ACS OMEGA

Open Access

http://pubs.acs.org/journal/acsodf

Article

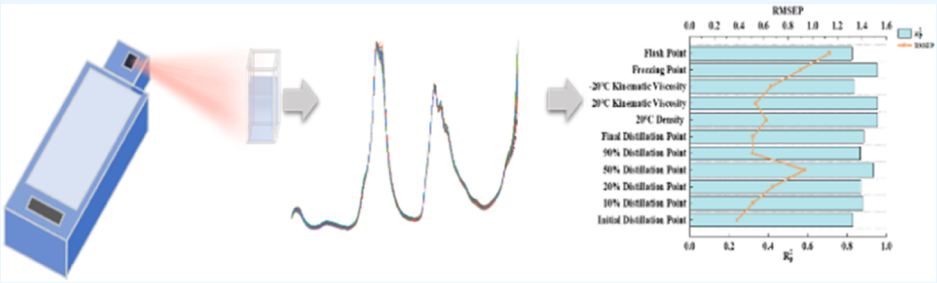
Simultaneous Rapid Detection of Multiple Physicochemical Properties of Jet Fuel Using Near-Infrared Spectroscopy

Ke Li, Xin Zhang, Jing Zhang, Biao Du, Xiaoping Song, Guixuan Wang, Qi Li, Yinglan Zhang, Fan Liu, and Zhengdong Zhang*

Cite This: *ACS Omega* 2024, 9, 16138−16146

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

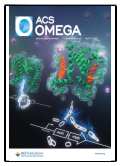


ABSTRACT: Jet fuel is the primary fuel used in the aviation industry, and its quality has a direct impact on the safety and operational efficiency of aircraft. The accurate quantitative detection and analysis of various physicochemical property indicators are important for improving and ensuring the quality of jet fuel in the domestic market. This study used near-infrared (NIR) spectroscopy to establish a suitable model for the simultaneous and rapid detection of multiple physicochemical properties in jet fuel. Using more than 40 different sources of jet fuel, a rapid detection model was established by optimizing the spectral processing methods. The measurement models were separately built using the partial least-squares (PLS) and orthogonal PLS algorithms, and the model parameters were optimized. The results show that after the Savitzky−Golay second derivative preprocessing, the PLS model built using the feature spectra selected by the uninformative variable elimination wavelength algorithm achieved the best measurement performance. Compared with the PLS model without preprocessing, the range of the resulting accuracy improvement was at least 15.01%. Under the optimal model parameters, the calibration set regression coefficient (R_c^2) of the 11 jet fuel property index models ranged from 0.9102 to 0.9763, with the root-mean-square error of calibration values up to 0.8468 °C (for flash points). The regression coefficient (R_p^2) of the validation set ranged from 0.8239 to 0.9557, with the root-mean-square error of prediction values up to 1.1354 °C (for flash points). The ratios of prediction to deviation (RPD) values were all in the range of 1.9−3.0, indicating high accuracy and reliability of the model. The rapid NIR analysis method established in this study enables the simultaneous and rapid detection of multiple physicochemical properties of jet fuel, thereby providing effective technical support for ensuring the quality of jet fuel in the market.

1. INTRODUCTION

Jet fuel is an indispensable resource in the aviation industry, and its quality directly affects jet engine performance. However, factors such as the high affinity between jet fuel and water can make the quality of jet fuel susceptible to environmental temperature and humidity fluctuations, leading to changes in its physical and chemical properties, which in turn affect the normal operation of the engine.[1] To ensure the quality of jet fuel in the market, the government has established a series of standards that strictly regulate over 40 physicochemical properties and their corresponding detection methods for jet fuel. Also, researchers have developed several conventional analytical methods, such as gas chromatography and liquid phase chromatography, to analyze jet fuel. These methods can achieve accurate detection of the physical and chemical properties of jet fuel, but they are hindered by a lengthy detection period, complex methodology, the need for the pretreatment of samples, and the lack of portability for the testing equipment, which fail to fulfill the requirements of quality inspection agencies and the petrochemical industry for the rapid on-site analysis of jet fuel. Some rapid detection methods were also developed, such as Raman and mid-infrared spectroscopy. However, the fluorescence signal interference of the Raman spectrum affects the accurate detection of jet fuel, and the mid-infrared spectrum is more suitable for the

Received: December 14, 2023
Revised: March 8, 2024
Accepted: March 18, 2024
Published: March 28, 2024



https://doi.org/10.1021/acsomega.3c09994
ACS Omega 2024, 9, 16138−16146

detection of physicochemical properties of jet fuel with definite characteristic absorption. Therefore, there is a need for a simple and reliable on-site analysis method to quickly determine the physicochemical properties of jet fuel and to ascertain its quality compliance.

Near-infrared (NIR) spectroscopic analysis can reflect the overtone and combination frequency absorption information of hydrogen-containing functional groups in compounds.[2] By combining chemometric algorithms, a rapid detection model can be established between the spectra and target properties, thus enabling the qualitative or quantitative analysis of samples.[2−4] NIR spectroscopy is highly efficient, rapid, and nondestructive and can simultaneously determine multiple components. It has been widely applied in various industries, such as chemical engineering, agriculture, and medicine.[5−11] In recent years, researchers have applied NIR spectroscopy to the characterization and quality detection of jet fuels. By utilizing NIR spectroscopy in conjunction with chemometrics, multivariate regression models have been constructed to rapidly detect the physicochemical properties of jet fuels.[12−14] Xiao et al.[15] developed a rapid analysis model for the total acid value of jet fuel based on NIR spectroscopy using a backpropagation neural network (BP-ANN) and partial least-squares (PLS) regression methods. Their results showed that the determination coefficient (R_p^2) of the BP-ANN model was 0.9778, and the root-mean-square error of prediction (RMSEP) was 0.0007 mg/g. Compared with the PLS model, the R_p^2 value increased by 3.9% and the RMSEP value decreased by 17.5%. These results indicate that the BP-ANN model method has a higher determination accuracy and can achieve quantitative analysis of the total acidity of jet fuels. The interval partial least-squares (iPLS) algorithm can simultaneously consider the correlation between multiple wavelengths and determine the advantages of wavelength combinations with the highest predictive ability. Yangjun et al.[16] developed a rapid analysis model for the initial boiling point of jet fuel based on NIR spectroscopy using the iPLS and BP-ANN algorithms. The model achieved an R_p^2 value of 0.9361 and validation standard error (SEV) of 1.0726 °C. Compared to the model built using full-wavelength spectroscopy, the R_p^2 value of this model improved by 36.23% and the SEV decreased by 54.85%. Therefore, the combination of iPLS and BP-ANN models can achieve an accurate analysis of the initial distillation point of jet fuel. Currently, most of the reported jet fuel property index determination models are only suitable for the single property index, with limited reports on models that can simultaneously determine multiple indices. The simultaneous detection of multiple physicochemical properties of jet fuel can simplify the on-site testing process, reduce analysis time to improve efficiency, and provide jet fuel quality data quickly, conveniently, and efficiently. This method can effectively manage the production and use of jet fuel and prevent accidents caused by quality issues. Therefore, establishing a rapid analysis method that is highly accurate and applicable to the simultaneous determination of multiple key property indices of jet fuel is crucial for ensuring the quality of jet fuel in the market.

