

A Transparent Tool for Seemingly Difficult Calibrations: The Parallel Calibration Method

Clifford H. Spiegelman,^{*,†} Jerome F. Bennett,[†] Marina Vannucci,[†] Michael J. McShane,^{‡,§} and Gerard L. Côté[‡]

Department of Statistics and Biomedical Engineering Program, Texas A&M University, College Station, Texas 77843

A new easy-to-understand calibration method for the analysis of spectral data is developed. The “parallel calibration” method is logically simple and intuitive yet often provides an improvement over more complex standard calibration methods. A description of the algorithm with a technical justification for the parallel algorithm is presented, underscoring the simplicity of the approach. In addition, performance as compared to that of the standard methods of classical least-squares (CLS) and partial least-squares (PLS) regression is studied. Calibrations are carried out on a computer-generated simulation data set as well as two scientific data sets. The results show that the parallel method gives results comparable to or better than those of CLS and PLS methods in terms of mean squared error.

Chemometrics is a cross-disciplinary field that takes methods and concepts from applied mathematics, statistics, and artificial intelligence and applies them to chemistry. Multivariate calibration is one of the areas in analytical chemistry and chemical instrumentation where chemometrics has made a large positive impact. In the modern chemometrics era, partial least squares (PLS) has played a central role because the method filled a void from classical statistical methods.¹

An experiment involving spectral measurements typically produces more data points per spectrum (wavelengths/channels) than calibration measurements (samples). Thus, each spectrum may contain hundreds to thousands of data points, while an instrument calibration may involve tens to hundreds of calibration standards. For example, consider an instrument calibration that uses 10 calibration standards, each containing five known ingredients, where three spectra of 1000 points are collected for each standard. Then, if the spectra were used to predict concentrations, there would be 1000 predictor variables and only 30 observations. Traditional stepwise regression methods fail miserably in these circumstances because of the lack of necessary sample sizes. Chemometricians often use principal component regression (PCR), partial least-squares (PLS) regression, multivariate least-squares, ridge regression, and neural nets (NN) to overcome these

limitations.² Each of these techniques provides a major improvement over stepwise regression and provides useful answers in many practical situations.

The main problem is that those techniques that most frequently are judged the best, partial least squares and neural networks, operate in nontransparent ways. Thus, for over two decades, chemometricians and other scientists have relied upon techniques that are usable but difficult to grasp. Of these, partial least squares is by far the most commonly used method. To properly implement a PLS calibration, an appropriate number of components (“factors”, or “latent variables”) must be chosen, sometimes a nontrivial task and one with which many users have difficulty. The mathematical foundation for PLS is difficult to visualize because a closed-form solution only exists in the case of a single calibration standard.³ Neural network procedures are typically complex as well, requiring a significant learning effort prior to use. Further details on these issues are provided elsewhere.^{4,5}

This paper presents a calibration method, called “parallel calibration”, whose properties appear transparent to anyone who understands multiple regression. A technical justification for the algorithm is presented to outline the utility and simplicity of the approach. Examples involving real and simulated data serve to support the use of the parallel method in multivariate calibration. In particular, three sets of data were used: a simulated data set generated to simulate spectroscopic data; a small-size data set containing near-infrared absorbance spectra of pseudogasoline samples; and a large set containing near-infrared absorbance spectra of aqueous solutions containing components commonly included in cell culture media. The mean squared error results obtained from the parallel method are compared to those from CLS and PLS. Finally, conclusions drawn from the comparisons among the three algorithms in terms of mathematics and performance are presented.

THEORY

Parallel Calibration Method Description. The model and notation assume a measurement device that produces a spectrum

[†] Department of Statistics.

[‡] Biomedical Engineering Program.

[§] Current address: Biomedical Engineering Program/Institute for Micro-manufacturing, Louisiana Tech University.

(1) Brown, S.; Sim, S.; Deshpande, F. *Anal. Chem.* **1996**, *68*, 21R–61R.

(2) Vandeginste, B. G. M.; et al. In *Handbook of Chemometric and Qualimetrics, Part B*; Vandeginste, B. G. M., Rutan, S. C., Eds.; Elsevier: Amsterdam, Lausanne, New York, Oxford, Shannon, Singapore, Tokyo, 1997; Chapter 36.

