Gasoline Blending Using Near-Infrared Analysis - Successful and Quick Implementation

Larry McDermott and Martha Ranc, Hamilton Sundstrand, USA, Tel: 909-593-3581 Fax: 909-392-3207 Email: ait@hs.utc.com Web Address: www.hs-ait.com

Near infrared analysers have been applied to numerous hydrocarbon processing applications over the past twenty years[1-9] including real-time gasoline and diesel blender analysis and analysis of feedstock, intermediate and finished products of hydrocarbon processing units. We will show that the startup time for a NIR blender project can be significantly reduced through the use of existing calibration sample databases that include analysed samples from global sources.

Near Infrared Technology

Near infrared (NIR) spectroscopic analysers have several attractive attributes that have lead to their widespread use in refineries. Spectroscopic techniques like NIR and FTIR are rapid, providing a typical analysis cycle time of 1 to 2 minutes. The accuracy of chemometric-based analysers is typically high and often only limited by the accuracy of primary method. A solid-state analyser, such as presented in this paper, provides very high precision because there is minimal or no operator interaction. In addition, the optical components of solid-state analysers are very stable and reliable with demonstrated up-times of close to 100%. Fiber optic cables are often used allowing for the placement of the analyser and electronics in a general-purpose area, reducing the equipment required to be installed in the hazardous sampling area. NIR analysers are also highly cost effective because one analyser often analyses from 2 through 15 properties, and one analyser can be optically or stream multiplexed to allow for the analysis of multiple streams.

Near infrared analysers are calibrated by developing a mathematical equation that relates the NIR spectrum to measured properties of interest (typically measured using ASTM or IP methods). This process is referred to as chemometric model development. In order to provide reliable results, the calibration equations must include a wide range of samples, covering the chemical and physical property ranges expected to be seen under production conditions.

Gasoline Blending

Gasoline blending requires a complex balance of feed stocks and blending components to meet regulatory and product specification requirements at the lowest possible cost. In a typical gasoline blending operation, between five and nine component streams are blended to produce the finished gasoline. A typical product blend may run for 4 to 30 hours. A refinery may also have several blend headers producing different grades of gasoline simultaneously. A medium complexity refinery may routinely work with more than 15 different products, each of which typically has a different formulation recipe. A schematic diagram of a seven-component blender is shown in Figure 1. Traditional blending operations calculate flow rates of the blend components based on laboratory sampling of component tanks. Often the component tanks are not static; as the blend component is withdrawn, additional product is added changing the composition. Real-time analysis of the blended product allows refiners the opportunity to immediately react to changes in component streams. An additional benefit of NIR based analysers is that real-time analysis and blend optimization offers the opportunity to eliminate or minimise the costly inventorying of finished product by blending directly to tankers or pipelines. For on-line analysis, a slipstream is typically installed at the output of the blender. This small stream of gasoline, typically 1/2" pipe, flows to a sample conditioning system where the flow and temperature is regulated and water and particulates are removed. The product stream then flows through a transmission cell for analysis. The flow cell is coupled to the analyser using fiber optic cables. In the example shown, the NIR analyser is placed in a control room or other general purpose area. The sample conditioning system and sample cell are placed in a hazardous area, typically ATEX Zone 1. A general purpose NIR analyser (Hamilton Sundstrand PIONIR MVP) is used in this example. An industrial version of this analyser, the PIONIR 1024 is also available for installation in hazardous areas such as ATEX Zone 1 and 2 or NEC Class I Division 1 or 2.
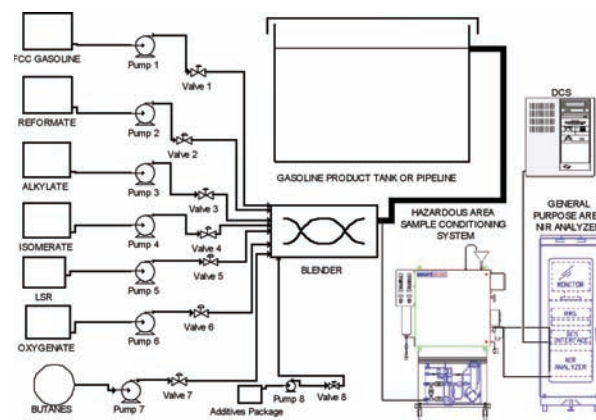

Figure 1. Schematic of a seven-component blender.

