ProUCL Utilization 2020

Trend Analysis

Presenters:

Travis Linscome-Hatfield,

Polona Carson

NEPTUNE AND COMPANY

Improving the quality of environmental decision making.

Summary of Session 1: ProUCL A to Z

- Objective:
  - Get familiar with ProUCL and some commonly used data analysis features
- We talked about:
  - Starting ProUCL
  - Preparing data for analysis and loading in ProUCL
  - Basics of dealing with missing values and NDs
  - Exploratory Data Analysis
  - Hypothesis testing

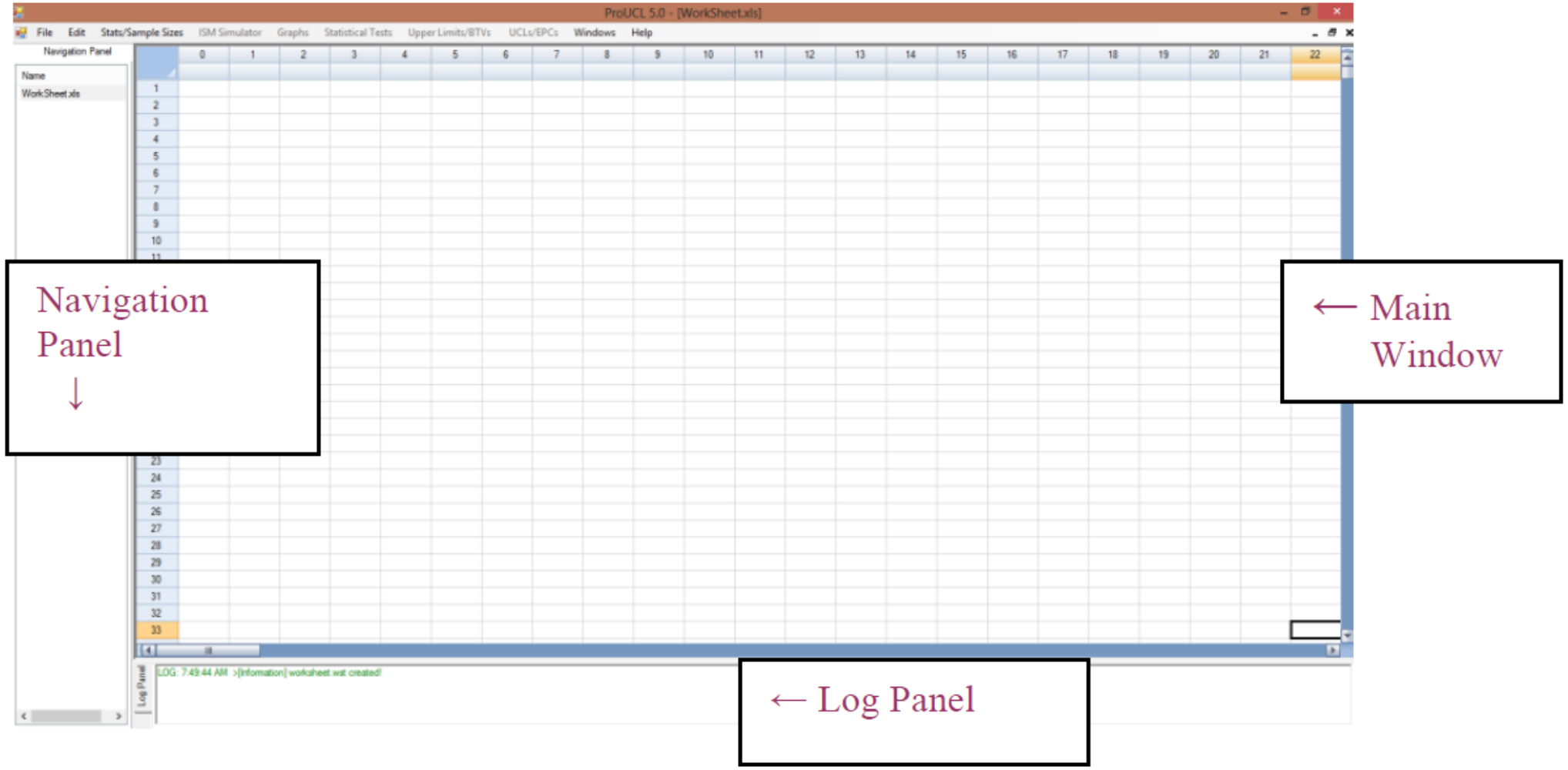
Learning Objectives

- Objectives
  - Get familiar with ProUCL features for trend analysis
- Today we will discuss:
  - Exploring the data with Time Series Plot
  - Mann-Kendall trend test
  - Theil-Sen trend test
  - Ordinary Least Square Regression
    - Analysis of residuals

ProUCL Software

- Statistical software for environmental data analysis
- User Guide
  - Provides instructions on how to use ProUCL
- Technical Guide
  - Provides detailed background on statistical methods

