

Calibration Standardization Algorithm for Partial Least-Squares Regression: Application to the Determination of Physiological Levels of Glucose by Near-Infrared Spectroscopy

Lin Zhang[†] and Gary W. Small*

Center for Intelligent Chemical Instrumentation, Department of Chemistry and Biochemistry, Clippinger Laboratories, Ohio University, Athens, Ohio 45701

Mark A. Arnold

Optical Science and Technology Center and Department of Chemistry, University of Iowa, Iowa City, Iowa 52242

Calibration standardization methodology for near-infrared (near-IR) spectroscopy is described for updating a partial least-squares calibration model to take into account changes in instrumental response. The guided model reoptimization (GMR) algorithm uses a transfer set of eight samples to characterize the new response and a database of previously acquired spectra used to develop the original calibration model. The samples in the transfer set need not have been measured under the old instrumental conditions, making the algorithm compatible with samples that change over time. The spectra comprising the transfer set are used to guide an iterative optimization procedure that (1) finds an optimal subset of samples from the original database to use in computing the updated model and (2) finds an optimal set of weights to apply to the spectral resolution elements in order to minimize the effects of instrumental changes on the computed model. The optimization relies on an alternating grid search and stepwise addition/deletion steps. The algorithm is evaluated through the use of combination region near-IR spectra to determine physiological levels of glucose in a synthetic biological matrix containing bovine serum albumin and triacetin in phosphate buffer. The ability to update a calibration to account for changes in the response of a Fourier transform spectrometer over four to six years is examined in this study. Separate spectral databases collected in 1994 and 1996 are used with a transfer set and separate test set of spectra collected in 2000. With the 1994 database, the standardization algorithm achieves a standard error of prediction (SEP) of 0.69 mM for the 2000 test set. This compares favorably to SEP values >2 mM when the original 1994 calibration model is used without standardization. A similar improvement in the prediction performance of the 2000 test set is obtained after standardization with the 1996 database (SEP = 0.70 mM).

Successful applications of near-infrared (near-IR) spectroscopy have been reported in the pharmaceutical, chemical, food, and petroleum industries.^{1–5} The advantages of this technique include rapidity, simplicity, nondestructive measurement, and the ability to measure many samples directly without pretreatment. To take advantage of these positive features, however, the analyst must overcome limitations in sensitivity and selectivity that arise from the relatively weak and highly overlapping spectral bands found in the near-IR region. A key step in the implementation of a successful analysis is the use of multivariate calibration models that allow analyte bands to be extracted from the background absorbance arising from the sample matrix.

While the use of multivariate calibration models has allowed many near-IR applications to be developed, two characteristics of these models are problematic. First, significant effort and expense are associated with the collection of sufficient calibration data for use in developing a reliable model. Second, the developed model may become invalid because of spectral differences between the calibration and prediction steps. When the calibration model is based on weak analyte signals, even small spectral changes can negatively impact model performance.

This lack of calibration stability can arise from changes in instrumental response or sample-based variation. Within an instrument, the response may change with time because of drifting (slow continuous changes, such as those caused by aging) or shifting (sudden changes arising from replacement or repair of a part). Between instruments, even small differences in components can lead to an overall difference in response. Possible sources of sample-based variation include changes in physical characteristics such as sample temperature or particle size. In this situation, the calibration model developed under one set of conditions may become invalid when the conditions change.

- (1) Martens, H.; Næs, T. *Multivariate Calibration*; Wiley: New York, 1989.
- (2) Arnold, M. A.; Small, G. W. *Anal. Chem.* **1990**, *62*, 1457–1464.
- (3) Wang, Y.; Veltkamp, D. J.; Kowalski, B. R. *Anal. Chem.* **1991**, *63*, 2750–2756.
- (4) Wang, Z.; Dean, T.; Kowalski, B. R. *Anal. Chem.* **1995**, *67*, 2379–2385.
- (5) Alam, M. K.; Franke, J. E.; Niemczyk, T. M.; Maynard, J. D.; Rohrscheib, M. R.; Robinson, M. R.; Eaton, R. P. *Appl. Spectrosc.* **1998**, *52*, 393–399.

[†] Present address: Pfizer Global R&D, Groton Laboratories, MS 4137, 558 Eastern Point Rd., Groton, CT 06340-5196.

The ideal solution to these problems would be to build a new calibration model based on a full recalibration. However, multivariate calibration models are generally based on a large number of samples measured over time. Thus, a full recalibration usually involves considerable effort, cost, and time. To make the old spectral database useful for the new situation and avoid the overhead of a full recalibration, many multivariate calibration standardization methods have been developed.^{6,7}

These standardization methods can be divided into two categories. Methods in the first category transform spectra collected in the new situation to match those measured in the old situation so that the original calibration model can still be used. Direct standardization (DS) transforms spectra using a global regression model,³ whereas piecewise direct standardization (PDS) is a related approach that transforms spectra by use of many local regression models.^{3,4} Blank et al. have used finite impulse response (FIR) filtering to map spectra measured in the new situation to the old situation by using a spectrum from the old situation.⁸ Walczak et al. proposed a standardization method based on transferring spectra in the wavelet domain,⁹ and Despagne et al. used neural networks to transfer spectra.¹⁰ Xie et al. treated the calibration standardization task as a data reconstruction problem based on positive matrix factorization.¹¹

Methods in the second category are used to attempt to build calibration models that are robust to differences between instrumental responses and experimental conditions. This approach can be implemented by incorporating all relevant sources of variation in the calibration design, eliminating variation by data preprocessing, or selecting wavelength regions robust to instrumental variation. Swierenga et al. enhanced the robustness of calibration models by variable selection with simulated annealing.¹² Sjöblom et al. used orthogonal signal correction as a preprocessing technique to make spectra less dependent on instrumental variation.¹³ Swierenga et al.¹⁴ and de Noord¹⁵ have also explored preprocessing strategies.

Most of these standardization methods require the measurement of a set of representative samples to calculate transfer parameters or help remove the differences between the two situations. The number of these samples must be significantly fewer than needed for a full recalibration if the method is to be practical. Except for the FIR filtering technique, all methods in the first category require the same set of representative samples to be measured in both situations. This requirement is problematic for samples that change over time (e.g., clinical or biological samples). Methods in the second category require a set of representative samples to be measured in the new situation. The

same set of samples need not be measured in the old and new situations, however.

In this paper, a new multivariate calibration standardization approach based on the idea of reoptimizing the original calibration model through wavelength and calibration sample selection is proposed. A small set of representative samples (termed the transfer set) is measured in the new situation and used to guide the reoptimization procedure. This guided model reoptimization (GMR) algorithm is investigated for use in the determination of physiological levels of glucose in a synthetic biological matrix.

EXPERIMENTAL SECTION

Design of Data Sets. Three data sets with the same constituents collected with the same instrument over a period of 6 years were used in this study. A synthetic biological matrix consisting of physiological levels of glucose, bovine serum albumin (BSA), and triacetin (glyceryl triacetate) in 0.1 M phosphate buffer was used. This sample matrix will be termed GTB in this discussion. The data sets were collected in 1994, 1996, and 2000 and will be termed data sets *1994*, *1996*, *2000*, respectively.

Data set *1994* employed 160 samples based on a full factorial design with 10 levels of glucose (1, 3, 5, 7, 9, 11, 13, 15, 17, and 19 mM), 4 levels of triacetin (1.4, 2.1, 2.8, and 3.5 g/L), and 4 levels of BSA (49.3, 64.4, 79.8, and 94.7 g/L). These concentration levels were selected to coincide with those encountered in clinical samples, such as blood or plasma.

Data set *1996* was designed by a genetic algorithm design program¹⁶ that chose optimal levels of glucose, BSA, and triacetin from a pool of possible concentrations to achieve the goal of minimizing correlations between any of the component concentrations and maximizing the number of glucose levels. The data set consisted of 100 samples with candidate glucose concentrations ranging from 1.00 to 20.00 mM at 0.50 mM intervals, candidate triacetin concentrations ranging from 1.40 to 3.50 g/L at 0.35 g/L intervals, and candidate BSA concentrations ranging from 49.40 to 95.00 g/L at 1.90 g/L intervals. The final design had 37, 6, and 18 levels of glucose, triacetin, and BSA, respectively. Across the 100 samples, the maximum correlation coefficient between any two components was -0.054 (glucose and triacetin).

Data set *2000* consisted of a test set and a transfer set designed by the uniform design (UD) method.^{17,18} The test set was designed to approximate a 20-level UD table and consisted of 20 samples with 19 levels of glucose, 12 levels of BSA, and 10 levels of triacetin. The transfer set was designed to approximate an 8-level UD table and consisted of 8 samples with 8 levels of glucose, 8 levels of BSA and 8 levels of triacetin. For each set, the levels of glucose, triacetin and BSA concentrations were uniformly distributed between 1 and 19 mM, 1.4 and 3.5 g/L, and 49.3 and 94.7 g/L, respectively.

