

Moment combined partial least squares (MC-PLS) as an improved quantitative calibration method: application to the analyses of petroleum and petrochemical products

Hoeil Chung,^{*a} Soohwa Cho,^a Yukihiro Toyoda,^b Kazushi Nakano^c and Manabu Maeda^d

Received 7th November 2005, Accepted 14th February 2006

First published as an Advance Article on the web 10th March 2006

DOI: 10.1039/b515761g

A new quantitative calibration algorithm, called “Moment Combined Partial Least Squares (MC-PLS)”, which combines the moment of spectrum and conventional PLS was proposed. Its calibration performance was evaluated for the analyses of three import petroleum and petrochemical products: gasoline, naphtha and polyol samples. The selected properties for these products included the research octane number (RON) and Reid vapor pressure (RVP) for gasoline, the distillation temperature at 10% (D 10%) for naphtha and the hydroxyl (OH) number for polyol. The major concept presented here used the moment to find the closest spectrum of a sample in a given dataset, and generate the difference spectrum and the corresponding difference in the property. These difference spectra and property differences were then used for PLS calibration. The moment has been employed in spectroscopic fields as a simple and effective “spectral feature characteristic” using just a few scalar values (moments). MC-PLS showed improved prediction performance over PLS for each case. In MC-PLS, the difference spectra generated using the moments were used as explained; therefore, additional detail in spectral variations can be utilized for calibrations. Additionally, the difference in the property was employed as reference data, so that its variation range was smaller when compared with that of the original property. Consequently, the MC-PLS performance could be better since the feature-enhanced spectra were used to model a narrower range of property variations. In the case of the D 10% prediction for naphtha, a non-linear prediction pattern that occurred in conventional PLS was effectively corrected using the MC-PLS method.

Introduction

The use of near-infrared (NIR) spectroscopy in the petroleum and petrochemical industries has consistently increased over the past fifteen years.^{1–3} The advantages of using NIR spectroscopy for process monitoring are based on its ability to be used in fast, non-destructive and simultaneous multi-compositional analyses in an on-line manner. When NIR spectroscopy is adopted, multivariate calibration methods^{4–6} are primarily chosen because the resulting spectra are highly complicated as a result of the overlapping of spectral features resulting from the compositional complexity of petroleum and petrochemical products. There are diverse applications for NIR which include the measurement of various properties of gasoline,^{7–12} diesel,^{13,14} kerosene,¹⁵ naphtha,^{16–18} and lube base oil.¹⁹ The critical issues when NIR spectroscopy in conjunction with multivariate calibration is employed for refining and petrochemical applications have been discussed.²

In many real application fields, partial least squares (PLS) regression^{20–22} is frequently adopted as a multivariate calibration method. This is indirect evidence of the superior performance of PLS for calibrations in refinery and petrochemical applications. Even though PLS has been quite successful, there are demands for the development of a new calibration algorithm which can improve upon the prediction accuracy of PLS. Improved prediction accuracy in the refining and petrochemical industries would lead to improved process quality control which has the potential to result in huge economic returns.

We present a new quantitative calibration algorithm that combines the moment of spectrum and conventional PLS, and is so called the “Moment Combined Partial Least Squares” (MC-PLS). The main strategy is to use the moment to find the closest spectrum to a sample in a given dataset, generate a difference spectrum and the corresponding difference in the property. Then, these difference spectra and property differences are used for PLS calibration. A moment is frequently adopted to describe the texture of an image in image processing fields such as image retrieval systems.^{23–27} It effectively describes the outline and overall shape of the object (image with large data size) as a few scalar scales (moments). Additionally, it is known to yield a relatively stable description under situations involving a minor shift or rotation of the image (this could be a wavelength shift or baseline variation of

^aDepartment of Chemistry, College of Natural Sciences, Hanyang University, Haengdang-Dong, Seoul, Korea 133-791.
E-mail: hoeil@hanyang.ac.kr

^bDepartment of Mechanical Engineering, Niihama National College of Technology Ehime, Japan 792-0805

^cDepartment of Electronic Engineering, The University of Electro-Communications, Tokyo, Japan 182-8585

^dMatsushita Electric Industries Co., Ltd. 1006, Kadoma, Kadoma City, Osaka, Japan 571-8501

the spectral data, respectively). The concept of the moment can be expanded to spectroscopic fields, especially for vibrational spectroscopy, and utilized as a simple and effective descriptor of spectral features. Additionally, the moment of a given spectrum is independent of other spectra in a dataset, while PLS scores in a dataset are strongly inter-related amongst each other and change by inclusion or exclusion of certain spectra.

Recently, several modified PLS algorithms have been proposed and summarized.²⁸ The recent developments include interactive variable selection PLS (IVS-PLS), GIFI-PLS and orthogonal signal correction PLS (OSC-PLS). IVS-PLS involves applying weights on x -variable (spectrum) based on the importance of feature before performing PLS. The strategy of GIFI is to mathematically transform each x -variable into a set of new variables to model nonlinear relationships with y -variable (property). OSC-PLS is designed to remove unwanted systematic spectral variations such as temperature and scattering variations (that are orthogonal to the analyte in feature or vector space) from the original spectrum. All the recent developments are mostly based on the pre-processing of the original spectrum to improve correlation between x and y variables before performing PLS. MC-PLS differs in the sense that it utilizes the new concept of spectral moment to topologically find the most matching spectrum, and difference spectrum and difference in property are used for the input for PLS. In this way, the minute spectral difference can be recognized and then PLS calibration accuracy can be improved.