Navigating ProUCL



Data Sets

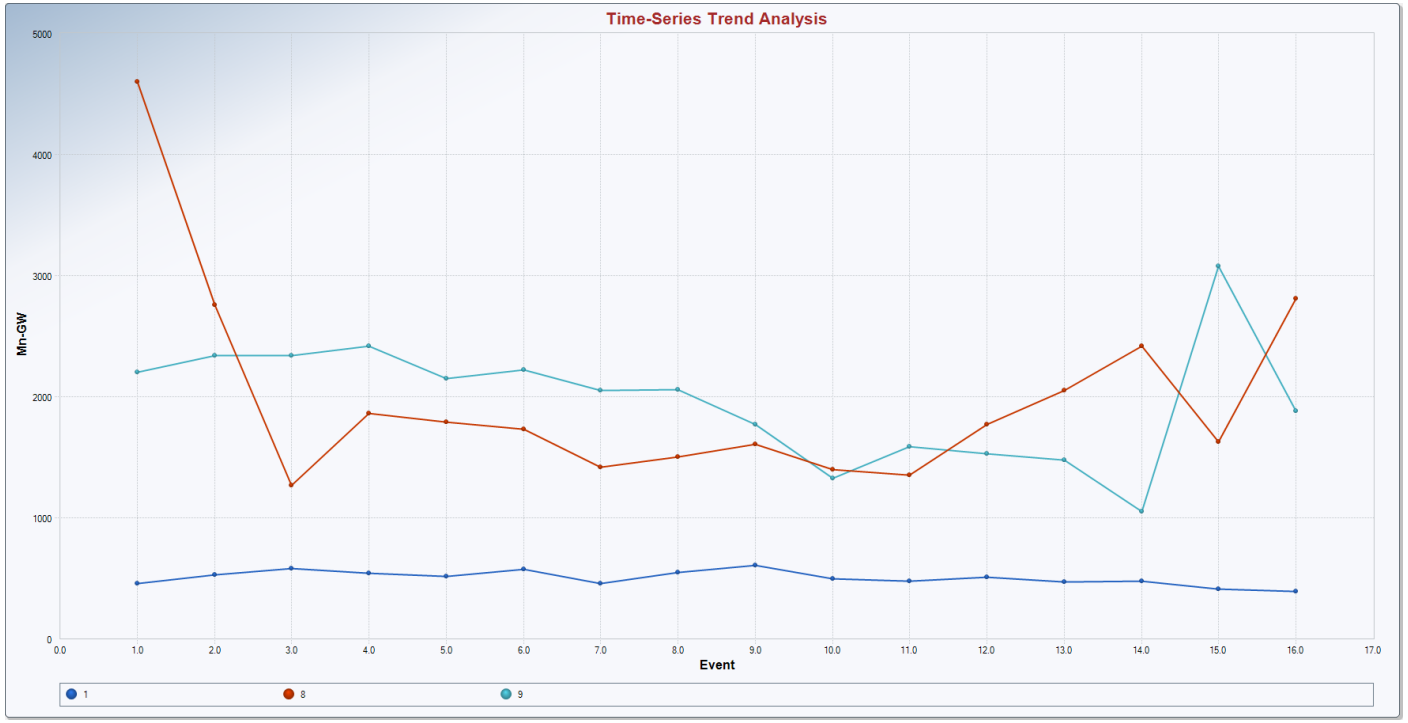
- MW-1-8-9.xls
  - In ProUCL Data folder
- Trend-MW-Real-data wln.xls
  - On training website

Nondetects (NDs)

- High non-detects always create severe problems in analyzing data
  - They should be removed from the data set before further analysis
- Rule of thumb to reject NDs
  - always reject NDs greater than the largest detect
  - Use judgement to reject NDs greater than 10x the smallest detect
- Reasonable substitutions for remaining NDs:
  - Substitute only if there is a small number of Nds (10-15%)
  - ½ of reporting limit (RL)
  - ½ of detection limit (DL)

Explore the Data with Time Series Plot

- Diagnostic and exploratory graphical tool
  - Is there a trend?
  - Linear or exponential?
- Grouping variable allows parallel display and comparison
- Helps to evaluate:
  - The presence of temporal / spatial variability
  - Is data transformation needed?

