


## Article

# Practical Online Characterization of the Properties of Hydrocracking Bottom Oil via Near-Infrared Spectroscopy

Jian Long <sup>1,2</sup> , Tiantian Xu <sup>1</sup> and Chen Fan <sup>1,2,\*</sup>

<sup>1</sup> Key Laboratory of Smart Manufacturing in Energy Chemical Process, East China University of Science and Technology, Ministry of Education, Shanghai 200237, China

<sup>2</sup> Engineering Research Center of Process System Engineering, East China University of Science and Technology, Ministry of Education, Shanghai 200237, China

\* Correspondence: fanchen@ecust.edu.cn; Tel.: +86-021-64253720

**Abstract:** Providing real-time information on the chemical properties of hydrocracking bottom oil (HBO) as the feedstock for ethylene cracker while minimizing processing time, is important to improve the real-time optimization of ethylene production. In this study, a novel approach for estimating the properties of HBO samples was developed on the basis of near-infrared (NIR) spectra. The main noise and extreme samples in the spectral data were removed by combining discrete wavelet transform with principal component analysis and Hotelling's T2 test. Kernel partial least squares (KPLS) regression was utilized to account for the nonlinearities between NIR data and the chemical properties of HBO. Compared with the principal component regression, partial least squares regression, and artificial neural network, the KPLS model had a better performance of obtaining acceptable values of root mean square error of prediction (RMSEP) and mean absolute relative error (MARE). All RMSEP and MARE values of density, Bureau of Mines correlation index, paraffins, isoparaffins, and naphthenes were less than 1.0 and 3.0, respectively. The accuracy of the industrial NIR online measurement system during consecutive running periods in predicting the chemical properties of HBO was satisfactory. The yield of high value-added products increased by 0.26 percentage points and coil outlet temperature decreased by 0.25 °C, which promoted economic benefits of the ethylene cracking process and boosted industrial reform from automation to digitization and intelligence.

**Keywords:** hydrocracking bottom oil; near-infrared; kernel partial least squares; online measurement; ethylene cracking



**Citation:** Long, J.; Xu, T.; Fan, C. Practical Online Characterization of the Properties of Hydrocracking Bottom Oil via Near-Infrared Spectroscopy. *Processes* **2023**, *11*, 829. <https://doi.org/10.3390/pr11030829>

Academic Editor: Albert Ratner

Received: 2 January 2023

Revised: 28 February 2023

Accepted: 7 March 2023

Published: 10 March 2023



**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

### 1.1. Background

Ethylene is an important organic compound in the petrochemical industry, and its production substantially dwarfs that of other petrochemical products [1]. Ethylene production plays a pivotal role in petrochemical production, and the extent of this trade is often regarded as a significant indicator of a country's level of industrialization [2]. Thermal cracking, as the primary force in the ethylene production process, utilizes cracking furnaces as major units to decompose feedstock into small molecules from a wide range of feedstock, such as ethane, propane, butane liquefied petroleum gas, naphtha, and hydrocracking bottom oil (HBO) [3,4].

HBO is obtained from the hydrocracking reaction unit (HCR), which is an essential process that converts heavy oil fractions with catalysts and excess hydrogen into a variety of valuable products [5]. Recently, the cost of feedstock accounts for over 90% of the total expense of ethylene production. The choice of ethylene feedstock largely depends on the availability of the raw material and profitability. Thus, HBO has become one of the preferred feedstocks due to its ability to produce high yields of ethylene, propylene, and butadiene.

Usually, the mass fraction of paraffins, isoparaffins, olefins, naphthenes, and aromatics (PIONA) is used to characterize the cracking performance of ethylene feedstock. In the cracking process, paraffins yield the highest amount of ethylene; isoparaffins readily form propylene in the cracking reaction; naphthenes tend to produce aromatic hydrocarbons; olefins generate acetylene and coke; and aromatics are prone to form heavy aromatic hydrocarbons [6,7]. Therefore, the properties of HBO, encompassing PIONA, density, boiling range, and the Bureau of Mines correlation indexes (*BMC I*), which change according to the properties of crude oil, processing technologies, and operating conditions, are crucial to the various operations involved in ethylene cracking, such as modeling, simulation, control, and optimization, especially in real-time monitoring and optimization [8].

