

# Prediction of Long and Short Residue Properties of Crude Oils from Their Infrared and Near-Infrared Spectra

PETER DE PEINDER, DEREK D. PETRAUSKAS, FRED SINGELENBERG, FABIEN SALVATORI, TOM VISSER,\* FOUAD SOULIMANI, and BERT M. WECKHUYSEN

*VibSpec, Tiel, The Netherlands (P.D.P.); Shell Global Solutions International B.V., Den Haag, The Netherlands (D.D.P., F.S., F.S.); and Inorganic Chemistry and Catalysis Group, Department of Chemistry, Utrecht University, The Netherlands (T.V., F.S., B.M.W.)*

Research has been carried out to determine the feasibility of chemometric modeling of infrared (IR) and near-infrared (NIR) spectra of crude oils to predict the long residue (LR) and short residue (SR) properties of these samples. A novel method is described to predict short residue properties at different flashing temperatures based on the IR spectrum of a crude oil measured at room temperature. The resulting method is the subject of European patent application number 07251853.3 filed by Shell Internationale Research Maatschappij B.V. The study has been carried out on 47 crude oils and 4 blends, representing a large variety of physical and chemical properties. From this set, 28 representative samples were selected by principle component analysis (PCA) and used for calibration. The remaining 23 samples were used as a test set to validate the obtained partial least squares (PLS) regression models. The results demonstrate that this integrated approach offers a fast and viable alternative for the currently applied elaborate ASTM (American Society for Testing and Materials) and IP (Institute of Petroleum) methods. IR spectra, in particular, were found to be a useful input for the prediction of different LR properties. Root mean square error of prediction values of the same order of magnitude as the reproducibility values of the ASTM methods were obtained for yield long on crude (YLC), density ( $D_{LR}$ ), viscosity ( $V_{LR}$ ), and pour point (PP), while the ability to predict the sulfur contents (S) and the carbon residue (CR) was found to be useful for indicative purposes. The prediction of SR properties is also promising. Modeling of the IR spectra, and to a lesser extent, the NIR spectra as a function of the average flash temperature (AFT) was particularly successful for the prediction of the short residue properties density ( $D_{SR}$ ) and viscosity ( $V_{SR}$ ). Similar results were obtained from the models to predict SR properties as a function of the yield short on crude (YSC) values. Finally, it was concluded that the applied protocol including sample pretreatment and instrumental measurement is highly reproducible and instrument and accessory independent.

Index Headings: Bitumen; Crude oils; Long residue properties; Short residue properties; Partial least squares; PLS; Average flash temperature; Yield on crude; Density; Viscosity; Sulfur; Attenuated total reflection; ATR; Infrared; Near infrared.

## INTRODUCTION

Bitumen is a complex mixture of hydrocarbons and hydrocarbon derivatives including aliphatic, olefinic, and aromatic compounds. It is usually pictured as a colloidal solution of high molecular carbon-hydrogens (so-called asphaltenes) finely dispersed in a mixture of resins and oils.<sup>1,2</sup> Bitumen is commonly known as an important component for road construction but is also used for roofing, waterproofing adhesive, and sealing applications. A substantial amount of bitumen is found in nature, i.e., buried in sand,<sup>3</sup> but, more generally, it is obtained as a residue of crude oil after atmospheric and vacuum distillation in the refinery. As such, it is an important product of the oil industry and of considerable

economical value. Therefore, Royal Dutch Shell is very careful to fully assess the bitumen from a crude oil to ensure it meets the market requirements before proceeding to approval for normal refinery production.

At present, a crude oil is assessed by laboratory preparation of actual bitumen samples, which are then passed through a range of further tests to confirm meeting the required quality. The assessment of these chemical and physical properties is carried out by standard methods developed by the American Society for Testing and Materials (ASTM) and the Energy Institute, formerly known as the Institute of Petroleum (IP). These methods are rather slow, elaborate, and expensive, requiring large sample amounts of the crude to be sent to the laboratory. As a fast and viable alternative, spectroscopic methods such as nuclear magnetic resonance (NMR),<sup>4–11</sup> Raman,<sup>12,13</sup> infrared (IR),<sup>14–31</sup> and near infrared (NIR)<sup>24,25</sup> have been proposed, since the corresponding spectra reflect the complete molecular composition of the crude oil.<sup>26</sup> Particularly, when combined with chemometric modeling techniques,<sup>27–29</sup> it might be possible to classify crude oils in terms of suitability for bitumen manufacture from their spectra. The aim of this study was to explore the suitability of IR and NIR spectroscopy to predict bitumen properties in terms of so-called long residue (LR) and short residue (SR) properties. LR properties are commonly defined as a series of physical and chemical properties of the residues flashed at 370 °C, while SR properties are described as physico-chemical properties of the residues flashed at higher temperatures, i.e., <490 °C Maxwell–Bonell. The approach of this study should link the more fundamental chemical composition of crude oils to short residue properties and thus to their bitumen quality. A novel method is described to predict short residue properties of a crude oil at different flashing temperatures based on the IR spectrum measured at room temperature.

## GENERAL METHODOLOGY

As a first step, methods for sample storage, sample pretreatment, and sample preparation were developed in experimental protocols to assure the acquisition of reproducible, high quality spectra. Although the high viscosity of several of the crude oils would make it reasonable to perform the measurements at elevated temperatures, for practical reasons it was chosen to carry out all measurements at room temperature (20 °C). As some samples contain considerable light-ends that flash off at low temperatures, spectra were acquired in closed cells. Next, experimental and instrumental settings were optimized to obtain sufficient spectral quality in terms of signal-to-noise ratio for each of the techniques. The final settings have been summarized in protocols and applied throughout the study. For IR, all samples have been analyzed

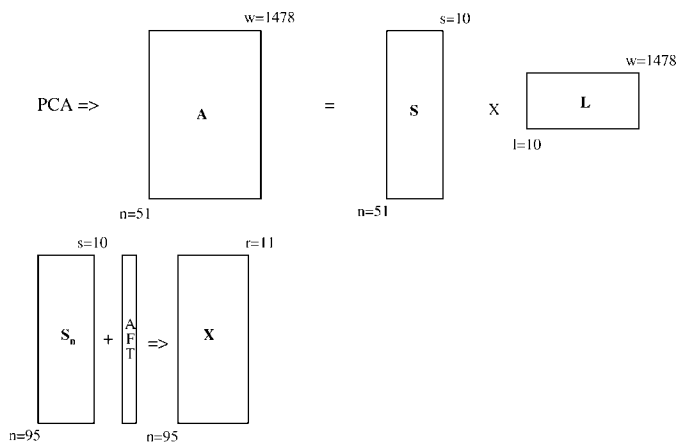
Received 5 November 2007; accepted 17 January 2008.