Apparatus and Reagents. The spectra for all data sets were collected with a Digilab FTS-60A Fourier transform (FT) spectrometer (Bio-Rad, Inc., Cambridge, MA) equipped with a 100 W tungsten-halogen lamp, CaF₂ beam splitter, and cryogenically cooled InSb detector. A K-band interference filter (Barr Associates,

- (6) de Noord, O. E. *Chemom. Intell. Lab. Syst.* **1994**, *25*, 85–87.
- (7) Bouveresse, E.; Massart, D. L. *Vib. Spectrosc.* **1996**, *11*, 3–15.
- (8) Blank, T. B.; Sum, S. T.; Brown, S. D.; Monfre, S. L. *Anal. Chem.* **1996**, *68*, 2987–2995.
- (9) Walczak, B.; Bouveresse, E.; Massart, D. L. *Chemom. Intell. Lab. Syst.* **1997**, *36*, 41–51.
- (10) Despagne, F.; Walczak, B.; Massart, D. L. *Appl. Spectrosc.* **1998**, *52*, 732–745.
- (11) Xie, Y.; Hopke, P. K. *Anal. Chim. Acta* **1999**, *384*, 193–205.
- (12) Swierenga, H.; de Groot, P. J.; de Weijer, A. P.; Derksen, M. W. J.; Buydens, L. M. C. *Chemom. Intell. Lab. Syst.* **1998**, *41*, 237–248.
- (13) Sjöblom, J.; Svensson, O.; Josefson, M.; Kullberg, H.; Wold, S. *Chemom. Intell. Lab. Syst.* **1998**, *44*, 229–244.
- (14) Swierenga, H.; Haanstra, W. G.; de Weijer, A. P.; Buydens, L. M. C. *Appl. Spectrosc.* **1998**, *52*, 7–16.
- (15) de Noord, O. E. *Chemom. Intell. Lab. Syst.* **1994**, *23*, 65–70.

- (16) Thathapudi, N. G. M.S. Thesis, Ohio University, 1998.
- (17) Fang, K. T.; Wang, Y. *Number-theoretic Methods in Statistics*; Chapman and Hall: London, 1993.
- (18) Fang, K. T. *Uniform Design and Uniform Design Tables*; Science Press: Beijing, 1994.

Westford, MA) was used to restrict the incident light to the first combination region of 5000–4000 cm⁻¹. Samples were contained in an Infrasil quartz transmission cell with a path length of 2 mm (International Crystal Laboratories, Garfield, NJ). A water-jacketed cell holder was used in conjunction with a refrigerated temperature bath (Fisher model 9100, Fisher Scientific, Pittsburgh, PA) to control sample temperatures to the physiological range near 37 °C. Sample temperatures were monitored by use of a type-T thermocouple and digital thermocouple meter (Omega Engineering, Stamford, CT). The mean values and associated standard deviations of the sample temperatures were 37.3 ± 0.1, 37.0 ± 0.1, and 37.0 ± 0.1 °C for data sets 1994, 1996, and 2000, respectively. During the six-year period, the reference He:Ne laser in the spectrometer was replaced twice.

The same sample preparation protocol was used for all data sets. Individual stock solutions were prepared by weighing glucose (ACS reagent, Fisher Scientific), triacetin (99%, Sigma Chemical Co., St. Louis, MO), or BSA (Cohn Fraction V powder, product no. A-4503, Sigma Chemical Co.) and diluting with 0.1 M phosphate buffer. For data sets 1994 and 1996, the phosphate buffers were at pH 7.4. For data set 2000, the buffer was at pH 7.5. For data set 1994, 0.044% w/w 5-fluorouracil was added to the buffer solution as a preservative. For data sets 1996 and 2000, 0.5% w/w sodium benzoate was used as a preservative instead. Buffer solutions were prepared with reagent-grade water obtained with a Milli-Q Plus water purification system (Millipore, Inc., Bedford, MA). Individual mixture samples were prepared by dilution of appropriate volumes of the stock solutions with the phosphate buffer. Class A volumetric glassware was used.

Procedures. Samples in data set 1994 were collected over 65 days during 18 individual data collection sessions with the exception of one remeasured sample. Samples in data set 1996 were collected over 53 days during 10 individual data collection sessions. Samples in data set 2000 were collected over 4 days during three data collection sessions. For data sets 1994 and 2000, spectra were collected as 16 384-point single-sided interferograms based on 256 co-added scans. For data set 1996, spectra were collected as 8192-point single-sided interferograms based on 512 co-added scans. Each spectral scan lasted ~2 min. In all cases, samples were run in a random order with respect to glucose concentrations to reduce the correlation between glucose concentration and any time-dependent artifacts. For each sample or buffer, the sample cell was filled and placed in the spectrometer, the solution was allowed to equilibrate to 37 °C, and three replicate spectra were measured consecutively. Buffer spectra were measured periodically throughout the data collection. Interferograms were Fourier-processed with triangular apodization and Mertz phase correction to generate single-beam spectra with a point spacing of 1.9 cm⁻¹. Interferograms in data set 1996 were zero-filled to 16 384 points to allow this point spacing to be achieved. Software resident on the Bio-Rad SPC-3200 computer controlling the spectrometer was used in the Fourier processing calculations.

All further computations were performed under MATLAB (version 6.0, The MathWorks, Inc., Natick, MA) using a Silicon Graphics ORIGIN 200 R12000 server (Silicon Graphics, Inc., Mountain View, CA). In this study, nine spectra in absorbance units (AU) were obtained for each sample by ratioing all combinations of the three replicate spectra of the sample to the three most

recently collected single-beam spectra of buffer and converting the resulting transmittance values to absorbance. The inclusion of replicate spectra in the analysis allowed some temperature variation to be incorporated into the data sets. All replicate spectra were treated together in the calculation and evaluation of the calibration models. Use of the word “sample” in the text refers to the group of nine spectra corresponding to one chemical sample. For example, during the operation of the standardization algorithm, individual samples were deleted from the calibration data. This entailed the deletion of the entire group of nine spectra corresponding to the chemical sample. Similarly, cross-validation calculations were performed on the basis of the iterative removal of one sample from the calibration data. During each step in this procedure, the group of nine spectra was withheld together.

Our previous research has demonstrated that the spectral signals at the ends of the 5000–4000 cm⁻¹ range are extremely noisy because of the high absorbance of water. Use of these regions can deteriorate the performance of calibration models. Accordingly, all calibration work performed here was restricted to the range of 4700–4300 cm⁻¹.

THEORY

According to convention in multivariate calibration standardization, the data set originally used in model building is termed the primary data set and the data set used to validate the ability to perform calibration standardization is termed the secondary data set. This terminology will also be used here.

The quality and robustness of a multivariate calibration model depends on the quality of the calibration samples and the quality of the spectral information available for use in constructing the model. The GMR algorithm presented here embodies these concepts through the use of calibration sample selection and a weighted partial least-squares (PLS) regression method. An overview of the procedures is provided in this section, whereas specific details regarding the algorithm steps are provided in the Appendix.

Weighting Strategy. The idea behind wavelength selection or weighting is that some spectral channels are more robust than others to instrumental changes. Wavelength selection can be considered an extreme case of weighting (i.e., the weights of some resolution elements are 0). By giving larger weights to some channels and smaller weights to others, the robustness of the calibration model can be improved. For multiple linear regression (MLR) models in near-IR spectroscopy, Mark and Workman developed a method for the selection of spectral channels that are robust to wavelength shifts.¹⁹ An iterative weighting procedure is incorporated into the GMR algorithm for use in conjunction with PLS regression.

Several implementations of the PLS algorithm have been reported.¹ For a calibration model between a concentration vector **c** and spectral data matrix **A**, each of these methods produces a model of the form

$$\mathbf{c}_{n \times 1} = \mathbf{A}_{n \times k} \times \mathbf{b}_{k \times 1} + \mathbf{e}_{n \times 1} \quad (1)$$

where *n* is the number of observations (i.e., spectra), *k* is the number of variables (i.e., spectral resolution elements), **b** is a

(19) Mark, H.; Workman, J. *Spectroscopy* **1998**, *11*, 28–36.

vector of PLS regression coefficients, and \mathbf{e} is the model residual vector. The PLS calculation employed here mean-centered the concentration and spectral variables, thereby producing a model with no intercept term. The computed means of the calibration data were saved and subsequently used in applying the model for prediction.

The regression coefficients vector, \mathbf{b} , contains useful information for variable selection/weighting. Centner et al. proposed a method to eliminate uninformative variables on the basis of a criterion using the \mathbf{b} coefficients.²⁰ The standard deviation of regression coefficient b_i , $s(b_i)$, is a useful index indicating the quality of the corresponding i th variable. Usually, $s(b_i)$ can be estimated empirically on the basis of n replicate regression coefficients obtained through leave-one-out jackknifing. Based on a similar idea, an add-one-in approach is incorporated in the GMR algorithm to calculate $s(b_i)$ making use of samples from the transfer set. During each iteration, one sample from the transfer set is added to the calibration set, and a PLS model is computed. An estimate of $s(b_i)$ is obtained from the resulting set of regression coefficient vectors. A smaller $s(b_i)$ indicates that resolution element i is less sensitive to differences in the samples measured in the new situation and is therefore considered to be more robust.