To evaluate the calibration capability of MC-PLS, we used three datasets of gasoline, naphtha and polyol samples. Gasoline and naphtha are the most important products that require fast near-infrared spectroscopic analysis in petroleum refining and petrochemical fields since a significant amount of cost-saving is expected when accurate on-line monitoring and advanced process control are harmonized. Polyol is an important raw material for the manufacture of numerous polyurethane products and it requires a simple, rapid and online spectroscopic monitoring method that can replace the current monitoring method which is a tedious and time-consuming wet chemical analysis. For all three datasets, MC-PLS provided a better standard error of prediction (SEP) over conventional PLS.

Moment combined partial least squares (MC-PLS)

As introduced, a key feature of MC-PLS is to find the most closely matching sample in a dataset by comparing the spectral moments of the various samples and generating the difference spectrum and the difference in property (between a given sample spectrum and the closest sample spectrum in a database set). Then, the generated difference spectra and the corresponding differences in the property (Δ property) are used as the inputs for the PLS calibration. Since the difference spectra, generated by subtracting a given sample spectrum from closest matching spectrum in a database, are used for calibration, even minute spectral variation can be highlighted and compared with the original spectra. Additionally, the magnitude of the Δ property is small since the two closest samples are chosen. Consequently, the feature-enhanced (difference) spectra with a narrower range of property variation are utilized for

PLS; therefore, it is expected that MC-PLS can recognize spectral variations more effectively and lead to improved calibration performance over conventional PLS. The samples used in this study (particularly gasoline and naphtha) are chemically complex mixtures with the presence of isomers and hydrocarbons of various chain lengths (there are more than 200 individual components in gasoline) and the resulting spectra will be highly overlapping with broad spectral features. To improve calibration for such cases, an algorithm should effectively recognize and address minute spectral variations.

The overall procedure of MC-PLS is as follows. First, a database set is constructed by collecting diverse samples that include possible compositional and property variations as typically done in the routine dataset preparation procedures for multivariate calibration. The spectral moments of samples in the training set are compared with those of the database set. Then, the difference spectra and the corresponding Δ property values are used to perform PLS. In the course of PLS, the optimal spectral range and number of factors are determined as usual. In the prediction step, the moment of the unknown sample is calculated and the sample with the closest corresponding moment in the database set is found. The difference spectrum is generated and predicted using a developed PLS model to yield the Δ property value. The final property of the unknown sample is the sum of the property of the closest sample in the database set and Δ property value. Detailed mathematical descriptions follow.

As indicated, we uniquely employed the n -th moment to describe the shape of each spectrum as the feature vector. Initially, the raw spectra (or preprocessed by, for example, baseline correction) $x(\lambda)$ are Fourier transformed to achieve $X(j\omega)$ by the use of equation (1):

$$X(j\omega) = \int_{-\infty}^{\infty} x(\lambda) \cdot e^{-j\omega\lambda} d\lambda \quad (1)$$

where $j = \sqrt{-1}$ and ω is the angular frequency. Maclaurin's expansion is applied to the exponential part in equation (1) to yield

$$X(j\omega) = \int_{-\infty}^{\infty} \left[\sum_{n=0}^{\infty} \frac{(-j\omega\lambda)^n}{n!} \right] \cdot x(\lambda) \cdot d\lambda = \sum_{n=0}^{\infty} (-j)^n \cdot m_n \frac{\omega^n}{n!} \quad (2)$$

where m_n is the n -th moment of $x(\lambda)$. The term m_n can be derived as:

$$m_n = \int_{-\infty}^{\infty} \lambda^n \cdot x(\lambda) \cdot d\lambda \quad (3)$$

The inverse Fourier transform is then performed on equation (2) to yield equation (4):

$$x(\lambda) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \sum_{n=0}^{\infty} (-j)^n m_n \frac{\omega^n}{n!} e^{j\omega\lambda} d\omega = \sum_{n=0}^{\infty} \left[\frac{1}{2\pi} \int_{-\infty}^{\infty} (-j)^n \frac{\omega^n}{n!} e^{j\omega\lambda} d\omega \right] \cdot m_n \quad (4)$$

Raw spectra $x(\lambda)$ can now be expressed as moments (m_1, m_2, \dots, m_n) that represent the spectral features in many fewer

dimensions (scalar values). The meaning of each individual moment is fairly difficult to translate into spectroscopic language; however, in the image processing field, the first, second, third and fourth moments represent the average, standard deviation, skewness and kurtosis, respectively. The higher order moments are more difficult to conceptualize. Nevertheless, the moments can be utilized as an effective “spectral feature characteristic” in spectroscopic fields.

To find the closest moment, we used multi-dimensional spaces constructed by several moments. The Mahalanobis distance (r_i) is then utilized to find the closest sample in the dataset using multi-dimensional moment spaces. After finding the closest sample, the difference spectrum ($S_{\text{difference}}$) and difference in the property ($P_{\text{difference}}$) are generated by following equations:

$$S_{\text{database set}} - S_{\text{training set}} = S_{\text{difference, training set}} \quad (5)$$

$$P_{\text{database set}} - P_{\text{training set}} = P_{\text{difference, training set}} \quad (6)$$

where capital S and P represent the spectrum and property value, respectively. This procedure repeated for n times (n : number of samples in training set). Finally, n numbers of $S_{\text{difference, training set}}$ and $P_{\text{difference, training set}}$ are used to perform PLS and finally a calibration model is constructed. In the prediction step, $S_{\text{difference, prediction}}$ is generated by

$$S_{\text{database set}} - S_{\text{prediction set}} = S_{\text{difference, prediction}} \quad (8)$$

Then, $S_{\text{difference, prediction}}$ is predicted by a developed PLS model to yield $P_{\text{difference, prediction}}$. The final property of prediction is:

$$\text{Final property} = P_{\text{database set}} + P_{\text{difference, prediction}} \quad (9)$$

The sign of $P_{\text{difference, prediction}}$ is either positive or negative.