\* Author to whom correspondence should be sent. E-mail: t.visser@uu.nl.

**TABLE I. Crude oil samples used for calibration (C1–C28) and validation (V1–V23) with corresponding long residue (LR) properties. YLC: yield long-on-crude; D<sub>LR</sub>: density; V<sub>LR</sub>: viscosity; S: sulfur content; PP: pour point; Asph: asphaltenes content; CR: carbon residue.**

	YLC	D <sub>LR</sub>	V <sub>LR</sub>	S	PP	Asph	CR
C1	34.7	0.8889	29.5	0.72	42	1.2	3.9
C2	55.7	0.9507	37.2	4.75	30	9.5	14.0
C3	23.9	0.8839	28.9	1.08	30	1.0	3.3
C4	24.1	0.8827	28.8	1.11	36	0.7	3.3
C5	64.1	0.8665	30.8	0.14	48	0.1	4.4
C6	40.8	0.9003	29.1	0.70	36	1.6	6.4
C7	61.8	0.9504	37.9	1.17	15	1.2	9.9
C8	35.2	0.8962	34.6	0.56	30	0.6	4.1
C9	47.1	0.9188	33.3	0.84	3	0.2	4.7
C10	49.0	0.9488	36.8	3.22	21	5.8	12.6
C11	58.0	0.9273	36.3	2.19	24	3.2	9.6
C12	43.2	0.8660	29.2	0.25	42	0.2	4.4
C13	80.7	0.9820	41.8	3.05	45	8.0	14.0
C14	69.3	0.9086	36.2	0.99	48	0.2	9.7
C15	56.8	0.9800	39.3	5.42	27	12.2	17.7
C16	79.7	0.9842	41.9	3.14	39	8.0	14.2
C17	49.4	0.9178	33.8	2.05	18	2.2	7.7
C18	47.4	0.9236	33.5	2.15	27	2.7	8.0
C19	32.7	0.9583	34.5	4.87	27	2.0	10.7
C20	78.0	0.9429	39.8	0.30	30	4.5	12.6
C21	39.1	0.9313	35.0	0.43	30	0.4	6.0
C22	80.2	0.9958	44.1	3.88	48	12.0	16.8
C23	36.9	0.8865	30.9	0.81	39	1.3	6.3
C24	49.1	0.9473	36.4	3.14	18	6.3	12.4
C25	45.7	0.9332	33.2	3.23	6	3.6	10.4
C26	65.5	0.9771	41.9	4.31	39	15.4	19.2
C27	66.9	0.9344	35.7	1.47	9	2.0	8.7
C28	55.6	0.9573	38.1	4.47	21	8.4	15.3
V1	46.8	0.9278	32.7	2.85	39	5.3	11.0
V2	46.1	0.9299	33.3	3.17	39	3.2	9.8
V3	41.8	0.8970	31.5	0.42	39	0.1	5.0
V4	49.9	0.9208	33.0	1.54	36	6.0	10.8
V5	49.7	0.9444	35.3	4.30	9	4.2	11.7
V6	58.7	0.9510	37.5	3.39	33	9.0	14.1
V7	31.1	0.8822	31.9	0.46	36	0.2	4.3
V8	61.7	0.9470	36.4	4.26	18	3.5	15.0
V9	49.9	0.9471	35.9	3.76	21	5.6	12.2
V10	51.6	0.9136	34.0	2.82	48	4.5	9.1
V11	39.0	0.8969	30.1	0.48	36	3.3	6.8
V12	51.4	0.9457	35.7	4.12	18	5.4	12.3
V13	57.1	0.9546	38.3	2.52	15	8.6	13.8
V14	66.7	0.9696	43.3	2.50	54	15.3	19.3
V15	60.1	0.9682	38.6	5.09	18	9.9	16.1
V16	41.4	0.9080	31.3	1.13	27	2.2	6.4
V17	42.5	0.9325	33.6	3.09	30	2.2	8.3
V18	48.1	0.9229	32.8	2.03	24	2.7	7.7
V19	33.9	0.9533	35.1	4.55	24	2.1	10.4
V20	45.2	0.9382	34.9	2.88	15	5.2	11.3
V21	45.5	0.9376	35.3	2.76	21	5.6	10.9
V22	43.7	0.9424	35.5	3.05	15	4.7	11.4
V23	46.1	0.9392	35.4	2.91	21	5.4	11.8

three times with time gaps of 1 month, using two different instruments and two different attenuated total reflection (ATR) accessories. The first set of spectra was used to optimize the experimental conditions and to develop the final protocol. The other sets were measured in accordance with this protocol to determine the reproducibility of the spectra. In a separate experiment, the effect of evaporation of volatile compounds from the crude oil during IR spectroscopic measurement was tested by comparison of spectra recorded with the closed sample cell with spectra taken with an open cell as a function of time. Finally, measurements were carried out to study the effect of time during spectra acquisition with a closed cell to

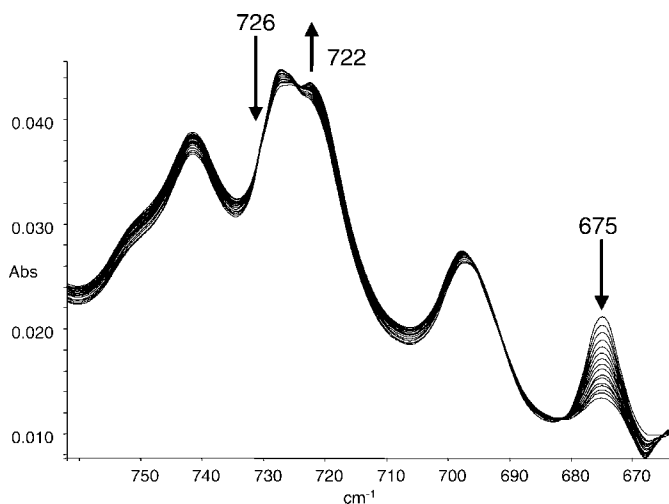


**FIG. 1.** Schematic representation of creation of the X matrix used as input for the PLS modeling of the crude SR property P at the available AFT values.

determine whether precipitation or a change in homogeneity occurred during spectrum acquisition.

