

## Evaluation of Bold and Main Basin Plus Main Basin and Reference Area Dioxin Data

### 1. Purpose

This memorandum describes the Dredged Material Management Program's (DMMP) statistical analysis of polychlorinated dioxin and furan data representing background in Puget Sound. This analysis was used by the DMMP agencies to select an appropriate background value to serve as a suitability criterion for open-water disposal of dredged material in Puget Sound. As noted in Attachment 1 of the 2009 SMARM Issue Paper, entitled "Agency Proposal for Establishing Dioxin Suitability Guidelines for Open-Water Disposal at Non-Dispersive Sites," the proposal selected guidance values at the 90th percentile of the population at the 90 percent upper confidence limit of the background data set. This memorandum provides the background on the methods used in these calculations.

The data sets considered here include high resolution dioxin and furan data from the following three sources:

- **Data collected by the DMMP using OSV Bold in August 2008** - The "Bold data set" consists of 70 stations located throughout Puget Sound.<sup>1</sup> Sample locations were all located outside urban embayments and away from known point sources of contamination. While 50 of the samples from this data set are classified as being within the Main Basin of Puget Sound, the remaining 20 samples are from reference bays sampled as part of that investigation.<sup>2</sup>
- **Data collected by the DMMP near the Anderson-Ketron (A/K) open water disposal site** - DMMP monitoring of the A/K site has occurred in 2005 and 2008. This site is the only open water disposal site that is located outside an urban bay in the main-basin of Puget Sound. Data for 13 samples were used from stations excluding onsite locations. Also excluded was the 2008 perimeter station that was impacted by an accidental disposal event (AKP01). The A/K disposal site data are included because they all met the minimum distance screen (>500 meters) relative to the disposal site boundary and because testing has shown that there is no difference between the site itself and the benchmark stations,<sup>3</sup> suggesting a lack of impact by dioxin.
- **Data from Ecology's Environmental Information Management System (EIM)** Thirteen data points from reference areas were derived from a query of Ecology's EIM database, as well as one reference sample associated with a recent dredge

---

<sup>1</sup> Five laboratory duplicates were also analyzed but only the primary samples are considered here.

<sup>2</sup> The Bold data set is posted on the DMMP web site.

<sup>3</sup> Benchmark stations are located well beyond the disposal site and region affected by disposal activity.

project sediment characterization (SamishRef). These 14 data points all met the screening requirements of being location outside of an urban embayment and having a minimum distance (500 m) from known cleanup sites, outfalls, and other known sources of contamination.

Combination of the aforementioned data resulted in a data set referred to as the “Bold plus existing main basin and reference bay data.” This combined data set was intended by the agencies for use in displaying options for a statistically-based suitability criterion.

Spreadsheets in Microsoft™ Excel and the freely-available EPA software, ProUCL v. 4<sup>4</sup> showing all data used to illustrate the calculations are available by request.

## **2. Methods and Materials**

2.1 The first step is calculation of nonparametric Kaplan-Meier (K-M) sample TEQ sums for all the below-detection-limit observations. This is done in accordance with the procedure suggested by Dr. Dennis Helsel at Practicalstats.com (formerly with USGS) during the Regional Sediment Evaluation Team Statistics Workshop in November, 2008.<sup>5</sup> This procedure, which is currently awaiting 2009 publication by Dr. Helsel in the *Journal of Environmental Science and Technology*, was selected due to the presence of many samples with below-detection-limit observations. EPA's ProUCL software is used in conjunction with Crystal Ball™ and Excel™ for organizing and summarizing the data. In addition to the method for K-M estimation, the statisticians strongly recommended nonparametric (distribution-free) techniques to establish a comparison for background “populations” of samples such as are treated here.

Appendix A documents the manipulations necessary to accomplish the K-M sum estimation and background calculations. It also points out the portions of the spreadsheets that contain the manipulations and calculations.

2.2 The second step is statistical summary of K-M TEQ sum results for the dataset. This step includes graphic comparison of the K-M TEQ sample sums to the sample sums derived from substituting 0 from below-detection-limit observations (ND=0) and substituting on-half the detection limits for such observations (ND=0.5DL). Statistical experts at the cited workshop recommended against these ND=0 and ND=0.5DL substitutions in favor of the nonparametric K-M method. The substitution methods have been the means by which dioxin/furan data are commonly summarized, and so are provided for comparison. ProUCL and Excel statistical functions were used to evaluate data.

---

<sup>4</sup> <http://www.epa.gov/nerlesd1/tsc/software.htm>

<sup>5</sup> [http://www.nws.usace.army.mil/Portals/27/docs/civilworks/dredging/Stats\\_Workshop\\_Report.pdf](http://www.nws.usace.army.mil/Portals/27/docs/civilworks/dredging/Stats_Workshop_Report.pdf)

2.3 The third step is calculation of a range of potential guidelines for “Bold plus existing main basin and reference bay” dataset by nonparametric estimation of the 90th and 95th upper confidence intervals of the 80<sup>th</sup>, 85<sup>th</sup>, and 90<sup>th</sup> percentile coverages or population proportion. As noted above, nonparametric methods such as K-M were recommended in the RSET statistics workshop. These were calculated using ProUCL. During this step, the potential contribution of within-sample and within-laboratory variability were also evaluated.

### 3. Results of Step 1, calculation of K-M sum of TEQ for individual samples

3.1 Figure 1 shows a quantile-quantile plot of these results. The K-M estimate falls between the ND=0 and ND=0.5DL estimates, and does not fundamentally alter the pattern of the data.