Weights for each variable are constructed as inversely proportional to $s(b_i)$. In the proposed algorithm, this is implemented by use of a linear mapping function between the $s(b_i)$ and weights. For a given set of regression coefficient vectors, different mappings can be achieved by using a different scaling value, α . The larger the value of α , the less the effect of weighting. When α is 1, every element is given weight 1 (i.e., no weighting is performed); when α is 0, the element which gives the maximum $s(b_i)$ is given a weight of 0. A grid search is embedded into the algorithm to search for the optimal α on the basis of the prediction error sum of squares (PRESS) for the transfer set.

$$\text{PRESS} = \sum_{i=1}^m (\hat{c}_i - c_i)^2 \quad (2)$$

For the i th of m spectra in the transfer set, the values of \hat{c}_i and c_i in eq 2 denote the analyte concentration predicted by the PLS model and the reference concentration, respectively. The details of the weighting process are presented in the Appendix.

Sample Selection Strategy. In the calibration standardization scenario, the goal of sample selection is to eliminate samples that are not helpful for building a calibration model for prediction in the new situation. Subset selection is a combinatorial optimization problem, and it is difficult to avoid the problem of becoming trapped in local optima. To simplify the algorithm, a sample selection method similar to variable selection in stepwise MLR was implemented instead of a global optimization method.

The sample selection procedure begins by defining a starting group of samples to delete from the calibration set. Each sample in the calibration set is removed one at a time, the calibration model is built, and PRESS for the transfer set is calculated. A threshold PRESS value is established, and those samples whose deletion results in a PRESS less than the threshold are placed in the initial deletion set.

The algorithm then attempts to add one sample from the deleted sample set back into the calibration set. If there is an improvement in PRESS, the sample will be kept. Similarly, the algorithm tries to delete one sample from the calibration set. If there is an improvement in PRESS, the sample will be deleted. These two steps alternate in an iterative manner and terminate when both adding and deleting samples cannot improve the PRESS value. Details of the addition and deletion processes are provided in the Appendix.

Selection of Latent Variables. During the weighting and sample selection steps of the GMR algorithm, the number of PLS latent variables is varied between 1 and h_{\max} , where h_{\max} is determined from an initial leave-one-sample-out cross-validation study performed on the calibration data. In determining h_{\max} , models are constructed on the basis of 1 to 20 latent variables, and values of PRESS are computed from the results of the cross-validation predictions. For each of the i models, the following test statistic is computed:

$$F_{\text{calc}} = \text{PRESS}_i / \text{PRESS}_{\min} \quad (3)$$

In eq 3, PRESS_i and PRESS_{\min} correspond to the PRESS values of the model based on i latent variables and the model that produced the minimum PRESS, respectively. For each model, the computed value of F_{calc} is compared to a tabulated value of the F distribution at a 95% confidence level. The value of h_{\max} is taken as the smallest i that produces a value of F_{calc} statistically indistinguishable from the tabulated F . This procedure assumes that the optimal model for the prediction of the transfer samples should require no more latent variables than the full calibration set. Given that samples are deleted from the calibration data during the standardization algorithm, this assumption seems justifiable. Without this restriction to h_{\max} latent variables, we observed cases of overfitting in which predictions with the transfer spectra were good, but subsequent predictions with the final test spectra were degraded.

The F -test procedure described above is also used to determine, h , the optimal number of latent variables to use with models developed during the iterative weighting and sample selection components of the GMR algorithm. In this case, PRESS values obtained from the predictions of the transfer samples are used with eq 3. Additional details regarding the implementation of this procedure are provided in the Appendix.

RESULTS AND DISCUSSION

Characterization of Spectral Information. As part of ongoing work in our laboratories directed to the development of near-IR glucose sensors, spectral data sets have been collected with a variety of synthetic and real biological matrixes.^{2,21–24} In the work described here, a synthetic sample matrix was employed in which BSA and triacetin were used to model blood proteins and

(20) Centner, V.; Massart, D. L.; de Noord, O. E.; de Jong, S.; Vandeginste, B. M.; Sterna, C. *Anal. Chem.* **1996**, *68*, 3851–3858.

(21) Small, G. W.; Arnold, M. A.; Marquardt, L. A. *Anal. Chem.* **1993**, *65*, 3279–3289.

(22) Pan, S.; Chung, H.; Arnold, M. A.; Small, G. W. *Anal. Chem.* **1996**, *68*, 1124–1135.

(23) Hazen, K. H.; Arnold, M. A.; Small, G. W. *Anal. Chim. Acta* **1998**, *371*, 255–267.

(24) Burmeister, J. J.; Arnold, M. A.; Small, G. W. *Diabetes Technol. Ther.* **2000**, *2*, 5–15.

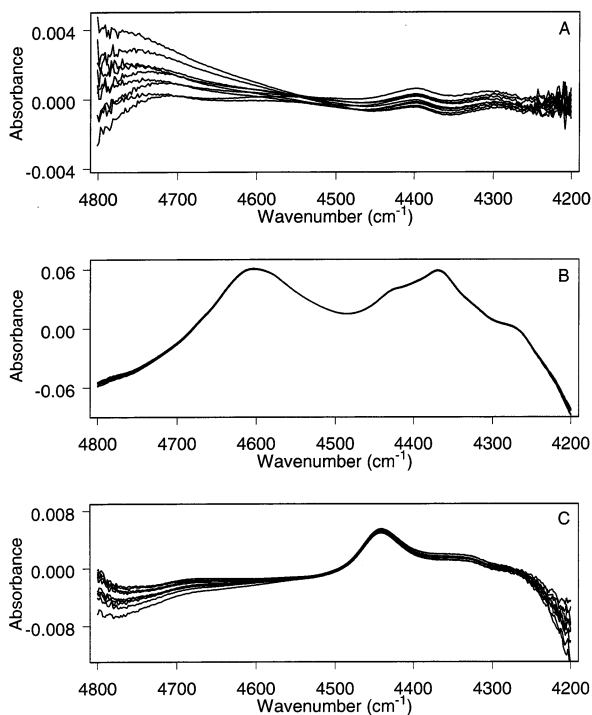


Figure 1. Spectra (AU) of 8.9 mM glucose (A), 70.0 g/L BSA (B), and 2.5 g/L triacetin (C) in phosphate buffer. For each compound, nine spectra are plotted corresponding to all combinations of the ratios of three replicate single-beam sample spectra to three replicate single-beam spectra of buffer. Note that the spectra of BSA have a significantly larger y-axis scale than those of the other two compounds. Negative absorbance values occur in regions of high water absorbance because of the lower concentration of water in the sample than in the phosphate buffer used as the spectral background in the absorbance calculation.

triglycerides, respectively. The quantitative analysis of glucose in this system is complicated by the small absorptivities associated with the near-IR glucose bands, a strong and temperature-sensitive background absorbance of water, and spectral interference from the other matrix constituents. At the concentrations and optical path lengths used here, glucose absorbance values are in the 10^{-3} – 10^{-4} AU range. This chemical system presents an excellent test case for calibration standardization, because even relatively minor instrumental drift can be significant when considered in the context of the small analyte signals that form the basis for the calibration model.

Absorbance spectra of pure component solutions of glucose (8.9 mM), BSA (70.0 g/L), and triacetin (2.5 g/L) in phosphate buffer are presented in Figure 1. These concentrations represent roughly the midpoint levels in the calibration designs employed in the three data sets used in this work. Data collection procedures were the same as those used with data set 2000, and a single-beam spectrum of phosphate buffer was used as the spectral background in the absorbance calculation. For each compound, nine spectra are plotted corresponding to the ratios of each of three replicate sample spectra to each of three replicate spectra of buffer. Solution temperatures were held to the range of 36.9–37.0 °C by use of the procedures described previously. Negative absorbance values observed in Figure 1 arise because of the differences in water concentration between the samples and the buffer solution.