Properties analysed

The most common properties NIR is used for in gasoline blending are the Research Octane Number (RON) and Motor Octane Number (MON). Other properties routinely analysed simultaneously with the octane numbers include Aromatics, Olefins, Oxygenates (individual or speciated), Benzene, Distillation Points, Reid Vapor Pressure, and Gravity among others.

Table 1 shows some typical properties and accuracies obtained for blended gasoline analysis. The SEP is a one-sigma standard error of prediction for the PIONIR. For ASTM this is the one-sigma reproducibility. This table shows that the accuracy of NIR methods often matches or approaches the accuracy of the primary ASTM method used to calibrate the NIR. Individual refinery results may be more accurate than the general results reported below when calibration models are developed for the customer's specific types of fuel.

Table 1. Typical Properties analysed by NIR

Property	Units	ASTM Method	SEP ASTM[a]	PIONIR
RON	ON	D2699	0.25	0.35
MON	ON	D2700	0.32	0.32
Aromatics (oxygenated)	vol%	D1319	1.1	1.2
Aromatics (non-xygenated)	vol%	D1319	1.3	1.2
Olefins[b]	vol%	D1319	$0.30X^{0.6}$	1.6
Saturates (oxygenated)	vol%	D1319	1.5	1.8
Saturates (non-xygenated)	vol%	D1319	2.0	1.8
MTBE[b]	vol%	D5599	$0.04(X^{0.67})$	0.36
Benzene[b]	vol%	D3606	0.05X+0.02	0.25
API Gravity	°API	D287	0.18	0.4
Sp. Gravity		D1298	0.0001	0.004
RVP	psi	D323	0.27	0.4
RVP	kPa	D323	1.9	2.8
IBP	°F	D86	5.5	8.0
T10[c]	°F	D86	1.95+0.95*S	7.6
T50[c]	°F	D86	1.88+1.43*S	6.2
T90[c]	°F	D86	1.30+0.91*S	6.0
End Point	°F	D86	6.8	8.0

[a] The ASTM SEP is calculated from the reported reproducibility (R); SEP = R / 2.77.
[b] X is the average of the values being compared.
[c] S is the slope.

Analyser Calibration

Multivariate calibration techniques are used to develop models that relate the NIR spectral data to the laboratory values of properties as determined by the approved ASTM, Institute of Petroleum (IP), or equivalent method. Chemometric models developed using algorithms such as Partial Least Squares (PLS) and Principal Components Regression (PCR) provide outlier statistics, which are used to insure that future samples analysed with the analyser are statistically similar to those used to calibrate the system. These statistics can also be used to automatically capture or flag "outlier" samples, which are used to improve the calibration models over time.

Key factors that impact the PIONIR or any other NIR analyser's accuracy include:
1) The strength of relationship between the spectral response and the laboratory data. For well-proven applications such as octane number, there is a strong correlation between the NIR response and the octane number. Not all properties can be accurately analysed using NIR. For example, NIR is not well suited for the analysis of low-level components such as additives or sulphur.
2) The instrument and sampling error. If the analyser is installed correctly and the sample conditioning system is properly maintained, these error sources are minimised.
3) The relevance and thoroughness of the calibration set. In the calibration phase where the analyser is "taught", a thorough set of samples covering the range of possible concentration and property ranges must be examined. Because there are so many variations in product formulation and blend components, the ability to attain a wide range of samples for calibration development is often the largest contributor to the timeline of NIR implementation.
4) Since the NIR analyser is calibrated against laboratory-determined values, the accuracy of the primary sample values is critical. Some customers choose to improve the accuracy of the samples used to develop NIR calibrations by performing the primary method analysis several times. By repeating the test three times, without biasing the operator to the "expected" results, and averaging the three results, the accuracy may improve by approximately a factor of 1.7.
5) Sample integrity. The samples to be scanned on the NIR must be the same as when they were run on the primary test. Our experience is that samples that are refrigerated and stored in dark bottles or cans with a good seal to prevent evaporation are usually stable for months. If the sample has been allowed to evaporate, to photo-oxidize, to be contaminated or otherwise change, we are introducing additional error to the NIR calibration.
6) Calibration Development. The process of chemometric model development includes many variables including numerous data processing techniques, numerous linear and nonlinear algorithms, evaluation of sample and sample set quality, evaluation of the range of applicability of models (or product specific models) and evaluation of the rank of the model to provide optimum results. In many cases, the refinery will contract the analyser vendor or a consultant to develop the models.

