

- (11) Aoki, K.; Kaneko, H. J. *Electroanal. Chem. Interfacial Electrochem.* **1988**, *247*, 17-27.
- (12) Nuwer, M.; Osteryoung, J. *Anal. Chem.* **1989**, *61*, 1954-1959.
- (13) Murphy, M. M.; Stojek, Z.; O'Dea, J. J.; Osteryoung, J. G. *Electrochim. Acta* **1991**, *36*, 1475-1484.
- (14) Sinru, L.; Osteryoung, J.; O'Dea, J. J.; Osteryoung, R. A. *Anal. Chem.* **1988**, *60*, 1135-1141.
- (15) Bond, A. M.; Henderson, T. L. E.; Mann, D. R.; Mann, T. F.; Thormann, W.; Zoski, C. G. *Anal. Chem.* **1988**, *60*, 1878-1882.
- (16) Kuwana, T.; Bubltz, D. E.; Hoh, G. J. *Am. Chem. Soc.* **1960**, *82*, 5811-5817.
- (17) Zara, A. J.; Machado, S. S.; Bulhoes, L. O.; Benedetti, A. V.; Rabockal, T. J. *Electroanal. Chem. Interfacial Electrochem.* **1987**, *221*, 165-174.
- (18) Montenegro, M. I.; Pletcher, D. J. *Electroanal. Chem. Interfacial Electrochem.* **1988**, *200*, 371-374.
- (19) Bond, A. M.; McLennan, E. A.; Stojanovic, R. S.; Thomas, F. G. *Anal. Chem.* **1987**, *59*, 2853-2860.
- (20) Gagne, R. R.; Koval, C. A.; Lisensky, G. C. *Inorg. Chem.* **1980**, *19*, 2854-2855.
- (21) Daschbach, J.; Blackwood, D.; Pons, J. W.; Pons, S. J. *Electroanal. Chem. Interfacial Electrochem.* **1987**, *237*, 269-273.
- (22) Crooks, R. M.; Bard, A. J. J. *Electroanal. Chem. Interfacial Electrochem.* **1988**, *243*, 117-131.
- (23) Sharp, M. *Electrochim. Acta* **1983**, *28*, 301-308.
- (24) Diggle, J. W.; Parker, A. J. *Electrochim. Acta* **1973**, *18*, 975-979.
- (25) Kolthoff, I. M.; Thomas, K. G. J. *Phys. Chem.* **1965**, *69*, 3049-3058.
- (26) O'Dea, J. J.; Lane, T.; Osteryoung, J. J. *Phys. Chem.* **1986**, *90*, 2761-2764.
- (27) Robinson, R. S.; McCurdy, C. W.; McCreery, R. L. *Anal. Chem.* **1982**, *54*, 2356-2361.
- (28) Wu, H. P.; McCreery, R. L. *Anal. Chem.* **1989**, *61*, 2347-2352.

RECEIVED for review May 29, 1991. Accepted September 23, 1991. This work was supported in part by the National Science Foundation under Grant Nos. CHE8521200 and CHE9024846.

## Multivariate Instrument Standardization

Yongdong Wang, David J. Veltkamp, and Bruce R. Kowalski\*

Laboratory for Chemometrics and Center for Process Analytical Chemistry, Department of Chemistry BG-10, University of Washington, Seattle, Washington 98195

Three closely related problems associated with using multivariate calibration methods in spectroscopy have surfaced recently. The first problem involves the desire to transport a calibration model developed on one instrument to a second or even multiple instruments. Differences between the primary and secondary instruments which can occur for a variety of reasons can lead to erroneous results, thereby prohibiting transferring the calibration model and necessitating the transport of the calibration samples. The second problem occurs when instruments change over time (e.g. wavelength shift) for any reason. Again, using a calibration model for analysis when the instrument responses are altered after the time calibration was performed is problematic. The third problem is caused by the variation between samples from different production batches. The calibration model built from one batch might not be applicable to another batch. Using the mathematics of multivariate calibration, four different approaches to solving these two problems have been derived and tested with computer simulation. The four standardization methods proceed by acquiring the spectra of a well-chosen subset of the calibration samples and then either correcting the primary calibration model for use on secondary instruments or correcting the spectra acquired on the secondary instrument to account for the response differences. While standardization does not outperform using the entire calibration set for recalibration, only a 1.2-1.6 times larger error is obtained by standardization in the simulations and a study of a near-infrared data set.

### INTRODUCTION

The use of multivariate calibration of analytical instruments is growing rapidly, primarily due to the development of so-called biased multivariate regression methods (e.g. principal component regression, PCR, and partial least-squares regression, PLS). One example of this growth is the use of

near-infrared spectroscopy (NIR) to predict the octane number of gasoline samples (1). The typical procedure in multivariate calibration is to obtain a series of calibration samples with known properties of interest (e.g. octane numbers) determined by a referee method (e.g. octane engine). The NIR spectra (reflectance or absorbance) are then measured for each sample, and a multivariate regression method is employed to build a mathematical model which allows the calibration of the property of interest as a function of the measured NIR spectral data. One goal of this procedure is to use the multivariate calibration model to predict the property of samples *not* included in the calibration set.

Three problems have been identified by investigators trying to use the above procedure. The first problem occurs when a calibration model (basically, a mathematical equation with a number of parameters determined during the calibration process) developed on one instrument is transported to another instrument. If the second instrument produces instrumental responses that differ from those which would have been obtained on the first instrument, the calibration model will usually produce erroneous results. This problem has necessitated recalibration of the second instrument, which requires that the calibration samples be physically transported to the site of the second instrument. This is a major burden when the samples are numerous, chemically or physically unstable, hazardous, or when several instruments at separate sites are involved. While manufacturers attempt to build their instruments to produce exactly the same responses from the same samples, the fact is that this goal is seldom reached and even small differences can cause problems (2).

