

IDENTIFICATION AND MEASUREMENT OF ADULTERATION IN MIZORAM GASOLINE USING FTIR SPECTROSCOPY AND CHEMOMETRIC METHODS

LAL BIAKTLUANGA

Department of Physics, Mizoram University, Aizawl, Mizoram, India.

JOSEF LALHRUAITLUANGA

Department of Physics, Mizoram University, Aizawl, Mizoram, India.

J. LALRAMNGHAKA

Department of Physics, Mizoram University, Aizawl, Mizoram, India.

H.H. THANGA*

Department of Physics, Mizoram University, Aizawl, Mizoram, India.

* Corresponding Author Email: hthanga@rediffmail.com

Abstract

Gasoline adulteration remains a global problem due to its economic, health and environmental impacts. The gasoline adulteration monitoring methods used in the present study were not effective in most developing countries due to the associated implementation costs. Therefore, a fast, reliable, and cheaper method of vibrational spectroscopy in combination with chemometric tools was developed. In this study, sensitive FTIR spectroscopy was proposed in combination with multivariate techniques to analyse adulteration in Mizoram gasoline fuels. Standard solutions were prepared by mixing the gasoline with different proportions of kerosene and methyl tert-butyl ether (MTBE), which are then used for model calibration. The chemometric techniques used were principal component analysis-linear discriminant analysis (PCA-LDA), classification of partial least squares discriminant analysis (PLS-DA) and the regression method (PLS). The developed PCA-LDA classification model has an overall error rate of 12.5%, while PLS-DA has an accuracy of 100% for model. The PLS regression models of kerosene and MTBE with a high significance value of R^2 0.996 and 0.986, RMSEP 0.275 and 0.291 were able to predict adulterant and MTBE concentrations, respectively. However, PCA-LDA achieved better accuracy than PLS-DA in classification analysis, especially on oxygenated gasoline samples.

Keywords: Gasoline, Adulteration, PCA-LDA, PLS-DA, PLSR.

1. INTRODUCTION

Fuel adulteration remains a problem across the world, particularly in developing countries. It refers to the illegal practice of blending of gasoline fuel with impurities or lower-grade quality, cheaper products, and solvents in attempts of increasing profits. It also often results in lower engine performance, leading to the harmful effects of engine tailpipe emissions on the environment. Hence, fuel quality monitoring has become more rigorous not just due to economic and social concerns, but also due to environmental concerns (Vempatapu & Kanaujia, 2017). To prevent fuel adulteration and ensure quality control, the government of India has implemented various measures and actions in accordance with the Marketing Discipline Guidelines (MDG) and a Dealership Agreement substituting between Retail Outlet (RO) dealers and public sector Oil Marketing Companies (OMCs).

OMCs have taken significant steps to improve fuel quality monitoring and surveillance at Retail Outlets. During the last three years of 2022, the government received complaints about supply problems from 3152 outlets from all other states except Mizoram, and 110 were suspected of adulteration

(<https://Pib.Gov.in/PressReleaselframePage.aspx?PRID=1805737>, n.d.).

In India, the most commonly used transport fuel of vehicles is gasoline or petrol. Among the largest petrol pump companies is Indian Oil Corporation Ltd. (IO), with over 34,000 retail outlets across the country (<https://ioil.Com/about-Indianoil>, n.d.). The other major companies are Hindustan Petroleum Corporation Ltd (HP) and Bharat Petroleum Corporation Ltd. (BP), also called reformulated gasoline (RFG). RFG is gasoline blended to burn more cleanly than conventional gasoline and to reduce smog-forming and toxic pollutants in air (<https://www.Epa.Gov/Gasoline-Standards/Reformulated-Gasoline>, n.d.). One of the popular chemical substances blended with gasoline is oxygenated compounds, namely, methyl tert-butyl ethers (MTBE) to reduce pollution carbon monoxide (CO) as well as to increase the octane number of engine (D. Li & Han, 2011). In the state of Mizoram, North-Eastern of India, the largest fuel pump station is also IO, which have over 80% outlets, and the rest of them are 20% of HP and BP. However, the other main problems of gasoline transport fuel are adulteration. In the past few years, the Supreme Court (SC) itself state that auto adulteration is rampant India and largest cases of such adulterated petrol pump outlets are IO and next followed by HP and BP (<https://Auto.EconomicTimes.Indiatimes.Com/News/Oil-and-Lubes/Auto-Fuel-Adulteration-Rampant-in-India-Says-Sc/53882362>, n.d.). Some of the different substances that blended gasoline adulteration are naphtha, kerosene, rubber solvents, aromatics, ethanol, light and heavy aromatics, thinners, white spirits, turpentine, and lubricants. However, in spite of these, adulteration is still remaining difficult to monitor, control and detection (Vempatapu & Kanaujia, 2017).

From this perspective, the detection and identification of adulterated gasoline are challenging tasks because the solvents used for this purpose are complex combinations of hundreds of different molecules, some of which may be found in unadulterated gasoline. The detection of adulteration in automotive fuels using gas chromatography (GC)/mass spectrometry is considered to be one of the most accurate and reliable analysis methods (Skrobot et al., 2005) (Doble et al., 2003) (Pedroso et al., 2008). However, there are several disadvantages, such as the high cost and time, which takes more than 30 minutes to generate a single chromatogram, and sometimes trained technicians are required for the analysis.

Various analytical method, which mostly combined vibrational spectroscopy such as (NIR, MIR, and RAMAN) with chemometric tools are commonly used to analyse gasoline fuels (Ardila et al., 2017; X. Li et al., 2020; Moura et al., 2019). Multivariate methods such as Principal Component Analysis (PCA), Partial Least Squares Discriminant Analysis (PLS-DA), and Partial Least Squares Regression (PLS) have been used for statistical analysis of NIR spectral data to detect and estimate gasoline adulteration (Mabood and Jabeen et al., 2017) (Ahmmed et al., 2022). Raman spectroscopy has also been

employed for checking the quality of fuels rapidly and non-destructively without any further purification (Oliveira et al., 2007) (Liu et al., 2019) (Marinović et al., 2012). However, the disadvantages of Raman spectrometer are fluorescence interfere with the acquisition of spectra, as well as more expensive compared to NIR and FTIR spectrometer.