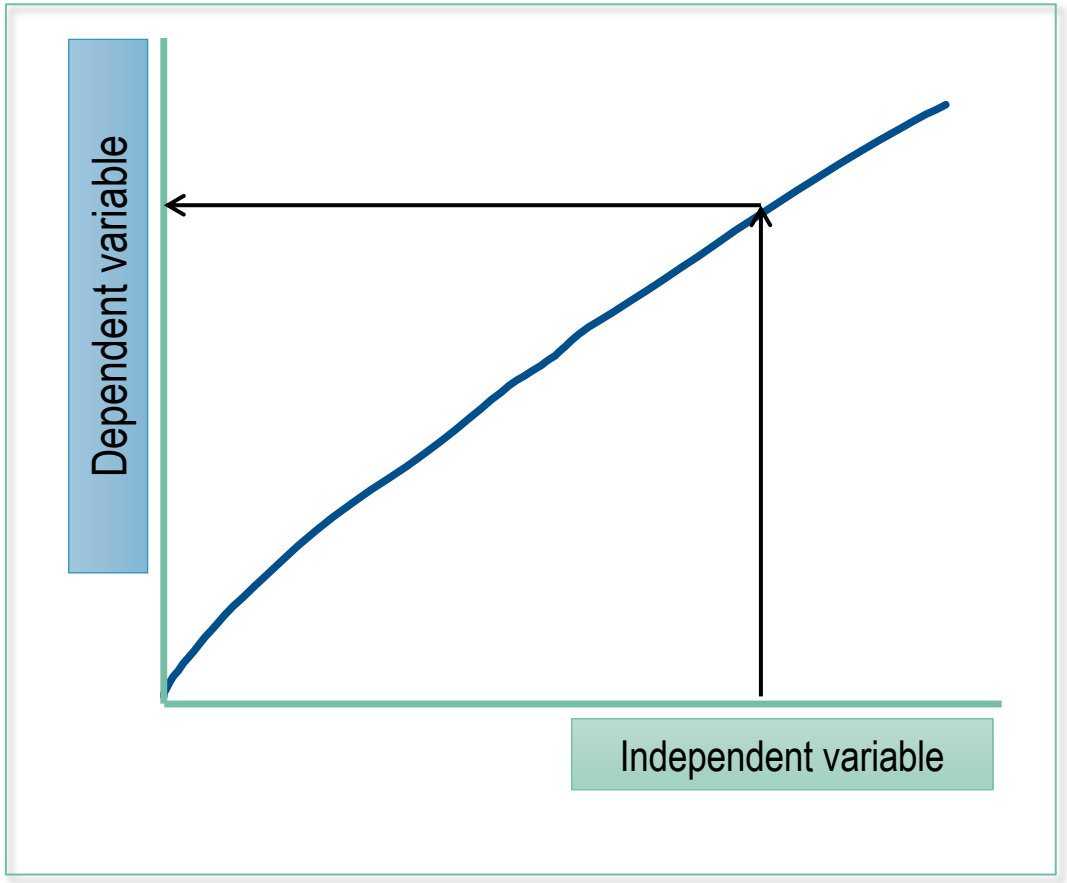


MW-1-8-9.xls

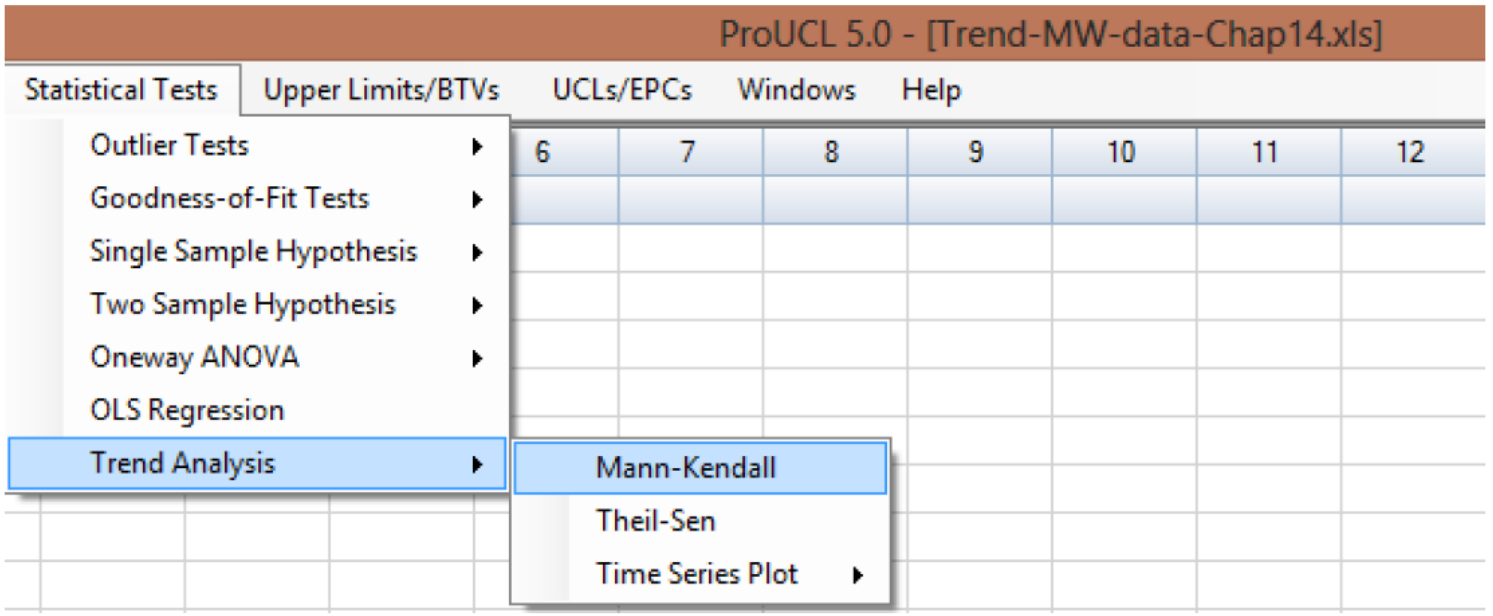
Are you Looking at your Data the Right Way?

