Selection of the Effective Characteristic Spectra Based on the Chemical Structure and Its Application in Rapid Analysis of Ethanol Content in Gasoline

Ke Li, Chi Zhang, Biao Du, Xiaoping Song, Qi Li, and Zhengdong Zhang*  

**Cite This:** ACS Omega 2022, 7, 20291–20297

**ABSTRACT:** Near-infrared (NIR) spectroscopy analysis is one of the most rapid detection methods for determining ethanol content in gasoline. Wavelength selection is a key step in the multivariate calibration analysis of NIR spectroscopy. To improve detection accuracy of ethanol content in gasoline and provide a simpler interpretation, we established NIR spectroscopy, a rapid analysis method based on the effective characteristic spectra. Five effective characteristic spectral bands were used according to the molecular structure of ethanol, followed by the development of four modeling schemes. The four modeling schemes spectra, NIR full spectra, and variable importance projection (VIP) spectra were used for modeling and analysis. The model was established based on the effective characteristic spectra without the interference spectra of aromatic hydrocarbons, achieving the best model performance. In addition, the model was further evaluated by internal cross-validation and external validation. The model’s evaluation parameters were as follows: the root mean square error of cross-validation (RMSECV) was 0.6193, the correlation coefficient of internal cross-validation ($R^2_{CV}$) was 0.9995, the root mean square error of prediction (RMSEP) was 0.5572, and the correlation coefficient of external prediction validation ($R^2_P$) was 0.9995. The effective characteristic spectra model had smaller RMSEP and RMSECV values, and larger $R^2_{CV}$ and $R^2_P$ values compared to the full spectra and VIP spectra models. In conclusion, the effective characteristic spectra model had the highest accuracy and could provide rapid analysis of the ethanol content in gasoline.

**INTRODUCTION**

Ethanol-gasoline is a new substitute fuel made by mixing predetermined volume ratios of ordinary gasoline and ethanol fuel. Ethanol fuel is a renewable resource, and use of ethanol-gasoline can significantly alleviate energy demand pressure. Ethanol-gasoline with a high oxygen content can be fully burned, and effectively reduces the emission of carbon monoxide, hydrocarbons, and other harmful substances in automobile exhausts. As a low-carbon, clean, and high-quality green fuel, ethanol-gasoline has received considerable attention in recent years. Random inspections of the ethanol-gasoline quality at various gas stations highlighted some problems, for example, the ethanol content was either too high or too low. However, use of substandard ethanol-gasoline led to engine speed instability and engine failure. Therefore, it is necessary to conduct spot checks to ensure that the quality of the refined oil including ethanol-gasoline in the market is of a high standard.

Currently, the conventional method for determining ethanol content in gasoline is gas chromatography, which is a reliable method for determining ethanol content in oil. However, this method has disadvantages, including long analysis times and the inability to meet the needs of on-site analysis and real-time detection due to bulky instruments that cannot be easily carried and moved around for analysis. To solve this problem, some rapid analytical methods for measuring ethanol content in gasoline have been developed, for example, Raman spectroscopy and Near-infrared (NIR) spectroscopy. NIR is a fast, non-destructive analytical method that consumes small amounts of reagents. NIR spectroscopy combined with multivariate statistical analytical methods, such as partial least squares or principal component analysis, has been widely used in rapid analysis of ethanol content in gasoline. Mabood et al., using partial least squares and principal component analysis, has been widely used in rapid analysis of ethanol content in gasoline. Mabood et al., using partial least squares and principal component analysis, has been widely used in rapid analysis of ethanol content in gasoline. Mabood et al., using partial least squares and principal component analysis, has been widely used in rapid analysis of ethanol content in gasoline. Mabood et al., using partial least squares and principal component analysis, has been widely used in rapid analysis of ethanol content in gasoline. Mabood et al., using partial least squares and principal component analysis, has been widely used in rapid analysis of ethanol content in gasoline. Mabood et al., using partial least squares and principal component analysis, has been widely used in rapid analysis of ethanol content in gasoline. Mabood et al., using partial least squares and principal component analysis, has been widely used in rapid analysis of ethanol content in gasoline. Mabood et al., using partial least squares and principal component analysis, has been widely used in rapid analysis of ethanol content in gasoline. Mabood et al., using partial least squares and principal component analysis, has been widely used in rapid analysis of ethanol content in gasoline. Mabood et al., using partial least squares and principal component analysis, has been widely used in rapid analysis of ethanol content in gasoline. Mabood et al., using partial least squares and principal component analysis, has been widely used in rapid analysis of ethanol content in gasoline. Mabood et al., using partial least squares and principal component analysis, has been widely used in rapid analysis of ethanol content in gasoline. Mabood et al., using partial least squares and principal component analysis, has been widely used in rapid analysis of ethanol content in gasoline. Mabood et al., using partial least squares and principal component analysis, has been widely used in rapid analysis of ethanol content in gasoline. Mabood et al., using partial least squares and principal component analysis, has been widely used in rapid analysis of ethanol content in gasoline. Mabood et al., using partial least squares and principal component analysis, has been widely used in rapid analysis of ethanol content in gasoline. Mabood et al., using partial least squares and principal component analysis, has been widely used in rapid analysis of ethanol content in gasoline. Mabood et al., using partial least squares and principal component analysis, has been widely used in rapid analysis of ethanol content in gasoline. Mabood et al., using partial least squares and principal component analysis, has been widely used in rapid analysis of ethanol content in gasoline. Mabood et al., using partial least squares and principal component analysis, has been widely used in rapid analysis of ethanol content in gasoline. Mabood et al., using partial least squares and principal component analysis, has been widely used in rapid analysis of ethanol content in gasoline.
Furthermore, Carneiro et al.\textsuperscript{15} compared and analyzed the difference between NIR spectroscopy and mid-infrared (MIR) spectroscopy in the rapid analysis of the methanol content in ethanol-gasoline, demonstrating a good prediction performance for the partial least squares model based on NIR spectroscopy.