(3) Spiegelman, C. H.; et al. *Anal. Chem.* **1998**, *70*, 35–44.

(4) DiFoggio, R. *Appl. Spectrosc.* **1995**, *49*, 67–75.

(5) Höskuldsson, A. J. *Chemometrics* **1995**, *9*, 91–124.

(multivariate measurement) y , ($q \times 1$), in response to possibly multivariate concentrations (inputs) c , ($p \times 1$). During the calibration experiment, the instrument responds to n samples with concentrations c_1, \dots, c_n to produce n independent spectra y_1, \dots, y_n . After calibration, a new measurement, z , is taken and the associated concentrations, c_z , are estimated from the model. Each spectrum y has two parts: (1) the response to the concentrations c without error, $f(c)$, which is assumed to be approximately linear for each component in c , and (2) the error, which varies from measurement to measurement and is assumed to have a mean of zero. Similarly z has the corresponding parts $f(c_z)$ and e_z . Let $f(c) = y_c$ denote the first part of y_i and e_i denote the second part. We will assume for pedagogical purposes that the errors are independent and identically distributed with mean 0 and variance, σ^2 . Thus, we have $y_i = y_c + e_i$. A linear relationship (e.g., Beer's law) between the concentrations and spectra is assumed.

Now we present the key model statement: Let the linear operator A , from $R^{p \times q}$ to R^q , be (a_1, \dots, a_n) , where we have $A(y_1, y_2, \dots, y_n) = \sum_{i=1}^n a_i y_i$. This operator describes the linear combination of n calibration spectra y_1, y_2, \dots, y_n . It is assumed that there is a corresponding operator A^* from $R^{n \times p}$ to R^p such that the unique c^* corresponding to $\sum_{i=1}^n a_i y_i$ is $\sum_{i=1}^n a_i c_i$. Note that, except for the fact that q usually does not equal p , $A = A^*$. This assumption is satisfied for full-rank linear models. Also, we require that the rank of (y_1, y_2, \dots, y_n) be n and that $q > n > p$.

The main idea behind the estimation method is as follows. We look for an operator \hat{A} so that $A(y_1, y_2, \dots, y_n) \approx z$ and then estimate c_z by $A(c_1, c_2, \dots, c_n) = \sum_{i=1}^n a_i c_i$. Thus, z , a newly measured spectrum, can be estimated by a linear combination of the calibration spectra. If linear superposition is applicable, then the ingredients of the new sample can be estimated using the same coefficients \hat{A} , because the main sources of the spectral variations in the samples are the ingredients (c).

In this paper, least squares is used to find the operator A . Let \hat{A} be the operator that minimizes $\|(z - A(y_1, y_2, \dots, y_n))\|^2 = (z - A(y_1, y_2, \dots, y_n))(z - A(y_1, y_2, \dots, y_n))$. It can be shown that \hat{A} is the usual least-squares estimator. Then for $Y = (y_1, y_2, \dots, y_n)$, $\hat{A} = (Y Y)^{-1} Y z$. Notice that the concentrations are *not* used to obtain \hat{A} —they are not needed because the information concerning the concentration is contained in the spectra. Thus, having some calibration inputs unknown does not affect the formula for the estimator \hat{A} . The formula given above for estimating \hat{A} is helpful in understanding the method; however, it is not a good way to compute least-squares estimates. We used the QR decomposition to compute our estimates.⁶

The following steps summarize the computation of c_z :

$$\begin{aligned} \text{step 1:} \quad & \text{compute} \quad \hat{A} = (Y Y)^{-1} Y z \\ \text{step 2:} \quad & \text{compute} \quad c_z = \hat{A} c, \quad \text{where} \quad c = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} \quad (1) \end{aligned}$$

The parallel method replaces a bias caused by model misspecification between the c 's and y 's with a bias due to errors in

the predictors. Thus, it is important that the calibration measurements Y have small errors relative to the range of responses to the inputs. (This effect is shown in the simulation experiment described below.) Under our pedagogical model, the estimator \hat{A} has the usual estimated asymptotic covariance matrix $\sigma^2(Y Y)^{-1}$, conditioned upon the calibration data. Unfortunately, the estimator \hat{A} has a bias caused by the error of the spectral measurements. This bias, conditioned on the calibration data, is approximately $[(Y Y)^{-1} - (Y Y + \sigma^2 I)^{-1}] Y f(c)$. If the responses have negligible error, the bias can be ignored.