Numerous studies have been conducted on the applications of spectroscopic techniques in combination with multivariate data analysis. In Mabood et al. investigation, the Super Premium 95 samples were blended with 18 different percentages of Premium 91 gasoline. NIR spectroscopy with multivariate techniques including PCA, PLS-DA and PLSR was used for statistical analysis and this technique was found to be suitable for the identification and quantification of adulterated gasoline (Mabood and Gilani et al., 2017). Barbeira et al. showed that the origin of 1328 gasoline samples sold in the Brazilian market can be determined using hierarchical cluster analysis (HCA) and linear discriminant analysis (LDA) applied to physiochemical parameters (Barbeira et al., 2007). Lasalvia et al. demonstrated that the PCA-LDA model was performed better than the PLS-DA model in terms of accuracy and specificity for FTIR spectrum classification of diesel fuel (Lasalvia et al., 2022). Silva et al. also presented a multivariate classification method of PLS-DA and LDA in conjunction with the variable selection algorithms to detect adulterants in hydrated ethyl alcohol fuel samples using NIR and MIR spectroscopies. The analysis results showed that the LDA and PLS-DA models in the test set achieved a correct prediction rate of 100% when using the MIR spectra. The qualitative analysis results achieved an accuracy of 100% and 98.66% accuracy on the calibration and prediction sets. The FTIR combined with PCA-LDA was also used to classified gasoline adulteration with an accuracy of 96% and the solvent used with an accuracy of 93% (Silva et al., 2012).

In previous work, gasoline adulteration was studied using FTIR spectroscopy in combination with multivariate methods such as PCA, SIMCA and PLS-DA (*Detection and Estimation of Adulterated Gasoline Fuel in India Using FTIR-ATR Spectroscopy with Chemometric Methods Lalbiaktluanga, J. Lalramnghaka, B. Lalremruata, R. Lalrempuia and H.H. Thanga (Communication)*, n.d.). Only gasoline samples collected from the IO company were analysed during the investigation. According to PLSR prediction, a relatively small amount of kerosene was found in a few samples. In the current study, the investigation of gasoline adulteration is conducted on the three companies such as IO, HP and BP. Therefore, the aim of the present work is to detect the adulteration of gasoline fuel from Mizoram (Northeast India) using FTIR in combination with multivariate data analysis of PCA, LDA, PLS-DA and PLSR. Moreover, there are no reports of vibrational analysis of gasoline adulterants using chemometric techniques in India, especially in the Northeastern area.

2. MATERIAL AND METHODS

This work is related to the FTIR spectroscopic combined with multivariate methods to classify and determine the adulterated and formulated gasoline samples, which were obtained from Mizoram.

2.1. Sample Collection and Preparation

A total of 78 gasoline samples were collected from filling stations of three companies, namely, Indian Oil (IO), Hindustan Petroleum (HP) and Bharat Petroleum (BP) in different parts of Aizawl, the capital of Mizoram, Northeastern India. Apart from this, one sample of conventional gasoline was collected from the Indian Oil Corporation Ltd of Marketing Division for standard sample, as well as 11 samples were also collected from nearby retail gasoline pumps of Silchar, Assam and taken as training set for comparison. The standard sample obtained from IOC Ltd, which was supposed to be pure and unadulterated, were mixed with kerosene at the percentage level of 5-20 % v/v. In addition, the oxygenate was also chosen as standard model in order to span slight variations because the RFG samples that differ from conventional gasoline are included in the analysis. To identify the amount of oxygenate in gasoline, 15 standard solutions were prepared by blending conventional fuel with MTBE in the range of 1-15% v/v. Therefore, the standard samples were grouped into two sets: a calibration set representing 80% of the total samples, to create a model for analysis and a validation set representing 20% of the samples, to evaluate the performance of the calibration model.

2.2. FTIR-ATR Spectroscopic Analysis

The infrared spectra of all samples were recorded in ABB Bomem FTIR Zinc Selenide (ZnSe) Attenuated Total Reflectance (ATR) with Horizon MB3000 data acquisition system in the mid-IR region of 3200-650 cm^{-1} , at a resolution of 4 cm^{-1} and 30 number of scans. Moreover, when every data spectrum was going to be recorded, the ATR was first cleaned with tissue paper of new sample at three times to remove any previous sample residue, and finally the sample of ATR cover was then closed to avoid evaporation.

2.3. Chemometric Analysis

For the statistical analysis, The Unscrambler X (version 10.4.1) of CAMO chemometric software was used. First of all, the pre-processing method of Savitzky-Golay algorithm and Standard Normal Variation (SNV) are used for reducing scattering effect (Rinnan, 2014). The Savitzky-Golay algorithm is method for removing spectral noise peaks while preserving spectral information (Zimmermann & Kohler, 2013). SNV removes multiplicative interferences from scattering and particle size effects from spectral data (Aljannahi et al., 2022) (Roger et al., 2020).

The Chemometric analyses are classified into three types: unsupervised PCA, supervised LDA and PLS-DA. The PCA is used to analyse the variance in gasoline fuel and was also performed to detect outliers in samples using the Hotelling T^2 statistics parameter [M Mohammidi]. The approach is to split a data matrix X into a structure and a noise component (Bro & Smilde, 2014), i.e.

$$X = TP^T + E \quad (1)$$

Where T and P represent the score and loading vector, while E are the noise components representing the residual matrix.

The discriminant analysis is a supervised classification method used to build classification models for a set of predefined classes. These models are used to assign unknown samples to the most likely class. Another important application of discriminant analysis is to help interpret differences between sample groups. LDA's discriminant technique is a classification method because the objects must be classified or categorized using the generated known model. The goal of LDA is to find the best fitting parameters for categorizing samples from a built model. The model is then used to categorize unknown samples (Kamble & Dale, 2022).