This study presents a rapid analysis method using NIR spectroscopy for the simultaneous detection of multiple physicochemical properties of jet fuels. Various spectral preprocessing methods, including Savitzky−Golay smoothing (SG smoothing), Savitzky−Golay first derivative (SG first D), Savitzky−Golay second derivative (SG second D), and multiple scatter correction (MSC), were applied to the spectral data. In addition, feature wavelength selection methods, including variable importance projection (VIP), uninformative variable elimination (UVE), and Monte Carlo UVE (MC-UVE), were used to select the feature variables. Separate PLS and orthogonal PLS (O-PLS) models were developed, and their predictive performances were compared. The model conditions were optimized, and the optimal model was used for the sample analysis. The aim of this study was to prove that the optimized model can be used to reduce the potential risk of accidents caused by fuel quality issues.

2. RESULTS AND DISCUSSION

2.1. Spectrum Data Preprocessing Method Selection.

In addition to information about the samples, the collected spectra contained information about environmental influences and instrument stability. The partial effects of these factors were visually reflected in the raw spectral plot of the jet fuel (Figure 1). For example, within the wavelength range of 900−

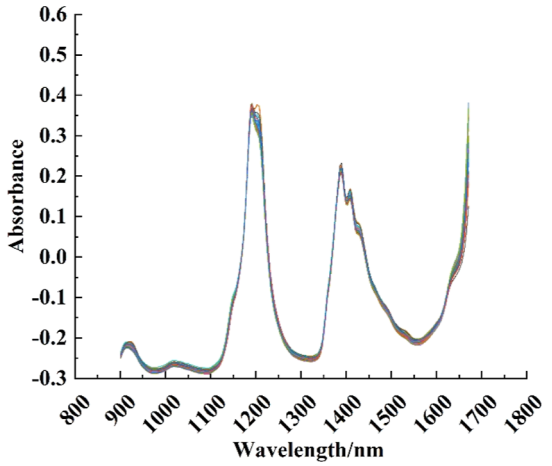


Figure 1. NIR spectra of the no. 3 jet fuel samples.

1700 nm, a noticeable difference was observed in the baselines of the spectra collected from different samples, which may be due to variations in instrument performance or the effects of different characteristics of samples.

To eliminate the influence of interference contained in the raw spectra and establish a stable and reliable measurement model, the preprocessing of raw spectra was performed. In this experiment, four methods (SG smoothing, SG first D, SG second D, and MSC) were employed for preprocessing the spectral data. SG smoothing was applied to remove fluctuations caused by noise in the spectra and produced a more compact NIR spectrum (Figure 2a).[17,18] Other than the primary absorption peaks, the fluctuations of the miscellaneous peaks at 1150 and 1500 nm became smoother, indicating that SG smoothing effectively reduced the spectral fluctuations caused by environmental factors that affected the instrument function. SG first D and SG second D are baseline correction methods that aim to eliminate the spectral baseline shift caused by instrument and sample background factors,[19] while enhancing the distinguishable differences between samples. The NIR spectra of the jet fuel after treatment using the SG first D method are shown in Figure 2b. The trend of spectral response changes occurred primarily in the range from 1100 to 1500 nm (the secondary octave region of methyl groups is

https://doi.org/10.1021/acsomega.3c09994
ACS Omega 2024, 9, 16138−16146

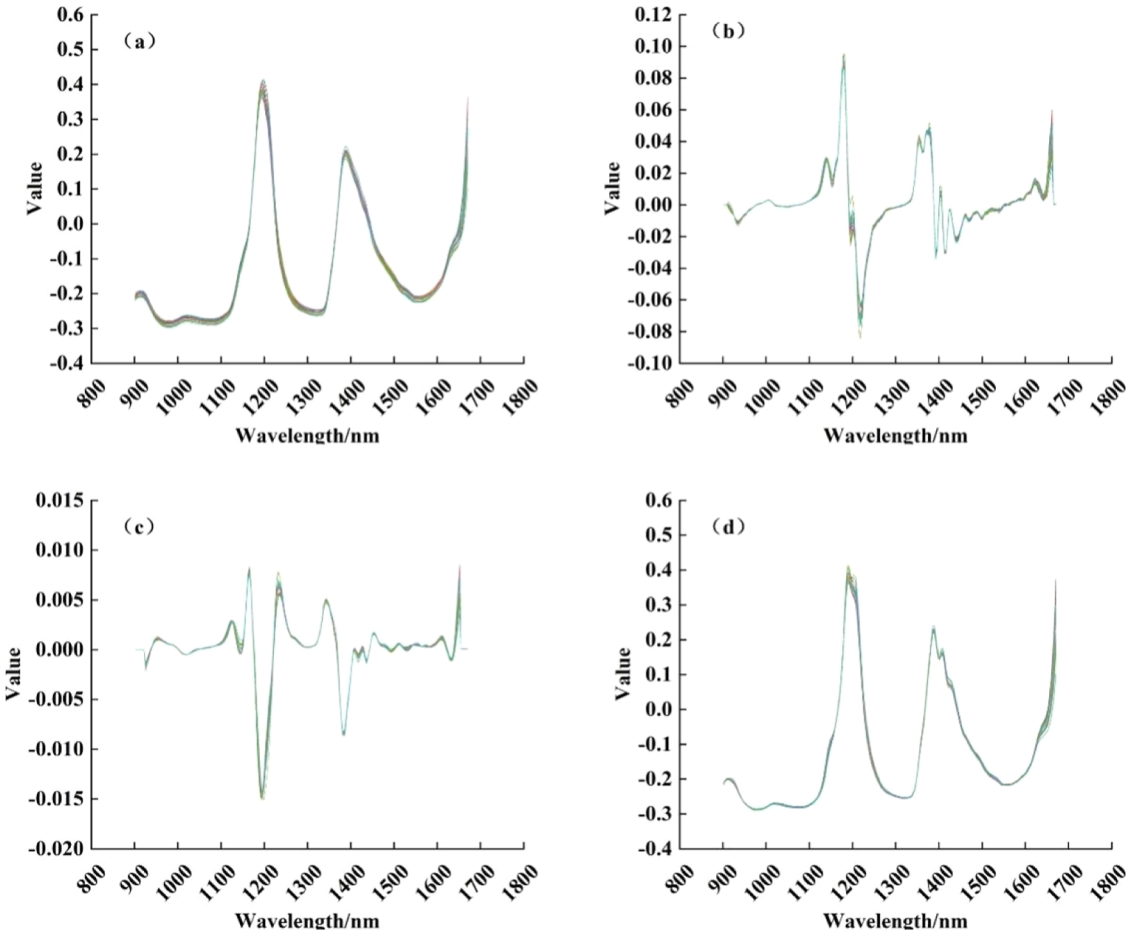


Figure 2. NIR spectra following the pretreatment of 40 no. 3 jet fuel samples: (a) SG smooth; (b) SG first D; (c) SG second D; and (d) MSC.

