

## Rapid Compositional Analysis of Naphtha by Near-Infrared Spectroscopy

Min-Sik Ku, Hoeil Chung\*, and Joon-Sik Lee

NIR Project TFT, Production Technology Center, SK Corporation,  
110 Nam-Gu, Kosa-Dong, Ulsan 680-130, Korea

Received

The determination of total paraffin, naphthene, and aromatic (PNA) contents in naphtha samples, which were directly obtained from actual refining process, has been studied using near-infrared (NIR) spectroscopy. Each of the total PNA concentrations in naphtha has been successfully analyzed using NIR spectroscopy. Partial least squares (PLS) regression method has been utilized to quantify the total PNA contents in naphtha from the NIR spectral bands. The NIR calibration results showed an excellent correlation with those of conventional gas chromatography (GC). Due to its rapidity and accuracy, NIR spectroscopy is appeared as a new analytical technique which can be substituted for the conventional GC method for the quantitative analysis of petrochemical products including naphtha.

### Introduction

Naphtha is one of the most important materials serving as a basic feed for the production of ethylene and BTX (Benzene Toluene Xylene) in petrochemical industries.<sup>1</sup> The range of boiling temperature of naphtha is typically from 120 to 150 °C. Naphtha is converted to ethylene and BTX by thermal cracking and catalytic reforming, respectively. Naphtha is a very complex mixture of more than one hundred components and composed of various kinds of hydrocarbons, such as alkanes, cycloalkanes (naphthenes), and aromatic hydrocarbons, which are present in an original crude oil. The individual components in naphtha, however, are fairly similar in structure.

The importance of naphtha in petrochemical industries inevitably requires an accurate and fast analytical method to determine its composition for the quality monitoring and process control when naphtha is used as a feed. The most typical method, for the compositional analysis of total paraffin, naphthene, and aromatic (so called PNA analysis) contents, is a conventional gas chromatography (GC).<sup>2-3</sup> The GC method, however, requires a long analysis time (over 90 minutes) due to the separation of huge numbers of hydrocarbons with similar molecular structures. Another disadvantage of the GC method is that the accuracy and reproducibility of the measurement depend on the separation efficiency of a GC column. Because the peaks of various components in chromatogram are very closely located each other, the minute change of retention time seriously affects the final calculation of composition by peak height or area.

Recently, near-infrared (NIR) spectroscopy<sup>4-5</sup> has been paid special attention as a new technique for the analysis of total PNA composition in naphtha. It is because this technique is not only fast but also allowing simultaneous measurements of multiple components with a great precision and accuracy. In order to measure total PNA composition in naphtha by NIR spectroscopy, the naphtha samples, which were directly obtained from actual refining process, have

been used in this study. Even though the spectral features of paraffin, naphthene, and aromatic hydrocarbons are quite different over the entire NIR region, the spectral bands of individual components in naphtha are overlapped each other with similar molar absorptivity. Using NIR spectroscopy, minute spectral differences among paraffin, naphthene, and aromatic hydrocarbons could be identified and those spectral differences provide an enough quantitative information which is necessary for the calibration using multivariate regressions. Partial least squares (PLS) regression<sup>6</sup> has been utilized to differentiate the spectral differences and to quantify total PNA composition.

The main goal of the present work is to evaluate NIR spectroscopic technique for the analysis of total PNA composition in naphtha. For this purpose, all the NIR prediction results have been compared with conventional GC data.

### Experimental

**Instrument and Apparatus.** NIR spectra were collected over the 1100 to 2500 nm region with 2 nm point spacing with a NIRSystems 6500 model spectrometer (NIR-Systems Inc., Silver Spring, MD) equipped with a quartz halogen lamp, PbS detector, and fiber optic interactance/reflectance probe (diameter: 2.54 cm). The fiber optic probe consists of illuminating fibers, receiving fibers, and a reflecting mirror. The length of the fiber optic was 2 m. The distance between the optical fibers and the reflecting mirror is 0.5 cm, resulting an actual pathlength of 1 cm.