A second problem is observed when the instrumental responses measured on a single instrument over a period of time (minutes, hours, or days) change for any reason (e.g. temperature fluctuations, electronic drift, wavelength or detector intensity instability, etc.). If this occurs following a calibration of the instrument, use of the calibration model may produce erroneous results. Finally, a third problem is caused by the differences between samples coming from different production batches. The calibration model built for one batch might not be applicable to another batch because of the differences in sample finish, particle size, density, etc.

\* To whom correspondence should be addressed.

Actually, these three problems are nearly identical although the response differences associated with each may have different sources. They all involve a calibration on a primary instrument and an attempt to use the calibration model on a secondary instrument that produces responses that differ in some way. These different responses may be the result of using the same instrument at a later time, applying the instrument to another type of sample, or using one or more other instruments.

While these problems pose rather severe restrictions on the successful application of multivariate calibration models, they have received very little attention in the literature. In the field of atomic spectroscopy, some of the early work on univariate standardization has been done to eliminate the sample-to-sample difference, using either internal standards or Zeeman effect (3), while most other work in transporting calibration models has been published in the area of NIR analysis as applied to agricultural products. Three publications are notable. Osborne and Fearn (4) investigated the effects of transferring single-wavelength calibration models between nine different instruments for the prediction of protein and moisture in wheat flour using NIR spectroscopy. Single-wavelength bias correction terms for the two calibration equations on each instrument were determined and the long-term stability of the calibration was studied. Later Shenk, Westerhaus, and Templeton (5) published results from a study where a large number of candidate calibration equations were developed on a single instrument for the prediction of several properties related to the forage quality of grasses and then transferred to six other instruments. The "best" equation was adjusted for bias, offset, and wavelength selection on the other instruments and the standard error of prediction (SEP) was compared between the original and the other instruments for a set of 60 common samples. Recently, Mark and Workman (6) published work describing the selection of wavelengths for NIR calibration based on their robustness toward wavelength shifts between instruments.

These methods all involve calibration utilizing a single, or sometimes a few, wavelength(s) and are not generally applicable to multivariate calibration based on full spectral responses. There is, however, a patented method (7) which attempts to correct for full spectral responses through a delicate univariate scheme. Unfortunately, the applicability of this method is limited by its univariate nature. In this paper, four possible approaches based on multivariate statistics are introduced to solve the problems of response variation between instruments when calibration models are transferred. Two of these methods correct the calibration model directly, while the other two correct the response from the secondary instrument to match that which would have been obtained if the sample had been measured on the primary instrument. Response preprocessing and subset selection are used to enhance the stability of standardization and improve the prediction accuracy. All methods allow the full response of the instrument to be utilized without restriction on the number of wavelengths which can be included in the calibration model. In order to control the type and level of difference between the instruments, simulated spectral data are used to test and compare all five standardization schemes. As an example, a standardization approach is applied to the NIR data measured on two different instruments for the determination of four analytes in gasoline samples. While the example used in this paper comes from NIR spectroscopy, the same techniques can also be applied to other spectroscopies, chromatography, FIA, and other analytical methods which generate multivariate responses.

## THEORY

Assume that the response matrix (dimensioned samples by

wavelengths) for a full calibration set,  $\mathbf{R}_1$ , has been measured, and  $\bar{\mathbf{R}}_1$  is a small subset (fewer samples) of  $\mathbf{R}_1$ . The response matrix of this subset is remeasured on a secondary instrument and denoted as  $\bar{\mathbf{R}}_2$ . Through standardization, it is hoped that the calibration information contained in  $\mathbf{R}_1$  could be transferred without measuring the response matrix ( $\mathbf{R}_2$ ) of the full calibration set on the secondary instrument.

**Standardization with the Classical Calibration Model.** Let the concentration matrix (dimensioned samples by analytes) for the full calibration set and standardization subset be  $\mathbf{C}$  and  $\bar{\mathbf{C}}$ , respectively. The linear relationship between the response and concentration is assumed, i.e.

$$\mathbf{R}_1 = \mathbf{C}\mathbf{K}_1 \quad (1)$$

$$\begin{aligned} \mathbf{R}_2 &= \mathbf{C}\mathbf{K}_2 \\ &= \mathbf{C}(\mathbf{K}_1 + \Delta\mathbf{K}) \end{aligned} \quad (2)$$

where  $\mathbf{K}_1$  and  $\mathbf{K}_2$  are matrices of sensitivities on both instruments (with rows being the pure component spectra) and  $\Delta\mathbf{K}$  is the difference matrix between them. The same relationship should be true for the standardization subset

$$\bar{\mathbf{R}}_1 = \bar{\mathbf{C}}\mathbf{K}_1 \quad (3)$$

$$\begin{aligned} \bar{\mathbf{R}}_2 &= \bar{\mathbf{C}}\mathbf{K}_2 \\ &= \bar{\mathbf{C}}(\mathbf{K}_1 + \Delta\mathbf{K}) \end{aligned} \quad (4)$$

Solving eqs 3 and 4 for  $\Delta\mathbf{K}$

$$\Delta\mathbf{K} = \bar{\mathbf{C}}^+(\bar{\mathbf{R}}_2 - \bar{\mathbf{R}}_1)$$

where  $\bar{\mathbf{C}}^+$  is the pseudoinverse of  $\bar{\mathbf{C}}$ . Substituting  $\Delta\mathbf{K}$  into eq 2 and using eq 1,  $\mathbf{R}_2$  is estimated as

$$\hat{\mathbf{R}}_2 = \mathbf{R}_1 + \mathbf{C}\bar{\mathbf{C}}^+(\bar{\mathbf{R}}_2 - \bar{\mathbf{R}}_1) \quad (5)$$

With  $\hat{\mathbf{R}}_2$  and  $\mathbf{C}$ , a new calibration model can be built for prediction on the secondary instrument, using various multivariate calibration methods (8). Two assumptions are implied in this method: the linear relationship should hold on both instruments and the concentrations for all analytes contributing to the response must be known.