### 1.2. Previous Research

These parameters can be assessed via standard procedures such as gas chromatograph-mass spectrometry; however, they tend to be time-consuming and costly. Nowadays, intelligent reformation of ethylene industries is of wide concern [9]. Fast and accurate property information of feedstock will improve aforementioned intelligence in the ethylene cracking process, leading to benefits for enterprises [10]. Therefore, real-time information on the chemical properties of HBO should be provided while minimizing the analysis and processing time to obtain a considerable amount of high-quality products.

Near infrared (NIR) spectroscopy has been proven to be highly valuable in predicting product quality as it can obtain information about multiple kinds of chemical bonds (e.g., C–H, O–H, or N–H) in molecules [11]. Thus, the application of this technique in the online detection of sample quality has received increasing attention due to its speed, low cost, and minimal sample preparation requirements [12–14]. It is extensively employed for predicting qualitative and quantitative properties of the products in various industries, such as agricultural, food, and pharmaceutical industries [15]. Optimal models on the basis of NIR spectroscopy have been developed to predict the quality of various types of materials, such as biorefinery feedstocks [16], tomatoes, and alternative fuels [17].

Recent studies have employed chemometric calibration modeling techniques combined with NIR spectroscopy to identify some of the most important properties of petroleum and its distillates [18], such as crude oil [19,20], gasoline [21], diesel, and naphtha [22]. By utilizing this technique, especially with a flexible NIR fiber probe [23,24], many critical parameters of petroleum and its products can be rapidly predicted. These predictions can be utilized to support rapid laboratory analysis of samples or online monitoring, controlling, and optimization of refinery processes, such as crude distillation, catalytic cracking, naphtha cracking, and gasoline and diesel blending [25–27].

The construction of high-quality quantitative models is challenging due to the large amounts of variables, noise, and overlapping bands [28]. To address this issue, some methods have been applied to accurately predict oil properties. For example, standard normal variate (SNV), wavelet transform (WT), and detrending have been employed to distinguish noise, improve resolution and sensitivity, and extract useful features in the NIR data preprocessing [29,30]. Methods that can be used to reduce the size of data sets used for training prediction models include correlation coefficient, two-dimensional correlation, principal component analysis (PCA), and Hotelling's T<sup>2</sup> [31–35]. Many modeling techniques are available for quantitative data analysis, including principal component regression (PCR), partial least squares (PLS), locally weighted regression, artificial neural networks (ANN), support vector regression, and the kernel partial least squares (KPLS), which fall under the category of multivariate or machine learning regression methods. They can be applied to analyze the correlation of NIR spectral data [36–42]. Additionally, bilinear approaches and the introduction of nonlinear kernels in support vector machine regression can assist in analyzing data with a strong nonlinear correlation. Recent studies indicate that novel chemometric approaches, including machine learning methods, can provide some advantage in the space nonlinearity and the time-varying issue of the process simultaneously [36,38].

However, unlike light oils or off-line measurements, HBO is a complicated hydrogenated oil with assorted internal molecular structures [43]. The NIR spectral peaks of the HBO samples overlap significantly. Their NIR spectral data fluctuate with the dynamic detection environment, such as the contamination degree of the NIR probe [44,45]. For a long-term online measurement of HBO properties in the dynamic ethylene cracking process, these factors pose a great challenge to the prediction accuracy and robustness of the HBO prediction model.

The aforementioned methods concentrate on addressing the nonlinearities of oil properties and NIR spectra data or ensuring the robustness of predictive models on an individual level. However, this approach may not yield a reliable predictive ability for HBO quality parameters. Generally, regression methods can provide a good solution to train the predictive model if the NIR spectra of the HBO samples can be denoised and abnormal oil samples can be removed. Meanwhile, there is interest in developing a model that incorporates the effects of oil properties on process control and profitability.

### 1.3. Motivation and Main Contributions

Real-time and accurate property information of HBO, such as PIONA, density, boiling range, and *BMCI* for a long run period, is essential and beneficial for intelligent operation and profitability in the ethylene cracking process. Aforementioned methods of NIR modeling are effective for offline prediction; however, the advantages of online prediction are not sufficient. The motivation of this study is to address the issue of using NIR data to measure the quality parameters including the physical and chemical properties of HBO in refineries, and to reveal the control and profitability improvement of this online measurement technique for real-time optimization of ethylene cracking.