3.2 Table 1 displays results of the three methods of determining sample TEQ sums.

Figure 1. Q-Q Plot of Three Methods for Calculating Dioxin TEQ

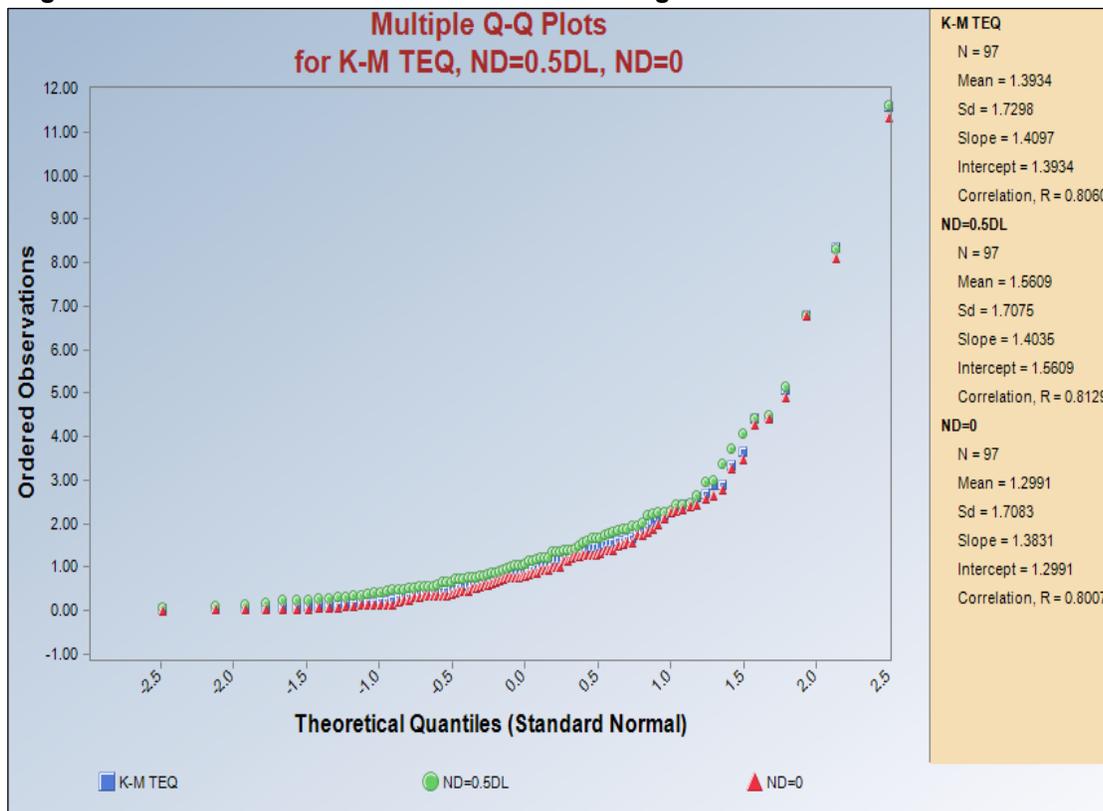


Table 1. Results of 3 Methods of Estimating Sample TEQ Sums. (Yellow cells are Bold data; blue cells are other reference areas.)

| Station    | K-M TEQ | K-M Note | ND=0.5DL | ND=0  |
|------------|---------|----------|----------|-------|
| AI_1       | 0.423   |          | 0.564    | 0.358 |
| AI_11_C    | 0.044   |          | 0.258    | 0.023 |
| AI_13_C    | 0.452   |          | 0.568    | 0.376 |
| AI_20_C_GS | 0.619   |          | 0.670    | 0.567 |
| AI_5_C     | 0.408   |          | 0.544    | 0.329 |
| AKB02-2005 | 1.625   |          | 1.685    | 1.564 |
| AKB02-2008 | 1.195   |          | 1.483    | 0.954 |
| AKB03-2005 | 2.618   |          | 2.637    | 2.560 |
| AKB03-2008 | 1.610   |          | 1.887    | 1.398 |
| AKP01-2005 | 6.794   | a        | 6.794    | 6.794 |
| AKP02-2005 | 2.346   |          | 2.420    | 2.266 |
| AKP02-2008 | 2.431   |          | 2.466    | 2.328 |
| AKP03-2005 | 4.404   | a        | 4.404    | 4.404 |
| AKP03-2008 | 2.091   |          | 1.956    | 1.377 |
| AKP04-2005 | 4.403   |          | 4.474    | 4.276 |
| AKP04-2008 | 1.887   |          | 2.271    | 1.796 |
| AKT01-2005 | 2.890   |          | 2.972    | 2.794 |
| AKT01-2008 | 1.503   |          | 1.793    | 1.297 |
| CPS_0      | 1.853   |          | 1.952    | 1.741 |
| CPS_1      | 2.193   |          | 2.217    | 2.127 |
| CPS_3      | 1.331   |          | 1.366    | 1.275 |
| CPS_4      | 0.949   |          | 1.036    | 0.844 |
| CPS_5      | 0.655   |          | 0.733    | 0.582 |
| CR02       | 2.278   |          | 2.290    | 2.285 |
| CR23       | 0.728   |          | 0.743    | 0.701 |
| CR23W      | 0.440   |          | 0.479    | 0.333 |
| CR24       | 0.780   |          | 0.784    | 0.770 |