Partial least squares (PLS) regression was applied to correlate the properties of long residues (LR) and short residues (SR) with the IR and NIR spectra of the crude oils. First, properties of residues flashed at 370 °C (i.e., long residues) have been modeled. The following LR properties were studied; the long residue yield-on-crude (YLC), density (D<sub>LR</sub>), viscosity (V<sub>LR</sub>), sulfur content (S), pour point (PP), asphaltenes (Asph), and carbon residue (CR). As input values for these properties, the data obtained with the following ASTM methods were used: D 2892 (YLC), D 4052 (D<sub>LR</sub>), D 445 (V<sub>LR</sub>), D2622 (S), D 97 (PP), and D 4530 (CR), while the values for Asph were determined by method IP 143.

In a second step, research was carried out to determine whether the short residue (bitumen) properties as a function of the average flash temperature (AFT) could be related to the spectral data of the crude oils in order to predict bitumen quality. As a starting point for the prediction of the short residue properties, the parameters penetration (P), softening point or ring and ball (R&B), viscosity (V<sub>SR</sub>) and density (D<sub>SR</sub>) were considered as the main properties. As input values for the



**FIG. 2.** The effect of evaporation of volatile components on the ATR-IR spectra of crude oil C1 when using an open sample cell.

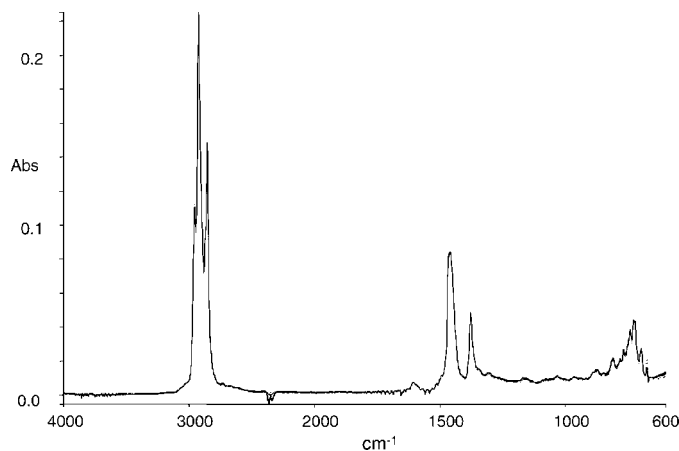


FIG. 3. Reproducibility of IR absorbance spectra of sample C1 in a closed ATR sample cell. 16 scans recorded over a 1 h time interval.

SR properties, the data obtained with the ASTM methods D 5 (P), D 36 (R&B), D 4052 ( $D_{SR}$ ), and D 445 ( $V_{SR}$ ) were used.

**Materials.** Spectra were acquired from a set of 47 crude oil samples and four blends. As can be seen from Table I, the set covers a wide range of LR properties. SR properties were known for smaller sets of samples. To reduce evaporation of volatile constituents from the original samples, all crude oils were stored in the refrigerator at 3 °C. Prior to analysis, samples were taken out of the refrigerator 8 h before measurement with caps closed to prevent condensation of moisture. Furthermore, samples were homogenized at least 1 h before measurement by shaking the sample every 10 min. From the total set, 28 samples were used for calibration purposes of the LR properties. The remaining 23 samples were used for validation of the models obtained. Modeling to predict SR properties was carried out on smaller sets since these data were not known for all compounds.

**Infrared Spectroscopy.** Infrared measurements were carried out at room temperature on two different instruments, a Bruker Tensor-27 and a Perkin-Elmer 2000 Fourier transform spectrometer. Both instruments were equipped with a DTGS detector and the sample compartments were flushed with dry air to reduce interference of  $H_2O$  and  $CO_2$ . Spectra were recorded either with a horizontal ATR accessory (MIRacle, Pike Technologies) with a ZnSe/diamond crystal as the internal reflection element (Perkin-Elmer 2000) or with a horizontal ATR accessory (FastIR™, Harrick Scientific Products) with a ZnSe crystal as the internal reflection element (Tensor-27 instrument). The spectral resolution was 4  $cm^{-1}$  for all spectra and 50 scans were accumulated with medium apodization for each spectrum. ATR intensity correction was not applied.

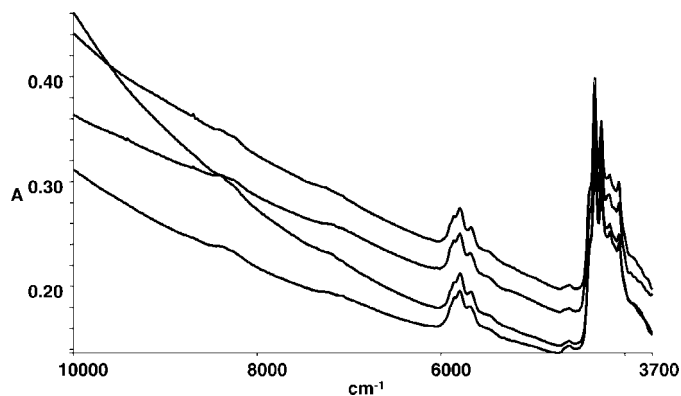


FIG. 4. NIR absorbance spectra of samples C9, C12, C13, and C22.

**Near-Infrared Spectroscopy.** Near-infrared spectra were recorded at room temperature on two different instruments, a Bruker Tensor-27 and a Perkin-Elmer 2000 Fourier transform infrared spectrometer. Both instruments were equipped with a (NIR) DTGS detector and the sample compartments were flushed with dry air to reduce interference of  $H_2O$  and  $CO_2$ . The spectral resolution was 16  $cm^{-1}$  and 100 scans were accumulated for each spectrum. Liquid samples were scanned in a fixed 0.1 mm cell equipped with  $CaF_2$  windows. Highly viscous samples were analyzed in a demountable cell with KBr windows and a path length of approximately 0.1 mm.

**Data Analysis.** Modeling was performed using the PLS Toolbox (Eigenvector Research, Inc.) for MatLab (The Math-Works, Inc.). Prior to modeling, the IR and NIR spectra were preprocessed by first taking the first derivative (25 pt. Savitzky–Golay smoothing), followed by multiplicative signal correction (MSC) and mean centering of the data. Preprocessing parameters were chosen based on prior knowledge for each spectroscopic technique combined with trial and error. For the IR spectra, the regions from 3500–2500  $cm^{-1}$  and 1760–650  $cm^{-1}$  were used as input, whereas for NIR the spectra from 7000  $cm^{-1}$  to 4000  $cm^{-1}$  (1429–2500 nm) was used. The obtained root mean square error of cross-validation (RMSECV) values were based on "leave-one-out" (LOO) cross-validation (CV). The number of latent variables (LV) used for a PLS model was based on the minimum value of the RMSECV. In case the minimum was reached at only one LV in the PLS cross-validation, the model was built with three LVs in order to obtain a more realistic PLS model.