Wavelength selection (or variable selection) is a key step in multivariate calibration analysis of NIR spectra.\textsuperscript{16} Appropriate wavelength selection can remove uninformative and interfering variables in the spectra to obtain better model prediction performance and improve interpretability.\textsuperscript{17} To date, several NIR wavelength selection methods have been developed. The most commonly used methods include uninformative variable elimination (UVE),\textsuperscript{18} successive projection analysis (SPA),\textsuperscript{19} moving window partial least squares (MWPLS),\textsuperscript{20} interval partial least squares (iPLS),\textsuperscript{21} simulated annealing (SA),\textsuperscript{22} genetic algorithm (GA),\textsuperscript{23} ant colony optimization (ACO),\textsuperscript{24} and variable importance in projection (VIP).\textsuperscript{25} These wavelength selection methods are based on spectral data, and they have their respective characteristics and advantages in selected applications. This means that none of these methods can achieve good model performance for all types of spectral data. In addition, these methods cannot solve the problem of “false correlations” between spectral variables and properties caused by environmental factors or instrument performance factors. Therefore, the above mentioned methods have no commercial application in the rapid analysis of ethanol-gasoline. To improve the accuracy of the rapid analysis of ethanol content in gasoline using NIR spectroscopy, it is necessary to establish an effective method for selecting characteristic spectra.

The chemical structure determines the properties of substances, and different functional groups or chemical bonds in chemical substances have their corresponding characteristic spectral bands.\textsuperscript{26–28} Modeling using characteristic spectral bands based on a chemical structure can improve the predictive performance and interpretability of a model.\textsuperscript{16,29} Based on this principle, we established a rapid analytical method for ethanol content analysis based on the effective characteristic spectra. Four calibration models were established by analyzing the effective characteristic spectral bands corresponding to each chemical bond of the ethanol molecule. The four models were compared with the models established by the full spectra and variables selected by the VIP method, to verify the accuracy of the effective characteristic spectra model.