| Station               | K-M TEQ | K-M Note | ND=0.5DL | ND=0  |
|-----------------------|---------|----------|----------|-------|
| Grays OM90 Sequim     | 2.890   |          | 2.026    | 0.030 |
| Grays OM90 West Beach | 0.149   |          | 4.065    | 2.655 |
| HC_0                  | 0.886   |          | 1.015    | 0.795 |
| HC_1                  | 0.802   |          | 0.871    | 0.721 |
| HC_2                  | 0.774   |          | 1.154    | 0.584 |
| HC_3                  | 0.444   |          | 0.646    | 0.355 |
| HC_6                  | 0.493   |          | 0.769    | 0.359 |
| MSMP-43               | 0.180   |          | 0.274    | 0.138 |
| NCPS_0                | 0.646   |          | 0.758    | 0.532 |
| NCPS_1                | 0.079   |          | 0.261    | 0.014 |
| NCPS_2                | 1.068   |          | 1.231    | 0.923 |
| NCPS_3                | 0.675   |          | 0.763    | 0.617 |
| NCPS_4                | 0.298   |          | 0.498    | 0.243 |
| PASED08 RF03A         | 0.857   |          | 0.857    | 0.856 |
| PASED08-RF01A         | 0.041   |          | 0.062    | 0.034 |
| PASED08-RF02A         | 0.055   |          | 0.101    | 0.041 |
| PSPS_1                | 2.040   |          | 2.194    | 1.879 |
| PSPS_2                | 2.686   |          | 2.980    | 2.387 |
| PSPS_3                | 0.862   |          | 1.056    | 0.749 |
| PSPS_8                | 0.105   |          | 0.327    | 0.067 |
| PSPS_9                | 1.464   |          | 1.670    | 1.325 |
| R_CAR_0               | 0.598   |          | 0.823    | 0.444 |
| R_CAR_1               | 1.040   |          | 1.215    | 0.865 |
| R_CAR_4               | 0.838   |          | 1.068    | 0.652 |
| R_CAR_5               | 5.066   |          | 5.152    | 4.920 |
| R_CAR_6_C             | 0.211   |          | 0.510    | 0.135 |

| Station       | K-M TEQ | K-M Note | ND=0.5DL | ND=0  |
|---------------|---------|----------|----------|-------|
| R_DAB_0       | 0.257   |          | 0.581    | 0.151 |
| R_DAB_1       | 1.578   |          | 1.685    | 1.486 |
| R_DAB_2       | 1.443   |          | 1.789    | 1.229 |
| R_DAB_5       | 1.532   |          | 1.848    | 1.297 |
| R_DAB_7_C     | 1.202   |          | 1.225    | 1.156 |
| R_HOL_0       | 0.121   |          | 0.295    | 0.080 |
| R_HOL_1       | 0.372   |          | 0.401    | 0.332 |
| R_HOL_3       | 0.100   |          | 0.241    | 0.061 |
| R_HOL_4       | 1.199   |          | 1.434    | 1.020 |
| R_HOL_7       | 0.862   |          | 0.946    | 0.778 |
| R_SAM_0       | 1.321   |          | 1.390    | 1.261 |
| R_SAM_1       | 1.561   |          | 1.604    | 1.523 |
| R_SAM_3       | 1.324   |          | 1.381    | 1.285 |
| R_SAM_4       | 0.877   |          | 1.035    | 0.814 |
| R_SAM_5       | 1.836   |          | 1.888    | 1.753 |
| RAYONR05S1-01 | 0.051   |          | 0.155    | 0.027 |
| RAYONR05S1-02 | 0.110   |          | 0.181    | 0.090 |
| RAYONR05S1-03 | 0.155   |          | 0.365    | 0.095 |
| SamishRef     | 2.439   | a        | 2.439    | 2.439 |
| SCPS_1        | 3.349   |          | 3.387    | 3.261 |
| SCPS_10_C     | 1.088   |          | 1.158    | 1.017 |
| SCPS_2        | 0.508   |          | 0.566    | 0.442 |
| SCPS_3        | 0.177   |          | 0.340    | 0.126 |
| SCPS_5        | 3.655   |          | 3.737    | 3.471 |

| Station     | K-M TEQ | K-M Note | ND=0.5DL | ND=0   |
|-------------|---------|----------|----------|--------|
| SJF_10_C    | 0.323   |          | 0.465    | 0.247  |
| SJF_12_C_GS | 1.678   |          | 1.747    | 1.575  |
| SJF_2       | 0.275   |          | 0.500    | 0.212  |
| SJF_3       | 0.163   |          | 0.420    | 0.123  |
| SJF_9_C     | 0.536   |          | 0.803    | 0.417  |
| SJI_0       | 0.677   |          | 0.991    | 0.516  |
| SJI_1       | 0.828   |          | 0.899    | 0.770  |
| SJI_20_C_GS | 1.149   |          | 1.341    | 1.019  |
| SJI_3       | 0.445   |          | 0.667    | 0.352  |
| SJI_8_C     | 0.556   |          | 0.722    | 0.439  |
| SPSB_0      | 1.457   |          | 1.577    | 1.403  |
| SPSB_1      | 1.265   |          | 1.384    | 1.164  |
| SPSB_2      | 2.142   |          | 2.271    | 1.968  |
| SPSB_3      | 0.192   |          | 0.421    | 0.132  |
| SPSB_8_C    | 0.073   |          | 0.319    | 0.029  |
| SS_0        | 8.347   |          | 8.311    | 8.121  |
| SS_1        | 0.393   |          | 0.525    | 0.331  |
| SS_2        | 1.331   |          | 1.367    | 1.258  |
| SS_8_C      | 1.103   |          | 1.186    | 0.954  |
| SS_9_C      | 11.594  |          | 11.626   | 11.354 |

<sup>a</sup> These samples had no congeners below detection limits. There was no corresponding K-M value, and the standard sum of TEQs is shown.

## 4. Results of Step 2

Table 2 displays a parametric estimation of the median and the 25<sup>th</sup>, 80<sup>th</sup>, 85<sup>th</sup>, and 90<sup>th</sup> percentiles of the distribution of the data.

**Table 2. Summary Statistics of Kaplan-Meier TEQs**

|                             | <b>Bold Plus Existing Main Basin and Reference Bays<sup>a</sup></b> |
|-----------------------------|---|
| 25 <sup>th</sup> Percentile | 0.42  |
| Median                      | 0.88  |
| 80th Percentile             | 2.26  |
| 85th Percentile             | 2.68  |
| 90th Percentile             | 3.24  |

<sup>a</sup> Gamma distribution recommended by ProUCL. Lognormal recommended by Crystal Ball.