predominant at 1100 nm, and the second combining region of C−H vibrations is predominant at 1500 nm),[20] and the processed spectrum contained more spectral information than the unprocessed spectrum. The spectrum (Figure 2c) after processing with the SG second D method showed no clear distinction in the overall baseline shift, indicating that the offset effects caused by environmental factors, instrument stability, or sample physical properties related for different scattering were effectively eliminated. The difference in the spectral line height between 1100 (C−H secondary octave absorption of cyclohexane) and 1250 nm (combination of the O−H stretching vibration and C−OH bending) reflects different response trends of the different samples at this wavelength.[20] The SG second D method visually represented the subtle numerical differences between the samples. MSC reduces the impact of multiple scattering and diffuse reflection components on the spectral quality by analyzing and simulating different scattering mechanisms, thereby improving the measurement accuracy of the model.[21,22] Following MSC processing, the original spectrum (Figure 2d) showed a noticeable correction for the inherent offset between the sample spectrum.

To validate the improvement effects of different preprocessing methods on the measurement performance of the model and select the optimal preprocessing method, the PLS and O-PLS algorithms were used separately to build measurement models by using full-wavelength raw and preprocessed spectra. The evaluation of each model was performed and is shown in Supporting Information Tables S1−S10. The models for the 11 physicochemical properties that were established based on the full-wavelength raw spectra had an average ratio of prediction to deviation (RPD) value of 1.3. However, the models that were established based on the preprocessed spectra had average RPD values greater than 1.3, with the highest value attaining 1.7. The models that were established based on the preprocessed spectra showed an improvement in the RPD values, ranging from 2.56 to 28.34%. Therefore, compared to the models that were based on the full-wavelength raw spectra, the models that were based on the preprocessed spectra demonstrated higher accuracy and reliability in the overall determination of the 11 physicochemical properties. The model built from the full-wavelength raw spectra after SG second D preprocessing (Supporting Information Table S4) had the highest values of R_c^2 and R_p^2 and the lowest values of the root-mean-square error of calibration (RMSEC) and RMSEP compared with the models built from spectra using SG smoothing, SG first D, and MSC preprocessing methods. These results indicate that the application of SG second D preprocessing to full-wavelength raw spectra can effectively reduce the influence of environmental factors, such as temperature and humidity, and improve the accuracy and reliability of the model. Therefore, SG second D was selected as the preprocessing method for NIR spectroscopy of the jet fuel.

2.2. Optimization of the Feature Band Selection Method. The full-wavelength spectrum contains both sample

https://doi.org/10.1021/acsomega.3c09994
ACS Omega 2024, 9, 16138−16146

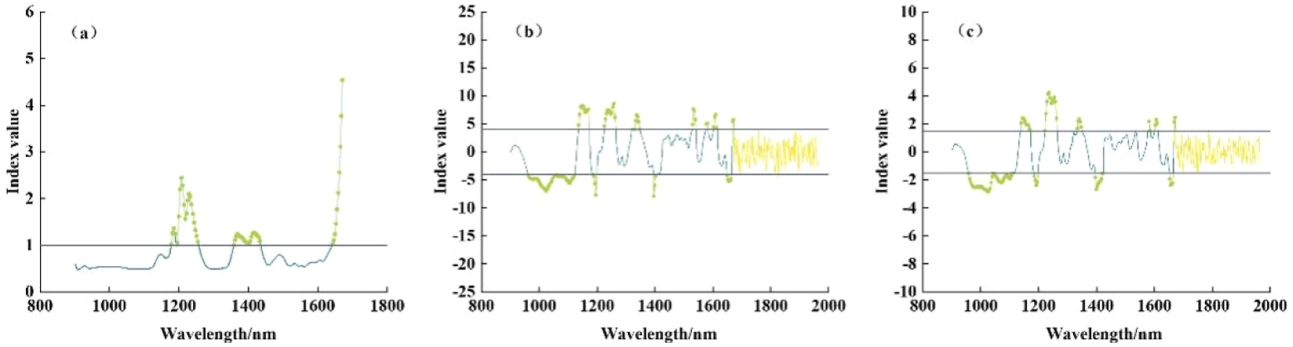


Figure 3. Feature selection images: (a) VIP, (b) UVE, and (c) MC-UVE.

and redundant information, which can complicate the process of building a measurement model and reduce its accuracy; therefore, to overcome these drawbacks, characteristic wavelength bands should be selected.[23,24] In this study, the three commonly used wavelength selection methods VIP, UVE, and MC-UVE were used to select feature bands from the full-wavelength NIR spectrum of jet fuel. In addition, the performances of the models built using the feature spectra selected by these three methods were compared.

VIP is a method that calculates the correlation between each wavelength and the response variable for variable selection and represents the explanatory power of each wavelength on the response variable. A VIP value of one is the threshold; thus, if the indicator value of a particular wavelength range is greater than one, this wavelength range significantly contributes to the modeling and determination results. Therefore, wavelength ranges with VIP values greater than one are selected as feature spectra.[25,26] Figure 3a shows that the selected spectral bands are primarily concentrated in the ranges of 1170−1250 nm (dominated by the secondary multiplication of C−H vibrations), 1380−1440 nm (dominated by the second combined frequency region of C−H vibrations), and 1640−1670 nm (corresponding to the first multiplication of C−H vibrations and the first multiplication of O−H vibrations).[20] This selected subset of spectral bands contains crucial information about the samples and can be used as a substitute for the full-wavelength spectral data to establish a quantitative model for jet fuel determination.

