Chemometrics

Application Note

Rating Octane of Gasoline by Near Infrared Spectroscopy

Abstract

In the petroleum industry, measurement of the physical properties of hydrocarbons forms the basis of commodity pricing and the assessment of process parameters. The key measurement performed on gasoline is the calculation of octane number. The ASTM standard for reporting this measurement is an internal combustion engine in which octane is measured by interpolating between the nearest standards above and below the unknown sample (1). The procedure is time consuming, involves expensive and maintenance-intensive equipment, requires skilled labor and is not well suited to on-line monitoring.

Prediction of gasoline octane numbers using features in the near infrared (NIR) has been studied by a number of research laboratories. One such study is reported here. The purpose of this study was to evaluate the chemometrics techniques of Principal Components Regression (PCR) and Partial Least Squares regression (PLS) and to assess the accuracy of the predictions as a function of wavelength range and spectral resolution. A plot of all of the spectra is shown in Figure 1.

Experimental

Fifty-seven unleaded gasoline samples were analyzed spectroscopically as described by Kelly *et al*. (2). Spectra were collected on a scanning NIR spectrophotometer at 1 nm resolution in the wavelength range of 900 to 1600 nm. Thus, this data set has 700 independent variables (individual spectral channels) plus one dependent variable (the octane value we wish to model: Pump Octane) which was assembled into an ASCII file.

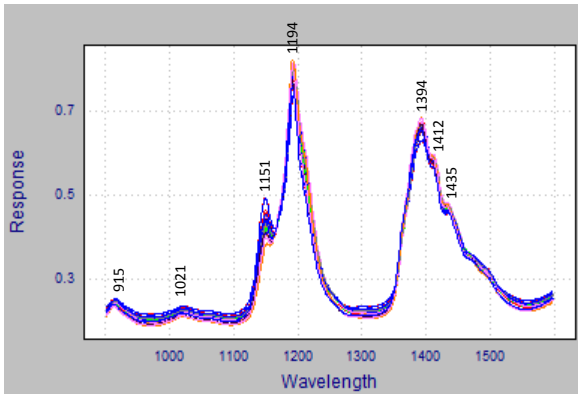

Figure 1. Graphical View of NIR Data (wavelength in nm)

The analysis was performed using *Pirouette 4.5* software running on a Dell dual-core computer running Windows 7.

Table 1. Spectral Peak Properties

Wavelength	Assignment
915 nm	Stretch for CH_2 (third overtone)
1021 nm	Combination bands for CH_2 and CH_3
1151 nm	Stretch for aromatic and CH_3 (second overtone)
1194 nm	Stretch for CH_3 (second overtone)
1394 nm	Combination bands for CH_2
1412 nm	Combination bands for aromatic and CH_2
1435 nm	Combination bands for aromatic and CH

Comparison of PLS and PCR

In Figure 2, the PLS results are displayed with a variance plot in the NW window followed around the clock by the PLS scores, the residuals and the predictions.

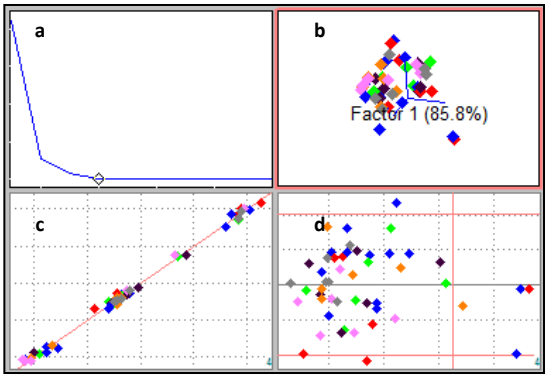

Figure 2. PLS Results for the Octane Data Set. (a) Variance; (b) 3D plot of first three scores; (c) PLS model predictions vs. octane engine reference values; and (d) studentized residuals vs. leverage values

The variance plot in the NW window shows that most of the variance is explained by the first principal component (roughly 87%) as would be expected and that 98.8% of the variance is contained in the first four principal components. The scores plot in the NE window shows a uniform distribution with the possible exception of the three samples at the lower right in the figure. These same three samples display high leverage in the SE window but the predictions seem well enough behaved that we were not concerned with these possible outliers. Finally, the SW window containing the PLS model predictions versus the analytical results from the octane engine show close agreement.

Results of PLS and PCR predictions are compared in Figure 3 by reading the variance and prediction data from the PCR results into the NE and SE windows, respectively.

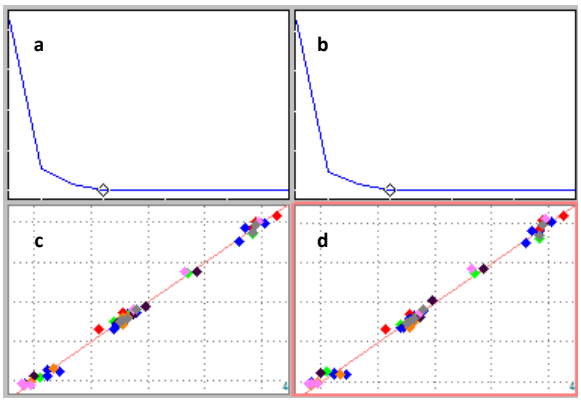

Figure 3. Comparison of PLS and PCR Predictions. (a) and (c) Results from PLS; (b) and (d) Results from PCR; (a) and (b) Variance; (c) and (d) Model predictions vs. octane engine reference values

No appreciable differences were shown to exist between the two methods of model building (confirmed also by leave-one-out cross validation), which is a sign that this model has good promise for field use. Both results were based on models using 4 principal components.