## 5. Results of Step 3

5.1 Development of K-M Nonparametric Estimates for Potential Background-based Suitability Values.

Table 3 displays the nonparametric estimation of the 90<sup>th</sup> and 95<sup>th</sup> upper confidence intervals for 80<sup>th</sup>, 85<sup>th</sup>, and 90<sup>th</sup> percentile coverages.

**Table 3. Nonparametric Estimation of Potential Guideline Values**

| <b>Coverage (% of Population)</b> | <b>Upper Confidence Limit</b> | <b>Bold Plus Existing Main Basin and Reference Bays</b> |
|-----------------------------------|-------------------------------|---|
| 80th Percentile                   | 90%                           | 2.28  |
|                                   | 95%                           | 2.35  |
| 85th Percentile                   | 90%                           | 2.62  |
|                                   | 95%                           | 2.89  |
| 90th Percentile                   | 90%                           | 3.66  |
|                                   | 95%                           | 3.66  |

5.2 Duplicate Sample Variability from the “Bold plus existing main basin and reference bay data.”

Analytical and field-sampling variability may be used to estimate a portion of the uncertainty associated with a data set. In these data sets, there were six relevant, field-split samples analyzed as single-blind laboratory duplicates (that is, the laboratory did not know they were duplicates.).

**Table 4. Evaluation of Duplicate Results from 2008 Bold and 2005 Anderson-Ketron Dioxin Investigations**

|                 | <b>K-M Sum</b> | <b>Relative Percent Difference<sup>6</sup></b> | <b>Difference (K-M Sums)</b> |
|-----------------|----------------|--|------------------------------|
|                 | ng TEQ/kg      | K-M to K-M, %                                  | ng TEQ/kg                    |
| CPS_3_Dup       | 1.55           | 15%  | 0.22                         |
| CPS_3           | 1.33           |  |                              |
| HC_2_Dup        | 3.33           | 125%   | 2.56                         |
| HC_2            | 0.77           |  |                              |
| NCPS_2_Dup      | 0.92           | 15%  | 0.14                         |
| NCPS_2          | 1.07           |  |                              |
| PSPS_1_Dup      | 0.95           | 73%  | 1.09                         |
| PSPS_1          | 2.04           |  |                              |
| SPSB_0_Dup      | 1.57           | 5%   | 0.12                         |
| SPSB_0          | 1.46           |  |                              |
| AKB03-2005      | 2.60           | 19%  | 0.44                         |
| AKB03-2005 Dup  | 2.16           |  |                              |
| Range           |                | 5 - 125%                                       | 0.12 - 2.56                  |
| Mean Difference |                | 27%  | 0.38                         |

## 6. Summary

The “Bold plus existing main basin and reference bay” data set was used to create a range of potential suitability guidelines using a nonparametric statistical evaluation. The range of 80<sup>th</sup> - 90<sup>th</sup> percentile coverage and 90<sup>th</sup> – 95<sup>th</sup> percentile confidence level of this data set was 2.3 - 3.7 ng TEQ/kg. The range of analytical- and heterogeneity-related uncertainty associated with the data set was 0.1-2.6 ng TEQ/kg, with an average of 0.4 ng TEQ/kg.

<sup>6</sup> The absolute difference of the 2 values divided by their average, expressed as a percent.

## Appendix A. Instructions for Using Pro-UCL to Create K-M Sums for Samples, and Key to the Spreadsheets.

1. Data as received resembles the following table. The order of the congeners is unimportant, but the pairing with the TEF (as well as use of the WHO 2005 TEFs) is critical.

| Sample ID<br>Analyte (pg/g) | TEF    | AI_1      |   | AI_11_C  |   |
|-----------------------------|--------|-----------|---|----------|---|
|                             |        | 7/31/2008 | Q | 8/1/2008 | Q |
| % Solids                    |        | 68.6      |   | 75.1     |   |
| 2,3,7,8-TCDD                | 1      | 0.144     | U | 0.141    | U |
| 1,2,3,7,8-PeCDD             | 1      | 0.21      | U | 0.144    | U |
| 1,2,3,4,7,8-HxCDD           | 0.1    | 0.227     | U | 0.209    | U |
| 1,2,3,6,7,8-HxCDD           | 0.1    | 0.431     | J | 0.217    | U |
| 1,2,3,7,8,9-HxCDD           | 0.1    | 0.349     | J | 0.225    | U |
| 1,2,3,4,6,7,8-<br>HpCDD     | 0.01   | 4.19      | J | 0.367    | J |
| OCDD                        | 0.0003 | 24.5      |   | 2.22     | U |
| 2,3,7,8-TCDF                | 0.1    | 0.404     | J | 0.192    | J |
| 1,2,3,7,8-PeCDF             | 0.03   | 0.122     | U | 0.118    | U |
| 2,3,4,7,8-PeCDF             | 0.3    | 0.269     | J | 0.117    | U |
| 1,2,3,4,7,8-HxCDF           | 0.1    | 0.264     | J | 0.174    | U |
| 1,2,3,6,7,8-HxCDF           | 0.1    | 0.317     | J | 0.163    | U |
| 1,2,3,7,8,9-HxCDF           | 0.1    | 0.275     | U | 0.243    | U |
| 2,3,4,6,7,8-HxCDF           | 0.1    | 0.315     | J | 0.171    | U |
| 1,2,3,4,6,7,8-<br>HpCDF     | 0.01   | 1.92      | J | 0.225    | U |
| 1,2,3,4,7,8,9-<br>HpCDF     | 0.01   | 0.383     | U | 0.348    | U |
| OCDF                        | 0.0003 | 2.57      | J | 0.54     | J |

### 2. Preparation for ProUCL.

2.1 Copy the data to a new worksheet. In the above example, no TEQs have been calculated. If TEQs had been provided and if the formulas used a substitution such as  $ND=0.5DL$ , those columns should be eliminated as should any unnecessary data (such as the % solids row).