**RESULTS AND DISCUSSION**

**Effective Characteristic Spectra Selection.** Figure 1 shows NIR spectra of 44 ethanol-gasoline samples. An increase in ethanol concentration/content in the gasoline causes an inconsistent variation in the absorption intensity at different wavenumbers. In some spectral ranges, the NIR absorbance increased significantly with an increase in ethanol concentration in the gasoline, while in other spectral ranges, the change was not noticeable. The absorption of ethanol in the NIR spectral region originates from the C–H and O–H bonds in the molecular structure. Gasoline is mainly composed of hydrocarbons, and changes in the absorption strength of the C–H bond were not apparent, whereas changes in the absorption strength of the O–H bond were apparent. In the characteristic absorption spectra of ethanol, the spectra ranged from 6060.171 to 7141.113 cm\(^{-1}\), covering the first overtone of O–H stretching.\textsuperscript{30–32} The absorption peaks in the region from 4661.104 to 5000.515 cm\(^{-1}\) are related to the combined absorption frequency of the O–H stretching and bending vibrations.\textsuperscript{33,34} The spectral range from 6450.422 to 7407.241 cm\(^{-1}\) resulted from the first overtone of the combination band from C–H + C–H and C–H + C–C stretching;\textsuperscript{35,36} and the spectral range from 5660.050 to 6001.389 cm\(^{-1}\) corresponds to the first overtone of C–H stretching.\textsuperscript{37} The absorption peaks in the region from 8300.121 to 8500.682 cm\(^{-1}\) are caused by the second overtone of C–H from the methyl group,\textsuperscript{15,38} while the peaks below 6060.171 cm\(^{-1}\) contain the first overtone from aromatic C–H stretching.\textsuperscript{35} Characteristic spectra of the other components may adversely affect ethanol content. The above characteristic absorption spectra of ethanol were ultimately selected as the effective characteristic spectra to build a calibration analysis model.

To explore the optimal effective characteristic spectra and eliminate the adverse effects of other interfering spectra, we established four modeling schemes that combined different effective characteristic spectra. The four modeling schemes are listed in Table 1. Scheme 1 contains all the characteristic spectral regions. Scheme 2 contains the characteristic spectral region of the hydroxyl group. Scheme 3 contains the characteristic spectral region of the hydroxyl group and excludes the interference spectra of aromatic groups. Scheme

![Figure 1. Raw NIR spectra of 44 ethanol-gasoline samples. The dashed rectangles 1, 2, 3, 4, and 5 represent the spectral ranges of 4661.104–5000.515 cm\(^{-1}\), 5660.050–6001.389 cm\(^{-1}\), 6000.171–7141.113 cm\(^{-1}\), 6450.422–7407.241 cm\(^{-1}\), and 8300.121–8500.682 cm\(^{-1}\), respectively.](image-url)

| Table 1. Effective Characteristic Spectral Region Modeling Scheme |
|---------------------|---------------------|
| modeling scheme | spectral combination (cm\(^{-1}\)) | combination strategy |
| 1 | 4661.104–5000.515 | all the characteristic spectra region |
| | 5660.050–7407.241 | |
| | 8300.121–8500.682 | |
| 2 | 4661.104–5000.515 | characteristic spectra region of hydroxyl group |
| | 6060.171–7141.113 | |
| 3 | 6060.171–7141.113 | characteristic spectra region of hydroxyl group and excludes the interference spectra of aromatic groups |
| 4 | 8300.121–8500.682 | all the characteristic spectra region and excludes the interference spectra of aromatic groups |

20292  
https://doi.org/10.1021/acsomega.2c02282
ACS Omega 2022, 7, 20291–20297
4 includes all the characteristic spectral regions and excludes the interference spectra of aromatic groups.

**Results of the Preprocess Methods’ Optimization.**

The spectral signal of ethanol-gasoline samples may be disturbed by noise, stray light, baseline drift, and other factors, which may result in irrelevant information in the NIR spectra and affect the accurate analysis of the ethanol content. Five preprocessing methods, namely savitzky-golay smoothing (SGM, 13 points with a second polynomial order), the savitzky-golay derivative (SGD, first derivative, 13 points with a second polynomial order), multiplicative signal correction (MSC), vector normalization (VN), and the standard normal variate (SNV), were used to process the NIR spectra of ethanol-gasoline.

Table 2 shows the performance of the calibration analysis model for the full spectra based on different preprocessing methods. Among the five preprocessing methods, the calibration analysis model based on the first derivative processing methods had the smallest root mean square error of cross-validation (RMSECV) and RMSEP values and the processing methods had the smallest root mean square error (RMSEP) values. Among the factors, which may result in irrelevant information in the NIR spectra, disturbed by noise, stray light, baseline drift, and other components, more accurately reflects the ethanol content in gasoline.