Experimental

Preparation of gasoline sample sets

An overall description of the three different datasets (gasoline, naphtha and polyol) is presented in Table 1. All gasoline samples were acquired from Nisseki-Mitsubishi Oil, Iwakuni, Japan. First, 225 samples for the database set were collected over a period of 12 months. The main purpose of collecting the samples over a year was to include compositional variations based on the four seasonal grade changes. The samples for the database set were collected in two different ways. First, the gasoline samples (165 samples) were collected on a regular basis. In this procedure, a significant effort was made to select

Table 1 The number of spectra in each dataset for gasoline, naphtha and polyol samples

	Gasoline	Naphtha	Polyol
Database set	225	90	34
Training set	60	40	32
Validation set	109	85	37
Total	394	215	103

samples that represented as diverse a range as possible by continuously and carefully examining changes in the gasoline components (feed stock). Property variations in parameters such as the research octane number (RON) and the Reid vapor pressure (RVP) of refined gasoline are usually small. In order to enhance the variations, we also acquired samples (which we call “step change” samples) that were artificially varied by changing the blending ratios to values that were wider than the normal upper and lower limits of each component to increase the range of the blends, as well as the chemical variations. By using this approach, we acquired 60 step change samples, which we added into the database set.

The training set was used to find optimum parameters, such as the optimal number of n -th moments, as well as factors. For this purpose, 60 final gasoline samples in addition to the step change samples were prepared over four months.

The developed models using MC-PLS and PLS were validated by 109 gasoline samples collected over an additional 12 months. In this period, only the refined gasoline samples were collected and predicted. In total, 28 months passed between the construction of the database set and its final validation. The RON and RVP for all of these samples were determined using a conventional knock engine²⁹ and an RVP analyzer,³⁰ respectively. The RON and RVP of gasoline are the most important properties for quality assurance and process control.

Preparation of naphtha and polyol sample sets

Two hundred and fifteen naphtha samples were obtained over a 9 month period from a petrochemical company in Korea. Samples were cautiously collected to introduce compositional variations into the datasets. The whole dataset was divided as described in Table 1. The distillation temperature at 10% (D 10%) of each sample was measured using the ASTM D86 method.³¹

A total of 103 polyol samples was acquired from Dongnam Chemical Co., Korea. These samples were collected over 8 months to give additional process related variations in the datasets. Each dataset was divided as described in Table 1. All the reference analyses (hydroxyl number determination) were performed using a wet analysis.³²

Spectral collection and data processing

All of the NIR spectra for gasoline and naphtha were collected using an ABB Bomem MB-160 bench-top FT-NIR spectrometer (Quebec City, Quebec, Canada) equipped with a tungsten-halogen source and a DTGS detector. A 0.5-mm pathlength flow cell (transmission) incorporating CaF₂ windows was used to collect the spectra. Air was used as a background for all of the sample spectra. Each NIR spectrum was obtained by accumulating 16 scans at a resolution of 4 cm⁻¹. NIR spectra were collected over a range of 5000–4000 cm⁻¹ at 25 °C.

The NIR spectra of polyol samples were collected using the same spectrometer with a different spectral collection device that was designed to use a disposable vial as a sample container. This is described in a previous publication.³³ The vial holder was used to hold a glass vial as well as to control

the temperature. A narrow rectangular aperture of 0.85 mm (width) was used to minimize the spectroscopic deviations from the round surface of the vial. NIR spectra were collected over a range of 9500–6000 cm^{-1} with a resolution of 8 cm^{-1} . The temperature of the sample was controlled at 75 °C. Air was used as the background for all samples.

PLS, MC-PLS and necessary spectral pre-processing were accomplished using Matlab Version 6.5 (The MathWorks Inc., MA, USA).

Results and discussion

NIR spectral features

Since gasoline and naphtha are very complex mixtures composed of C_4 – C_9 hydrocarbons, the NIR spectra of both gasoline and naphtha display highly overlapping and broad features because of the inherently broad spectral features of the NIR spectral bands (overtones and combinations) in addition to the compositional complexity present. When NIR spectroscopy is used for such applications, it is necessary to use a spectral range that has enough spectral sensitivity (variation) to resolve such a complex mixture. For this purpose we used the spectral region from 5000 to 4000 cm^{-1} (including mostly combination bands of $-\text{CH}$ vibrations), which provided the highest sensitivity as well as selectivity in the NIR region. Fig. 1 shows the NIR spectra of gasoline and naphtha samples (40 randomly selected samples for each case). The overall spectral features are similar to each other. The bands around 4600 cm^{-1} are more distinct for gasoline due to its higher concentration of aromatic compounds. These bands arise from pure aromatic or olefin compounds, particularly from a combination of $=\text{C}-\text{H}$ stretching at 3100–3000 cm^{-1} and $\text{C}=\text{C}$ ring stretching at 1600–1450 cm^{-1} (e.g., 3050 cm^{-1} + 1550 cm^{-1} = 4600 cm^{-1}). The spectral variations are relatively small, even in the most sensitive range in the NIR region.