In our experience, the collection of a diverse and representative calibration set for a specific refinery is the dominant factor that impacts the performance and the timeline. The other factors mentioned can be optimized by properly installing the system and implementing standard operating procedures that the refinery staff can follow. For some customers, it may take 3 to 12 months of sample collection to obtain an appropriate calibration set. Suggestions to shrink the implementation timeline are made in the next section.

Currently, the analytical laboratory typically analyses between 1 and 3 finished gasoline samples over the course of a blend to insure blend product specifications are being met. These samples are typically analysed for the properties of interest, for example octane numbers, as well as a variety of other measurements. Project schedules for complex blender analyser systems often include a three to five month period between the awarding of the contract and the

delivery and physical installation of the system. In order to speed the calibration development, we have used this time lag to build up calibration data sets by either running these routine samples on a laboratory version of the process analyser, or by retaining a small volume (75-150 mL) of the routine samples that can be run when the analyser is delivered. It is essential that each sample have a unique identification, typically from a Laboratory Information Management System (LIMS), that can be used to mate the laboratory-determined values with the spectrum. Running the samples on a laboratory NIR analyser is a simple process that takes less than 5 minutes per sample. If a large set of samples has been properly saved for running with system delivery, we have been able to run from 80 – 120 samples a day depending on the system layout. For a refinery producing two blends a day over a typical four month system delivery period this results in an "instant" set of more than 200 site calibration samples for calibration development. Another option would be to ship the retained samples to the vendor where they can be scanned in the laboratory. Because of the patented instrument matching technology employed by the PIONIR and MVP systems (AVI™), data can be seamlessly transferred between analysers.

Another approach that some of our customers have used to develop a diverse calibration set involves preparing samples in the laboratory by manually blending samples of your blend components to simulate the finished products. By knowing the blend formulas expected, an experimental design can be developed to minimise the number of blends required to be prepared. The prepared samples are scanned by the NIR analyser and tested by the laboratory reference methods.

ASTM Validation

The ASTM has released one key method and one key standard practice for implementing NIR and infrared analyser systems. The *Standard Practice for Infrared Quantitative Analysis*, (E1655) recommends procedures for calibrating systems. The *Standard Practice for the Validation of Multivariate Process Infrared Spectrophotometers*, (D6122) is designed to provide control charting tools and statistical analysis of on-line systems and allows users to separate modeling problems from analyser system problems. The PIONIR 1024 and MVP analyser systems are typically calibrated using the recommendations in E1655. An optional automated analyser validation system, or manual sample introduction can be used to meet the requirements of D6122. Several methods have also been approved for FTIR analysis including the *Test Method for Benzene in Motor and Aviation Gasoline by Infrared Spectroscopy* (D4053) and the *Test Method for the Determination of MTBE, ETBE, TAME, DIPR, Methanol, Ethanol and tert-Butanol in Gasoline by Infrared Spectroscopy* (D5845).

Quick Implementation

The Hamilton Sundstand PIONIR 1024 and MVP analysers have been used in refinery applications including blender analysis since 1993. We have accumulated over this time a vast database of samples (spectra + analytical values) from global sources including over 10,000 blended gasoline samples. This allows us to develop models for customers that will have a reasonable accuracy for many properties without extensive site sample selection. Table 2 shows the ranges contained in the database for some routine properties. A trade-off between the accuracy of a chemometric model and the diversity of samples in the model often exists. In order to optimise the models for a refinery, we select a subset of the full database based on product specifications for gasoline made at the customer site. For example, if the refinery does not blend MTBE or TAME we would not use samples containing MTBE or TAME from our database for the site models. Similarly if the refinery only makes gasoline with a RON between 91 and 97, we would exclude samples with a RON less 90 or greater than 98.

Table 2. Property Ranges in the Database.

Property	Units	Min	Max
RON	ON	71.2	100.9
MON	ON	69.1	88.7
IBP	°F	63	120
T10	°F	89	169
T50	°F	126	314
T90	°F	233	407
EndPoint	°F	291	475
RVP	kPa	29.6	110.3
AROMATICS	vol%	2.3	53.8
OLEFINS	vol%	0.4	45.5
BENZENE	vol%	0	5
ETOH	vol%	0	11