In PLS-DA, the rules of Partial Least Squares Regression (PLS-R) are used for discrimination or classification. This is used to classify different gasoline samples based on the model created. The PLS models X and Y matrices together to find the hidden variables in X that best predict the hidden variables in Y. Random cross-validations were employed for the internal cross-validation of LDA, PLS-DA and PLSR. While external cross-validation was applied using the test set. To check the predictive ability of the PLSR calibration model, RMSECV and RMSEP were used for internal and external cross validations. In the regression model, the measured RMSECV and RMSEP values should be minimal with a smaller number of factors and with the highest number of R^2 (Mabood, Gilani, et al., 2017). Thus, the root means square error of prediction (RMSEP) and the determinant coefficients (R^2) were evaluated according to Equations 2 and 3 (Câmara et al., 2017).

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

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \quad (3)$$

In classification, specificity and sensitivity are the two important statistical parameters to measure the performance of the classification model. The specificity of a test, also known as the true negative rate, is the proportion of truly negative samples that give a negative result. The sensitivity of a test is also referred to as the true positive rate and is the proportion of actually positive samples that deliver a positive result. Equations 4 and 5 below are used to calculate the specificity and sensitivity of a test.

$$Specificity = \frac{TN}{FP+TN} \quad (4)$$

$$Sensitivity = \frac{TP}{FN+TP} \quad (5)$$

where TP and TN represent the number of true positive and true negative results in the samples, FP and FN represent the number of false positive and false negative results, respectively, in the samples (Santos et al., 2021).

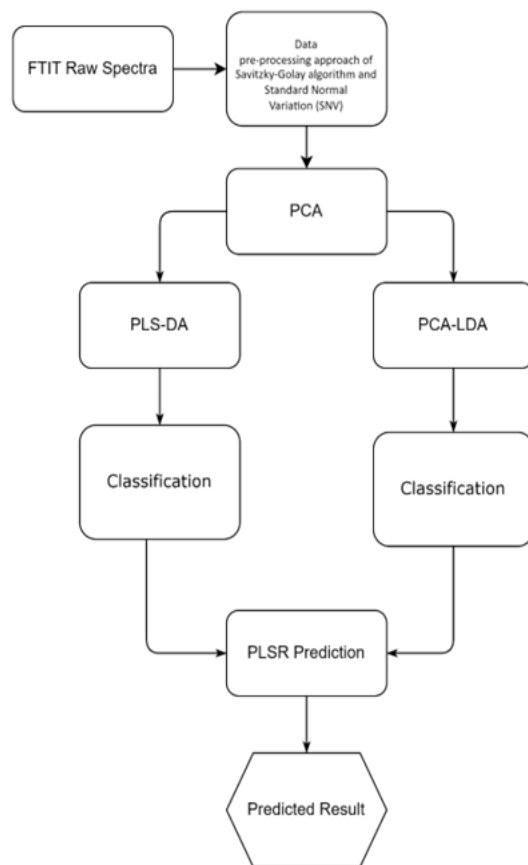


Figure 1: Schematic workflow of this study

3. RESULT AND DISCUSSION

3.1. IR Spectra of gasoline

Figure 2 (a) and (b) represent the original spectra of all gasoline samples before and after pre-processing in the region of 3200–650 cm^{-1} . In order to obtain a better result and reduce scattering noise in the spectra, the transformation of Savitzky-Golay smoothing and Standard Normal Variate (SNV) is applied to the MIR spectra of all samples. The spectrum is characterised by a broad absorption around 3026 cm^{-1} of C-H stretching. The group of strong bands in the 3000–2750 cm^{-1} region is mainly due to the vibrations of C-H stretching. The IR absorption peaks observed in the 1700–1300 cm^{-1} region is due to the C=C and C-H of the skeletal modes. The broad absorption in the region of 1200-1000 cm^{-1} is a characteristic of functional group O-H stretching vibrations. So, the absorption peaks below 1000 cm^{-1} are mostly due to out-of-plane vibrations of aromatic and other cyclic hydrocarbons. Figure 1 clearly shows the differences in formulated and non-formulated samples in the 1200-1000 cm^{-1} region. However, it is difficult to discern between adulterated and unadulterated gasoline samples; hence, chemometric classification methods are used to check the quality of the test samples.

