ELSEVIER

Chemometrics and Intelligent Laboratory Systems 51 (2000) 163–173

Chemometrics and intelligent laboratory systems

www.elsevier.com/locate/chemometrics

Rapid determination of FeO content in sinter ores using DRIFT spectra and multivariate calibrations

Kwang-Su Park [a], Hyeseon Lee [a], Chi-Hyuck Jun [a,*], Kwang-Hyun Park [b], Jae-Won Jung [b], Seung-Bin Kim [b,1]

[a] *Laboratory for Applied Statistics, Department of Industrial Engineering, Pohang University of Science and Technology, San 31, Hyoja-dong, Nam-gu, Pohang 790-784, South Korea*

[b] *Laboratory for Vibrational Spectroscopy, Department of Chemistry, Pohang University of Science and Technology, San 31, Hyoja-dong, Nam-gu, Pohang 790-784, South Korea*

Received 23 September 1999; accepted 15 March 2000

Abstract

Using diffuse reflectance infrared Fourier transform spectroscopy (DRIFTS), several multivariate calibration methods are explored for the quantitative determination of FeO content in sinter ores. The multivariate calibrations include ridge regression with variable selection, principal component regression, ridge principal component regression, and partial least square regression with the linear and the nonlinear mapping using neural networks. Spectral data are preprocessed by signal correction and scaling prior to the modeling. Cross validation is employed to obtain the optimal biasing parameter in ridge-related regression and to obtain the optimal number of principal components (or latent variables) in component-related modeling. We consider the possibility of reducing the number of variables involved in models while maintaining the prediction power to propose a final prediction model. For the quantitative determination of FeO content in sinter ores, component related regressions on auto-scaled orthogonal signal correction are suggested as appropriate calibration methods.

Keywords: FeO content; Sinter ores; DRIFT spectra; Multivariate calibration; Signal correction

1. Introduction

Chemometrics [1–3] coupled with infrared spectroscopy (mid-IR and NIR) has been widely used for the rapid determination of the chemical composition of various materials (e.g. gasoline [4,5], palm olein [6], meat [7], enzyme [8,9], resinate [10], vitamins [11], pharmaceutical components [12]). In particular, DRIFTS (diffuse reflectance infrared Fourier transform spectroscopy) has gained popularity for the study of powders, solids, and species adsorbed on solids [13] and in quality control applications due to its minimal sample preparation requirement. It has, however, been reported that spectral measurements are influenced by the parameters such as particle size

* Corresponding author.

E-mail addresses: chjun@postech.ac.kr (C.-H. Jun), sbkim@postech.ac.kr (S.-B. Kim).

[1] Also corresponding author.

0169-7439/00/$ - see front matter
PII: S0169-7439(00)00067-8

[14], sample packing density [15,16], optical properties of the powder [17], and the type of DRIFTS accessory [18]. Nevertheless, quantitative results for a powdered sample can be obtained from DRIFT spectra if proper sample preparations are made. Using DRIFT spectra for a quantitative purpose usually requires the use of a multivariate calibration method to construct a relationship between the DRIFT spectra and the contents or properties of the powdered sample.

Fredericks et al. showed in their works on the application of factor analysis (in this case, PCR) to the FTIR in the mid-IR region that the analyses of coals, minerals, and iron ores for their constituents and properties could be realized using DRIFTS [19–21]. Since coals and iron ores, the major raw material in sinter ores, are obtained from diverse sites, mixtures of them should be controlled carefully to obtain the optimal composition that would lead to a good quality of iron in a blast furnace. In particular, the increase of FeO content in sinter ores may cause low pH in furnace and makes deoxidization difficult, which results to require more fuel materials in furnace. Although FeO in sinter ores is one of the major constituents in making iron and steel, there have been little attempts for the prediction of FeO content in sinter ores using DRIFTS and a multivariate calibration, which have several advantages over conventional analysis methods.

Since FeO itself, designated as wustite, is not stable to exist in sinter ores, various forms of FeO species combined with other constituents such as Fe_3O_4, $CaO \cdot FeO \cdot Fe_2O_3$, $CaO \cdot FeO \cdot SiO_2$, etc., are present in sinter ores instead of wustite species [22]. In particular, the principal form of FeO in sinter ores is usually magnetite, Fe_3O_4 ($FeO \cdot Fe_2O_3$). Two different methods are used to determine the amount of FeO in sinter ores at POSCO (Pohang Iron and Steel): a wet analysis based on the redox reaction between Fe(II) and standard potassium dichromate titrant [23] and an instrumental method based on the magnetic field permeability of FeO in sinter ores [22]. The former, a standard method generally adopted in iron-making industry, is accurate but time-consuming (ca. 4 h) and uses chemicals which may make hazardous effects on operators and environment. The instrumental method is rapid and environment-friendly, but not as accurate as wet analysis due to the permeability difference caused by various forms of wustite combined with other constituents. A rapid, simple, yet accurate determination of FeO content in sinter ores is essential to provide appropriate materials in making iron and steel.

In this paper, we describe the application of DRIFTS for the determination of FeO content in sinter ores that would be an efficient and environment-friendly technique without acid treatment (wet analysis) due to the minimal sample preparation. We investigate the possibility of DRIFTS combined with an improved multivariate analysis that may replace the conventional analysis methods. The purposes of this paper are: (1) to propose the most appropriate prediction model through the comparison of various multivariate methods, and (2) to demonstrate the effect of signal correction methods on the prediction model with DRIFT spectra.