First, modeling for the LR properties was carried out on the spectra (IR and NIR) of the set of 51 samples (47 crudes, 4 blends). This set was divided into a group of 28 spectra for calibration and 23 spectra for validation. The validation set was

TABLE II. Results of PLS modeling and prediction of the LR properties of crude oils based on IR spectroscopy. Models are based on a calibration set of 28 spectra and an independent test set of 23 spectra. Mean values and reproducibility values obtained by ASTM and IP methods.

LR property	Mean value	Reproducibility (+/–)	RMSECV	LVs	RMSEP
YLC (wt %)	52.5	1.2	3.5	7	3.5
$D_{LR}$ (g/mL)	0.930	0.0005	0.012	7	0.008
$V_{LR}$ (-)	35.3	-	2.1	6	1.3
S (wt %)	2.2	$0.09 \times S$ ( $S > 0.9\%$ )	0.5	8	0.2
PP (°C)	29.6	9	9.1	6	11.2
Asph (wt %)	4.1	$0.2 \times \text{Asph}$	2.1	6	2.4
CR (wt %)	9.7	$0.046 \times (3 + CR)$	2.0	6	1.4

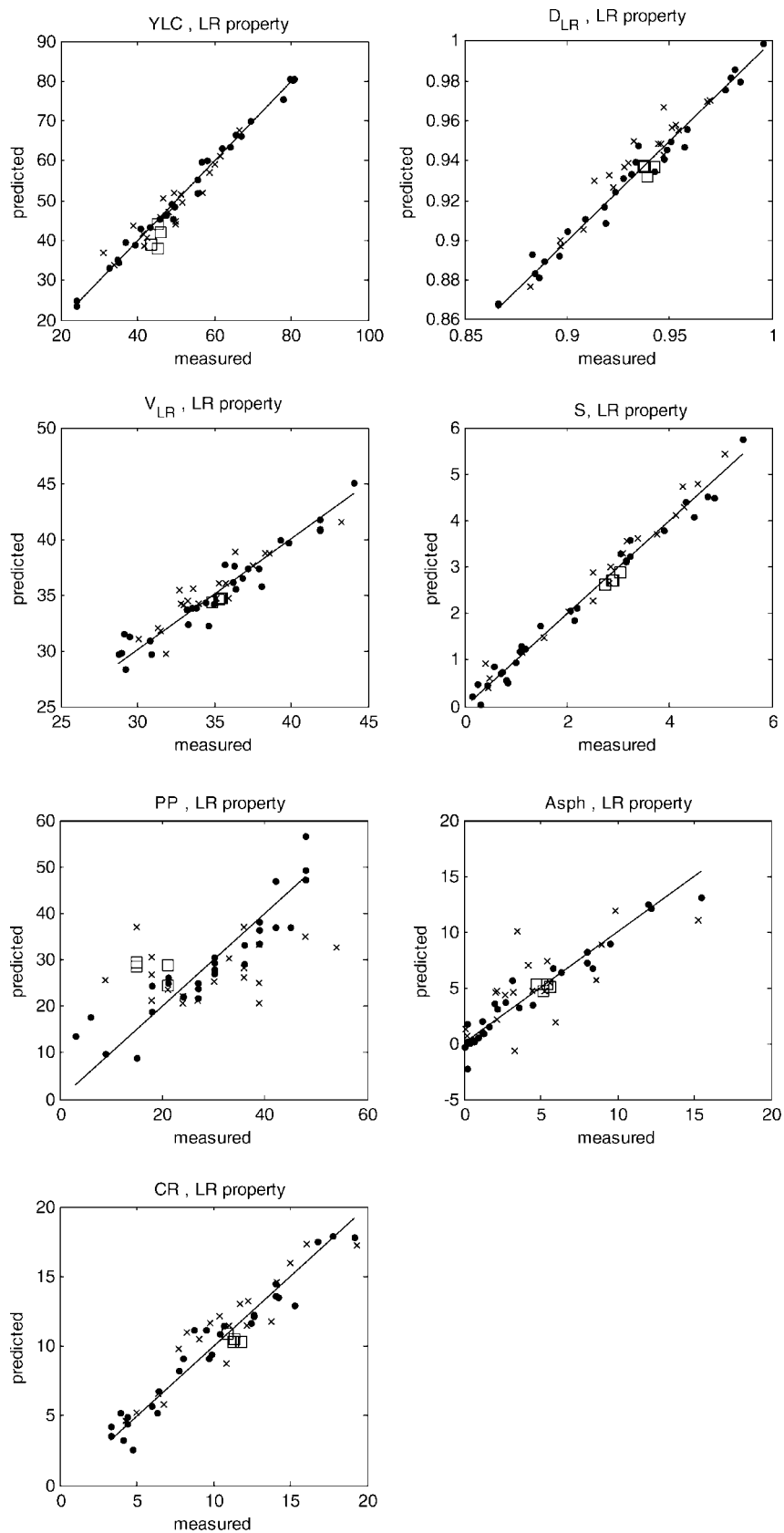


FIG. 5. Prediction plots of PLS modeling of the LR properties based on the IR spectra of crude oils. Models are based on a calibration set of 28 spectra (dots). The crosses and squares originate from the test set of 23 spectra. The squares represent the four blends.

**TABLE III. Results of PLS modeling and prediction of the LR properties of crude oils based on NIR spectroscopy. Models are based on a calibration set of 28 spectra and an independent test set of 23 spectra. Mean values and reproducibility values obtained by ASTM and IP methods.**

LR property	Mean value	Reproducibility (+/-)	RMSECV	LVs	RMSEP
YLC (wt %)	52.5	1.2	3.0	5	2.4
D <sub>LR</sub> (g/mL)	0.930	0.0005	0.017	4	0.008
V <sub>LR</sub> (-)	35.3	-	2.2	4	1.6
S (wt %)	2.2	0.09 × S (S > 0.9%)	0.7	7	0.5
PP (°C)	29.6	9	8.7	6	11.6
Asph (wt %)	4.1	0.2 × Asph	3.0	4	2.9
CR (wt %)	9.7	0.046 × (3 + CR)	2.8	4	2.2

based on the results of a principal component analysis (PCA) of the total collection without the four blends to obtain a representative subset. The Mahalanobis distance based on the first six scores was used to select 21 extreme samples with the largest distance and two samples with the smallest distance to the center of the scores. The five additional samples of the calibration set were chosen randomly and the blends were added to the test set. Secondly, the seven LR properties were modeled with PLS-1 with the described calibration and validation set of, respectively, 28 and 23 spectra (IR and NIR).