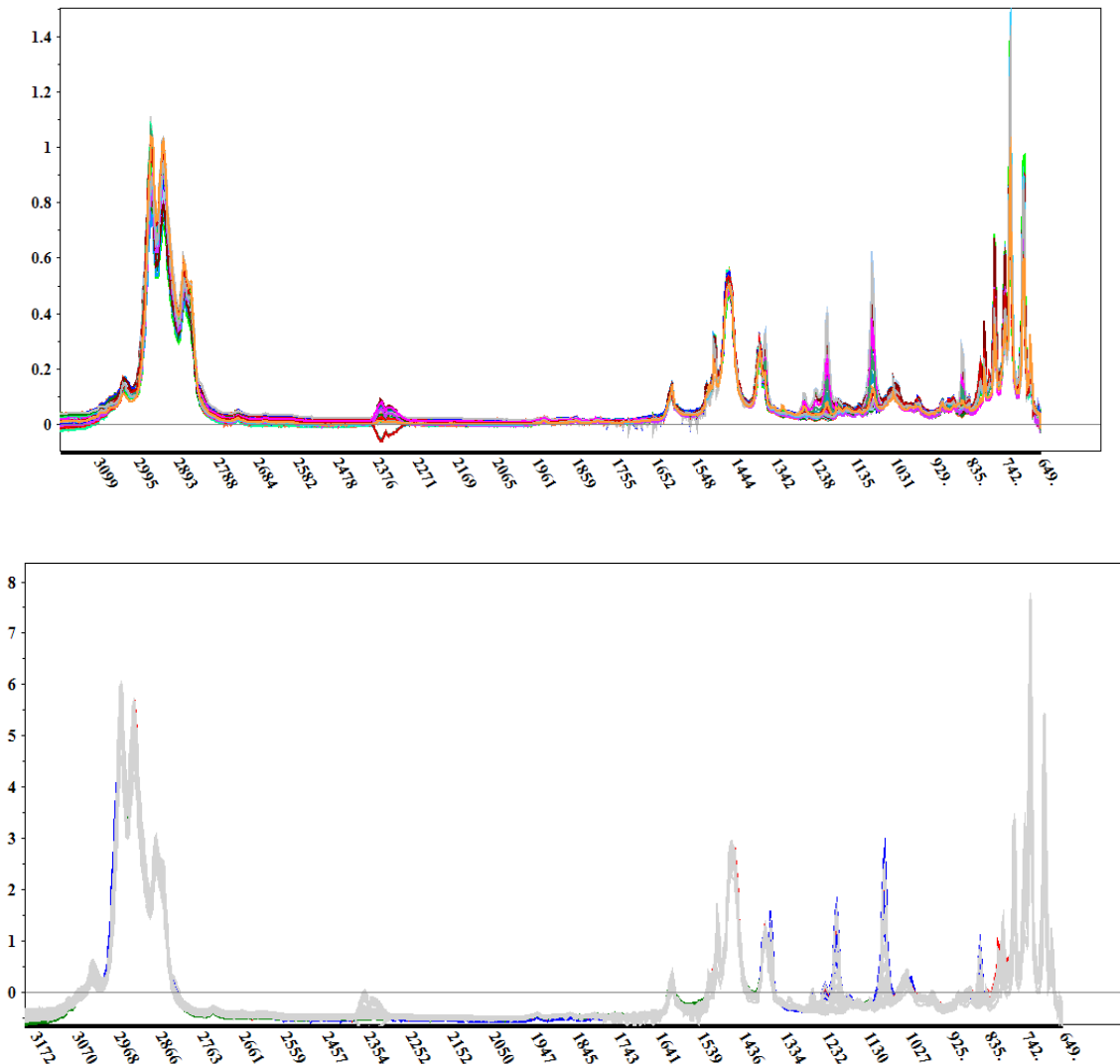


Figure 2 (a) and (b): FTIR spectra of before and after transformation of gasoline samples in the region of 3200-650 cm^{-1}

3.2. PCA Outlier Detection

PCA is an unsupervised technique that was used as an exploratory technique as well as for outlier discovery. The PCA results in Figure 3 shows the score plot, which explains 78% of the total variation. It shows that only the reformulated samples and a portion of the test samples can be differentiated from other samples. There are not any other clear distinctions between the samples. Therefore, the PCA result only demonstrates a slight degree of variation and class overlap rather than clearly demonstrating how the classes are varied.

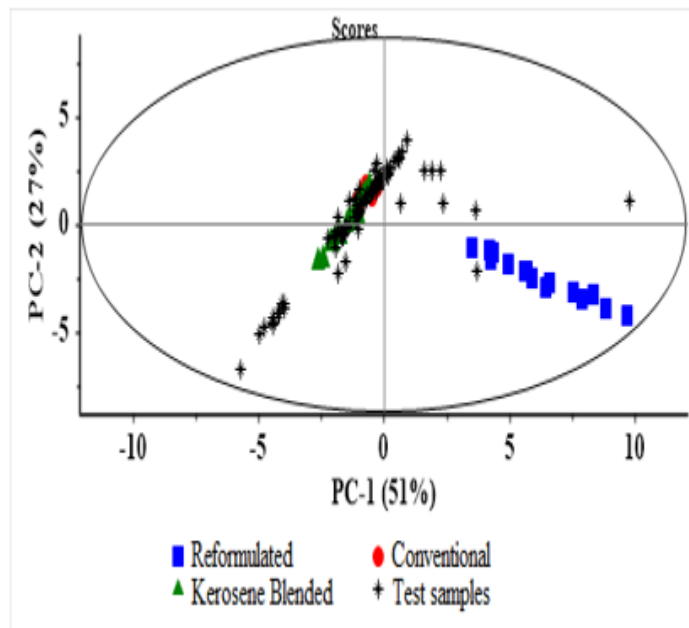


Figure 3: PCA Score Plot of the first two components

3.3. Classification of LDA and PLS-DA Techniques

The supervised technique of LDA is a well-known and useful pattern recognition approach that characterises a separation hyperplane by calculating linear discriminant functions, resulting in optimal class discrimination. In this work, the LDA in combination with PCA was obtained for all samples in the spectral region of 3200–650 cm^{-1} . When it comes to training samples, the first two PCs explaining 90% of the total variation were used for the analysis. The training sets are then conducted with 28 samples (7 conventional, 10 reformulated, and 11 kerosene blended), and the remaining 14 samples are introduced to the validation set (4 conventional, 5 reformulated, and 5 kerosene blended) in order to predict 78 unknown samples. The training sets are used to predict the validation samples by using the LDA technique. The discrimination plot with 80% accuracy is projected on two PCs, as shown in Figure 4. The calibration model has great specificity and sensitivity, as shown in the classification results of validation samples given in Tables 1 and 2. It has been proven that it is suitable for classifying a large data set of commercial gasoline samples. The classification of gasoline samples was done by training sets of LDA, which are shown in Table 3. The result shows that out of 78 test samples, 30 samples from conventional and 6 samples from reformulated samples are assigned to the normal class. An additional 3 samples from the reformulated are placed in the oxygenate class. In addition, 37 conventional fuel samples and 2 reformulated fuel samples were classified as contaminated with kerosene.