Based on blend formulas provided by the refinery, we selected 1368 samples from our database to develop models for the site. These spectra were imported into a Pirouette (Infometrix, Inc., Woodinville, Washington, USA), a leading chemometric modeling software package. The site collected an additional 183 samples for model development. These samples were also imported into Pirouette for evaluation. Figure 2 shows the NIR spectra of our database samples (in blue) along with the site samples (in red). The baseline variation observed in the spectra was removed by using a simple Savisky-Golay 9-point first derivative.
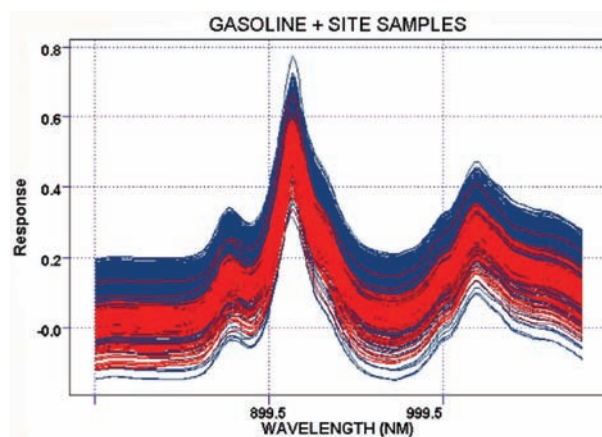

Figure 2. Spectra of samples. The blue spectra are from the database, the red spectra are from the site.

The site samples were evaluated using Partial Least Squares regression analysis. Figure 3 shows several of the graphical tools available to evaluate samples. Two samples were flagged as outliers, meaning they are suspect. One sample failed the statistical outlier tests based on its very high leverage. A second sample had an abnormally large disagreement between the measured RON and the predicted RON using the model developed. Based on the outlier evaluation these two samples were eliminated. One had a bad spectrum, and we suspect the other sample was either a bad octane engine analysis or a transcription error of the measured RON. The balance of the spectra (181 site samples) were judged to be of high quality and retained for analysis.
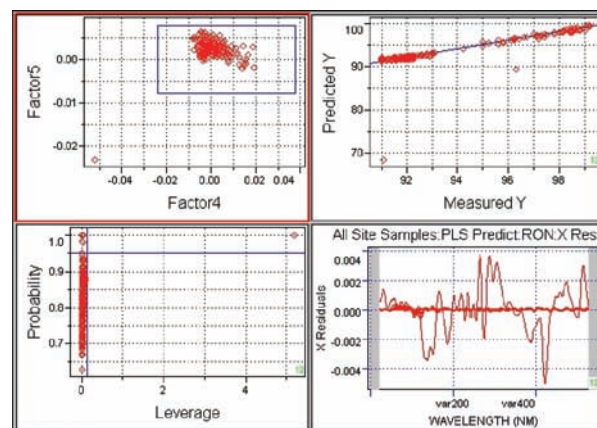

Figure 3. Graphical statistical analysis tools.

The 181 site samples were collected over a long time frame (six months). A PLS calibration model was developed for only the site samples to determine a baseline for the accuracy of the RON measurement. The standard error of prediction based on a cross validation of the site samples was 0.25 RON. Cross validation is a common technique to validate models. One sample is iteratively excluded from the calibration set and a model is developed for the remaining samples. The excluded sample is then predicted as a true "unknown" sample. The standard error of prediction (SEP) obtained in this method is an estimation of the accuracy that will be attained for similar samples. The SEP obtained for the site samples matches the reported accuracy of the knock research octane engine, which is the best accuracy that can be demonstrated. A plot of the model estimated RON versus the laboratory RON is presented in Figure 4.
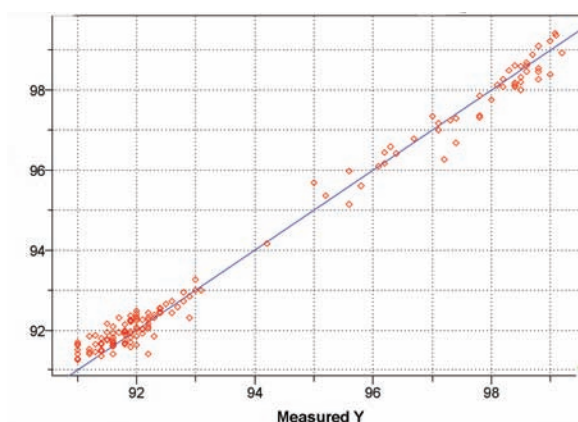

Figure 4. Research Octane Number.

