Weighted Partial Least Squares Method to Improve Calibration Precision for Spectroscopic Noise-Limited Data

David M. Haaland and Howland D. T. Jones

Sandia National Laboratories, Albuquerque, NM 87185-0342

Multivariate calibration methods have been applied extensively to the quantitative analysis of Fourier transform infrared (FT-IR) spectral data. Partial least squares (PLS) methods have become the most widely used multivariate method for quantitative spectroscopic analyses. Most often these methods are limited by model error or the accuracy or precision of the reference methods. However, in some cases, the precision of the quantitative analysis is limited by the noise in the spectroscopic signal. In these situations, the precision of the PLS calibrations and predictions can be improved by the incorporation of weighting in the PLS algorithm. If the spectral noise of the system is known (e.g., in the case of detector-noise-limited cases), then appropriate weighting can be incorporated into the multivariate spectral calibrations and predictions. A weighted PLS (WPLS) algorithm was developed to improve the precision of the analyses in the case of spectral-noise-limited data. This new PLS algorithm was then tested with real and simulated data, and the results compared with the unweighted PLS algorithm. Using near-infrared (NIR) spectra obtained from a series of dilute aqueous solutions, the simulated data produced calibrations that demonstrate improved calibration precision when the WPLS algorithm was applied. The best WPLS method improved prediction precision for the analysis of one of the minor components by a factor of nearly 9 relative to the unweighted PLS algorithm.

INTRODUCTION

When multivariate spectral calibrations are limited by spectral noise and when the noise is nonuniform across the spectral region used in the analysis, the partial least squares (PLS) algorithm can be suboptimal. This suboptimal performance arises from the fact that PLS assumes that the spectral errors are normally distributed with zero mean and constant variance throughout the spectral region. For infrared spectra, the high-frequency spectral noise of transmittance spectra is often limited by detector noise that exhibits constant variance that is independent of the spectral intensity. However, once these transmittance spectra are converted to absorbance, the spectral noise variance is no longer constant. This nonuniform noise variance is accentuated in those cases where there are large variations in absorbance. In this case, the precision of the concentration prediction estimates should be improved using a WPLS algorithm relative to that of the traditional unweighted PLS algorithm. In this paper, we demonstrate one version of a weighted PLS (WPLS) algorithm and compare the prediction precisions for weighted and unweighted PLS analyses applied to both real and simulated data obtained from a set of dilute aqueous solutions.

EXPERIMENTAL PROCEDURES

The spectral data in this paper have been described previously (1). The data are from a set of spectra of urea, creatinine, and NaCl at 0 to 3000 mg/dL in water. 31 calibration samples were prepared in sealed cuvettes with each component varied at 13 concentration levels in a Latin hypercube experimental calibration design (1). The near-infrared spectra were obtained in 10-mm path-length cuvettes maintained at 23° C in a Nicolet 800 FT-IR spectrometer operated at 16 cm^{-1} resolution. The spectrometer was equipped with a liquid-nitrogen-cooled InSb detector, a quartz beam splitter, and a 75 W tungsten-halogen source. The run order of the samples was randomized. The spectra are exhibited in Fig. 1 for the analysis region (7500 to 11000 cm^{-1}). Figure 2 shows the same calibration spectra after mean centering.
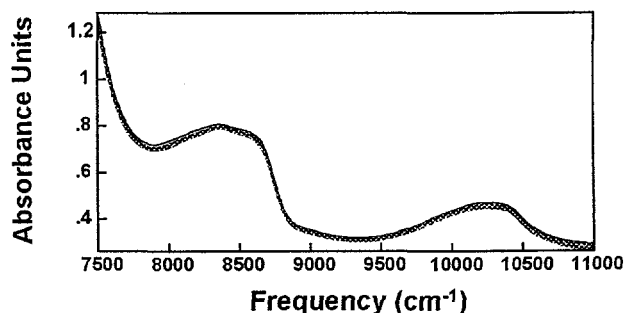

FIGURE 1. Spectra of 31 calibration samples.
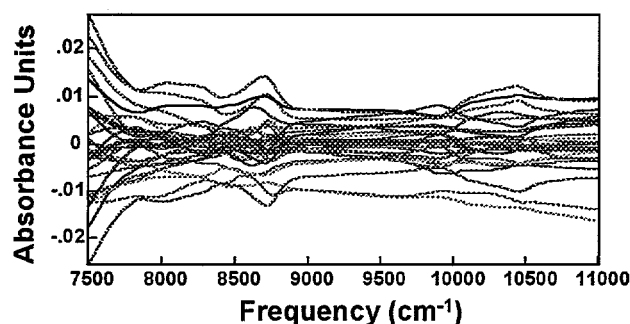
DISCLAIMER

DISCLAIMER

FIGURE 2. Mean-centered spectra of 31 calibration samples.

