Diesel cetane number estimation from NIR spectra of hydrocracking total effluent

Jhon Buendia Garcia, Marion Lacoue-Nègre, Julien Gornay, Sílvia Mas García, Ryad Bendoula, Jean-Michel Roger

Diesel cetane number estimation from NIR spectra of hydrocracking total effluent

J. Buendia Garcia[a,c], M. Lacoue-Negre[a,c], J. Gornay[a], S. Mas Garcia[b,c], R. Bendoula[b,c], J.M Roger[b,c]

[a] IFP Energies Nouvelles, Rond Point de l'échangeur de Solaize, France

[b] ITAP-INRAE, Institut Agro, University Montpellier, Montpellier, France

[c] ChemHouse Research Group, Montpellier, France

Corresponding Authors:

Marion Lacoue-Negre (marion.lacoue-negre@ifpen.fr)

Abstract

The work shown in this paper offers a fast and efficient alternative for estimating the cetane number of the diesel obtained from the distillation of the hydrocracking total effluent. In this study, the estimation of this diesel property was achieved through a partial least squares regression (PLSR) model using only the NIR spectrum of the hydrocracking total effluent. For calibrating and validating the PLS model, it was used a database containing the NIR spectra acquired on 98 total effluent samples and the cetane number measured on the 98 diesel fractions recovered from each total effluent sample distillation. The database was divided into the calibration and test data sets using the Kennard-Stone algorithm. The regression model developed exhibited good performance in estimating the studied property with errors of calibration (1.3), cross-validation (2.2), and prediction (2.0), close to the reproducibility of the reference method (±3.6). The alternative method for diesel cetane number estimation discussed in this article evidences its feasibility in optimizing diesel fuel characterization by reducing the necessity of the total effluent distillation. Furthermore, the results also show the potential of the alternative proposed to be applied in predicting other properties of fuels obtained from the hydrocracking process.

Keywords

Hydrocracking, Total effluent, Diesel, Cetane Number, Near-Infrared (NIR), Chemometrics.

1. Introduction

The shift in consumption from gasoline to diesel has led over the last 20 years to a strong worldwide increase in demand for middle distillates (kerosene and diesel) [1]. At the same time, the increasing heavy crude oil production [2] has resulted in low-quality feedstocks being processed. The outlined issues and the constant demand for high-quality products have raised the need for flexible refining processes that maximize the production of middle distillates from heavy feedstocks while ensuring their quality for compliance with environmental and commercial legislations[2,3]. Given its extensive flexibility in processing heavy feedstocks, the hydrocracking (HCK) process is essential in addressing the need described [4]. Moreover, as an extensively implemented process nowadays, it is the subject of ongoing research.

The research on the HCK process is conducted by implementing experimental designs in pilot plants and laboratory facilities under controlled conditions. The implemented experimentation contributes to determining the best process configuration by processing different types of residues, mostly vacuum gas oil (VGO), under different operating conditions. In general, the experimentation is carried out in two main steps. In the first step, a hydrotreating stage (HDT) is applied to remove heteroatoms, saturate the olefins, and partially hydrogenate the aromatics. Subsequently, the hydrotreated effluent is sent to a reactor where, in the presence of a specific catalyst, the hydrocracking reactions occur [5] (See block #1 – Figure 1). In the second step, the liquid product obtained from the reaction section, known as total effluent, is distilled under atmospheric conditions to obtain the middle distillates, particularly diesel. These cuts are characterized using different standard norms such as the American Society for Testing and Materials (ASTM) and the International Organization for Standardization (ISO) (See block #2 – Figure 1). Finally, the analytical information obtained from this last step is gathered and analyzed to evaluate the impact of the operating conditions, including the catalytic system parameters, on the yield and quality of the diesel as a function of the processed feedstock.

In contrast to the reaction section, the characterization of the products is performed on a discontinuous time basis. Firstly, the laboratory analyses are conducted offline and are conditioned to the different laboratories' response times. Moreover, to perform the laboratory analyses based on the standards mentioned above, the physical product sample must be obtained from the total effluent distillation, which is also conducted in a non-continuous sequence. The products characterization is a fundamental task in the HCK process research. However, as previously discussed, the analytical workflow traditionally followed is both time- and volume-consuming. Therefore, a fast and efficient alternative for diesel fuel characterization is of great interest.

