

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



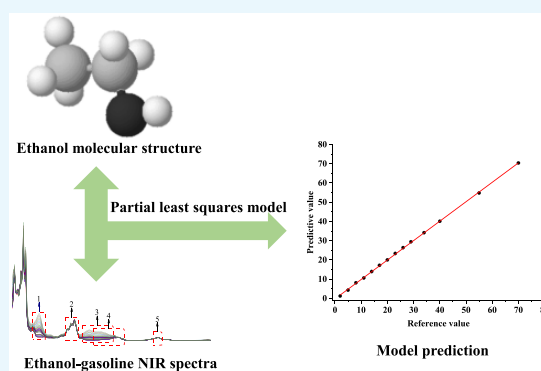
Read Online

ACCESS |

Metrics & More

Article Recommendations

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_{CV}^2) was 0.9995, the root mean square error of prediction (RMSEP) was 0.5572, and the correlation coefficient of external prediction validation (R_p^2) was 0.9995. The effective characteristic spectra model had smaller RMSEP and RMSECV values, and larger R_{CV}^2 and R_p^2 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 fuel¹ 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.^{2–4} As a low-carbon, clean, and high-quality green fuel, ethanol-gasoline has received considerable attention in recent years.⁵ 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.⁶ 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^{7,8} and Near-infrared (NIR) spectroscopy.^{9,10} NIR is a fast, non-destructive analytical method that consumes small amounts of reagents.^{11–14} 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.,² using partial least squares and principal component analysis, established a rapid analytical method to determine ethanol content in gasoline using NIR spectroscopy, achieving a low root mean square error of prediction (RMSEP).

Received: April 12, 2022

Accepted: May 16, 2022

Published: May 30, 2022



Furthermore, Carneiro et al.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.¹⁶ Appropriate wavelength selection can remove uninformative and interfering variables in the spectra to obtain better model prediction performance and improve interpretability.¹⁷ To date, several NIR wavelength selection methods have been developed. The most commonly used methods include uninformative variable elimination (UVE),¹⁸ successive projection analysis (SPA),¹⁹ moving window partial least squares (MWPLS),²⁰ interval partial least squares (iPLS),²¹ simulated annealing (SA),²² genetic algorithm (GA),²³ ant colony optimization (ACO),²⁴ and variable importance in projection (VIP).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.^{26–28} Modeling using characteristic spectral bands based on a chemical structure can improve the predictive performance and interpretability of a model.^{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.^{30–32} The absorption peaks in the region from 4661.104 to 5000.515 cm^{-1} are related to the combined

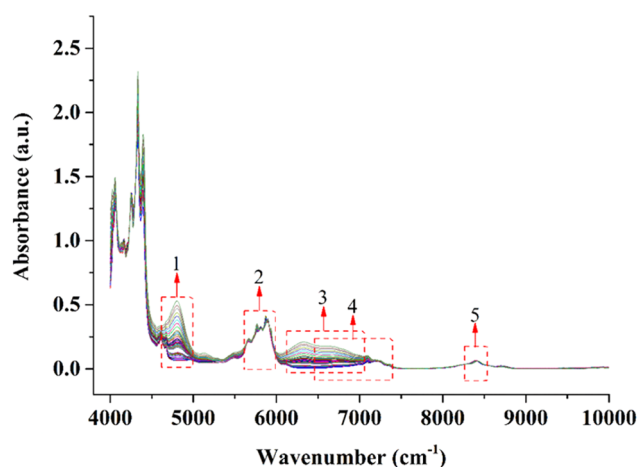


Figure 1. Raw NIR spectra of 44 ethanol-gasoline samples. The dashed rectangles 1, 2, 3, 4, and 5 represent the spectral ranges of 4661.104–5000.515 cm^{-1} , 5660.050–6001.389 cm^{-1} , 6060.171–7141.113 cm^{-1} , 6450.422–7407.241 cm^{-1} , and 8300.121–8500.682 cm^{-1} , respectively.