The parallel method responds differently from all other commonly used calibration methods to errors in the spectra and standards. First, since the entire fitting is done using the spectra and no fitting is done using the concentrations, the parallel method is less sensitive to errors in the measured concentrations. Second, the parallel method is less sensitive to unspecified input compounds. The fitting of the spectra correctly adjusts for all input compounds whether or not they are known. It is this ability that helps it on many occasions to outperform PLS. Of course, the more compounds present (known or unknown), the greater the variance of the concentration estimates for the known inputs.

The parallel method is inefficient under typical independent and identically distributed error models. This is not a practical limitation because, while these models are common in statistics texts, they do not seem appropriate for spectral measurements. For all spectra that we have measured or generated, the responses at the wavelengths are highly correlated. The parallel method is in fact helped by a more typical factor model error structure. (This model is sometimes referred to as a random baseline model.) In our experience, this structure is in many cases more realistic than the independent error structure.

Technical Justification. The reason that the parallel method works exceedingly well with spectroscopic data can be explained qualitatively by considering the following. Spectroscopic data may be represented as $Y_{ij} = c_i \beta_j + \epsilon_i + \delta_{ij}$, where ϵ_i is a common error term for all wavelengths at input concentration levels c_i . The common error is responsible for the error correlation among the spectra at various wavelengths and can be viewed as coming from an unspecified input (with mean zero and a slope that is constant across wavelengths). The terms δ_{ij} also represent errors which are independent for all input levels and wavelengths. We will see that the common error term has a stabilizing effect on the parallel estimate, whereas the independent error has a destabilizing effect.

The parallel estimator has the form $(Y Y)^{-1} Y z \approx (\beta' c \beta + q[\text{diag}(\epsilon_1, \dots, \epsilon_n)] + \text{asymptotically smaller terms})^{-1} Y z$. A common estimator used by statisticians in ill-conditioned problems, as for example spectral calibration, is a ridge estimate of the form $(\beta' c \beta + \text{diag}(\lambda_1, \dots, \lambda_n))^{-1} Y z$, where the λ 's are chosen to minimize the estimated mean squared error.⁷ By inspection, we can see that the common error terms (ϵ) act as ridge parameters and help to stabilize the parallel estimate. Thus, if the common error term dominates the independent error term, the parallel method will work reasonably well. Neither PLS nor the parallel method has distribution theory developed for the common error model.

(6) Seber, G. A. F. In *Multivariate Observations*; Wiley: New York, 1984; Chapter 10, Section 1b.

(7) Hocking, R. R. *Methods and Applications of Linear Models*; Wiley: New York 1996; p 292.

Table 1. Mean Squared Errors for Pseudogasoline Data^a

method	glucose	lactate	ammonia	glutamate	glutamine	ave ^b	corr ^c
PLS	0.41	0.19	0.08	0.26	0.33	0.25	0.99
PAR	0.12	0.01	0.00	0.01	0.12	0.05	1.00
CLS	3.62	0.13	0.09	0.23	2.70	1.36	0.59

^a Leave-one-out cross-validation was used to estimate mean squared errors. ^b The average MSE over the ingredients for each method. ^c The correlation between the predicted concentrations and the true concentration for each method.

However, Yeh found the limiting distribution for PLS under the independent and identically distributed error model.⁸

MATERIALS AND METHODS

Two sets of spectroscopic data were tested with parallel, PLS, and CLS calibration methods. In addition, a simulated data set was used to show comparisons among the methods in some highly controlled situations. This section describes the data and the data analysis procedures.

Description of Data. Near-Infrared Spectra of Gasoline Samples.

The first data set used in this study is available in the Matlab PLS_Toolbox produced by Eigenvector Research, Inc., and is described elsewhere.⁹ Thirty near-infrared absorbance spectra, recorded at 401 equally spaced wavelengths in the range 800–1600 nm, were collected using a scanning NIR spectrometer in the transmission mode. The 30 pseudogasoline samples were prepared with the use of characterized batches of hydrocarbons containing five components, two aromatics (G1, G2) and three saturates (G3, G4, G5).⁹ The concentration ranges for these components were 29.9–50.0, 3.9–15.2, 9.8–29.9, 9.9–29.9, and 0.9–33.0 vol %, respectively.