2. Experimental

2.1. Instrumentation

DRIFT spectra are recorded on a Bomem MB100 FTIR spectrometer equipped with a DRIFT attachment (Praying Mantis Model, Harrick, USA), an FTIR purge gas generator (Whatman Model 74-504), and a wide band MCT detector (400 cm^{-1} cutoff). Purge rate is 15 l/min in the sample part and 2.5 l/min in the interferometer part of FTIR spectrometer. Sinter ores are ground by a Disc Grinding Mill operating at 900 RPM (Model TS250, Siebtechnik, Germany).

2.2. Analysis of samples

FeO content (in weight percentage) of sinter ores related to DRIFT spectra are measured by Chemical Testing and Inspection Section (Kwangyang Steel Works, POSCO) using the conventional wet analysis [23]. Among 95 samples collected, 70 observations are used for the calibration and 25 observations for the validation. Fig. 1 shows the distributions of FeO content in the calibration and validation data sets whose means are 8.183, 8.044 and standard deviations are 0.858, 1.032, respectively.

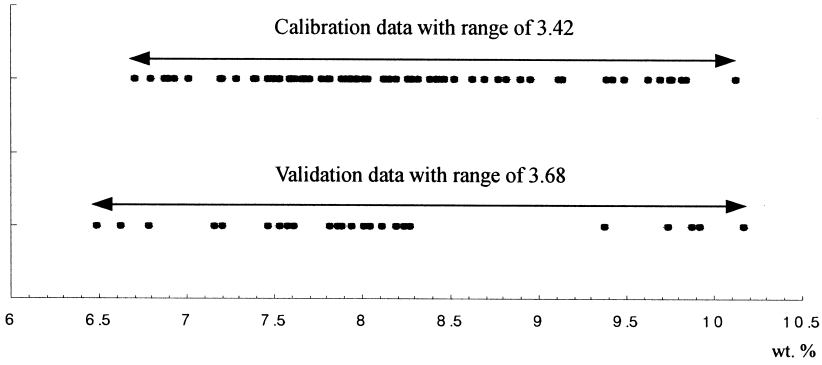


Fig. 1. Distributions of FeO content in calibration and validation data sets.

2.3. Measurement of DRIFT spectra

Because it is difficult to obtain a specific particle size in the powdered samples, sinter ores are milled for 3 min to take a specific size distribution (74–105 μm) and dried at 107 ± 2°C for 2 h to be made as laboratory samples consistent with the ones taken in real application situation. Sample cups are overfilled and lightly tapped on the bench and straight edge drawn across the surface. This sample preparation method renders good reproducibility in the DRIFT spectra. A total of 128 scans of nominal resolution of 4 cm^{-1} are collected to obtain a diffuse reflectance spectrum in the range of 4000–450 cm^{-1}. The diffuse reflectance spectra of sinter ores are obtained with the background spectra of a Harrick reference mirror and the resultant spectra are transformed into DRIFT spectra by Kubelka–Munk function. The procedure is repeated three times and three spectra are averaged to obtain the final spectrum giving a total of 1842 data points. Infrared bands of several iron oxides (e.g., α-Fe_2O_3, γ-Fe_2O_3, Fe_3O_4, FeO) in KI discs appeared in the region 1100–300 cm^{-1} [24]. Overlapped weak and broad bands (1600–1430 cm^{-1}) in Fig. 2 seem to come from overtone (combination) transitions of iron oxides and/or fundamental transitions of other minor constituents in sinter ores. The spectral region used for statistical

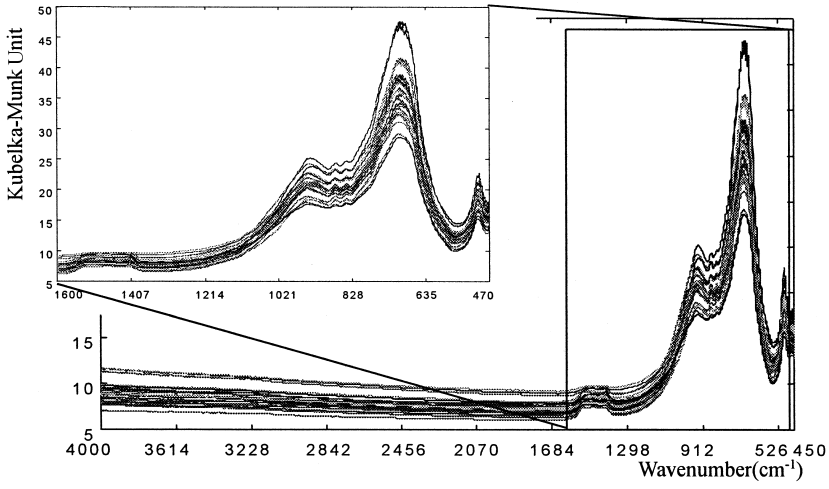


Fig. 2. FTIR spectra of sinter ores.

modeling is 1600–470 cm^{-1} (587 variables), which includes most of the significant reflectance bands. The general appearance of the DRIFT spectra is shown in Fig. 2. The magnified portion of the spectra is used for the modeling hereafter. Several multivariate methods are used to derive a relationship between the DRIFT spectra (**X** matrices) and the FeO content of sinter ores samples ($\boldsymbol{y}$ vector).