The NIR spectra of (9100–6000 cm^{-1}) of three selected polyol samples (hydroxyl (OH) number: 89.5, 122.8 and 155.1 mg KOH g^{-1}) using disposable vials (pathlength: 6.11 mm) are shown also in Fig. 1. The 9100–7800 cm^{-1} bands correspond to the second overtone bands of $-\text{CH}$ and the band centred at 6900 cm^{-1} corresponds to an OH combination. As shown, the absorption of the hydroxyl bands increased with the number of OH. The 9100–6000 cm^{-1} range including the 9100–7500 cm^{-1} hydrocarbon bands as well as the 7500–6000 cm^{-1} range was used for calibration.

Spectral representation by moment and principal component scores

To demonstrate the spectral description capability of a moment, we have initially generated a two dimensional (2D) sample space constructed by two moments (moment 3 and 4) as well as two principal component (PC) scores (score 1 and 2). These results are shown in Fig. 2. On each 2D space, gasoline samples in the database set are represented by moments and PC scores, respectively. Among them, two samples that are located at the both edges of the cluster and one sample in the middle of the cluster were selected (total 3 samples and designated as A, B and C). The two samples (A and C) located

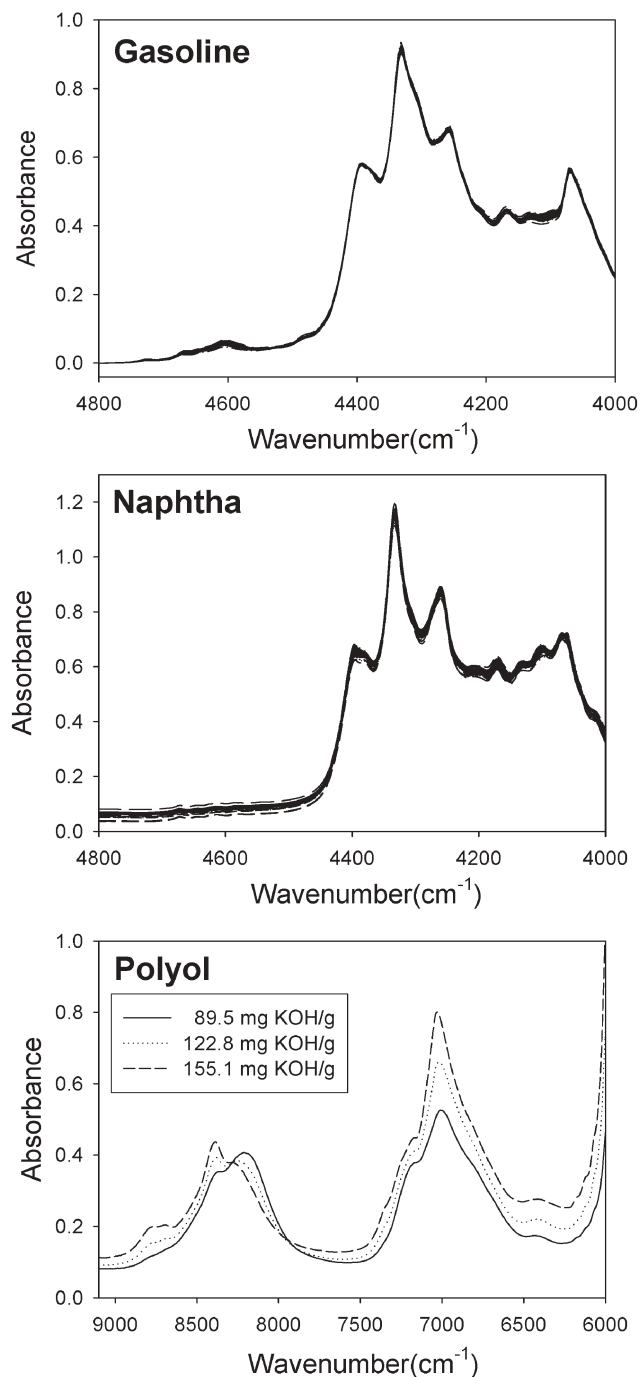


Fig. 1 The NIR spectra of gasoline and naphtha samples (40 randomly selected samples for each case), and three selected polyol samples collected using disposable vials.

on both edges of the cluster should represent considerable spectral differences. The corresponding three spectra for each case are shown in the same figure. It is clear that the spectral variation was successfully represented with the use of moments. As shown, the RON variation and spectral variation of the three samples was distinct since these were located far apart in the cluster. In particular, the aromatic bands centred around 4600 cm^{-1} were clearly increasing with respect to the RON since aromatic compounds usually have high RON

