

# Improvement of Multivariate Calibration through Instrument Standardization

Yongdong Wang, Michael J. Lysaght, and Bruce R. Kowalski\*

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

The successful application of multivariate calibration in solving industrial analytical problems requires a protocol to standardize calibration models between different instruments, e.g., from a central laboratory instrument to a process instrument. In a previous study, several standardization methods based on the measurement of a small set of standards have been developed and compared to the recalibration on individual instruments. It is shown that comparable standard error for prediction can be obtained through standardization, when both instruments involved are of the same quality in terms of prediction performance. Based on the study of a shortwave near-infrared (SW-NIR) jet fuel data set and computer simulation, this paper shows that the standard error for prediction from standardization can be better than that of full set recalibration, when the response on one instrument is standardized into that of a higher quality instrument and the calibration model built on the latter is used for the prediction. This instrument improvement through standardization provides an effective way to avoid on-site recalibration for process analyzers.

The real world application of various multivariate calibration methods, such as principal component regression (PCR) and partial least squares (PLS), requires that a calibration model built on one instrument be transferred to the other, e.g., from the one located in the laboratory to that on an industrial process. In a previous study,<sup>1</sup> four methods have been presented for the standardization of multivariate calibration models, based on the measurement of a small common set of standards, called a transfer set.<sup>2</sup> Of the four methods, piecewise direct standardization (PDS) gave the best standard error for prediction (SEP), which is comparable to that of full set recalibration on individual instruments. PDS proceeds by calculating a transformation matrix using the responses of the transfer set on both instruments, followed by the standardization of the instrument responses (spectra) measured on one instrument into that as if it were measured on the other, such that the calibration model built on the latter could still be utilized for the prediction.

If the instrument to which the response is standardized is of higher quality in terms of prediction performance, as is the case studied in this paper, the prediction from standardization could be better than a separate full set recalibration on the poor instrument due to an effect of information repartition during the standardization and the utilization of a better calibration model. An example from the shortwave near-infrared (SW-NIR) study of a set of jet fuel samples is presented along with results of a computer simulation conducted to investigate this instrument improvement through standardization.

\* To whom correspondence should be addressed.

## THEORY

Since all the methods have been fully described elsewhere,<sup>1</sup> only a brief description for the piecewise direct standardization (PDS) is given here. Assume that the response matrix for a full calibration set,  $\mathbf{R}_1$  (dimensioned samples by wavelengths), has been measured on a higher quality instrument along with the properties of interest,  $\mathbf{C}$  (dimensioned samples by properties); and  $\mathbf{R}_2$  is the response matrix for the transfer set. The response matrix of this transfer set is remeasured on a poor instrument and denoted as  $\mathbf{R}_2$ . In the subsequent description, all vectors are column vectors denoted by a boldface lowercase letter, e.g.,  $\mathbf{r}$ . For row vectors, a superscript T is used, e.g.,  $\mathbf{r}^T$ .

In piecewise direct standardization (PDS), the response matrix of the transfer set on both instruments can be related to each other by

$$\mathbf{R}_1 = \mathbf{R}_2 \mathbf{F} \quad (1)$$

where  $\mathbf{F}$  is a transformation matrix to be determined. In most applications, the response intensity is continuous with respect to response channels. Therefore, when the response on the poor instrument is shifted with respect to the response channel on the higher quality instrument, the corresponding response channels are most likely to be found in a restrictive local region. This leads to a strong structure in the transformation matrix  $\mathbf{F}$  which can be determined in a piecewise manner from eq 1 using various multivariate regression methods. The final form of  $\mathbf{F}$  is a banded diagonal matrix

$$\mathbf{F} = \text{diag}(\mathbf{b}_1^T, \mathbf{b}_2^T, \dots, \mathbf{b}_i^T, \dots, \mathbf{b}_p^T)$$

where  $\mathbf{b}_i^T$  is a vector of transformation coefficients for the  $i$ th response channel and  $p$  is the total number of channels included. The number of elements in  $\mathbf{b}_i^T$  equals the window size used for PDS, which can be determined by examining the responses from both instruments.

Once the transformation matrix  $\mathbf{F}$  is determined, any response measured on the poor instrument for an unknown sample,  $\mathbf{r}_{2,\text{un}}^T$ , can be standardized into the response as if it were measured on the higher quality instrument through

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

which can be used to predict the unknown properties of interest, utilizing the calibration model built on the higher quality instrument.