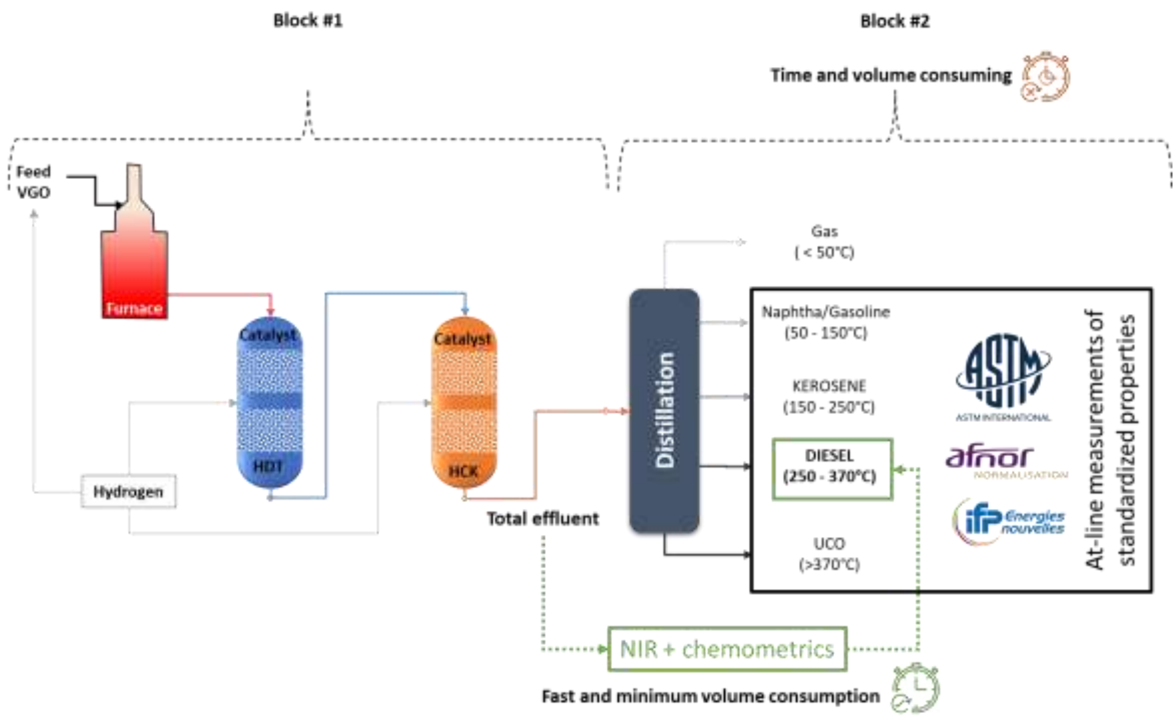


Figure 1 Workflow scheme for the characterization of fuels obtained from the HCK process

In the last decades, combining infrared spectroscopy analysis and chemometric methods has drastically increased for fuels characterization, from crude oils to refined cuts such as gasoline [6], diesel [7,8], biodiesel [7–10], and lubricants [11]. On the one hand, the main advantage of applying multivariate calibration methods to analytical techniques such as vibrational spectroscopy is both money- and time-saving. On the other hand, the sample volume required is quite low (up to a few milliliters) compared to some normalized methods generally used to characterize fuels. A recent review from Moro et al. [12] points out the growing use of infrared spectroscopy (IRS) to predict crude oils properties using chemometrics methods. To our knowledge, there is no existing equivalent review for other petroleum fractions. However, a plethora of interesting studies can be found showing the interest in using IRS and chemometrics to rapidly estimate fuel properties with statistical performance close to the reference methods [13–16].

Due to its extensive set of applications [17], NIR spectroscopy is particularly popular in laboratories to characterize fuels. Concerning diesel fuel, Hradecká et al. [15] recently demonstrated the feasibility of employing this vibrational technique to assess its quality. Using the partial least squares (PLS) algorithm, they estimated the kinematic viscosity, the cold filter plugging point, the pour point, and the sulfur and aromatics content from the NIR spectra acquired on different diesel samples. Each of the developed models enabled fast and reliable property predictions. Another recent study was developed by Yu et al. [18], where the estimation of diesel density from NIR spectra acquired on diesel samples was achieved using a "novel automatic model construction method." The resulting errors and squared correlation coefficients of the cited studies corroborated that an adequate application of chemometric methods on spectroscopic information leads to an accurate fuel properties estimation.

Among all the diesel fuel properties that can be investigated, the study shown in this article was focused on the diesel cetane number [19]. This property determines the ignitability of the diesel fuel using a standardized engine and a reference fuel. The cetane number is determined by comparing the ignition time of a mixture of cetane and hepta-methyl-nonane having the same ignition time delay as the tested sample. The cetane number on diesel is generally measured using the ASTM D613-01 standard [19], a destructive test that requires a significant volume of sample (500 ml), and its response time is a couple of hours.

Regarding the diesel cetane number estimation using NIR spectroscopy, the most recent studies are reported by Zhan et al. [20] and Barra et al. [21]. In the first study, a least squares-support vector machine (LS-SVM) regression model was developed with errors of calibration (1.8) and prediction (2.0) lower than the reproducibility of the ASTM D613-01 standard method (~3.3). However, the squared correlation coefficients of calibration (r^2c) and prediction (r^2p) were quite low (0.66). In the second study, diesel cetane number estimations with prediction errors around 0.5 and an r^2p value higher than 0.9 were achieved using a PLS regression model with 8 latent variables (LVs). Another study worth mentioning is the one developed by Zanier-Szydlowski et al. [22], who worked on predicting various fuel properties, including the diesel cetane number, developing a PLS model with a standard error of prediction (SEP) of 2.0.

All studies before-reported show that using NIR spectroscopy combined with proper chemometric methods in diesel properties estimation reduces the required sample volume and response time. However, the dependence on the distillation step of crude oil or HCK total effluent to obtain the diesel fraction and its subsequent characterization remains since the developed models are based on the NIR spectra acquired on the diesel cut. Therefore, aiming to go a step further in optimizing the analysis response time, this study presents an alternative for the cetane number estimation consisting of using the NIR spectra acquired on the HCK total effluent, avoiding the distillation step (see Figure 1). This main objective was achieved through four work steps. First, the total effluent samples obtained in different experimental tests of the HCK process conducted at a pilot level were identified and recovered. Next, the cetane number was measured on the diesel cuts corresponding to the total effluent samples. Then, the NIR spectra were acquired on the total effluent samples to finally perform all the necessary chemometric analysis, which included the preprocessing of the information and the calibration of the predictive model. To our knowledge, no comparative research has been reported.