Three bands are present in the glucose spectra in Figure 1A: a broad O–H combination band centered at 4700 cm^{-1} and C–H combination bands near 4400 and 4300 cm^{-1} . The glucose bands are located in a valley between two dominant absorption bands of water centered at 5200 and 3800 cm^{-1} . The effect of the tails of these water bands can be observed in the large spectral variation below 4300 and above 4700 cm^{-1} in Figure 1A. The magnitude of this variation reinforces the decision to restrict the calibration calculations to the 4700 – 4300 spectral range. Baseline variation among the nine replicate spectra is also caused by the slight temperature variation between the sample and buffer spectra used to compute the absorbance values. One can observe clearly in Figure 1A that this baseline variation is similar in magnitude to the glucose absorption features at this midlevel concentration of $\sim 9\text{ mM}$. Of the three glucose bands, the C–H combination band at 4400 cm^{-1} has been found to be the most useful for quantitative analysis because of its more distinct profile and its location near the minimum in the water absorbance.^{2,25}

The BSA spectrum (Figure 1B) has broad N–H and C–H combination bands near 4610 and 4360 cm^{-1} , respectively. The triacetin spectrum (Figure 1C) shows a large C–H combination band near 4450 cm^{-1} . Significantly less variation is observed in the BSA spectrum because of the higher concentration and subsequently larger absorbance signal. Baseline variation of the magnitude observed in Figure 1A is virtually undetectable on this absorbance scale. The glucose band at 4400 cm^{-1} overlaps with the 4360 cm^{-1} band of BSA and the 4450 cm^{-1} band of triacetin. Because of this extensive spectral overlap, multivariate calibration methods are necessary for quantitative analysis of glucose in this system. Furthermore, the degree of spectral variation observed in Figure 1A for even the midlevel glucose concentration underscores the difficulty of this calibration problem.

Characterization of Data Sets. Figure 2A presents absorbance spectra of GTB mixture samples at similar concentrations from data sets 1994 (solid line), 1996 (long-dashed line), and 2000 (short-dashed line). Across the three samples, the glucose, BSA, and triacetin concentrations range from 12.50 to 13.20 mM, 79.80 to 79.88 g/L, and 2.10 to 2.16 g/L, respectively. Spectra of phosphate buffer were used as the reference measurement in the absorbance calculation. For each data set, three replicate spectra are plotted corresponding to the ratio of each of three sample single-beam spectra to a single spectrum of buffer collected during the same day. Comparison of these spectra with the spectra in Figure 1 reveals that the GTB mixture spectra appear qualitatively identical to a spectrum of BSA.

The spectral differences in Figure 2A arise primarily from baseline artifacts induced by temperature variation. Figure 2B plots the same set of spectra after a simple two-parameter linear least-squares baseline correction across the 4800 – 4200 cm^{-1} range. After baseline correction, the spectra from the three data sets are very similar. It must be stressed, however, that the calibration problem being investigated in this work involves glucose spectral bands that cannot be visually observed in the spectra in Figure 2. Very small changes in the spectra can invalidate the calibration model when the analyte absorbances are so small.

(25) Hazen, K. H.; Arnold, M. A.; Small, G. W. *Appl. Spectrosc.* **1994**, *48*, 477–483.

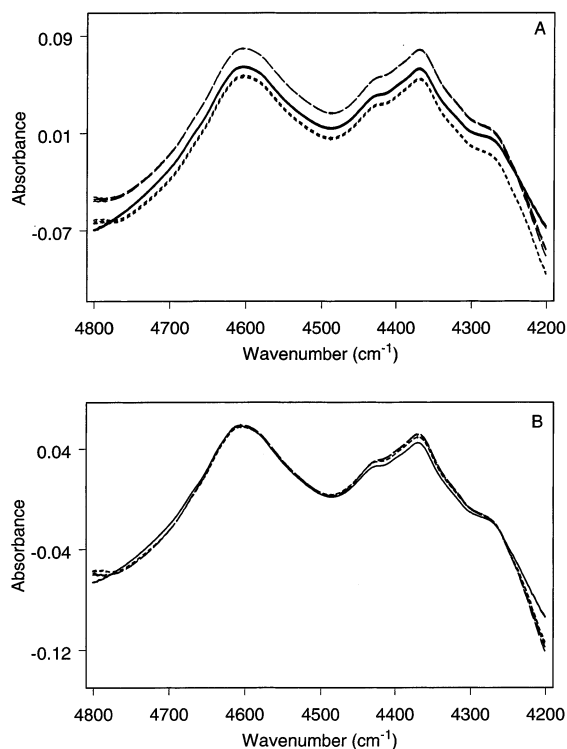


Figure 2. Spectra (AU) of GTB mixture samples of similar composition from data sets 1994 (solid line, 12.99 mM glucose, 79.88 g/L BSA, 2.10 g/L triacetin), 1996 (long dashed line, 12.50 mM glucose, 79.80 g/L BSA, 2.14 g/L triacetin), and 2000 (short dashed line, 13.20 mM glucose, 79.82 g/L BSA, 2.16 g/L triacetin). For each data set, three spectra are plotted corresponding to the ratios of three replicate single-beam sample spectra to a single-beam spectrum of phosphate buffer collected during the same data collection session. A. Spectra are presented without any preprocessing. B. Spectra are presented following a two-parameter linear least-squares baseline correction performed across the 4800–4200 cm^{-1} range. The mixture spectra are visually identical to the spectra of pure BSA in Figure 1B. No glucose or triacetin bands can be observed in the mixture spectra because of the dominant absorption of BSA.

Characterization of Spectral Quality. The performance of any multivariate calibration model is keyed to the quality of the spectra used in building and applying the model. Spectral quality is of similar importance in determining the success of calibration standardization. To evaluate the quality of the spectra used here, both short-term and long-term spectral variation were studied in each data set.

Analysis of Short-Term Spectral Variation. Short-term variation was studied through the calculation of noise levels in spectral 100% lines. Employing the three replicate single-beam spectra for each sample, three noise spectra in AU were computed by ratioing all combinations of the replicate spectra to each other and converting the resulting transmittance values to absorbance. For all three data sets, the collection of the three replicate spectra spanned approximately 6 min.

In the absence of intensity drift and variation in sample temperature among the replicate spectra, these noise absorbance spectra will be centered at 0 AU, and the noise value at resolution element i will be random in sign. If the principles of error propagation are applied to the absorbance calculation, the noise will be proportional in magnitude to s_i/I_i , where s_i is the noise in the single-beam spectrum (assumed to be constant with i for

measurements made with an FT spectrometer) and I_i is the single-beam spectral intensity at i (assumed to be constant for each replicate spectrum). Noise in absorbance is, thus, lowest where the single-beam intensity is highest. Simple intensity drift between the replicate single-beam spectra will cause the noise absorbance spectra to shift above or below 0 AU. Variation in sample temperature produces both wavenumber shift and a change in magnitude of the water absorbance. This effect is magnified here because of the high water concentration. The glucose and BSA spectra also exhibit some temperature sensitivity, although the effect is insignificant compared to that of water. When this occurs across the replicate single-beam spectra used in the noise calculation, the resulting noise spectra will exhibit a curved baseline.

To characterize random noise, drift-induced variation, and temperature variation, three root-mean-square (RMS) calculations were implemented over the range of 4500–4300 cm^{-1} . This spectral region encompasses the important glucose band centered near 4400 cm^{-1} . For each calculation, the RMS values produced by the three pairwise combinations of the replicate single-beam spectra were pooled to produce a single RMS value for each sample. By employing a mean of 0 in the RMS calculation, all variation about 0 AU is encoded ($\text{RMS}_{\text{total}}$). By employing the computed mean value across the 4500–4300 cm^{-1} range in the RMS calculation, simple drift-induced variation is removed. The resulting noise is termed RMS_{mean} . Finally, by modeling the curvature of the noise spectrum through a least-squares fit to a second-order polynomial function, both shift and curvature can be largely removed. The RMS calculation is then performed about this fitted baseline, producing $\text{RMS}_{\text{second-order}}$. Although this calculation is effective in most cases, some structure remains in the noise spectrum when the temperature range spanned by the single-beam spectra exceeds 0.5 °C. The curvature is not fully modeled by the second-order polynomial in this case.

These calculations were applied to each data set, and the distributions of the resulting $\text{RMS}_{\text{total}}$, RMS_{mean} , and $\text{RMS}_{\text{second-order}}$ values were characterized. The total short-term variation ($\text{RMS}_{\text{total}}$) represented by the 6-min data collection was on the order of 1000 μAU (0.001 AU) and was similar for each data set. Data set 2000 had the largest median value of $\text{RMS}_{\text{total}}$, but the narrowest distribution. The narrow width reflects a consistent level of instrumental performance over the short four-day time span of the data collection. The larger distribution widths for data sets 1994 and 1996 indicate a greater range in instrument stability across the larger number of samples and significantly greater time span associated with these data sets.