**Sample Preparation.** Fifty naphtha samples were obtained over 3 month period at SK Corporation (Ulsan, Korea). Over this period, the naphtha samples have been cautiously collected for NIR analysis. This sample collection is also very important to include the process related variations into calibration data set which would result in more robust calibration model in final. The concentration range of total paraffin, naphthene, and aromatic hydrocarbons were 43.84 to 60.81, 25.34 to 38.24, and 11.43 to 17.82 percent by volume, respectively. Immediately after

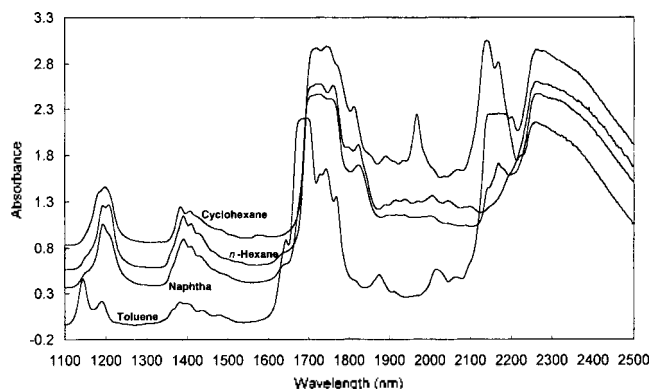
\*To whom correspondence should be addressed.

collection, the samples were sealed and stored in a refrigerator at 4 °C to prevent the evaporation of hydrocarbons. The concentration of each sample was measured by a Hewlett-Packard 5890 GC (Reformulyzer™) equipped with three separation columns and a flame ionization detector (FID). Initially, the first column (OV-275, 3 m × 3.1 mm i.d.) was utilized to separate naphtha to aromatic and non-aromatic hydrocarbons. Then pre-separated aromatic and non-aromatic hydrocarbons were transferred to OV-101 column (dimethylpolysiloxane, 4 m × 3.1 mm i.d.) and 13X molecular sieve column (1.55 m × 2 mm i.d.), respectively, where separation by carbon number was achieved. For analysis, 0.2 μL of naphtha sample was injected using a syringe and helium was used as a carrier gas. The temperature conditions of oven, injection port, and detector were 130, 120, and 150 °C, respectively.

**NIR Spectra.** Spectra were collected by dipping the fiber optic probe into each sample in a sealed bottle. Samples were allowed to equilibrate to room temperature and then corresponding single beam spectra were collected. Background spectra of air were collected for every sample immediately before the collection of the sample single beam spectra. In total, 50 absorbance spectra were collected. All data processings including PLS regression and second derivative algorithm were accomplished using NSAS ver. 3.5 software (NIRSystems Inc., Silver Spring, MD).

## Results and Discussion

**Near-Infrared Spectral Features.** For the selective and simultaneous quantification of total paraffin, naphthene, and aromatic contents in naphtha, the minute spectral differences between components in the NIR region should be analyzed. NIR spectra of typical naphtha, *n*-hexane, cyclohexane, and toluene are shown in Figure 1. The spectra were linearly offset for the purpose of qualitative comparison. *n*-Hexane, cyclohexane, and toluene were selected to examine the general spectral differences among paraffin, naphthene, and aromatic hydrocarbons. The most useful spectral information is located in the 1100-1650 nm and 1800-2100 nm spectral ranges. The 1650-1800 nm and 2100-2500 nm ranges contain no useful spectral information due to the strong and saturated absorption of the NIR radiation from a long optical pathlength. The bands around 1200 nm correspond to the second overtone of CH band at 3000-2700