2.2 A new column should be inserted to the right of the raw concentration data. Then, the sample identifier (AI\_1) should be copied over the new column and the Q column for that sample. It will look as follows:

|    | A                   | B      | C         | D    | E    |
|----|---------------------|--------|-----------|------|------|
| 1  | Sample ID           |        | AI_1      |      |      |
| 2  | Analyte (pg/g)      | TEF    | 7/31/2008 | AI_1 | AI_1 |
| 3  | 2,3,7,8-TCDD        | 1      | 0.144     |      |      |
| 4  | 1,2,3,7,8-PeCDD     | 1      | 0.21      |      | U    |
| 5  | 1,2,3,4,7,8-HxCDD   | 0.1    | 0.227     |      | U    |
| 6  | 1,2,3,6,7,8-HxCDD   | 0.1    | 0.431     |      | U    |
| 7  | 1,2,3,7,8,9-HxCDD   | 0.1    | 0.349     |      | J    |
| 8  | 1,2,3,4,6,7,8-HpCDD | 0.01   | 4.19      |      | J    |
| 9  | OCDD                | 0.0003 | 24.5      |      | J    |
| 10 | 2,3,7,8-TCDF        | 0.1    | 0.404     |      |      |
| 11 | 1,2,3,7,8-PeCDF     | 0.03   | 0.122     |      | J    |
| 12 | 2,3,4,7,8-PeCDF     | 0.3    | 0.269     |      | U    |
| 13 | 1,2,3,4,7,8-HxCDF   | 0.1    | 0.264     |      | J    |
| 14 | 1,2,3,6,7,8-HxCDF   | 0.1    | 0.317     |      | J    |
| 15 | 1,2,3,7,8,9-HxCDF   | 0.1    | 0.275     |      | J    |
| 16 | 2,3,4,6,7,8-HxCDF   | 0.1    | 0.315     |      | U    |
| 17 | 1,2,3,4,6,7,8-HpCDF | 0.01   | 1.92      |      | J    |
| 18 | 1,2,3,4,7,8,9-HpCDF | 0.01   | 0.383     |      | J    |
| 19 | OCDF                | 0.0003 | 2.57      |      | U    |

2.3 Next, create a formula to multiply the concentration times the TEF. In the present case, in cell D3 the formula is =C3\*\$B\$3. The \$ code is an absolute reference, and permits one to copy this to other cells for other samples at the same row. Copy this formula down that column. Check that it always references the correct TEF to the left of the concentration (e.g., for cell D13, the formula should be =C13\*\$B\$13). Now it should resemble the following:

| Sample ID           |        | AI_1      |         |      |
|---------------------|--------|-----------|---------|------|
| Analyte (pg/g)      | TEF    | 7/31/2008 | AI_1    | AI_1 |
| 2,3,7,8-TCDD        | 1      | 0.144     | 0.144   |      |
| 1,2,3,7,8-PeCDD     | 1      | 0.21      | 0.21    | U    |
| 1,2,3,4,7,8-HxCDD   | 0.1    | 0.227     | 0.0227  | U    |
| 1,2,3,6,7,8-HxCDD   | 0.1    | 0.431     | 0.0431  | U    |
| 1,2,3,7,8,9-HxCDD   | 0.1    | 0.349     | 0.0349  | J    |
| 1,2,3,4,6,7,8-HpCDD | 0.01   | 4.19      | 0.0419  | J    |
| OCDD                | 0.0003 | 24.5      | 0.00735 | J    |
| 2,3,7,8-TCDF        | 0.1    | 0.404     | 0.0404  |      |
| 1,2,3,7,8-PeCDF     | 0.03   | 0.122     | 0.00366 | J    |
| 2,3,4,7,8-PeCDF     | 0.3    | 0.269     | 0.0807  | U    |
| 1,2,3,4,7,8-HxCDF   | 0.1    | 0.264     | 0.0264  | J    |

| Sample ID           |        | AI_1      |          |      |
|---------------------|--------|-----------|----------|------|
| Analyte (pg/g)      | TEF    | 7/31/2008 | AI_1     | AI_1 |
| 1,2,3,6,7,8-HxCDF   | 0.1    | 0.317     | 0.0317   | J    |
| 1,2,3,7,8,9-HxCDF   | 0.1    | 0.275     | 0.0275   | J    |
| 2,3,4,6,7,8-HxCDF   | 0.1    | 0.315     | 0.0315   | U    |
| 1,2,3,4,6,7,8-HpCDF | 0.01   | 1.92      | 0.0192   | J    |
| 1,2,3,4,7,8,9-HpCDF | 0.01   | 0.383     | 0.00383  | J    |
| OCDF                | 0.0003 | 2.57      | 0.000771 | U    |