To improve the accuracy and robustness of the property prediction model and realize the integration of online measurement and online optimization of ethylene cracking, a practical online characterization of HBO properties was proposed in this paper. An NIR absorption spectrometer that was equipped with a transmission probe was applied to implement the approach. The main noise of the spectral data was removed by applying discrete wavelet transform (DWT). Extreme samples were marked out to solve the robustness problem by combining principal component analysis (PCA) and Hotelling's T2 test. The nonlinearities between the quality parameters of HBO and NIR data were constructed by KPLS algorithm to improve prediction accuracy. The method can improve the predictive ability of the models, thereby reducing the risk of performance degradation. Moreover, integrated applications of the aforementioned rapid measurement methods and real-time optimization of ethylene cracking unit were carried out.

The main contributions of the paper are as follows:

- (1) A novel approach of DWT combining PCA with Hotelling's T2 test was proposed to solve the robustness problem of NIR prediction model.
- (2) NIR prediction models based on KPLS were constructed to predict HBO properties more accurately.
- (3) An optimization framework integrating HBO online analysis and real-time optimization module was proposed.

## 2. Experimental Section

### 2.1. HBO Sample Collection

Over 130 HBO samples, which consisted of mixtures of bottom oils from a medium-pressure HCR unit and a high-pressure HCR unit, were collected from the feed pipeline of an ethylene cracking furnace located in an eastern Chinese refinery. An automatic sampler was utilized to collect HBO samples in the form of oil drops from the pipeline for a sampling period of 1 h. Approximately 100 mL of the HBO sample was continuously collected to ensure a greater number of HBO droplets were obtained from the sample. This step ensured that the as-obtained HBO samples were composed of many small HBO oil droplets at different time periods, which reflected the properties of the HBO sample as

it flowed through the pipeline until relatively stable oil properties were achieved. The interval between collecting different samples, which lasted for 2–3 days, depended on changes in the properties of crude oil processed by the crude distillation unit and the operating conditions of the hydrocracking unit. Consequently, the different HBO samples had vastly different properties.

## 2.2. Laboratory Analysis of HBO Properties

The *BMCI* is widely used as an aromatic index for oil cracking, and the optimal *BMCI* value for hydrocracking of HBO is normally lower than 12. The *BMCI* was calculated as follows.

$$BMCI = \frac{48,640}{T + 273} + 473.7 \times d_{15.6}^{15.6} - 456.8 \quad (1)$$

where  $T$  is the average boiling point ( $^{\circ}\text{C}$ ) and  $d_{15.6}^{15.6}$  is the specific gravity at  $15.6^{\circ}\text{C}$  ( $\text{g}/\text{cm}^3$ ).

The density, *BMCI*, and PIONA content of the liquid feedstock are crucial for the optimization and control of the ethylene cracking process.

Density and average boiling point ( $T$ ) (such as 10% point, 30% point, 50% point, 70% point, and 90% point) were determined following the American Society for Testing and Materials (ASTM) D1298 and D86, respectively. The PIONA content of HBO samples was determined via the reproductive laboratory analysis method [46]. The chemical compositions of the HBO samples in terms of hydrocarbon families and carbon number distributions were analyzed via gas chromatograph [47] (Agilent 7890). Mass spectral data were obtained using an Agilent GC–MS/7890-5975C operating in the electron impact positive ion mode within the pyrolysis temperature range of  $350$ – $750^{\circ}\text{C}$ . The Agilent 7890 GC was equipped with a programmed temperature vaporization system, a flame ionization detector, a cryogenic cooled oven compartment, and a capillary column of fused silica, 30 m long and  $0.25\ \mu\text{m}$  wide. The mass range was  $1\ m/z$ – $1050\ \text{amu}\ m/z$ , the resolution was unit mass resolution, the sensitivity was 100 pg octafluoronaphthalene,  $S/N \geq 10:1$  at  $272\ m/z$ . The sample was extracted with a  $0.4$ – $1.0\ \mu\text{L}$  microsyringe during the experiment. The working condition of the instrument is to adjust the repulsion pole of the ion source to maximize the molecular ion peak  $m/e\ 226$  of *n*-cetane. The magnetic field scanning was done from  $m/e\ 65$ – $410$ . The ionization voltage was 70 eV and the ionization current was  $10$ – $70\ \mu\text{A}$ . PIONA content was calculated using software (Carbon Distribution of Diesel Oil 2005) for calculating wax oil carbon number distribution and an MS software (Beijing Research Institute of Petroleum Processing of China Petroleum & Chemical Corporation, Beijing, China) for analyzing the hydrocarbon composition of vacuum gas oil.