**Standardization with the Inverse Calibration Model.** Instead of using the classical calibration model, the inverse calibration model corresponding to eq 1 through eq 4 is used

$$\mathbf{c} = \mathbf{R}_1\mathbf{b}_1 \quad (6)$$

$$\begin{aligned} \mathbf{c} &= \mathbf{R}_2\mathbf{b}_2 \\ &= \mathbf{R}_2(\mathbf{b}_1 + \Delta\mathbf{b}) \end{aligned} \quad (7)$$

and

$$\bar{\mathbf{c}} = \bar{\mathbf{R}}_1\bar{\mathbf{b}}_1 \quad (8)$$

$$\begin{aligned} \bar{\mathbf{c}} &= \bar{\mathbf{R}}_2\bar{\mathbf{b}}_2 \\ &= \bar{\mathbf{R}}_2(\bar{\mathbf{b}}_1 + \Delta\mathbf{b}) \end{aligned} \quad (9)$$

where  $\mathbf{b}_1$  and  $\mathbf{b}_2$  are the two regression vectors for an analyte on both instruments,  $\bar{\mathbf{b}}_1$  and  $\bar{\mathbf{b}}_2$  are the corresponding regression vectors calculated from the subset,  $\mathbf{c}$  and  $\bar{\mathbf{c}}$  represent the concentration vectors (or other properties) for this analyte in full set and subset. Using eq 6 to calculate  $\mathbf{b}_1$  and combining eq 8 with eq 9 to estimate  $\Delta\mathbf{b}$ , a standardized regression vector can be calculated as

$$\begin{aligned} \hat{\mathbf{b}}_2 &= \mathbf{b}_1 + \Delta\mathbf{b} \\ &= \mathbf{b}_1 + (\bar{\mathbf{b}}_2 - \bar{\mathbf{b}}_1) \\ &= \mathbf{R}_1^+\mathbf{c} + (\bar{\mathbf{R}}_2^+ - \bar{\mathbf{R}}_1^+)\bar{\mathbf{c}} \end{aligned} \quad (10)$$

With the standardized regression vector  $\hat{\mathbf{b}}_2$  (an estimate of  $\mathbf{b}_2$ ), it is straightforward to predict concentrations or other properties for samples measured on the secondary instrument. Different from classical model standardization, only the concentration of the analyte of interest needs to be known. Nevertheless, the method can be expanded to work with more

than one analyte at a time by making  $c$  and  $\bar{c}$  matrices and using a matrix of regression vectors.

**Direct Standardization.** The above two methods correct the calibration model so that it can be applied on the secondary instrument. Here another approach called direct standardization is presented in which the spectra measured on the secondary instrument are corrected to match spectra measured on the primary instrument while the calibration model remains unchanged. In direct standardization, response matrices on both instruments are related to each other by a transformation matrix  $F$ , i.e.

$$\bar{R}_1 = \bar{R}_2 F \quad (11)$$

where  $F$  is a square matrix dimensioned wavelengths by wavelengths.

From eq 11, the transformation matrix  $F$  is calculated as

$$F = \bar{R}_2^+ \bar{R}_1 \quad (12)$$

And the response vector of an unknown sample measured on the secondary instrument,  $r_{2,un}^T$ , is standardized to the response vector  $\hat{r}_{1,un}^T$  expected from the primary instrument

$$\hat{r}_{1,un}^T = r_{2,un}^T F \quad (13)$$

With  $\hat{r}_{1,un}^T$ , the model constructed using  $R_1$  and  $C$  (on the primary instrument) can be used for the prediction of unknown concentrations. Notice that no concentration information is used in the standardization step, which makes it possible that general standard materials (e.g., NIST standards) could be used instead of subset samples. This possibility is under active investigation.

It should be noted that the  $F$  matrix is calculated from a relatively small subset and eq 11 is usually underdetermined with respect to estimating  $F$ . To get a determined  $F$  matrix, the raw spectra  $R_1$ ,  $\bar{R}_1$ , and  $\bar{R}_2$  are projected onto the column space of  $R_1$  calculated from singular value decomposition. Then in the standardization and subsequent calibration, the scores obtained by projection are used in place of original response matrices. Since the projection also has a noise-filtering effect, the same projection is also applied to classical and inverse model standardization.

**Piecewise Direct Standardization (PDS).** In the previous formulation of standardization, the number of distinct subset samples must be at least equal to the rank of  $R_1$ , in order to obtain an adequate standardization (see eqs 5, 10, and 12). This could be a difficulty in real applications. Also, it is noticed that in direct standardization, the whole spectrum on the secondary instrument is used to fit each spectral point on the primary instrument (eq 11). For real spectroscopic data, however, spectral variations are often limited to a smaller region. Therefore, each spectral point on the primary instrument would more likely be related to the spectral measurement at nearby wavelengths than the full spectrum on the secondary instrument. On the basis of these considerations, a piecewise standardization method is developed to reconstruct each spectral point on the primary instrument from several measurements in a small window on the secondary instrument. For subset measurement  $r_{1,i}$  at wavelength index  $i$  on the primary instrument, subset measurements on the secondary instrument  $r_{2,i-j}$ ,  $r_{2,i-j+1}$ , ...,  $r_{2,i+k-1}$ , and  $r_{2,i+k}$  at nearby wavelengths from index point  $i-j$  to  $i+k$  are chosen and put into a matrix

$$X_i = [r_{2,i-j}, r_{2,i-j+1}, \dots, r_{2,i+k-1}, r_{2,i+k}] \quad (14)$$

For a nominal shift and linear intensity change from primary to secondary instrument

$$r_{1,i} \propto r_{2,l}$$

with  $i-j \leq l \leq i+k$ . For the case where  $l$  lies between two

nominally denoted index points with a nonlinear intensity change, this equation will not hold. However, a multivariate regression in the form of

$$r_{1,i} = X_i b_i$$

can perform interpolation and provide a reasonable approximation to the nonlinear intensity change. It is the fact that  $\text{rank}(X_i) \leq \text{rank}(R_1)$  which makes it possible to reduce the number of subset samples. The regression vector  $b_i$  can be calculated via various multivariate calibration methods such as principal component regression (PCR) or partial least squares (PLS) and subsequently used to transfer  $r_{2,un}$  piece by piece into the spectrum as if it were measured on the primary instrument. For each end on primary spectra, a window of certain size on secondary spectra cannot be constructed and thus both ends are simply discarded. Another alternative is extrapolation at each end using part of the window.