To further prove the effectiveness of the calibration analysis model established by the effective characteristic spectra, it was compared with the modeling results of spectral variables screened by the VIP method. After 3112 spectral variables were preprocessed using the first derivative, 640 variables (VIP spectra) with VIP values greater than 1 were screened out. A partial least squares model was then established to conduct predictive analysis on the validation set samples. As shown in Table 3, the RMSECV value of the VIP spectra model is lower than that of the full spectra model, while the RMSEP value is higher. The \( R^2_{CV} \) and \( R^2_p \) values of the full spectra and VIP spectra models were nearly consistent. Compared with Scheme 3, the RMSECV value of the VIP spectra model decreased from 0.6288 to 0.6193, the RMSEP value decreased from 0.8859 to 0.5572, while \( R^2_{CV} \) increased from 0.9991 to 0.9995, and \( R^2_p \) increased from 0.9977 to 0.9991. The results show a superior prediction performance of the model established in Scheme 3. The effective characteristic spectra model based on the chemical structure achieved highly accurate prediction results.

Thus, Scheme 3 was selected for modeling and analyzing the ethanol content in gasoline.

**Application of the Models Established By Effective Characteristic Spectra, Full Spectra, and VIP Spectra.** The 14 ethanol-gasoline samples in the prediction set were analyzed using a calibration analysis model established based on Scheme 3, full spectra, and VIP spectra. Internal cross-
validation and external validation were used to evaluate and analyze the results. Figure 3 shows the internal cross-validation results. The RMSECV value of the partial least squares model established based on Scheme 3 or VIP spectra is better than that of the model based on the full spectra, indicating that the characteristic spectra model can eliminate irrelevant spectral variables to improve accuracy. Compared to the full spectra and the VIP spectra models, the RMSECV value of the model established based on Scheme 3 decreased by 10.71 and 1.51%, respectively, indicating that the spectral information of the effective characteristic spectra is more “characteristic” and could more accurately reflect the sample content. In addition, the model based on Scheme 3 has the largest $R^2_{CV}$ value (close to 1), indicating that the model has the highest accuracy, and the greatest correlation between the predictive value and the reference value during the internal cross-validation.

External validation results are shown in Figure 4. In Figure 4c, all sample points fell on the fitting curve after the prediction of the model based on the effective characteristic spectra. For the full spectra model (Figure 4a) and the VIP spectra model (Figure 4b), some sample points did not fall on the fitting curve. This indicates that the smallest deviation between the predicted and reference values was obtained by the effective characteristic spectra model. Furthermore, the RMSEP values of the three models also explain this result. Compared to the full spectra and VIP spectra models, the RMSEP value of the effective characteristic spectra model decreased by 49.14 and 37.10%, respectively. The effective characteristic spectra model achieved the highest prediction accuracy. Using the effective characteristic spectra model to predict the validation set sample, the linear equation between the predicted value and the reference value is $Y = 1.0039X - 0.1275$, and the correlation coefficient of the external validation prediction is $R^2_{P} = 0.9991$, indicating that the predicted value is consistent with the reference value. In conclusion, modeling based on the

**Figure 3.** Internal cross-validation results of the ethanol content in gasoline using partial least squares models based on different spectral ranges.

**Figure 4.** External validation results of the ethanol content in gasoline using partial least squares models based on different spectral ranges. (a) Full spectra; (b): VIP characteristic spectra; (c) effective characteristic spectra.
effective characteristic spectra yielded the best prediction performance.

■ CONCLUSIONS

A rapid NIR analysis method based on effective characteristic spectra was developed and successfully applied to detect ethanol content in gasoline. Four effective characteristic spectral modeling schemes were developed based on the effective characteristic spectra of ethanol. To compare the model performance, models based on full spectra and VIP spectra were also established.

After optimizing five main preprocess methods, the model based on the first derivative showed the best performance. The spectral range is 6060.171–7141.113 cm⁻¹ for the model established by Scheme 3, which is the characteristic spectra that only contain the spectra of hydroxyl groups without the interference of aromatic hydrocarbon group spectra, had the smallest RMSECV and RMSEP values and the best model performance. These results of the application demonstrated that, compared to the full spectra model, the RMSECV and RMSEP values of the Scheme 3 model decreased by 10.71 and 49.14%, respectively. Compared to the VIP spectra model, the RMSECV and RMSEP values of the Scheme 3 model decreased by 1.51 and 37.10%, respectively. The ethanol content in gasoline was accurately and rapidly analyzed using the optimal effective characteristic spectral modeling scheme, with $R^2_{CV} = 0.9995$, RMSECV = 0.6193, $R^2_P = 0.9991$, and RMSEP = 0.5572.

Thus, the rapid analysis method of NIR based on effective characteristic spectra is a specific analysis method with high accuracy, which can be applied to the rapid analysis of more characteristic indicators in complex sample systems such as gasoline and diesel.