2. Materials and methods

This section gives the origins and details of the sample physicochemical characterization. As a reminder, two sets of samples were considered: (i) the total effluents produced from HCK process reactors and (ii) the recovered diesel fractions.

2.1 Total effluent

In this study, 27 different feedstocks, mainly VGO, were processed in the HCK pilot plant units at IFPEN (Solaize, France) under various operating conditions involving different catalytic systems. The process variability ensured the physicochemical properties diversity of the 98 total effluent samples used in this research, as shown in Table 1. This table summarizes four relevant physicochemical properties of the obtained samples: the density,[23] the simulated initial boiling point (IBP), and distillation temperatures range to recover both 5% and 95% of sample distillate (Simulated Distillation T5 and T95)[24]. Table 1 also shows the fraction of the total effluent corresponding to the diesel cut.

Table 1 Summary of physicochemical properties measured on the total effluent samples obtained from the hydrocracking process experimental tests.

	Méthod	Minimum	Maximum	Mean	Standard Deviation
Density (g/mL)	ASTM D1218-12[23]	0.79	0.94	0.85	0.043
IBP (°C)	ASTM D2887-19[24]	38	205	106	42.1
SimDis T5 (°C)	ASTM D2887-19	69	345	179	83.3
SimDis T95 (°C)	ASTM D2887-19	401	585	503	44.6
Diesel yield (%)	ASTM D2892-20[25]	5.6	45.7	23.6	9.73

Near-infrared analysis

Before NIR spectra acquisition, the samples were first heated in a water bath at 60°C in a closed flask for one hour and then manually shaken to ensure homogeneity. Subsequently, NIR analysis was performed on each of the total effluents obtained using a Falcata Lab6 immersion reflectance probe (Hellma GmbH & Co. KG, Müllheim – Germany) with an optical path fixed at 2 mm. A spectrometer NIRS XDS Process Analyzer (Metrohm, Villebon - France) recording wavelengths within the 800 - 2200 nm spectral range with a resolution of 0.5 nm was used to acquire the spectra. Each final spectrum obtained was the average of 32 scans performed on the sample. The software used with the spectrometer was VISION (Metrohm, Villebon - France).

2.2 Diesel

The diesel samples used in this study were recovered from the atmospheric distillation of each of the 98 total effluents according to the ASTM D2892-20[25] standard. The cetane number was measured on each diesel sample recovered using an IFPEN internal method, which estimates this property from diesel NIR spectra through a PLS model based on Zanier-Szydlowski et al. work [22], with a larger database and equivalent performance. The internal method outlined was developed using the cetane numbers measured using the ASTMD613-01 standard [19] analysis as the reference method and validated against the reproducibility limits defined by this norm. Table 2 summarizes the general statistical information of the cetane number, the density and the Simulated Distillation SimDis T5 and T95 of the diesel samples considered in this study.

Table 2 General statistical information of the cetane number, density and simulated distillation measured on 98 diesel samples recovered from the total effluent distillation.

	Method	Minimum	Maximum	Mean	Standard Deviation
Cetane Number (CN)	ASTM D5949	30.3	69.5	51.6	11.07
Density (g/mL)	ASTM D1218-12	0.81	0.91	0.86	0.031
SimDis T5 (°C)	ASTM D2887-19	213	258	245	9.1
SimDis T95 (°C)	ASTM D2887-19	246	431	367	15.3

2.3 Modelling

An analysis to determine the best preprocessing scheme to be used was conducted. This study analyzed eight of the most common preprocessing methods applied to NIR spectra (see Table 3) [26] using an in-house MATLAB script. Each method was evaluated, taking their different parameter settings and possible combinations into account, based on the performance of different PLS regression models built using the root mean square error of cross-validation (RMSECV) and the squared coefficient of correlation (r^2C) as the figures of merit. For all models, the RMSECV was determined using the Venetian blind 10-fold.

Table 3 Pre-processing method evaluated on the NIR spectra of the HCK total effluent

#	Category	Method	Acronym	Parameters
1	Normalization	Variable Sorting for Normalization[27]	VSN	Automatic calculation
2		Standard Normal Variate[28]	SNV	
3		Multiplicative Signal Correction[29]	MSC	Reference data = mean of data, whole spectral range
4		Probabilistic Quotient Normalization[30]	PQN	
5	Filtering	Automatic Weighted Least Squares Baseline[26]	AWLS-B	
6		Detrend[28]	Dt	Polynomial order (1-3)
7		Extended Multiplicative Scatter/Signal Correction[31]	EMSC	Reference spectrum (basis to remove the scatter) = mean of each matrix generated, polynomial order = (1-4), whole spectral range, algorithm (CLS, ILS)*
8		Savitsky-Golay Derivative[32]	SG-D	Window points (9-25), polynomial order = (1-4), derivative order (1-4)

* CLS = Classical Least Squares, ILS = Inverse Least Squares.

