

Prediction of the refractive index of hydrotreated gas oils with near infrared spectroscopy

Nathalie Zanier

Institut Français du Pétrole (IFP), 1 à 4 avenue Bois Préau, 92500 Reuil-Malmaison, France.

Introduction

Road vehicle gas oil is a worldwide product with properties that are important to gas oil producers, automobile manufacturers and consumers. The target properties of interest start with a good cetane number (measuring the self-ignition of the gas oil), good low temperature properties (influencing the way of introduction of the gas oil) and a low sulfur content (affecting catalyst durability and environmental pollution). They also include other properties such as the density (affecting the power of the engine and the emission rate), the distillation parameters (affecting the volatility of the gas oil and thus the overall engine performance), the viscosity (affecting the combustion procedure), the nitrogen content (affecting the stability of the gas oil) and the aromatic content (affecting the ignition characteristics). The legislation on gas oil is evolving in Europe and in the US. The constraints on the composition become more and more important to reduce emissions or to modify them.

Gas oils with different characteristics can be obtained by different refining configurations and the refiners have to find the best way to meet the specified targets and to continue to make an operating profit. It is a complex problem because of the wide variety of gas oil blending stocks, ranging from paraffinic straight run gas oil to coker gas oils and FCC (fluid catalytic cracking) cycle oils.

In the laboratory, there are several normalised methods available for characterising gas oils.¹ However, many of these methods are time-consuming and need tedious procedures that require large volumes of sample and well-trained, experienced analysts. Moreover, each method requires a specific instrument which adds to cost. The measurement of these gas oil properties can also be required for the optimisation and the advanced control of processes. This is not possible without a fast and complete characterisation of the overall properties of interest. Thus, clearly, an alternative method for determining gas oils properties is desirable. It has to be more rapid, less expensive and more suitable and if possible, able to operate on-line.

Near infrared (NIR) spectroscopy fulfills all these conditions and, over the last several years, its potential has been explored and largely demonstrated in the fields of the determination of chemical and physical properties of hydrocarbons.²⁻⁶ Its suitability for the determination of such properties comes from different points. First, near infrared spectrometers are reliable in terms of high signal to noise ratio and reproducibility. Second, it is possible to use long optical fibres for on-line plant analysis without significant loss of sensitivity and to multiplex the fibres in order to analyse at different points of the plant with a single spectrometer. Third, the NIR region contains mainly bands that result from overtones and combinations of $-CH$ vibrations such as methyl,

methylene, olefinic and aromatic –CHs. Finally, with chemometrics, it is possible to determine the composition of a sample or a physical property related to the molecular structure.

Unfortunately, there are also some drawbacks which, up to now, have limited the use of this technique in petroleum refineries. It is necessary to develop a good and strong background experience both in the analytical area and in the modeling step.

The scope of this paper is to describe the different steps of the modeling of the refractive index with the partial least square (PLS) algorithm. The refractive index was chosen as an academic example to test the potential of the PLS method to predict a physical parameter determined with high accuracy by a normalised method. Refiners normally use the refractive index to determine the composition of aromatic, paraffinic and naphthenic compounds in gas oils with the so-called *ndM* (refractive index *n*, density *d*, average molecular mass *M*) method referenced ASTM (American Society for Testing and Materials) D-3238.

Experimental section

Gas oil samples

One hundred gas oil samples were provided by the centre of industrial developments (CEDI) of IFP situated in Lyon (France). They were all issued from an hydrotreatment process either from the same feedstock hydrotreated under different conditions with different types of catalysts or from feedstocks with different origins but hydrotreated under the same conditions. The refractive indices of the samples were determined by refractometry at 20°C using the ASTM D-1218 method. The refractive index range was between 1.4470 and 1.5210. The repeatability and the reproducibility of the normalised method are 6×10^{-5} .

Spectroscopy

The NIR spectra (10,000–4000 cm^{-1}) were recorded on a Bomem MB160 spectrometer equipped with a InAs Peltier cooled detector and a halogen source. The spectrum of each sample was recorded twice in a random order and the two spectra were averaged. The resolution was 4 cm^{-1} and the number of accumulations 100. The optical pathlength of the quartz cell (quality QX) was 2 ± 0.02 mm. The spectrometer was purged with dry nitrogen for 10 minutes with a nitrogen flow of 5 L h^{-1} . The sample temperature was 30 ± 0.1 °C and the room temperature 24 ± 2 °C.