Near-Infrared Spectra of Five-Component Aqueous Mixtures.

Near-infrared absorbance spectra were recorded for the 2.0–2.5 μm range using a Fourier transform infrared spectrometer. Fifty unique mixtures of glucose, lactate, ammonium chloride, glutamate, and glutamine were prepared from stock solutions consisting of dry chemicals dissolved in pH 7 buffer. Samples were mixed such that no correlation (small R^2) existed between the concentrations of individual components. Total concentration ranges for glucose, lactate, total ammonium nitrogen, glutamate, and glutamine were 1.733–27.733, 0.083–11, 0.133–7.333, 0.583–14.250, and 0.083–11.5 mM, respectively.

Three consecutive single-beam spectra were collected for each sample without removal from the cell holder. Single-beam spectra were ratioed with background spectra of plain buffer solutions, which were collected every four samples. Only the 2.06–2.36 μm spectral range was used because of the excessive noise in the areas outside this range. The data were randomly split into calibration and validation sets consisting of 120 spectra and 27 spectra, respectively, after three spectra were removed because of obvious anomalies. All replicate spectra for a given sample were moved together into the calibration or validation sets. A more complete description of the data can be found elsewhere.¹⁰

Simulations. The simulated data comprised 32 standards with three ingredients whose concentrations were randomly selected (according to a uniform distribution in the range 0%–20%) for each ingredient. Half of the measurements were used for calibration, and half of the measurements were used as a test set. (Similar results were obtained when a full-factorial design was used for the ingredients.) From each spectrum, 250 absorbance points (Y) were calculated using a linear relationship between the ingredients (x) and generated coefficients (β). The coefficients were generated to allow three peaks in the spectra and were highly correlated among ingredients. A data set was created using $Y = c\beta + \text{error}(\text{spectral}) + \text{error}(\text{specific})$, with $\text{error}(\text{spectral}) \sim \text{Normal}(0, 0.3^2)$ being a constant for all wavelengths in a given spectrum and with $\text{error}(\text{specific}) \sim \text{Normal}(0, 0.03)$. For each wavelength and each spectrum, all of the specific errors are independent. Together, these two errors account for the randomness among different samples, the correlation among the spectra at all wavelengths, and the noise which a spectrometer might produce for each wavelength and each spectrum. The correlation among wavelength errors was 0.9901. The estimated final average correlation among wavelengths was 0.95. The correlation structure was chosen to follow what was observed in the second example. The simulation replicated the base experiment 100 times and allowed us to estimate a standard error.

Calibration Procedures. For each set of data described above, the calibration set was used as the input for partial least-squares (PLS), our parallel estimator (PAR), and classical least-squares (CLS) calibration.¹ A prediction mean squared error was calculated for each sample ingredient. For the first data set (gasoline) and the simulation data, a “leave-one-out” cross-validation method was used to compare the three methods of calibration, since an out-of-sample prediction is crucial for parallel calibration. A CVMSE was calculated by

$$\text{CVMSE} = \sum_{i=1}^n (c_i - \hat{c}_i)^2 / n$$

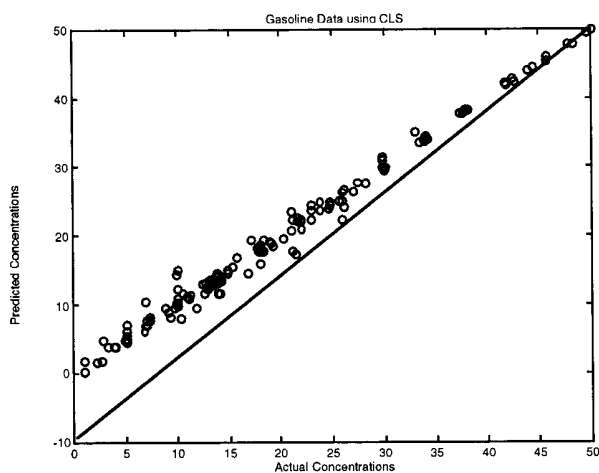
where \hat{c}_i is the predicted chemical concentration for the i th sample in a model developed without the i th sample, c_i is the actual concentration of the i th sample, and n is the total number of samples.