For building and testing the regression models, the database was split into two datasets using the Kennard-Stone (KS) algorithm[33]: the calibration set (70% samples), which was used in model calibration and internal validation (cross-validation), and the independent test set (30% samples), which was used in the performance evaluation of the final developed model. For each PLS model developed, the number of latent variables (LVs) with the lowest RMSECV was retained as long as the cross-validation and calibration error ratio (RMSECV/RMSEC) did not exceed 1.7. This criterion was established empirically through previous modelling results to avoid model overfitting. In addition, analogous statistics were calculated on the test set (RMSEP, r^2P) to evaluate the model performance. The model errors were calculated using the Eq. (1), where y_i and $\hat{y}_i$ are the cetane number measured and predicted on sample i respectively, and n is the number of samples. For the squared correlation coefficients calculation Eq. (2) was utilized, where Cov and Var correspond to the covariance and variance respectively.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \qquad (1)$$

$$r^2 = \left(\frac{Cov(y,\hat{y})}{\sqrt{Var(y)Var(\hat{y})}}\right)^2 \qquad (2)$$

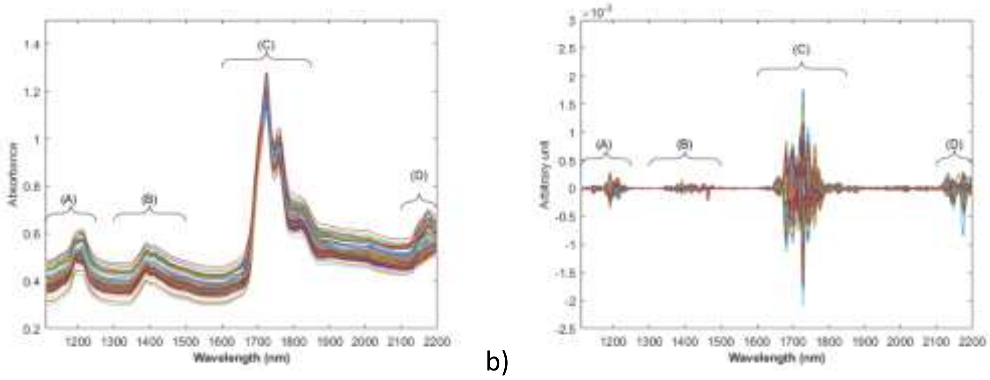
The models were developed with the PLS_Toolbox V.8.9 (Eigenvector Research Inc. Wenatchee, WA, USA) and MATLAB V.2020b (The MathWorks, Inc., Natick, MA, USA).

3. Results and discussion

3.1 Preliminary spectral analysis

Before developing the regression models, a preliminary analysis of the NIR spectra was performed to determine the spectral range used. Based on the studies conducted by Yalvac et al.[34] and Kelly et al.[35], it was established that the spectral region between 1100 and 2200 nm provides the most informative spectral features for hydrocarbon samples. Figure 2a shows the absorbance spectra of the total effluent samples in this spectral range. Although assigning each band of a near-infrared spectrum to a hydrocarbon molecule is difficult, a global attribution can be done as follows: (A) the bands around 1200 nm correspond to the second overtone of the CH bands; (B) the bands in the spectral region 1300-1500 nm can be attributed to the combinations of vibrational modes for the stretching of CH bonds; (C) the bands in the spectral interval 1600-1850 nm correspond to the first overtone bands of -CH stretch in $-CH_2$ and $-CH_3$; (D) the bands around 2200 nm can be attributed to the combination absorption bands of -CH stretching bonds and C=C stretching bonds in the aromatic ring. According to the previously outlined information, it was decided to develop the models on the 1110-2200 nm spectral region.

The different preprocessing methods summarized in Table 3 were evaluated using the spectral range defined. The best performance scenario obtained for this study was the combination of the Standard Normal Variate (SNV) and the second derivative of Savitzky-Golay with a third polynomial order (SavGol[23,3,2]). The preprocessing scheme was completed by centering the matrix by columns (mean center). Figure 2b shows the spectra preprocessed where the four spectral zones identified before can be observed.



a) b)

Figure 2 a) NIR spectra in absorbance, b) NIR preprocessed spectra. Spectral range used in modelling (1110-2200 nm). Highlighted regions: (A)(1100-1250 nm), (B)(1300-1500 nm), (C)(1600-1850 nm), (D)(2100-2200 nm)

3.2 Model performance analysis

After data preprocessing, a PLS model for the cetane number estimation was calibrated from the NIR spectra of 67 hydrocracked total effluent samples. The 67 corresponding diesel samples had a cetane number between 30.3 and 69.5. The external test set consisted of 31 total effluent spectra with an associated diesel cetane number ranging from 37.3 to 69.3. The score plot of the first two LVs of the developed PLS model shows a homogeneous distribution between the calibration and test samples (see Figure 3). This distribution ensures a representative evaluation of the model performance within the domain used in the model calibration. The distribution remains homogeneous throughout the other LVs (information not shown).

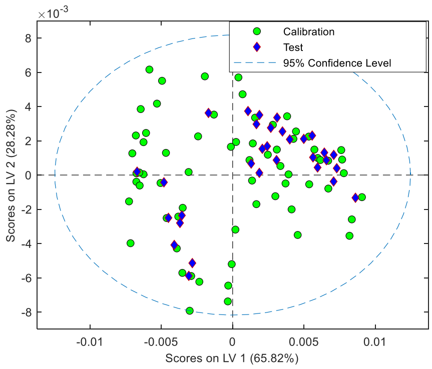


Figure 3 Projection of the calibration and test sets over the first and second latent variables (Score-plot).