When compared to the direct standardization method, piecewise standardization is in fact a calculation of the transformation  $F$  (eq 11) by setting most of the off-diagonal elements to zero, i.e.

$$F = \text{diag}(b^T_1, b^T_2, \dots, b^T_i, \dots, b^T_p)$$

where  $p$  is the number of spectral channels included. If there is only a linear intensity change and no wavelength shift (e.g., path length change),  $F$  will be a diagonal matrix; if there are both linear intensity change and +1 unit wavelength shift (e.g., misalignment of monochromator),  $F$  will be a matrix with nonzero elements on the subdiagonal and zeros elsewhere. For the more general case where intensity change is nonlinear and shift varies with wavelength (e.g. peak broadening),  $F$  will be a banded matrix with nonzero elements in a window along the main diagonal and zeros elsewhere.

**Patented Method.** The patent by Shenk et al. provides another way for full spectral standardization. This approach will be briefly summarized here for comparison. For more details, readers are referred to the original literature (7). Basically, the approach does a filtering for every point on the secondary spectra. The filterings take two steps: wavelength index correction followed by spectral intensity correction.

In the first step, the vector  $r_{1,i}$  composed of subset measurements on a primary instrument at wavelength index  $i$  is correlated to the measurement of the same subset on a secondary instrument,  $r_{2,i-j}$ ,  $r_{2,i-j+1}$ , ...,  $r_{2,i+k-1}$ , and  $r_{2,i+k}$ , at nearby wavelengths. The wavelength index  $l$  ( $i-j \leq l \leq i+k$ ) at which the vector  $r_{2,l}$  gives maximum correlation to  $r_{1,i}$  is chosen to be the wavelength index (on the secondary instrument) corresponding to the index  $i$  (on the primary instrument). To get a more precise estimation of  $l$ , a quadratic function of the form  $a + bx + cx^2$  is fitted to all the correlation coefficients being considered and the maxima of the function evaluated to give  $l$ . The wavelength on the secondary instrument,  $\lambda_l$ , is then regressed against the corresponding wavelength on the primary instrument,  $\lambda_i$ , via a simple linear equation

$$\lambda_l = a + b \lambda_i \quad (15)$$

to yield the wavelength correction parameters  $a$  and  $b$ . The  $\lambda_l$  thus estimated (denoted as  $\hat{\lambda}_l$ ) would normally be expected not to coincide with a wavelength index on the secondary instrument, but to lie between two index points. Linear interpolation is therefore used to calculate the spectral measurement at  $\hat{\lambda}_l$ ,  $r_{2,l}$ .

After the above wavelength index correction,  $r_{2,l}$  is fitted to the corresponding  $r_{1,i}$  through a simple linear equation

$$r_{1,i} = c + d r_{2,l} \quad (16)$$

to correct for spectral intensity change between the two in-

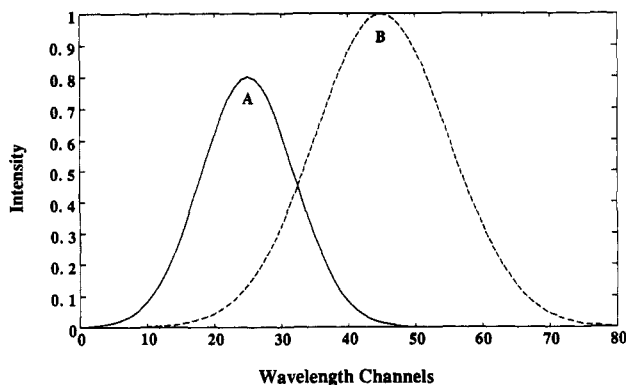


Figure 1. Simulated pure component spectra for analytes A and B.

struments. As for the missing ends where there is no measurement  $r_{2,i}$  corresponding to  $r_{1,i}$ , a multiple linear regression method is used to calculate  $r_{1,i}$  from the four nearest available data points. The standardization file thus obtained will include the wavelength index correction, intensity correction, and missing end handling parameters, which can be used to standardize the spectrum  $r_{2,un}^T$  measured on the secondary instrument into the spectrum as if it were measured on the primary instrument. Due to its univariate nature, this method can only correct for simple wavelength shifting and linear intensity changes.

**Selection of Subset for Standardization.** The subset used for standardization must contain enough information to describe the difference between two instruments. A stepwise procedure is employed here to select the sample with the highest leverage (assuming outliers have already been detected and deleted from the calibration set). The information contained in the sample just selected is then removed from the rest of the samples by a linear transformation so that they are all orthogonal to this selected sample (9). This procedure continues until the desired number of samples have been included in the subset. The algorithm is as follows:

Step A: Calculate the hat matrix of the calibration set

$$H = R_1 R_1^+$$

Step B: Select the sample with the highest leverage (maximum  $h_{ii}$ ), and orthogonalize the row selected against every other row (sample) to obtain a new  $R_1$ , i.e., calculate the linear transformation

$$r'_j = \alpha r_i + \beta r_j \quad \text{for } j \neq i$$

subject to

$$r'_j \cdot r_i = 0$$

Step C: Check if enough samples have been included for the subset: if not, go back to step A using  $R_1$  with the appropriate row removed and the rest transformed; otherwise stop.