The UVE method calculates the coefficient of variation for each wavelength in the spectral matrix, resulting in a noise matrix with the same dimensions as the original spectral matrix. This noise matrix is then added to the spectral matrix. The coefficient of variation of the spectral and noise matrix was compared and it was determined whether each wavelength can be used for model establishment.[27,28] Based on the extremum values of the noise matrix, a threshold range of 4.8 was set, and the selected spectral bands had absolute coefficients of variation within the range of 5−10, indicating a pronounced response. The primary feature response occurred within the range of 900−1200 nm (corresponding mainly to the secondary frequency doubling region of C−H vibrations with part of the tertiary frequency doubling region of C−H vibrations), with small concentrations at 1350−1400, 1500, and 1600 nm (1350−1400 nm is included in the combined frequency region of C−H vibrations and 1500 nm is the second) (Figure 3b).[20]

The MC-UVE method is an extension of the UVE method that incorporates a Monte Carlo simulation to randomly resample the data, perform feature selection, and train the model. This method was used to assess the importance of the feature bands selected by UVE on different data sets, thereby avoiding errors in the model measurement caused by the randomness of the validation data set. This helps to evaluate the responsiveness of different feature wavelengths to different sample sets.[29] Based on the range of the simulated noise response values, a threshold was set to perform feature selection on the full-wavelength raw spectral data (Figure 3c). The trends of the change and distribution of the feature bands are similar to those shown in Figure 3b. The selected feature bands were concentrated from 900 to 1200 (mainly secondary multiplication of C−H vibrations), 1230 (secondary multiplication of C−H vibrations), and 1380 nm Figure 2c.[20] Compared with the results obtained after UVE screening of the full-wavelength raw spectra, the range from 1500 to 1600 nm was not selected. This phenomenon suggests that during the feature selection process, MC-UVE considered the feature response in the 1500−1600 nm range to be unsuitable for all samples and had a low contribution rate.

After feature bands were selected using different methods, measurement models for 11 physicochemical property indices were established and evaluated using the PLS and O-PLS algorithms (Supporting Information Tables S11−S16). Compared with the measurement models that were established using full-wavelength spectra, the average fluctuations of R_c^2 and R_p^2 for the measurement models that were established using characteristic wavelength bands that ranged from 0.83 to 2.04% and 0.43 to 14.65%, respectively. The RMSEC and RMSEP values decreased by 2.31−26.43% and 12.61−32.54%, respectively. The PLS model that was generated by the spectrum processed by using the UVE algorithm showed the most significant performance improvement (Supporting Information Table S12). The mean R_p^2 and RPD values of the model increased from 0.7940 and 1.3 to 0.8193 and 1.6, respectively, compared to those of the full-wavelength spectrum model. The improvements in R_p^2 and RPD values for the 11 physicochemical property indicator models ranged from 0.11 to 9.87% and from 1.51 to 87.91%, respectively. These results indicate that using the UVE algorithm for feature band selection in full-wavelength spectra can effectively improve the decrease in model measurement accuracy caused by the introduction of redundant information, thereby enhancing the accuracy and reliability of the model. Therefore, the UVE algorithm was chosen as the feature band selection method and used in subsequent studies to explore the optimal modeling method for determining multiple physicochemical properties.

https://doi.org/10.1021/acsomega.3c09994
ACS Omega 2024, 9, 16138−16146

2.3. Modeling Method Comparison. Choosing an appropriate modeling method can adapt to data characteristics, reduce modeling complexity, and improve model measurement accuracy. To investigate the optimal modeling approach to meet the requirements of the rapid and simultaneous detection of multiple physicochemical properties of the jet fuel, this study used the PLS and O-PLS methods for modeling and model evaluation. PLS is a multivariate statistical analysis method that considers the correlations between independent and dependent variables to establish a linear relationship model. O-PLS, based on PLS, separates the data into predictive (T) and orthogonal (TO) components and selects the predictive component to build the model.[30,31]

After processing the full spectrum using the SG second D and UVE algorithms, modeling was performed using the PLS and O-PLS methods. The modeling results are presented in Supporting Information Tables S17 and S18. After predictive (T) and nonpredictive (TO) decomposition of the spectral data, the performance of the O-PLS model decreased compared with when it was not decomposed. This phenomenon can be attributed to the combined effect of feature band selection and the orthogonal transformation of the O-PLS, thus resulting in insufficient dimensions of the input data for the model, which decreases the accuracy of the measurement. After processing the spectral data using the SG second D and UVE methods, the number of wavelength points was reduced by approximately 80%. The models built using PLS had a maximum RMSEP value of 1.1354 °C (for flash points) and an average RPD value of 2.3. In contrast, the O-PLS models had a maximum RMSEP value of 3.9566 °C (for the final distillation point) and an average RPD value of 1.6. Compared to the models built with O-PLS, the models built with PLS showed a 70.98% decrease in the average RMSEP and 41.59% increase in the average RPD. The accuracy and reliability of the PLS model are higher than those of the O-PLS model. Therefore, the PLS method was selected to establish a rapid analysis model for several physicochemical properties of jet fuel. In Supporting Information Table S17, the R_p^2 of the SG second D-UVE-PLS model ranged from 0.8329 to 0.9557 with a maximum RMSEP value of 1.1354 °C (for flash points). Compared to the PLS models obtained using other preprocessing and feature selection methods (Supporting Information Tables S19−S26), the SG second D-UVE-PLS model improved the R_p^2 values by 0.9−85.41% and reduced the RMSEP values by 25.31−92.60% for the determination of various physicochemical properties of the jet fuel. To determine the optimal modeling scheme suitable for multiple indicators, the parameters of the SG second D-UVE-PLS model were optimized.

2.4. Parameter Optimization and Evaluation of the Measurement Model. To further improve the measurement performance of the model, the latent variable factors of the SG second D-UVE-PLS model were optimized. The evaluation criterion for the number of latent variable factors was based on the average root-mean-square error of cross-validation (RMSECV) of the 11 trait-index models. Figure 4 shows the curve of the model measurement performance as a function of the number of latent variable factors. When the number of latent variable factors was six, the curve of the average RMSECV values flattened, indicating that increasing the number of latent variable factors did not significantly improve the measurement performance of the model. Six latent variable factors were used to establish the SG second D-UVE-PLS

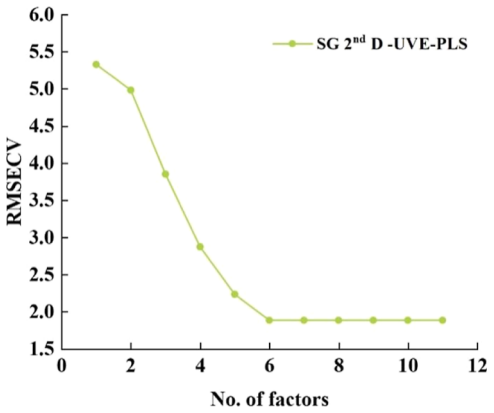


Figure 4. Mean of RMSECV as a function of the number of factors.

prediction model (Table 1), which demonstrated low RMSEP values for various physicochemical properties of the jet fuel. Moreover, most of the property indices had RPD values greater than two, indicating high accuracy and reliability of the model. Therefore, under the condition of six latent variable factors, the SG second D-UVE-PLS model accurately analyzed 11 physicochemical property indices of the jet fuel.