The developed model uses 9 LVs to explain by about 99% the variance of the studied property. Considering the most recent studies regarding the estimation of diesel cetane number from NIR spectroscopy, the model developed in this study presents an RMSEP (2.0) comparable to the one obtained by Zhan et al.[20] (2.0) but presenting a better r^2P (0.96 vs. 0.55). Additionally, compared to the regression method employed by them (LS-SVM), by using the PLS method in this study, the obtained model was interpretable, helping to understand the chemical information of the total effluent having an impact on the diesel cetane number. Regarding the study done by Barra et al. [21], which presents a lower RMSEP (0.42) using a PLS model of 8LVs, it should be noted that the data set used for testing their model is smaller (10 vs. 31) with a narrower cetane number range. The limited application range of the models reported in the two previously analyzed studies highlights another advantage of the model described in this article. While in the studies of Zhan and Barra the applicable model range is between 20.4-49.5 and 49-59, respectively, for the model developed is between 30.9-69.5.

Although the results of these studies are not rigorously comparable with the research shown in this paper due to the type of sample used for the NIR spectra acquisition (diesel vs. HCK total effluent), it can be observed that improvements in certain aspects are achieved. Furthermore, it is worth emphasizing that the alternative investigated in this study optimizes the diesel characterization response time, which was restricted by the distillation step. Finally, compared to ASTM D613-01 [19], the RMSEP of the developed model is below the reproducibility of all the cetane number ranges established by this standard.

In summary, using a PLS regression model with 9 LVs, it is possible to estimate the diesel cetane number from the spectroscopic information of the HCK total effluent with errors below the reproducibility limit of the IFPEN internal reference method (±3.6) and the ASTM D613-01 norm [19]. Moreover, the developed model ensures a reliable prediction throughout the entire range of property evaluation by presenting squared correlation coefficients higher than 0.95, showing a good correlation between the reference and predicted values. Table 4 shows the main information describing the chemometric model developed.

Table 4 Statistical parameters and model information for predicting diesel cetane number (CN)

Regression method	PLS
Latent variables	9
X Explained Variance	99.4%
Y Explained Variance	98.6%
RMSEC	1.3
RMSECV	2.2
RMSEP	2.0
r^2C	0.986
r^2CV	0.959
r^2P	0.955
Prediction Bias	-0.6

The satisfactory performance of the model obtained is reflected in the parity and residual plots shown in Figure 4a and Figure 4b, respectively. Figure 4a shows that out of the 31 samples used in the model test set, 30 were predicted between the lower and upper limits of the reference method reproducibility, resulting in a prediction effectiveness of approximately 97%. In turn, Figure 4b illustrates the homogeneous distribution of the residual values obtained in both the calibration and the test of the model, showing its homoscedasticity in the whole evaluation range of the studied property, and evidencing the absence of model overtraining.

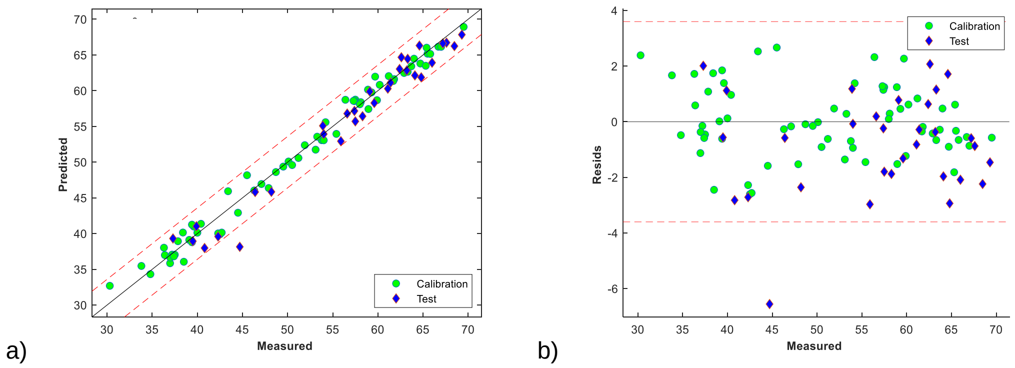


Figure 4 a) Parity plot, b) prediction residuals plot of PLS model for predicting the diesel cetane number. Red dotted lines: upper and lower limits of the reproducibility of the reference method (±3.6)

A graphical analysis combining the Q residual and the Hotelling T^2 statistical analyses was performed to establish if the predicted sample outside the reproducibility limits of the reference method corresponds to an outlier. The Q residual test determines the samples with atypical behavior by measuring the difference between a sample and its projection into the LVs retained in the model [36]. If the residual Q value of a sample exceeds the unit, this sample can be considered a weak outlier, and its cause would be mainly related to the acquisition spectrum quality. Analogously, Hotelling's T^2 determines the atypicality of the samples using the measure of the variation in each sample within the model [36]. If the resulting test value of a sample exceeds the unit, it could be considered a strong outlier, and the cause would be mostly related either to the quality of the studied variable measurement or to the physicochemical properties of the sample. Finally, if a sample simultaneously exceeds the established thresholds of the two tests, the information from this sample could substantially impact the model performance. Therefore, its use in the model should be reconsidered.