Our approach assumes that RMS_{mean} represents the removal of drift-induced variation from $\text{RMS}_{\text{total}}$ and that $\text{RMS}_{\text{second-order}}$ represents the further removal of most of the temperature-induced variation. Comparison of the distributions formed from the three noise calculations across the data sets revealed that $\text{RMS}_{\text{total}}$ was dominated by intensity drift and that the temperature-induced variation ($\sim 100 \mu\text{AU}$) exceeded the random noise level ($\sim 20 \mu\text{AU}$) by a factor of ~ 5 . The level of temperature-induced variation was similar across the data sets. This is reasonable, given that the same equipment and procedures for controlling the sample temperature were used with each data set. The random noise levels were also similar across the data sets. Thus, the overall

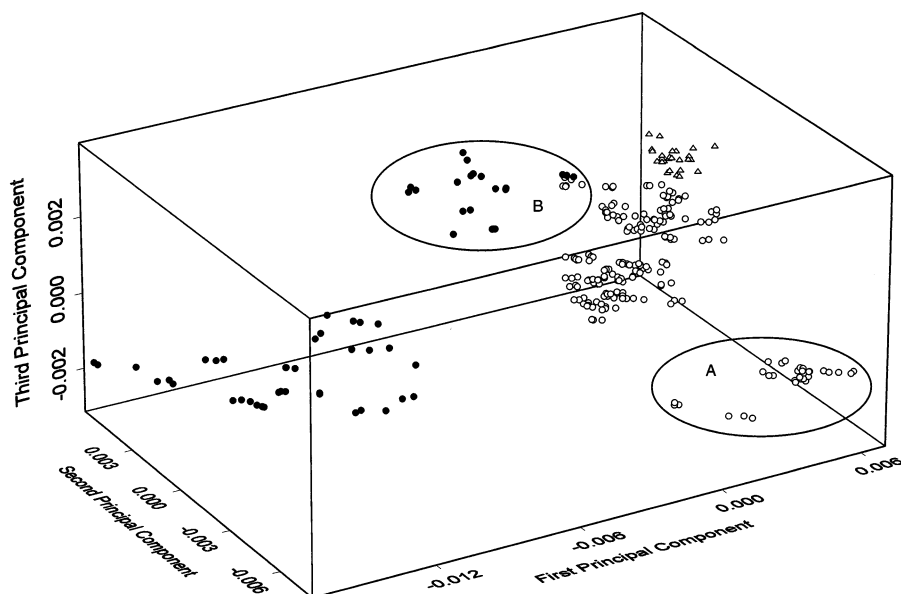


Figure 3. Principal components score plot computed from the 300 single-beam spectra of phosphate buffer that span data sets 1994 (○), 1996 (●), and 2000 (△). The first three principal components account for 98.6% of the data variance. Within each data set, the points are clustered somewhat according to the time of data collection. For example, the points in region A correspond to the last three data collection sessions of data set 1994, whereas the solid circles in region B derive from the last three data collection sessions of data set 1996. Data set 1996 exhibits the greatest overall spectral variation, signaling the presence of significant variability in instrumental response. Greater similarity in clustering is observed between data sets 1994 and 2000 than between data sets 1996 and 2000.

conclusion from this analysis is that the three data sets are similar with respect to short-term spectral variation.

Analysis of Long-Term Spectral Variation. Long-term variation in the data sets was evaluated through analysis of the single-beam spectra of phosphate buffer collected several times during each data collection session in which spectra of the GTB samples were acquired. Except in three cases, these buffer spectra were collected in groups of three replicates. Two to three groups of spectra were typically acquired during each data collection session. The temperature precision of the buffer spectra was analogous to that obtained with the GTB samples. Across the time span of the data set, these spectra provide an assessment of the variation in overall instrumental response.

For data sets 1994, 1996, and 2000, 201, 70, and 29 single-beam buffer spectra were available for analysis. These spectra were normalized in the vector sense (sum of squares = 1.0) across the full 5000–4000 cm^{-1} region to correct for simple intensity drift. Principal component analysis (PCA)²⁶ was applied to this range in the full set of 300 mean-centered buffer spectra. PCA performs an orthogonal decomposition of the spectral data matrix into factors that are ordered according to the amount of variance they explain. The three most significant principal components computed from the 300 buffer spectra accounted for 98.6% of the data variance.

Figure 3 is a score plot that represents the projection of each spectrum onto the first three principal components. Open circles, solid circles, and open triangles denote the buffer spectra from data sets 1994, 1996, and 2000, respectively. Inspection of this figure reveals that the spectra from data set 1996 exhibit much greater variation than those from the other two data sets. The spectra from data set 2000 are the most tightly clustered,

indicative of the short four-day data collection. As detailed in the figure caption, there is general clustering in the score plot according to the time of data collection within each set. Overall, the spectra from data sets 1994 and 2000 are clustered more closely together than are data sets 1996 and 2000. This suggests that data sets 1994 and 2000 are perhaps a better match for the purpose of calibration standardization than data sets 1996 and 2000. In the evaluation of the GMR algorithm, data sets 1994 and 1996 will be used separately to evaluate the ability to tune the calibration model to data set 2000.

Calibration Standardization between Data Sets 1994 and 2000. Data set 1994 was used as the primary data set, and data set 2000 was used as the secondary data set. An initial study was performed to investigate the deterioration of the calibration model built with the spectral data collected 6 years previously. The calibration set contained 128 samples (80%) randomly selected from the 160 samples in the primary data set. The first test set consisted of the remaining 32 samples (20%). The second test set contained the 20 samples from the test set of the secondary data set.

By use of the 128 samples in the calibration set, leave-one-sample-out cross-validation results were obtained with PLS models on the basis of 1 to 20 latent variables. The minimum PRESS value for the cross-validation predictions occurred at 19 latent variables. Application of the *F*-test described previously resulted in the selection of 14 and 16 latent variables when the test was performed at the 95 and 70% levels, respectively. Employing 19, 16, and 14 latent variables, calibration models were built with the calibration set and predictions were made with all three sets. Values for the standard error of calibration (SEC), standard error of prediction (SEP) and bias for each set and model are listed in Table 1. The SEC is calculated as

(26) Jolliffe, I. T. *Principal Component Analysis*; Springer-Verlag: New York, 1986.

Table 1. Calibration and Prediction Results without Calibration Standardization

data set	calibration data ^a		internal test set ^b		test set from data set 2000	
	latent variables	SEC (mM)	SEP (mM)	Bias (mM)	SEP (mM)	bias (mM)
1994	19	0.49	0.66	-0.09	2.12	1.08
1994	16	0.55	0.63	-0.04	2.74	2.17
1994	14	0.61	0.68	-0.02	8.69	8.37
1996	15	1.16	1.06	-0.45	1.72	-1.13
1996	14	1.20	1.09	-0.50	4.15	-3.86

^a The calibration sets for data sets 1994 and 1996 consisted of 128 and 80 samples, respectively. ^b The internal test sets for data sets 1994 and 1996 consisted of 32 and 20 samples, respectively, that were withheld from the calibration sets.

$$SEC = \sqrt{\frac{\sum_{i=1}^n (c_i - \hat{c}_i)^2}{n - h - 1}} \quad (4)$$

where c_i is the measured analyte concentration, \hat{c}_i is the predicted analyte concentration, n is the number of calibration spectra, and h is the number of PLS factors used in the calibration model. Similarly, the SEP is calculated as

$$SEP = \sqrt{\frac{\sum_{i=1}^m (c_i - \hat{c}_i)^2}{m}} \quad (5)$$

where c_i and \hat{c}_i are the same as in eq 4, and m is the number of prediction spectra. The bias is calculated as

$$\text{bias} = \frac{\sum_{i=1}^n (\hat{c}_i - c_i)}{n} \quad (6)$$

where n is either the number of calibration or prediction spectra. The bias will always be effectively 0 for the calibration set, but it can reveal systematic error in the performance of the model when applied to a prediction set. Figure 4 provides a graphical display of the results for the model based on 16 latent variables. Concentration correlation plots are shown for the prediction results obtained with the (A) calibration set, (B) first test set, and (C) second test set.

The results in Table 1 and Figure 4 show that the calibration model works satisfactorily for predicting data collected during the same period. However, the model is incapable of predicting samples collected 6 years later. In Figure 4C, the obvious deviation of the predictions toward larger glucose concentrations and the bias value listed in Table 1 for the second test set show the inability of the calibration model to predict the new samples. To make use of the old calibration data, a calibration standardization procedure is clearly necessary.