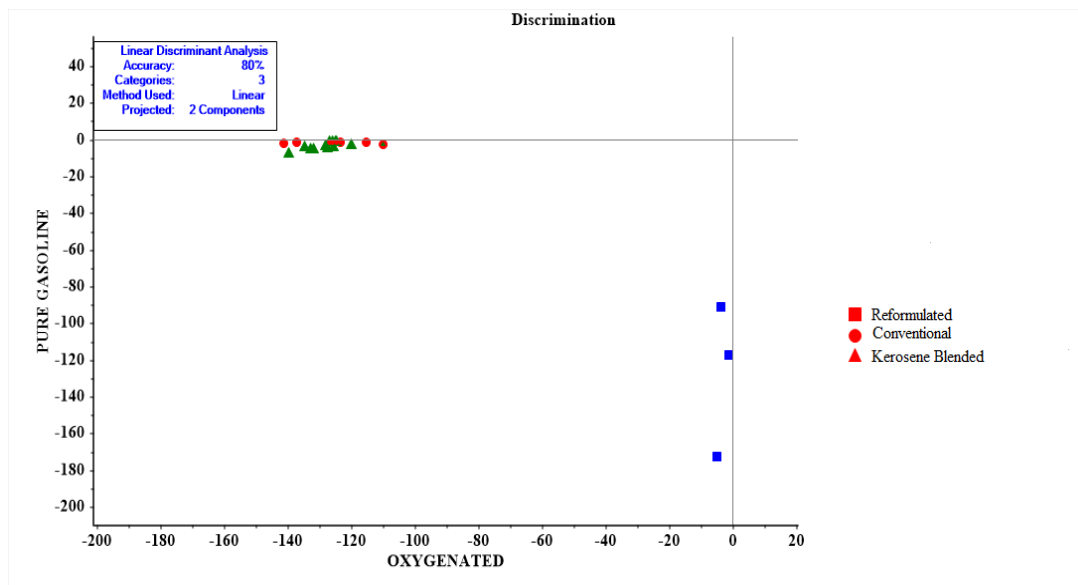


Figure 4: The LDA Discrimination plot of Training Set

Table 1: The LDA predicted result of validation sets

	No of Samples	Classified as Normal Class	Classified as Oxygenate	Classified as Kerosene Blended Class
Conventional	2	1	0	1
Reformulated	2	0	2	0
Kerosene Blended	4	1	0	3

Table 2: The statistical parameters for LDA training set

	No of Samples	TPR	FPR	TNR	FNR	Specificity	Sensitivity
conventional	2	0.50	0	1.00	0.50	0.50	0.50
Reformulated	2	1.00	0	1.00	0	1.00	1.00
Kerosene Blended	4	0.75	0	1.00	0.25	0.75	0.75

Table 3: The LDA classification result of commercial gasoline samples

	Total No of samples	Classified as Normal Class	Classified as Oxygenate Class	Classified as kerosene contaminated
Conventional	67	30	0	37
Reformulated	11	6	3	2

PLS-DA is one of the most recognised supervised classification algorithms in chemometrics. This method is based on partial least-squares regression of continuous predictor variables to find optimal latent variables with the highest covariance. The PLS-DA technique was also employed on the IR spectra of the same samples. In order to predict the unknown samples, the same training and validation sets were also conducted for the analysis. The performance of the training model is shown in Figures 5 (a) and (b). The predicted with deviation plot shows that the validation samples that fall on 1 are predicted as oxygenate or reformulate, samples falling on 0 are conventional gasoline, and -1 are kerosene blended samples. The classification results of validation samples are

given in tables 4 and 5, with a 0% rate of error. The calibration model was shown to be sufficiently suited for data analysis when it was found that both models had 100% Specificity and Sensitivity. Thus, the constructed PLS-DA calibration model was used for analysing each and every commercial gasoline sample. From the classification result of 78 commercial gasoline samples, as shown in Table 6, it became clear that 54 conventional gasoline samples and 6 reformulated gasoline samples are classified in the normal class group. Another 4 reformulated samples are categorized as oxygenate class. However, 13 samples from conventional fuel and 1 sample from reformulated fuel are identified as adulterated with kerosene.

Table 4: The predicted result of PLS-DA validation set

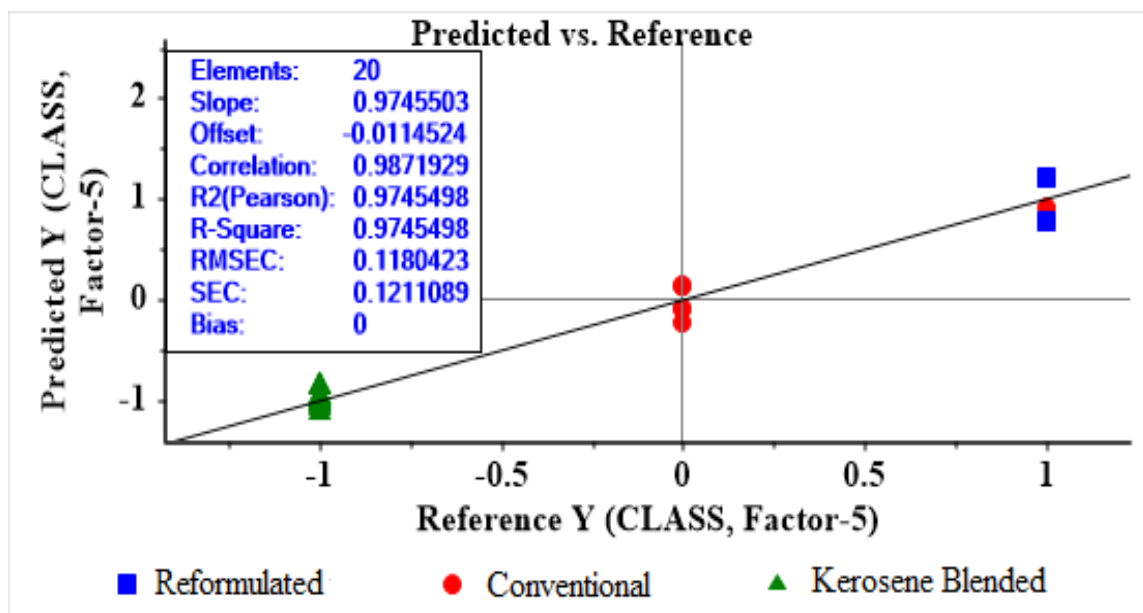
	No of Samples	Classified as Normal Class	Classified as Oxygenate Class	Predicted as Kerosene Blended Class
Conventional	2	2	0	0
Reformulated	2	0	2	0
Kerosene Blended	4	0	0	4

Table 5: The statistical parameters of PLS-DA Training set

	No of Samples	TPR	FPR	TNR	FNR	Specificity	Sensitivity
Conventional	2	1.00	0	1.00	0	1.00	1.00
Reformulated	2	1.00	0	1.00	0	1.00	1.00
Kerosene Blended	4	1.00	0	1.00	0	1.00	1.00

Table 6: The PLS-DA classification result of commercial gasoline samples

	Total No of samples	Classified as Normal Class	Classified as Oxygenate Class	Classified as kerosene contaminated Class
Conventional	67	54	0	13
Reformulated	11	6	4	1