Spectral Range and Resolution

Now that we can show that the NIR spectra have sufficient modeling power to predict the octane ratings of unleaded gasoline samples, we can investigate the effects of wavelength range and resolution on these predictions. The first issue is to compare octane models created using the shortwave NIR to the full range of these data. The subset range to be studied is highlighted in Figure 4.

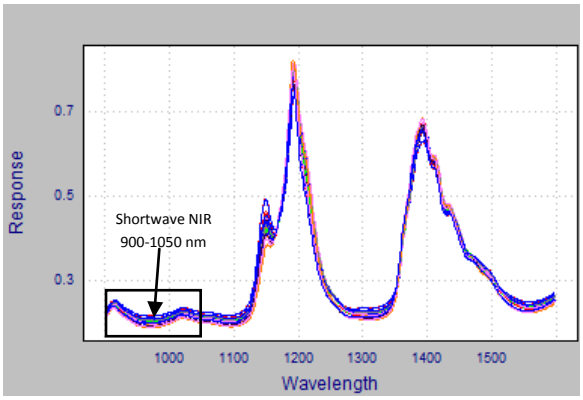

Figure 4. Shortwave region of the near infrared spectra

To develop this comparison between shortwave NIR and the entire wavelength range, we created a subset of the data by excluding all wavelengths above 1050 nm. Excluding every other channel created a 2 nm subset of the same shortwave NIR region (more typical of the silicon based diode arrays in current instrument systems). We similarly generated a 4 nm subset for comparison. Running PLS predictions on all three subsets and contrasting the results to the original PLS results for the entire range at 1 nm resolution creates the picture shown in Figure 5 below.

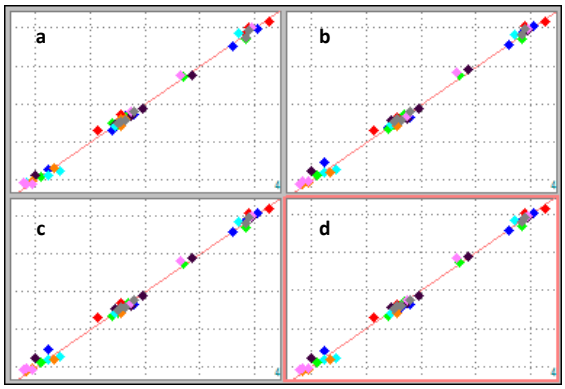

Figure 5. Comparison of intermediate vs shortwave NIR. (a) 900-1600 nm @ 1 nm; (b) 900-1050 nm @ 1 nm; (c) 900-1050 nm @ 2 nm; and (d) 900-1050 nm @ 4 nm

The prediction accuracy is not apparently affected by restricting the wavelengths from the second plus third overtones (900-1600 nm) to only the third overtone region of 900-1050 nm. This shows that the shortwave near infrared, centered around 1000 nm, despite the lower intensity, retains the information about octane level in gasoline. This feature of the third overtone region was also demonstrated by Kelly, et al., who found via stepwise multiple linear regression, that the three best wavelengths for predicting octane level all occur in this region: 895 nm, 932 nm, and 1032 nm (2).

Continuing the resolution study, we returned to the original 900 to 1600 nm range, used exclusion sets to create 10 nm, 20 nm and 40 nm subsets of the original data, and compared the PLS predictions to the original 1 nm data (Figure 6). We found no significant degradation

of the prediction among the resolution simulations spanning 1 nm up to 40 nm resolution.

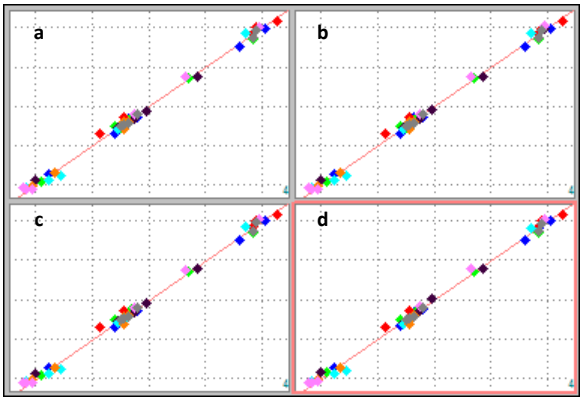

Figure 6. Octane Resolution Study. (a) 1 nm resolution; (b) 10 nm resolution; (c) 20 nm resolution; and (d) 40 nm resolution

Summary

We have seen that PLS and PCR can competently model data collected on unleaded gasoline samples for the purpose of predicting the pump octane rating of gasoline. We have also found that the information contained in the major absorbance bands in the 900 to 1600 nm span is sufficiently retained in the shortwave (900-1050 nm) region. In addition, we have demonstrated for these data that greater resolution is not crucial to the prediction accuracy.

(1) Annual Book of ASTM Standards, (1985) *Vol. 05.04.*

(2) Kelly, J.J.; Barlow, C.H.; Jinguji, T.M. and Callis, J.B. *Anal. Chem.* (1989) *61 (4)*: 313-320.