- Dependent variable or response
  - Contaminant
    - MW-28 column
- Independent variable or predictor
  - Time
    - Index column (time in days)

Is contaminant (dependent variable) decreasing / increasing over time (independent variable)

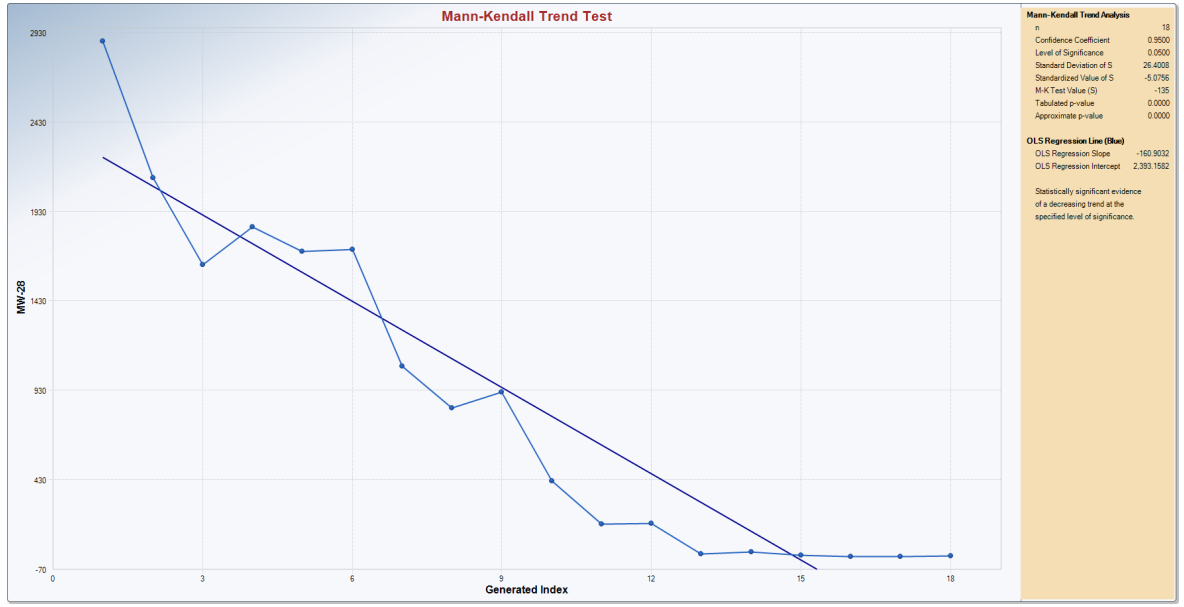


1. Click **Statistical Tests ►Trend Analysis ► Mann-Kendall.**

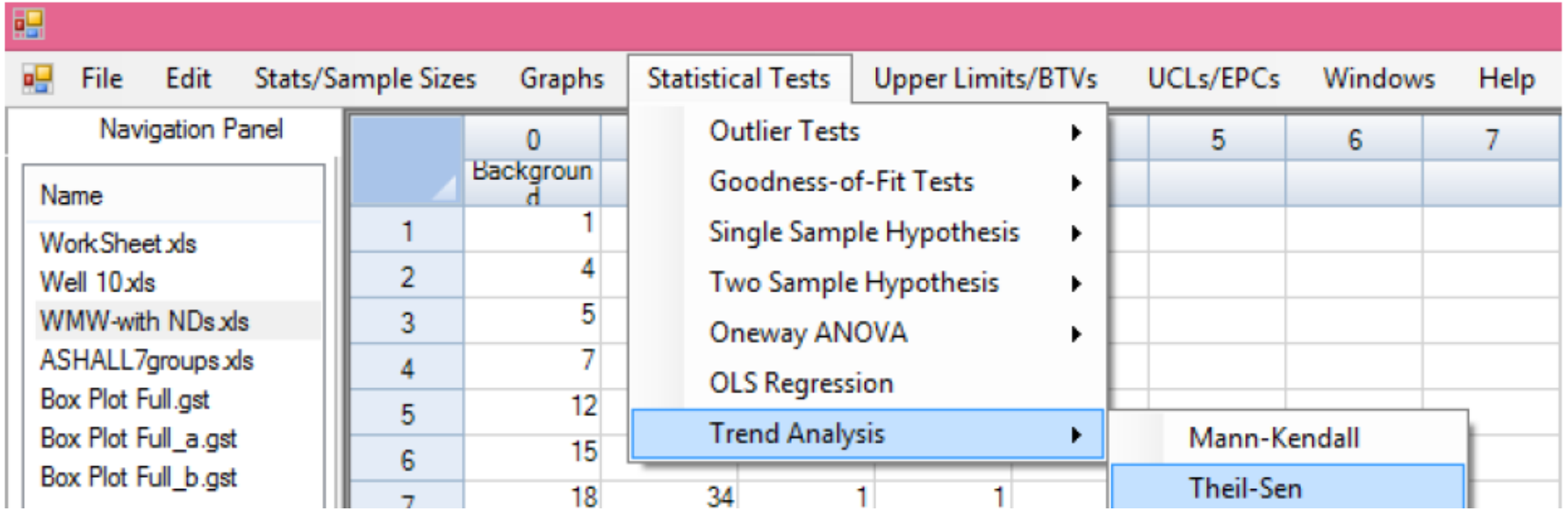


Advantages / Disadvantages

- Good for detecting monotone trends
- Identifies trend
- No information about the slope

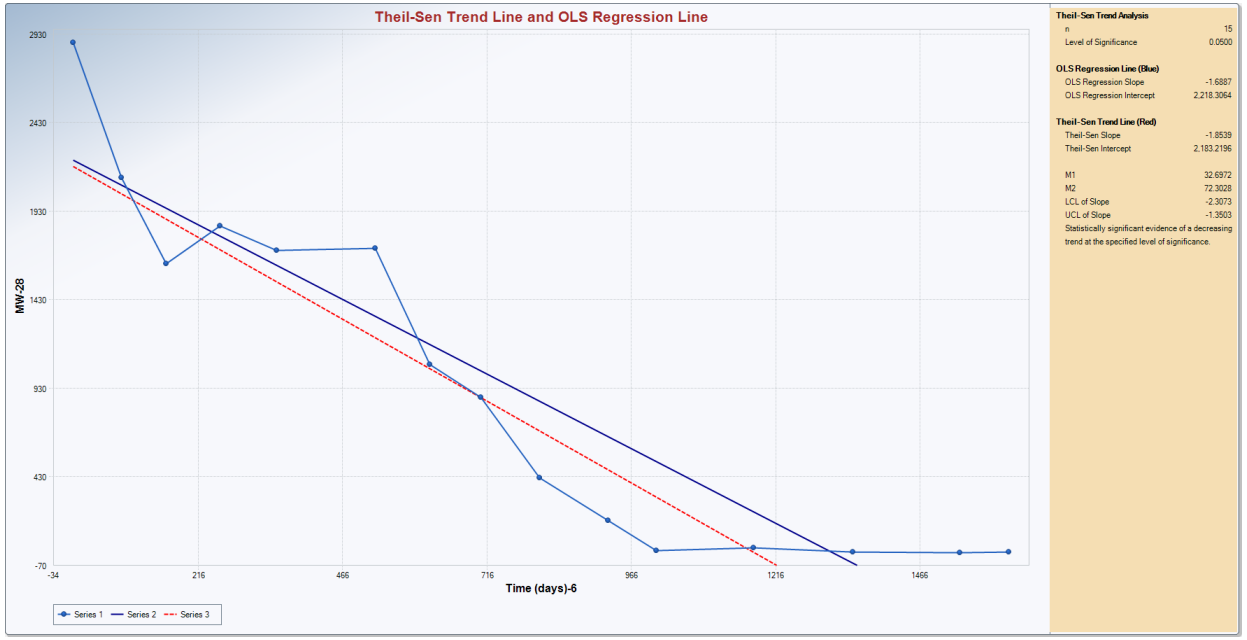