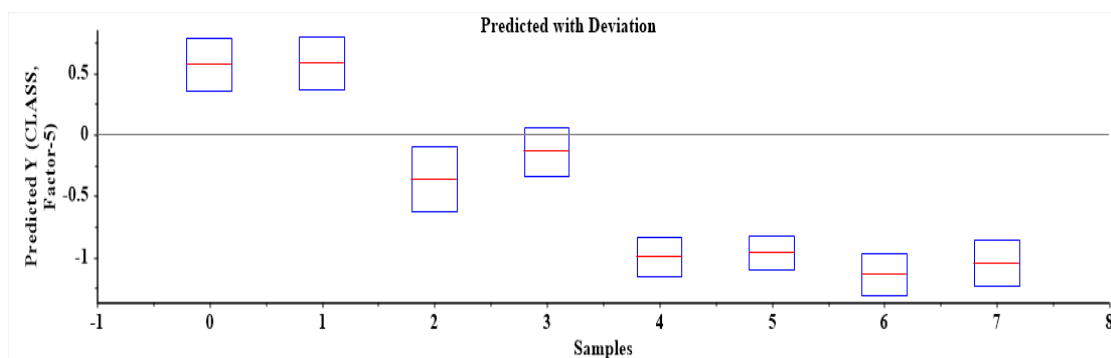


Figure 5: The PLS-DA (a) predicted vs reference plot, (b) predicted vs deviation plot

3.4. Prediction of PLS-R

In order to investigate the differences in oxygenation and adulteration levels based on the classification results of test samples. Therefore, PLS regression models were developed for oxygenated and adulterated samples. Two types of models, kerosene-blend samples and MTBE-oxygenated samples, were constructed for the calibration model, ranging from 5–20% v/v and 6–13% v/v, respectively. The two optimum latent variables (factors) were chosen for the model. It was able to explain 95% and 99% of the total variability for kerosene-blend and Oxygenated calibration samples, respectively. Whereas other parameters of sensitivity and specificity remain the same as in PLS-DA. The PLSR models of both oxygenated and kerosene-blended samples showed a high significant value of R-Square 0.996 and 0.986, while RMSEP 0.275 and 0.291, indicating an effective quantification of the oxygenation and adulteration levels. Therefore, the model is then used to predict the concentrations of MTBE and kerosene in gasoline samples. According to the PLSR predicted results of MTBE in Table 7 and 8, the classification done by LDA in oxygenated samples ranged from 4 to 11% v/v; however, the PLS-DA classified samples ranged from -4 to 11% v/v. The result obtained is very closed to the investigation done by the previous work of Lalramnghaka et al., which was in a range of 1-7% v/v (Lalramnghaka et al., 2023). It is also important to note that the average content of MTBE in Germany was found at 11.9 % w/w (Achten & Püttmann, 2001). In the case of kerosene adulteration, as shown Tables 7 and 8, there are few differences between conventional and reformulated samples. The kerosene content in conventional fuels ranges from 1-16 % v/v for both LDA and PLS-DA categorised samples. The predicted result is found very close to the gasoline adulteration with kerosene in Brazil with maximum concentration of 13.4 % v/v (Teixeira et al., 2008). However, 26-31% were found in each of the reformulated samples. It has also been noted that the result is close to adulteration with kerosene in Ghana in the range of 17-32 % v/v (Dadson et al., 2018). So that, concentration of kerosene is higher in reformulated samples compared to conventional samples. According to the classification and predicted PLSR results, the LDA classifications are more accurate than PLS-DA, particularly in oxygenated samples.

Table 7: The PLSR predicted result of MTBE and Kerosene content from the LDA classification result

	Predicted as Kerosene Contaminated samples	Predicted as Oxygenated samples	Predicted Conc. of MTBE	Predicted Conc. of Kerosene
Conventional	37	0	0	1 to 16 % v/v
Reformulated	2	3	4 to 11 %v/v	30 and 31 % v/v

Table 8: The PLSR predicted result of MTBE and Kerosene content from the PLS-DA classification result

	Predicted as Kerosene Contaminated samples	Predicted as Oxygenated samples	Predicted Conc. of MTBE	Predicted Conc. of Kerosene
Conventional	13	0	0	1 to 16 % v/v
Reformulated	1	4	4 to 11 %v/v	26 %v/v

4. CONCLUSION

In this study, the vibrational spectra of FTIR combined with chemometric tools were used to investigate Mizoram gasoline samples. The spectra of gasoline samples were analysed by supervised PCA-LDA and PLS-DA techniques. The analysis was carried out in two successive phases: First, classification algorithms were built with the three datasets as a training set; then, the performance of the constructed models was evaluated by analysing the classifications performed on the validation set. The obtained results were evaluated according to the values of sensitivity and specificity in the classification of the spectra of the three datasets. According to the expected validation result in classification, PLS-DA has 100% accuracy for the training set. The LDA training set, on the other hand, exhibits less error. It indicates that the performance of the PLS-DA model is more sensitive than that of the LDA model. However, the PLS-R model with a good fitting value of R-Square and RMSE predicted result proved that the classification performed by LDA on gasoline samples is more accurate than PLS-DA, particularly in oxygenate or reformulate samples.

Acknowledgement

Financial support from the project no. EMR/2016/005525 of Science and Engineering Research Board (SERB), Department of Science and Technology (DST), New Delhi, India is gratefully acknowledged. The authors are also thankful to Ministry of Tribal Affairs, Govt. of India for awarding National Fellowship for ST students (NFST) to Mr. J. Lalramnghaka and Mr. Lalbiaktluanga.