Next, the GMR algorithm was employed to perform calibration standardization. All 160 samples from data set 1994 were included in the calibration set. The transfer set of eight samples from data

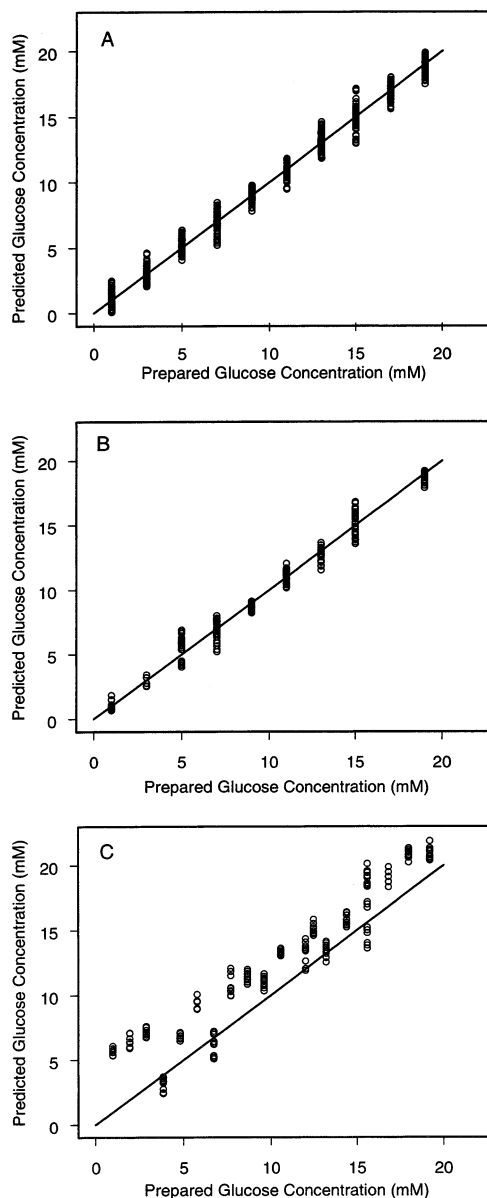


Figure 4. Correlation plots of predicted vs prepared glucose concentrations (mM) produced by the application of the calibration model computed from 128 samples in data set 1994 (16 latent variables). (A) Prediction of calibration samples. (B) Prediction of the first test set corresponding to 32 samples in data set 1994 withheld from the calibration set. (C) Prediction of the second test set consisting of 20 samples from data set 2000. Values of SEC, SEP, and bias corresponding to these correlation plots are listed in Table 1.

Table 2. Calibration and Prediction Results with Standardization Algorithm

data set	calibration set		transfer set 2000				test set 2000			
	beginning ^a	end ^b	beginning ^a		end ^b		beginning ^a		end ^b	
	SEC (mM)	SEC (mM)	SEP (mM)	bias (mM)	SEP (mM)	bias (mM)	SEP (mM)	bias (mM)	SEP (mM)	bias (mM)
1994	0.90	1.02	1.83	-0.77	0.67	-7.8×10^{-4}	1.87	-1.16	0.69	-0.02
1996	1.68	1.38	1.22	-0.05	0.77	-0.11	1.09	0.15	0.70	0.19

^a Results obtained with the calibration model corresponding to the start of the standardization algorithm (cycle 1). ^b Results obtained with the final optimized calibration model.

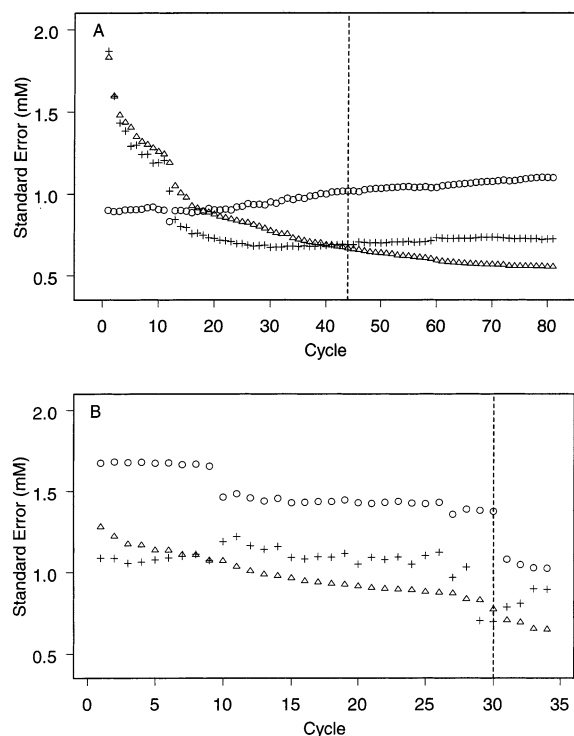


Figure 5. Calibration and prediction performance during the operation of the standardization algorithm with (A) calibration spectra from data set 1994 and transfer and test spectra from data set 2000 and (B) calibration spectra from data set 1996 and transfer and test spectra from data set 2000. Values listed in each plot are SEC (○), SEP for the transfer set (△), and SEP for the test set (+). These statistics were computed for each cycle of the algorithm. The vertical dashed line in each plot denotes the optimal model selected by application of the pruning procedure to the results of the transfer set.

set 2000 was used to guide the standardization algorithm, and the test set of 20 samples from data set 2000 was used as an independent set to evaluate the performance of the optimized calibration model. By use of the procedures described previously, the value of h_{\max} was set to 14.

Figure 5A plots traces of SEC (○), SEP for the transfer set (△), and SEP for the test set (+) for each cycle of the algorithm. The algorithm ended after 81 cycles, corresponding to the termination condition that PRESS for the transfer set could no longer be improved through the weighting or sample deletion/addition procedures. Inspection of Figure 5A reveals that the SEP for the transfer set continuously decreases, because this value is the driving force for the optimization. To minimize the possibility of overfitting to the transfer set, the optimal model for use in subsequent predictions was taken as the model for the earliest

cycle for which the corresponding PRESS value for the transfer set was statistically indistinguishable from the minimum PRESS value at cycle 81. The *F*-test described previously in the discussion of eq 3 (95% level) was used to select the model for cycle 44 as optimal. The vertical dashed line in Figure 5A identifies this cycle. For this model, 130 of the 160 samples remained in the calibration set. Comparison of the traces of the SEP values for the transfer and test sets reveals that the *F*-test identifies the crossover point in the two traces very effectively. Beyond the crossover point, the models are clearly overfitting the predictions performed with the transfer spectra.

Table 2 reports the values of SEC, along with the SEP and bias values for the transfer and test sets. These statistics are provided for cycles 1 and 44. The values of SEC reported in Table 2 are larger than that obtained previously when both calibration and prediction focused solely on spectra from data set 1994 and employed all 160 samples (Table 1). This arises primarily from a decrease in the number of latent variables found to be optimal in the two cases. At the end of the standardization algorithm, the number of latent variables was 10, as compared to the previously selected value of $h_{\max} = 14$. The larger number of PLS factors reflects the need to model the long-term spectral variation present in the full data set when the optimization criterion is the prediction performance across that data set. The higher-order factors likely contain specific information for data set 1994 that is not relevant for the prediction of samples in data set 2000.

As displayed in Figure 5A, the SEC is also observed to increase somewhat (0.90–1.02 mM) through the course of the optimization. This suggests the algorithm is forced to accept a tradeoff between modeling the calibration and transfer sets. Given that the goal of the optimization is to produce a calibration model for use with data set 2000, the increase in SEC is not problematic.

Plots A and B in Figure 6 display concentration correlation plots for the test set at the beginning and end, respectively, of the standardization algorithm. The two plots clearly illustrate the improvement in prediction performance obtained through the optimization of the model.

Calibration Standardization between Data Sets 1996 and 2000. To evaluate the algorithm further, data set 1996 was used as the primary data set, and data set 2000 was used as the secondary data set. As before, the transfer set from data set 2000 was used to guide the algorithm, and the test set from data set 2000 was used to evaluate the prediction performance of the optimized model.

To evaluate the internal performance of the calibration data, the set of 100 samples was randomly split into a calibration subset of 80 samples and an internal test set of 20 samples. Leave-one-