Statistical method

PLS regression was performed with a 486/60 MHz PC computer and the analysis routine called PLSplus supplied by Galactic Software. The averaged absorbance-type spectra of each sample were mean-centered by the average spectrum of the database and variance scaled before each modelisation step. We have evaluated the predictive power of the models by obtaining the standard error of prediction (SEP) of the data by cross-validation with one “left-out” sample.^{7–8}

Methodology used to optimise the “refractive index” model using the PLS algorithm

Relationship between the variations of the refractive index and those of the near infrared spectra

The refractive index is a physical property measured in the visible region at 0.583 μm by refractometry. It varies in the region of the absorption bands. It depends on the chemical composition of the sample and namely on the concentrations in aromatic, olefinic and naphthenic compounds and on the average molecular mass.

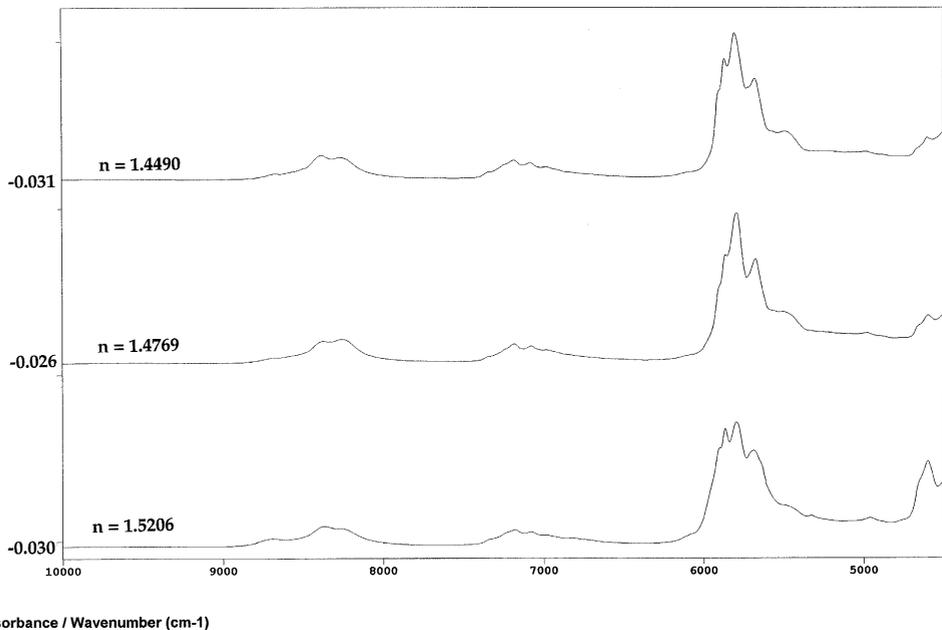


Figure 1. Near infrared spectra of hydrotreated gas oils with different refractive indices.

It also depends on the temperature. For the database samples and for the temperature range 20–30°C, it follows the relation: $\Delta n/\Delta T = 0.0004$. This relation is linear but the variation of the refractive index is not negligible. The near infrared measurements thus have to be realised at a controlled temperature. Before any recordings, the samples were brought to room temperature and left for 10 minutes under the infrared beam. The measured temperature of the sample was, in those conditions, $30 \pm 0.1^\circ\text{C}$. Any temperature variation in this range would involve a negligible refractive index variation compared with the repeatability of the standard method.

A variation of the refractive index may involve different optical phenomena which are added to the chemical fingerprints of the different chemicals. As the spectra were recorded with reference to the empty cell, the difference in the refractive index between the air and the solution implies a change in the reflexion coefficient at the interface between the liquid and the quartz wall of the cell and a defocalisation of the infrared beam which can affect the overall baseline of the spectrum. As illustrated in Figure 1, representing three near infrared spectra of gas oils with different indices, these baseline variations are important and obviously non-linear. A change in absorption coefficients can also occur, due to changes in the local fields and in the optical pathlength. These two latter effects are certainly negligible.

So, this analysis of the property of interest, i.e. the refractive index, shows that it correlates with the sample composition whose fingerprint is present in the spectrum. We can already expect to have to model non-linear effects due to the baseline translation and certainly due to the extent of the database values. It will therefore be interesting to evaluate the performances of different models by previously pretreating the data. The quality of the models will be evaluated according to the values of target parameters referenced to the standard method, i.e. the target SEP and the target R^2 .