## 2.3. Online NIR Detection System and HBO Spectra Collection

The HBO sampling site and the online method for measuring the NIR spectra of the HBO samples are shown in Figure 1. An online NIR Fourier transform spectrometer (Bruker, Germany), equipped with a tungsten halogen lamp, was used to online collect NIR spectra. A liquid immersion fiber optic probe with a 1.0 mm optical path length was installed on the feeding pipeline before the cracking furnace. The probe was connected to a spectrometer via a 50 m fiber optics cable. When the spectrometer and illumination are properly installed, this transmission mode of the NIR probe could collect more information about the HBO samples, as the light can penetrate through them. It can be used for a comprehensive assessment of the quality of HBO if satisfactory spectral information can be obtained and processed. The NIR spectra were scanned within the wavelength range of  $4500\ \text{cm}^{-1}$ – $12,000\ \text{cm}^{-1}$ . During a 1 h HBO sampling period, 30 NIR spectra of the HBO sample were recorded at 2 min intervals. Each spectrum consisted of 32 accumulation scans, resulting in over 4000 spectra collected in total. The NIR spectrum of a sample used for modeling was constructed by taking the average of the 30 spectra, ensuring spectral reproducibility.



**Figure 1.** Sampling site and online method for measuring the NIR spectra of the HBO samples.

## 2.4. Preprocessing and Analysis of the Spectral Data

### 2.4.1. WT for NIR Spectrum Preprocessing

A previous study had shown that WT can be used to denoise NIR data [48]. In the present study, the spectral data were denoised via DWT. The signals transformed by DWT were represented using Equation (2) below:

$$X(t) = \sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}} C(j, k) 2^{-\frac{j}{2}} \psi(2^{-\frac{j}{2}} t - k) \quad (2)$$

where  $X(t)$  is the function transferred by WT,  $\psi$  denotes the mother function, and  $j$  and  $k$  are dimensions, wherein  $k$  is an integer value.

The NIR spectra of the HBO samples were preprocessed by DWT to denoise and improve resolution. In this study, one-dimensional discrete wavelet denoising was employed. The wavelet discrete Meyer family has been shown to be relatively accurate in decomposing temporal wind signals and time-series data [49,50]. Moving average transform was utilized to smoothen the raw spectral data of different segment sizes. SNV, multiplicative scattering correction (MSC), SNV-minus a straight line (SNV-MASL), min-max normalization (MMN), and DWT were used for preprocessing spectral data.

### 2.4.2. PCA and Hotelling's T2 Test

PCA can be used to find combinations of variables that capture the main trends in the data set and to visualize the most important information [51]. Given a data matrix  $\mathbf{X}$ , its covariance matrix is calculated as follows [52].

$$\text{cov}(\mathbf{X}) = \frac{\mathbf{X}^T \mathbf{X}}{m - 1} \quad (3)$$

where  $m$  indicates samples. The result of the PCA procedure is given as follows:

$$\mathbf{X}_{n \times m} = \mathbf{t}_1 \mathbf{p}_1^T + \mathbf{t}_2 \mathbf{p}_2^T + \mathbf{t}_i \mathbf{p}_i^T + \dots + \mathbf{t}_k \mathbf{p}_k^T + \mathbf{E}_{n \times m} \quad (4)$$

where  $\mathbf{t}_i$  is the score vector,  $\mathbf{p}_i^T$  is the loading vector, and  $\mathbf{E}$  is the residual matrix. The original dataset can be adequately described by several orthogonal principal components (PCs) without losing substantial information, and the relationships between samples can be easily detected when the PCs are relative to each other.

Extreme samples can affect the accuracy of the established model, so they should be identified. One way of doing this is by using the Hotelling's T2 statistic. Hotelling's T2 is commonly used to detect extreme sample values and has relatively high power in all unbiased tests [53]. Assuming that  $s_{11}, s_{12}, \dots, s_{1N_1}$  and  $s_{21}, s_{22}, \dots, s_{2N_2}$  are independent normally distributed samples with parameters  $\lambda_1, \lambda_2$  and  $\sigma^2$ , i.e.,  $N(\lambda_1, \sigma^2)$  and  $N(\lambda_2, \sigma^2)$ , Hotelling's T2 test analyzes the null hypothesis  $H_0: \bar{\lambda}_1 = \bar{\lambda}_2$  against the alternative one,  $H_1: \bar{\lambda}_1 \neq \bar{\lambda}_2$ . In performing this test, the mean ( $\bar{s}_i$ ) and standard deviation ( $s_i$ ) of each sample are calculated as follows [54].