2.4 ProUCL needs to have columns that show the result with ND=DL, and a code that says whether it is detected or not, i.e., 1 for detected and 0 for nondetected. First, look over the data set and identify the codes that represent nondetected congeners. These are typically U and UJ, but there may be others such as UJK. Then, be sure to highlight a range of data that doesn't include your sample ID names. (Otherwise, "Sequim" becomes "SeqOim") Substitute for nondetected codes a "0" using ctrl-F search and replace. It is recommended that the longer strings are done first, such as UJ, because Excel will replace *within* a cell and if done in another order (say, U first), it will not see the changed value, OJ, as nondetected. Note: where there are no qualifiers, one must manually insert "1"s. One may alternatively wish to write "=if" functions to do this entire operation. Upon completion, it should appear as following:

| Sample ID           |        | AI_1      |          |      |
|---------------------|--------|-----------|----------|------|
| Analyte (pg/g)      | TEF    | 7/31/2008 | AI_1     | AI_1 |
| 2,3,7,8-TCDD        | 1      | 0.144     | 0.144    | 1    |
| 1,2,3,7,8-PeCDD     | 1      | 0.21      | 0.21     | 0    |
| 1,2,3,4,7,8-HxCDD   | 0.1    | 0.227     | 0.0227   | 0    |
| 1,2,3,6,7,8-HxCDD   | 0.1    | 0.431     | 0.0431   | 0    |
| 1,2,3,7,8,9-HxCDD   | 0.1    | 0.349     | 0.0349   | 1    |
| 1,2,3,4,6,7,8-HpCDD | 0.01   | 4.19      | 0.0419   | 1    |
| OCDD                | 0.0003 | 24.5      | 0.00735  | 1    |
| 2,3,7,8-TCDF        | 0.1    | 0.404     | 0.0404   | 1    |
| 1,2,3,7,8-PeCDF     | 0.03   | 0.122     | 0.00366  | 1    |
| 2,3,4,7,8-PeCDF     | 0.3    | 0.269     | 0.0807   | 0    |
| 1,2,3,4,7,8-HxCDF   | 0.1    | 0.264     | 0.0264   | 1    |
| 1,2,3,6,7,8-HxCDF   | 0.1    | 0.317     | 0.0317   | 1    |
| 1,2,3,7,8,9-HxCDF   | 0.1    | 0.275     | 0.0275   | 0    |
| 2,3,4,6,7,8-HxCDF   | 0.1    | 0.315     | 0.0315   | 1    |
| 1,2,3,4,6,7,8-HpCDF | 0.01   | 1.92      | 0.0192   | 1    |
| 1,2,3,4,7,8,9-HpCDF | 0.01   | 0.383     | 0.00383  | 1    |
| OCDF                | 0.0003 | 2.57      | 0.000771 | 0    |

2.5 There are two data columns (actually three, but the one with the date will be eliminated shortly) that say AI\_1. ProUCL has to have a "D\_" or a "d\_" code in front of

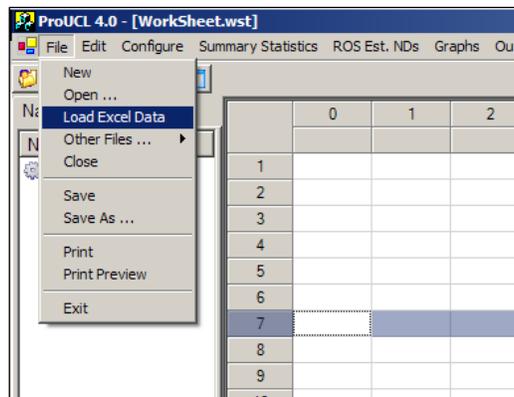
the detect/nondetect column. Edit the right-most column to say **D\_AI\_1**. Repeat this activity across the spreadsheet for every sample. Don't forget at the end of this time to zoom out and assure that there is a detect/nondetect code in every cell. If there isn't, ProUCL will assume the value is "1", and calculate a K-M value for the congener.

2.6 Copy this to another worksheet. It is important to copy the data using "Paste Special" and check the box that says "values only" (to prevent the dependant cells from recalculating after deleting columns). Remove the first three columns and all columns with dates in them. This is almost ready for import to ProUCL.

| AI_1     | D_AI_1 |
|----------|--------|
| 0.144    | 1      |
| 0.21     | 0      |
| 0.0227   | 0      |
| 0.0431   | 0      |
| 0.0349   | 1      |
| 0.0419   | 1      |
| 0.00735  | 1      |
| 0.0404   | 1      |
| 0.00366  | 1      |
| 0.0807   | 0      |
| 0.0264   | 1      |
| 0.0317   | 1      |
| 0.0275   | 0      |
| 0.0315   | 1      |
| 0.0192   | 1      |
| 0.00383  | 1      |
| 0.000771 | 0      |

2.7 Copy that worksheet to a new Excel file. ProUCL can import data directly, but it cannot deal with multiple worksheets. Save the Excel file. Since this spreadsheet is just a copy, after it is imported to ProUCL, it can be deleted.