## EXPERIMENTAL SECTION

**Simulated Data.** Two partially overlapped pure component spectra  $K$  with 80 channels are generated using Gaussian functions (Figure 1). Then a 10 by 10 full concentration design is constructed (Figure 2), from which 100 mixture spectra  $R_1$  are formed by

$$R_1 = CK$$

in which  $C$  is the concentration matrix with two columns each representing the concentration of one analyte, and  $K$  is the sensitivity matrix with two pure spectra arranged in two rows. To simulate a Lambert-Beer's law deviation, a 30% quadratic term is subtracted from each element in  $R_1$ . After 1% normally distributed random noise is added, cross validation is used to determine the optimal rank for prediction. It is found that the

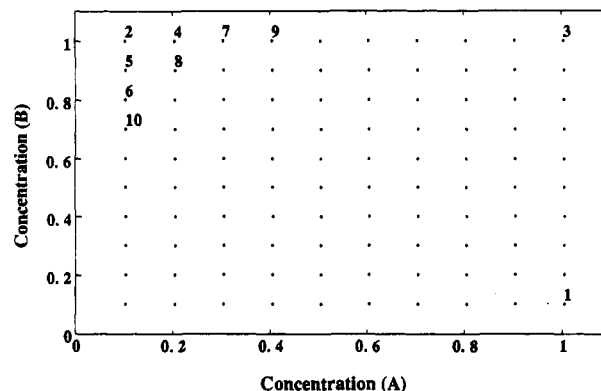


Figure 2. Full experimental design (10 by 10) and the first 10 subset samples selected.

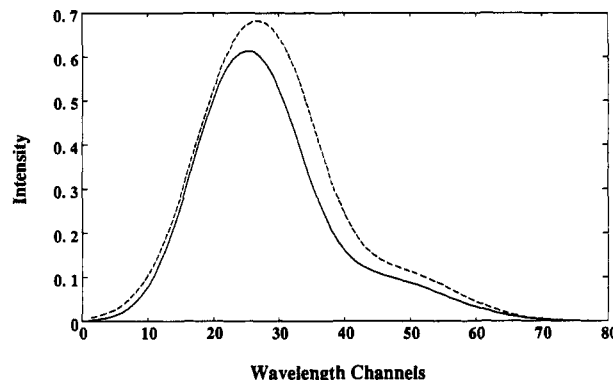


Figure 3. Simulated spectra of the first subset sample measured on the primary (solid) and secondary (dashed) instrument.

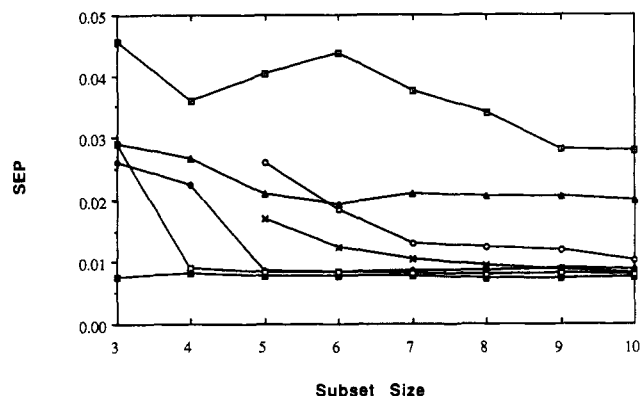
optimal rank for PLS can be chosen as 5 (with SEP for analyte A and B being 0.0055 and 0.0038, respectively). This data set will serve as the large calibration set  $R_1$  (on a primary instrument).

To simulate a linear gain difference on a secondary instrument, a 20% linear increase in the signal is added, followed by a subtraction of a 10% quadratic term to generate a nonlinear response variation. A nonlinear shift is simulated using a quadratic function, the parameters of which are determined such that the shift at both ends is +2 channels and the shift at the extrema of the quadratic function is -2 channels. The spectra are subject to this nonlinear wavelength shift via linear interpolation, to simulate the full calibration set on the secondary instrument,  $R_2$ . With 1% noise added, the cross validation indicates that the optimal rank remains the same 5 with approximately unchanged SEP for both analytes.

The idea of standardization is to select a best subset  $\bar{R}_1$  composed of only a few samples from  $R_1$  and to remeasure these as  $\bar{R}_2$  on the secondary instrument. The first 10 subset samples selected are (arranged in order of decreasing leverage) indicated in Figure 2. The spectra of the first subset sample from both instruments are plotted on top of each other in Figure 3. It is seen that there is significant variation between the two instruments.

Besides standardization, there are several other options available when the problem of response variation occurs. For example, one can make a combined calibration by putting  $R_1$  and  $\bar{R}_2$  together into one set. Another option is to make a new calibration based on  $\bar{R}_2$  only (herein called subset recalibration). The former is found to cause enormous degradation for prediction and thus only the latter is included for comparison with standardization. In the simulation, the subset set size is changed from 3 to 10, in order to study its effect on standardization. To assure the accuracy of SEP used for comparison, each simulation is repeated 250 times and the average SEP calculated.

**NIR Data Set.** A total of 52 gasoline samples were collected and measured on two NIR instruments by two manufacturers. Since the purpose of the study is to demonstrate the application of standardization and not to evaluate different instruments, names of instrument manufacturers are omitted here, and the



**Figure 4.** SEP for analyte A from different procedures: (□) subset recalibration; (◆) classical model standardization; (○) inverse model standardization; (×) direct standardization; (■) PDS; (□) PDS with a quadratic term; (▲) the patented method.

designations A and B are used for the first and second instrument. The measurements were carried out in the transmission mode with a sample cell of 1 cm path length. While instrument A has more data points per nanometer than instrument B and each covers a different spectral region, a common 850–1550 nm wavelength range was chosen for the standardization study. These NIR spectra are utilized in their second derivative form for the PLS prediction of four analytes of interest, labeled as A, B, C, and D.

