CALIBRATION TRANSFER ALGORITHM
FOR NIR SPECTROSCOPY AS AN ONLINE ANALYZER

Jeongah Yoon, Hoeil Chung*, Chonghun Han

*Department of Chemical Engineering, Pohang University of Science and Technology, San 31, Hyoja, Pohang, Kyungbuk, 790-784, Korea, Email: chan@postech.ac.kr
NIR Project Team, SK Corporation, 110 Nam-Gu, Kosa-Dong, Ulsan, Korea**

Abstract: The use of near-infrared (NIR) spectroscopy has increased significantly as an online analyzer to monitor the quality of products in chemical plants because of its fast analysis speed and no or little sample treatment required. This paper proposes new calibration transfer algorithm between NIR instruments to update the degraded model or to maintain the performance of NIR permanently. The proposed transfer method performs two steps as follows: the preprocessing based on first-difference (DIFF1) and orthogonal signal correction (OSC) as feature correction and the model modification using piecewise direct standardization. The results have shown good performances for transfer with both high accuracy and simple structure.

Keywords: NIR spectroscopy, calibration transfer, data preprocessing, difference, orthogonal signal correction, multivariate calibration

1. INTRODUCTION

To analyze qualities of products or properties of materials, instruments such as gas chromatography (GC) have been used in laboratories. However, it is difficult to measure the properties real-time because it takes a long time to analyze due to a sampling procedure. Currently, the necessity of real-time analysis has increased gradually and NIR spectroscopy has been widely applied to industrial processes as an on-line analyzer due to its fast assaying time that takes less than 1 minute. The core element in NIR spectroscopy is the multivariate statistical calibration model that can predict the quality of products based on the spectra obtained on-line from NIR spectroscopy. However, the state of NIR analyzer changes as either environmental conditions change or NIR instrument itself deteriorates. Consequently, the performance of the existing calibration model degrades. Therefore, the transfer of a calibration model, which corrects those changes and as result allows the model used irrespective of the differences between instrumental responses of NIR instruments, has become an important research issue. The effective calibration transfer method allows us to save a significant amount of time and cost by making expensive recalibration or remodeling unnecessary.

Various methods for calibration transfer have been developed and published. There are two different ways for automatic calibration transfer. First, if collected sample spectra are corrected properly through preprocessing, the resulting prediction value will be the same as before. On the other way, the calibration model can be adjusted and used to predict sample spectra. By understanding instrumental variation in advance using a reference and adjusting a calibration model, the constant prediction result can be achieved. The above first perspective is on the feature correction, including data preprocessing and variable selection (Savitzky, et. al., 1964; Geladi, et. al., 1985; Wold, et. al., 1998). This step indirectly accompanies calibration transfer to improve the performance of the model itself. This step has the importance of how to select and collect characteristics of data. The above second perspective is on the modification of the constructed model, called standardization method. (Wang, et. al., 1991,1992; Shenk, et. al., 1991; Bouveresse, et. al., 1996a; Forina, et. al., 1995; Stephen, et. al., 1998; Sjöblom, et. al., 1998; Walczak, et. al., 1997). The mathematical relationship between two NIR spectra obtained from each instrument can lead to the model modification. The adequate number of standardization samples for model modification or

calibration transfer methods is very important and should be determined carefully.

This study proposed the approach based upon the feature correction and model modification as mentioned above. The PLS calibration model is preprocessed using two preprocessing methods: first-difference (DIFF1) and orthogonal signal correction (OSC), as feature correction. The model is transferred through piecewise direct standardization (PDS) as model modification. The industrial tatoray NIR data that predicts the benzene concentration are used to validate the proposed algorithm. The spectral data are divided into two groups: the calibration and the test data set. The model and the transfer results are presented for four cases: RAW (no preprocessing), DIFF1, OSC, and DIFF1+OSC. The overall scheme consists of the following steps: a) NIR spectra collection and preprocessing, b) calibration model using PLS from the reference data set, c) on-line monitoring and automatic adjustment of the model by the proposed method. The proposed algorithm guarantees the good prediction performance for NIR spectroscopy which is used as an on-line measuring tool.

2. THEORETICAL BACKGROUND

2.1. Preprocessing techniques

Most of information in NIR spectral data are distributed over a broad wavelength region because the frequency of the NIR spectra is very low. The dimension of the observed data is so large and the data are strongly correlated that the extraction of critical information and removal of the sources of variation is needed to obtain the data with the better quality. The two techniques are employed as preprocessing methods prior to the standardization.

