

Slope and Bias (S and B) factors

The difficulty posed to any instrument manufacturer is that no two instruments can be made the same. Subtle differences exist which may skew or offset the spectral data of the newly manufactured instrument, the slave instrument, with respect to the instrument the calibration was developed on, the master instrument.

Such differences can sometimes make the transfer of calibrations ineffective. Fortunately, mathematical manipulations exist which are used to correct spectral data of slave instruments to match them to the master. The process involves firstly correcting for skew with a slope factor followed by an offset correction to adjust for the skew. This produces a set of factors called slope and bias (S and B) factors, which correct individual spectral points.

Whole grain analysers are subject to samples which do not display consistent packing characteristics and it was found that the use of reference materials such as neutral density filters do not provide satisfactory correction factors, as they do not display the variability shown in whole grains.

The method employed to correct for instrumental differences is described as follows:

- 9 reference wheat samples were chosen based on sample absorption characteristics ranging from 2 to 4 abs units.
- The samples were scanned on the master instrument (5 replicates per analysis, rotating the sample each time to account for sample variability).
- The same 9 samples were scanned on the slave instrument in the same way they were scanned on the master instrument.

In order to compare and correct the data, a Cat-file (Calibration Transfer) file is constructed using Microsoft Excel with the following format.

The spectral data for the master instrument (45 rows x 38 spectral data columns) is placed at the top of the spreadsheet. Underneath this, the data for the slave instrument is placed for the same samples, in the same order as the master data file.

Two methods may be employed to determine the best correlation between the data

Method 1: Bias and Slope

The first step of this method is to use the spreadsheet function SLOPE, to calculate a correction for skew for each spectral point (ie. for each of the 38-elements of the detector, referred to as pixels). The formula used is:

$$Slope_i = \mathbf{SLOPE} (Sp_{i,1} : Sp_{i,45}, Sp_{i,x} : Sp_{i,x+45})$$

Where:

$Slope_i$ = the skew correction for pixel i (i=1-38)

Sp_i = the spectral point for pixel i (the first term of the equation relates to the master instrument)

$Sp_{i,x}$ = the spectral point for pixel i starting at row x where the slave spectral data begins.

For an effective matching, the slope value should be within the limits 0.9 to 1.1 for each pixel.

When any correction for skew has been performed, an adjustment for bias usually is required. The formula for this correction is:

$$Bias_i = \mathbf{INTERCEPT} (Sp_{i,1} : Sp_{i,45}, Sp_{i,x} : Sp_{i,x+45})$$

Where the definitions of this equation are similar to the slope equation described above.

Using the values of bias and slope for each pixel, a corrected set of data Sp_{corr} , is generated via the following function:

$$Sp_{(corr)i,x} = Sp_{i,x} \times Slope_i + Bias_i$$

Method 2: Line of Best Fit

A second and sometimes more accurate correction factor is the line of best fit calculated for the two sets of spectral data using a least-squares regression fitting the data through the origin. The purpose of the least-squares function is to calculate the line of best fit by minimising the error term in trying to fit the data.

The equation for the line is

$$y = m_1x_1 + m_2x_2 + \dots + m_{38}x_{38} + b$$

For this calculation $b=0$ and hence the general form of the straight line is reduced to

$$y=mx$$

The formula used to calculate the line of best fit is:

$$\textit{Line.of.best.fit} = \mathbf{LINEST}(Sp_{i,1} : Sp_{i,45}, Sp_{i,x} : Sp_{i,x+45}, \textit{const}, \textit{stats})$$

Where:

Sp_i = the spectral data for the master and slave instruments

const = logic value. When const is set to **FALSE**, the intercept is set to 0

stats = logic value. When stats is set to **TRUE**, the regression statistics for the operation are displayed.

The formula should then have the following structure:

$$\textit{Line.of.best.fit}_i = \mathbf{LINEST}(Sp_{i,1} : Sp_{i,45}, Sp_{i,x} : Sp_{i,x+45}, \mathbf{FALSE}, \mathbf{TRUE})$$

To obtain the set of corrected spectral data, the following formula is used:

$$Sp_{(corr)i} = Sp_{1,x} \times \textit{Line.of.best.fit}_i$$

Correlation

Microsoft Excel also has a function, which is used to determine if the two instruments are closely related to each other. This is known as correlation and it is defined as:

$$r_{x,y} = \frac{Cov(X,Y)}{\sigma_x \cdot \sigma_y}$$

Where

$$-1 \leq r_{x,y} \leq 1$$

And Cov (X,Y) is the covariance between the data sets defined as:

$$Cov(X, Y) = \frac{1}{n} \sum_{i=1}^n (x_i - \mu_x)(y_i - \mu_y)$$

Where:

x_i = the master instrument spectral data

y_i = the slave instrument spectral data

The closer the $r_{x,y}$ value is to 1, the higher the correlation.

Specifications

- Slope values must be in the range 0.9 to 1.1
- Bias values must be in the range ± 0.2
- Correlation must be >0.97

If all these conditions are met, a good transfer of calibration should result.

Calibration Transfer

The Cat-file now contains a set of corrected spectral data Sp_{corr} , which when multiplied by the calibration equation, should produce accurate prediction results. The regression coefficients for the calibration are added to the Cat-file and each corrected spectral point is multiplied by the factor associated with that point using the following formula.

$$\% \text{ Protein} = \sum_{i=1}^{38} (Sp_{(corr)i} \times b_i) + b_0$$

Where the b_i are the calibration regression coefficients for each spectral point and b_0 is the usual factor associated with partial least squares (PLS) regression.

The predicted data generated here are compared to the predicted values from the master instrument, and to the laboratory data. At this point, it is decided whether to use method 1 or method 2. The criteria used to determine the best method is to see which one gives a better correlation (R^2) value.

In general, correlations >0.95 are acceptable for successful calibration transfer. When satisfied that the instruments are matched, the correction factors must be converted into a text file, which the instrument can read. The S and B factors are copied from their present position and pasted as their transpose as values only at the bottom of the Cat-file. The column of B factors is cut and pasted at the bottom of the S factors to form a single column of data containing 76 rows. If method 1 was used, the S and B factors should contain 76 rows of discrete correction factors. If method 2 was used, the data was fit to pass through the origin, hence the last 38 rows, ie. the B factors should all be zero. The following procedure outlines this process.

The prn file

The S and B factors are copied into a new spreadsheet formatted into 3 columns. The first column consists of 76 rows. The first 38 rows contain S1-S38 and the second 38 contain B1-B38. Column 2 contains 76 rows consisting of = signs. The third column contains the S and B factors calculated by either method 1 or method 2. The column containing the correction factors is left justified and the document is saved as the instrument number with a prn extension (formatted text).

This file is now opened using the Wordpad application program. The file is formatted in such a way that any multiple spaces between the S_i and the = sign are converted to a single space. This is readily achieved by highlighting the spaces between them and pasting this into the replace function (found in the edit menu) in the box labelled **Find**. In the box labelled **Replace With**, type in a single space and click on the **Replace all** button. A similar process is performed with the spaces between the = sign and the correction factors. Once the data has been formatted it is now ready to be copied and pasted into the instruments ini file.

The data, when pasted into the ini file is then saved as **wga(instrument number).ini** and using the WGA-Test application, the file is downloaded into the instrument under the directory **c:\wga\wga.ini**. After downloading, the instrument is shut down and then restarted to activate the new ini file. The final step is to scan a certain number of samples with accurate laboratory data and fine tune the instrument accordingly to these samples.