■ MATERIALS AND METHODS

Preparation of the Samples. Octane-rated gasoline samples (92 and 95) without ethanol were obtained from different gas stations in Beijing, including different batches of gasoline from various gas stations under PetroChina, Sinopec, and Sinochem groups. Forty-four ethanol-gasoline samples were prepared by adding a predetermined amount of ethanol (Analytical Reagent, Fuchen (Tianjin) Chemical Reagent Co., Ltd.) to gasoline (sample details in Table 4). The prepared ethanol-gasoline samples have different solvent components, which can simulate actual ethanol-gasoline samples. The ethanol-gasoline samples were grouped into two sets, namely, the training set (samples not marked with * in Table 4), which was used to establish the calibration analysis model, and the prediction set (samples marked with * in Table 4), which was used to evaluate the prediction performance of the model. The ethanol concentration range in the calibration set was 0.5–80% (volume fraction), covering the range of ethanol in normal and adulterated ethanol-gasoline, indicating that the established calibration model is highly representative.

Acquisition of NIR spectra. The NIR spectra of the ethanol-gasoline samples were measured using a Frontier NIR spectrometer (Antaris II, Thermo Fisher Scientific (China) Co., Ltd.) equipped with an InGaAs detector and a tungsten-halogen source. The spectra were obtained by co-adding 32 scans in the transmission mode. The spectral scan range was 4000–10 000 cm⁻¹ at 4 cm⁻¹ digital resolution. Each sample was loaded and tested three times in a 1 mm pathlength cuvette, and average spectra were used for further analysis. The spectra were recorded at room temperature and the humidity was maintained at 40%.

Data Preprocessing and Model Evaluation. The Unscrambler X software was used for the preprocessing and statistical analysis of the NIR spectral data. SGM, SGD, MSC, VN, and SNV were selected to denoise the spectral data and then compared with the preprocess results. SGM can effectively reduce noise in the spectral signal and improve the signal-to-noise ratio. SGD can reduce the drift of NIR spectroscopy and interference of certain background signals. MSC is used to reduce the scattering effects of mechanical impurities in gasoline samples. VN can improve the role of characteristic spectral segments in modeling and eliminate the adverse effects caused by large-scale differences, while SNV can reduce the scattering effects of sample surfaces and the influence of optical path changes. After the spectral data were preprocessed using the five methods described above, the partial least squares (P) method was used to build the calibration model. Finally, an optimal preprocessing method was selected.

$R^2_{CV}$ and RMSECV were selected as parameters to evaluate the quality of the calibration model. The precision of the calibration model was better when the $R^2_{CV}$ value was closer to 1 and the RMSECV value was low. $R^2_P$ and RMSEP are typically used as parameters to evaluate the predictive ability of the calibration model. The accuracy of the prediction model improves when the $R^2_P$ value is closer to 1, and the RMSEP value is low. For an optimal calibration model, the RMSECV and RMSEP values should be close. Therefore, in this study, we chose four parameters, RMSECV, $R^2_{CV}$, RMSEP, and $R^2_P$ to compare the effects of the full spectra, VIP spectra, and effective characteristic spectra models. RMSECV, RMSEP, and $R^2$ ($R^2_{CV}$ or $R^2_P$) were calculated using the following equations.

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

(1)

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

(2)

Table 4. Ethanol Volume Fraction in Ethanol-gasoline Samples

| no. | volume fraction (%) | no. | volume fraction (%) | no. | volume fraction (%) | no. | volume fraction (%) |
|-----|---------------------|-----|---------------------|-----|---------------------|-----|---------------------|
| 1   | 0.5                 | 11  | 23                  | 22  | 34                  | 36  |
| 2   | 1.0                 | 13  | 24                  | 23  | 35                  | 38  |
| 3   | 3.0                 | 15  | 25                  | 26  | 37                  | 40  |
| 4   | 4.0                 | 16  | 27                  | 25  | 38                  | 50  |
| 5   | 5.0                 | 17  | 26                  | 27  | 39                  | 55  |
| 6   | 6.0                 | 18  | 25                  | 28  | 40                  | 60  |
| 7   | 7.0                 | 19  | 29                  | 28  | 41                  | 65  |
| 9   | 8.0                 | 20  | 31                  | 30  | 42                  | 70  |
| 10  | 9.0                 | 21  | 32                  | 32  | 43                  | 75  |
| 11  | 10.0                | 22  | 33                  | 34  | 44                  | 80  |

Note: Samples marked with * are included in the validation set.
\[ R^2 = \frac{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \] (3)

Note: \( n \) is the number of training set samples, \( m \) is the number of validation set samples, \( y_i \) indicates the measurement results obtained by standard methods for sample \( i \), \( \bar{y} \) is the mean value of \( y_i \) and \( \hat{y}_i \) indicates the predicted results of sample \( i \) based on spectral modeling.