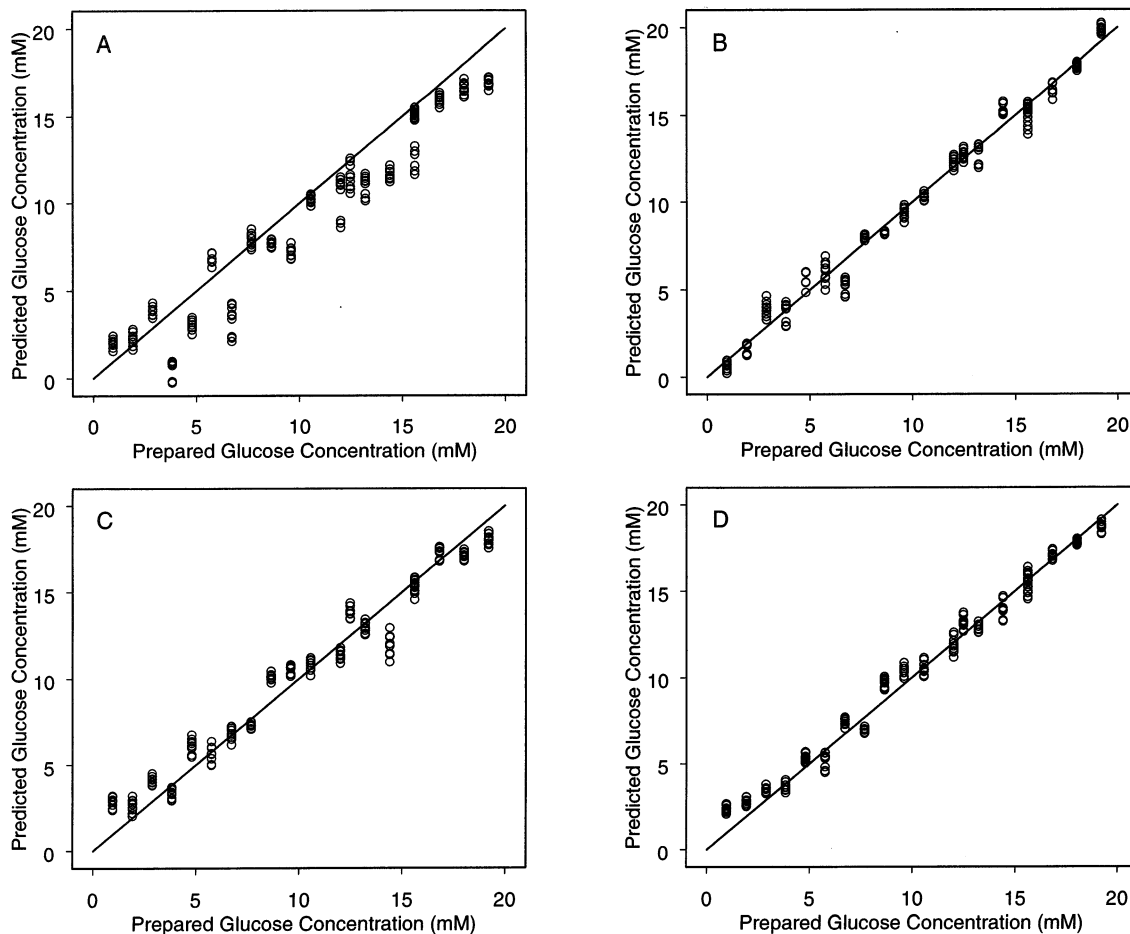


Figure 6. Correlation plots of predicted vs prepared glucose concentrations (mM) for spectra in the test set of data set 2000. (A) Prediction results after cycle 1 of the standardization algorithm when calibration spectra from data set 1994 were used. (B) Prediction results with the final optimized model based on calibration spectra from data set 1994. (C) Prediction results after cycle 1 of the standardization algorithm when calibration spectra from data set 1996 were used. (D) Prediction results with the final optimized model based on calibration spectra from data set 1996. Values of SEC, SEP, and bias corresponding to these models are listed in Table 2.

sample-out cross-validation was applied to the 80 samples to assess the number of latent variables required. The minimum cross-validation PRESS value occurred at 15 latent variables. Application of the F -test described previously at the 95% level produced a value of $h_{\max} = 14$. For models based on 15 and 14 latent variables, Table 1 lists the values of SEC for the 80-sample calibration subset. In addition, SEP and bias values are provided for the application of these models to the 20-sample internal test set and the test set from data set 2000. As observed previously when data set 1994 was used, models built without any standardization perform poorly when applied to data set 2000. The internal calibration and prediction results for data set 1996 are poorer than those obtained with data set 1994. As described previously in the discussion of Figure 3, this degraded performance can be attributed to the greater long-term spectral variation in this data set.

The GMR algorithm was next applied to the 100 samples in data set 1996, with the transfer set of eight samples from data set 2000 being used as before to guide the optimization. The algorithm terminated after 34 cycles. The model from cycle 30 was selected as optimal by use of the F -test described previously at the 95% level. At cycle 30, 79 of the 100 samples remained in the calibration set. The optimal number of latent variables was 12.

Figure 5B plots the optimization traces for SEC and SEP in a manner analogous to Figure 5A. The vertical dashed line denotes the optimal model at cycle 30. As before, the trace for SEP of the transfer set steadily decreases through the course of the algorithm. The trace for SEC is relatively stable except for two sharp dips. The drop in SEC between cycles 9 and 10 corresponds to a change in the optimal number of latent variables from 11 to 12. The change in SEC between cycles 30 and 31 occurs after deletion of sample 51. The F -test for selection of the final model is again effective in identifying the point at which overfitting to the transfer set occurs.

Plots C and D in Figure 6 display concentration correlation plots for the prediction of glucose concentrations in the test set with the calibration models corresponding to cycles 1 and 30. Improved prediction results are clearly obtained through the standardization procedure. Values for SEC, SEP, and bias are listed in Table 2 as before. The prediction results for the test set of data set 2000 are virtually identical to those obtained previously when data set 1994 was used. Despite the poorer calibration and internal prediction results (Table 1) that appear to result from the greater long-term spectral variation in data set 1996, the GMR algorithm is still able to tune the model effectively to data set 2000.

CONCLUSIONS

The results show that the GMR algorithm significantly improves prediction performance in the new situation by using calibration sample selection and a weighting procedure. Unlike standardization methods, such as DS or PDS, the identical set of samples need not be measured in both old and new situations. Clearly, it would be desirable to compare the performance of the GMR algorithm with that of DS and PDS. This was not possible with the data used here because the same samples were not available for measurement in the old and new situations. A comparison study of this type is currently underway in our laboratories with a new set of samples and a more complex background matrix.

Attempts were made to interpret the computed weights and spectral subsets as well as the PLS regression coefficients obtained at the beginning and end of the standardization algorithm. No simple interpretations were apparent, however. Interpretation of the algorithm decisions is particularly difficult for the data sets used here, because the number of possible spectral subsets is huge, the models are based on a relatively large number of latent variables, and the analyte signal represents a very small component of the overall spectral information.

This study did not incorporate optimization of the spectral range supplied to the PLS calculation or the use of preprocessing techniques, such as digital filtering.²¹ Incorporation of these additional processing and optimization steps into the standardization algorithm represents a potentially useful extension of the work reported here. In addition, no work was performed to optimize the number of transfer samples. Eight was chosen here as a reasonable number that could adequately span the concentration space of the sample matrix and provide a statistical sample large enough to be used to drive the optimization procedure. In the comparison study currently underway, this variable is being investigated. Finally, although the algorithm presented here was based on the use of PLS regression to compute the calibration model, the same weighting and subset selection strategies should be compatible with other model calculation procedures, such as principal component regression.

APPENDIX

Initialization of algorithm. The standardization algorithm consists of two major blocks, a weighted PLS block and a calibration sample selection block. The initial calibration set is $S_c = \{s_{c(1)}, s_{c(2)}, \dots, s_{c(n)}\}$ with n samples corresponding to the primary instrument/conditions, and the transfer set is $S_s = \{s_{s(1)}, s_{s(2)}, \dots, s_{s(m)}\}$ with m samples corresponding to the secondary instrument/conditions. Each spectrum has k resolution elements. The rows in spectral data matrix \mathbf{A} contain calibration sample spectra measured with the primary instrument/conditions and transfer sample spectra measured with the secondary instrument/conditions.

Leave-one-sample-out cross-validation is applied to set S_c to establish h_{\max} , the maximum number of latent variables to be investigated during the standardization calculations. The F test (95% level) described previously in the discussion of eq 3 is used to define h_{\max} .

To initialize the standardization calculations, the $k \times k$ weight matrix \mathbf{W}_{new} is set to the identity matrix, $S_{c(1)} = S_c$, the deleted sample set $S_{d(1)} = \{\}$, and cycle = 1. Note that $S_{c(1)}$ is the

calibration set for cycle 1, whereas $s_{c(1)}$ is the first sample in the calibration set. Except where mentioned explicitly, all spectral data are taken from spectral data matrix \mathbf{A} when performing calculations.

Step 1. Apply weights to the spectral data matrix according to the following equations.

$$\mathbf{A}_{\text{old}} = \mathbf{A} \quad (\text{A-1})$$

$$\mathbf{A} = \mathbf{A}_{\text{old}} \mathbf{W}_{\text{new}} \quad (\text{A-2})$$

Step 2. Implement the weight selection process.

Estimate the standard deviation of the regression coefficient for each resolution element.

(1) Determine h , the number of latent variables in the PLS model on the basis of calibration data set $S_{c(\text{cycle})}$ and prediction with the transfer set S_s using the F -test (95% level) described previously in the discussion of eq 3. In choosing h , evaluate models with latent variables from 1 to h_{\max} .

(2) Perform the following calculation for each of the m samples in the transfer set.

i. Set $S_c^* = S_{c(\text{cycle})} \cup \{s_{s(i)}\}$, where $s_{s(i)}$ is the i th of the m transfer samples.

ii. Build the i th PLS model based on set S_c^* using h factors.

iii. Calculate the regression coefficient vector of the i th PLS model, \mathbf{b}_i .

(3). Pool the resulting m regression coefficient vectors into an $m \times k$ matrix \mathbf{B} according to the following equation, where ' indicates transpose.

$$\mathbf{B} = \begin{bmatrix} \mathbf{b}_1' \\ \mathbf{b}_2' \\ \dots \\ \mathbf{b}_m' \end{bmatrix} \quad (\text{A-3})$$

(4) From \mathbf{B} , calculate the standard deviation of each column to yield $\mathbf{d} = [d_1 \ d_2 \ \dots \ d_k]'$.