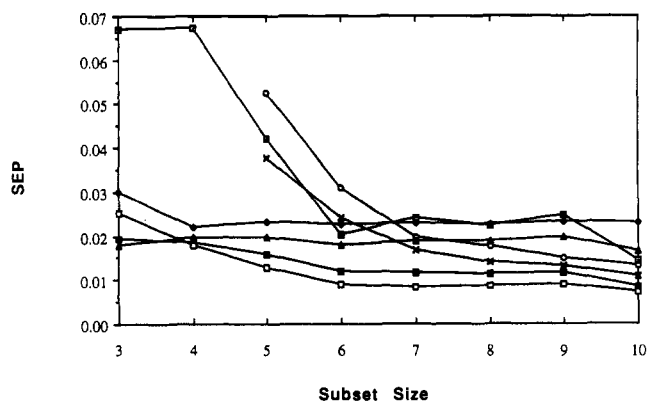
## RESULTS AND DISCUSSION

### Comparison of Different Standardization Schemes.

Five standardization methods mentioned above are tested using the simulation data along with the subset recalibration. Figure 4 illustrates the performance of each method on analyte A when different numbers of subset samples are used. All standardization methods do significantly better than subset recalibration, demonstrating that the calibration information in  $R_1$  is indeed transferred through standardization. As expected, both direct and inverse model standardization give large SEP when the number of subset samples is smaller than 5, the rank of  $R_1$ . In fact, the SEP's are so large that they had to be left out of the figure. As the number of subset samples increases from 5, the prediction accuracy is consistently improved, indicating that both methods have the intrinsic ability to handle the response variation but a larger size of subset samples is preferred for better noise filtering. Between these two methods, direct standardization outperforms inverse model standardization, which is believed to be due to the unfavorable error propagation in the calculation of  $\Delta b$  as the difference of two similar regression vectors,  $\hat{b}_1$  and  $\hat{b}_2$  (eq 10).

Classical model standardization, surprisingly, works better than both methods just discussed. It works even when the subset size is smaller than the rank of  $R_1$ . Remember that the data set simulated is a nonlinear set, where the rank of the response matrix  $R_1$  is larger than the rank of the concentration matrix  $C$ . On the other hand, for the classical model standardization, it is the pseudoinverse of  $C$ , not  $R_1$ , which is calculated and used for the standardization (eq 5). For  $\Delta K$  to be mathematically determined, therefore, the number of subset samples needed should be at least 2 (not 5), which is satisfied here. However, the model will not provide a sufficient fit to standardize the spectra for calibration and prediction. This explains the significant improvement of prediction when the subset size changes from 3 to 5 eventually.

Though being better than subset recalibration, the patented method is not able to attain the accuracy of other standardization approaches. Since this method does not put any ex-



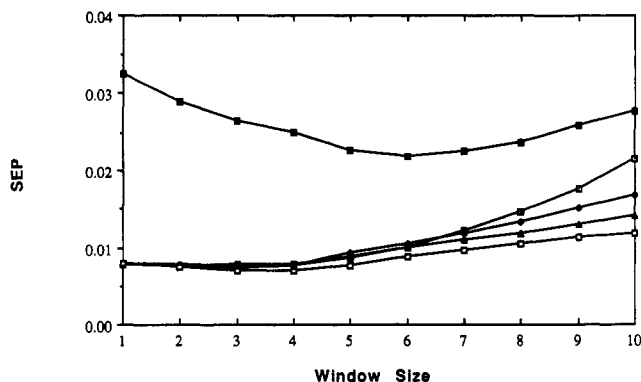
**Figure 5.** SEP for analyte B from different procedures (see Figure 4 for legend).

PLICIT requirement on the subset size, it also works when the number of subset samples is smaller than 5. However, the prediction accuracy is not significantly improved when more subset samples are added, indicating a lack of intrinsic modeling ability (lack of fit). Indeed, a simulation in the absence of noise does show that this method cannot provide a valid mathematical correction for response variations caused by nonlinear shift and nonlinear intensity change.

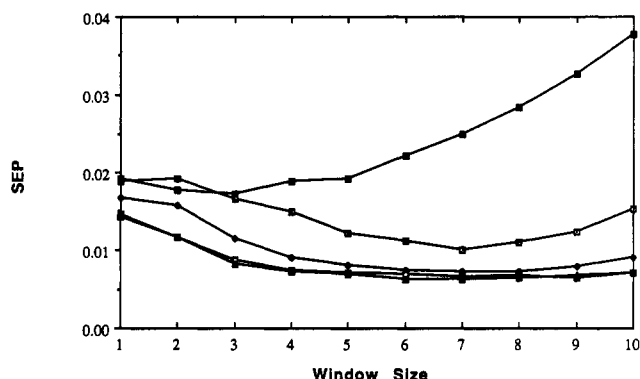
Among all the methods, piecewise direct standardization gives the best results. And more importantly, a very small SEP is reached with only three subset samples. In order to examine its ability to handle nonlinear response variation, a slightly different version of this method was tested. A vector composed of the quadratic terms of subset measurements at wavelength index  $i$  is added to the window matrix  $X_i$  in eq 14. Since the response variation simulated is truly quadratic, it is expected that the inclusion of the quadratic term would provide a better correction. Unfortunately, this is not the case according to Figure 4, which shows that PDS with a quadratic term does slightly worse than without a quadratic term, especially when the subset size is small. The dramatic decrease in SEP when subset size changes from 3 to 4 suggests the possible overfitting. As the subset size increases, the overfitting is reduced and similar prediction accuracy obtained. From the comparison of these two versions, it is clear that, like many multivariate calibration methods, piecewise standardization does not need a nonlinear term to sufficiently account for a nonlinear response change.