Next, PLS models (SIMPLS algorithm, PLS-1) were created from the IR spectra of 39 crude oils for which SR properties were known. PLS prediction models were made for the SR properties penetration (P), ring-and-ball (R&B), density (D<sub>SR</sub>), and viscosity (V<sub>SR</sub>) at various flash temperatures. The corresponding data sets consisted of 32 (P), 36 (R&B), 39 (D<sub>SR</sub>), and 37 (V<sub>SR</sub>) IR spectra with 95, 115, 125, and 115 properties, respectively. Because of the much smaller number of spectra for which SR properties were known, LOO cross-validation was applied instead of creating an independent validation set. In addition, modeling of the spectra was carried out after randomizing the SR properties as a test for the validity of the models obtained.<sup>30</sup> Finally, the SR properties were further determined at various AFT and this AFT value was used as discriminator.

As schematically depicted in Fig. 1, the data matrix **X** was created by PCA on the available IR spectra (**A**) after preprocessing (first derivative and 25 pt. Savitzky-Golay smoothing, MSC, and mean centering). The first ten scores of this PCA model (**S**) were used to describe the spectra for which the SR properties were available, resulting in a score matrix **S<sub>n</sub>**. The discriminating AFT values were then added to **S<sub>n</sub>**, resulting in 11 variables in the new data matrix **X**, which were variance scaled for each crude. Next, the corresponding SR properties were used as the **Y** matrix in the PLS regression. The same procedure was followed using the “yield short on crude” (YSC) instead of the AFT values as additional input of the data matrix.

**TABLE IV. Results of PLS modeling of the SR properties at different flash temperatures (AFT input) of crude oils based on IR spectroscopy. Models are based on 36 spectra for property P, 40 for R&B, 43 for D<sub>SR</sub>, and 41 for V<sub>SR</sub>. Mean values and reproducibility values obtained by ASTM methods.**

SR property (AFT)	Mean value	Reproducibility (+/-)	RMSECV	LVs
Log P (0.1 mm)	1.73	0.01–0.5	0.30	3
Log R&B (°C)	1.66	0.04	0.07	3
D <sub>SR</sub> (g/mL)	1.008	0.005	0.009	3
V <sub>SR</sub> (-)	42.7	-	1.5	3

Principal component analysis modeling was used to examine the differences between the different IR methods that were used.

## RESULTS AND DISCUSSION

**Infrared Spectroscopy.** All IR spectra are characterized by strong absorptions around 2900 and 1400 cm<sup>-1</sup>, originating from the C–H stretching and bending vibrations, respectively. Furthermore, all spectra show bands around 725 cm<sup>-1</sup> due to aromatic ring deformation and out-of-plane vibrations and long alkyl chain CH<sub>2</sub> rocking modes. As illustrated in Fig. 2 by sample C1, the IR spectra of most crude oils, recorded in an open cell as a function of time, revealed considerable changes, particularly in the fingerprint region. For that reason, only closed sample cell data were used for further modeling experiments. Under these conditions the spectra were found to be highly reproducible, as shown in Fig. 3.

**Near-Infrared Spectroscopy.** As shown in Fig. 4, the NIR spectra of the crude oils are characterized by absorption bands related to overtones/combinations of –C–H stretching and bending vibrations around 4100 cm<sup>-1</sup> (2439 nm) and 5500 cm<sup>-1</sup> (1818 nm).

Also, weaker bands can be observed at 7600 cm<sup>-1</sup> (1316 nm) and 8300 cm<sup>-1</sup> (1206 nm). Furthermore, considerable baseline differences are present, which can be attributed to different scattering effects and electronic transitions at higher wave numbers. In accordance with the literature, the effect was found to be the strongest for the dark and viscous heavy oils that usually contain relatively high amounts of aromatic compounds, including the solid asphaltene.<sup>31</sup> It should be noted that the highly viscous samples require the application of a demountable cell of which the path length cannot be adjusted with high accuracy, thus limiting a fair mutual comparison. This problem was partially circumvented in the modeling phase by normalizing the spectra on the strongest band present, i.e., around 4100 cm<sup>-1</sup> (2439 nm) or by applying MSC.

**Spectral Quality and Reproducibility.** The effect of the

**TABLE V. Results of PLS modeling of the SR properties at different flash temperatures (YSC input) of crude oils based on IR spectroscopy. Models are based on 36 spectra for property P, 40 for R&B, 43 for D<sub>SR</sub>, and 41 for V<sub>SR</sub>. Mean values and reproducibility values obtained by ASTM methods.**

SR property (YSC)	Mean value	Reproducibility (+/-)	RMSECV	LVs
Log P (0.1 mm)	1.73	0.01–0.5	0.39	3
Log R&B (°C)	1.66	0.04	0.09	3
D <sub>SR</sub> (g/mL)	1.008	0.005	0.015	3
V <sub>SR</sub> (-)	42.7	-	2.5	3