2.5. Practical Applications. The effectiveness of applying the SG second D-UVE-PLS model was investigated by using four real samples. The model measurement values of the 11 indicators for the four samples and reference values obtained using the national standard methods (specified in "GB 6537-2018 no. 3 jet fuel") are listed in Table 2. The reproducibility of each physicochemical property indicator was based on standard methods. The measured values were obtained using the SG second D-UVE-PLS model, and the difference between the reference and measured values represented the error. The accuracy of the model measurements was evaluated based on the magnitude of the error. The errors in the determination results for the four unknown samples were within the reproducibility range required by the national standard method, indicating that the SG second D-UVE-PLS model accurately determined the 11 jet fuel indicators. Therefore, it can satisfy the requirements for the rapid on-site detection of the no. 3 jet fuel.

3. CONCLUSIONS

In this study, we developed a rapid analytical method for the simultaneous detection of multiple physicochemical properties of jet fuels using NIR spectroscopy. This method reduces field-testing time, improves field-testing efficiency, and provides a simultaneous, rapid, and accurate determination of 11 key physicochemical indicators of the no. 3 jet fuel. This study used and compared the effects of four preprocessing methods and three feature selection methods for spectral data processing. In addition, the PLS and O-PLS algorithms were compared, and the best modeling method was selected by comparing and optimizing the performance indicators of each model. Among them, the SG second D-UVE-PLS model showed the best performance, with R_c^2 ranging from 0.9102 to 0.9763, and RMSEC values of 0.3846−0.9364 °C (distillation values), 0.6227 kg/m^3 (density), 0.5353−0.6603 mm^2/s (kinematic viscosity), 0.8978 °C (freezing point), and 1.1354 °C (flash point), for the 11 physicochemical property indicators of the calibration set. For the validation set, R_p^2 ranged from 0.8239 to 0.9557, with RMSEP values of 0.3846−

https://doi.org/10.1021/acsomega.3c09994
ACS Omega 2024, 9, 16138−16146

Table 1. SG 2nd D-UVE-PLS Optimization Modeling Results of 11 Indicators

projects	R_c^2	RMSEC	R_p^2	RMSEP	RMSECV	RPD
initial distillation point (°C)	0.9216	0.3569	0.8288	0.3846	0.3701	2.0
10% distillation point (°C)	0.9323	0.2618	0.8801	0.5114	0.3210	1.8
20% distillation point (°C)	0.9102	0.4260	0.8696	0.6741	0.4622	2.5
50% distillation point (°C)	0.9724	0.5498	0.9347	0.9364	0.6518	2.2
90% distillation point (°C)	0.9225	0.4056	0.8662	0.5096	0.4219	2.2
final distillation point (°C)	0.9618	0.2591	0.8844	0.5144	0.3215	2.2
20 °C density (kg/m^3)	0.9724	0.1815	0.9557	0.6227	0.3305	3.0
20 °C kinematic viscosity (mm^2/s)	0.9763	0.0156	0.9523	0.5353	0.1208	2.7
−20 °C kinematic viscosity (mm^2/s)	0.9254	0.1275	0.8349	0.6603	0.3651	2.5
freezing point (°C)	0.9681	0.2615	0.9521	0.8978	0.3012	2.2
flash point (°C)	0.9533	0.8468	0.8239	1.1354	0.9567	1.9

0.9364 °C (distillation values), 0.6227 kg/m^3 (density), 0.5353−0.6603 mm^2/s (kinematic viscosity), 0.8978 °C (freezing point), and 1.1354 °C (flash point) and the RPD values ranged from 1.8 to 3.0. Compared with the PLS model that was established using full-wavelength spectra, the R_c^2 and R_p^2 values improved by 2.10−82.17 and 2.40−91.65%, respectively. The RMSEC and RMSEP values decreased by 32.30−99.14 and 15.01−88.64%, respectively. Therefore, the SG second D-UVE-PLS model method can reduce the influence of external factors on the spectral quality and effectively improve the measurement accuracy and reliability of the no. 3 jet fuel.

This study successfully established a rapid NIR analysis method for the simultaneous determination of 11 physicochemical properties of jet fuels. The differences between the measured and reference values for each physicochemical property indicator met the reproducibility requirements of the national standard methods, demonstrating the high accuracy and reliability of the measurements. This method provides efficient and accurate technical support for rapid on-site analysis in the production and use of jet fuels, which is essential for research into simultaneous rapid analysis models for multiple physicochemical properties of fuels.

4. EXPERIMENTAL SECTION

4.1. Materials. We used 2 L capacity metal buckets specifically designed for oil sampling to collect two buckets each of 40 samples of the no. 3 jet fuel over a period of one month. These samples originated from 31 representative refineries in China, including PetroChina Daqing Petrochemical, SINOPEC Qilu Petrochemical, and SINOPEC Jingmen Petrochemical, with 9 sets of samples being from different production batches of the same refinery. Of the 40 samples collected, one bucket was designated for sample preservation and sealing, while the other bucket was used for experimental purposes. After all samples were numbered, they were placed in the sample reserve room, which is situated far from suburban areas, in a dry, cool location that consistently maintains a temperature of 15−25 °C throughout the year. The ratio of the calibration set to the validation set is 3:1. After measuring all samples with standard methods, we selected 10 groups of samples, whose measurement indicators were evenly distributed among all samples' measurement results, to act as the validation set from the 40 groups of samples. The remaining 30 groups of samples were designated as the calibration set.

4.2. NIR Spectroscopy Acquisition. Spectral acquisition was performed using a portable NIM-260R NIR spectrometer manufactured by the National Institute of Metrology, China. This instrument used a halogen light source (L1024, International Light Technologies, USA) with fiber-optic add-ons and a reflectance detector with a 1 cm optical path. The spectrometer was preheated for 30 min and set to a wavelength resolution of 12 nm with a wavelength range of 800−1700 nm for spectral acquisition. The tests were performed by using air as the background. Each sample was scanned seven times, and the spectral scanning results were determined by averaging the scans. To minimize the influence of instrument conditions and environmental factors, each sample was tested five times, and the average value was used as the input spectral variable for modeling.