Figure 5 shows the reduced Q residual and Hotelling T^2 analysis applied to the test set. Firstly, it can be observed in this figure that no sample is above the threshold of the two tests simultaneously. Secondly, two of the samples used in testing the model are above the threshold of the residual Q test. However, neither of these two samples corresponds to the sample predicted outside the limits. On the contrary, this sample is between the threshold limits of both tests (red point Figure 5). Consequently, it cannot be identified as an outlier. By a deeper analysis of this sample information regarding the operating and spectrum acquisition conditions, it was found that the total effluent sample analyzed was produced during a test with particular operating conditions in comparison to the rest of the sample set (feedstock with a high content of paraffinic carbon (>60%) processed under lower operating pressure). Thereby, the poor prediction of this sample could be attributed to the fact that the spectroscopic information used in the model calibration is not capturing the sample chemical description given by the particularity of the sample's origin. The present study focuses on estimating the studied property using NIR spectroscopy. The results indicate that this estimation is possible but also exhibit that some external parameters, as operating conditions, can influence the prediction. This issue, related to the calibration robustness [37], could be addressed by developing predictive models that simultaneously use the information of the total effluent NIR spectra and the operating conditions employed in obtaining the sample.

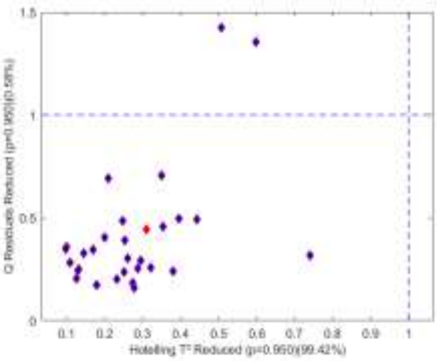
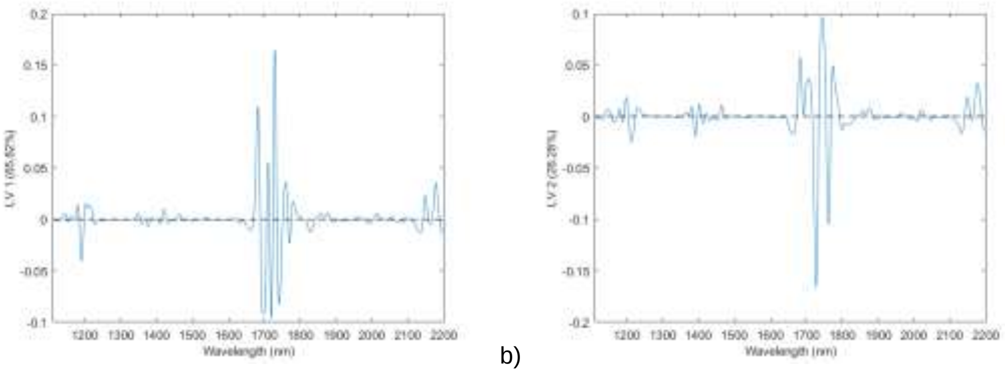


Figure 5 Reduced Q residual and Hotelling T^2 analysis using a 9 LVs PLS model

3.3 Model interpretation analysis

As mentioned before, one advantage of using the PLS regression method is to obtain predictive models helping to have a more detailed understanding of the effect that the different chemical compounds present in the sample may have on the estimation of the studied property. Figure 6 shows the PLS model loadings of the first 2 LVs, explaining 94% of the variance of the investigated property. This figure shows that the four zones previously identified influence the cetane number estimation. The zone between 1610 and 1810 nm is the one that presents the greatest impact. As mentioned formerly, this zone corresponds to the first overtone of the -CH stretching bands in $-CH_2$ and $-CH_3$. The behavior of the diesel cetane number is directly related to the type of isomerization, the length, and the amount of the identified linear hydrocarbons compounds. Therefore, the coherent relationship between the studied property and the chemical information extracted from the NIR spectra acquired on the total effluent is demonstrated. This consistency suggests the possibility of applying the alternative proposed in this study to estimate other diesel properties.



a) b)

Figure 6 PLS model loadings plot for a) 1st latent variable (65.8% of Y variable variance explained), b) 2nd latent variable (28.3% of Y variable variance explained)

The previous description and results analysis validated the suitability of applying the alternative investigated in this article for estimating middle distillate properties with errors close to the reproducibility of the reference method. The diesel characterization alternative discussed in this study is based on exploiting the NIR spectra acquired on the HCK total effluent. The predictive model calibration represented a challenge during the research work due to the complex extraction and exploitation of the total effluent chemical information for correctly describing the studied property. Nonetheless, compared to the models for diesel cetane number estimation reported in the literature, the model developed in this study showed satisfactory performance. In addition, the model presents some further advantages concerning its homoscedasticity and its application range. Finally, as discussed in the introduction section, the interest in employing the total effluent NIR spectra was motivated by the need to go a step further in the response time optimization when characterizing the diesel fuel. Through the approach developed, this need is fully addressed as the distillation of the total effluent to recover the physical cuts is not required, offering the possibility of performing the properties estimation in real-time.

Conclusions

The proper application of chemometric methods enables the physicochemical properties estimation of a crude oil cut using spectral information from another related product. This study developed a chemometric model for predicting the diesel cetane number using NIR spectroscopy information acquired on the total effluent obtained from the hydrocracking process. Hence, a fast and efficient alternative for fuel properties estimation was presented. The PLS regression model obtained provides a reliable and fast estimation of the diesel cetane number with errors within the reproducibility of the reference method and correlation squared coefficients above 0.95. These results demonstrate the potential of the alternative investigated to minimize the required sample volume and the response time for property estimation by reducing the necessity to perform the total effluent distillation. Furthermore, this optimization could lead to performing a time- and cost-effective research of the hydrocracking process by real-time estimating the studied property.

When estimating diesel properties using the spectroscopic information acquired on the total effluent, the predictive performance could be affected by the total effluent properties, which are impacted by parameters related to the feedstock quality and operating conditions. Therefore, it is important to address the model robustness constraint to ensure reliable performance over time and under different analytical conditions.

