# Appendix A Calculations Used for Regression Analysis

## A1.1 Orthogonal Regression

The response of all samplers used within this study is assumed to be linear and therefore follow the equation of a straight line:

$$y_i = a + bx_i \qquad \text{(A1)}$$

Where:

$x_i$ = concentration of Low sampler measured at point $i$, where $i$ > 0.

$y_i$ = concentration of High sampler measure at point $i$, where $i$ > 0.

$a$ = intercept.

$b$ = slope.

The between-sampler uncertainty ($u_{bs}$) is calculated using the equation below:

$$u_{bs} = \left(\frac{\sum_{i=1}^{n_{bs}}(y_{i,1} - y_{i,2})^2}{2n_{bs}}\right)^{\frac{1}{2}} \qquad \text{(A2)}$$

Where:

$y_{i,1}$ and $y_{i,2}$ are parallel measurements for a single 1-minute, 5 minute, 15 minute or 1 hr period i.

$n_{bs}$ = the number of between sampler measurements.

The Mean Absolute Error is calculated using the following equation:

$$MAE = \sum_{i=1}^{i=n}|\hat{y}_i - y_i| \qquad \text{(A3)}$$

Where:

$\hat{y}_i$ is the predicted value from the regression model.

$y_i$ is the measured value.

$n$ = the number of paired measurements.

The following equations were used to calculate the slope and intercept for the best-fit orthogonal regression line:

$$b = \frac{S_{yy} - S_{xx} + \left[\left(S_{yy} - S_{xx}\right)^2 + 4\left(S_{xy}\right)^2\right]^{\frac{1}{2}}}{2S_{xy}} \tag{A4}$$

$$a = \bar{y} - b \cdot \bar{x} \tag{A5}$$

Where:

$$S_{xx} = \sum_{i=1}^{i=n} (x_i - \bar{x})^2 \tag{A6}$$

$$S_{yy} = \sum_{i=1}^{i=n} (y_i - \bar{y})^2 \tag{A7}$$

$$S_{xy} = \sum_{i=1}^{i=n} (x_i - \bar{x}) \cdot (y_i - \bar{y}) \tag{A8}$$

Where:

$n_{l-h}$ = number of paired measurements recorded at 0.80 m, Low *(x).* and 1.68 m ,High *(y)* samplers.

$$\bar{x} = \frac{\sum_{i=1}^{i=n} x_i}{n_{l-h}} \tag{A9}$$

$$\bar{y} = \frac{\sum_{i=1}^{i=n} y_i}{n_{l-h}} \tag{A10}$$

The coefficient of determination, $r^2$, is calculated using the following equation:

$$r^2 = \left[\frac{\left[n \sum_{i=1}^{i=n}(xy)\right] - \left[\sum_{i=1}^{i=n} x \cdot \sum_{i=1}^{i=n} y\right]}{\left[\left(n \sum_{i=1}^{i=n}(x^2)\right) - \left(\sum_{i=1}^{i=n} x\right)^2\right] \cdot \left[\left(n \sum_{i=1}^{i=n}(y^2)\right) - \left(\sum_{i=1}^{i=n} y\right)^2\right]}\right]^2 \tag{A11}$$

## A1.2 Uncertainty in Slope and Intercept – Uncorrected Datasets

The uncertainty in the slope $(u_b)$ is calculated using:

$$u_b = \left[\frac{S_{yy} - ({S_{xy}}^2/S_{xx}}{(n_{l-h} - 2)S_{xx}}\right]^{\frac{1}{2}} \qquad \text{(A12)}$$

The uncertainty in the intercept $(u_a)$ is calculated using:

$$u_a = \left[{u_b}^2 \; \frac{\sum_{i=1}^{i=n} {x_i}^2}{n_{l-h}}\right]^{\frac{1}{2}} \qquad \text{(A13)}$$

## A1.3 Significance of Intercept and Slope

In order to determine if data need to be corrected for intercept (a) or slope (b) using the colocation results, the following criteria have been used to define if these are significant. The calculate intercepts and slopes are ***not*** deemed significant if:

$$|a| \leq 2u_a \qquad \text{(A14)}$$

$$|b - 1| \leq 2u_b \qquad \text{(A15)}$$

In this case no correction for intercept or slope is made.