Successful transfer of models

To evaluate the "out of the box" performance of the models based on Hamilton Sundstand's database for the site samples, the site samples were analysed using a PLS model developed on 1368 database samples. The expected SEP of this broad model was 0.40 RON, and the SEP obtained for the site samples was 0.38 RON. Considering the breadth of the calibration set and the fact that no site samples were included, this demonstrates a strong ability to transfer models.

Because there is often a tradeoff between the breadth of application of a model and the accuracy, we chose to select samples from our database that were most similar to the site samples. This was done by using qualitative analysis tools available in Pirouette. A principle components analysis (PCA) was performed on the 1368 database samples and the 181 site samples. A useful output of PCA analysis that is frequently used for qualitative data analysis are scores plots. In scores plots the principal component axes represent a compressed view of the multivariate data. The scores are a mapping of the original sample data onto a coordinate system specified by the loadings. Similarities between samples can be clearly seen in these plots, and we used them to only select calibration samples from our database that are similar to the site samples. Samples that were drastically different (far away on the plot) from the site samples are very different, and may contain components not used at the site such as oxygenates. Figure 5 shows a scores plot for the global database samples in blue and the site samples in red. We selected the samples approximately encompassed in the three-dimensional sphere containing the site samples. Figure 6 shows the result of this sample selection. By excluding samples dramatically different from the site samples, we expect to obtain more specific and accurate models.
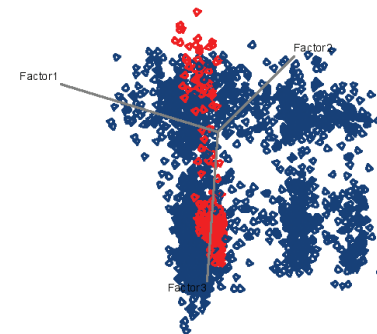

Figure 5. Scores plot of global database samples (blue) with site samples (red)
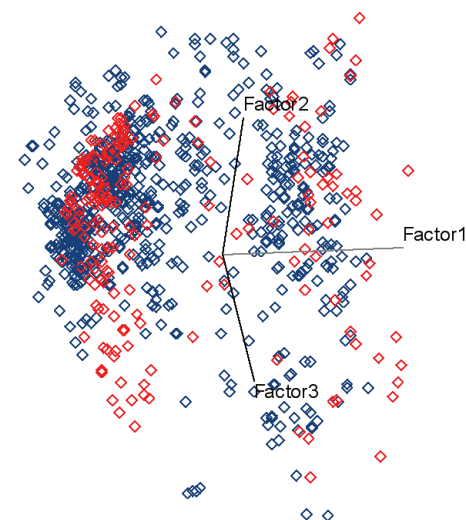

Figure 6. Scores plot of localised database samples (blue) with site samples (red).

By evaluating the PCA scores plots, the 690 samples in the database most similar to the site samples were selected. A PLS model was developed for these 690 database samples, and the 181 site samples were predicted with a resulting SEP of 0.39 RON, resulting in no improvement over the 0.38 SEP observed using the broader model.

A subset of the site samples (37 of 181) was selected to cover the range of RON, and merged with the 690 database samples. The PLS model developed using this calibration set was used to analyse the remaining 144 site samples. Inclusion of these 37 site samples reduced the SEP obtained for the "unknown" site samples to 0.30 RON. For many refineries, this would be acceptable performance.

In order to attempt to improve the accuracy of the models, a PLS model was developed containing only the regular grade gasolines (RON less than 95.5). This model contained 490 database samples plus 25 site samples. The SEP obtained on prediction of the remaining 108 site regular grade samples was 0.26 RON, equal to the performance (0.25 RON) obtained for the site specific model.

The full set of site samples (181) was merged with the 690 database samples and a PLS model was developed. The SEP for the site samples in this model was 0.30 RON. The fit between the laboratory RON and the estimated RON from the cross-validated model is shown in Figure 7. The database samples are plotted in blue and the site samples are plotted in red. Inclusion of external database samples with the site samples expands the applicability of the model, minimising the chance of future samples occurring as outliers, which would result in larger errors. Table 3 summarises the model results discussed here.
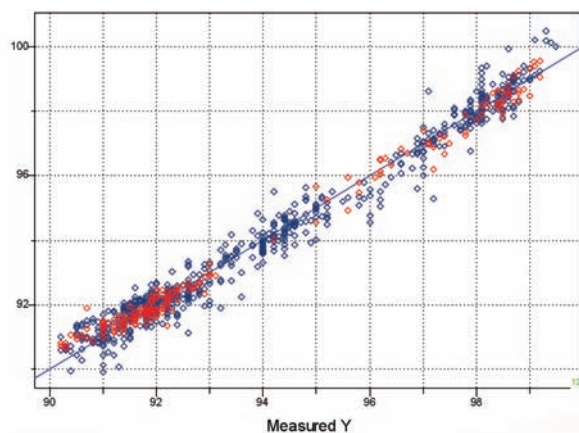
Figure 7. Estimated versus Laboratory RON. The blue samples are from the database while the red samples are from the site.