3. Import to ProUCL. Open ProUCL. The following shows the approach to import from the spreadsheet.



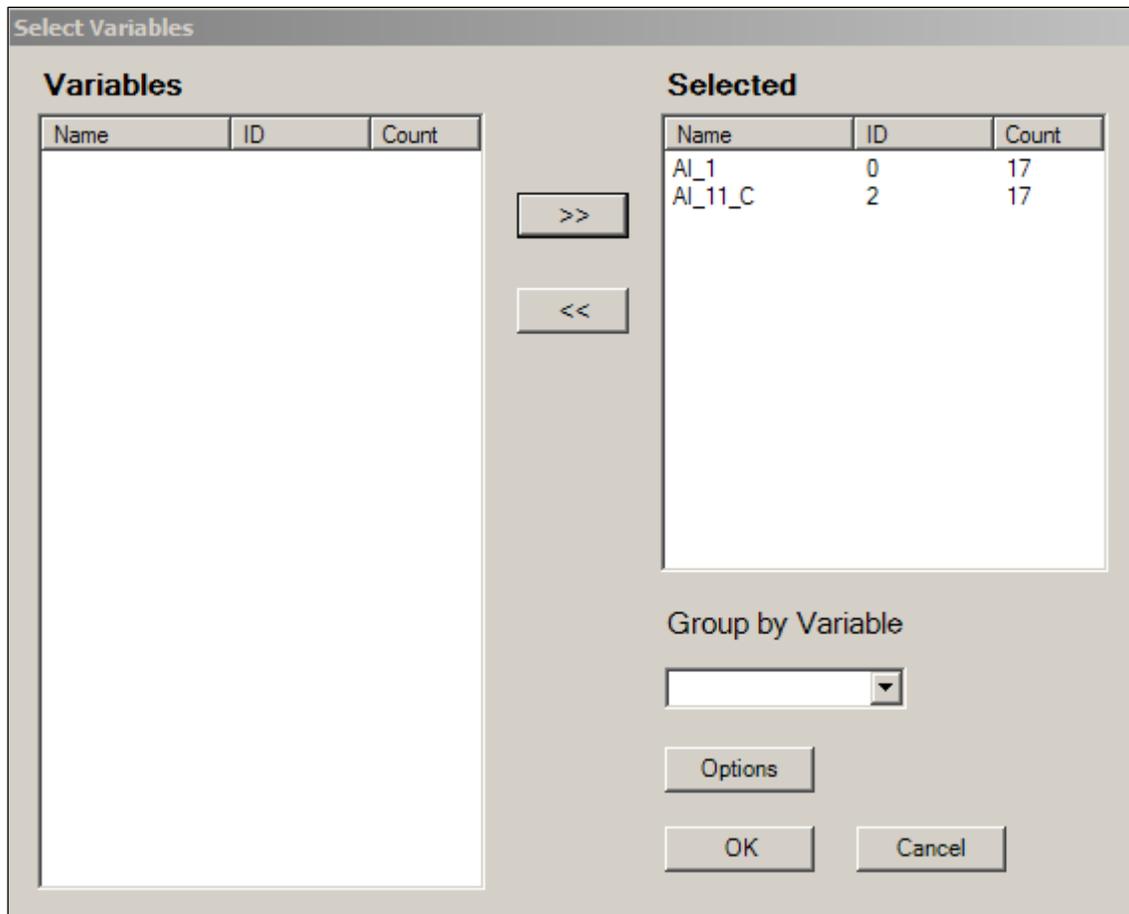
The following shows what it should look like (using the example data):

|    | 0        | 1      |  |
|----|----------|--------|--|
|    | AI_1     | D_AI_1 |  |
| 1  | 0.144    | 1      |  |
| 2  | 0.21     | 0      |  |
| 3  | 0.0227   | 0      |  |
| 4  | 0.0431   | 0      |  |
| 5  | 0.0349   | 1      |  |
| 6  | 0.0419   | 1      |  |
| 7  | 0.00735  | 1      |  |
| 8  | 0.0404   | 1      |  |
| 9  | 0.00366  | 1      |  |
| 10 | 0.0807   | 0      |  |
| 11 | 0.0264   | 1      |  |
| 12 | 0.0317   | 1      |  |
| 13 | 0.0275   | 0      |  |
| 14 | 0.0315   | 1      |  |
| 15 | 0.0192   | 1      |  |
| 16 | 0.00383  | 1      |  |
| 17 | 0.000771 | 0      |  |

4. Calculate Sums of TEQs using Kaplan-Meier. Create a sum for each sample. ProUCL does not do this directly. Instead it creates a mean for each sample, which when multiplied by the number of congeners (17) is the sum. Mean = sum/n; sum= Mean\*n. Here is the process. First, calculate a Background with NDs/Nonparametric using the following approach:

|    | 0        | 1      | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|----|----------|--------|---|---|---|---|---|---|---|---|----|----|----|
|    | AI_1     | D_AI_1 |   |   |   |   |   |   |   |   |    |    |    |
| 1  | 0.144    | 1      |   |   |   |   |   |   |   |   |    |    |    |
| 2  | 0.21     | 0      |   |   |   |   |   |   |   |   |    |    |    |
| 3  | 0.0227   | 0      |   |   |   |   |   |   |   |   |    |    |    |
| 4  | 0.0431   | 0      |   |   |   |   |   |   |   |   |    |    |    |
| 5  | 0.0349   | 1      |   |   |   |   |   |   |   |   |    |    |    |
| 6  | 0.0419   | 1      |   |   |   |   |   |   |   |   |    |    |    |
| 7  | 0.00735  | 1      |   |   |   |   |   |   |   |   |    |    |    |
| 8  | 0.0404   | 1      |   |   |   |   |   |   |   |   |    |    |    |
| 9  | 0.00366  | 1      |   |   |   |   |   |   |   |   |    |    |    |
| 10 | 0.0807   | 0      |   |   |   |   |   |   |   |   |    |    |    |
| 11 | 0.0264   | 1      |   |   |   |   |   |   |   |   |    |    |    |
| 12 | 0.0317   | 1      |   |   |   |   |   |   |   |   |    |    |    |
| 13 | 0.0275   | 0      |   |   |   |   |   |   |   |   |    |    |    |
| 14 | 0.0315   | 1      |   |   |   |   |   |   |   |   |    |    |    |
| 15 | 0.0192   | 1      |   |   |   |   |   |   |   |   |    |    |    |
| 16 | 0.00383  | 1      |   |   |   |   |   |   |   |   |    |    |    |
| 17 | 0.000771 | 0      |   |   |   |   |   |   |   |   |    |    |    |

ProUCL opens a dialog box showing the two samples. (If the D\_AI\_1 wasn't entered correctly, that sample will not appear in the Variables window.) Highlight and "arrow" your samples to the Selected window. Click OK.