1. Click **Statistical Tests ►Trend Analysis ► Theil-Sen.**



Advantages / Disadvantages

- Identifies trend
- Quantifies slope
  - The rate of decrease / increase
- Robust to presence of extreme values
- Handles only 1 observation at each sampling event



When to transform the data?

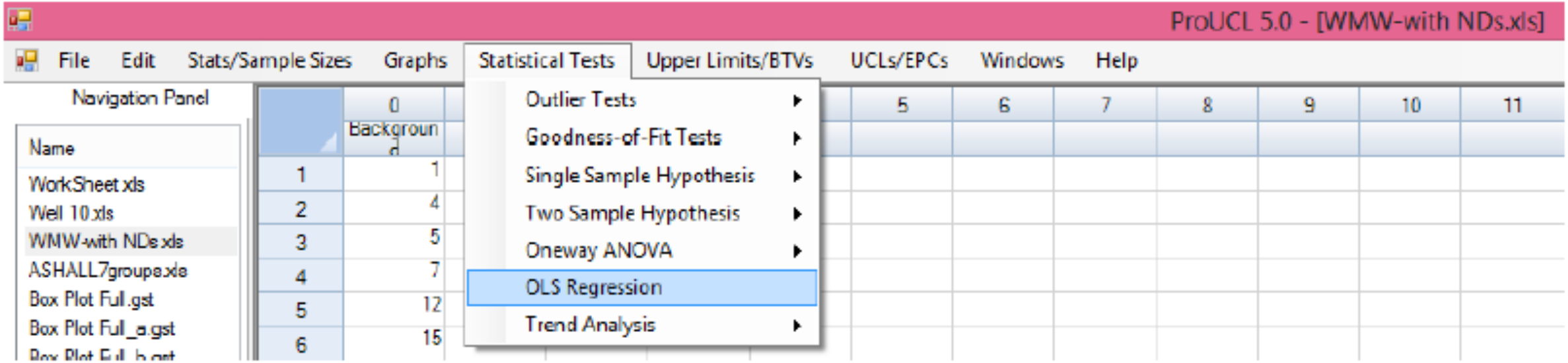
Visible exponential decay trend

- 1st order reactions
  - Rate of decay proportional to amount of contaminant present
  - Dilution with little or no desorption from soil
- Log transformation is needed

