

## SAVITZKY-GOLAY DERIVATIVES : A SYSTEMATIC APPROACH TO REMOVING VARIABILITY BEFORE APPLYING CHEMOMETRICS

David W. Hopkins

*NIR Consultant, Battle Creek, MI 49015 USA*

*dwh @ voyager.net*

Removal of variability in spectra data before the application of chemometric modeling will generally result in simpler (and presumably more robust) models. Particularly for sparsely sampled data, such as typically encountered in diode array instruments, the use of Savitzky-Golay (S-G) derivatives offers an effective method to remove effects of shifting baselines and sloping or curving apparent baselines often observed with scattering samples. The application of these convolution functions is equivalent to fitting a selected polynomial to a number of points in the spectrum, usually 5 to 25 points. The value of the polynomial evaluated at its mid-point, or its derivative, is taken as the (smoothed) spectrum or its derivative at the mid-point of the wavelength window. The process is continued for successive windows along the spectrum. The original paper, published in 1964 [1] presented these convolution functions as integers to be used as multipliers for the spectral values at equal intervals in the window, with a normalization integer to divide the sum of the products, to determine the result for each point. Steinier et al. [2] published corrections to errors in the original presentation [1], and a vector formulation for obtaining the coefficients.

The actual selection of the degree of polynomial and number of points in the window determines whether closely situated bands and shoulders are resolved in the derivatives. Furthermore, the actual noise reduction in the derivatives may be estimated from the square root of the sums of the coefficients, divided by the NORM value. A simple technique to evaluate the actual convolution factors employed in the calculation by the software will be presented. It has been found that some software packages do not properly account for the sampling interval of the spectral data (Equation VII in [1]). While this is not a problem in the construction and implementation of chemometric models, it may be noticed in comparing models at differing spectral resolutions. Also, the effects on parameters of PLS models of choosing various polynomials and numbers of points in the window will be presented.

1. Savitzky, A. and M.J.E. Golay. 1964. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. *Anal. Chem.* 36 (8) 1627-1639.
2. Steinier, J. Y. Termonia and J. Deltour. 1972. Comments on Smoothing and Differentiation of Data by Simplified Least Square Procedure. *Anal. Chem.* 44 (11) 1906-1909.