absorption frequency of the O–H stretching and bending vibrations.^{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;^{35,36} and the spectral range from 5660.050 to 6001.389 cm^{-1} corresponds to the first overtone of C–H stretching.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,^{15,38} while the peaks below 6060.171 cm^{-1} contain the first overtone from aromatic C–H stretching.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

Table 1. Effective Characteristic Spectral Region Modeling Scheme

modeling scheme	spectral combination (cm^{-1})	combination strategy
1	4661.104–5000.515 5660.050–7407.241 8300.121–8500.682	all the characteristic spectra region
2	4661.104–5000.515 6060.171–7141.113	characteristic spectra region of hydroxyl group
3	6060.171–7141.113	characteristic spectra region of hydroxyl group and excludes the interference spectra of aromatic groups
4	6060.171–7407.241 8300.121–8500.682	all the characteristic spectra region and excludes the interference spectra of aromatic groups

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,³⁹ 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

Table 2. Performance Comparison of Full Spectra Models using Different Preprocessing Methods

preprocess method	training set		validation set	
	RMSECV	R_{CV}^2	RMSEP	R_p^2
without	0.7682	0.9987	1.0955	0.9966
first derivative	0.6396	0.9991	0.7824	0.9982
SNV	0.9581	0.9979	0.9676	0.9973
VN	1.8401	0.9927	2.6418	0.9800
MSC	0.7597	0.9987	1.0501	0.9968
SGM	0.7677	0.9987	1.1954	0.9959

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 largest correlation coefficient of internal cross-validation (R_{CV}^2) and correlation coefficient of external prediction validation (R_p^2) values, indicating that this model has the highest prediction accuracy. Compared to the model based on raw spectra, the RMSECV value of the model established by the first derivative spectra decreased from 0.7682 to 0.6396, the RMSEP value decreased from 1.0955 to 0.7824, the R_{CV}^2 value increased from 0.9987 to 0.9991, and the R_p^2 value increased from 0.9966 to 0.9982. This indicates that the model based on the first derivative spectra has a higher prediction accuracy than the model based on the raw spectra. The first derivative was finally proven to be used for the optimal preprocessing method. The preprocessed NIR spectra of 44 ethanol-gasoline samples are shown in Figure 2.

Comparison of the Model Performance Between Effective Characteristic Spectra, Full Spectra, and VIP Spectra. After preprocessing using the optimal method, the calibration analysis models were established by full spectra and the four effective characteristic spectral schemes, and then the validation set samples were predicted. As shown in Table 3, the models established by the effective characteristic spectra have better parameters than the model established by the full spectra. However, the RMSEP values of Schemes 2 and 4 are slightly larger than those of the full spectral model, which may be due to the interference of other components in the complex gasoline samples. Scheme 3 has the best performance, the values of the model parameters, RMSECV, RMSEP, R_{CV}^2 , and R_p^2 were 0.6193, 0.5572, 0.9995, and 0.9991, respectively. The close values of RMSECV and RMSEP mean that the model established by Scheme 3 has high stability and accuracy. These

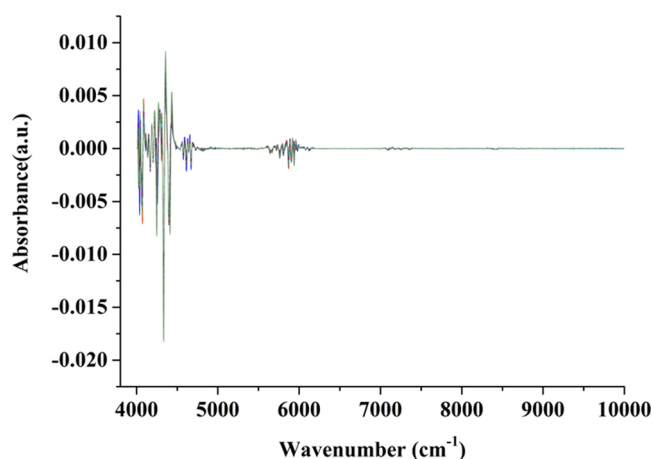


Figure 2. Preprocessed NIR spectra by SGM (13 points with a second polynomial order).