Since the standardization is from a poor instrument to a higher quality instrument, the transformation matrix  $\mathbf{F}$  calculated from eq 1 will tend to rearrange the information distribution in  $\mathbf{R}_2$  to match that in  $\mathbf{R}_1$  which corresponds to a better calibration model. Thus the prediction from standardization can be better than a separate full set recalibration on the poor instrument. For example, if the difference between the two instruments is in the signal-to-noise ratio, a cross-validation PCR or PLS will choose a fewer number of latent variables (rank) for the poor instrument due to its higher noise, though the same information is contained in the

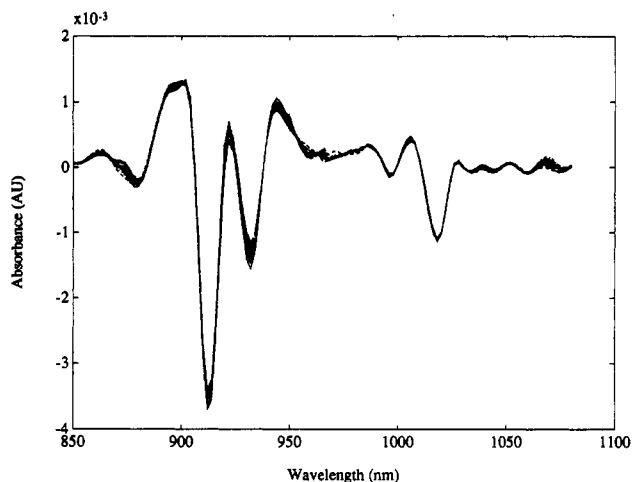


Figure 1. Second derivative spectra of 24 JP-4 fuel samples.

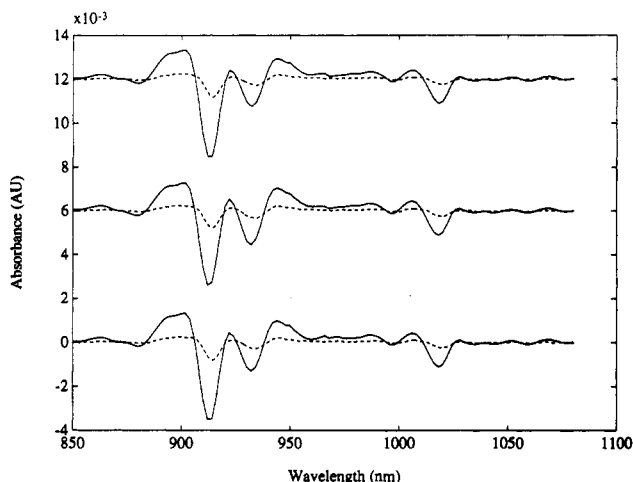


Figure 2. Spectra of three subset samples measured on spectrometer A (solid) and B (dashed). An offset has been added to each spectrum.

responses from both instruments. This is called the information leakage (into the latent variables not included in the calibration model), due to the artificial truncation in PCR or PLS decomposition.<sup>3</sup> A transformation in the form of eq 2 helps to retain as much rank as possible from the poor instrument and therefore to improve the prediction.

### EXPERIMENTAL SECTION

**JP-4 Fuel Samples.** A total of 33 JP-4 samples were circulated among six independent laboratories for the round-robin analysis of volume percent aromatics, volume percent saturates, and freezing point. A complete description of the test procedure has been given elsewhere.<sup>4</sup> Two different spectrometers (denoted as A and B, respectively) were used to collect the NIR transmission spectra, using a 1.00-cm sample cuvette. Spectrometer A is known to have about 10 times higher *S/N* and 2 times higher resolution. According to the calibration study in ref 4, the second-derivative spectra gave the best standard error for prediction (SEP) for all three properties, therefore all the subsequent calculations are based on the second-derivative spectra. Since each spectrometer covers a different spectral region, a common wavelength range from 850 to 1080 nm is chosen for the standardization study. An exploratory study of the 33 spectra in this selected spectral region revealed two somewhat separated clusters composed of 24 and 9 samples, respectively. Since the purpose of this study is not to build a general calibration model for all samples, only the first cluster of 24 samples are used in the standardization study.