Table 1. Optimisation of the wavelength region for modeling the refractive index.

100 samples l = 2 mm averaged, mean-centered and variance scaled spectra.	Wavelength region in cm^{-1}	Number of eigenvectors (P: the probability associated with the F-test applied to compare the SEPs)	SEP	R^2
Regions 1 + 2 + 3	10,000–4500	13 (P = 0.980) 14 (P = 0.768)	0.00022 0.00019	0.9997
Region 1	6400–4500	12 (P = 0.996) 14 (P = 0.699)	0.00019 0.00015	0.9998
Region 2	7800–6400	11 (P = 0.809) 12 (P = 0.72)	0.00044 0.00042	0.9987
Region 3	9000–7800	8 (P = 0.9290) 9 (P = 0.731)	0.00054 0.00049	0.9983

The target *SEP* can be evaluated by the formula:

$$\text{target } SEP^2 = SEP_{\text{standard method}}^2 + SEP_{\text{NIR}}^2$$

with $SEP_{\text{standard method}}$, the *SEP* of the standard method equal to 2.1×10^{-5} and SEP_{NIR} , the *SEP* of the near infrared measurements evaluated via the calibration model equal to 5.6×10^{-5} . So, the target *SEP* is equal to 6×10^{-5} .

The target R^2 can be calculated by the formula:

$$R^2_{\text{target}} = 1 - (SEP_{\text{standard method}}^2 / \text{variance of the database})$$

This parameter is equal to $[1 - (4 \times 10^{-10} / 1.6129 \times 10^{-4})] = 0.999996$ in our case. This is very close to 1 and could, by now, predict that we will need a large number of eigenvectors to model the refractive index.

Optimisation of the model parameters by performing the PLS algorithm

Selection of the wavelength region

Although it is normal practice to perform the PLS regression on the overall wavelength region, we tried to evaluate the potential information content of the different wavelength regions. The 10,000–4000 cm^{-1} region can be divided into three regions respectively attributed to the first harmonic (2ν) and combination ($\nu + 2\delta$) of the $-\text{CH}_2$'s stretching and deformation vibration between 6400 and 4500 cm^{-1} , to the ($2\nu + \delta$) bands in the 7800 and 6400 cm^{-1} region and finally to the (3ν) and [$2(\nu + \delta)$] bands in the 9000 and 7800 cm^{-1} region. The main results obtained by truncature and combination of the different regions are illustrated in Table 1. In terms of best *SEP*, we can see that the best results were obtained in the 6400–4500 cm^{-1} wavelength region. This is not due to the fact that the absorbance is of the order of 1 in this region instead of 0.1 in the others. All the spectra were recorded with an optical pathlength of 1 cm on a Bran+Luebbe InfraProver spectrometer and similar results were obtained. So, the best wavelength region to model the

Table 2. Effect of the size of the database on the modeling of the refractive index.

Region 1 6400–4500 cm ⁻¹ pathlength = 2 mm	Number of eigenvectors (P: the probability associated with the F-test applied to compare the SEPs)	SEP	r ²
42 samples 1.449 < n < 1.4896	10 (P = 0.9794) 11 (P = 0.5)	0.00020 0.00014	0.99969
34 samples 1.4660 < n < 1.469	8 (P = 0.99997) 12 (P = 0.7487)	0.00019 0.00009	0.9543
29 samples 1.4630 < n < 1.467	9 (P = 0.9932) 10 (P = 0.71)	0.00025 0.00018	0.9898

refractive index property seems to be the 6400–4500 cm⁻¹. Nevertheless, the number of eigenvalues necessary to model it with precision is equal to 14 and this seems very high. In the following section, we try to understand this point.

Effect of the data pretreatment

By previously analysing the property, we underlined the effect of the non-linear variation of the spectra baseline. These variations should disappear by derivation. Taking the first one decreases the number of eigenvalues necessary to model the refractive index to 9. Nevertheless, the *SEP* increases from 0.00015 for the model obtained without taking the first derivative to 0.00043. Results obtained by applying a second derivative to the spectra gave even worst results in terms of *SEP* quality [*SEP* = 0.0008, *n* = 6 (*P* = 0.599)]. So, either the baseline translation is at the origin of the important number of eigenvalues and contains an important part of the information necessary to model the property of interest or the derivative introduces too much noise forbidding a precise model. This uncertainty should be solved by looking at the effect of the smoothing of the spectra because it could decrease the noise introduced by the derivative. Decreasing the resolution from 4 to 80 cm⁻¹ or applying a multiplicative scatter correction (MSC) had no effect on the quality of the results.