First-difference method The derivative method is generally used for handling the base-line variations of spectral data. Thus, the first-difference (DIFF1), similar to first derivative, is employed to remove baseline variations in NIR spectrum. Assume that the full spectral data (X) is represented as a $M \times N$ matrix where the rows and columns represent samples and wavelengths, respectively. X_i is the column vector for the i th variable. $X_{i,diff}$, i th column vector for the first-difference data, is computed as follows:

$$X_{i,diff} = X_i - X_{i-1} \quad i = 2, 3, 4, \ldots, N \quad (1)$$

Orthogonal signal correction (OSC) method This preprocessing method was developed by Wold et.al(1998) to remove systematic variations existing in an NIR instrument. Predictor variables contain useless noise as well as unrelated information which is not noise, but unrelated to the response variable (Y). The filtering step by OSC preprocessing should not remove information about Y from the spectra X. Therefore, the essential goal of OSC algorithm is to remove the part of information in X, unrelated to Y, without losing important information in X. This is expressed as the following equation:

$$Y = CX + E$$
$$\text{where } X = X_{related} + X_{unrelated} + X_{noise}, \quad (2)$$
$$C \text{ is a coefficient matrix,}$$
$$\text{and } E \text{ is the error of estimation}$$

Noise information, X_{noise}, such as spectra of high frequency or outlier spectral samples, has to be removed before OSC preprocessing. After that, unrelated X-block to Y, $X_{unrelated}$, is removed by the OSC algorithm. Mathematically, this $X_{unrelated}$ removed by OSC is called orthogonal information, which means it has no correlation with Y. One of the key characteristics of OSC algorithm is this orthogonalization step to achieve the minimum covariance between X and Y in the PLS NIPALS (nonlinear iterative partial least squares) algorithm.

2.2. Calibration model and Transfer method

PLS model is developed to identify whether NIR instrument can measure correctly the quality of products as an on-line analyzer. If the performance of the model is degraded, model modification, called calibration transfer, is needed to compensate the difference due to the degradation. The transfer method is accompanied by the piecewise direct standardization (PDS) method (Wang, et.al, 1991) that is useful to correct the spectral differences with linear characteristics of variation.

2.3. Subset selection

The subset that will be used for standardization should be carefully selected so that the subset can represent the population (Bouveresse, et. al., 1996b). Kennard-Stone algorithm (Kennard, et. al., 1969), a method for the selection of representative samples, is used to spread the standardization data over the experimental domain. The stability of the subset samples is also

important to make a model with good transfer results.

2.4. Proposed method

The proposed procedure is shown in Figure 1.

Step 1. Carry out the first-difference preprocessing for original NIR spectra.

Step 2. Apply the OSC preprocessing. The optimal number of OSC components to be removed is determined based on the minimum value of root mean squared error of prediction of test data set (RMSET).

Step 3. Make a PLS calibration model from the calibration set and validate the model with the test set from the master instrument.

Step 4. Select standardization samples in order from the calibration set using Kennard-Stone algorithm for the transfer of the developed PLS model into another instrument.

Step 5. Transfer the PLS model using PDS algorithm. Two parameters for PDS algorithm, subset size and window size, should be decided. . The window size should be an odd number. Then, the recursive loop starts from the number of minimum subset samples and window sizes to the maximum number of samples and window sizes. The minimum number of subset size should be determined carefully to prevent the occurrence of singular problem due to multicollinearity. In our case study, the subset samples range from at least four samples to at most fifteen samples. From this step, all pairs of subset and window size can be investigated and the result is evaluated according to root mean squared error after calibration transfer (RMSE_aft).

Step 6. Analyze and compare the transfer results between the proposed method and the conventional methods.

2.5. Performance measures

To evaluate the performance of the calibration model, the RMSE value is used and defined as follows.