The near-infrared aqueous mixture data set had natural training and test data sets that were reported by McShane et al.¹⁰ For these data, independent test sets were used to estimate the mean squared error. The data were split into separate calibration (80%) and validation (20%) sets for model definition and evaluation,

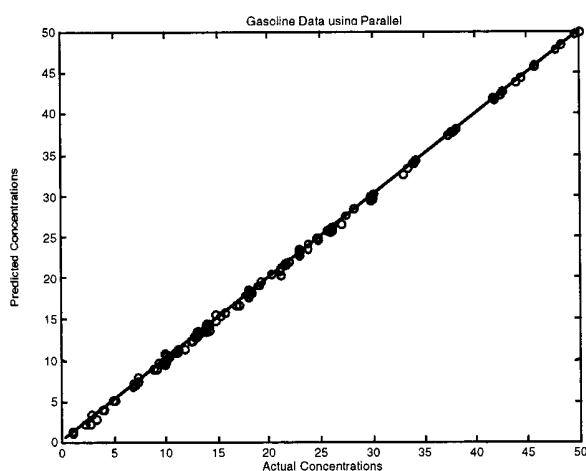
(8) Yeh, C.-H. Solving Classification Problems with the Partial Least Squares Method and Classification Trees. Ph.D. Dissertation, Texas A&M University, 1991.

(9) Wang, Y.; Kowalski, B. R. *Appl. Spectrosc.* **1992**, *46*, 776–771.

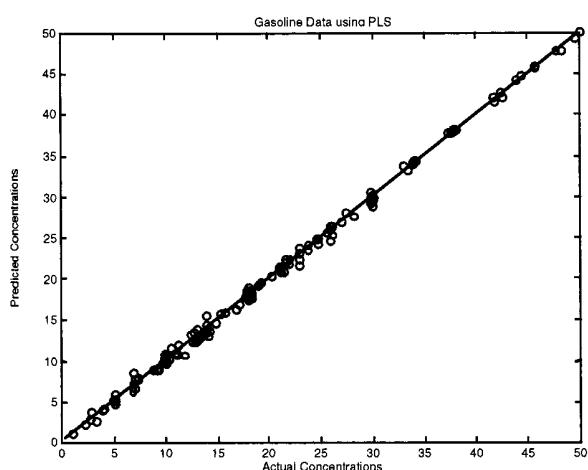
(10) McShane, M. J.; Coté, G. L.; Spiegelman, C. H. *Appl. Spectrosc.* **1998**, *52*, 878–884.



(a)



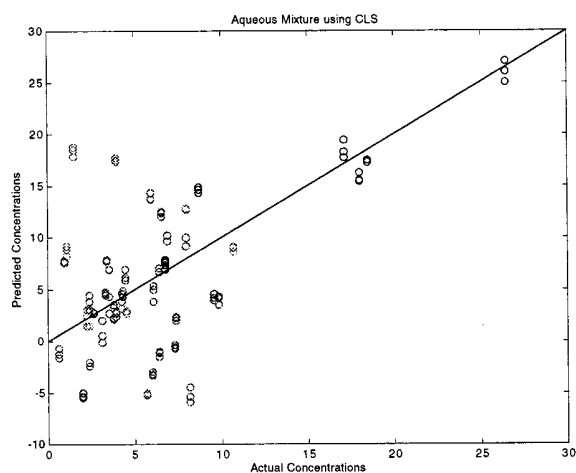
(b)



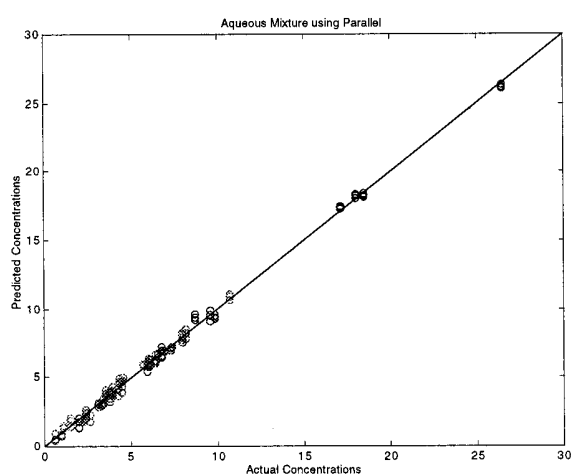
(c)

Figure 1. Correlation plots for gasoline data using (a) CLS, (b) PAR, and (c) PLS.