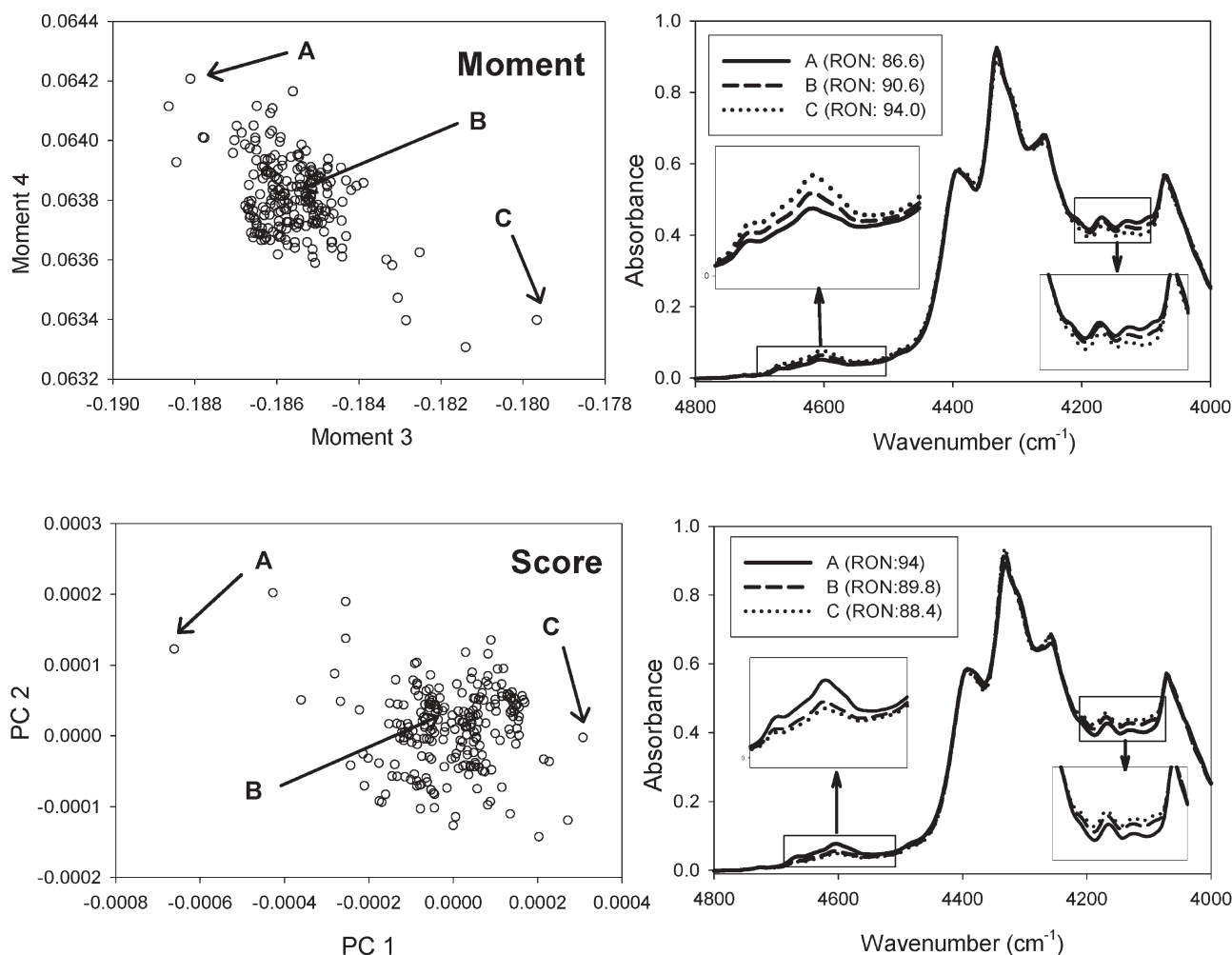


Fig. 2 Two dimensional (2D) sample spaces constructed by two moments (moment 3 and 4) as well as two principal component (PC) scores (score 1 and 2) for gasoline samples in the database set. Three samples were selected and the corresponding spectra are also shown.

values. Also, similar results were observed for the PC scores, while the trend was fairly different from that of the moment. The overall results in Fig. 2 show that the moment successfully represents the spectral variation and can be used as an alternative and effective spectral representation method. Similar results were also achieved for naphtha and polyol datasets.

Dataset preparation

Table 1 summarizes the overall spectral dataset construction for gasoline, naphtha and polyol samples. The database set contains samples that include wide compositional variations which span any expected variation that could be possessed by

unknown samples. For an unknown sample, the closest matching sample in the database set was found by comparing moments of spectra. The training set was used to develop an optimal PLS model that describes the relationship between difference spectra ($S_{\text{difference}}$) and the difference in property ($P_{\text{difference}}$). In this step, an optimal number of PLS factors and the proper spectral ranges were determined. The validation set is for the independent evaluation of prediction performance of a developed calibration model.

Table 2 summarizes the selected properties for each sample and corresponding range of variation in each dataset. It is very important for these properties to be accurately measured from an industrial viewpoint, especially the RON and RVP

Table 2 Selected properties for each sample and corresponding ranges of variation in each dataset

		Range		
		Database set	Training set	Validation set
Gasoline	RON	85.6–94.0	87.6–92.1	89.5–90.8
	RVP/kPa	65.3–87.5	66.2–86.3	68.5–85.6
Naphtha	D 10%/°C	37.2–65.3	40.0–60.4	37.9–65.6
Polyol	OH number/mg KOH g ⁻¹	21.4–155.1	23.7–155.1	21.4–155.1

Table 3 The overall calibration results using MC-PLS. Numbers in parentheses correspond to the number of factors used

		Spectral range	Number of moments used	SECV	SEP
Gasoline	RON	4800–4000 cm^{-1}	9	0.19 (4)	0.20
	RVP/kPa	4800–4000 cm^{-1}	6	1.15 (5)	1.04
Naphtha	D 10%/°C	4800–4000 cm^{-1}	6	0.94 (5)	0.93
Polyol	OH number/mg KOH g^{-1}	9100–6000 cm^{-1}	7	1.60 (4)	1.64

of gasoline. Even a slight improvement in the accuracy of measuring these properties can result in substantial economic savings.