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{n}} \quad (3)$$

y_i is the original i th sample value and $\hat{y}_i$ is the i th value estimated from the model. Next, to show the transfer results of PDS, the RMSE
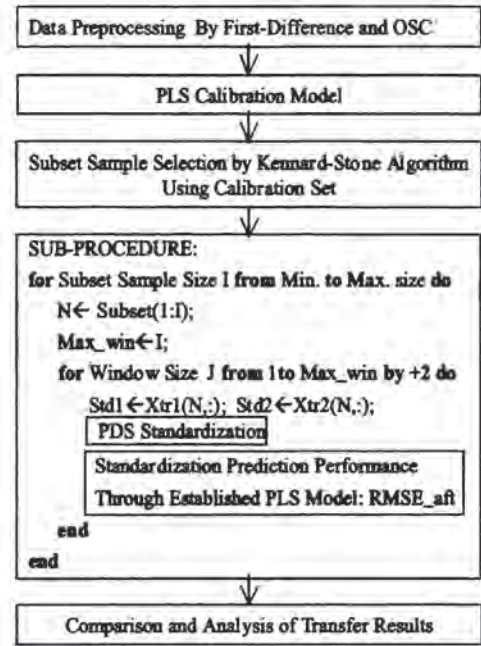

Fig. 1. Proposed calibration transfer algorithm by PDS based upon the first-difference and the OSC preprocessing method.

before calibration transfer (RMSE_bef) and the RMSE after calibration transfer (RMSE_aft) are calculated by the following equation.

$$RMSE\_bef = \sqrt{\frac{\sum_{i=1}^{n}(y_{M,i} - y_{S,i})^2}{n}} \quad (4)$$

$$RMSE\_aft = \sqrt{\frac{\sum_{i=1}^{n}(y_{M,i} - \hat{y}_{S,i})^2}{n}} \quad (5)$$

y_M is the original value from the master instrument, y_S is the value from the slave instrument, and $\hat{y}_S$ is the estimated value from the slave instrument after standardization. RMSE_bef and RMSE_aft represent the total mean variation of two instruments before and after standardization.

3. EXPERIMENTAL

3.1. NIR data set

The tatoray data were collected from SK corporation (Ulsan, Korea) to predict the benzene

concentration. NIRSystems model 6500 spectrometer (Foss-NIRSystems Inc., Silver Spring, MD) equipped with a quartz halogen lamp and PbS detector has been used. There are 700 data points (independent variables) with 2nm data intervals in each NIR specta. Out of initial 700 wavelength points, 427 data points were selected after discarding the wavelengths which are physically meaningless. Total 88 spectral samples were collected from each instrument, randomly mixed, and then divided into 44 calibration sets and 44 test sets. Tatorary NIR spectral data are shown in Figure 2.

3.2. Software

All the algorithms for the analysis of the NIR spectra were implemented in MATLAB 5.2 environment (MathWorks Inc. USA). The PLS_Toolbox version 1.5 (Eigenvector Research Inc. Manson, USA) was used for the PDS and PLS calibrations.

4. RESULTS AND DISCUSSION

4.1. NIR calibration model

With the calibration set, PLS calibration model based on the first-difference and OSC preprocessing methods was constructed to predict the benzene concentration (response variable) of the chemical plant. For the four cases (RAW, no preprocessing; DIFF1, first-difference; OSC; DIFF1+OSC, first-difference and OSC), the model performances and characteristics have been analyzed and compared as shown in Table 1 and Figure 3. Note that for all cases the mean centering was applied to the spectral data (X) and auto scaling to the benzene concentration data (Y) before building the PLS model. The model parameter, namely PLS factor, for the four cases, was determined at the minimal RMSET value which was also used to determine the optimal number of OSC components. In Table 2, the optimal dimension for each PLS model is different although the prediction accuracy are not much different. As shown in Table 2, OSC preprocessing enables the model compact and to have the best accuracy among the four cases. This best RMSET value results from the robustness of the model which is not sensitive to unknown errors within the test data. In addition, the RMSEC of calibration obtained by the first-difference (DIFF1) preprocessing is the best when it has seven PLS factors. It contributes to maintain the high accuracy of model. Therefore, the application of the two combined preprocessing
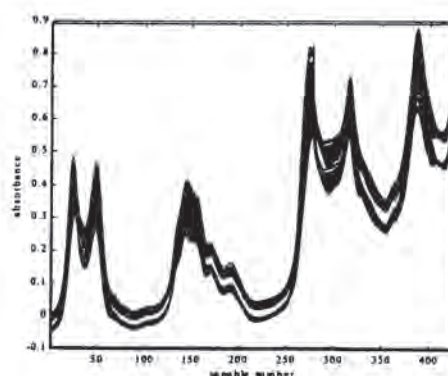

Fig. 2. Spectra of tatorary data (above: instrument 1, below: instrument 2)

Table 1. Comparison of RMSEC values from the calibration set and RMSET values from the test set