(B) Perform a grid search to choose the optimal mapping from standard deviations to weights. Let β be a predefined grid search interval value with the constraint of $\beta = 1/(q - 1)$ where q is the user-specified number of grid search levels. A value of $q = 11$ was used in this work.

(1) Calculate the minimum and maximum standard deviations of the regression coefficients, d_{\min} and d_{\max} .

(2) Construct vector $\alpha = [\alpha_1 \ \alpha_2 \ \dots \ \alpha_q]'$ where $\alpha_i = (i - 1)\beta$. Note $\alpha_1 = 0$ and $\alpha_q = 1$.

(3) For $i = 1:q$

i. Set the i th weight matrix \mathbf{W}_i to the $k \times k$ identity matrix.

ii. For $j = 1:k$, set the weight for the j th resolution element w_{ij} in \mathbf{W}_i according to the following equation:

$$w_{ij} = 1 + \frac{(d_j - d_{\min})(1 - \alpha_j)}{d_{\min} - d_{\max}} \quad (\text{A-4})$$

iii. Obtain the i th weighted spectral data matrix \mathbf{A}_i according to the following equation:

$$\mathbf{A}_i = \mathbf{A} \cdot \mathbf{W}_i \quad (\text{A-5})$$

iv. Using corresponding spectra from the i th weighted spectral data matrix, \mathbf{A}_i , build a PLS model with calibration set $\mathcal{S}_{c(\text{cycle})}$, and choose the number of PLS factors from the range of 1 to h_{max} on the basis of the prediction with transfer set \mathcal{S}_s and the F -test described previously. Compute the PRESS value for set \mathcal{S}_s, p_i .

(4) Let $\mathbf{p} = [p_1 \ p_2 \ \dots \ p_q]$. Choose $\mathbf{W}_\alpha = \mathbf{W}_i$ and $\alpha_{\text{opt}} = \alpha_i$ where $p_i = \min(\mathbf{p})$.

Step 3. Update \mathbf{W}_{new} by the following equations:

$$\mathbf{W}_{\text{old}} = \mathbf{W}_{\text{new}} \quad (\text{A-6})$$

$$\mathbf{W}_{\text{new}} = \mathbf{W}_{\text{old}} \mathbf{W}_\alpha \quad (\text{A-7})$$

Step 4. If $\alpha_{\text{opt}} = 1$, go to step 5. Because each iteration updates the old weights by use of eq A-7, $\alpha_{\text{opt}} = 1$ signals that the weights are no longer changing. Otherwise, go to step 1. For the data sets studied, three or fewer iterations through this section of the algorithm were required for the weights to converge.

Step 5. If cycle = 1, go to step 6; otherwise, go to step 7.

Step 6. Implement the initial deletion process. Assume γ to be a user-specified threshold percentage value. A value of $\gamma = 0.05$ was used here.

(1) For $i = 1:n$,

i. Set $\mathcal{S}_c^* = \mathcal{S}_{c(\text{cycle})} - \{s_{c(i)}\}$.

ii. Build a PLS model with set \mathcal{S}_c^* , perform prediction with set \mathcal{S}_s , and use the F -test described previously to determine the optimal number of latent variables from the range of 1 to h_{max} . Calculate the PRESS value for set \mathcal{S}_s, p_i .

(2) Let $\mathbf{p} = [p_1 \ p_2 \ \dots \ p_n]$. Calculate a threshold PRESS value.

$$\text{PRESS}_0 = \min(\mathbf{p}) + (\max(\mathbf{p}) - \min(\mathbf{p}))\gamma \quad (\text{A-8})$$

(3) For $i = 1:n$, if $p_i < \text{PRESS}_0$, $\mathcal{S}_{d(\text{cycle})} = \mathcal{S}_{d(\text{cycle})} \cup \{s_{c(i)}\}$. This procedure allows multiple samples to be deleted in this step if the corresponding PRESS values are within the specified tolerance of $\min(\mathbf{p})$. For the data sets investigated here, however, the tolerance of $\gamma = 0.05$ never allowed more than one sample to be deleted in this step.

(4) $\mathcal{S}_{c(\text{cycle})} = \mathcal{S}_{c(\text{cycle})} - \mathcal{S}_{d(\text{cycle})}$

(5) Go to step 10.

Step 7. Implement the add-one-sample process. Assume $\mathcal{S}_{d(\text{cycle})} = \{s_{d(1)}, s_{d(2)}, \dots, s_{d(l)}\}$, where l is the size of the current set of deleted samples.

(1) Build a PLS model with set $\mathcal{S}_{c(\text{cycle})}$, perform prediction with set \mathcal{S}_s , and optimize the number of latent variables as before.

(2) Calculate PRESS for set $\mathcal{S}_s, \text{PRESS}_{\text{cur}}$.

(3) For $i = 1:l$,

i. Set $\mathcal{S}_c^* = \mathcal{S}_{c(\text{cycle})} \cup \{s_{d(i)}\}$

ii. Build a PLS model based on set \mathcal{S}_c^* , perform prediction with set \mathcal{S}_s , optimize the number of latent variables as before from the range of 1 to h_{max} , and calculate the PRESS value for set \mathcal{S}_s, p_i .

(4) Let $\mathbf{p} = [p_1 \ p_2 \ \dots \ p_l]$. If $\min(\mathbf{p}) < \text{PRESS}_{\text{cur}}$, set $\mathcal{S}_{d(\text{cycle})} = \mathcal{S}_{d(\text{cycle})} - \{s_{d(i)}\}$ and $\mathcal{S}_{c(\text{cycle})} = \mathcal{S}_{c(\text{cycle})} \cup \{s_{d(i)}\}$ where $p_i = \min(\mathbf{p})$.

Step 8. Implement the delete-one-sample process. Assume $\mathcal{S}_{c(\text{cycle})} = \{s_{c(1)}, s_{c(2)}, \dots, s_{c(r)}\}$, where r is the number of samples in the current calibration set.

(1) Build PLS model with set $\mathcal{S}_{c(\text{cycle})}$ and perform prediction with set \mathcal{S}_s .

(2) Calculate PRESS for set $\mathcal{S}_s, \text{PRESS}_{\text{cur}}$. Select the number of latent variables as before from the range of 1 to h_{max} .

(3) For $i = 1:r$,

i. Set $\mathcal{S}_c^* = \mathcal{S}_{c(\text{cycle})} - \{s_{c(i)}\}$.

ii. Build a PLS model based on calibration set \mathcal{S}_c^* , perform prediction with set \mathcal{S}_s , select the number of latent variables as before from the range of 1 to h_{max} , and calculate the PRESS value for set \mathcal{S}_s, p_i .

(4) Let $\mathbf{p} = [p_1 \ p_2 \ \dots \ p_r]$. If $\min(\mathbf{p}) < \text{PRESS}_{\text{cur}}$, set $\mathcal{S}_{d(\text{cycle})} = \mathcal{S}_{d(\text{cycle})} \cup \{s_{c(i)}\}$ and $\mathcal{S}_{c(\text{cycle})} = \mathcal{S}_{c(\text{cycle})} - \{s_{c(i)}\}$ where $p_i = \min(\mathbf{p})$.

Step 9. If $\mathcal{S}_{c(\text{cycle})} = \mathcal{S}_{c(\text{cycle}-1)}$ (i.e., no sample additions or deletions were made), terminate the iterative phase of the algorithm.

Step 10. Set $\mathcal{S}_{c(\text{cycle}+1)} = \mathcal{S}_{c(\text{cycle})}$, $\mathcal{S}_{d(\text{cycle}+1)} = \mathcal{S}_{d(\text{cycle})}$, save the parameters of the PLS model for this cycle, and set cycle = cycle + 1. Go to step 1.

Final Model Selection. Set $\mathbf{p} = [p_1, \dots, p_{\text{cycle}}]$, where p_i is the optimal PRESS value for the transfer set, \mathcal{S}_s , obtained during each cycle. The algorithm structure dictates that the minimum PRESS value, $\min(\mathbf{p})$, will occur in the last cycle.

(1). For $i = (\text{cycle} - 1):1$ by -1 ,

i. If $(p_i / \min(\mathbf{p})) > F(95\%, \nu_m, \nu_m)$, break and use the model corresponding to cycle $i + 1$ for future predictions.

The F value used corresponds to the 95% level and degrees of freedom (ν_m, ν_m) , where ν_m is the degrees of freedom associated with the m samples in the transfer set (i.e., $\nu_m = mq$ if a constant replication factor of q was used in assembling the spectra of the transfer samples).

ACKNOWLEDGMENT

This work was supported by the National Institutes of Health under Grant DK45126. Preliminary results from this work were presented at the meeting of the Federation of Analytical Chemistry and Spectroscopy Societies, Nashville, TN, October 2000, and at the Eastern Analytical Symposium, Atlantic City, NJ, October 2001. Kirsten Kramer is acknowledged for collecting the pure-component spectra of glucose, triacetin, and BSA presented in Figure 1.

Received for review January 10, 2002. Accepted May 31, 2002.

AC020023R