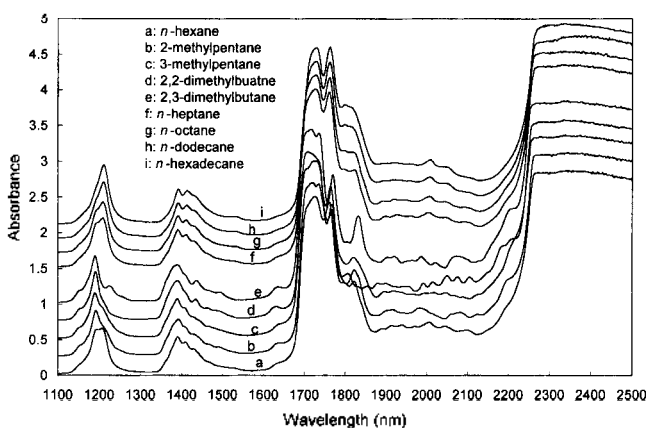


**Figure 1.** NIR spectra (1100-2500 nm) of typical naphtha, *n*-hexane, cyclohexane, and toluene. (offset for clarity)

cm<sup>-1</sup> in mid-infrared region (e.g. 2870 cm<sup>-1</sup> × 3 = 8610 cm<sup>-1</sup> = 1161 nm). The bands around 1400 nm correspond to the combination bands between the first overtone of CH stretching from the 3000-2700 cm<sup>-1</sup> range and the CH<sub>3</sub>/CH<sub>2</sub> bending around 1450 cm<sup>-1</sup> (e.g. 2870 cm<sup>-1</sup> × 2 + 1450 cm<sup>-1</sup> = 7190 cm<sup>-1</sup> = 1390 nm). The absorption bands in the 1800-2100 nm range correspond to the shoulder of the strongly absorbing combination band at 2300 nm which is from the coupling between the CH stretching band at the 3000-2700 cm<sup>-1</sup> region and CH<sub>3</sub>/CH<sub>2</sub> bending band around 1450 cm<sup>-1</sup> region (e.g. 2870 cm<sup>-1</sup> + 1450 cm<sup>-1</sup> = 4320 cm<sup>-1</sup> = 2314 nm). The spectral variations in the 1800-2100 nm range are more dominant than those in the 1100-1650 nm range.

The spectral features of toluene are significantly different from those of aliphatic hydrocarbons due to the clear structural differences between them. In the case of toluene, the second overtone band of CH stretching (1140 nm) attached to benzene ring has been clearly observed. *n*-Hexane and cyclohexane also show the noticeable spectral differences in the 1800-2100 nm range. Consequently, it is found that there are distinct spectral differences among paraffin, naphthene, and aromatic compounds in NIR region. Since naphtha is a very complex mixture, the absorption bands of all components are overlapped each other with similar molar absorptivity. The spectral features of each component in Figure 1 are very broad compared to those in the mid-infrared region, however the spectral differences are clearly noticeable and demonstrated to be sufficient to build up NIR calibration models.

Even though the distinct spectral variations among paraffin, naphthene, and aromatic hydrocarbons were found, the spectral features of normal and branched paraffin have been expected to be very similar due to the structural similarity of each compound. Normal linear hydrocarbons and C6 structural isomers were selected to examine qualitative spectral difference. For this purpose, NIR absorption spectra for the normal linear hydrocarbons of *n*-hexane, *n*-heptane, *n*-octane, *n*-dodecane, and *n*-hexadecane, and the C6 isomers of *n*-hexane, e.g., 2-methylpentane, 3-methylpentane, 2,2-dimethylbutane, and 2,3-dimethylbutane, have been collected. The NIR spectra of all compounds are shown in Figure 2. All the spectra have been linearly offset for the purpose of qualitative comparison. As shown in



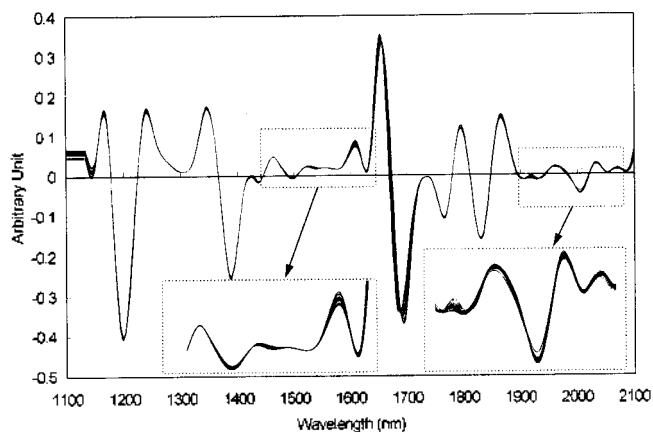
**Figure 2.** NIR spectra (1100-2500 nm) of different chain length of normal linear hydrocarbons and C6 structural isomers.