2.4. Calibration and validation methods with preprocessing

The calibration and validation steps are carried out with multivariate methods including ridge regression (RR), principal component regression (PCR), ridge combined PCR (RPCR), and partial least squares regression (PLSR or PLS) with the linear and the nonlinear mapping. The nonlinear mapping for inner relation of PLS is implemented by neural networks (NN). Prior to the calibration modeling, we applied data preprocessing methods such as scaling and signal correction. Programs for the preprocessing and modeling are written in MATLAB® (The Language of Technical Computing, Version 5.2.0, the MathWorks, 1998) and the *C* programming language.

3. Results and discussion

3.1. Preprocessing

For powder samples as like in many others, the surface scattering and the specular reflection cause baseline variations and wavelength shifts in measured spectra, and the resulting spectra inevitably contain noises. In multivariate calibration of FTIR data, preprocessing is commonly employed to reduce systematic variations unrelated to the response [25]. Two types of preprocessing, the differentiation and the signal correction (SC) are often used. We focused on the SC techniques in this paper. It is generally recommended that scaling (auto-scaling or mean-centering) is first applied to the original data and the SC is followed.

In essence, the SC should not remove information about $\boldsymbol{y}$ from the spectra (**X**) so that this non-removing part can be stringently formulated mathematically: that is, the information of $\boldsymbol{y}$ should be unrelated to what is removed from **X** by the SC. We employ two methods of SC, multiplicative signal correction (MSC) and orthogonal signal correction (OSC).

3.1.1. Multiplicative signal correction

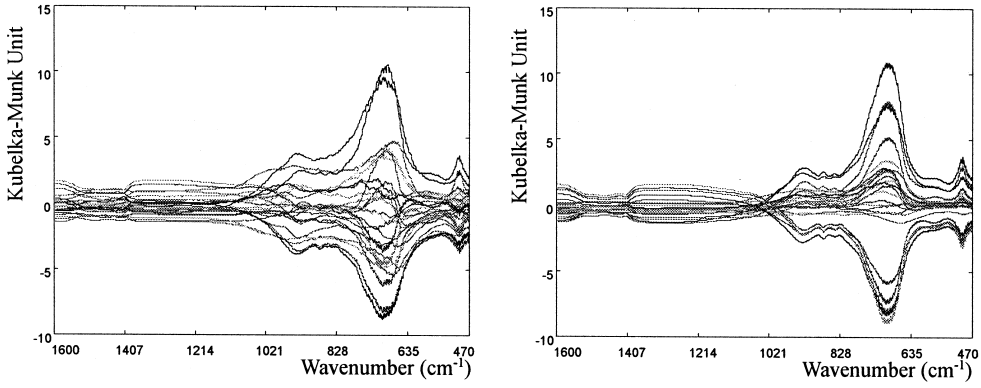
MSC is an additive and multiplicative signal correction method [7]. Each wavelength spectrum, $x_i^T = (x_{i1}, x_{i2}, \ldots, x_{ip})$, is corrected by regressing it against the average spectrum of the calibration set ($m_k = \Sigma x_{ik}/n$) with $x_{ik} = a_i + b_i m_k + \varepsilon_{ik}$, $i = 1, \ldots, n$; $k = 1, \ldots, p$, where n is the number of observations and p is the number of wavelengths. New corrected spectrum of each wavelength is obtained by $x_{i,\text{corrected}}^T = (x_i^T - a_i)/b_i$.

3.1.2. Orthogonal signal correction

As Wold et al. [26] proposed, OSC removes only the part that is unrelated to $\boldsymbol{y}$ from the spectral matrix (**X**). This is ensured by enforcing that the removed part is mathematically orthogonal to $\boldsymbol{y}$, or close to orthogonal as much as possible. Implementing OSC, removing of two orthogonal components is often sufficient. These removals are commonly considered as the base line correction and the multiplicative effect correction [26]. After two OSC components are removed from the mean-centered DRIFT spectra, the resulting spectra look quite different from the original spectra and become smoothed as in Fig. 3.

3.2. Multivariate modeling: calibration and validation

Several algorithms are considered to develop predictive models: RR, PCR, RPCR, and PLS with the linear and the nonlinear inner relation. The model parameters (e.g. the number of latent variables in PLS) are chosen as the one giving the minimum value of predicted residual error sum of squares ($PRESS_m$). It is obtained by $\Sigma_{i=1}^{n_c} (\hat{y}_{(m,i)i} - y_i)^2$, where n_c is the number of the calibration samples and $\hat{y}_{(m,i)i}$ is the ith predicted value by the model using m latent variables except ith observation from calibration set in cross-validation process by leave-one-out method. In calibration step, the predictive ability of each model is evaluated using the root mean square error of



(a) spectra (X) before OSC (b) spectra (X) after OSC

Fig. 3. Raw and OSC-treated DRIFT spectra after mean-centering.

cross validation (RMSECV $= \sqrt{\text{PRESS}/n_c}$) and the root mean squared error of calibration (RMSEC $= \sqrt{\Sigma_{i=1}^{n_c}(\hat{y}_i - y_i)^2/n_c}$). Once the statistical (or mathematical) model is developed using 70 calibration samples, we obtain the root mean squared error of validation (RMSEV $= \sqrt{\Sigma_{i=1}^{n_v}(\hat{y}_i - y_i)^2/n_v}$), where n_v is the number of validation samples) using the rest model-independent 25 samples for the model validation (testing).

