Ultra-low Sulfur Diesel Classification with Near-Infrared Spectroscopy and Partial Least Squares

4 AUTHORS, INCLUDING:

Robert E. Morris
United States Naval Research Laboratory
73 PUBLICATIONS **394** CITATIONS

Susan Lynn Rose-Pehrsson
United States Naval Research Laboratory
109 PUBLICATIONS **1,376** CITATIONS

Communications

Ultra-low Sulfur Diesel Classification with Near-Infrared Spectroscopy and Partial Least Squares

Jeffrey A. Cramer, Robert E. Morris,* Mark H. Hammond, and Susan L. Rose-Pehrsson

United States Naval Research Laboratory, Code 6181, 4555 Overlook Avenue, Southwest, Washington, District of Columbia 20375

The United States Environmental Protection Agency (EPA) will have regulations in effect no later than 2010 requiring sulfur content to be no greater than 15 parts per million (ppm) for on-road, off-road, and marine diesel fuel applications.[1] Hydrotreatment[2] will remove sulfur, but it also removes other polar compounds that impart fuel lubricity. The rapid and accurate discrimination of ultra-low sulfur diesel (ULSD) fuels is then important for both regulation compliance and lubricity assessment. While near-infrared (NIR) spectroscopy has not yet been able to accurately predict the sulfur content of fuels,[3,4] partial least-squares (PLS)[5] models can be constructed to predict ULSD identity indirectly through the other chemical changes caused by hydrotreatment[6,7] that do, in fact, affect NIR instrument responses, albeit only subtly (see the Supporting Information). Therefore, it is possible to develop relatively low-cost portable NIR field instrumentation for the rapid identification of fuels that have undergone hydrotreatment, which, by virtue of the inevitably low resulting sulfur content, are ULSD fuels.

Data were collected from a set of 391 worldwide diesel fuel samples, consisting of 251 Naval distillate (NATO F-76), 129 marine gas oil (MGO), and 11 ULSD fuels from various North American sources. The non-ULSD fuels had measured sulfur contents ranging from 200 to over 9000 ppm, and the ULSD fuels contained 10 ppm or less sulfur. NIR absorbance spectra were collected from 1000−1600 nm with a fiber optic reflectance probe coupled to a custom Bruker Optics NIR spectrometer, which employed a thermoelectrically cooled 512 element GaAs detector array. Spectra were acquired using custom software written in compiled LabVIEW 8.5 (National Instru-

ments Corporation, Austin, TX). Signal preprocessing and chemometric analyses were performed with in-house algorithms using MATLAB 2008a (The MathWorks, Inc., Natick, MA), using chemometric functionality provided by PLS Toolbox 4.2.1 (Eigenvector Research, Inc., Manson, WA).

NIR absorbance spectra from the 512 element detector arrays were baseline-corrected and adjusted with a wavelength calibration to 600 points to provide a 1 nm resolution from 1000 to 1600 nm. The spectra were normalized to unit length and mean-centered prior to PLS model construction. PLS discriminant[8] models were constructed in MATLAB by correlating the NIR data with a calibration vector of values equal to either −1 (in the case of non-ULSD samples) or 1 (in the case of ULSD samples). This approach produced a qualitative model separating one class of sample (ULSD samples, class 1) from another (all of the other samples, class −1).

A PLS model was constructed on the basis of the first 10 latent variables (LVs) or underlying linear factors derived from the training data. This number of LVs was determined automatically using the *F*-test statistic[4] with an 85% confidence interval, in a manner similar to that previously used in this laboratory to produce fuel property models.[9] Model efficacy was confirmed via leave-one-out cross-validation,[10] which recreates the model without each sample in turn and predicts the classification of each sample without the benefit of its presence in the training data. Predictions made in this manner better demonstrate how the model will function with new incoming samples that are not part of the original calibration set used to construct the model. This model accurately separated the training data into the ULSD (1) and non-ULSD (−1) classes as shown by the boundary at 0 in Figure 1.

The model loadings (or LVs, as described previously) can reveal which variables in the data set have the greatest modeling significance. This is shown graphically in Figure 2, where the NIR calibration spectra were averaged and the portions of the spectrum that are most important to the ULSD modeling are set apart. In this figure, important regions correspond to those areas in which the sums of the absolute values of the model loadings were 50% or more of the maximum value. An additional spectral comparison can be found in the Supporting Information.