Figure 1 shows the second-derivative spectra of 24 samples measured on spectrometer A, from which a subset can be selected as transfer set, using a leverage-based procedure.<sup>1</sup> The spectra of the first three transfer samples as measured on both spectrometers are plotted in Figure 2, which clearly indicates the

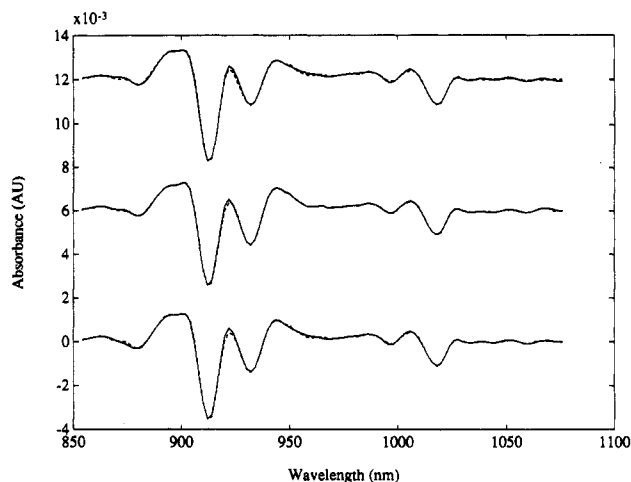


Figure 3. Spectra of three typical samples after standardization using the first three subset samples as a transfer set.

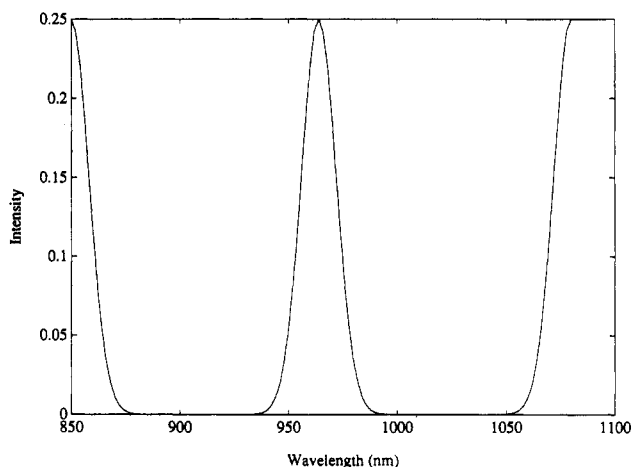
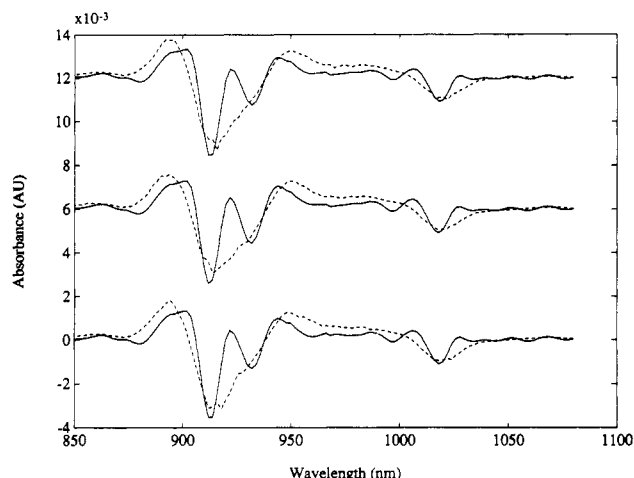


Figure 4. Transfer function used for the first, central, and last wavelength channels.

significant differences between the two spectrometers. In the standardization, the spectra from B are standardized to match the spectra measured on A. Examination of the spectra from both spectrometers decided the window size to be used as 4 nm. Figure 3 shows three typical samples after standardization using the first three subset samples as the transfer set. The standardized samples after excluding those included in the transfer set are used as a test set for the calibration model built on A, and the SEP is calculated as a measure of standardization performance.

**Computer Simulation.** To understand the instrument improvement through standardization, a computer simulation is carried out by making use of the SW-NIR spectra described above. A singular value decomposition (SVD) of the spectra from spectrometer A is done, and only the first nine singular vectors are retained to reconstruct a set of spectra with rank 9, from which three properties of interest are calculated for each sample by utilizing the regression vectors obtained in the above PLS calibration. A data set thus simulated (on spectrometer A) will have characteristics very similar to the above experimental data, yet contain no noise. The degradation of resolution on spectrometer B is simulated by multiplying the spectra from A with a transformation matrix, the columns of which are composed of a set of Gaussian functions. Figure 4 shows three transfer functions for the first, central, and last wavelength channels. The *S/N* difference is simulated by adding 0.5% and 5% normally distributed random noise to the spectra from A and B, respectively. To compare the spectra from both spectrometers, the same three subset samples are plotted in Figure 5, which clearly shows the dramatic resolution and *S/N* difference between the two spectrometers. The same standardization procedure was used for the computer simulation, except that each simulation is repeated 100 times and an average SEP calculated.



