed.

N values are tabulated in table A16 of Gilbert 1987.

The function calculates Ri+1, and if Ri+1 > N, then x(i) is removed as an outlier.

# **GreatLakesMDN.m**

This inputs and prepares MDN great lakes files for analysis. It reads the MDN Excel files and outputs a file sorted by year, 1996-2014, as well as those with values for each day per year filled in. Other files include those sorted for sktt.m analysis for precipitation, hg conc, and hg deposition data. There are also those sorted for analysis from 2005 onwards.

For sorted matrices, there are 76 columns. Each set of 4 columns represents data for one year, the first column being dates, the second precipitation values, the third concentrations, and the fourth deposition. The data years run from 1996 to 2014.

# **MDNmonth.m**

This is the main file to run for data input and analysis. It calls GreatLakesMDN and outputs all the files needed for analysis. It also holds the baseline for plotting data by year, and for running sktt.

# **Sort2005.m**

This function only takes values of precipitation, hg concentration, and hg deposition matrices from 2005 onwards and returns them in similar output vectors.

# **SortallMDN.m**

SortallMDN holds all functions necessary for sorting data. It is called in GreatLakesMDN, and does not need to be run independently. The input should be in the form MDN\_site generated by GreatLakesMDN. Outputs sorted MDN data, and some matrices prepped for analysis via sktt.

*Sorting of data*

Data is first sorted into years using the function sortMDN, returned to site\_sort.

The function dayavg(MDNsort) calculates an average for each day of data, returning the file Site\_dy

The function monmed(MDNsort) calculates monthly medians for each year of data, returning Site\_mpre, Site\_mcon, and Site\_mdep.

The function monavg(MDNsort) calculates monthly means for each month of data, returning Site\_pre, Site\_con, and Site\_dep.

The function changeformat(monthavg) arranges files into the format needed for sktt.m and ktaub.m. The first column is year of collection, the second the month of collection, and the third the calculated value.

# **Files of Trend Summaries**

These are Excel files with saved summaries of all important trends across all sites. Different pages describe those for different time periods and types of data.

# **GLworkspace.mat**

This is a saved workspace of all MDN files for Great Lakes sites. If these are used, GreatLakesMDN does not need to be run, and the Excel files for sites do not need to be downloaded. Calling this saved workspace will prepare all the files automatically.

Types of variables

* *SITE#con* is the concentration array prepared for sktt.m. The first column is year of collection, the second is month, and third is the monthly mean.
* *SITE#con05* is the same concentration array as *SITE#con*, but with data only from 2005 onwards.
* *SITE#dep* is the deposition array prepared for sktt.m. The first column is year of collection, the second is month, and the third is the monthly mean.
* *SITE#dep05* is the same deposition array as *SITE#dep*, but with data only from 2005 onwards.
* *SITE#pre* is the precipitation array prepared for sktt.m. The first column is year of collection, the second is month, and the third is the monthly mean.
* *SITE#pre05* is the same precipitation array as *SITE#pre*, but with data only from 2005 onwards.
* *SITE#mpre, SITE#mcon, SITE#mdep* are also prepared for sktt.m, only with monthly medians instead of means.
* *SITE#dy* is an array where each group of four columns represent a years’ worth of daily average data, where the first column is date, the second is precipitation, the third is concentration, and the fourth is deposition.
* *SITE#sort* is an array where data is once again sorted by year, but without daily averaging. The same scheme of columns is used.

Setup & Initial Running

* Open MATLAB and set the directory as the MercuryTrendsAnalysis repository.
* Use **load(‘GLworkspace’)** to load the data files.
* To run trend analysis:
  + Refer to MDNmonth for functions
  + Use **[taubsea tausea Sens h sig sigAdj Zs Zmod Ss Sigmas CIlower CIupper] = sktt(Site#\_data, 0.05, 1)**
  + Replace **Site#\_data** with the site name you want for the type of data you want.
  + If you want to analyze precipitation at 48 Michigan, you insert MI48pre; for concentration you input MI48con; for deposition MI48dep
* To plot data for a certain year across all sites:
  + Run the Great Lakes by Year section of MDNmonth, replacing the column number with the correct values for the year and type of data you wish to plot.
  + For n=1:19, where n represents the number of years above 1995 for the year you desire to plot (for 1997, n=2), replace the column numbers for each site as follows:
  + **Plot(site#\_sort(:,(4\*n)-3), site#\_sort(:,k))**
  + where k depends on the type of data you wish to plot.  
    For precipitation: k = 4\*n-2  
    For concentration: k= 4\*n-1  
    For deposition: k= 4\*n
  + After all column values in the plot function have been replaced, run the section. The plots will display on four separate figures. You can also select for certain sites and plot only those.