* To whom correspondence should be addressed. Telephone: 202-767-3845. Fax: 202-767-1716. E-mail: robert.morris@nrl.navy.mil.

(1) United States Environmental Protection Agency (EPA). *Clean Air Non-Road Diesel Rule Summary.* U.S. EPA, May 2004; EPA420-F-04-029.

(2) Gary, J. H.; Handwerk, G. E. *Petroleum Refining: Technology and Economics*, 2nd ed.; Marcel Dekker, Inc.: New York, 1984; pp 130−136.

(3) Johnson, K. J.; Morris, R. E.; Rose-Pehrsson, S. L. *Energy Fuels* **2006,** *20*, 727−733.

(4) Kramer, K. E.; Johnson, K. J.; Cramer, J. A.; Morris, R. E.; Rose-Pehrsson, S. L. *Energy Fuels* **2008,** *22*, 523–534.

(5) Brereton, R. G. *Chemometrics: Data Analysis for the Laboratory and Chemical Plant*; John Wiley and Sons, Ltd.: Chichester, U.K., 2003; pp 297−313.

(6) Kalichevsky, V. A.; Kobe, K. A. *Petroleum Refining with Chemicals*; Elsevier Publishing Company: Amsterdam, The Netherlands, 1956; pp 275−280.
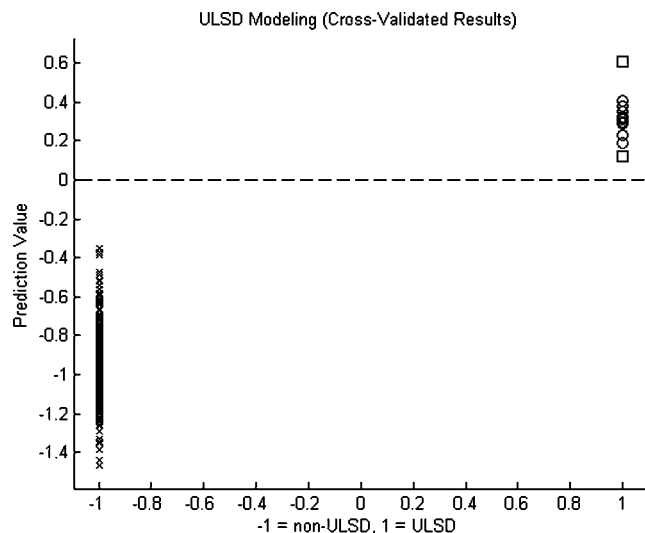
(7) *Advances in Petroleum Chemistry and Refining*; McKetta, J. J., Jr., Ed.; Interscience Publishers, Inc.: New York, 1960; Vol. 3, pp 193−277.

(8) Kemsley, E. K. *Chemom. Intell. Lab. Syst.* **1996,** *33*, 47−61.

(9) Cramer, J. A.; Kramer, K. E.; Johnson, K. J.; Morris, R. E.; Rose-Pehrsson, S. L. *Chemom. Intell. Lab. Syst.* **2008,** *92*, 13−21.

(10) Beebe, K. R.; Pell, R. J.; Seasholtz, M. B. *Chemometrics: A Practical Guide*; Wiley: New York, 1998; pp 93−96.

			Form Approved OMB No. 0704-0188

Report Documentation Page

1. REPORT DATE **2009**	2. REPORT TYPE	3. DATES COVERED **00-00-2009 to 00-00-2009**

4. TITLE AND SUBTITLE **Ultra-low Sulfur Diesel Classification with Near-Infrared Spectroscopy and Partial Least Squares**	5a. CONTRACT NUMBER
	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **United States Naval Research Laboratory,Code 6181,,4555 Overlook Avenue,Washington,DC,20375**	8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**6**	