3.2.1. Stepwise ridge regression

If we state the multiple linear regression (MLR) model in the form of $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, where $\mathbf{X}$ is spectral data matrix, $\boldsymbol{y}$ is the vector of chemical FeO content, $\boldsymbol{\beta}$ is the vector of regression coefficients, and $\boldsymbol{\varepsilon}$ is the vector of random errors. For spectral data, the number of variables (p) in the matrix $\mathbf{X}$ often exceeds the number of observations (n). In the presence of high degree of collinearity among the x variables, the ordinary least square (OLS) estimates of regression coefficients, $\hat{\beta} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{y}$, may be unstable and lead to a poor prediction.

In the works on response surface related to ridge regression, Horel [27] and Horel and Kennard [28] provided ridge estimator, $\hat{\beta}_r = (\mathbf{X}^T\mathbf{X} + k\mathbf{I})^{-1}\mathbf{X}^T\boldsymbol{y} = (\mathbf{I} + k(\mathbf{X}^T\mathbf{X})^{-1})^{-1}\hat{\beta}$, where k is a biasing parameter to be determined. It is well known from the works of Draper and Nostrand [29] that $\Sigma_i \text{MSE}(\hat{\beta}_{ir}) \leq \Sigma_i \text{MSE}(\hat{\beta}_i)$ is guaranteed for some $k > 0$ where $\text{MSE}(\hat{\beta}_i)$ is the mean square error of the estimate $\hat{\beta}_i$ of parameter β_i.

We use the stepwise method to select a set of variables: either adding or deleting one variable in each step. The significance level for entering a candidate variable, 0.05, is used here. If the partial F-statistic of a candidate variable for exiting is less than the cutoff, that variable is dropped from the model. This is similar to the standard stepwise method except for the consideration of the biasing parameter. In each step after adding or deleting a variable, for a set of selected variables, the biasing parameter k is determined according to the grid search by leave-quarter-out cross-validation because of the computation time. The leave-quarter-out cross-validation means that samples randomly shuffled are split into four and each quarter is repeatedly used for the validation with the model obtained by other three-quarters. The grid algorithm first searches for the best one (which is minimizing RMSECV) among the values $\{0, 1, 2, \ldots, 9\} \times 10^{-1}$. If the selected value is $l \times 10^{-1}$, then do this searching procedure among the values $\{0, 1, 2, \ldots, 9\} \times 10^{-2} + l \times 10^{-1}$. This procedure is repeated until the best value in the precision of 10^{-6} is determined.

After 11 (14, 10) steps under no SC (MSC, OSC), we have obtained the results listed in Table 1, where PEIY denotes the percentage of explained variance (= information) in the $\boldsymbol{y}$ block. In this case, data is auto-scaled before applying SC methods.

The biasing parameters in OSC and no SC are 0, which suggest using MLR. We find that the SC gives better performance with smaller number of variables and that OSC reduces RMSEV of MSC by 12%. In

Table 1
Result of the RR models under auto-scaling

	No SC	MSC	OSC
Number of variables	9	6	6
Selected wavenumbers (cm^{-1})	613, 618, 721, 726, 858, 994, 998, 1201, 1598	562, 601, 622, 721, 994, 1519	572, 574, 618, 721, 998, 1201
k	0	0.000226	0
RMSECV (wt.%)	0.2010	0.2114	0.0548
PEIY (%)	96.86	96.01	99.75
RMSEC (wt.%)	0.1511	0.1704	0.0425
RMSEV (wt.%)	0.2203	0.2066	0.1815

the case of OSC, the RMSEC is much smaller than the RMSEV, which may be suspect overfitting. However, the minimum value of RMSECV is similar to RMSEC and it is also smaller value than RMSEV. It implies that the modeling may not be overfitting but pertinent. This phenomenon is mainly because the OSC removes part of validation **X** that is orthogonal to $\boldsymbol{y}$ of calibration data set. Though both data sets have similar distribution, the OSC result is not same when the OSC weight and loading matrix from the calibration set is applied to the validation set. Signal correction of validation **X** is slightly different from the perfect one using the validation $\boldsymbol{y}$. Though the model of OSC does not overfit, it gives much greater RMSEV than RMSECV. In spite of the loss for the prediction of new model independent data, the OSC tends to show better result than the no SC or the MSC. Similar results are obtained in another modeling methods (e.g. PCR, PLS).

3.2.2. Principal component regression

The MLR model may be expressed as $\boldsymbol{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{X}\mathbf{P}\mathbf{P}^T\boldsymbol{\beta} + \boldsymbol{\varepsilon} = \mathbf{T}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}$, where $\mathbf{T} = \mathbf{X}\mathbf{P}$ (matrix of component scores), $\boldsymbol{\alpha} = \mathbf{P}^T\boldsymbol{\beta}$, and **P** is a loading matrix. As commented by Vigneau et al. [30], q (= rank of the **X** matrix) principal components are computable and the OLS estimator of the coefficient vector $\boldsymbol{\alpha}$ is $\hat{\alpha} = \mathbf{D}^{-1}\mathbf{T}^T\boldsymbol{y}$, where $\mathbf{D} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_q)$, and λ_i's are nonzero eigenvalues with descending order. In PCR using a portion of the **P** matrix, the estimation of $\boldsymbol{\alpha}$ is achieved by retaining only the first c components, namely, replacing **D** of $\hat{\alpha}$ by $\mathbf{D}_{0,c} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_c, 0, \ldots, 0)$.

The PCR models with minimum in PRESS (or RMSECV) are listed in Table 2 under each preprocessing method, where PC denotes the number of principal components in the model and PEIX represents the percentage of explained information in the **X** block.