**AUTHOR INFORMATION**

Corresponding Author
Zhengdong Zhang — Center for Environmental Metrology, National Institute of Metrology, Beijing 100029, P. R. China; orcid.org/0000-0001-8754-3822; Email: zhangzhengdong@nim.ac.cn

Authors
Ke Li — Center for Environmental Metrology, National Institute of Metrology, Beijing 100029, P. R. China; orcid.org/0000-0001-9015-7279
Chi Zhang — Sinochem Oil Marketing Co., Ltd., Beijing 100069, P. R. China
Biao Du — Beijing Yixingyuan Petrochemical Technology Co., Ltd., Beijing 101301, P. R. China
Xiaoping Song — Center for Environmental Metrology, National Institute of Metrology, Beijing 100029, P. R. China
Qi Li — Center for Environmental Metrology, National Institute of Metrology, Beijing 100029, P. R. China

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.2c02282

Notes
The authors declare no competing financial interest.

**ACKNOWLEDGMENTS**

This research was financially by the State Administration for Market Regulation Science and Technology Program, China (No. 2021MK153), the National Institute of Metrology, China (No. AKYZZZ2131), and the Institute-Enterprise Horizontal Scientific Research Program (No. JFSW2102).

**NOMENCLATURE**

NIR near-infrared
MIR mid-infrared
RMSECV root mean square error of cross-validation
RMSEP root mean square error of prediction
\( R^2 \) correlation coefficient of internal cross-validation
\( R^2_e \) correlation coefficient of external prediction validation
UVE uninformative variable elimination
SPA successive projection analysis
MWPLS moving window partial least squares
iPLS interval partial least squares
SA simulated annealing
GA genetic algorithm
ACO ant colony optimization
VIP variable importance projection
SGM savitzky-golay smoothing
SGD savitzky-golay derivative
MSC multiplicative signal correction
VN vector normalization
SNV standard normal variate

**REFERENCES**

(1) Safieddin Ardebili, S. M.; Solmaz, H.; Ipci, D.; Calam, A.; Mostafaei, M. A Review on Higher Alcohol of Fusel Oil as a Renewable Fuel for Internal Combustion Engines: Applications, Challenges, and Global Potential. *Fuel* 2020, 279, No. 118516.

(2) Mabood, F.; Boqué, R.; Hamaed, A.; Jabeen, F.; Al-Harrasi, A.; Hussain, J.; Alameri, S.; Albroumi, M.; Al Nabhan, M. M. O.; Naureen, Z.; et al. Near-Infrared Spectroscopy Coupled with Multivariate Methods for the Characterization of Ethanol Adulteration in Premium 91 Gasoline. *Energy Fuels* 2017, 31, 7591–7597.

(3) Hu, J.; Liu, Y. D.; Hao, Y.; Sun, X. D.; Ouyang, A. G. Qualitative Discrimination and Quantitative Determination Model Research of Methanol Gasoline and Ethanol Gasoline. *Spectros. Spectros. Anal.* 2020, 40, 1640.

(4) Zhao, L.; Wang, D. Combined Effects of a Biobutanol/Ethanol–Gasoline (E10) Blend and Exhaust Gas Recirculation on Performance and Pollutant Emissions. *ACS Omega* 2020, 5, 3250–3257.

(5) Awad, O. I.; Mamat, R.; Ali, O. M.; Sidik, N. A. C.; Yusaf, T.; Kadrigama, K.; Kettner, M. Alcohol and Ether As Alternative Fuels in Spark Ignition Engine: A Review. *Renewable Sustainable Energy Rev.* 2018, 82, 2586–2605.

(6) Zhou, K. P.; Bi, W. H.; Xing, Y. H.; Chen, J. G.; Zhou, T.; Fu, X. H. Multi Spectral Detection of Ethanol Content in Gasoline Based on SiPLS Feature Extraction and Information Fusion. *Spectros. Spectros. Anal.* 2017, 37, 429–434.