Figure 1. Cross-validated results of the 10 LV model, showing clear discrimination between the ULSD fuels (○) and F-76 and MGO fuels (×). Two additional ULSD fuels not included in the calibration data (□) were also introduced to the model for validation purposes.
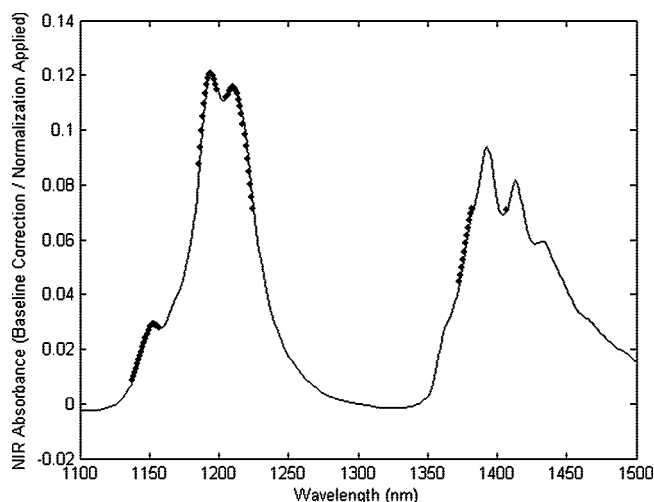
Figure 2. Average NIR calibration spectra showing portions of the spectrum most relevant to ULSD modeling.

The 10 LV PLS discriminant model constructed here effectively discriminates between ULSD and non-ULSD fuels with the original training data, as shown by the cross-validation results in Figure 1. The cross-validation also indicates that the model is capable of classifying new incoming diesel fuel samples correctly. It is possible with large model sizes (i.e., larger numbers of LVs) that resultant PLS models are only as effective as they appear to be with calibration data and that unknown future samples will not be interpreted correctly because of the specificity of the model for the calibration data, a situation known as overfitting. The cross-validation as performed here is evidence against overfitting, but a 10 LV model interpreting 11 ULSD samples may still be approaching the practical limit. However, when two additional ULSD samples not part of the training set are introduced to the model (Figure 1), they are still correctly classified as being greater than zero. This indicates that, despite the model size, overfitting was not occurring and this modeling approach would continue to be effective as a means for practical ULSD detection. A further evaluation of the use of 10 LVs for ULSD modeling as applied to smaller amounts of non-ULSD training data can also be found in the Supporting Information.

From the aggregate results, it has been shown that ULSD fuels can be indirectly identified from general fuel populations by taking advantage of the spectral artifacts produced by the hydrotreatment used to refine ULSD fuel from standard diesel. This identification relies on the construction of a PLS discriminant model that effectively separates ULSD and non-ULSD fuels into two distinct classes. The initial discrimination is entirely effective, and it is anticipated that additional ULSD training fuel samples will increase the robustness of this model toward discriminating between ULSD and high-sulfur diesel fuels. This research is presented here to rapidly disseminate the information to interested parties to keep pace with the 2010 implementation of EPA regulations.

Supporting Information Available: NIR spectra of the high-sulfur diesel and ULSD fuels plotted against each other to illustrate their similarities as well as the number of non-ULSD calibration fuels sequentially reduced to provide additional uncalibrated validation data and thus to illustrate the number of non-ULSD fuel samples that are necessary to obtain a useful discriminant model. This material is available free of charge via the Internet at http://pubs.acs.org.

EF8007739

ULTRA-LOW SULFUR DIESEL CLASSIFICATION WITH NEAR-INFRARED SPECTROSCOPY AND PARTIAL LEAST SQUARES

SUPPORTING INFORMATION

Jeffrey A. Cramer, Robert E. Morris[*], Mark H. Hammond and Susan L. Rose-Pehrsson

U.S. Naval Research Laboratory, Code 6181, 4555 Overlook Ave. SW,
Washington, D.C. 20375

**Ultra-Low Sulfur Diesel (ULSD) and Non-ULSD Spectra Plotted Together**

Figure 1-SI shows the NIR spectra of ULSD and non-ULSD samples plotted concurrently to better show the lack of simple difference between the two sample populations visually. This presents evidence (in addition to that found in the main Communication's Figure 2) indicating that the use of a multivariate analysis technique such as partial least squares (PLS) is a necessary analysis step due to the multivariate nature of the spectroscopic changes produced by the hydrotreatment associated with ULSD production.