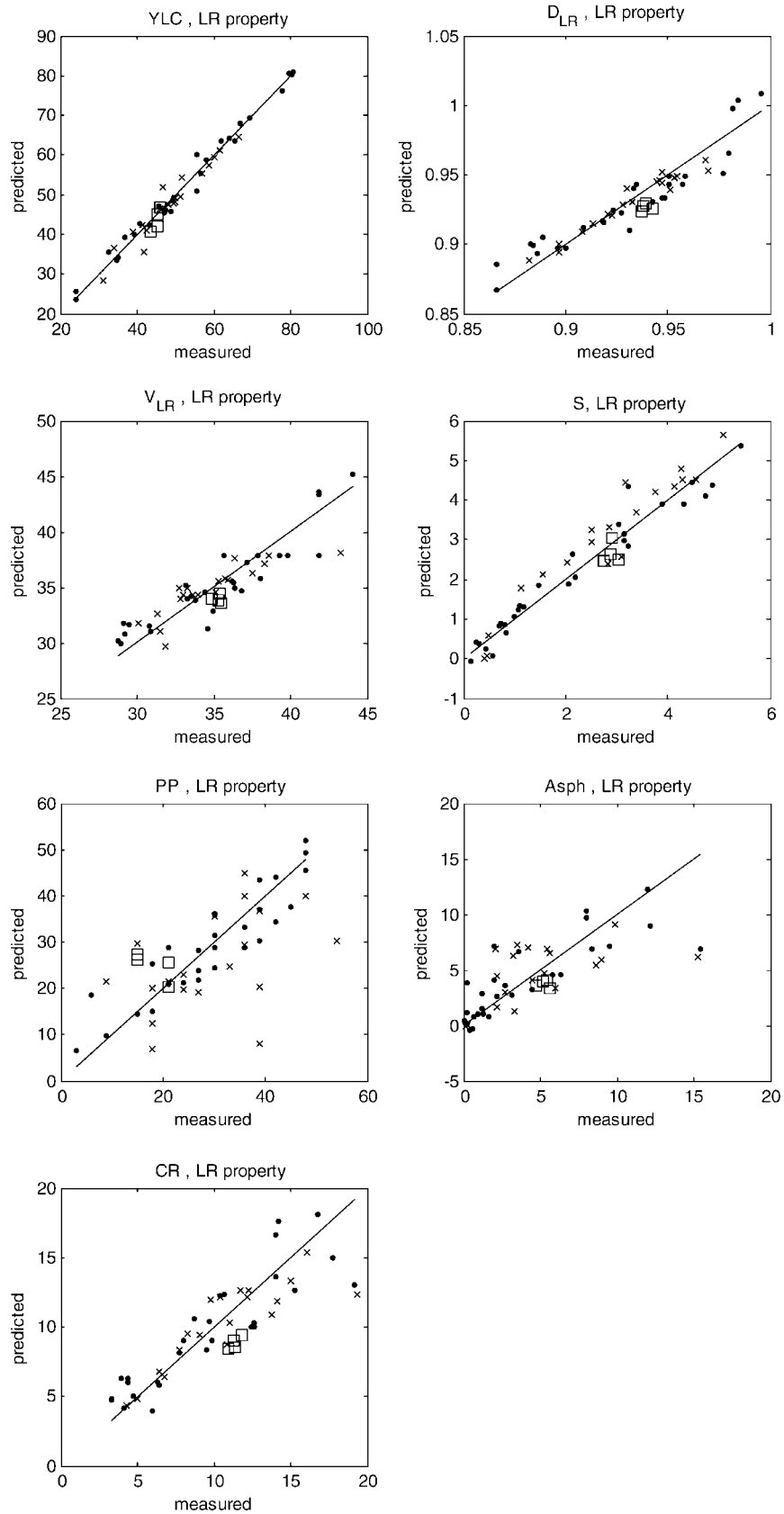


FIG. 6. Prediction plots of PLS modeling of the LR properties based on the NIR spectra of crude oils. All models were based on a calibration set of 28 spectra (dots). The crosses and squares originate from the test set of 23 spectra. The squares represent the four blends.

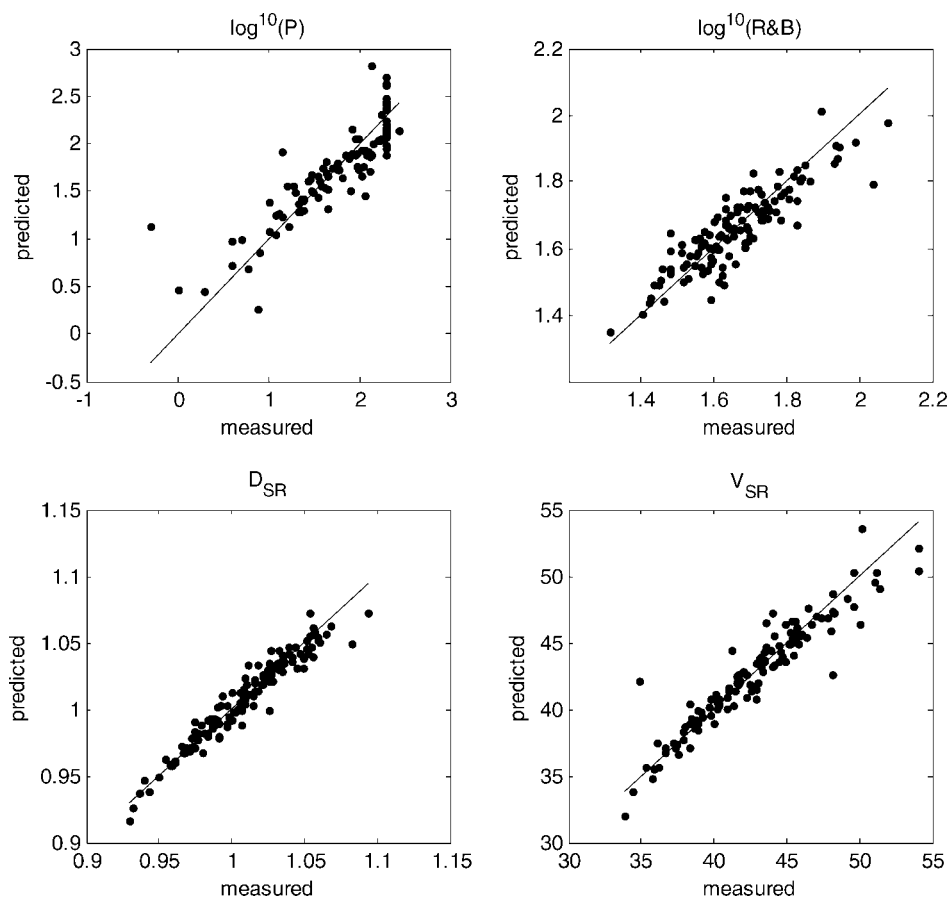


FIG. 7. PLS correlation plot of SR properties  $\log^{10}(P)$ ,  $\log^{10}(R\&B)$ ,  $D_{SR}$ , and  $V_{SR}$  obtained from 36, 40, 43, and 41 IR spectra combined with 103, 123, 134, and 122 AFT values, respectively.

spectrometer and the measurement cell on the reproducibility and the quality of the IR and NIR spectra appeared to be negligible. For both spectroscopic techniques, the RMSECV values of modeling the LR properties were found to be very similar for the duplicate measurements. This was confirmed by the results of PCA and partial least squares discriminant analysis (PLS-DA) of the complete data set of replicates forming two classes. The measurements on the two IR spectrometers with different ATR units also showed hardly any difference. The average peak-to-peak noise in the IR spectra (range 2200–2000  $\text{cm}^{-1}$ ) of one combination was better than for the other setup (0.0004 against 0.0006 absorbance units), but the differences in the results of modeling turned out to be negligibly small. Therefore, it was concluded that the experimental settings and scan parameters were correct for the purpose of this study.