Figure 2, the overall shapes of all spectra are similar to each other, however, reasonable spectral variations are observed in the 1100-1650 nm and 1800-2100 nm ranges, respectively. Furthermore, the most systematic spectral variations are observed in the 1120-1270 nm range. As the chain length of normal linear hydrocarbons is increased, the  $\text{CH}_2$  second overtone band at 1210 nm becomes more dominant while the  $\text{CH}_3$  second overtone band at 1194 nm becomes less dominant. In *n*-hexane spectrum, both 1194 and 1210 nm bands are observed, however, the 1194 nm band is clearly decreased in the spectrum of *n*-hexadecane. On the contrary, between C6 isomers, as increasing the number of  $\text{CH}_3$  groups (more branched structure) in a given molecular weight,  $\text{CH}_2$  second overtone band at 1210 nm becomes less dominant while  $\text{CH}_3$  second overtone band at 1194 nm becomes more dominant. Similarly in a *n*-hexane spectrum, both 1194 and 1210 nm bands are observed, however, 1210 nm band is clearly decreased in the spectrum of 2,3-dimethylbutane. In addition, the other minor systematic spectral variations are observed in the 1350-1650 nm range as the difference of ratios between  $\text{CH}_2$  and  $\text{CH}_3$  in C6 isomers increases. In C6 isomer spectra, considerable spectral variations are observed from 1800 to 2100 nm range, however no systematic variations based on the ratio between  $\text{CH}_2$  and  $\text{CH}_3$  are observed in normal linear hydrocarbons.

As shown in Figure 1, naphtha is a very complex mixture and the corresponding NIR spectrum is highly overlapped; however unique spectral features can provide selective information for the quantification of each component. Obviously, instead of simple regression methods, a multivariate calibration method is required to resolve these mixtures due to the highly overlapping spectral features.

**PLS Calibration Models.** PLS (partial least squares) regression is the method of reconstructing the spectral data to give the greatest variation related to analyte. It involves projecting spectral data onto a new coordinate defined by factors (eigenvectors). The set of variables (scores) in the new coordinate system is then used for calibration. Factors in PLS regression are being built using both spectral data and concentration information to give the greatest spectral variation and best correlation with concentration; therefore, PLS can extract the analyte-related information and generate a calibration model even in highly complex mixture by excluding irrelevant information and spectral noise.

For PLS calibration, a total of 50 spectra from 50 naphtha samples were divided into 35 spectra for calibration and 15 spectra for the prediction data set serving as an independent validation set. Spectra in the prediction set were totally independent from the calibration set and randomly chosen. Before utilizing the PLS algorithm, all of the raw absorbance spectra were converted to their second derivative spectra. It is well known that second derivatization of raw spectra helps to enhance the spectral features and remove (or at least reduce) baseline variation.<sup>7</sup> Figure 3 shows all 35 second derivative spectra in calibration data set. The spectral variations are effectively enhanced while keeping the same qualitative information and removing the baseline variations. The spectral variations from 1450 to 1640 nm and 1900 to 2080 nm ranges are enlarged in the same figure. Noticeable spectral variations in terms of position and intensity of peaks are observed around major bands in the 1120-1300, 1350-