In the case of PCR, though we select the number of PCs at the minimum PRESS, the models need too large number of PCs. It requires reduction of the number of PCs with some other selection policy. It can be seen that significant difference exists between the mean-centered and the auto-scaled data in OSC whereas similar results are obtained in MSC and no

Table 2
Result of the PCR models

	No SC		MSC		OSC	
	Mean-centered	Auto-scaled	Mean-centered	Auto-scaled	Mean-centered	Auto-scaled
PC	12	11	14	15	13	11
RMSECV (wt.%)	0.2075	0.1990	0.2475	0.2800	0.0266	0.0588
PEIX (%)	99.95	99.91	99.66	99.60	100	99.98
PEIY (%)	96.00	96.20	94.73	93.90	99.94	99.76
RMSEC (wt.%)	0.1703	0.1661	0.1957	0.2119	0.0430	0.0422
RMSEV (wt.%)	0.2064	0.2127	0.2004	0.2060	0.1962	0.1820

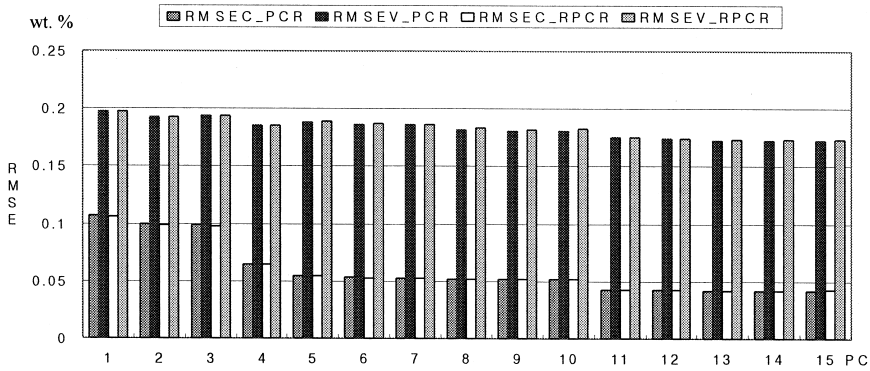


Fig. 4. Comparison of RPCR with PCR according to the number of principal components.

SC. Again, we see that the SC leads to a better performance than without using it.

3.2.3. Ridge combined principal component regression

Based on the fact that RR is implemented by using $\mathbf{D}_{k,q} = \text{diag}\ (\lambda_1 + k,\ \lambda_2 + k,\ \ldots,\ \lambda_q + k)$ instead of $\mathbf{D}$, in RPCR, $\mathbf{D}_{k,c} = \text{diag}\ (\lambda_1 + k,\ \lambda_2 + k,\ \ldots,\ \lambda_c + k,\ 0,\ \ldots,\ 0)$ is used instead of $\mathbf{D}_{0,c}$. Fig. 4 shows RMSEC and RMSEV for PCR and RPCR according to the number of principal components. It indicates that RPCR does not improve PCR in this case.

The result using auto-scaled OSC in Fig. 4 indicates that the effect of biasing parameters is almost negligible. Similar results are obtained in MSC and no SC when applying PCR and RPCR.

3.2.4. PLS with linear mapping

One of the most frequently used multivariate modeling techniques is PLS, which intends to extract most of the information present in the predictor variables related to the response variables. Linear mapping (or ordinary PLS) assumes a linear relationship between the score vectors of the $\mathbf{X}$ matrix and those of the $\boldsymbol{y}$ vector (the inner relation).

A PLS model is constructed by minimizing PRESS and the final selected model is applied to the validation set. Table 3 summarizes the results of the ordinary PLS models, where LV denotes the number of latent variables in the model.

In the case of MSC PLS models with the minimum PRESS show slight overfitting. It indicates that, to obtain pertinent number of LVs, some other determination rules for the MSC should be employed. When seven LVs for mean-centered and nine LVs for auto-scaled data are used for the prediction, the performance of MSC is similar to the OSC. SC gives better performance than no SC in PLS. The auto-scaled OSC gets the minimum error of prediction, which reduces RMSEV of auto-scaled data without

Table 3
Result of the PLS models

	No SC		MSC		OSC	
	Mean-centered	Auto-scaled	Mean-centered	Auto-scaled	Mean-centered	Auto-scaled
LV	5	8	14	15	9	8
RMSECV (wt.%)	0.2130	0.2077	0.2207	0.2234	0.0543	0.0615
PEIX (%)	99.50	99.72	99.57	99.46	99.97	99.93
PEIY (%)	95.22	96.71	99.33	99.49	99.80	99.79
RMSEC (wt.%)	0.1863	0.1546	0.0697	0.0608	0.0382	0.0393
RMSEV (wt.%)	0.2201	0.1934	0.189	0.2325	0.1971	0.1782

Table 4
Learning parameters in neural networks according to the number of latent variables

LV	No. of hidden nodes	Learning rate	Repetitions of learning
3	13	0.002137	17,958
4	40	0.000208	210
5	50	0.000133	377
6	48	0.000116	1447
7	28	0.00017	906
8	40	0.000104	1903
9	31	0.000119	1631
10	70	0.000119	5783
11	71	0.000043	10,226
12	90	0.000309	1578
13	91	0.000282	1860
14	91	0.000262	8070
15	15	0.000148	2118

SC by 8%. Considering it with the PCR results together, the OSC data would be good for our case of component regression.