**Data Analysis of Long Residue Properties.** The calibration set of 28 IR spectra was used to build a PLS model for each of the seven LR properties. The results, as summarized in Table II, show the mean value for each of the properties, the RMSECV based on the PLS model that was built, and the number of latent variables (LVs) used. For all models, the RMSECV is based on "leave-one-out" (LOO) cross-validation. All spectra were preprocessed before calibration was carried out. Preprocessing parameters were chosen based on prior knowledge of the spectroscopic technique combined with trial and error. The number of latent variables (LV) or principal components used for a PLS model was based on the minimum

value of the RMSECV. In case the minimum was reached at only one LV in the PLS cross-validation, the model was built with three LVs in order to obtain a more realistic PLS model. Validation of the models was carried out by the independent set of 23 test spectra. The results of prediction in terms of the root mean square error of prediction (RMSEP) values are included in Table II and depicted as prediction plots in Fig. 5. As it appears, the plots obtained for the properties S, PP, Asph, and CR are somewhat scattered. For PP, the corresponding RMSEP value is relatively large, but of the same order of magnitude as the reproducibility of the ASTM method. The RMSEP values for S, Asph, and CR are larger than the reproducibility of the physico-chemical methods, which implies that these models are more useful for indicative purposes. However, the models obtained for the properties YLC,  $D_{LR}$ , and  $V_{LR}$  perform very well. The high reliability of the predictions of these properties is reflected in the small RMSEP values, particularly regarding the small number of latent variables that was used in these models. The RMSEP values are not as good as the reproducibility values of the corresponding ASTM methods but are within reasonable limits. The RMSEP values are even smaller than the corresponding RMSECVs. This result and the fact that the properties of the test set are close to the average values confirm the validity of the models. Besides, the more extreme samples were selected for the calibration set. It follows that the RMSEP values are a better indication of the performance of the models than the RMSECV values.

The same preprocessing and modeling procedure as for IR

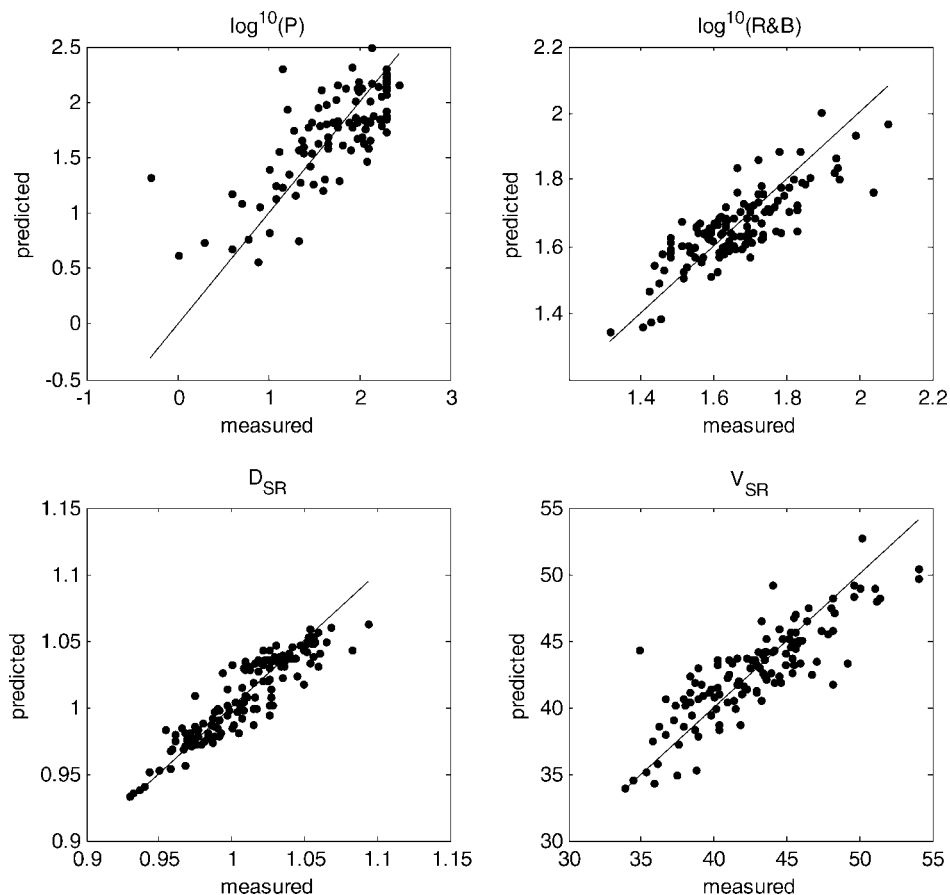


FIG. 8. PLS correlation plot of SR properties  $\log^{10}(P)$ ,  $\log^{10}(R\&B)$ ,  $D_{SR}$ , and  $V_{SR}$  obtained from 36, 40, 43, and 41 IR spectra combined with 103, 123, 134, and 122 YSC values, respectively.

was applied to build PLS models for the different LR properties from the calibration set of 28 NIR spectra. The results are presented in Table III. The independent set of 23 NIR test spectra was used for validation of the models. The results expressed as RMSEP values are added to Table III and the corresponding prediction plots are shown in Fig. 6. Similar to IR, the properties YLC,  $D_{LR}$ , and  $V_{LR}$  are predicted very well. Furthermore, the RMSEP values are again smaller than the corresponding RMSECVs and fairly close to the reproducibility values of the ASTM methods. The latter is also valid for the property PP, while the models for S, Asph, and CR are more useful for rough though very fast estimates.

Comparison of the results obtained for IR and NIR spectra demonstrate that the IR models perform better than the ones based on the NIR spectra for all properties. In all cases the RMSEP values are better and the plots less scattered.

At first glance, one could attribute the lower performance of the NIR models to the strong baseline differences at higher wavenumbers in the NIR spectra. However, as indicated by Mullins et al.,<sup>31</sup> these effects are partly due to differences in electronic transitions and become most apparent for the heavy crudes, since these contain relatively high amounts of aromatic compounds, including asphaltenes. For this reason, one might expect a better performance of the NIR model to predict the property Asph, but this is not observed. Furthermore, it should be noted that the NIR spectra have been recorded on an extended-range IR instrument. Possibly a dedicated high quality NIR instrument would have provided a better

performance of the models. For that reason, it is not possible to judge from this study whether IR is a better technique to predict LR and SR properties than NIR. However, the practical limitation that such an NIR spectrometer was not available led to the decision to focus further research on IR spectroscopy.

First, the use of a reduced spectral region was explored. However, the best results were obtained when using all parts of the spectrum that contain absorption bands, which implies the range 3500 to 650  $\text{cm}^{-1}$ . This region was used in all further modeling experiments on IR data.