Training and validation

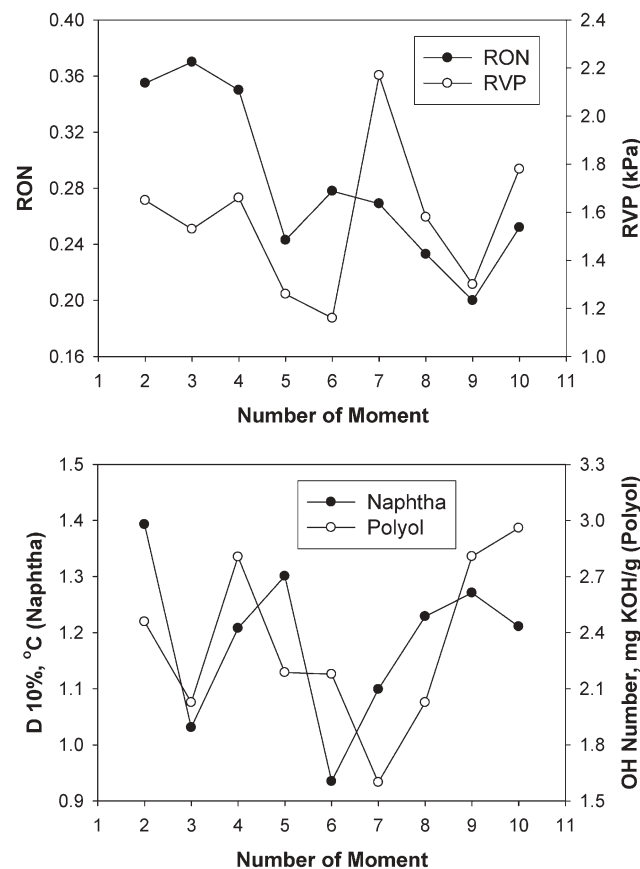
To determine the optimal number of moments, we generated the first through tenth moments and continuously added one by one from first moment. In this way, a total of 10 sets of difference spectra ($S_{\text{difference}}$) and differences in the property ($P_{\text{difference}}$) from the training set were generated. PLS was then performed and the corresponding SECVs (Standard Error of Cross Validations) were evaluated at each number of moments to determine the optimal result. The spectral ranges used for each dataset are summarized in Table 3. The optimum number of factors was identified as the one that yielded the minimum SECV. The SECV is a useful index to determine the optimal number of PLS factors and prevents over-modelling. The relationship of the SECV to the number of PLS factors was

examined. The SECV decreased sharply for the first several factors and then gradually decreased for the remaining factors. After a certain factor, it started to increase. The factor before this increase in the SECV was chosen as the optimum number of factors. The resulting SECVs were compared at each number of moments. Simultaneously, using the original spectra, conventional PLS was performed in order to compare its results with those from MC-PLS. The combined set which was composed of the original database and the training dataset was used to build the PLS model using the same spectral ranges as MC-PLS.

Fig. 3 (top) shows the variation of the SECV based on the number of moments used for determination of the RON and RVP. As discussed, each moment is independent of the others therefore there was no systematic pattern as moments were added. Nonetheless, the moment corresponding to lowest SECV was clearly identifiable. For the RON, the lowest SECV was achieved at 9 moments and the lowest SECV was clearly identified at 6 moments for the RVP. The optimal numbers of moments for naphtha and polyol were also determined by generating and examining the same plot shown Fig. 3 (bottom).

The overall MC-PLS and PLS calibration results are summarized in Tables 3 and 4, respectively. The numbers in parentheses correspond to the number of factors used. The SECV was estimated using the training set and the standard error of prediction (SEP) was calculated by predicting the samples contained in the validation set. In every case, the SEPs using the MC-PLS were improved over the SEPs from the PLS. The difference spectra of three training sets generated from each optimized number of moments are shown in Fig. 4. When compared with the spectral features in Fig. 1, the spectral variations are much more descriptive and effectively enhanced.

The improved SEPs using MC-PLS for the RON and RVP of gasoline could be attributed to two factors. First, since gasoline is a very complex mixture containing more than 200 individual hydrocarbon components, the resulting NIR spectra are highly overlapping without any distinct spectral features, as shown in Fig. 1. Therefore, the spectral variations that correspond to a change in the properties of the gasoline would not be significant in the original raw spectra. In

**Fig. 3** The variation of the SECV based on the number of moments used for the determination of the RON and RVP of gasoline (top), and D 10% of naphtha and OH number of polyol (bottom).**Table 4** The overall calibration results using PLS. Numbers in parentheses correspond to the number of factors used

		Spectral range	SECV	SEP
Gasoline	RON	4800–4000 cm^{-1}	0.18 (6)	0.27
	RVP/kPa	4800–4000 cm^{-1}	1.09 (7)	1.66
Naphtha	D 10%/°C	4800–4000 cm^{-1}	1.38 (6)	1.22
Polyol	OH number/mg KOH g^{-1}	9100–6000 cm^{-1}	2.01 (6)	2.21