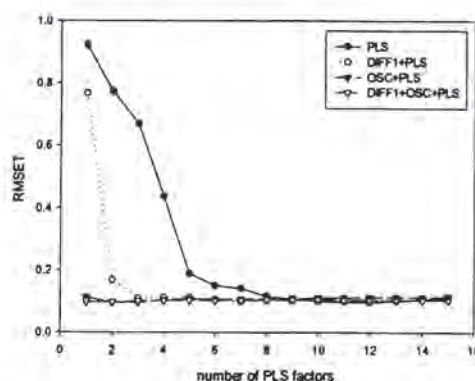| Calibration	Number of latent variable	RMSEC	RMSET
PLS	12	0.0542	0.0966
DIFF1+PLS	7	0.0447	0.0977
OSC+PLS	3	0.0488	0.0953
DIFF1+OSC+PLS	2	0.0457	0.0973

Fig. 3. Change of RMSET values resulted from the model over the number of PLS factors.

methods to make PLS model can offer advantages of simple model structure by OSC and simultaneously high accuracy by DIFF1.

4.2. Instrument standardization

PDS is used as a standardization method to correct linear spectral differences. Table 2 shows that the RMSE_bef (RMS error before standardization) value based on the two combined preprocessing techniques is four times better than the spectral difference before standardization.

Table 2. RMSE_bef values between the master instrument and the slave instrument before standardization

(Rank)	RAW (12)	DIFF1 (7)	OSC (3)	DIFF1 +OSC (2)
RMSE_bef	4.4251	1.0876	3.2108	0.9912

Thus, the RMSE_bef results show that those preprocessing methods can contribute to more robust performance before calibration transfer.

For RMSE_aft (RMSE after standardization) result, the effects of two parameters, window size and subset sample size, on the final transfer performance have been analyzed as shown in Figure 4. The effects of the two parameters may generate a locally bad point region (LBP), where RMSE_aft values have locally large values as shown in the ranges (5 < subset sample size < 9, 3 < window size < 9) in the Figure 4(a)-RAW, 4(b)-DIFF1 and 4(d)-DIFF1+OSC. In case of (a) RAW, the global minimum in the RMSE_aft value is found at the point, (window size = 3, subset sample size = 12). In Figure 4(b) and Figure 4(d), the transfer performance in terms of the RMSE_aft in the LBP region is much better than the case of no preprocessing. However, there is no significant change in the location of LBP region. On the other hand, the case of OSC as shown in Figure 4(c) has no LBP region because the insignificant information in X for Y has been removed during the preprocessing process.

The heuristic guidance for the decision of optimal parameter values is as follows. First, more subset samples produce the better results, regardless of the proper choice of window size. For the case study, the better results were obtained when more than 10 subset samples were used. However, when only small subset samples are available for the transfer due to the difficulties in the collection such as physically or chemically unstable state or hazardous state, the smallest window size, 1, is desirable for transfer. This minimum window size allows us to obtain the tolerable prediction performance even with small number of subset sample and an excellent appromixate transfer results. For some spectral data with linear variations like tatoray data, the best result can obtained at the smallest window size irrespective of the subset sample size. As a result, the best transfer result from model modification for NIR as an on-line analyzer is obtained from the two combined preprocessing methods of DIFF1 and OSC.

5. CONCLUSIONS

This paper proposes the calibration transfer method for NIR spectroscopy as an on-line analyzer which can improve the degraded performance of an existing PLS model. The modification of the model should be done using a transfer model, rather than recalibration from the scratch. When two combined preprocessing methods, DIFF1+OSC, are applied to the model modification method of PDS as an initial data filtering step, the good transfer results are obtained in terms of both model structure and model accuracy, although the model size is more compact. As a result, to maximize the performance of the model transfer using model modification methods, these two combined preprocessing methods are recommended.

REFERENCES

Bouveresse, E., C. Hartmann, and D. L. Massart (1996a). Standardization of Near-Infrared Spectrometric Instruments, *Anal. Chem.*, **68**, 982.

Bouveresse, E. and D. L. Massart (1996b). Standardization of Near-Infrared Spectrometric Instruments: A Review, *Vibrational Spectroscopy*, **11**, 3.

Forina, M., G. Drava, C. Armanino, R. Boggia, S. Lanteri, R. Leardi, P. Conti, R. Giangiacomo, C. Galliena, R. Bigoni, I. Quartari, C. Serra, D. Ferri, O. Leoni, and L. Lazzeri (1995). Transfer of Calibration Function in Near-Infrared Spectroscopy, *Chemom. Intell. Lab. Syst.*, **27**, 189

Geladi, P., D. Macdougall, and H. Martens (1985) Linearization and Scatter-Correction for Near-Inrared Reflectance Spectra of Meat, *Appl. Spectrosc.*, **39**, 3491.