**Short Residue Properties as a Function of the Flash Temperature.** Next, the IR data were combined with the average flash temperature (AFT) values in order to predict the SR properties P, R&B,  $D_{SR}$ , and  $V_{SR}$  from the IR spectra at different AFTs. First, PCA was applied to the 51 available crude oil IR spectra. The first ten scores were used to describe each spectrum and the AFT value was added as the 11th variable and variance scaling was applied. The resulting data set consists of 36 unique crude spectra for P, 40 for R&B, 43 for  $D_{SR}$ , and 41 for  $V_{SR}$  combined with 103, 123, 134, and 122 SR properties for the different AFT values. Subsequently a PLS regression was performed against the available SR property values. The results, expressed as RMSECV values and the corresponding number of LVs used, are presented in Table IV. In addition, the plots of the values predicted by the PLS model versus the measured SR properties are presented in Fig. 7. As can be seen, the RMSECV values for  $D_{SR}$  and  $V_{SR}$  are small, offering a good alternative for the corresponding



ASTM methods, whereas P and R&B are not predicted very well.

The results of the prediction models with the YSC values as input are presented in Table V and Fig. 8. They illustrate that the performance of the prediction models is better for all SR properties when AFT is used as discriminator.

## CONCLUSION

Infrared and near-infrared spectroscopy combined with PLS modeling have high potential to predict long residue (LR) and short residue (SR) properties as a function of the flash temperature. Both techniques offer a very fast alternative for the time-consuming ASTM and IP methods currently used to obtain this type of information. In accordance with vibrational spectroscopy theory, IR and NIR techniques score best on the same properties as a result of redundancy in the information content. However, the results obtained from IR models are better than for NIR in all cases. Even though pronounced electronic absorption profiles in the NIR spectra of heavy crude oils are present, significantly better predictions for the property Asph are not observed. Possibly, a dedicated NIR instrument might improve the predictive power of the NIR models.

As demonstrated by the prediction plots and RMSEP values obtained from the set of 23 validation spectra, the prediction of the LR properties YLC,  $D_{LR}$ , and  $V_{LR}$  from the IR data is quite reliable, while the inaccuracy in the prediction of PP is similar to that of the corresponding ASTM method. It follows that IR spectra of crude oils can be used to predict these LR properties with reasonable accuracy. The results to predict SR properties from the IR spectra of the crude oils are also promising. Modeling as a function of the average flash temperature (AFT) was particularly successful for the density ( $D_{SR}$ ) and the viscosity ( $V_{SR}$ ). It is expected that a higher accuracy and reproducibility of the input property values on the one hand, and further optimization of the models by deliberate wavelength selections on the other hand, will lead to even better results. Other potential options for improvement are the application of other preprocessing methods and outlier removal. The performance of the models to predict SR properties decreases when the yield short on crude (YSC) is used as input instead of the AFT.

The results of modeling the IR and NIR spectra recorded with different instruments and with different sample cells show

that the applied analysis protocol including sample pretreatment is well chosen and leads to highly reproducible results.

## ACKNOWLEDGMENT

This work has been supported by research grants from Shell Global Solutions BV.

1. M. H. B. Hayes, M. Stacey, and J. Standley, *Fuel* **51**, 27 (1972).
2. A. N. Ratov and A. P. Krylov, *Neftekhimiya* **36**, 195 (1996).
3. N. Berkowitz and J. G. Speight, *Fuel* **54**, 138 (1975).
4. J. G. Speight, *Fuel* **49**, 76 (1970).
5. M. H. B. Hayes, M. Stacey, and J. Standley, *Fuel* **51**, 32 (1972).
6. J. M. Jacobson and M. R. Gray, *Fuel* **66**, 753 (1987).
7. H. A. Akrami, M. F. Yardim, A. Akar, and E. Ekinci, *Fuel* **76**, 1389 (1997).
8. M. U. Hasan, M. F. Ali, and A. Bukhari, *Fuel* **62**, 518 (1983).
9. J. M. Jacobson, M. F. Ali, and A. Bukhari, *Fuel* **64**, 839 (1985).
10. L. Michon, D. Martin, J.-P. Planche, and B. Hanquet, *Fuel* **76**, 9 (1997).
11. C. Giavarini and C. Vecchi, *Fuel* **66**, 868 (1987).
12. L. C. T. Shoute, K. J. Schmidt, R. H. Hall, M. A. Webb, S. Rifai, P. Abel, P. H. Arboleda, A. Savage, J. T. Bulmer, and G. R. Lopnow, *Appl. Spectrosc.* **56**, 1308 (2002).
13. K. H. Michaelian, S. L. Zhang, R. H. Hall, and J. T. Bulmer, *Spectrochim. Acta, Part A* **57**, 73 (2001).
14. J. Lamontagne, P. Dumas, V. Mouillet, and J. Kister, *Fuel* **80**, 483 (2001).
15. L. W. Corbett, *Anal. Chem.* **41**, 576 (1969).
16. J. M. Jacobson and M. R. Gray, *Fuel* **66**, 749 (1987).
17. H. Tahmassebi, *Energy* **11**, 343 (1986).
18. J. Jehlička, O. Urban, and J. Pokorný, *Spectrochim. Acta, Part A* **59**, 2341 (2003).
19. E. D. Hondros, *Phil. Trans. R. Soc. Lond. A* **295**, 9 (1980).
20. K. H. Michaelian, R. H. Hall, and K. I. Kenny, *Spectrochim. Acta, Part A* **64**, 703 (2006).
21. P. W. Yang, H. H. Mantsch, L. S. Kotlyar, and J. R. Woods, *Energy Fuels* **2**, 26 (1988).
22. T. A. Adedosu and O. O. Sonibare, *J. Appl. Sci.* **5**, 906 (2005).
23. K. G. Martin, *Aust. Road Res. Board, Proc. Conf.*, 3rd, **Pt. 2**, 1433 (1966).
24. S. E. Moschopedis, R. W. Hawkins, and J. G. Speight, *Fuel* **60**, 397 (1981).
25. R. J. Mikula, V. A. Munoz, N. Wang, B. Bjornson, D. Cox, B. Moisan, and K. Wiwchar, *J. Can. Petrol. Technol.* **42**, 50 (2003).
26. K. A. Lintemann, *Anal. Chem.* **67**, 327R (1995).
27. K. J. Johnson, R. E. Morris, and S. L. Rose-Pehrsson, *Energy Fuels* **20**, 727 (2006).
28. M. Blanco, S. MasPOCH, I. Villarroya, X. Peralta, J. M. González, and J. Torres, *Appl. Spectrosc.* **55**, 834 (2001).
29. T. Zhang, Y. Liang, B. Li, H. Cui, and F. Gong, *Anal. Sci.* **20**, 717 (2000).
30. H. van der Voet, *Chemom. Intell. Lab. Syst.* **25**, 313 (1994).
31. O. C. Mullins, S. Mitra-Kirtley, and Y. Zhu, *Appl. Spectrosc.* **46**, 1405 (1992).