(7) Bezerra, A. C. d. M.; Silva, D. d. O.; Matos, G. H. M. d.; dos Santos, J. P.; Borges, C. N.; Silveira, L.; Pacheco, M. T. T. Quantification of Anhydrous Ethanol and Detection of Adulterants in Commercial Brazilian Gasoline by Raman Spectroscopy. *Instrum. Sci. Technol.* 2019, 47, 90–106.

(8) Legner, R.; Voigt, M.; Wirtz, A.; Friesen, A.; Haefner, S.; Jaeger, M. Using Compact Proton Nuclear Magnetic Resonance at 80 MHz and Vibrational Spectroscopies and Data Fusion for Research Octane Number and Gasoline Additive Determination. *Energy Fuels* 2020, 34, 103–110.

(9) Chen, H.; Zhang, Y.; Qi, H.; Li, D. Detection of Ethanol Content in Ethanol Diesel Based on PLS and Multispectral Method. *Optik* 2019, 195, No. 162861.

(10) Santos, F. D.; Santos, L. P.; Cunha, P. H. P.; Borgli, F. T.; Romão, W.; de Castro, E. V. R.; de Oliveira, E. C.; Filgueiras, P. R. Discrimination of Oils and Fuels Using a Portable NIR Spectrometer. *Fuel* 2021, 283, No. 118854.

(11) Wang, H.; Chu, X.; Chen, P.; Li, J.; Liu, D.; Xu, Y. Partial Least Squares Regression Residual Extreme Learning Machine (PLSRL-ELM) Calibration Algorithm Applied in Fast Determination of Gasoline Octane Number with Near-Infrared Spectroscopy. *Fuel* 2022, 309, No. 122224.

(12) Moreira, M.; Felício, A.; de França, J. Evaluation of the Influence of Sample Variability On the Calibration of A NIR Photometer for Quantification of Ethanol in Gasoline. *Meas. Sci. Technol.* 2019, 30, No. 075301.

(13) Castro, R. C.; Ribeiro, D. S. M.; Santos, J. L. M.; Pascoa, R. N. M. J. Comparison of Near Infrared Spectroscopy and Raman Spectroscopy for the Identification and Quantification Through MCR-ALS and PLS of Peanut Oil Adulterants. *Talanta* 2021, 230, No. 122373.

(14) Novikova, N. I.; Matthews, H.; Williams, I.; Sewell, M. A.; Nieuwoudt, M. K.; Simpson, M. C.; Broderick, N. G. R. Detecting Phytoplankton Cell Viability Using NIR Raman Spectroscopy and PCA. *ACS Omega* 2022, 7, 5962–5971.

(15) Carneiro, H. S. P.; Medeiros, A. R. B.; Oliveira, F. C. C.; Aguiar, G. H. M.; Rubim, J. C.; Suarez, P. A. Z. Determination of Ethanol Fuel Adulteration by Methanol Using Partial Least-Squares Models Based on Fourier Transform Techniques. *Energy Fuels* 2008, 22, 2767–2770.

(16) Yun, Y. H.; Li, H. D.; Deng, B. C.; Cao, D. S. An Overview of Variable Selection Methods in Multivariate Analysis of Near-Infrared Spectra. *TrAC, Trends Anal. Chem.* 2019, 113, 102–115.
Spectroscopic Characterization of Novel Benzofuran-Based Hydrocarbons: Structural and Spectroscopic Analysis.

J. Mol. Struct. 2020, 125, 2101–2113.

Xiaobo, Z.; Jiewen, Z.; Povey, M. J. W.; Holmes, M.; Hanpin, M. Variables Selection Methods in Near-Infrared Spectroscopy. Anal. Chim. Acta 2010, 667, 14–32.

Jiang, W.; Lu, C.; Zhang, Y.; Ju, W.; Wang, J.; Hong, F.; Wang, T.; Ou, C. Moving-Window-Improved Monte Carlo Uninformative Variable Elimination Combining Successive Projections Algorithm for Near-Infrared Spectroscopy (NIRS). J. Spectrosc. 2020, 2020, No. 3590301.

Sun, Z.; Fan, J.; Wang, J.; Wang, F.; Nie, L.; Li, L.; Dong, Q.; Li, C.; Du, R.; Quan, S.; Zang, H. Assessment of the Human Albumin Concentration Using Hemoglobin Acceptor in Acid Precipitation Process Using NIRS and Multi-Variable Selection Methods Combined with SPA. J. Mol. Struct. 2020, 1199, No. 126942.