A somewhat different picture appears when analyte B is considered (Figure 5). While direct standardization still outperforms inverse model standardization, both methods cannot give better results than subset recalibration until the subset size reaches 7. From its pure spectrum (see Figure 1), it is noticed that this analyte has a higher sensitivity. In the simulation of nonlinear response change, the signal at each channel, after a 20% linear gain, is subtracted by 10% of its quadratic form to generate the response on the secondary instrument. Thus in the spectral region where the signal for analyte B dominates, a larger nonlinearity is involved in the instrumental response variation. The results from Figure 5 indicate that more subset samples are needed for both methods to obtain a sufficient correction for this larger nonlinearity.

As in Figure 4, while direct and inverse model standardization cause enormous errors when subset size is smaller than 5, classical model standardization can still provide a reasonable approximation and outperform subset recalibration. However, when the subset size increases to 6, results similar to subset recalibration are obtained, demonstrating again an intrinsic limitation of its modeling ability for nonlinear data. Unlike in Figure 4, the patented method does better than classical



**Figure 6.** Effect of window size on SEP for analyte A using piecewise direct standardization: ( $\square$ ) subset size = 3; ( $\blacklozenge$ ) subset size = 5; ( $\blacktriangle$ ) subset size = 7; ( $\blacksquare$ ) the patented method with subset size = 3.

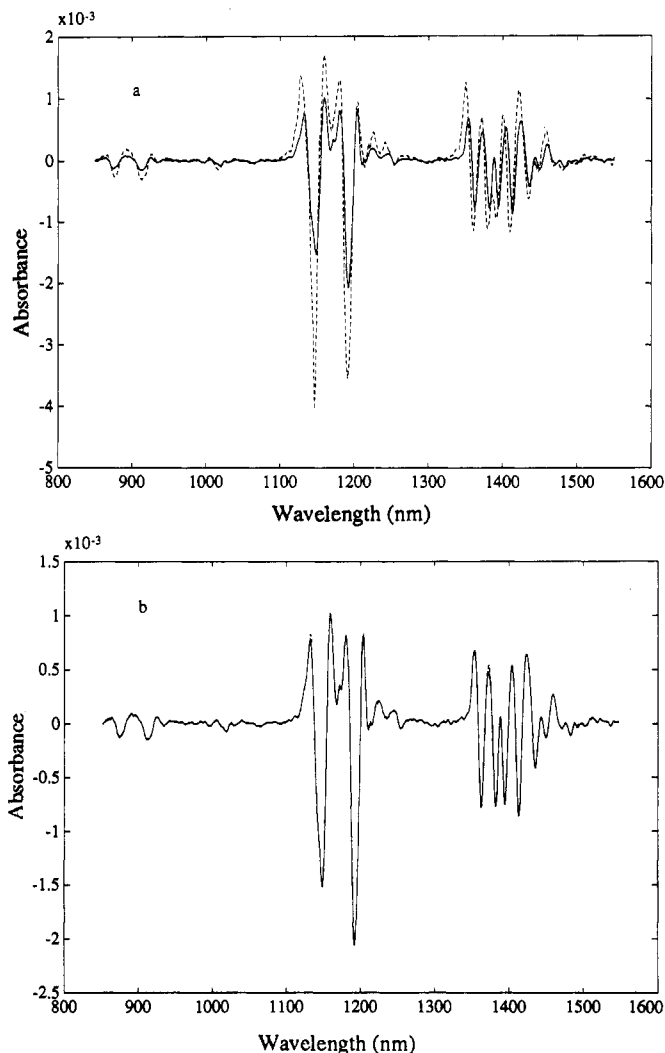


**Figure 7.** Effect of window size on SEP for analyte B using piecewise direct standardization (see Figure 6 for legend).

model standardization, due possibly to a better handling of the wavelength shift, since the nonlinear shift simulated has its extrema at channel 40, near the peak of analyte B. Again, a lack of fit becomes obvious when the subset size increases, which is consistent with the above discussion for analyte A.

Piecewise direct standardization again outperforms all other methods in all but one case where the patented method gives slightly better result with three subset samples, which is related to the window size used and will be discussed later. The standardization with a quadratic term, however, does slightly better than without the quadratic term, except for the case of three subset samples, where overfitting is believed to occur again. As is mentioned above, the nonlinear response change associated with analyte B is larger; therefore an explicit modeling does help somewhat at the risk of overfitting.

**Window Size Optimization.** In the above simulation for piecewise standardization, the shift is known to be between  $-2$  to  $+2$  channels, and the window parameters  $j$  and  $k$  in eq 14 are accordingly chosen as  $j = k = 2$ . This might not be appropriate when nonlinear response variation is involved. The window size is thus changed from  $j = k = 1$  to  $j = k = 10$  with three, five, seven, and nine subset samples to study its effect on standardization. The patented method with three subset samples is also included for comparison. The results are illustrated in Figures 6 (for analyte A) and 7 (for analyte B). From Figure 6, it is seen that PDS works well from  $j = k = 1$  to  $j = k = 4$ , and the increase of subset samples does not have significant influence on standardization, showing that an effective correction has been obtained with this method. If the window size increases to above 4, the inclusion of additional subset samples starts to have a more and more significant effect on standardization. This suggests the occurrence of overfitting when too many nonrelevant channels are



**Figure 8.** Second derivative NIR spectra of a typical sample on instruments A (solid) and B (dashed) before (a) and after (b) standardization.

included, which is the main problem of direct standardization where the full spectral region is included. Although the patented method is slightly improved when the window size changes, the error is still well above that from piecewise standardization. This is in agreement with the above discussion about the lack of intrinsic modeling ability with this method. From Figure 7, it is seen that for analyte B, the optimal window for PDS is  $j = k = 7$ , substantially larger than that for analyte A, confirming that a larger nonlinear response change is associated with this analyte. The nonlinear response change is so large that at least five samples are needed to attain an adequate correction with this method. While the increase of window size improves the piecewise standardization, the patented method is degraded when more wavelength channels are included due again to its limited modeling ability. This explains why slightly better results are obtained when the patented method is used with  $j = k = 2$  and three subset samples. The SEP's for analyte A and B at their optimal window size with five subset samples are 0.0077 and 0.0065, respectively. Compared to the full set cross validation SEP's, 0.0055 and 0.0038, this accuracy is quite satisfactory, since only five subset samples need to be remeasured.