For the purposes of this study, the Low sampler has been used as the reference sampler. Corrections for intercept and slope are made to data recorded by the High sampler.

## A1.4 Uncertainty in Measurements

The residual sum of squares (RSS) from the orthogonal regression is calculated using:

$$RSS = \sum_{i=1}^{i=n} (y_i - a - bx_i)^2 \qquad \text{(A16)}$$

The uncertainty in each y-value $(\sigma)$ is calculated using:

$$\sigma = \left[\frac{1}{n-2}\sum_{i=1}^{i=n}(y_i - a - bx_i)^2\right]^{\frac{1}{2}} \quad \text{(A17)}$$

The uncertainty in the results in the results from the High sampler $(u_{l-h}(y_i))$ using the Low sampler as the reference is calculated using:

$$[u_{l-h}(y_i)]^2 = \sigma^2 + [a + ((b-1)z_i]^2 \quad \text{(A18)}$$

Where $z_i$ is the reference concentration at which the uncertainty is calculated.

The combined relative uncertainty of the results as measured by the High sampler $(w_h(y_i))$ relative to the results from the Low sampler is calculated using:

$$w_h(y_i) = 100 \cdot \left(\frac{u_{l-h}(y_i)}{z_i}\right) \quad \text{(A19)}$$

The expanded uncertainty is calculated using a coverage factor of $k = 2$ reflecting a 95% confidence interval with a normal distribution associated with the large number of measurements. Therefore:

$$W(y_i) = k \cdot w_h(y_i) = 2w_h(y_i) \quad \text{(A20)}$$

## A1.5 Uncertainty in Intercept Corrected Datasets

If it is found from the colocation exercises that the data from a High sampler requires a correction for intercept then the following equation is used to calculate the corrected data $(y_{i,corr})$ :

$$y_{i,corr} = y_i - a \quad \text{(A21)}$$

Orthogonal regression then carried out using the corrected data with the equation of the straight line:

$$y_{i,corr} = c + dx_i \quad \text{(A22)}$$

Where:

$c =$ intercept calculated using the corrected data.

$d =$ slope calculated using the corrected data.

In order to take account of the uncertainty in the intercept introduced during the colocation exercise regression ($u_a$) this is added to the uncertainty calculation:

$$\left[u_{l-h}(y_{i,corr})\right]^2 = \sigma^2 + \left[c + ((d-1)z_i\right]^2 + u_a{}^2 \quad \text{(A23)}$$

## A1.6 Uncertainty in Slope Corrected Datasets

If it is found from the colocation exercises that the data from a High sampler requires a correction for slope then the following equation is used to calculate the corrected data $(y_{i,corr})$ :

$$y_{i,corr} = \frac{y_i}{b} \quad \text{(A24)}$$

In order to take account of the uncertainty in the intercept introduced during the colocation exercise regression ($u_b$) this is added to the uncertainty calculation:

$$\left[u_{l-h}(y_{i,corr})\right]^2 = \sigma^2 + \left[c + ((d-1)z_i\right]^2 + z_i{}^2 u_b{}^2 \quad \text{(A25)}$$

## A1.7 Uncertainty in Intercept and Slope Corrected Datasets

If it is found from the colocation exercises that the data from a High sampler requires a correction for slope and intercept then the following equation is used to calculate the corrected data $(y_{i,corr})$ :

$$y_{i,corr} = \frac{y_i - a}{b} \quad \text{(A26)}$$

In order to take account of the uncertainty in the intercept ($u_a$) and slope ($u_b$) introduced during the colocation exercise regression this is added to the uncertainty calculation:

$$\left[u_{l-h}(y_{i,corr})\right]^2 = \sigma^2 + \left[c + ((d-1)z_i\right]^2 + z_i{}^2 u_b{}^2 + u_a{}^2 \quad \text{(A27)}$$