The study exposed in this paper highlights the wide application field of chemometrics, which facilitates the use of spectral information in the development of prediction models and enables the analysis and identification of atypical behaviors that fuel properties may have, helping to establish and understand the possible causes. Therefore, a better description of the influence that different process parameters and variables would have on the studied properties can be achieved, contributing to efficient process optimization.

The results obtained raise the prospect of using the alternative presented in this study for estimating other diesel properties as well as for properties prediction of different fuel products, namely, kerosene.

Finally, it should be highlighted that no regression model was found in the literature to predict diesel cetane number from NIR spectroscopy information of the hydrocracking total effluent, making this work the first one developed.

Acknowledgments

The authors would like to thank IFP Energies Nouvelles for providing the total effluent samples from the HCK process reactors, the facilities for the distillation to obtain the diesel samples, and the facilities for spectra acquisition and data analysis. Thanks to Axel One Analysis for providing the probe used on the NIR spectra acquisition.

Funding

This work was supported by IFPEN Energies Nouvelles

CRediT authorship contribution statement

J. Buendia Garcia: Conceptualization, data curation, Writing - original draft. **M. Lacoue-Negre:** Conceptualization, Writing - original draft. **J. Gornay:** Conceptualization, Writing - original draft. **S. Mas Garcia:** Writing - original draft. **R. Bendoula:** Writing - original draft, **J.M Roger:** Conceptualization, Writing - original draft

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

REFERENCES

References

[1] Fuels Europe. Statistical Report 2018. Study. Belgium; 2018.

[2] Marafi A, Albazzaz H, Rana MS. Hydroprocessing of heavy residual oil: Opportunities and challenges. Catalysis Today 2019;329:125–34. https://doi.org/10.1016/j.cattod.2018.10.067.

[3] Rana MS. Heavy Oil Refining Processes and Petrochemicals: A Role of Catalysis. Recent Adv Petrochem Sci 2017;2. https://doi.org/10.19080/RAPSCI.2017.01.555580.

[4] Elshout R, Bailey J, Brown L, Nick P. Upgrading the bottom of the barrel. Hydrocarbon Processing 2018, March 2018; Available from: https://www.hydrocarbonprocessing.com/magazine/2018/march-2018/special-focus-clean-fuels/upgrading-the-bottom-of-the-barrel.

[5] Vivas-Báez JC, Servia A, Pirngruber GD, Dubreuil A-C, Pérez-Martínez DJ. Insights in the phenomena involved in deactivation of industrial hydrocracking catalysts through an accelerated deactivation protocol. Fuel 2021;303:120681. https://doi.org/10.1016/j.fuel.2021.120681.

[6] Mabood F, Gilani SA, Albroumi M, Alameri S, Al Nabhani MM, Jabeen F. Detection and estimation of Super premium 95 gasoline adulteration with Premium 91 gasoline using new NIR spectroscopy combined with multivariate methods. Fuel 2017;197:388–96. https://doi.org/10.1016/j.fuel.2017.02.041.

[7] Balabin RM, Lomakina EI, Safieva R. Neural network (ANN) approach to biodiesel analysis: Analysis of biodiesel density, kinematic viscosity, methanol and water contents using near infrared (NIR) spectroscopy. Fuel 2011, 2011:2007–15.

[8] Rocabruno-Valdés CI, Ramírez-Verduzco LF, Hernández JA. Artificial neural network models to predict density, dynamic viscosity, and cetane number of biodiesel. Fuel 2015;147:9–17. https://doi.org/10.1016/j.fuel.2015.01.024.

[9] Bemani A, Xiong Q, Baghban A, Habibzadeh S, Mohammadi AH, Doranehgard MH. Modeling of cetane number of biodiesel from fatty acid methyl ester (FAME) information using GA-, PSO-, and HGAPSO- LSSVM models. Renewable Energy 2020;150:924–34. https://doi.org/10.1016/j.renene.2019.12.086.

[10] Nabipour N, Daneshfar R, Rezvanjou O, Mohammadi-Khanaposhtani M, Baghban A, Xiong Q et al. Estimating biofuel density via a soft computing approach based on intermolecular interactions. Renewable Energy 2020;152:1086–98. https://doi.org/10.1016/j.renene.2020.01.140.

[11] Pinheiro CT, Rendall R, Quina MJ, Reis MS, Gando-Ferreira LM. Assessment and Prediction of Lubricant Oil Properties Using Infrared Spectroscopy and Advanced Predictive Analytics. Energy Fuels 2017;31(1):179–87. https://doi.org/10.1021/acs.energyfuels.6b01958.

[12] Moro MK, dos Santos FD, Folli GS, Romão W, Filgueiras PR. A review of chemometrics models to predict crude oil properties from nuclear magnetic resonance and infrared spectroscopy. Fuel 2021;303. https://doi.org/10.1016/j.fuel.2021.121283.