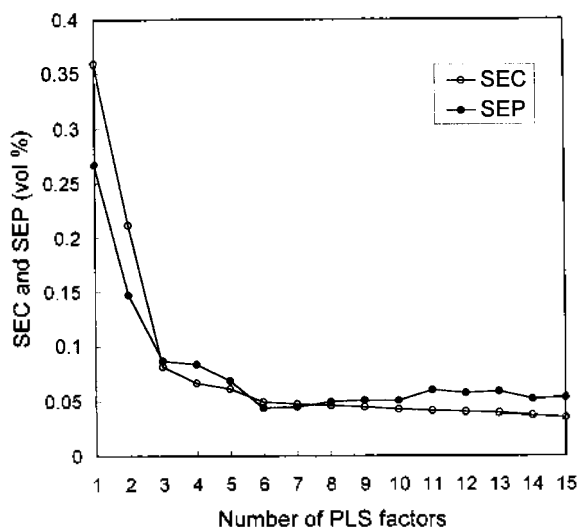


**Figure 3.** Second derivative spectra of all 35 samples in calibration data set. The 1450-1640 nm and 1900-2080 nm regions are enlarged.

1650, and 1850-2100 nm ranges. The 1650-1800 nm range shows significant spectral variations, however, it contains no useful spectral information due to the strong saturation of the NIR radiation from a long optical pathlength.

In the application of the PLS algorithm, it is generally known that the spectral range and number of PLS factors are critical parameters. It has been previously determined that the calibration performance depends on the spectral range utilized.<sup>8-9</sup> Five different spectral ranges of 1100-1340, 1340-1650, 1800-2100, 1100-1650, and the combination of 1100-1650 and 1800-2100 nm were examined in this study. These spectral ranges were selected to include combinations of the three major absorption regions (1100 to 1340, 1340 to 1650, and 1800 to 2100 nm).

The optimum number of factors can be identified as the number of factors that gives a minimum standard error of prediction (SEP) from a separate set of validation samples not present in the calibration set. The SEP is an excellent index to represent the ability of a calibration model to



**Figure 4.** The effect of number of PLS factors on standard errors of calibration (open circle) and prediction (filled circle) for total aromatic content in the 1100-1650 nm combined with 1800-2100 nm range.

**Table 1.** PLS calibration results for each component at different spectral ranges

Range		1100-1340 (nm)	1340-1650 (nm)	1800-2100 (nm)	1100-1650 (nm)	1100-1650 & 1800-2100 (nm)
	factors	8	7	8	8	6
Paraffin	SEC (vol%)	0.14	0.09	0.12	0.11	0.10
	SEP (vol%)	0.12	0.11	0.21	0.10	0.10
	factors	8	8	8	9	6
Naphthene	SEC (vol%)	0.12	0.08	0.12	0.08	0.09
	SEP (vol%)	0.13	0.09	0.20	0.08	0.10
	factors	7	7	7	6	6
Aromatic	SEC (vol%)	0.06	0.06	0.05	0.07	0.05
	SEP (vol%)	0.04	0.04	0.05	0.05	0.04

predict analyte concentrations independently. Figure 4 shows standard error of calibration (SEC, open circle) and SEP (filled circle) plotted as functions of the number of PLS factors used for the determination of total aromatic content within the 1100-1650 nm combined with 1800-2100 nm range. As expected, both the SEC and SEP decrease sharply with the initial factors and gradually decrease as more analyte-dependent spectral variation is incorporated into the calibration model. As the number of factors increase further, the SEC continues to drop slightly while the SEP begins to increase at 7 PLS factors. An increase in the SEP indicates that the data has been over-fitted by incorporating spectral variation into the calibration model that is not present in the prediction data set. In this case, therefore, 6 PLS factors were chosen as the optimum. The optimal number of PLS factors for other spectral ranges and other components is determined by generating and examining the same type of plot.