Table 3. Comparison of Different Wavelength Selection Methods for Modeling and Analysis of Ethanol Content

modeling scheme	training set		validation set	
	RMSECV	R_{CV}^2	RMSEP	R_p^2
full spectra	0.6936	0.9991	0.7824	0.9982
1	0.6012	0.9992	0.6491	0.9988
2	0.6204	0.9991	0.9011	0.9977
3	0.6193	0.9995	0.5572	0.9991
4	0.6592	0.9991	0.8958	0.9977
VIP spectra	0.6288	0.9991	0.8859	0.9977

results indicate that the characteristic spectra of Scheme 3 are less disturbed by other components. This also proves that the hydroxyl group spectra, without the interference of aromatic groups, more accurately reflect 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_{CV}^2 and R_p^2 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 0.8859 to 0.5572, the RMSEP value decreased from 0.8859 to 0.5572, while R_{CV}^2 increased from 0.9991 to 0.9995, and R_p^2 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

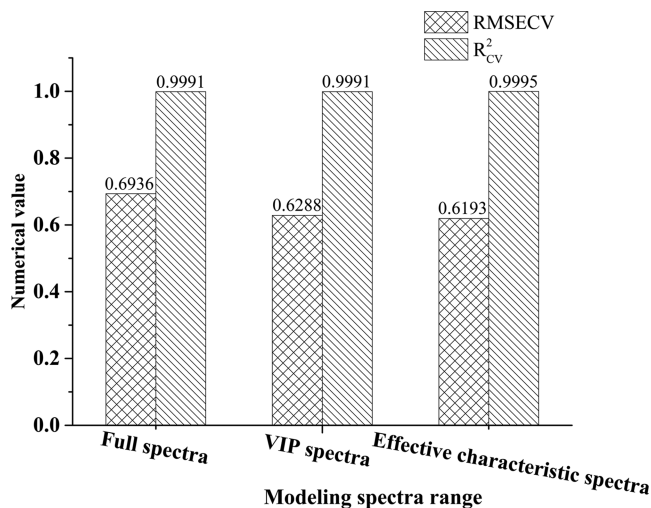


Figure 3. Internal cross-validation results of the ethanol content in gasoline using partial least squares models based on different spectral ranges.

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 $R_p^2 = 0.9991$, indicating that the predicted value is consistent with the reference value. In conclusion, modeling based on the