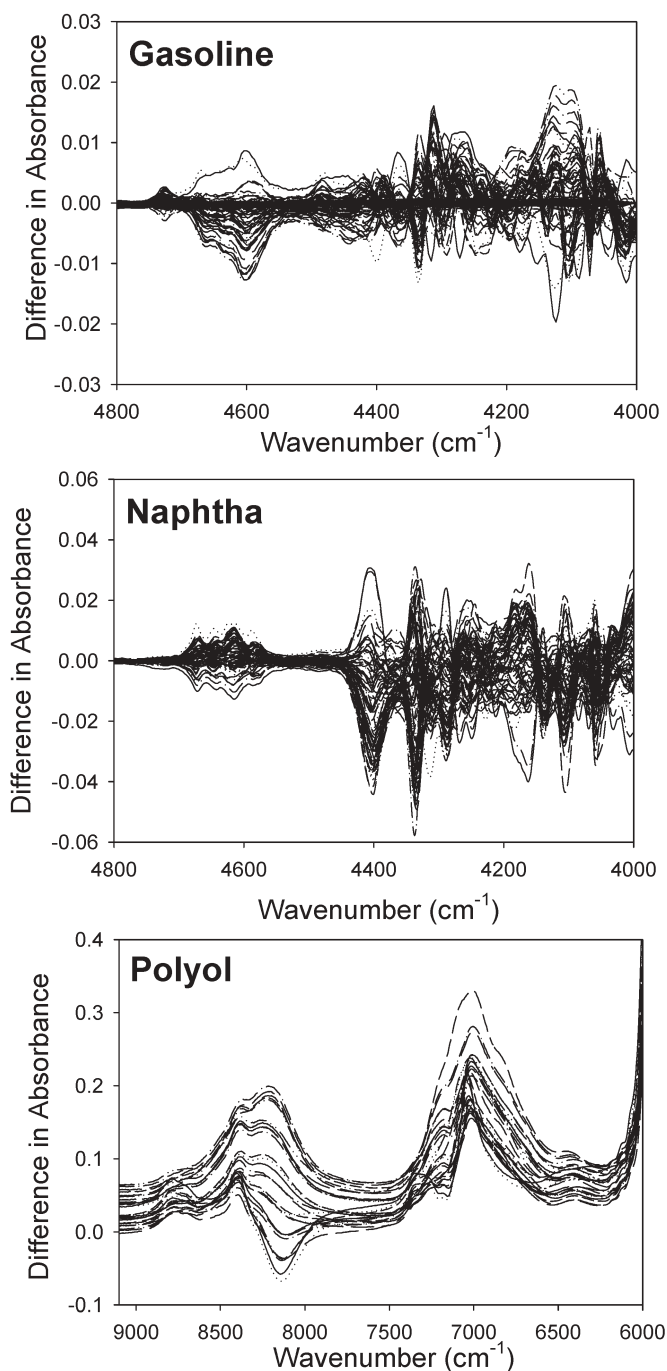


Fig. 4 The difference spectra of the training sets of gasoline, naphtha and polyol generated from each optimized number of moments. Three, four and five moments are used for gasoline, naphtha and polyol, respectively.

MC-PLS, the difference spectra generated using the moments are used as shown in Fig. 4 and a more enhanced spectral variation can be realized for calibration. Second, when PLS is performed using the difference spectra, the difference in the property is employed as reference data. Its property variation range is smaller than when compared to that of the original property. Consequently, from a modelling standpoint, the PLS performance could be improved since feature-enhanced spectra are used to model smaller ranges of property variation.

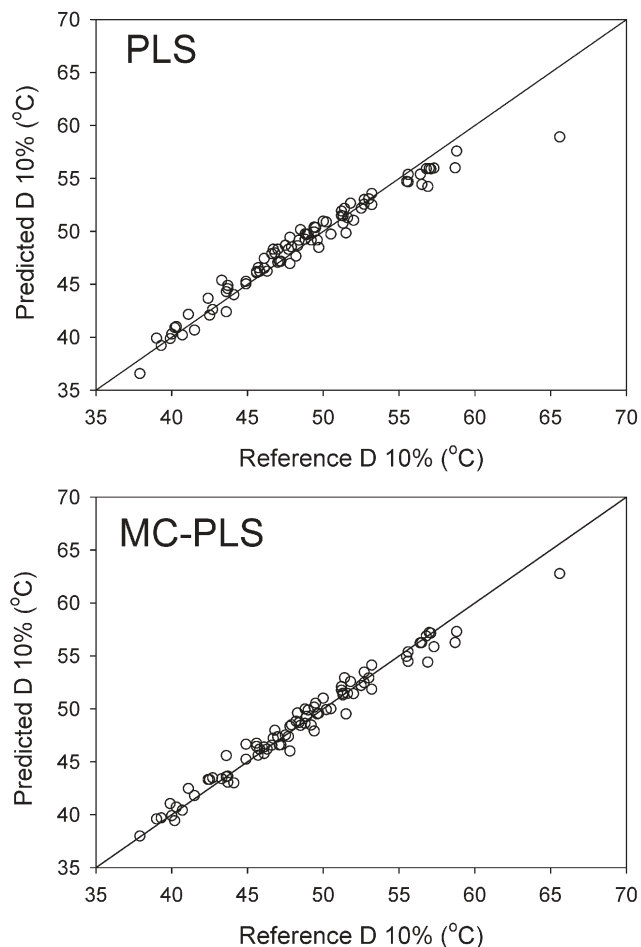


Fig. 5 The prediction results of D 10% for naphtha using both MC-PLS and PLS.

Fig. 5 shows the prediction results of the D 10% of naphtha for PLS (top) and MC-PLS (bottom). A non-linear relationship in the PLS prediction was apparently observed while this was considerably reduced in the MC-PLS. When physical properties, such as distillation temperature, are measured using spectroscopic techniques, the variation of these physical properties may not be linear with respect to spectral variations based on the chemical composition. It is expected that there are possible unforeseen non-linear physical interactions among each component that could make the prediction of the final property inaccurate. MC-PLS uses the differences in the property that can be regarded as linear because its variation is narrow in range; therefore, the non-linear prediction pattern that occurred in conventional PLS is effectively corrected. This is another distinct advantage of MC-PLS; it can be applied to model various non-linear properties, especially those which occur in petroleum refining and petrochemical fields.

For the OH number of polyol, which is a fairly linear property with respect to spectral variation, the MC-PLS also shows improved prediction results over PLS. As is seen in the case of gasoline, the improved prediction of the OH number using MC-PLS can be attributed to the use of feature-enhanced difference spectra within a narrower range of property variation.

Based on the overall results, MC-PLS shows an improved prediction performance for all three cases, especially when the spectral variation of a target sample is small due to a highly complex matrix. This method also excels when it involves the determination of a physical property that is not linear with respect to compositional variation.

