Comment on:
"Utilizing Artificial Neural Networks in MATLAB to Achieve Parts-Per-Billion Mass Measurement Accuracy with a Fourier Transform Ion Cyclotron Resonance Mass Spectrometer" by D. Keith Williams Jr., Alexander L. Kovach, David C. Muddiman, and Kenneth W. Hanck. *J. Am. Soc. Mass Spectrom*. 20, 1303–1310 (2009)

When calibrating a FT-ICR MS system [1], the authors report that "results …demonstrated that a fit using artificial neural networks (ANN) provided a better fit of data than the multiple linear regression (MLR) method previously published" [2]. However, the article's Figure 3a (that "illustrates the residuals generated by the fitting of the calibration data by MLR") shows the largest two oligomers to be outliers. The residuals shown in Figure 3b from the ANN show the outlier extreme displacements are gone from the largest two oligomers, but now displacements along the line of smaller oligomers reflect an underlying curvature in the ANN fit.

Five statistical tests were run that showed non-normality when ML was fit but not when ANN was run, but the tests were likely detecting kurtosis (long tailed distributions) due to the outliers. The kurtosis was diminished after fitting the compliant ANN, which masked the outliers. However, the above-mentioned displacements testify that passing normality tests does not mean that there was no over-fitting.

The ANN reportedly used 9.7 or 12.3 "parameters" (compared with the 3 or 4 of MLR) and since 55 observations were fit, this number may seem small. In Figure 2a, the 55 observations, when plotted, effectively fall at only 11 positions. Thus, the critical fitting was done to the 11 oligomer averages of predictions and now, with 12.3 larger than 11, the over-fitting by the ANN becomes apparent.

The Figure 2a fitted line based on 12 parameters (the "appropriate number of neurons") compared with the Figure 2b line based on "too many neurons" illustrates clearly that more parameters mean more curvature. What is missing is the MLR line to appear in Figure 2a. It is apparent from Figure 3a that this MLR line would not pass anywhere near the clusters of observations for the two largest oligomers. Thus, there is curvature in the ANN line that allows it to pass close to all data clusters. The ANN is a weighted average of S-curves and although very useful in detecting pattern outlines, is not suited for a calibration curve.

The appropriate way to deal with outliers in reference materials would be to correct them and use MLR or to omit them and use MLR.

Charles Proctor
North Carolina State University
Raleigh, NC, USA

References

1. Williams, D.K., Kovach, A.L., Muddiman, D.C., Hanck, K.W.: Utilizing artificial neural networks in MATLAB to achieve parts-per-billion mass measurement accuracy with a Fourier transform ion cyclotron resonance mass spectrometer. J. Am. Soc. Mass Spectrom. **20**, 1303–1310 (2009)
2. Williams, D.K., Muddiman, D.C.: Parts-per-billion mass measurement accuracy achieved through the combination of multiple linear regression and automatic gain control in Fourier transform ion cyclotron resonance mass spectrometer. Anal. Chem. **79**, 5058–5063 (2007)

Correspondence to: Charles Proctor; *e-mail:* chproctor@msn.com

© 2013 American Society for Mass Spectrometry
doi:10.1007/s13361-013-0805-8
(J Am Soc Mass Spectrom 2014)