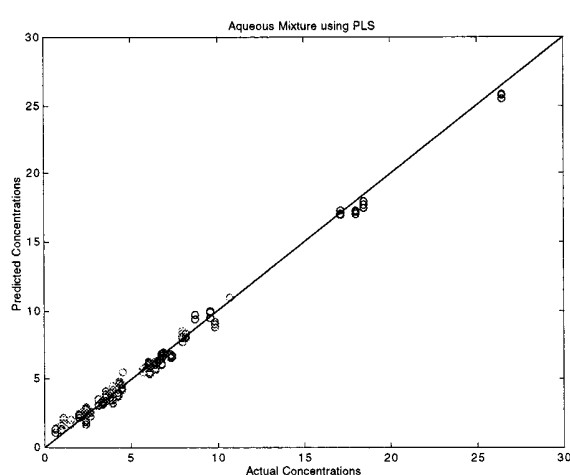
respectively. For the simulation experiment, 50% were used as calibration data and 50% were used as independent test data. Latent variables were chosen for PLS using the first near minimum



(a)



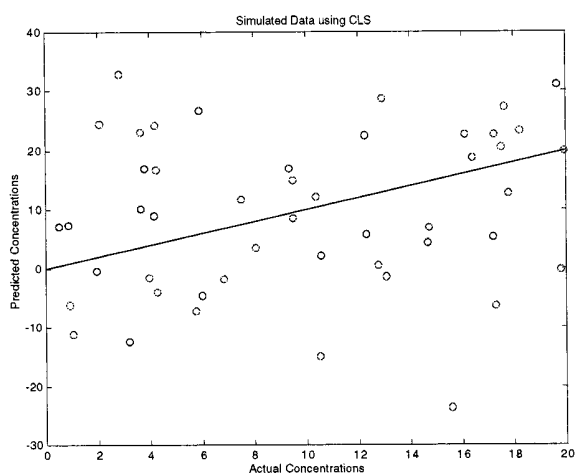
(b)



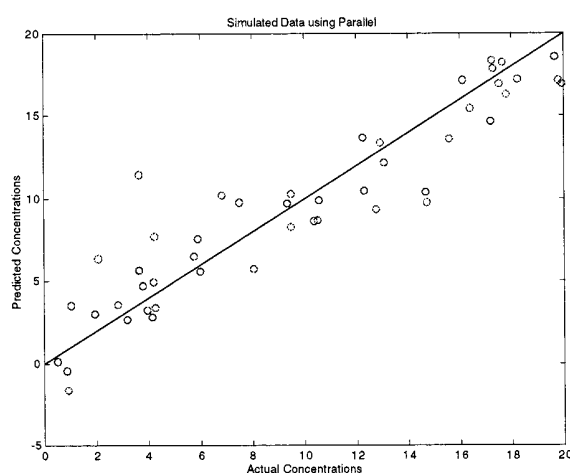
(c)

Figure 2. Correlation plots for aqueous five-component mixture data using (a) CLS, (b) PAR, and (c) PLS.

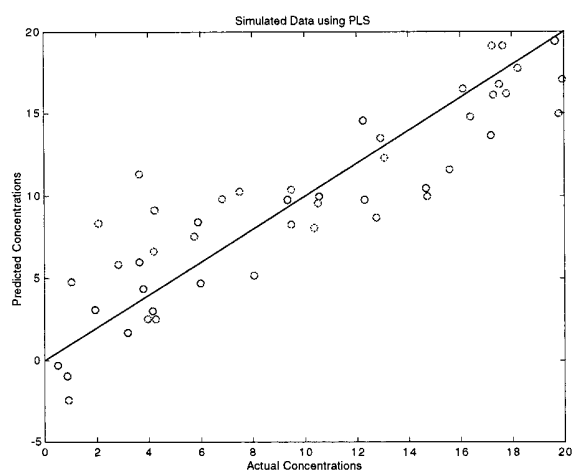
PRESS value obtained for each set of data individually. (With the absolute minimum chosen, PLS would perform poorly on test sets because of too many latent variables.) Five latent variables were



(a)



(b)



(c)

Figure 3. Correlation plots for three-component simulated data using (a) CLS, (b) PAR, and (c) PLS.

determined for the pseudogasoline data, while seven and three factors were used for the aqueous mixture and simulation data sets, respectively.