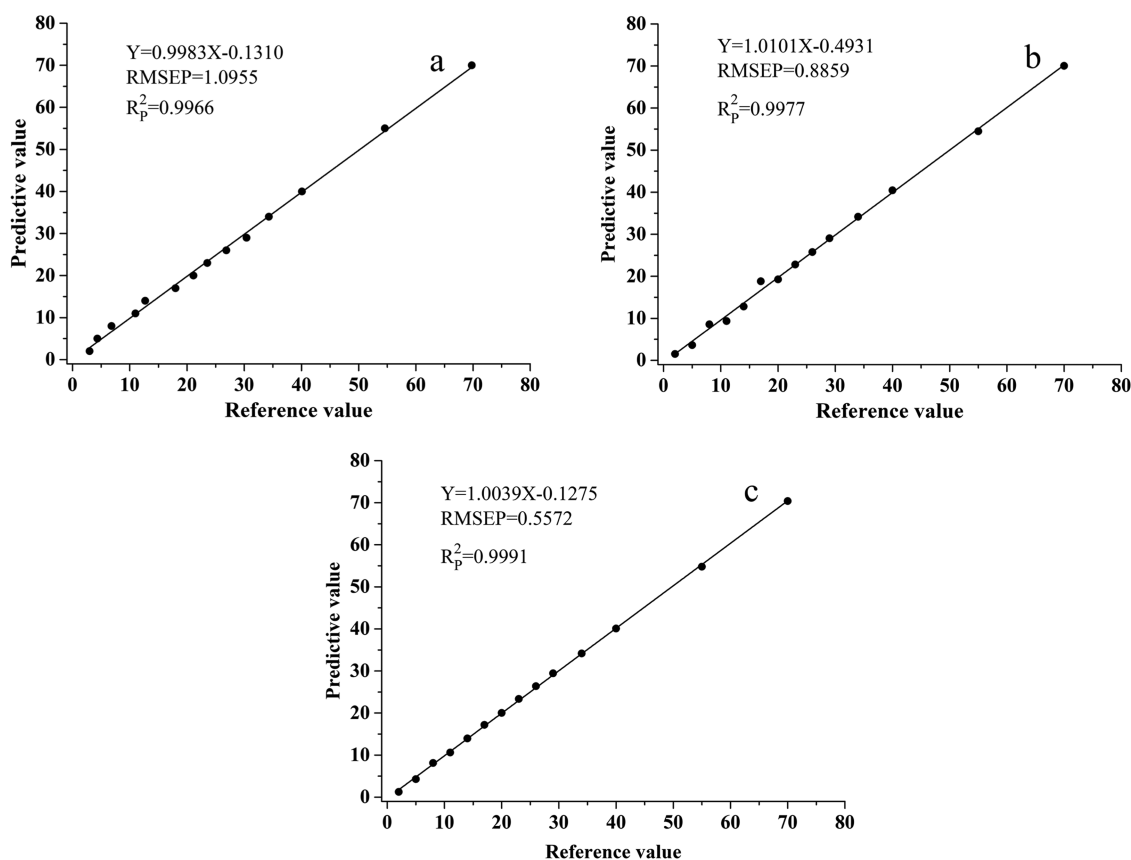


Figure 4. External validation results of the ethanol content in gasoline using partial least squares models based on different spectral ranges. (a) Full spectra; (b) VIP characteristic spectra; (c) effective characteristic spectra.

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 $R_{\text{CV}}^2 = 0.9995$, $\text{RMSECV} = 0.6193$, $R_{\text{p}}^2 = 0.9991$, and $\text{RMSEP} = 0.5572$.

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^{40,41} method was used to build the calibration model. Finally, an optimal preprocessing method was selected.

R_{CV}^2 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_{CV}^2 value was closer to 1 and the RMSECV value was low. R_{p}^2 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_{p}^2 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_{CV}^2 , RMSEP, and R_{p}^2 to compare the effects of the full spectra, VIP spectra, and effective characteristic spectra models. RMSECV, RMSEP, and R^2 (R_{CV}^2 or R_{p}^2) were calculated using the following equations

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

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

Table 4. Ethanol Volume Fraction in Ethanol-gasoline Samples

no.	volume fraction (%)	no.	volume fraction (%)	no.	volume fraction (%)	no.	volume fraction (%)
1	0.5	12*	11	23	22	34	36
2	1	13	12	24*	23	35	38
3*	2	14	13	25	24	36*	40
4	3	15*	14	26	25	37	45
5	4	16	15	27*	26	38	50
6*	5	17	16	28	27	39*	55
7	6	18*	17	29	28	40	60
8	7	19	18	30*	29	41	65
9*	8	20	19	31	30	42*	70
10	9	21*	20	32	32	43	75
11	10	22	21	33*	34	44	80

Note: Samples marked with * are included in the validation set.

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

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. AKYZZ2131), and the Institute-Enterprise Horizontal Scientific Research Program (No. JSFW2102).

NOMENCLATURE

NIR near-infrared
 MIR mid-infrared
 RMSECV root mean square error of cross-validation
 RMSEP root mean square error of prediction
 R_{CV}^2 correlation coefficient of internal cross-validation
 R_p^2 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 Nabhani, 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. *Spectrosc. Spectrosc. 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. *Spectrosc. Spectral Anal.* **2017**, *37*, 429–434.
- (7) Bezerra, A. C. d. M.; Silva, D. dO.; 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.; Borghi, 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 (PLSRR-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. 075501.
- (13) Castro, R. C.; Ribeiro, D. S. M.; Santos, J. L. M.; Páscoa, 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.