Effect of the size of the data base

The values of indices in the database are widely spread out. The refractive index measured at 20°C varies from 1.4470 to 1.5210. Almost 85% of the refractive indices values are between 1.455 and 1.475. One idea was thus to reduce the extent of the refractive index by keeping enough spectra. Table 2 summarises the main-results and shows that the number of eigenvalues can be reduced to 10–11 keeping a *SEP* of ~0.00018. This illustrates the effect of the size of the database for which the variation of the refractive index is certainly not linear over the whole refractive index domain. Nevertheless, at this point we are limited by the number of possible samples and from a lab or industrial application viewpoint it is easier to use the most “universal” model. This is the reason why, keeping the model with the 14 eigenvalues, we chose to validate it by different methods and to evaluate its robustness.

Validity and robustness tests of the developed model

Validity tests

The validity of the model developed by PLS regression on 100 averaged/mean-centered/variance scaled/non-derivated near infrared spectra, with 14 eigenvectors, $SEP = 0.00015$ and $r^2 = 0.9998$ was checked either (i) by dividing the database in two (the statistical results obtained by cross-validation on 50 spectra were the same as the one obtained on the 100 spectra. The 50 "left-out" samples were predicted with a SEP equal to 0.00015) or (ii) by predicting non-hydrotreated gas oils produced by different refinery processes.

Outlier detection F-tests⁸ allowed determination of the characteristics of the samples which can be precisely predicted by the model. The samples characteristics have to fulfill these conditions:

$$1.4470 < n_{20^\circ\text{C}} < 1.5210; 7 < \% \text{ aromatics} < 30; 220 < \text{average molecular mass} < 280; \\ \% \text{ olefins} < 0.3; \% \text{ naphthens} < 5 \text{ and } 33 < \% \text{ paraffins} < 51.$$

Robustness tests

The robustness of the model has been evaluated by (i) increasing and decreasing by a factor of two the S/N ratio of the spectrum to be predicted by changing the number of scans from 100 to 25 and 400, (ii) modifying the quantity of residual water in the sampling compartment of the spectrometer and (iii) recording the spectrum of the unknown sample in another cell than the one used for the calibration. The optical pathlength of the cell was equal to the calibration one ± 0.02 mm. Changing these parameters, did not modify the quality of the prediction of the unknown sample.

Transferability of the model

The transfer of calibration of the "refractive index" model ($n = 14$, $SEP = 0.00015$) has been successfully realised on three different Bomem MB160 spectrometers without any previous corrections of the spectra or of the calibration model.

Acknowledgements

The author would like to thank M. Berger for his cooperation in the modeling, Dr F. Wahl for helpful discussions, Bomem Canada for testing the model transferability on different MB160 units, Mr Ottensmeyer (Bran+Luebbe) for testing the model quality on the Infracprover with a 1 cm pathlength cell and D. Allinquant for the overheads design of the corresponding talk.

References

1. J.P. Wauquier, *Le Raffinage du Pétrole*. Editions Technip, Paris, p. 478 (1994).
2. S.J. Foulk and B.E. DeSimas, *Process Control and Quality* **2**, 69 (1990).
3. J.J. Kelly, C.H. Barlow, T.M. Jinguji and J.B. Callis, *Anal. Chem.* **61**, 313 (1989).
4. S.J. Swarin and C.A. Drumm, *Spectroscopy* **7(7)**, 42 (1992).
5. M. Ichikawa, N. Nonaka, I. Takada, and S. Ishimori, *Appl. Spectrosc.* **46(6)**, 966 (1992).
6. A.F. Parisi, L. Nogueiras and H. Prieto, *Analytica Chimica Acta* **238**, 95 (1990).
7. H. Martens and T. Næs, *Multivariate Calibration*. John Wiley and Sons, Chichester, p. 250 (1989).
8. D.M. Haaland and E.V. Thomas, *Anal. Chem.* **60(11)**, 1193 (1988).