Kennard, R. W. and L.A. Stone (1969). Computer Aided Design of Experiments, *Technometrics*, **11**, 137.

Savitzky, A and M. J. E. Golay (1964). Smoothing and Differentiation of Data by Simplified Least Squares Procedures, *Anal. Chem.*, **36**, 1627.

Shenk, J.S. and M.O. Westerhaus (1991). New Standardization and Calibration Procedures for NIRS Analytical Systems, *Crop Sci.*, **31**,
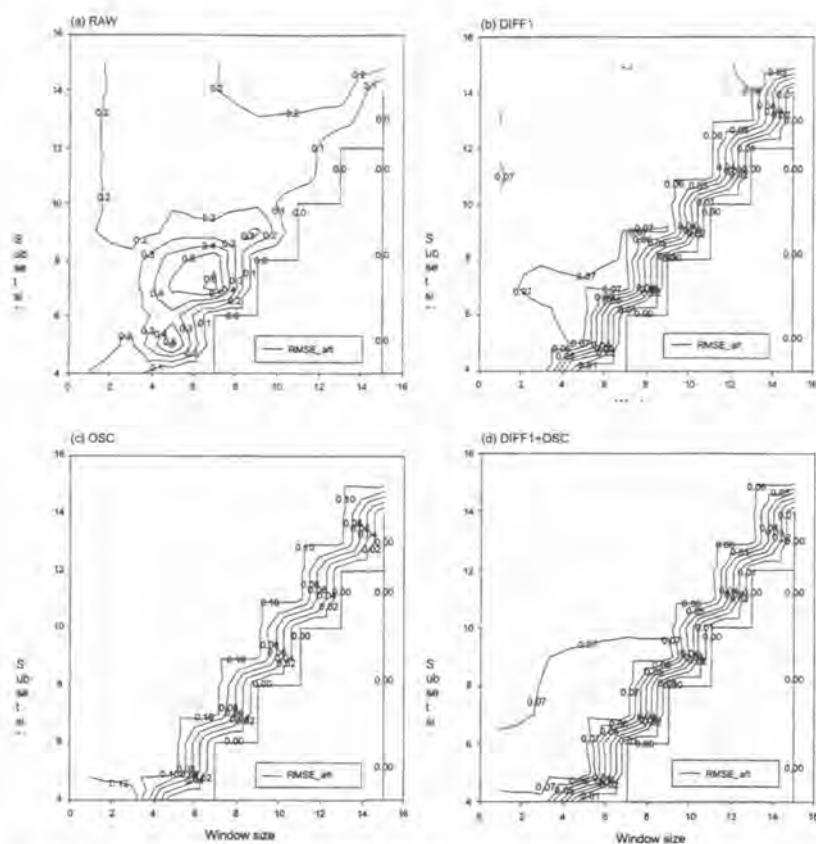
Fig. 6. Contour plot of the RMSE_aft values after standardization according to the change of the number of subset samples and window sizes (a) RAW (12) (b) DIFF1 (7) (c) OSC (3) (d) DIFF1+OSC (2) where the number in parenthesis is rank.

1694.

Sjöblom, J., O. Svensson, M. Josefson, H. Kullberg and S. Wold (1998). An Evaluation of Orthogonal Signal Correction Applied to Calibration Transfer of Near Infrared Spectra, *Chemom. Intell. Lab. Syst.*, **44**, 229.

Stephen, S. and B. Steven (1998). Standardization of Fiber-Optic Probes for Near-Infrared Multivariate Calibration, *Appli. Spectrosc*, **52**, 897.

Walczak, B., E. Bouveresse, and D. L. Massart (1997). Standardization of Near-Infrared Spectra in the Wavelet Domain, *Chemom.*

Intell. Lab. Syst., **36**, 41.

Wang, Y, D. J. Veltkamp, and B. R. Kowalski (1991). Multivariate Instrument Standardization. *Anal. Chem.*, **63**, 2750.

Wang, Y and B.R. Kowalski (1992). Calibration Transfer and Measurement Stability of Near-Infrared Spectrometers, *Appl. Spectrosc.*, **46**, 764.

Wold, S., H. Antti, F. Lindgren and J. Öhman (1998). Orthogonal Signal Correction of Near-Infrared Spectra, *Chemom. Intell. Lab. Syst.*, **44**, 175.