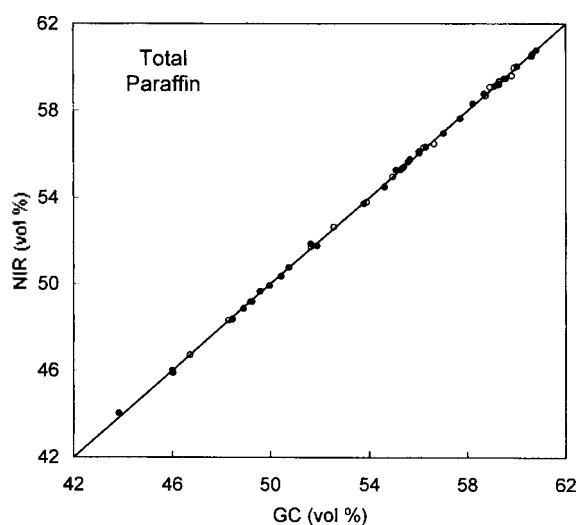
The results of the PLS calibration employing second derivative spectra for each spectral range are summarized in Table 1. For total paraffin, the 1100-1650 nm region combined with 1800-2100 nm provided the best calibration model. The spectral features of paraffins are slightly different each other, however they are fairly different from those of naphthene and aromatic compounds in many areas of the spectrum; therefore, the inclusion of wider spectral range provides more paraffin-related information in the calibration model. The calibration model using the 1100-1650 nm range shows similar results, however it requires 8 PLS factors even with narrower spectral range. The calibration model using the 1100-1650 nm region combined with 1800-2100 nm is superior to that with 1100-1650 nm region because it shows similar calibration performance by incorporating more spectral information with less factors. Using only small spectral regions does not seem to give acceptable calibration results for paraffin, since each spectral region shows only limited spectral differences between paraffins and some of the other components.

For total naphthene, similar to total paraffin, the calibration models using 1340-1650 nm, 1100-1650 nm, and 1100-1650 nm region combined with 1800-2100 nm show similar calibration results. The spectral features of naphthene are clearly different from those of paraffin and aromatic compounds throughout NIR region; therefore, those spectral ranges provide similar calibration results. The calibration model using the 1100-1650 nm region combined with 1800-2100 nm is selected as the best and robust calibration model

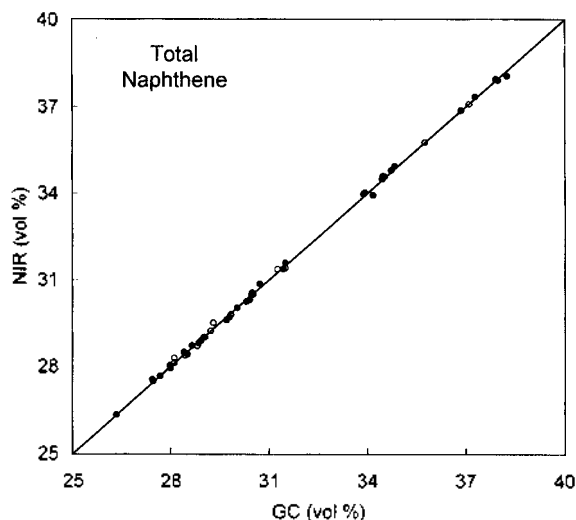
because it uses fewer number of factors without seriously sacrificing the accuracy of model. The calibration model using 1100-1650 nm range shows slightly better performance in SEC and SEP, however it requires 9 factors even with less amount spectral data which could be too tightly fitted and sensitively influenced by even minor sample matrix change.

In contrast to the cases of paraffin and naphthene, the calibration of total aromatic shows similar results for all five different spectral ranges. For example, using the 1800-2100 nm range provides an excellent calibration model due to the unique and distinguishable spectral features of aromatic compounds compared to other aliphatic hydrocarbons in this range. There are clear qualitative differences in all of the NIR spectral regions which lead to the acceptable calibration performance regardless of region or combination of regions used. However, the calibration model using the 1100-1650 nm region combined with 1800-2100 nm is selected as the best calibration model because it requires less number of factors and both SEC and SEP are small. The calibration models using 1100-1340 nm and 1340-1650 nm show excellent prediction ability, however, corresponding SECs are high relatively. The SEC and SEP of an optimal calibration model should be small as well as similar in scale.