Wang, S. H.; Zhao, Y.; Hu, R.; Zhang, Y. Y.; Han, X.-H. Analysis of Near-Infrared Spectra of Coal Using Deep Synergy Adaptive Moving Window Partial Least Square Method Based on Genetic Algorithm. Chin. J. Anal. Chem. 2019, 47, e19034–e19044.

Song, X.; Du, G.; Li, Q.; Tang, H.; Huang, Y. Rapid Spectral Analysis of Agro-Products Using an Optimal Strategy: Dynamic Backward Interval PLS–Competitive Adaptive Reweighted Sampling. Anal. Bioanal. Chem. 2020, 412, 2795–2804.

Pourdarbani, R.; Sabzi, S.; Kalantari, D.; Karimzadeh, R.; Ilbeysi, E.; Arribas, J. I. Automatic Non-Destructive Video Estimation of Maturation Levels in Fuji Apple (Malus Malus Pumila) Fruit in Orchard Based on Colour (Vis) and Spectral (NIR) data. Biopest. Eng. 2020, 195, 156–161.

Kawamura, K.; Tsujimoto, Y.; Nishigaki, T.; Andriamananjara, A.; Rabenarivo, M.; Asai, H.; Rakotoson, T.; Razafimbelo, T. Laboratory Visible and Near-Infrared Spectroscopy with Genetic Algorithm-Based Partial Least Squares Regression for Assessing the Soil Phosphorus Content of Upland and Lowland Rice Fields in Madagascar. Remote Sens. 2019, 11, 506.

Zhang, Y.; Li, M.; Zheng, L.; Qin, Q.; Lee, W. S. Spectral Features Extraction for Estimation of Soil Total Nitrogen Content Based on Modified Ant Colony Optimization Algorithm. Geoderma 2019, 333, 23–34.

Kaneko, H.; Kono, S.; Nojima, A.; Kambayashi, T. Transfer Learning and Wavelength Selection Method in NIR Spectroscopy to Predict Glucose and Lactate Concentrations in Culture Media Using VIP-Boruta. Anal. Sci. Adv. 2021, 2, 470–479.

Khan, B.; Khalid, M.; Shah, M. R.; Tahir, M. N.; Asif, H. M.; Rahnamayee Alibad, H. A.; Hussain, A. Synthetic, Spectroscopic, SC-XRD and Nonlinear Optical Analysis of Potent Hydraside Derivatives: A Comparative Experimental and DFT/TD-DFT Exploration. J. Mol. Struct. 2020, 1200, No. 127140.

Haroon, M.; Khalid, M.; Akhtar, T.; Tahir, M. N.; Khan, M. U.; Muhammad, S.; Al-Sehemi, A. G.; Hameed, S. Synthesis, Crystal Structure, Spectroscopic, Electronic and Nonlinear Optical Properties of Potent Thiadiazole Based Derivatives: Joint Experimental and Computational Insight. J. Mol. Struct. 2020, 1202, No. 127354.

Khalid, M.; Arshad, M. N.; Tahir, M. N.; Asiri, A. M.; Naseer, M. M.; Ishaq, M.; Khan, M. U.; Shafiq, Z. An Efficient Synthesis, Structural (SC-XRD) and Spectroscopic (FTIR, 1HNMR, MS Spectroscopic) Characterization of Novel Benzofuran-Based Hydrazones: An Experimental and Theoretical Studies. J. Mol. Struct. 2020, 1216, No. 128318.

Yun, Y. H.; Wang, W. T.; Tan, M. L.; Liang, Y. Z.; Li, H. D.; Cao, D. S.; Lu, H. M.; Xu, Q. S. A Strategy that Iteratively Retains Informative Variables for Selecting Optimal Variable Subset in Multivariate Calibration. Anal. Chim. Acta 2014, 807, 36–43.

Prasad, P. R.; Rama Rao, K. S.; Bhuvaneswari, K.; Praveena, N.; Srikanth, Y. V. V. Determination of Ethanol in Blend Petrol by Gas Chromatography and Fourier Transform Infrared Spectroscopy. Energy Sources, Part A 2008, 30, 1534–1539.

Tsukui, M.; Furudate, S.; Kugaya, Y.; Yabushita, S. Graphical Transition Moment Decomposition and Conceptual Density Functional Theory Approaches to Study the Fundamental and Lower-Level Overtone Absorption Intensities of Some OH Stretching Vibrations. J. Phys. Chem. A 2021, 125, 2101–2113.