$$\bar{s}_j = \frac{1}{N} \sum_{j=1}^{N_i} s_{ij} \quad i = 1, 2 \quad (5)$$

$$s_i^2 = \frac{1}{n_j} \sum_{j=1}^{N_i} (x_{ij} - \bar{x}_i)^2 \quad i = 1, 2 \quad (6)$$

where  $N_i$  is the sample size, and  $n_i = N_i - 1$ . The pooled variance is then represented by

$$S_p^2 = \frac{n_1 S_1^2 + n_2 S_2^2}{f} \quad (7)$$

where  $f = n_1 + n_2 = N_1 + N_2 - 2$ . The null hypothesis is rejected if and only if

$$\left[ \left( \frac{N_1 N_2}{N_1 + N_2} \right) \frac{(\bar{x}_1 - \bar{x}_2)^2}{S_p^2} \right] > t_{f, \alpha/2}^2 = F_{1, f, \alpha} \quad (8)$$

where  $\bar{x}_i$  is the mean of each independent normally distributed sample,  $S_p^2$  is the pooled variance,  $t_{f, \alpha/2}$  is the critical value derived from the samples' t-distribution,  $\alpha$  is the significance level, and  $F_{1, f, \alpha}$  is the critical value derived from the Fisher distribution. The basic assumption of Hotelling's T2 test is that the distribution of data is normal.

In this study, PCA and Hotelling's T2 test were utilized to analyze the variation in spectral data and remove extreme samples.

#### 2.4.3. Selection of Characteristic Variables

The NIR spectrum region extends from 12,000  $\text{cm}^{-1}$  to 4000  $\text{cm}^{-1}$ , but the region generally used for rapid HBO analysis is between 10,000  $\text{cm}^{-1}$  and 4500  $\text{cm}^{-1}$ . Due to the large amounts of redundant information unrelated to the prediction of density, *BMCI*, and *PIONA* content, spectral regions related to specific properties should be selected, and spectral regions that do not overlap with absorption peaks of other functional groups should be avoided. In spectral modeling, the spectral data obtained according to the spectral wavelength are input variables for the model, and several important wavelengths related to response information should be selected as far as possible to enhance the predictive ability of the constructed calibration model. Different wavelength ranges were chosen based on the correlation between HBO properties and NIR absorbance and chemical properties, which are discussed in detail in Section 3.2.

#### 2.5. Development and Evaluation of Prediction Models

The NIR data and a corresponding estimation for the properties of the HBO are expressed as follows.

$$\mathbf{x} = [x_1, \dots, x_n]^T \quad (9)$$

$$\hat{\mathbf{y}} = [\hat{y}_1, \dots, \hat{y}_l]^T \quad (10)$$

where  $x_i, i = 1, \dots, n$  denotes the NIR absorbance at the  $i$ th wavelength;  $\hat{y}_j, j = 1, \dots, l$  denotes the  $j$ th estimated property.

$$\mathbf{x}_k = [x_{k1}, \dots, x_{kn}]^T, k = 1, \dots, m \quad (11)$$

$$\mathbf{y}_k = [x_{k1}, \dots, x_{kl}]^T, k = 1, \dots, m \quad (12)$$

where  $\mathbf{x}_k$  and  $\mathbf{y}_k$  are oil samples.  $m$  is the number of oil samples.

Therefore, the NIR spectral peak height data matrix  $\mathbf{X}$  and the corresponding property matrix  $\mathbf{Y}$  are calculated as follows.

$$\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m]^T \quad (13)$$

$$\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_m]^T \quad (14)$$

Five prediction models were constructed by PCR [55], PLS [56], ANN, and KPLS. Their results were compared and the best prediction approach was selected.

### 2.5.1. KPLS

The KPLS models can project the input  $\mathbf{X}$  to a high-dimensional figure space, which are represented as follows [57]:

$$\begin{cases} \Phi(\mathbf{X}) = \mathbf{TP}^T + \mathbf{E} \\ \mathbf{Y} = \mathbf{UQ}^T + \mathbf{F} \end{cases} \quad (15)$$

where  $\Phi$  indicates a nonlinear transformation.