To further evaluate the advantages of the WPLS method, simulated spectra were generated to model the same system in a 25-mm path-length cell. These simulation spectra assured that the multivariate data analysis would be dominated by spectral noise. The simulation was accomplished by estimating the pure-component spectra from the 31 calibration spectra by CLS multivariate methods. Simulated absorbance spectra were generated by multiplying the pure-component spectra by 2.5 and then adding these path-length-corrected pure-component spectra according to their concentrations in each calibration sample. The baseline offset was also increased to further accentuate spectral noise differences across the spectrum. These simulated absorbance spectra were then converted to single-beam spectra and random noise (0 mean, normally distributed, and constant variance) was added. The resulting single-beam spectra were then converted back to absorbance. In this manner, the spectral noise mimics that expected for 25-mm path-length aqueous samples obtained on an FT-IR spectrometer with linear response. Figure 3 shows the mean-centered spectra of the 31 simulated samples. Note the greater noise in Fig. 3 relative to that in Fig. 2.
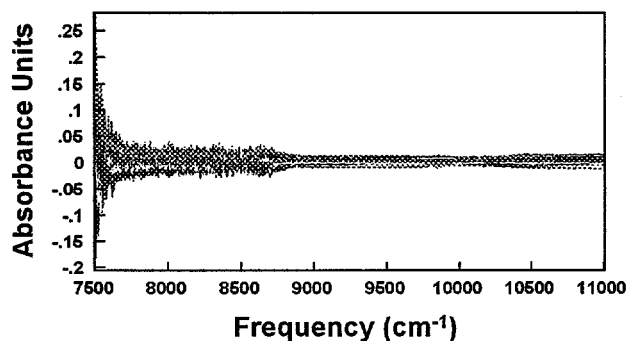

FIGURE 3. Mean-centered spectra of 31 simulated calibration samples.

All of the weighted and unweighted PLS analyses were performed using in-house software generated in Matlab 5.0. The cross-validated standard errors of prediction (CVSEP) for the two types of PLS calibrations were used

for comparing the results.

THEORY

When the spectral noise variance is not constant and the multivariate calibration prediction precision is limited by this spectral noise variance, then a weighted multivariate calibration should be applied. Wentzell et al. (2) have recently presented a weighted approach to principal component regression. However, to date, a weighted approach to PLS has not been published. In this paper, we present our methods for performing a weighted PLS analysis.

As reported by Haaland and Thomas (3), the PLS algorithm can be broken down into a series of CLS calibration and prediction steps and an inverse least squares (ILS) step. Haaland and coworkers (4,5) have presented the methods for performing CLS analyses in both the calibration and prediction phases. By analogy, the WPLS algorithm is readily constructed by adding weighting to the CLS portions of the PLS algorithm in the calculation of each PLS loading vector. Since the errors are presumed to be dominant in spectral rather in concentration space, weighting is not applied to the ILS steps of the PLS algorithm. Two basic unweighted PLS algorithms are currently available, one that generates orthogonal scores and one that generates orthogonal loading vectors (6). Without weighting, these two algorithms are equivalent. However, in the WPLS algorithms, it can be shown that the two PLS methods no longer yield identical results. We report in this paper results for the WPLS method that uses as its basis the unweighted PLS method with orthogonal loading vectors. The resulting WPLS algorithm has neither the scores nor the loading vectors orthogonal. However, orthogonal loading vectors can be generated during the WPLS algorithm simply by adding a Gram-Schmidt orthogonalization procedure after the second and subsequent loading vectors have been determined.

In Haaland and Easterling (4), the appropriate weighting function for detector-noise-limited FT-IR absorbance spectra is given. The proposed weighting factor was T_i^2 where T_i is the transmittance of the sample at frequency i. This weighting factor was determined by expanding the absorbance signal as a Taylor series about the transmittance and retaining only the first two terms. The variance of the transmittance noise was then found to be proportional to T_i^{-2}, and the appropriate weighting factor is inversely proportional to this variance. This weighting factor is the appropriate for the case presented in Ref. 3 where the analysis was performed over a small spectral window and the single-beam intensity of the spectrometer response was essentially constant with frequency.

A more general weighting factor can be determined for those cases where the spectrometer response is highly variable over the spectral region of interest. In this case, the Taylor series expansion of the absorbance is made about the single-beam intensity. Retaining the first two terms of this expansion shows that the appropriate weighting function is the square of the single-beam intensity spectrum (I_i^2). This weighting takes into account variations in spectral noise with sensitivity of the spectrometer as a function of frequency. In this paper, we used the square of the single-beam spectrum for weighting each step of the WPLS algorithm.