Visible linear decay

- 0 order reactions
  - Decay rate constant rather than proportional to amount of contaminant
- No transformation needed

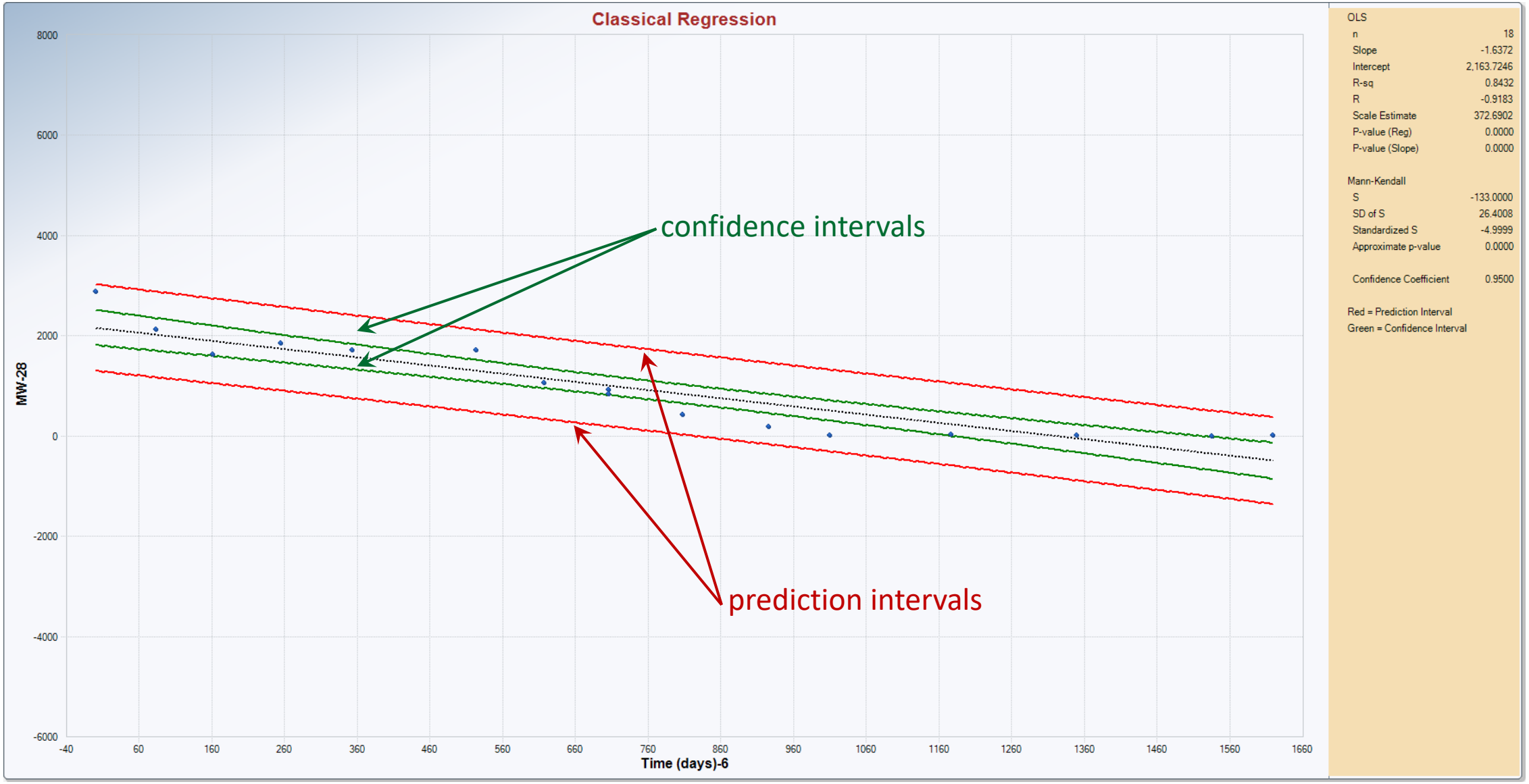
. Click **Statistical Tests▶ OLS Regression.**



Ordinary Least Square (OLS) Regression

- Select regression variables
- Fit regression
- Check assumptions
- Evaluate model and Interpret Results