4.3. Determination of Standard Values for Physical and Chemical Properties. The reference values of 11 physicochemical properties of jet fuel no. 3 were tested in accordance with the standard methods specified in the "GB 6537-2018 no. 3 jet fuel". Specific properties include: initial and final boiling points, 10, 20, 50, and 90% distillation points, density at 20 °C, kinematic viscosities at 20 and −20 °C, and freezing and flash points.

4.4. Model Establishment and Evaluation. The NIMSPEC spectrometry software (National Institute of Metrology, China) was used to complete the preprocessing of the spectral data, selection of feature bands, and model establishment. The spectral data were preprocessed using SG smoothing, SG first D, SG second D, and MSC. Three commonly used feature wavelength selection methods, VIP, UVE, and MC-UVE, were employed to select feature bands from the full-wavelength NIR spectra of the jet fuel. The preprocessed spectral data of the calibration set samples and the corresponding reference values of the 11 physicochemical properties were used as input variables for the model. Jet fuel multi-index determination models were established using PLS and O-PLS algorithms (the PLS modeling methods used were PLS1). Using leave-one-out cross-validation (LOOCV), we sequentially extracted sample data from the calibration data set, established the model with the remaining data, and then predicted with the extracted data. After all data in the calibration set had been cycled through once, the RMSECV was calculated using the cross-validation predicted values and the optimal number of factors was determined for the regression model based on the RMSECV.[32−34] The performance of the established measurement model was evaluated using the coefficient of determination for the calibration set (R_c^2), coefficient of determination for the validation set (R_p^2), RMSEP, and RMSEC. The accuracy of the model was assessed using RPD. Under optimal conditions, predictive models for 11 physicochemical indices were established, and the performance evaluation referred to

https://doi.org/10.1021/acsomega.3c09994
ACS Omega 2024, 9, 16138−16146

Table 2. Results of 11 Metrics Predicted Using the SG 2nd D-UVE-PLS Model

indicator	reproducibility	unknown sample no. 1			unknown sample no. 2			unknown sample no. 3			unknown sample no. 4		
		reference	measurement	error	reference	measurement	error	reference	measurement	error	reference	measurement	error
initial distillation point (°C)	±7.2	159.0	156.7	−2.3	156.5	156.7	0.2	154.0	155.4	1.4	152.5	155.5	3.0
10% distillation point (°C)	±5.64	170.0	165.7	−4.3	169.5	167.7	−1.8	170.5	168.2	−2.3	180.0	178.0	−2.0
20% distillation point (°C)	±6.87	175.5	170.3	−5.2	174.4	175.3	0.9	175.0	172.4	−2.6	187.5	187.0	−0.5
50% distillation point (°C)	±6.87	187.0	183.9	−3.1	185.0	183.9	−1.1	190.0	188.4	−1.6	199.0	198.6	−0.4
90% distillation point (°C)	±4.53	210.5	210.6	0.1	208.3	209.6	1.3	218.0	218.7	0.7	217.5	219.9	2.4
final distillation point (°C)	±10.5	247.0	244.3	−2.7	246.0	244.3	−1.7	251.0	254.3	3.3	239.0	236.1	−2.9
20 °C density (kg/m^3)	±1.2	792.4	791.9	−0.5	791.9	791.9	0.0	787.9	788.1	0.3	797.2	797.8	0.6
20 °C kinematic viscosity[a] (mm^2/s)	±2.2%	1.479	1.488	0.6%	1.509	1.498	−0.7%	1.570	1.578	0.5%	1.750	1.781	1.8%
−20 °C kinematic viscosity[a] (mm^2/s)		3.169	3.140	−0.9%	3.399	3.412	0.4%	3.453	3.388	−1.9%	3.860	3.920	1.6%
freezing point (°C)	±5.5	−63.0	−65.0	−2.0	−61.7	−65.0	−3.3	−58.4	−60.2	−1.8	−65.9	−66.2	−0.3
flash point (°C)	±4	45.0	45.4	0.4	44.5	45.4	0.9	44.5	45.4	0.9	45.0	45.7	0.7

[a]Standard "Petroleum product-determination of kinematic viscosity and calculation of dynamic viscosity" does not have explicit requirements for the reproducibility of the −20 °C kinematic viscosity test.

the average evaluation indicators of each model. The corresponding formulas are as follows

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

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

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

$$\mathrm{RMSECV} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad (4)$$

$$\mathrm{RMSEP} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \hat{y}_i)^2} \quad (5)$$

$$\mathrm{RPD} = \sqrt{\frac{\sum_{i=1}^{m}(\hat{y}_i - \bar{y})^2}{\sum_{i=1}^{m}(\hat{y}_i - y_i)^2}} \quad (6)$$

the number of samples in the calibration set is n, and the number of samples in the prediction set is m. y_i is the reference value of sample i, $\bar{y}$ is the average value of y_i, and $\hat{y}_i$ is the value of sample i predicted by the model.

A high R^2 value and low RMSEC and RMSEP values indicate that the model has a strong predictive ability.[35−37] A large RPD value reflects the reliability of the model; a reliable model can accurately determine new unknown samples. RPD values < 1.4, between 1.4 and 2.0, and > 2 indicate that the constructed model is unreliable, relatively reliable, and has a high degree of reliability, respectively.[38−40]

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsomega.3c09994.

Result of PLS and O-PLS modeling on 11 indicators using raw spectra, and spectra processed by various preprocessing methods and feature spectral band selection techniques (PDF)

■ AUTHOR INFORMATION

Corresponding Author

Zhengdong Zhang − *Center for Environmental Metrology, National Institute of Metrology, Beijing 100029, 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, China;* orcid.org/0000-0001-9015-7279

Xin Zhang − *College of Environmental and Chemical Engineering, Dalian University, Dalian 116622, China*

Jing Zhang − *College of Environmental and Chemical Engineering, Dalian University, Dalian 116622, China*

Biao Du − *Beijing Yixingyuan Petrochemical Technology Co., Ltd., Beijing 101301, China*

https://doi.org/10.1021/acsomega.3c09994
ACS Omega 2024, 9, 16138−16146

Xiaoping Song − *Center for Environmental Metrology, National Institute of Metrology, Beijing 100029, China*
Guixuan Wang − *Beijing Yixingyuan Petrochemical Technology Co., Ltd., Beijing 101301, China*
Qi Li − *Center for Environmental Metrology, National Institute of Metrology, Beijing 100029, China*
Yinglan Zhang − *Leibniz Institut für Polymerforschung Dresden e.V., Dresden 01069, Germany; Institut für Werkstoffwissenschaft, Technische Universität Dresden, Dresden 01062, Germany*
Fan Liu − *Center for Environmental Metrology, National Institute of Metrology, Beijing 100029, China*

Complete contact information is available at: https://pubs.acs.org/10.1021/acsomega.3c09994

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

The work was supported by the State Administration for Market Regulation for Technical Guarantee of Market Supervision Special Program, China (grant no. 2023YJ03), the State Administration for Market Regulation Science and Technology Program, China (grant no. 2021MK153), and the National Institute of Metrology, China (grant nos. AKYZZ2131 and AKYZZ2332).