The WPLS algorithm is considerably slower than the standard PLS method. For the data presented in this paper, the WPLS method was a factor of seven slower than the traditional PLS method. Of the weighted steps in the PLS algorithm, by far the slowest is the weighted CLS calibration step of the WPLS method. However, in those cases where the spectral differences between samples are small relative to the range of absorbance variations within a sample, then this slow weighted CLS calibration step has relatively little influence on the predictions. Thus, in the case described above, the weighted portion of the CLS calibration step can be skipped. With this modification of the algorithm, the analyses time of the data presented in this paper differed by less than a factor of two between the WPLS and PLS algorithms.

RESULTS AND DISCUSSION

The cross-validated SEP values for the weighted and unweighted PLS analyses were not significantly different for the three components analyzed (CVSEP for the three components ranged from 13 to 24 mg/dL). The lack of significant difference between the two methods suggests that the spectral analysis of the original absorbance data is not dominated by spectral noise. Therefore, simulated spectra were generated as described above to assure that the analysis would be limited by spectral noise. The urea prediction results for the PLS and WPLS analyses are presented in Fig. 5 and Fig. 6, respectively. The results for the analyses of these simulated data are given in Table I for all three minor components in the calibration

samples. The enormous benefit of WPLS over the normal PLS is observed in Figs. 4 and 5 and in Table I where the system is clearly limited by spectral noise and where the weighting function is highly variable over the spectral analysis region. Improvements in precision vary from factors of 3.8 to 8.9. In addition, the model sizes are approximately half the size for the WPLS models where the spectral noise is normalized. These results dramatically demonstrate the potential value of applying WPLS to systems where the errors are dominated by spectral noise.
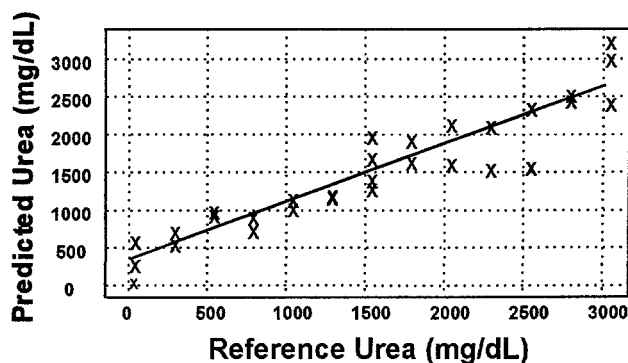

FIGURE 4. Prediction results for urea using the traditional PLS algorithm. The line is the linear least squares fit to the data.
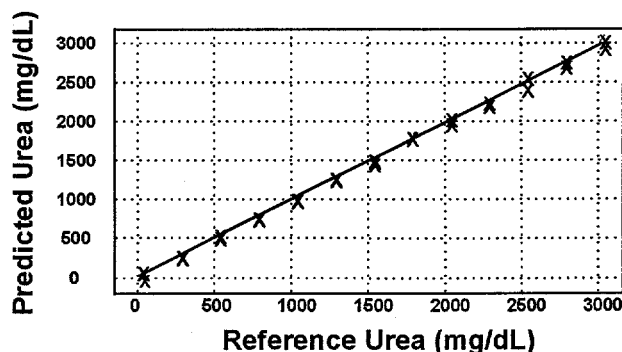

FIGURE 5. Prediction results for urea using the weighted PLS algorithm. The line is the linear least squares fit to the data.

Table I: Results of Applying PLS and WPLS to Simulated Data

Component	PLS			WPLS		
	CVSEP (mg/dL)	Factors	R^2	CVSEP (mg/dL)	Factors	R^2
Urea	350	7	0.8800	39.4	3	0.9986
Creatinine	216	8	0.9480	53.8	4	0.9967
NaCl	166	7	0.9689	43.5	4	0.9982

ACKNOWLEDGMENTS

REFERENCES

1. Haaland, D. M. and Jones, H. D. T., 9th International Conference on Fourier Transform Spectroscopy, Bertie, J. E. and Wisser, H., Editors, Proc. SPIE **2089**, 448-449 (1993).
2. Wentzell, P. D., Andrews, D. T., Hamilton, D. C., Faber, K., and Kowalski, B. R., J. Chemometrics **11**, 339-366 (1997).
3. Haaland, D. M. and Thomas, E. V., Analytical Chemistry **60**, 1193-1202 (1988).
4. Haaland, D. M. and Easterling, R. G., Applied Spectroscopy **36**, 665-673, (1982).
5. Haaland, D. M., Easterling, R. G., and Vopicka, D. A., Applied Spectroscopy **39**, 73-84, (1985).
6. H. Martens and T. Naes, Multivariate Calibration, New York: John Wiley & Sons, 1989, ch. 3, pp. 121-125.