Appendix B Grubbs' Test for Outliers

Grubbs' outlier test (NIST/SEMATECH, 2015) at the 95% confidence level was used to identify and remove outlier data pairs recorded during the colocation exercises.

The absolute difference (Δ_i) between each data pair (x_i, y_i, where x_i = Low and y_i = High) was first calculated and the maximum difference ($\Delta_{i=max}$) was identified using the following equations:

$$\Delta_i = |x_i - y_i| \quad \text{(B1)}$$

$$\Delta_{i=max} = |x_i - y_i|_{i=max} \quad \text{(B2)}$$

The Grubbs' test statistic (G) is given by:

$$G = \frac{|\Delta_{i=max} - \bar{\Delta}|}{s} \quad \text{(B3)}$$

Where $\bar{\Delta}$ is the mean of the calculated absolute differences between data pairs and s is the standard deviation, both calculated using the equations below:

$$\bar{\Delta} = \frac{\sum_{i=1}^{i=n_{l-h}} \Delta_i}{n_{l-h}} \quad \text{(B3)}$$

$$s = \frac{\sum_{i=1}^{i=n_{l-h}} \Delta_i - \bar{\Delta}}{n_{l-h} - 1} \quad \text{(B4)}$$

Where n_{l-h} = the number of data pairs recorded by the low and high samplers.

The Grubbs critical value (GCV) is calculated using:

$$GCV = \frac{n_{l-h} - 1}{(n_{l-h})^{\frac{1}{2}}} \left(\frac{\left(t_{\alpha/2n_{l-h}, n_{l-h}-2}\right)^2}{n_{l-h} - 2 + \left(t_{\alpha/2n_{l-h}, n_{l-h}-2}\right)^2} \right)^{\frac{1}{2}} \quad \text{(B5)}$$

Where $t_{\alpha/2n_{l-h}, n_{l-h}-2}$ is the critical value (tCV) of the t-distribution with $(n_{l-h} - 2)$ degrees of freedom and a significance level of $\alpha/2n_{l-h}$, where α is the significance level (0.05 in this case). The TINV() function[1] within Microsoft Excel 2013 was used to calculate tCV.

If G > GCV then the data pair is considered an outlier. The test was applied iteratively until no further outliers were identified.

[1] TINV() Function, Microsoft Excel 2013 [Online]
Available at http://office.microsoft.com/en-gb/excel-help/tinv-HP005209317.aspx [accessed on 30/12/2014]