References

- 1) Achten, C., & Püttmann, W. (2001). Method for determination of methyl tert-butyl ether in gasoline by gas chromatography. *Journal of Chromatography A*, 910(2), 377–383. [https://doi.org/10.1016/S0021-9673\(00\)01220-6](https://doi.org/10.1016/S0021-9673(00)01220-6)
- 2) Ahmmed, F., Killeen, D. P., Gordon, K. C., & Fraser-Miller, S. J. (2022). Rapid Quantitation of Adulterants in Premium Marine Oils by Raman and IR Spectroscopy: A Data Fusion Approach. *Molecules*, 27(14), 4534. <https://doi.org/10.3390/molecules27144534>
- 3) Aljannahi, A., Alblooshi, R. A., Alremeithi, R. H., Karamitsos, I., Ahli, N. A., Askar, A. M., Albastaki, I. M., Ahli, M. M., & Modak, S. (2022). Forensic Analysis of Textile Synthetic Fibers Using a FT-IR Spectroscopy Approach. *Molecules*, 27(13), 4281. <https://doi.org/10.3390/molecules27134281>

- 4) Ardila, J. A., Soares, F. L. F., Dos Santos Farias, M. A., & Carneiro, R. L. (2017). Characterization of Gasoline by Raman Spectroscopy with Chemometric Analysis. *Analytical Letters*, 50(7), 1126–1138. <https://doi.org/10.1080/00032719.2016.1210616>
- 5) Barbeira, P. J. S., Pereira, R. C. C., & Corgozinho, C. N. C. (2007). Identification of Gasoline Origin by Physical and Chemical Properties and Multivariate Analysis. *Energy & Fuels*, 21(4), 2212–2215. <https://doi.org/10.1021/ef060436l>
- 6) Bro, R., & Smilde, A. K. (2014). Principal component analysis. *Anal. Methods*, 6(9), 2812–2831. <https://doi.org/10.1039/C3AY41907J>
- 7) Câmara, A. B. F., De Carvalho, L. S., De Moraes, C. L. M., De Lima, L. A. S., De Araújo, H. O. M., De Oliveira, F. M., & De Lima, K. M. G. (2017). MCR-ALS and PLS coupled to NIR/MIR spectroscopies for quantification and identification of adulterant in biodiesel-diesel blends. *Fuel*, 210, 497–506. <https://doi.org/10.1016/j.fuel.2017.08.072>
- 8) Dadson, J., Pandam, S., & Asiedu, N. (2018). Modeling the characteristics and quantification of adulterants in gasoline using FTIR spectroscopy and chemometric calibrations. *Cogent Chemistry*, 4(1), 1482637. <https://doi.org/10.1080/23312009.2018.1482637>
- 9) *Detection and Estimation of Adulterated Gasoline Fuel in India using FTIR-ATR Spectroscopy with Chemometric Methods Lalbiaktluanga, J. Lalramnghaka, B. Lalremruata, R. Lalrempuia and H.H. Thanga (communication)*. (n.d.).
- 10) Doble, P., Sandercock, M., Du Pasquier, E., Petocz, P., Roux, C., & Dawson, M. (2003). Classification of premium and regular gasoline by gas chromatography/mass spectrometry, principal component analysis and artificial neural networks. *Forensic Science International*, 132(1), 26–39. [https://doi.org/10.1016/S0379-0738\(03\)00002-1](https://doi.org/10.1016/S0379-0738(03)00002-1)
- 11) <https://auto.economictimes.indiatimes.com/news/oil-and-lubes/auto-fuel-adulteration-rampant-in-india-says-sc/53882362>. (n.d.).
- 12) <https://iocl.com/about-indianoil>. (n.d.).
- 13) <https://pib.gov.in/PressReleaselframePage.aspx?PRID=1805737>. (n.d.).
- 14) <https://www.epa.gov/gasoline-standards/reformulated-gasoline>. (n.d.).
- 15) Kamble, V. H., & Dale, M. P. (2022). Machine learning approach for longitudinal face recognition of children. In *Machine Learning for Biometrics* (pp. 1–27). Elsevier. <https://doi.org/10.1016/B978-0-323-85209-8.00011-0>
- 16) Lalramnghaka, J., Thanga, H. H., & Biaktluanga, L. (2023). Evaluation of gasoline fuel quality using FTIR spectroscopy and multivariate technique: A case study in Aizawl city. *Petroleum Science and Technology*, 41(6), 677–699. <https://doi.org/10.1080/10916466.2022.2091596>
- 17) Lasalvia, M., Capozzi, V., & Perna, G. (2022). A Comparison of PCA-LDA and PLS-DA Techniques for Classification of Vibrational Spectra. *Applied Sciences*, 12(11), 5345. <https://doi.org/10.3390/app12115345>
- 18) Li, D., & Han, X. (2011). Methyl tert-butyl ether. In *Reproductive and Developmental Toxicology* (pp. 617–621). Elsevier. <https://doi.org/10.1016/B978-0-12-382032-7.10046-3>
- 19) Li, X., Liu, Y., Jiang, X., Ouyang, A., Sun, X., & Wang, G. (2020). Determination and quantification of kerosene in gasoline by mid-infrared and Raman spectroscopy. *Journal of Molecular Structure*, 1210, 127760. <https://doi.org/10.1016/j.molstruc.2020.127760>
- 20) Liu, Z., Luo, N., Shi, J., Zhang, Y., Xie, C., Zhang, W., Wang, H., He, X., & Chen, Z. (2019). Raman spectroscopy for the discrimination and quantification of fuel blends. *Journal of Raman Spectroscopy*, jrs.5602. <https://doi.org/10.1002/jrs.5602>