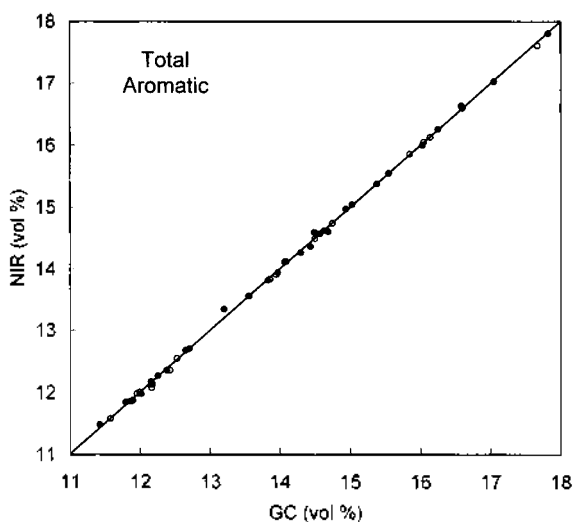
In overall, a wider spectral range gives better calibration



**Figure 5.** Scatter plot showing correlation between NIR and GC analyses for total paraffin content. Filled and open circles correspond to calibration and prediction data, respectively.



**Figure 6.** Scatter plot showing correlation between NIR and GC analyses for total naphthene content. Filled and open circles correspond to calibration and prediction data, respectively.



**Figure 7.** Scatter plot showing correlation between NIR and GC analyses for total aromatic content. Filled and open circles correspond to calibration and prediction data, respectively.

models by incorporating more spectral information with fewer number of factors. Scatter plots showing the correlation between NIR and GC analyses for the best selected calibration models of each component are shown in Figure 5 through 7. Filled circles and open circles represent the calibration and prediction data, respectively. In all cases, the

calibration and prediction data has an excellent correlation with conventional GC data and many points fall on or close to the unity line. Concentrations in the prediction data set were accurately predicted with SEP of 0.10 vol% for total paraffin, 0.10 vol% for total naphthene, and 0.04 vol% for total aromatic content.

### Conclusion

The total paraffin, naphthene, and aromatic contents in naphtha have been successfully measured by NIR spectroscopy less than two minutes. Using NIR spectroscopy, even minute spectral differences among paraffin, naphthene, and aromatic hydrocarbons can provide the enough quantitative information for the calibration using PLS regression. Using wider spectral range gives better calibration models by incorporating more spectral information with fewer number of factors. The NIR prediction results showed excellent correlation with those of GC. Thus, the conventional GC has been successfully replaced by NIR spectroscopic method for the quantitative analysis in naphtha. With the help of fast NIR analysis, it can be expanded to the excellent process control which is more accurately and rapidly reflecting the state of the process at a given time. Future work will investigate NIR spectroscopy for the determination of detailed compositions in naphtha including branched paraffin content based on carbon chain length.

### References

1. Wiseman, P. *An Introduction to Industrial Organic Chemistry*; Applied Science Publishers: London, U. K., 1976; p 193.
2. Leveque, R. E. *Anal. Chem.* **1967**, *39*, 1811.
3. Kosal, N.; Bhairi, A.; Ashraf Ali, M. *Fuel* **1990**, *69*, 1012.
4. Burns, D. A.; Ciurczak, E. W. *Handbook of Near-Infrared Analysis*; Marcel Dekker: New York, U. S. A., 1992.
5. Wetzel, D. L. *Anal. Chem.* **1983**, *55*, 1165A.
6. Martens, H.; Naes, T. M. *Multivariate Calibration*; John Wiley and Sons: New York, U. S. A., 1989; p 116.
7. Thomas, E. V.; Haaland, D. M. *Anal. Chem.* **1990**, *62*, 1091.
8. Chung, H.; Arnold, M. A.; Rhiel, M.; Murhammer, D. W. *Appl. Spectrosc.* **1996**, *50*, 270.
9. Chung, H.; Lee, J. S.; Ku, M. S. *Appl. Spectrosc.* **1998**, *52*, 885.