- (17) 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.
- (18) 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.
- (19) 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 in Acid Precipitation Process Using NIRS and Multi-Variable Selection Methods Combined with SPA. *J. Mol. Struct.* **2020**, *1199*, No. 126942.
- (20) 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.
- (21) Song, X.; Du, G.; Li, Q.; Tang, G.; 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.
- (22) Pourdarbani, R.; Sabzi, S.; Kalantari, D.; Karimzadeh, R.; Ilbeygi, 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. *Biosyst. Eng.* **2020**, *195*, 136–151.
- (23) 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.
- (24) 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.
- (25) 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.
- (26) Khan, B.; Khalid, M.; Shah, M. R.; Tahir, M. N.; Asif, H. M.; Rahnamaye Aliabad, H. A.; Hussain, A. Synthetic, Spectroscopic, SC-XRD and Nonlinear Optical Analysis of Potent Hydrazone Derivatives: A Comparative Experimental and DFT/TD-DFT Exploration. *J. Mol. Struct.* **2020**, *1200*, No. 127140.
- (27) 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 Thiazole Based Derivatives: Joint Experimental and Computational Insight. *J. Mol. Struct.* **2020**, *1202*, No. 127354.
- (28) 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.
- (29) 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.
- (30) Prasad, P. R.; Rama Rao, K. S.; Bhuvaneshwari, 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.
- (31) Tsuyuki, 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.
- (32) Li, M.; Xue, J.; Du, Y.; Zhang, T.; Li, H. Data Fusion of Raman and Near-Infrared Spectroscopies for the Rapid Quantitative Analysis of Methanol Content in Methanol–Gasoline. *Energy Fuels* **2019**, *33*, 12286–12294.
- (33) Aiguo, O.; Tianyi, T.; Haiyang, W.; Yan-de, L. Detection of Key Performance Indicators of Ethanol Diesel by the Infrared Spectroscopy Method. *Chin. Opt.* **2017**, *10*, 363–369.
- (34) Wang, X.; Bao, Y.; Liu, G.; Li, G.; Lin, L. Study On the Best Analysis Spectral Section of NIR to Detect Alcohol Concentration Based on SiPLS. *Procedia Eng.* **2012**, *29*, 2285–2290.
- (35) Fernandes, H. L.; Raimundo, I. M., Jr; Pasquini, C.; Rohwedder, J. J. Simultaneous Determination of Methanol and Ethanol in Gasoline Using NIR Spectroscopy: Effect of Gasoline Composition. *Talanta* **2008**, *75*, 804–810.
- (36) Cai, J.; Chen, Q.; Wan, X.; Zhao, J. Determination of Total Volatile Basic Nitrogen (TVB-N) Content and Warner–Brazler Shear Force (WBSF) in Pork Using Fourier Transform Near Infrared (FT-NIR) Spectroscopy. *Food Chem.* **2011**, *126*, 1354–1360.
- (37) Beć, K. B.; Grabska, J.; Huck, C. W.; Czarniecki, M. A. Effect of Conformational Isomerism on NIR Spectra of Ethanol Isotopologues. Spectroscopic and Anharmonic DFT study. *J. Mol. Liq.* **2020**, *310*, No. 113271.
- (38) Czarniecki, M. A.; Morisawa, Y.; Futami, Y.; Ozaki, Y. Advances in Molecular Structure and Interaction Studies Using Near-Infrared Spectroscopy. *Chem. Rev.* **2015**, *115*, 9707–9744.
- (39) Diwu, P. Y.; Bian, X. H.; Wang, Z. F.; Wei, L. Study On the Selection of Spectral Preprocessing Methods. *Spectrosc. Spectral Anal.* **2019**, *39*, 2800–2806.
- (40) Geladi, P.; Kowalski, B. R. Partial Least-Squares Regression: a Tutorial. *Anal. Chim. Acta* **1986**, *185*, 1–17.
- (41) Kaneko, H. Genetic Algorithm-Based Partial Least-Squares with Only the First Component for Model Interpretation. *ACS Omega* **2022**, *7*, 8968–8979.