Regression Plot



Evaluate Model – Regression Equation

Regression Estimates and Inference Table				
Parameter	Estimates	Std. Error	T-values	p-values
intercept	2164	165.3	13.09	5.793E-10
Time (days)-6 (slope)	-1.637	0.176	-9.276	7.7292E-8

Small p- values indicate significance $p<0.05$

Contaminant = 2164 -1.637 * Time

intercept

slope

Evaluate model – ANOVA table and R^2

OLS ANOVA Table					
Source of Variation	SS	DOF	MS	F-Value	P-Value
Regression	11952431	1	11952431	86.05	0.0000
Error	2222368	16	138898		
Total	14174799	17			

Small p-value indicates significance $p<0.05$

R Square	0.843
Adjusted R Square	0.833
Sqrt(MSE) = Scale	372.7

% of variability explained by regression model

Advantages / Disadvantages

- Standard statistical method for identifying trends
- Assumptions for residuals need to be satisfied
- Confidence bands can be used to determine compliance with fixed standards even when a trend is apparent
  - Concentration of contaminant is changing
- Sensitive to extreme values (outliers)

Assumptions for Residuals

$$Residuals = observed\ value\ - fitted\ value$$

- Constant variance (homoscedasticity of residuals)
- Independent
- Normally distributed

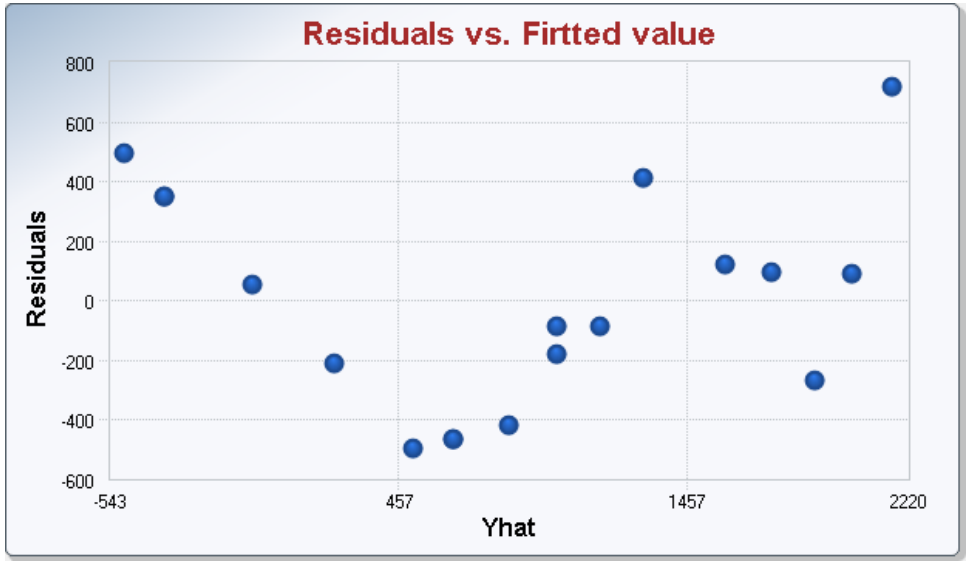
Prepare OLS Regression Table to Check Assumptions

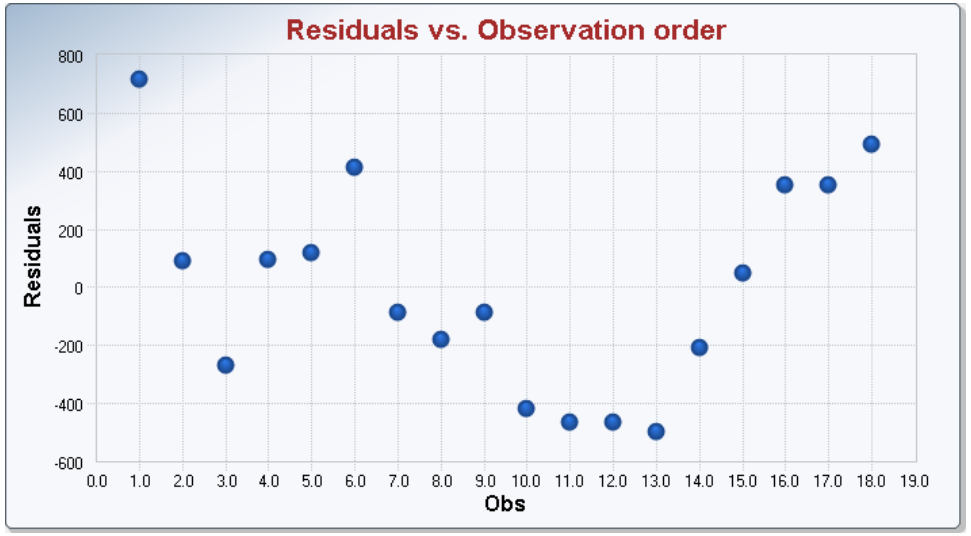
- Copy Regression Table to EXCEL
  - May have leading spaces!!
- Helpful trick to remove leading spaces:
  - select Y vector to Res/Scale columns
  - choose replace
  - hit spacebar in find field,
  - Click replace all