■ ABBREVIATIONS

NIR, near-infrared; PLS, partial least-squares; O-PLS, orthogonal partial least-squares; R_c^2, calibration set regression coefficient; R_p^2, prediction set regression coefficient; RMSEC, root-mean-square error of calibration; RMSECV, root-mean-square error of cross-validation; RMSEP, root-mean-square error of prediction; RPD, residual prediction deviation; BP-ANN, backpropagation neural network; iPLS, interval partial least-squares; SEV, validation standard error; SG smoothing, Savitzky−Golay smoothing; SG first D, Savitzky−Golay first derivative; SG second D, Savitzky−Golay second derivative; MSC, multiple scatter correction; VIP, variable importance projection; UVE, uninformative variable elimination; MC-UVE, Monte Carlo uninformative variable elimination

■ REFERENCES

(1) Chen, T.; Xu, X.; Hu, J. Q.; Guo, L.; Yang, S. Z.; Zhao, T. X.; Ma, J. Water behavior of current jet fuel versus operating conditions: Storage time, temperature, relative humidity and anti-icing agent. *Fuel* **2022**, *309*, 122088.
(2) Jin, X.; Ba, W. J.; Wang, L. L.; Zhang, T.; Zhang, X. D.; Li, S. W.; Rao, Y.; Liu, L. A Novel Tran_NAS Method for the Identification of Fe-and Mg-Deficient Pear Leaves from N-and P-Deficient Pear Leaf Data. *ACS Omega* **2022**, *7* (44), 39727−39741.
(3) Chen, S.; Ma, M.; Peng, J.; He, X.; Wang, Q.; Chu, G. Rapid prediction method of ZIF-8 immobilized Candida rugosa lipase activity by near-infrared spectroscopy. *Spectrochim. Acta, Part A* **2023**, *302*, 123072.
(4) Yan, Z.; Liu, H.; Li, J.; Wang, Y. Qualitative and quantitative analysis of Lanmaoa asiatica in different storage years based on FT-NIR combined with chemometrics. *Microchem. J.* **2023**, *189*, 108580.
(5) Khodasevich, M. A.; Borisevich, D. A. Identification of Flax Oil by Linear Multivariate Spectral Analysis. *J. Appl. Spectrosc.* **2020**, *86* (6), 996−999.
(6) Barbosa, M. F.; Santos, J. R. B.; Silva, A. N.; Soares, S. F. C.; Araujo, M. C. U. A cheap handheld NIR spectrometric system for automatic determination of methane, ethane, and propane in natural gas and biogas. *Microchem. J.* **2021**, *170*, 106752.
(7) Wang, L.; Li, J. Q.; Li, T.; Liu, H. G.; Wang, Y. Z. Method Superior to Traditional Spectral Identification: FT-NIR Two-Dimensional Correlation Spectroscopy Combined with Deep Learning to Identify the Shelf Life of Fresh Phlebopus portentosus. *ACS Omega* **2021**, *6* (30), 19665−19674.
(8) Ma, B.; Cai, W. S.; Shao, X. G. Analyzing the Water Confined in Hydrogel Using Near-Infrared Spectroscopy. *Appl. Spectrosc.* **2022**, *76* (7), 773−782.
(9) Su, P. F.; Liang, W. H.; Zhang, G.; Wen, X. Y.; Chang, H.; Meng, Z. H.; Xue, M.; Qiu, L. L. Quantitative Detection of Components in Polymer-Bonded Explosives through Near-Infrared Spectroscopy with Partial Least Square Regression. *ACS Omega* **2021**, *6* (36), 23163−23169.
(10) Assi, S.; Arafat, B.; Lawson-Wood, K.; Robertson, I. Authentication of Antibiotics Using Portable Near-Infrared Spectroscopy and Multivariate Data Analysis. *Appl. Spectrosc.* **2021**, *75* (4), 434−444.
(11) Li, K.; Zhang, C.; Du, B. A.; Song, X. P.; Li, Q.; Zhang, Z. D. Selection of the Effective Characteristic Spectra Based on the Chemical Structure and Its Application in Rapid Analysis of Ethanol Content in Gasoline. *ACS Omega* **2022**, *7* (23), 20291−20297.
(12) Zhang, H.; Hu, X.; Liu, L.; Wei, J.; Bian, X. Near infrared spectroscopy combined with chemometrics for quantitative analysis of corn oil in edible blend oil. *Spectrochim. Acta, Part A* **2022**, *270*, 120841.
(13) Hooshyari, M.; Casale, M. Identification of Base Stock in Engine Oils by Near Infrared and Fluorescence Spectroscopies Coupled with Chemometrics. *Surv. Geophys.* **2021**, *42* (2), 451−465.
(14) Leal, A. L.; Silva, A. M. S.; Ribeiro, J. C.; Martins, F. G. Using Spectroscopy and Support Vector Regression to Predict Gasoline Characteristics: A Comparison of ^{1}H NMR and NIR. *Energy Fuels* **2020**, *34* (10), 12173−12181.
(15) Xiao, H.; Ju-xiang, W.; Jie, L. Quantitative Prediction of Total Acid Number in Aviation Kerosene by BP-ANN and NIR Spectroscopy. *J. Anal. Sci.* **2011**, *27* (06), 751−754.
(16) Yangjun, L.; Juxiang, W.; Zhina, X. Study on Near Infrared Analysis Model of Initial Distillation Point of Jet Fuel Based on iPLS and BP-ANN. *J. Ordnance Equip. Eng.* **2020**, *41* (06), 212−215.
(17) Yao, Y.-N.; Ma, K.; Zhu, J. F.; Huang, F.; Kuang, L.; Wang, X.; Li, S. Non-destructive determination of soluble solids content in intact apples using a self-made portable NIR diffuse reflectance instrument. *Infrared Phys. Technol.* **2023**, *132*, 104714.
(18) Tang, T.; Chen, C.; Wu, W.; Zhang, Y.; Han, C.; Li, J.; Gao, T.; Li, J. Hyperspectral Inversion Model of Relative Heavy Metal Content in Pennisetum sinese Roxb via EEMD-db3 Algorithm. *Remote Sens.* **2023**, *15* (1), 251.
(19) Lee, L. C.; Liong, C.-Y.; Jemain, A. A. A contemporary review on Data Preprocessing (DP) practice strategy in ATR-FTIR spectrum. *Chemom. Intell. Lab. Syst.* **2017**, *163*, 64−75.
(20) Jerry, W.; Lois, W. *Practical Guide To Interpretive Near infrared Spectroscopy. Translated by CIU Xiao-li, XU Yu-peng, TAN Gao-you*; Chemical Industry Press: Beijing, 2009.
(21) Liu, L.; Jiang, H.; Chen, Q. Rapid determination of acidity index of peanut during storage by a portable near-infrared spectroscopy system. *Infrared Phys. Technol.* **2022**, *127*, 104472.
(22) Peng, D.; Liu, Y.; Yang, J.; Bi, Y.; Chen, J. Nondestructive Detection of Moisture Content in Walnut Kernel by Near-Infrared Diffuse Reflectance Spectroscopy. *J. Spectrosc.* **2021**, *2021*, 1−9.
(23) Sanchez-Pinto, L. N.; Venable, L. R.; Fahrenbach, J.; Churpek, M. M. Comparison of variable selection methods for clinical predictive modeling. *Int. J. Med. Inf.* **2018**, *116*, 10−17.
(24) Luo, M.; Wang, Y.; Xie, Y.; Zhou, L.; Qiao, J.; Qiu, S.; Sun, Y. Combination of Feature Selection and CatBoost for Prediction: The First Application to the Estimation of Aboveground Biomass. *Forests* **2021**, *12* (2), 216.
(25) Ottavian, M.; Fasolato, L.; Serva, L.; Facco, P.; Barolo, M. Data Fusion for Food Authentication: Fresh/Frozen-Thawed Discrim-