3.2.5. PLS with nonlinear mapping

One way to deal with non-linearity is to find the inner relationship by nonlinear mapping method such as NN. In nonlinear PLS (NLPLS), when outer relations are defined as follows: $\mathbf{X} = \mathbf{TP} + \mathbf{E}$ and $\boldsymbol{y} = \boldsymbol{Uq} + \boldsymbol{f}$, the (inner) relation between $\boldsymbol{U}$ and $\mathbf{T}$ is often implemented by NN. However, we here modified NN to relate $\mathbf{T}$ and $\boldsymbol{y}$ directly for easier implementation.

A variety of NN models are constructed to obtain the best mapping parameters by changing the number of hidden nodes (37 kinds: {2, 2.5, 3, 3.5, ..., 20} × number of latent variables), and the learning speed of NN [10 kinds: ({3, 6, ..., 30} × number of latent variables × number of hidden nodes)$^{-1}$] for each of the number of latent variables (from 3 to 15) in $\mathbf{T}$. The back-propagation learning method and the sigmoid function as an activation function are used. During the learning, we employ the stopping rule based on the difference of prediction error between before and after every epoch learning less than 10^{-6}. The selected learning parameters according to the number of latent variables are shown in Table 4.

Fig. 5 compares RMSEC and RMSEV of linear (LN) mapping with neural networks (NN) mapping according to the number of latent variables by PLS algorithm.

It can be seen in Fig. 5 that there is little difference in RMSE between linear and nonlinear PLS under auto-scaled data with no SC. It means that nonlinearity between $\mathbf{X}$ (or $\mathbf{T}$) and $\boldsymbol{y}$ is not evident in this case, that is, linear mapping between them may be sufficient for the calibration.

3.2.6. Reduced models and their performance

As shown in Tables 2 and 3, the number of components involved in previous models is quite large. The model with larger number of components may cause over-fitting and give poor prediction. Since OSC generally shows good performance, we will

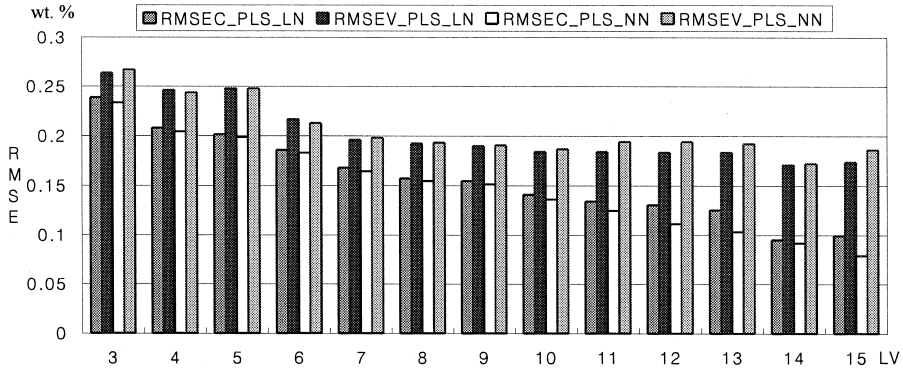


Fig. 5. Comparison of NLPLS with linear PLS according to the number of latent variables.

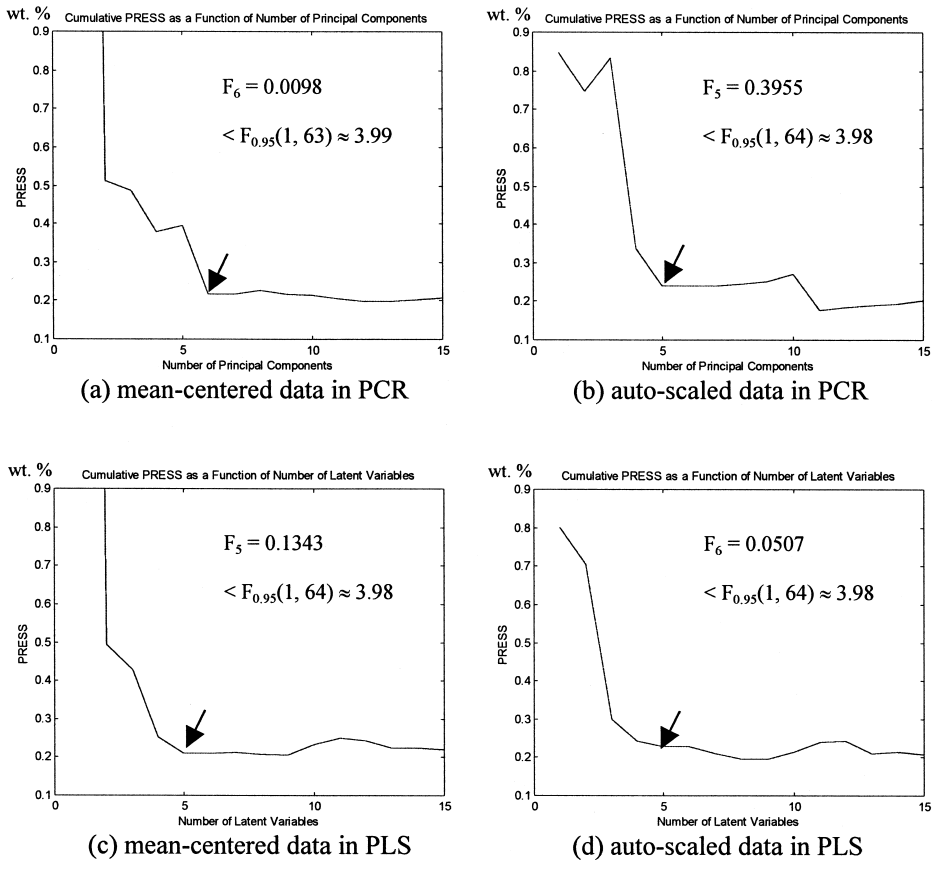


Fig. 6. PRESS of PCR ((a) and (b)) and PLS ((c) and (d)) with OSC.

investigate the possibility of reducing the number components in PCR and PLS and the number of variables in RR under OSC.