Table 3 shows the results obtained for the accuracy of analysis of site samples using different data sets in the model. It can be seen that the use of our database samples with the inclusion of only 25 site samples (for regular grade gasoline) results in models that perform as well as more extensive site specific models that take significantly longer to attain.

Table 3. Summary of Model Combinations

Model	Calibration	Database Samples	Site Samples in model	Time to Results	SEP of site samples
1	Only Database	1368	0	0 – pre-existing data	0.38 RON
2	Database samples similar to site	685	0	6 months	0.30 RON
3	Database + 37 site	685	37	1 month	0.30 RON
4	Database + 25 site, regular grade only	690	25	1 month	0.26 RON
5	Database + 181 site	685	181	6 months	0.30 RON
6	Site only samples	0	181	6 months	0.25 RON

Additional properties were evaluated using the models developed from the database samples plus 37 site samples. The results are presented in Table 4. The reported line sample SEP is based on independent analysis of the remaining 144 site samples.

Table 4. Analysis of Additional Properties

Property	Model Error (SEP)	Line Samples SEP
RON	0.40	0.30
MON	0.39	0.30
50% Distillation (T50)	5.5 $^\circ$F	5.3 $^\circ$F
Benzene	0.17%	0.18%
Aromatics	1.45%	1.53%
Olefins	1.7%	1.4%

Customer Experiences

Some examples of customer experiences using this database-based approach to reduce commissioning and startup time are listed below:

- A European refinery commissioned their multi-stream PIONIR system in under three weeks using gasoline from our database along with retained site samples.
- A Gulf Coast refiner added a new blend component. Using models based on our dataset they were able to get good property estimates as soon as the unit came on-line.
- A Midwest refiner successfully used models based on our database and samples from a sister refinery when implementing an NIR-based blend analyser.
- A Gulf Coast refiner who does not have on-line engines has estimated that the on-line use of the PIONIR reduced the number of reblends by over 75%. They used combined models until sufficient samples had been collected to adequately represent their grade and seasonal changes.

Some additional results of a few users of the PIONIR blend analysis system are listed below:

- A Gulf Coast refiner has sold their on-line knock engines and now relies solely on the PIONIR.
- A European refinery has stated that their unleaded gasoline model has a standard error of prediction of 0.22 RON and 0.19 MON[7].
- A Scandinavian refinery that blends gasoline directly to ships has reported close to 100% uptime for the analysis of 10 properties[8].

Conclusion

Near infrared analysis can provide rapid, accurate real-time analysis of refinery products such as blended gasoline. Feedback control based on analysed values allows refineries to meet economic, regulatory and inventory constraints. Creating calibration models can take up to 6 months. The use of existing database samples for calibration alongwith the use of routine laboratory samples can provide successful, reliable implementation of a NIR-based blend analyser in a reduced time frame (a month or less), which may also significantly reduce the cost of implementing the project.

References:

1) Bonanno, A and Griffiths, P., Journal of Near Infrared Spectroscopy 1, 13-23 (1993).

2) Mayes, D., and Callis, J., Applied Spectroscopy 43, 27-32 (1989).

3) Kelly, J., and Callis, J., Analytical Chemistry 62, 1444-1451 (1990).

4) Kelly, J., et al., Analytical Chemistry 61, 313-320 (1989).

5) McDermott, L., "Cetane Benefits", Hydrocarbon Engineering, July 2003.

6) DiFoggio, R., Sadhukhan, M., and Ranc, M., Oil and Gas Journal, May 3, 1993, 87-90.

7) Centner et.al. "Comparison of Multivariate Regression Techniques Applied to Experimental NIR Data Sets", Applied Spectroscopy, Vol. 54, Number 4, 2000.

8) Uvland, K.A., Oil and Gas Process International, 1997.

9) Vickers, G.H., "On-Line Determination of Naphtha Properties to Control a Refinery Process Using Near Infrared Spectroscopy.", IFPAC Conference, 2001.