ination in West African Goat-fish (Pseudupeneus prayensis) Fillets. *Food Bioprocess Technol.* **2014**, *7* (4), 1025−1036.
(26) Huang, X.; Xia, L. Application of iterative distance correlation and PLS for wavelength interval selection in near infrared spectroscopy. *Chemom. Intell. Lab. Syst.* **2023**, *234*, 104756.
(27) Lu, B.; Liu, N.; Li, H.; Yang, K.; Hu, C.; Wang, X.; Li, Z.; Shen, Z.; Tang, X. Quantitative determination and characteristic wavelength selection of available nitrogen in coco-peat by NIR spectroscopy. *Soil Tillage Res.* **2019**, *191*, 266−274.
(28) Zhang, L.; Li, Y.; Huang, W.; Ni, L.; Ge, J. The method of calibration model transfer by optimizing wavelength combinations based on consistent and stable spectral signals. *Spectrochim. Acta, Part A* **2020**, *227*, 117647.
(29) Li, H.-D.; Liang, Y.-Z.; Cao, D.-S.; Xu, Q.-S. Model-population analysis and its applications in chemical and biological modeling. *TrAC, Trends Anal. Chem.* **2012**, *38*, 154−162.
(30) Gagnebin, Y.; Pezzatti, J.; Lescuyer, P.; Boccard, J.; Ponte, B.; Rudaz, S. Combining the advantages of multilevel and orthogonal partial least squares data analysis for longitudinal metabolomics: Application to kidney transplantation. *Anal. Chim. Acta* **2020**, *1099*, 26−38.
(31) Makraduli, L.; Makreski, P.; Goracinova, K.; Stefov, S.; Anevska, M.; Geskovski, N. A Comparative Approach to Screen the Capability of Raman and Infrared (Mid- and Near-) Spectroscopy for Quantification of Low-Active Pharmaceutical Ingredient Content Solid Dosage Forms: The Case of Alprazolam. *Appl. Spectrosc.* **2020**, *74* (6), 661−673.
(32) Li, H.-D.; Xu, Q.-S.; Liang, Y.-Z. libPLS: An integrated library for partial least squares regression and linear discriminant analysis. *Chemom. Intell. Lab. Syst.* **2018**, *176*, 34−43.
(33) Tan, J.; Li, R.; Jiang, Z.-T.; Tang, S.-H.; Wang, Y.; Shi, M.; Xiao, Y.-Q.; Jia, B.; Lu, T.-X.; Wang, H. Synchronous front-face fluorescence spectroscopy for authentication of the adulteration of edible vegetable oil with refined used frying oil. *Food Chem.* **2017**, *217*, 274−280.
(34) Levate Macedo, L.; da Silva Araújo, C.; Costa Vimercati, W.; Gherardi Hein, P. R.; Pimenta, C. J.; Henriques Saraiva, S. Evaluation of chemical properties of intact green coffee beans using near-infrared spectroscopy. *J. Sci. Food Agric.* **2021**, *101* (8), 3500−3507.
(35) Caporaso, N.; Whitworth, M. B.; Fisk, I. D. Protein content prediction in single wheat kernels using hyperspectral imaging. *Food Chem.* **2018**, *240*, 32−42.
(36) Luo, D.; Gao, Y.; Wang, Y.; Shi, Y.; Chen, S.; Ding, Z.; Fan, K. Using UAV image data to monitor the effects of different nitrogen application rates on tea quality. *J. Sci. Food Agric.* **2022**, *102* (4), 1540−1549.
(37) Li, M.; Han, D. H.; Liu, W. Non-destructive measurement of soluble solids content of three melon cultivars using portable visible/near infrared spectroscopy. *Biosyst. Eng.* **2019**, *188*, 31−39.
(38) Hu, Y.; Huang, P.; Wang, Y.; Sun, J.; Wu, Y.; Kang, Z. Determination of Tibetan tea quality by hyperspectral imaging technology and multivariate analysis. *J. Food Compos. Anal.* **2023**, *117*, 105136.
(39) Liu, G.; Zhou, X.; Li, Q.; Shi, Y.; Guo, G.; Zhao, L.; Wang, J.; Su, Y.; Zhang, C. Spatial distribution prediction of soil As in a large-scale arsenic slag contaminated site based on an integrated model and multi-source environmental data. *Environ. Pollut.* **2020**, *267*, 115631.
(40) Liu, J.; Dong, Z.; Xia, J.; Wang, H.; Meng, T.; Zhang, R.; Han, J.; Wang, N.; Xie, J. Estimation of soil organic matter content based on CARS algorithm coupled with random forest. *Spectrochim. Acta, Part A* **2021**, *258*, 119823.

https://doi.org/10.1021/acsomega.3c09994
ACS Omega 2024, 9, 16138−16146