The PRESS as a function of the number of components through the cross-validation is shown in Fig. 6. From this plot, we can see the possibility of reduc-

Table 5
Result of the reduced PCR and PLS models with OSC

	PCR		PLS	
	Mean-centered	Auto-scaled	Mean-centered	Auto-scaled
PC (or LV)	6	5	5	5
RMSECV (wt.%)	0.0280	0.0682	0.0548	0.0571
PEIX (%)	99.98	99.84	99.88	99.83
PEIY (%)	99.91	99.6	99.69	99.64
RMSEC (wt.%)	0.0508	0.0541	0.0476	0.0510
RMSEV (wt.%)	0.1970	0.1958	0.1983	0.1927

ing the number of components in PCR and latent variables in PLS while maintaining the PRESS at the same level. Since an F-test can determine whether two PRESS values are significantly different, we need only compare those models having fewer PC or LV (m) than the minimum PRESS model. After the statistic $F_m = (|\mathrm{PRESS}_m - \mathrm{PRESS}_{m+1}|(n - m - 1))/(\mathrm{PRESS}_m)$ is compared against tabulated values $F_{0.95}\ (1,\ n - m - 1)$, we select the number of components (or latent variables) for each model as minimum m satisfying $F_m < F_{0.95}\ (1,\ n - m - 1)$. The selected number is marked by arrow and F-statistic and tabulated F-value are shown in the Fig. 6.

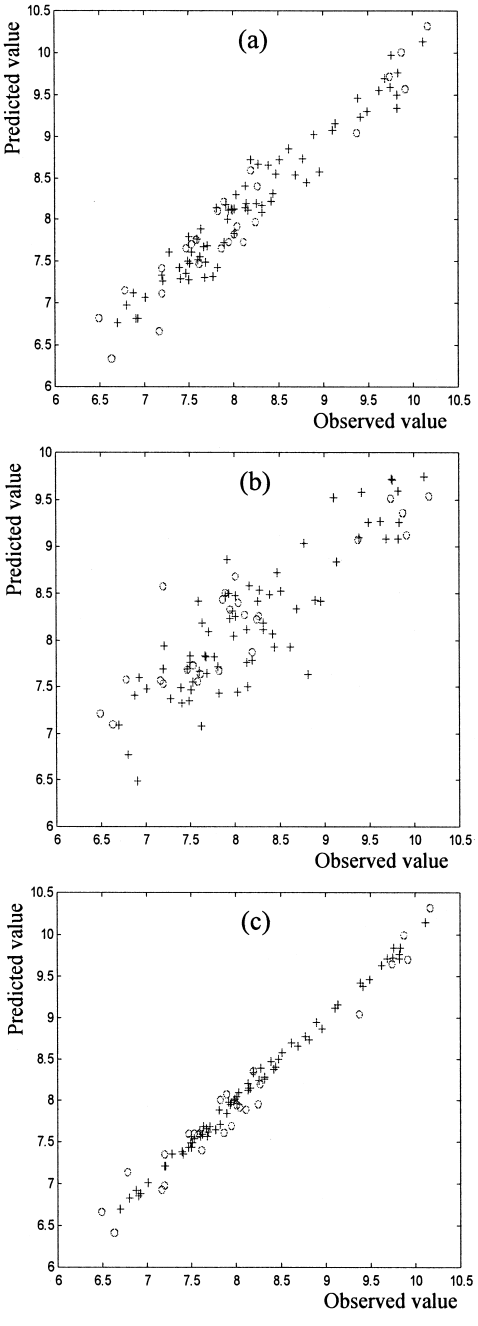


Fig. 7. Observed vs. predicted FeO values for the PCR models applying to calibration set (+), and validation set (○) which are auto-scaled (a) no SC, (b) MSC and (c) OSC using five PCs.

Table 5 summarizes results of the reduced PCR and PLS models, which show the reduced models may still be acceptable with sufficient accuracy. In the reduced models under OSC, auto-scaled PCR and PLS perform better than those of mean-centered. In PCR we can reduce the number of components from 11 to 5 with 7.6% increase in RMSEV.

RR model is applied to OSC auto-scaled data with variable selection. If the significance level for entering a candidate variable is 0.01, then we may choose the model with only two variables. Using the two variables (998, 1598 cm^{-1}), the RR model with biasing parameter $k = 0.0013$ shows the performance of PEIY = 99.32, RMSEC(70) = 0.0701, and RMSEV(25) = 0.1983. This RR model reduces the degree of collinearity significantly. Though the RR with these two variables gives similar performance with the PLS using five latent variables in prediction, it requires a great deal of attentions when specific wavenumbers should be used in a real application since there exists a possibility of the wavenumber shift afterward.