[13] Morris RE, Hammond MH, Cramer JA, Johnson KJ, Giordano BC, Kramer KE et al. Rapid Fuel Quality Surveillance through Chemometric Modeling of Near-Infrared Spectra. Energy Fuels 2009;23(3):1610–8. https://doi.org/10.1021/ef800869t.
[14] Al Ibrahim E, Farooq A. Octane Prediction from Infrared Spectroscopic Data. Energy Fuels 2020;34(1):817–26. https://doi.org/10.1021/acs.energyfuels.9b02816.
[15] Hradecká I, Velvarská R, Dlasková Jaklová K, Vráblík A. Rapid determination of diesel fuel properties by near-infrared spectroscopy. Infrared Physics & Technology 2021. https://doi.org/10.1016/j.infrared.2021.103933.
[16] Feng F, Wu Q, Zeng L. Rapid analysis of diesel fuel properties by near infrared reflectance spectra. Spectrochim Acta A Mol Biomol Spectrosc 2015;149:271–8. https://doi.org/10.1016/j.saa.2015.04.095.
[17] Chung H. Applications of Near-Infrared Spectroscopy in Refineries and Important Issues to Address. Applied Spectroscopy Reviews 2007;42(3):251–85. https://doi.org/10.1080/05704920701293778.
[18] Yu H, Wang X, Shen F, Long J, Du W. Novel automatic model construction method for the rapid characterization of petroleum properties from near-infrared spectroscopy. Fuel 2022;316. https://doi.org/10.1016/j.fuel.2021.123101.
[19] ASTM D613-01. Test Method for Cetane Number of Diesel Fuel Oil. West Conshohocken, PA: ASTM International; 2001. https://doi.org/10.1520/D0613-01.
[20] Zhan B, Yang J. Measurement of Diesel Cetane Number Using Near Infrared Spectra and Multivariate Calibration. Advances in Engineering 2017;100:270-247. https://doi.org/10.2991/icmeim-17.2017.41.
[21] Barra I, Kharbach M, Qannari EM, Hanafi M, Cherrah Y, Bouklouze A. Predicting cetane number in diesel fuels using FTIR spectroscopy and PLS regression. Vibrational Spectroscopy 2020;111:103157. https://doi.org/10.1016/j.vibspec.2020.103157.
[22] Zanier-Szydlowski N, Quignard A, Baco F, Biguerd H, Carpot L, Whal F. Control of Refining Processes on Mid-Distillates by Near Infrared Spectroscopy. Oil & Gas Science and Technology - Rev. IFP 1999;54(4):463–72. https://doi.org/10.2516/ogst:1999040.
[23] ASTM D1218 - 12. Standard Test Method for Refractive Index and Refractive Dispersion of Hydrocarbon Liquids; Available from: https://www.astm.org/Standards/D1218.htm.
[24] ASTM D2887 - 19ae1. Standard Test Method for Boiling Range Distribution of Petroleum Fractions by Gas Chromatography; Available from: https://www.astm.org/Standards/D2887.htm.
[25] ASTM D 2892-20. Standard Test Method for Distillation of Crude Petroleum (15-Theoretical Plate Column). West Conshohocken, PA: ASTM International; 2020. https://doi.org/10.1520/D2892-20.
[26] Rinnan Å, van Berg F den, Engelsen SB. Review of the most common pre-processing techniques for near-infrared spectra. TrAC Trends in Analytical Chemistry 2009;28(10):1201–22. https://doi.org/10.1016/j.trac.2009.07.007.
[27] Rabatel G, Marini F, Walczak B, Roger J-M. VSN: Variable sorting for normalization. Journal of Chemometrics 2020;34(2):205. https://doi.org/10.1002/cem.3164.
[28] Barnes RJ, Dhanoa MS, Lister SJ. Standard Normal Variate Transformation and De-Trending of Near-Infrared Diffuse Reflectance Spectra. Appl Spectrosc 1989;43(5):772–7. https://doi.org/10.1366/0003702894202201.
[29] Martens H, Naes T. Multivariate calibration. Chichester: Wiley; 1989.
[30] Dieterle F, Ross A, Schlotterbeck G, Senn H. Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in 1H NMR metabonomics. Analytical Chemistry 2006;78(13):4281–90. https://doi.org/10.1021/ac051632c.
[31] Harald M, Edward Stark. Extended multiplicative signal correction and spectral interference subtraction: new preprocessing methods for near infrared spectroscopy. Journal of Pharmaceutical & Biomedical Analysis 1991;9.
[32] Abraham. Savitzky/M. J. E. Golay. Smoothing and Differentiation of Data by Simplified Least Squares Procedures. Analytical Chemistry 1964;36.

[33] Kennard RW, Stone LA. Computer Aided Design of Experiments. Technometrics 1969;11(1):137–48. https://doi.org/10.1080/00401706.1969.10490666.
[34] Yalvac ED, Seasholtz MB, Crouch SR. Evaluation of Fourier Transform Near-Infrared for the Simultaneous Analysis of Light Alkene Mixtures. Appl. Spectrosc., AS 1997;51(9):1303–10. https://doi.org/10.1366/0003702971942303.
[35] Kelly JJ, Callis JB. Nondestructive analytical procedure for simultaneous estimation of the major classes of hydrocarbon constituents of finished gasolines. Anal. Chem. 1990;62(14):1444–51. https://doi.org/10.1021/ac00213a019.
[36] Le Mujica, Rodellar J, Fernández A, Güemes A. Q-statistic and T2-statistic PCA-based measures for damage assessment in structures. Structural Health Monitoring 2011;10(5):539–53. https://doi.org/10.1177/1475921710388972.
[37] Zeaiter M, Roger J-M, Bellon-Maurel V, Rutledge DN. Robustness of models developed by multivariate calibration. Part I. TrAC Trends in Analytical Chemistry 2004;23(2):157–70. https://doi.org/10.1016/S0165-9936(04)00307-3.