Conclusion

We have presented a new quantitative calibration method, MC-PLS, which can possibly provide enhanced prediction accuracy over PLS. Using the three datasets acquired from real industrial fields was a practical way to evaluate the applicability of MC-PLS. Nonetheless, additional results that corroborate those presented here will be necessary to ensure that MC-PLS will be accepted in the NIR spectroscopic community. Future work will be directed to investigate (1) the performance of MC-PLS in different fields such as agricultural and pharmaceutical areas, especially those which use diffuse reflectance NIR spectra, and (2) the other vibrational spectroscopic data from IR and Raman spectroscopy.

Acknowledgements

The authors thank Mr. Nishioka and his team from Nisseki-Mitsubishi Oil, Japan, for their continual supply of samples, necessary data, and advice. This work was supported by the Korea Science and Engineering Foundation (Grant No. R-14-2002-004-01000-0).

References

- 1 D. A. Burns and E. W. Ciurczak, in *Handbook of Near-Infrared Analysis*, Marcel Dekker Inc., New York, 1992.
- 2 S. Macho and M. S. Larrechi, *Trends Anal. Chem.*, 2002, **21**, 12, 799–806.
- 3 J. Workman, Jr., *J. Near Infrared Spectrosc.*, 1996, **4**, 69–74.
- 4 H. Martens and T. M. Naes, *Multivariate Calibration*, John Wiley and Sons, New York, 1989.
- 5 K. R. Beebe, R. J. Pell and M. B. Seasholtz, *Chemometrics: A Practical Guide*, John Wiley and Sons, New York, 1998.
- 6 B. K. Lavine and J. Workman, Jr., *Anal. Chem.*, 2002, **74**, 2763–2770.
- 7 H. Chung, H. Lee and C. H. Jun, *Bull. Korean Chem. Soc.*, 2001, **22**, 37–42.
- 8 G. Bohacs, Z. Ovadi and A. Salgo, *J. Near Infrared Spectrosc.*, 1998, **6**, 341–348.
- 9 N. M. Faber, D. L. Duewer, S. J. Choquette, T. L. Green and S. N. Chesler, *Anal. Chem.*, 1998, **70**, 2972–2982.
- 10 G. Buttner, *Process Control Qual.*, 1997, **9**, 197–203.
- 11 I. Litani-Barzilai, I. Sela, V. Bulatov, I. Zilberman and I. Schechter, *Anal. Chim. Acta*, 1997, **339**, 193–199.
- 12 J. J. Kelly, C. H. Barlow, T. M. Jinguji and J. B. Callis, *Anal. Chem.*, 1989, **61**, 313–320.
- 13 M. C. Breikreitz, I. M. Raimundo, J. J. R. Rohwedder, C. Pasquini, H. A. Dantas Filho, G. E. Jose and M. C. U. Araujo, *Analyst*, 2003, **128**, 8, 1204–1207.
- 14 S. J. Foulk and B. E. DeSimas, *Process Control Qual.*, 1992, **2**, 69–72.
- 15 H. Chung, M. S. Ku and J. S. Lee, *Vib. Spectrosc.*, 1999, **20**, 2, 155–163.
- 16 H. Chung, J. S. Lee and M. S. Ku, *Am. Lab.*, 1999, **31**, 23, 24–25.
- 17 M. S. Ku and H. Chung, *Appl. Spectrosc.*, 1999, **53**, 5, 557–564.
- 18 M. S. Ku, H. Chung and J. S. Lee, *Bull. Korean Chem. Soc.*, 1998, **19**, 1189–1193.
- 19 H. Chung and M. S. Ku, *Appl. Spectrosc.*, 2003, **57**, 5, 545–550.
- 20 P. M. Fredericks, J. B. Lee, P. R. Osborn and D. A. Swinkels, *Appl. Spectrosc.*, 1985, **39**, 303–310.
- 21 D. M. Haaland and E. V. Thomas, *Anal. Chem.*, 1988, **60**, 1193–1202.
- 22 S. Wold, M. Sjöström and Lennart Eriksson, *Chemom. Intell. Lab. Syst.*, 2001, **58**, 109–130.
- 23 M. Gibaldi and D. Perrier, 'Noncompartmental analysis based on statistical moment theory', in *Pharmacokinetics*, Marcel Dekker Inc., New York, 1982.
- 24 L. Gupta and M. D. Srinath, *Pattern Recognit.*, 1987, **20**, 3, 267–272.
- 25 J. Tajima, *Pattern Recognit.*, 1981, **14**, 211–217.
- 26 D. W. Thomas and B. R. Wilkins, *Pattern Recognit.*, 1972, **4**, 4, 379–389.
- 27 M. F. Zakaria, L. J. Vroomen, P. J. A. Zsombor, J. M. Murray and H. M. Kessel, *Pattern Recognit.*, 1987, **20**, 6, 639–643.
- 28 S. Wold, J. Trygg, A. Berglund and H. Antti, *Chemom. Intell. Lab. Syst.*, 2001, **58**, 131–150.
- 29 American Society for Testing and Materials, *Annual Book of ASTM Standards*, Philadelphia, 1985, Method D 2699.
- 30 American Society for Testing and Materials, *Annual Book of ASTM Standards*, Philadelphia, 1985, Method D 323.
- 31 American Society for Testing and Materials, *Annual Book of ASTM Standards*, Philadelphia, 1985, Method D 86.
- 32 American Society for Testing and Materials, *Annual Book of ASTM Standards*, Philadelphia, 1982, Method D 2848.
- 33 S. Cho, H. Chung and Y. Lee, *Microchem. J.*, 2005, **80**, 189–193.