**Figure 5.** Spectra of three simulated subset samples measured on spectrometer A (solid) and B (dashed). An offset has been added to each spectrum.

**Table I.** Standard Error for Prediction (SEP) from the JP-4 Fuel Study

		v/v % aromatics	v/v % saturates	freezing point, °C
cross validation (B)		0.94	0.91	1.76
(rank)		(4)	(4)	(3)
subset size (PDS)	3	1.67	1.68	1.74
	5	0.75	0.72	2.07
	7	0.46	0.46	1.53
	9	0.48	0.47	1.44
cross validation (A)		0.44	0.37	1.47
(rank)		(5)	(5)	(6)

## RESULTS AND DISCUSSION

**Jet-4 Fuel Samples.** The PLS cross validation SEPs for spectrometers B and A are listed in the second and last row of Table I. It is seen that the ranks of the calibration models determined by leave-one-out cross validation for spectrometer B are indeed smaller than those for A, consistently for all three properties of interest. It is also seen that the prediction performance for A is (on the average) about 2 times better than B. Table I also shows that as the number of transfer samples increases, PDS standardization gets better. At a subset size of 5, standardization becomes better than the full set cross validation on B. When more subset samples are included in the transfer set, the standardization is further improved, with prediction error closer to the full set cross validation on A. It is noticed that there is some fluctuation in the SEP of the freezing point at the subset size of 5, due possibly to its weak correlation to the NIR spectra.<sup>4</sup>

**Computer Simulation.** The results from computer simulation are listed in Table II. The cross validation on A and B shows that the ranks used for A are larger than or equal to those on the poor instrument B and the prediction on A is consistently better than that on B. When the spectra from B are standardized to match those on A, better prediction than

**Table II.** Standard Error for Prediction (SEP) from the Computer Simulation

		v/v % aromatics	v/v % saturates	freezing point, °C
cross validation (B)		1.40	1.36	1.42
(rank)		(2)	(2)	(6)
subset size (PDS)	3	0.95	1.02	1.88
	5	0.88	0.98	1.71
	7	0.83	0.96	1.62
	9	0.76	0.90	1.27
cross validation (A)		0.13	0.19	0.36
(rank)		(8)	(6)	(6)

a separate recalibration on B is achieved with only three transfer samples for aromatics and saturates. For the prediction of freezing point, however, nine transfer samples are needed for a better prediction, which is again believed to be related to its weak correlation and highly nonlinearity with the NIR spectra.

## CONCLUSIONS

Through the study of the SW-NIR jet fuel data set and computer simulation, it is demonstrated that standardization to a better instrument can provide better prediction than a separate full set recalibration, since the standardization takes advantage of the larger rank on a better calibration model. The implication of this instrument improvement is significant: it is possible to maintain one high-quality instrument (in terms of *S/N*, resolution, etc.) in the central lab, to which many robust and inexpensive field instruments are standardized, using a few transfer samples. Thus the need for on-site recalibration is effectively eliminated, solving one of the most difficult problems in process analytical chemistry. Future work will be devoted to the error propagation in standardization and design of transfer samples.

## ACKNOWLEDGMENT

We thank Ronald Butler and Steve Anderson of the United States Air Force Wright Aeronautical Laboratories for supplying the fuel samples used in the round-robin analysis and also for conducting ASTM reference tests on these samples. We also thank Richard Kiene of ARCO Petroleum Products Co., Jay Hubert of Chevron Research and Technology Co., Rocco DiFoggio and Baron Munchausen of Core Laboratories, Lawrence Altman of Mobil Research and Development Corp., and Walter Hellmuth of Texaco Inc. for their participation in the round-robin analysis. 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.

## REFERENCES

- (1) Wang, Y.; Veltkamp, D. J.; Kowalski, B. R. *Anal. Chem.* **1991**, *63*, 2750-2756.
- (2) Wang, Y.; Kowalski, B. R. *Appl. Spectrosc.*, in press.
- (3) Borgen, O. S.; Kowalski, B. R. Unpublished work, 1990.
- (4) Lysaght, M. J.; Kelly, J. J.; Callis, J. B. *Fuel*, submitted for publication.

RECEIVED for review October 23, 1991. Accepted November 22, 1991.