**Application to the NIR Data Set.** From the NIR data set described in the Experimental Section, three samples were selected by the subset selection procedure. The spectra of these three standardization samples are carefully examined (e.g., for peak positions) and compared to those from the other

**Table I. Results from the Cross Validation and Standardization Study of a NIR Data Set Measured on Two Instruments<sup>a</sup>**

analyte	instrument A				instrument B			
	A	B	C	D	A	B	C	D
concn range	2.3–5.2	10.6–15.6	23.6–36.6	37.0–56.3	2.3–5.2	10.6–15.6	23.6–36.6	37.0–56.3
SEP (optimal rank)	0.08 (4)	0.22 (4)	0.51 (3)	0.64 (4)	0.10 (4)	0.27 (3)	0.64 (3)	0.76 (2)
SEP (PDS)	0.10	0.33	0.67	0.83	0.12	0.35	0.98	1.09
SEP (subset recalibration)	0.21	0.39	1.13	1.36	0.20	0.45	1.61	1.77

<sup>a</sup> All analytes are in volume %.

instrument, and a window with  $j$  and  $k$  corresponding to 3 and 4 nm is chosen for piecewise direct standardization.

The standardization is done in two directions, from instrument B to A and vice versa. Figure 8 illustrates spectra of the most unique nonsubset sample, i.e., the fourth sample selected by the subset selection procedure, from both instruments before and after standardization from instrument B to A. The dramatic difference shown in Figure 8a completely rules out the possibility of applying one calibration model to data from the other instrument. Standardization with only three samples corrects the spectrum from the instrument B to be almost the same as that from instrument A (Figure 8b), making possible the application of the calibration model on instrument A to spectra from instrument B.

In order to test the prediction performance of the calibration model for standardized spectra measured on the other instrument, a test set is formed by excluding the three subset samples and SEP calculated. These SEP's are listed in Table I, along with those from cross validation and subset recalibration on each individual instrument. From Table I, it seems that instrument A gives consistently better predictions for the four analytes of interest. The standardization from instrument A to B (listed under instrument A in Table I) is consistently better than from B to A. This might be due to the fact that instrument A has more data points per nanometer, which favors its multivariate calibration and standardization. As is expected, PDS outperforms subset recalibration in all cases. According to the SEP's listed, the error from standardization is about 1.2–1.6 times larger than that from cross validation using the entire set. This standardization performance is in good agreement with the simulation results presented above.

### CONCLUSIONS

The simulation and real data study have demonstrated that the various standardization methods introduced can provide a sound alternative when recalibration using the entire calibration set is not desired. The standardization performance, however, varies significantly from one method to the other. Direct and inverse model standardization seems to work well with a relatively larger subset, with the former being a more favorable choice. Though limited in their modeling capacity, both the patented method and classical model standardization can work with fewer samples than the rank of the full calibration set. Among all the methods tested, piecewise direct

standardization gives the best results and provides a promising method for standardization. As is expected, standardization does not outperform entire set recalibration, with a 1.2–1.6 times larger error for piecewise direct standardization in both simulation and real data study. Considering that only a few subset samples need to be remeasured on an instrument which produces a response of pronounced difference, this prediction error could be accepted as satisfactory in real-world applications. A patent including the four standardization methods described has thus been filed.

Future study will concentrate on improving these methods and on noise propagation through standardization and the ways to limit this propagation. As has been mentioned, the possibility of utilization of some general standards as standardization samples is of practical interest and research is underway in this laboratory to test such a possibility.

### ACKNOWLEDGMENT

We express our appreciation for the many inspiring discussions with members of the Laboratory for Chemometrics. In addition, we thank Eugenio Sanchez and Larry J. Altman, both at Mobil Research and Development Corp., for collecting the NIR data used in this study and for many encouraging discussions. Finally, Y.W. thanks the Mobil Research and Development Corp. for providing an internship opportunity during which part of the NIR data analysis was conducted.

### LITERATURE CITED

- (1) Kelly, J. J.; Barlow, C. H.; Jingui, T. M.; Callis, J. B. *Anal. Chem.* **1989**, *61*, 313.
- (2) Veltkamp, D. J.; Wang, Y.; Kowalski, B. R. Technical Report No. 78; Center for Process Analytical Chemistry, University of Washington: Seattle, 1989.
- (3) Robinson, J. W. *Atomic Spectroscopy*; Marcel Dekker, Inc.: New York and Basel, 1990; pp 128, 236.
- (4) Osborne, B. G.; Fearn, T. J. *Food Technol.* **1983**, *18*, 453–460.
- (5) Shenk, J. S.; Westerhaus, M. O.; Templeton, W. C. *Crop Sci.* **1985**, *25*, 159.
- (6) Mark, H.; Workman, J., Jr. *Spectroscopy* **1988**, *3*, No. 11, 28.
- (7) Shenk, J. S.; Westerhaus, U.S. Patent No. 4866644, Sept 12, 1989.
- (8) Beebe, K. R.; Kowalski, B. R. *Anal. Chem.* **1987**, *59*, 1007A.
- (9) Sharaf, M. A.; Illman, D. L.; Kowalski, B. R. *Chemometrics*; John Wiley & Sons: New York, 1986; p 239.

RECEIVED for review June 1, 1991. Accepted September 9, 1991. This research was supported by the Center for Process Analytical Chemistry, a National Science Foundation Industry/University Cooperative Research Center at the University of Washington.