Obs	Y Vector	Yhat	Residuals
1	2880.0	2164.0	716.3
2	2117.0	2028.0	89.17
3	1633.0	1900.0	-267.6
4	1845.0	1748.0	97.13
5	1706.0	1587.0	118.2
6	1719.0	1307.0	411.1
7	1065.0	1154.0	-88.55
8	831.8	1009.0	-177.7
9	920.6	1009.0	-88.87
10	424.6	842.5	-417.9
11	181.1	647.6	-466.5
12	184.9	647.6	-462.7
13	14.0	511.8	-497.8
14	26.8	236.7	-209.9
15	5.9	-44.9	50.81
16	1.7	-349.4	351.1
17	1.8	-349.4	351.2
18	5.5	-487	492.5

Residual Plots

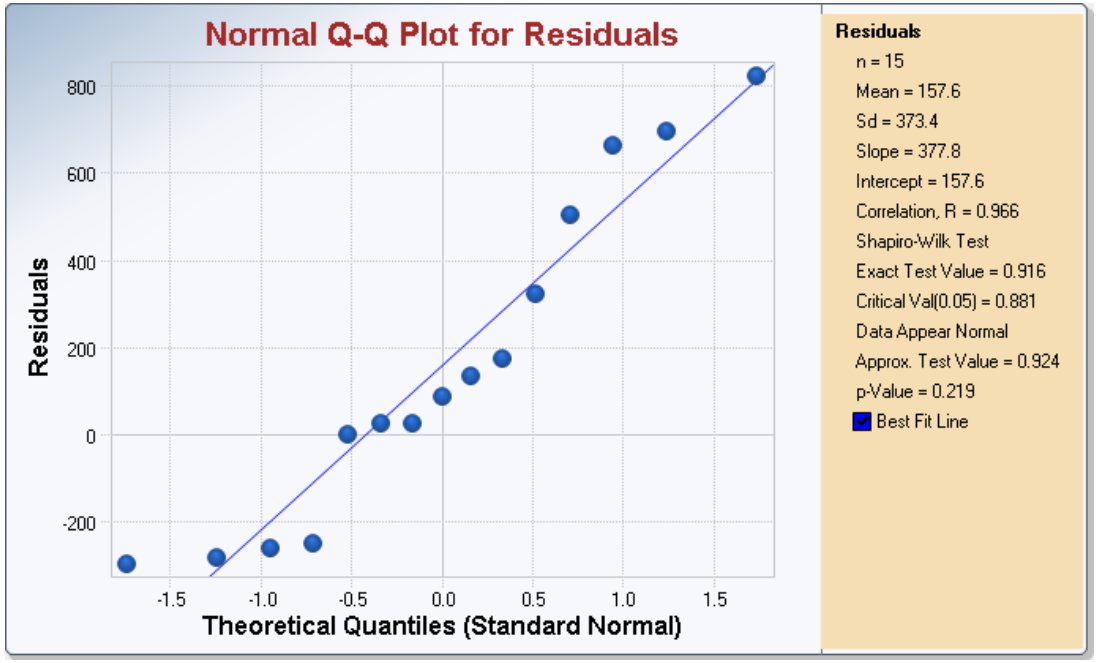
- Plot 1
  - Y= Residuals
  - X= Fitted values (Estimates)
  - Indicator of constant variance
    - Random scatter
- Plot 2
  - Y = Residuals
  - X = Observation order (Event #)
  - Indicator of independence
    - Random scatter of residuals





Explore Normality of Residuals

- Import Regression table from Excel to Pro UCL
  - Goodness of Fit test
- Normality required for reliable:
  - trend test, it assumes normality of residuals!!
  - confidence and prediction intervals
- Non-normal residuals
  - Likely indicate lack of fit
    - Need to improve the model



Trend analysis Summary

Method	Assumptions	Information	Limitations
Theil-Sen	None	Identifies trend Evaluates slope	Fair amount of data 1 observation per sampling
Mann - Kendal	None	Is trend monotone?	Min 4 points
OLS	Normality Constant variance Independence	Quantifies slope and intercept	Fair amount of data Valid assumptions

Final remarks

- Take time to explore the data
- Knowing something about chemical & bio processes can be helpful
- Residuals analysis provides a lot of information
- When in doubt consult statistician
- Document well steps of analysis and decisions you make

Next session: Background Level Calculations

- Objective: Explore ProUCL features for Background to Site comparison
  - Sampling design and sample size selection
  - UTLs and background threshold values
  - Comparison of BTVs to site data

Next ProUCL Webinar

ProUCL Utilization 2020: Part 3: Background Level Calculations

Mar 9, 2020

1:00PM-2:30PM EST

Contact Information for ProUCL

Felicia Barnett, EPA SCMTSC

barnett.felicia@epa.gov

Travis Linscome-Hatfield, Neptune and Company, Inc

travis@neptuneinc.org

Polona Carson, Neptune and Company, Inc

pcarson@neptuneinc.org