Non-ULSD Sample Evaluation

In the Communication, it was shown that ULSD model overfitting is unlikely due to the fact that two ULSDs not included in the calibration data were predicted correctly when introduced to the 10-latent variable (LV) partial least squares (PLS) model. In order to fully evaluate the use of 10 LVs, however, it is also prudent to determine if a 10 LV model can be reproduced using a smaller number of non-ULSD calibration samples than the 384 samples used in the main Communication. In effect, this would determine the rough ratio of non-ULSD and ULSD required for an effective model to be constructed.

Figure 2-SI shows the cross-validated results obtained by sequentially removing increasingly larger portions of the non-ULSD training data. The percentage shown in the Figure

dictates how much of the data was eliminated by random selection (it should be made explicit that the removed samples were reintroduced as validation data to ascertain full model utility). These removal and subsequent evaluation operations were carried out at each percentage through five replicates to minimize random selection errors. What can be seen is that up to 60% of the non-ULSD data can be removed while still preserving the analytic utility of the 10 LV model. Also, although false positives (i.e. non-ULSDs not detected as such) begin to appear when 65% of the data is removed, the number of false positives doesn't become 1% of the total non-ULSD sample population (i.e. about 4 samples out of the original 380 samples) until about 85% of the original non-ULSD samples are removed from the training data. It should also be noted that, during the course of this sequential non-ULSD evaluation and all repetitions, no false negative results (i.e. ULSDs not detected as such) were obtained from either the 11 ULSD samples included in the training data or the two additional ULSD fuels used for model confirmation in the main Communication. Although this shows that model utility can in fact be preserved with much less data than was actually used in the main Communication, the larger amount of training data is used to maintain a prediction model that will remain robust during the analysis of the most diverse populations of fuel samples.
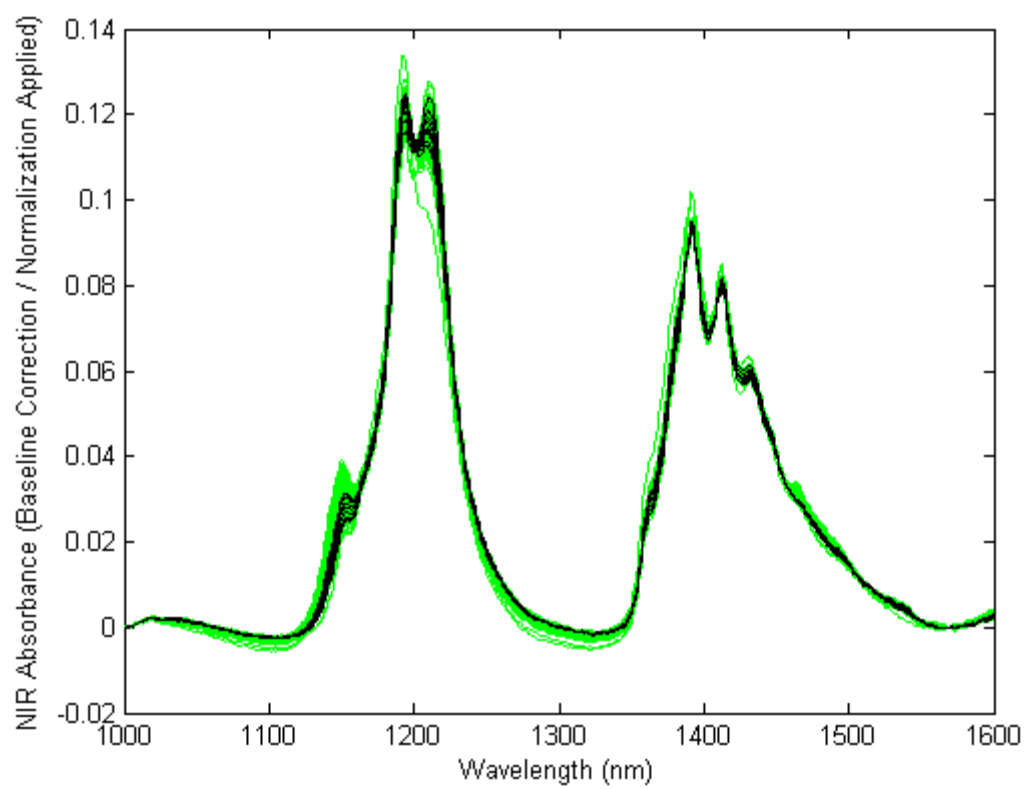
Figure 1-SI. Concurrent plot of ULSD (black) and non-ULSD (green) sample populations.
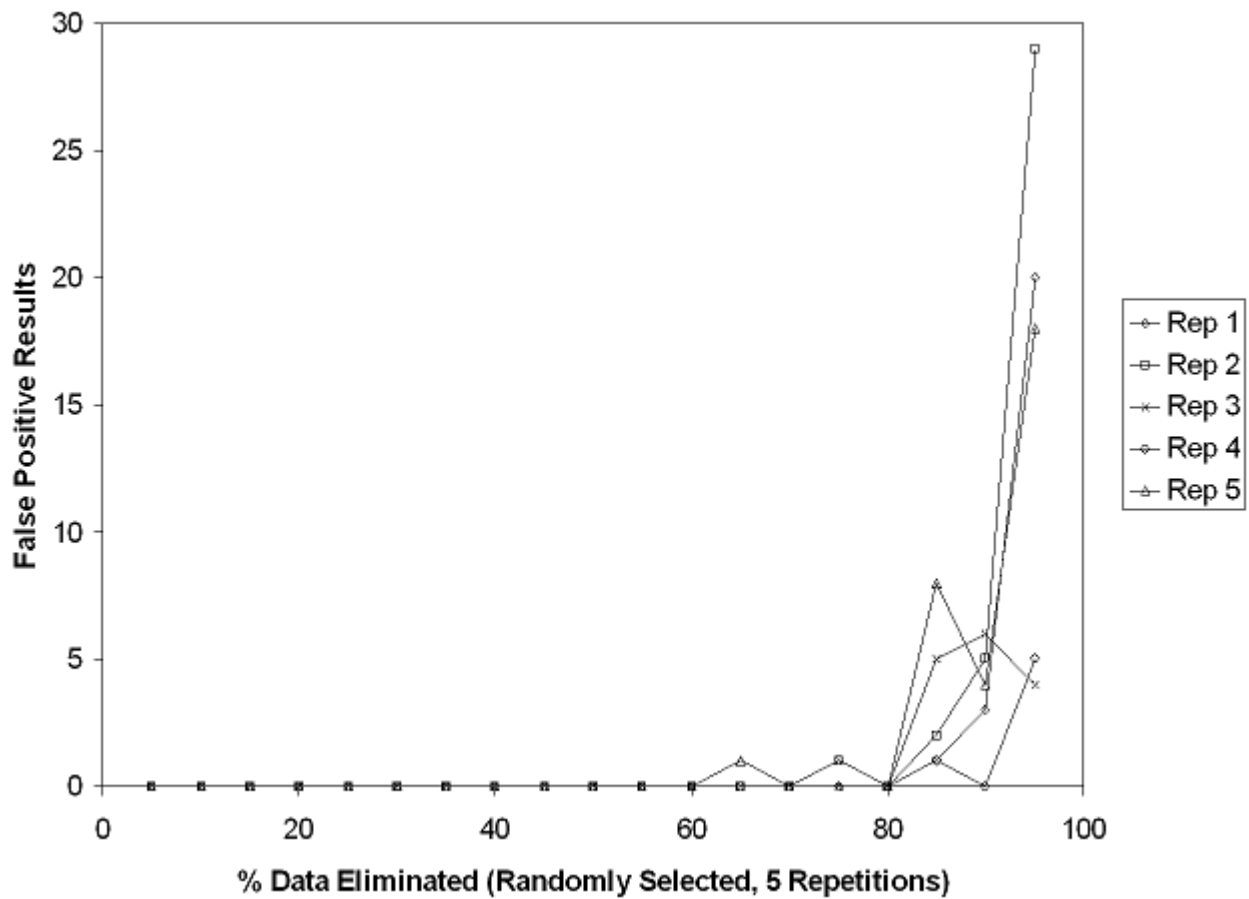
Figure 2-SI. Plot of the number of false positive results obtained from eliminating a certain percentage of the training data (randomly selected, five replicates) used to construct the main Communication's 10 LV model. Note that false positives only begin to appear when 35% of the original data remains.

REFERENCES

* Corresponding Author. Phone: 202-767-3845, fax: 202-767-1716, e-mail: robert.morris@nrl.navy.mil.