Table 2. Mean Squared Errors for Aqueous Mixture Data^a

method	G1	G2	G3	G4	G5	ave ^b	corr ^c
PLS	0.39	0.11	0.09	0.34	0.21	0.23	1.00
PAR	0.12	0.07	0.08	0.10	0.07	0.09	1.00
CLS	13.95	39.96	9.78	74.86	25.19	32.75	0.59

^a Mean squared errors were computed for test sets comprising 20% of the total sample sizes. ^b The average MSE over the ingredients for each method. ^c The correlation between the predicted concentrations and the true concentration for each method.

Table 3. Simulation with $\sigma = 0.03$ ^a

method	ingr1	ingr2	ingr3	ave ^b	corr ^c
PLS	3.37	4.97	13.64	7.33	0.91
PAR	2.68	4.17	9.93	5.59	0.92
CLS	119.7	192.8	449.3	254.0	0.34

^a Mean squared errors were computed for test sets comprising 50% of the total sample sizes. ^b The average MSE over the ingredients for each method. ^c The correlation between the predicted concentrations and the true concentration for each method.

RESULTS AND DISCUSSION

The mean squared error results for the gasoline data are presented in Table 1. As shown in the table, on average and for each ingredient, the PAR method outperforms both PLS and CLS. The correlation values near unity also point to the accuracy of both the PLS and parallel methods. The classical calibration method is not at all competitive for these data. Figure 1 contains correlation plots of predicted versus actual concentration values for the gasoline data using CLS, PLS, and PAR. Upon visual inspection of these plots, the CLS predictions are clearly biased, while the PLS and PAR predictions are much more accurate. Both PLS and PAR give predictions grouped tightly around the unit slope line.

The MSE values for the aqueous solutions are shown in Table 2, and the correlation plots are found in Figure 2. Again, the parallel method is better than PLS in terms of MSE for all of the ingredients. It is also clearly superior to PLS when the average mean squared errors are compared. The classical calibration method is more competitive for these data but still lags behind PLS and PAR. Consideration of the correlation plots again shows that CLS provides poor predictions while PLS and PAR are clearly superior. In this case, PAR again shows slight improvement in adherence to the unit slope line.

For the simulated data depicted in Table 3, the parallel method produces better mean squared errors in all cases and thus has a smaller average estimated mean squared error than PLS. With the added noise in this data set, the classical least-squares calibration performs very poorly for all the ingredients. The plot of Figure 3a emphasizes the poor fit. However, the PLS and PAR correlation plots in Figure 3a and Figure 3b, respectively, also contain large amounts of scatter. All the differences in mean squared errors shown in Table 3 are significantly different at the 0.05 level. In fact, the *p* values for all the paired *t* tests were all small (<0.01).

CONCLUSION

We have proposed a transparent and effective alternative for spectral calibration. The parallel method brings a much-needed understandable methodology to the analyst's choice of calibration methods and opens a new window to enhancing spectrometric measurements. The applicability of the parallel calibration procedure has been demonstrated with both "real-life" and simulated spectroscopic data. Results indicate the parallel calibration as a wise choice for well-designed experiments using linear instruments. In the cases tested here, it is comparable or superior to PLS and far superior to CLS. The parallel algorithm is straightforward in operation and does not require the user to

select the proper number of abstract latent variables, unlike PLS.

ACKNOWLEDGMENT

This work was supported in part by grants from the National Science Foundation (DMS-9523878), from the Amarillo National Resource Center for Plutonium, from the NAS (NAG-821), and from the NIEHS (P30-ES09106) and by a Whitaker Foundation fellowship.

Received for review June 1, 1999. Accepted October 26, 1999.

AC990584R