The next page shows the output from this example. The critical value is the K-M mean (highlighted). Note that only one sample is shown.

| <b>Nonparametric Background Statistics for Data Sets with Non-Detects</b> |     |          |
|---|-----|----------|
| Confidence Coefficient  | 95% |          |
| Coverage  | 90% |          |
| Different or Future K Values  | 1   |          |
| AI_1  |     |          |
| Total Number of Data  |     | 17       |
| Number of Non-Detect Data   |     | 6        |
| Number of Detected Data   |     | 11       |
| Minimum Detected  |     | 7.71E-04 |
| Maximum Detected  |     | 0.0807   |
| Percent Non-Detects   |     | 35.29%   |
| Minimum Non-detect  |     | 0.00366  |
| Maximum Non-detect  |     | 0.21     |
| Mean of Detected Data   |     | 0.0325   |
| SD of Detected Data   |     | 0.0211   |
| Mean of Log-Transformed Detected Data                                     |     | -3.816   |
| SD of Log-Transformed Detected Data                                       |     | 1.26     |
| Data appear Normal at 5% Significance Level                               |     |          |
| Nonparametric Background Statistics                                       |     |          |
| 95% UTL with 90% Coverage   |     |          |
| Order Statistic   |     | 17       |
| Achieved CC   |     | 1        |
| UTL   |     | 0.21     |
| Warning: Largest Non-detect at Order                                      |     | 17       |
| 95% UPL   |     |          |
| 95% UPL   |     | 0.21     |
| Kaplan-Meier (KM) Method  |     |          |
| Mean  |     | 0.0249   |
| SD  |     | 0.0216   |
| Standard Error of Mean  |     | 0.00594  |
| 95% UTL 90% Coverage  |     | 0.0683   |
| 95% KM Chebyshev UPL  |     | 0.122    |
| 95% KM UPL (t)  |     | 0.0638   |
| 90% KM Percentile (z)   |     | 0.0527   |
| 95% KM Percentile (z)   |     | 0.0605   |
| 99% KM Percentile (z)   |     | 0.0753   |

Copy the K-M mean to the spreadsheet of record (not the one you are going to delete), and multiply it by 17. The value is the sum of TEQs of the sample.

5. Calculate background from a set of sample K-M sums.

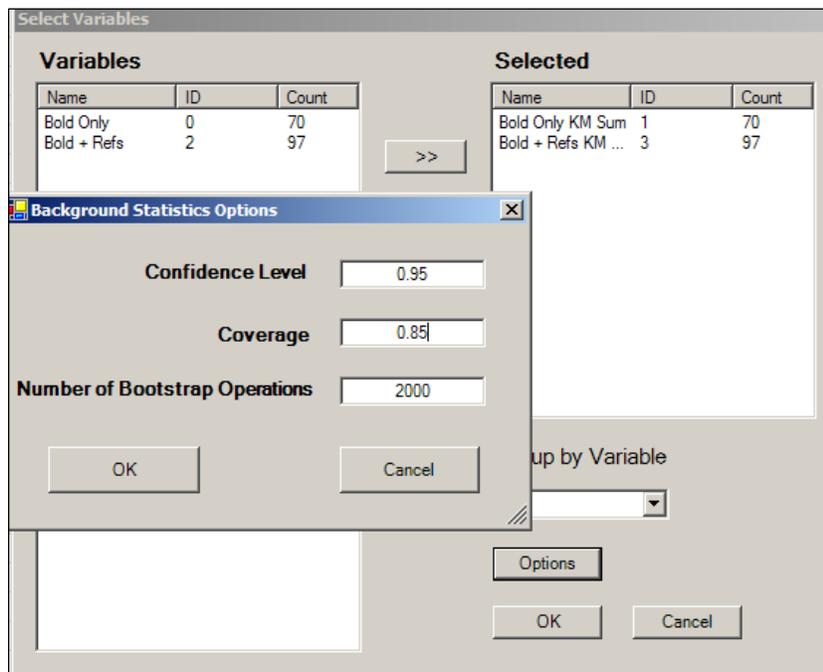
5.1 Copy the K-M sums into a blank spreadsheet, and load it described earlier.

5.2 Use this setting: calculate a Background/Full (no NDs) Background Statistics/Nonparametric. (The “With NDs form” isn’t needed as it was dealt with earlier.)

5.3 As before, select samples by variable name and arrow them into the Selected



box. If it was desired to calculate a UTL for a 95% confidence interval on the 85<sup>th</sup> percentile (coverage), the Option box would look like this.



5.4 The output will look like this. The desired value is highlighted.

| <b>Nonparametric Background Statistics for Full Data Sets</b> |      |              |
|---|------|--------------|
| Full Precision  | OFF  |              |
| Confidence Coefficient  | 95%  |              |
| Coverage  | 85%  |              |
| Number of Bootstrap Operations                                | 2000 |              |
|   |      |              |
| Bold Only KM Sum  |      |              |
|   |      |              |
| Some Non-Parametric Statistics                                |      |              |
| Number of Valid Observations                                  |      | 70           |
| Number of Distinct Observations                               |      | 68           |
| Minimum   |      | 0.044        |
| Maximum   |      | 11.59        |
| Second Largest  |      | 8.347        |
| Mean  |      | 1.268        |
| First Quartile  |      | 0.419        |
| Median  |      | 0.85         |
| Third Quartile  |      | 1.447        |
| SD  |      | 1.765        |
| Variance  |      | 3.115        |
| Coefficient of Variation                                      |      | 1.392        |
| Skewness  |      | 4.103        |
| Mean of Log-Transformed data                                  |      | -0.305       |
| SD of Log-Transformed data                                    |      | 1.07         |
|   |      |              |
| Data appear Lognormal at 5% Significance Level                |      |              |
|   |      |              |
| Non-Parametric Background Statistics                          |      |              |
| 90% Percentile  |      | 2.188        |
| 95% Percentile  |      | 4.29         |
| 99% Percentile  |      | 11.59        |
|   |      |              |
| 95% UTL with 85% Coverage                                     |      |              |
| Order Statistic   |      | 64           |
| Achieved CC   |      | 0.962        |
| <b>UTL</b>  |      | <b>2.193</b> |