C27 Fitting Calibration Data

Willem Schreüder, PhD, Principia Mathematica, Inc., 405 Urban Street, Suite 305, Lakewood, CO; and Leslie Eng, PhD*, Trillium, Inc., 3623 Ady Road, Street, MD*

The goal of this presentation is to present to the forensic community the concepts involved in selecting appropriate curves to accurately describe instrument calibration data. Attendees will have an understanding of mathematical foundations of linear regression using functions more complex than a linear polynomial and how it can be applied to calibration curves of analytical instruments.

An improper or poor selection of a mathematical curve to characterize the responses of an analytical instrument to known concentrations of a material can introduce errors in the determination of concentrations of unknown field samples which, in turn, could lead to misinterpretation of the character of the material being investigated.

When generating calibration curves for analytical instruments, samples with various known concentrations of a material are introduced into the instrument, and the instrumental responses are measured. The dependency of the responses to the concentrations is and is used as the instrument calibration curve. Responses from samples with unknown concentrations of the material are then measured and compared to the calibration curve to determine the concentration of the material in the sample.

In constructing a calibration curve, a normal linear regression is typically used to fit a simple curve to the calibration data. In most cases the linear polynomial, $y = a + bx$, is used. This asserts a linear relationship between the responses of the instrument and the measured quantities. Many users examine only the correlation coefficient, R, associated with the linear regression to determine if a linear relationship exists. A correlation coefficient of 0.99 or greater is often taken as an indication of a linear relationship between the instrument response and the concentration in the sample.

In reality this is not always the case. Often the response is non-linear, especially when calibration curves extend over large concentration ranges. Depending on the instrument, non-linearity can exist even over only two orders, or less, of magnitude. Forcing a straight line through calibration data that are inherently non-linear introduces errors into certain portions of the curve. The higher responses associated with the higher concentrations tend to numerically bias the calculated curve at the expense of the lower responses and concentrations. Yet, large errors can often exist in calibration curves having correlation coefficients of 0.99 or greater.

To produce a “better” fit of the data, 11x weighting and similar techniques are often used. While this results in a correlation coefficient R closer to one and a reduced error for small values of x, the error for large values of x is increased by the use of the 11x weighting technique. Any such techniques are inherently flawed since it reduces the error for some values of x at the expense of others, typically at one or the other end of the calibration range. Moreover, the error tends to vary over the calibration range making the error term also dependent on the concentration being measured.

The fundamental problem is the assertion that $y=a + bx$. If, instead, it is acknowledged that the response is nonlinear and $y=f(x)$, better instrument calibration can be obtained and more accurate determinations of unknown concentrations can be obtained. This paper shows that by choosing appropriate basis functions, linear regression can be performed using non-linear basis functions such that the approximation error over the entire calibration range is reduced. The calibration curve may be as simple as the quadratic polynomial $y=a + bx + cx^2$, or more complex such as $y=a+ bx + ce^{Bx}$. As long as the basis functions are linearly independent and functions of only x, a linear combination of these basis functions can be fitted using linear regression techniques. By choosing appropriate basis functions, optimal regression coefficients can be uniquely determined that best represent the non-linearity inherent in the detector.

C**alibration, Regression, Error**

 ** Presenting Author*