4. Conclusions

We have presented various applications of multivariate modeling with preprocessing such as scaling and signal correction to the DRIFT spectra for the prediction of FeO content in sinter ores. Since projection modeling such as PCR, PLS and RR are affected by the variation in predictor matrix ($\mathbf{X}$) that is unrelated to the response vector ($\boldsymbol{y}$), it may be necessary to perform preprocessing $\mathbf{X}$ before calibration modeling is obtained.

Through the analysis, we propose that signal corrected auto-scaled data is appropriate to the FeO prediction from DRIFTS. Projection based modeling methods are relatively simple, rapid and accurate as compared with other modeling techniques after data preprocessing. In the plot of observed vs. predicted values from the reduced PCR model on auto-scaled data (Fig. 7), it can be seen that OSC tends to give good result while MSC may be less effective when using five PCs. The similar results seem to be repeated when other modeling techniques are employed to the spectral data of sinter ores.

From those results of prediction modeling for FeO content by multivariate techniques, we conclude that prediction modeling using DRIFT spectral data needs SC technique, and the pertinent model is based on the auto-scaled OSC data using five principal components (or latent variables). Though the RR model shows similar prediction performance power to the component related regressions, we hesitate to suggest using it since there is some difficulties in the maintenance of the model and the selection of the appropriate model.

Acknowledgements

This project was supported by POSCO under Grant No. 98Y701. The authors are grateful to Mr. Woo Il Chung (the Chemical Testing and Inspection Section, Kwangyang Steelworks, POSCO) for complementary supply of the sinter ores samples with the list of FeO content together.

References

[1] K.R. Beebe, R.J. Pell, M.B. Seasholtz, Chemometrics: A Practical Guide, Wiley, New York, 1998.
[2] R. Kramer, Chemometric Techniques for Quantitative Analysis, Marcel Dekker, New York, 1998.
[3] B.K. Lavine, Anal. Chem. 70 (1999) 209R–228R.
[4] J.M. Andrade, M.S. Sánchez, L.A. Sarabia, Chemom. Intell. Lab. Syst. 46 (1999) 41–55.
[5] A. Iob, R. Buenafe, N.M. Abbas, Fuel 77 (15) (1998) 1861–1864.
[6] Y.B. Che Man, G. Setiowaty, Food Chem.: Anal. Nutr. Clin. Methods Sect. 66 (1999) 109–114.
[7] A. Geladi, D. MacDougall, H. Martens, Appl. Spectrosc. 3 (1985) 491–500.
[8] R. Schindler, B. Lendl, R. Kellner, Anal. Chim. Acta 366 (1998) 35–43.
[9] R. Schindler, B. Lendl, Anal. Chim. Acta 391 (1999) 19–28.
[10] S. Sundqvist, M. Leppämäki, E. Paatero, P. Minkkinen, Anal. Chim. Acta 391 (1999) 269–276.
[11] C. Wojciechowski, N. Dupuy, C.D. Ta, J.P. Huvenne, P. Legrand, 63 (1) (1998) 133–140.
[12] M. Blanco, J. Coello, H. Iturriage, S. Maspoch, C. de la Pezuela, Anal. Chim. Acta 333 (1996) 147–156.
[13] M.A. Kaise, D.B. Chase, Anal. Chem. 52 (1980) 1849–1851.
[14] M.P. Fuller, P.R. Griffiths, Anal. Chem. 50 (1978) 1906–1910.
[15] S. Agyare, S. Wang, P.R. Griffiths, Appl. Spectrosc. 38 (1984) 259.
[16] Z. Kkrivacsy, J. Hlavay, Spectrochim. Acta 50A (1994) 49–55.
[17] G. Kortüm, Reflectance Spectroscopy, Springer, New York, 1969.
[18] P.W. Yang, H.H. Mantsch, F. Baudais, Appl. Spectrosc. 40 (1986) 974–978.
[19] P.M. Fredericks, J.B. Lee, J.B. Osborn, D.A. Swinkels, Appl. Spectrosc. 39 (1985) 303–310.
[20] P.M. Fredericks, J.B. Lee, J.B. Osborn, D.A. Swinkels, Appl. Spectrosc. 39 (1985) 311–316.
[21] P.M. Fredericks, J.B. Osborn, D.A. Swinkels, Anal. Chem. 57 (1985) 1947–1950.
[22] D.H. Lee, Development of Analysis System for FeO Content in Sinter Ores and Iron Ores, Report 93A004, POSCO, 1994.
[23] POSCO, WK-680-306-030 Determination of FeO, 1998; KS, E 3016 Determination of ferrous oxide in iron ores, Korean Standard, 1993; JIS, M 8213 Iron ores — Method for determination of acid soluble iron (II) content, Japanese Industrial Standard, 1995.
[24] G.W. Poling, J. Electrochem. Soc. 116 (1969) 958–963.
[25] J. Sun, J. Chemom. 11 (1997) 525–532.
[26] S. Wold, H. Antti, F. Lindgren, J. Öhman, Chemom. Intell. Lab. Syst. 44 (1998) 175–185.
[27] A.E. Horel, Chem. Eng. Process 60 (1962) 54–59.
[28] A.E. Horel, R.W. Kennard, Technometrics 12 (1970) 55–67.
[29] N.R. Draper, R.C. Van Nostrand, Technometrics 21 (4) (1979) 451–466.
[30] E. Vigneau, M.F. Devaux, E.M. Qannari, P. Robert, J. Chemom. 11 (1997) 239–249.