- 21) Mabood, F., Gilani, S. A., Albroumi, M., Alameri, S., Al Nabhani, M. M. O., Jabeen, F., Hussain, J., Al-Harrasi, A., Boqué, R., Farooq, S., Hamaed, A. M., Naureen, Z., Khan, A., & Hussain, Z. (2017). Detection and estimation of Super premium 95 gasoline adulteration with Premium 91 gasoline using new NIR spectroscopy combined with multivariate methods. *Fuel*, 197, 388–396. <https://doi.org/10.1016/j.fuel.2017.02.041>
- 22) Mabood, F., Jabeen, F., Ahmed, M., Hussain, J., Al Mashaykhi, S. A. A., Al Rubaiey, Z. M. A., Farooq, S., Boqué, R., Ali, L., Hussain, Z., Al-Harrasi, A., Khan, A. L., Naureen, Z., Idrees, M., & Manzoor, S. (2017). Development of new NIR-spectroscopy method combined with multivariate analysis for detection of adulteration in camel milk with goat milk. *Food Chemistry*, 221, 746–750. <https://doi.org/10.1016/j.foodchem.2016.11.109>
- 23) Marinović, S., Krištović, M., Špehar, B., Rukavina, V., & Jukić, A. (2012). Prediction of diesel fuel properties by vibrational spectroscopy using multivariate analysis. *Journal of Analytical Chemistry*, 67(12), 939–949. <https://doi.org/10.1134/S1061934812120039>
- 24) Moura, H. O. M. A., Câmara, A. B. F., Santos, M. C. D., Morais, C. L. M., De Lima, L. A. S., Lima, K. M. G., & De Carvalho, L. S. (2019). Advances in chemometric control of commercial diesel adulteration by kerosene using IR spectroscopy. *Analytical and Bioanalytical Chemistry*, 411(11), 2301–2315. <https://doi.org/10.1007/s00216-019-01671-y>
- 25) Oliveira, F. C. C., Brandão, C. R. R., Ramalho, H. F., Da Costa, L. A. F., Suarez, P. A. Z., & Rubim, J. C. (2007). Adulteration of diesel/biodiesel blends by vegetable oil as determined by Fourier transform (FT) near infrared spectrometry and FT-Raman spectroscopy. *Analytica Chimica Acta*, 587(2), 194–199. <https://doi.org/10.1016/j.aca.2007.01.045>
- 26) Pedroso, M. P., De Godoy, L. A. F., Ferreira, E. C., Poppi, R. J., & Augusto, F. (2008). Identification of gasoline adulteration using comprehensive two-dimensional gas chromatography combined to multivariate data processing. *Journal of Chromatography A*, 1201(2), 176–182. <https://doi.org/10.1016/j.chroma.2008.05.092>
- 27) Rinnan, Å. (2014). Pre-processing in vibrational spectroscopy – when, why and how. *Anal. Methods*, 6(18), 7124–7129. <https://doi.org/10.1039/C3AY42270D>
- 28) Roger, J.-M., Boulet, J.-C., Zeaiter, M., & Rutledge, D. N. (2020). Pre-processing Methods. In *Comprehensive Chemometrics* (pp. 1–75). Elsevier. <https://doi.org/10.1016/B978-0-12-409547-2.14878-4>
- 29) Santos, F. D., Santos, L. P., Cunha, P. H. P., Borghi, F. T., Romão, W., De Castro, E. V. R., De Oliveira, E. C., & Filgueiras, P. R. (2021). Discrimination of oils and fuels using a portable NIR spectrometer. *Fuel*, 283, 118854. <https://doi.org/10.1016/j.fuel.2020.118854>
- 30) Silva, A. C., Lira Pontes, L. F. B., Pimentel, M. F., & Pontes, M. J. C. (2012). Detection of adulteration in hydrated ethyl alcohol fuel using infrared spectroscopy and supervised pattern recognition methods. *Talanta*, 93, 129–134. <https://doi.org/10.1016/j.talanta.2012.01.060>
- 31) Skrobot, V. L., Castro, E. V. R., Pereira, R. C. C., Pasa, V. M. D., & Fortes, I. C. P. (2005). Identification of Adulteration of Gasoline Applying Multivariate Data Analysis Techniques HCA and KNN in Chromatographic Data. *Energy & Fuels*, 19(6), 2350–2356. <https://doi.org/10.1021/ef050031l>
- 32) Teixeira, L., Oliveira, F., Dossantos, H., Cordeiro, P., & Almeida, S. (2008). Multivariate calibration in Fourier transform infrared spectrometry as a tool to detect adulterations in Brazilian gasoline. *Fuel*, 87(3), 346–352. <https://doi.org/10.1016/j.fuel.2007.05.016>
- 33) Vempatapu, B. P., & Kanaujia, P. K. (2017). Monitoring petroleum fuel adulteration: A review of analytical methods. *TrAC Trends in Analytical Chemistry*, 92, 1–11. <https://doi.org/10.1016/j.trac.2017.04.011>

- 34) Zimmermann, B., & Kohler, A. (2013). Optimizing Savitzky–Golay Parameters for Improving Spectral Resolution and Quantification in Infrared Spectroscopy. *Applied Spectroscopy*, 67(8), 892–902. <https://doi.org/10.1366/12-06723>

Nomenclature: Abbreviations and Symbols with descriptions.

1. MDG - Marketing Discipline Guidelines.
2. OMC - Oil Marketing Companies.
3. SC – Supreme Court.
4. IO - Indian Oil Corporation Limited.
5. HP - Hindustan Petroleum Corporation Limited.
6. BP - Bharat Petroleum Corporation Limited.
7. MTBE - Methyl tert-Butyl Ether.
8. RFG - Reformulated Gasoline.
9. CO – Carbon Monoxide.
10. Ltd. – Limited.
11. ZnSe – Zinc Selenide.
12. ATR - Attenuated Total Reflectance.
13. NIR - Near Infrared radiation.
14. MIR - Mid Infrared Radiation.
15. GC – Gas Chromatography.
16. FTIR - Fourier Transform Infrared Radiation.
17. PCA - Principal Component Analysis.
18. SIMCA - Soft Independent Modelling of Class Analogy.
19. LDA - Linear Discriminant Analysis.
20. PLS-DA - Partial Least Squares Discriminant Analysis.
21. PLSR - Partial Least Square Regression.
22. HCA - Hierarchical Cluster Analysis.
23. SNV - Standard Normal Variate.
24. TP - True Positive.
25. FP - False Positive.
26. TN - True Negative.
27. FN - False Negative.
28. RMSE - Root Mean squared Error.
29. R^2 - Squared correlation coefficient – explained variance.
30. T - Transpose of a matrix or vector.
31. X - Uses to describe explanatory, input or independent variables.

32. Y - Output or response variables.
33. P - Represent the score and loading vector.
34. E - The noise components representing the residual matrix.
35. % - Percentage.
36. % v/v - Volume concentration, which stands for volume per volume.
37. n – Number of samples.
38. y_i – Represents the label vector of the i th sample.
39. \hat{y} – Represents the predictions of the samples using the proposed method.
40. cm^{-1} - Centimetre inverse, SI unit of wavenumber.