The relationship of score matrices  $\mathbf{T}$  and  $\mathbf{U}$  can be represented as Equation (16).

$$\mathbf{T} = \Phi(\mathbf{X})\Phi(\mathbf{X})^T \mathbf{U}(\mathbf{T}^T \Phi(\mathbf{X})\Phi(\mathbf{X})^T \mathbf{U})^{-1} \quad (16)$$

The kernel function  $\mathbf{K}$  is represented as Equation (17).

$$\mathbf{K}(\mathbf{x}, \mathbf{y}) = \langle \Phi(\mathbf{x}), \Phi(\mathbf{y}) \rangle = \Phi\Phi^T \quad (17)$$

where  $\mathbf{x}$  and  $\mathbf{y}$  are the inputs and outputs, respectively.  $\langle \Phi(\mathbf{x}), \Phi(\mathbf{y}) \rangle$  denotes the dot product. The deflation of it is calculated by Equations (18) and (19).

$$\tilde{\mathbf{K}} = \left( I - \frac{1}{m} \mathbf{1}_m \mathbf{1}_m^T \right) \mathbf{K} \left( I - \frac{1}{m} \mathbf{1}_m \mathbf{1}_m^T \right) \quad (18)$$

$$I = \begin{bmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{bmatrix}_{m \times m}, \mathbf{1}_m = \begin{bmatrix} 1 & \dots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \dots & 1 \end{bmatrix}_{m \times m} \quad (19)$$

In this research, the dot product calculated by Gaussian kernel function is represented as Equation (20).

$$\mathbf{K}(\mathbf{x}, \mathbf{y}) = \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\sigma^2}\right) \quad (20)$$

where  $\sigma$  is the kernel parameter.

### 2.5.2. Significance Test on Model Accuracy

In building a prediction model, relevant samples from the dataset were randomly applied to construct a calibration model, and the remaining samples were applied to test the performance of the model. In the calibration subset, every method (i.e., PCR, PLS, and ANN) was employed to evaluate the methods by cross-validation [58]. This study used 10-fold cross-validation and averaged (or combined) the results of 10-fold results to produce

a single estimation. The root mean square errors of cross validation ( $RMSECV$ ), the root mean squared errors of prediction set ( $RMSEP$ ), and the mean absolute relative error ( $MARE$ ) were calculated using Equations (21)–(24), respectively, and they were applied to evaluate the constructed models. For prediction set samples, smaller  $RMSEP$  and  $MARE$  values indicate higher accuracy.  $R$  measured the correlation between predicted and true values. The closer the value of  $R^2$  was to 1, the better the model performance was.

$$RMSECV = \sqrt{\frac{\sum_{i=1}^n (y_{i,\text{exp}} - y_{i,\text{pre}})^2}{n}} \quad (21)$$

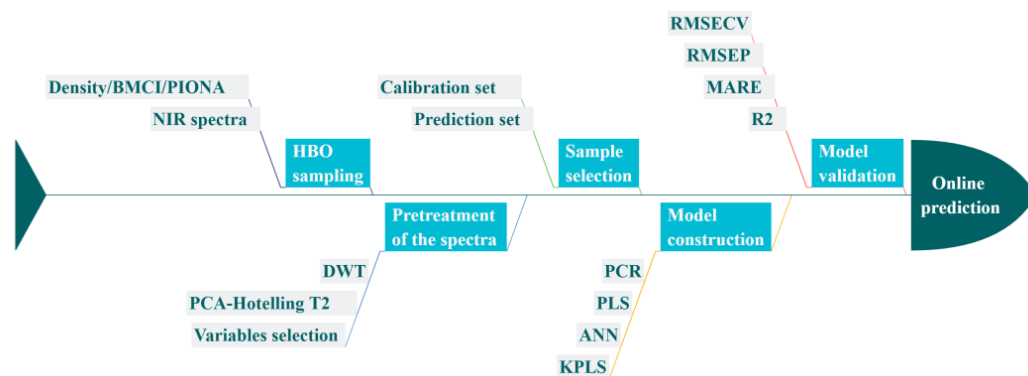
$$RMSEP = \sqrt{\frac{\sum_{i=1}^n (y_{i,\text{exp}} - y_{i,\text{pre}})^2}{n}} \quad (22)$$

$$MARE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_{i,\text{pre}} - y_{i,\text{exp}}}{y_{i,\text{exp}}} \times 100\% \right| \quad (23)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_{i,\text{exp}} - y_{i,\text{pre}})^2}{\sum_{i=1}^n (y_{i,\text{exp}} - \bar{y}_{i,\text{pre}})^2} \quad (24)$$

where  $n$  denotes the number of calibration/prediction set samples,  $y_{i,\text{pre}}$  denotes the predicted value for a given sample  $i$ ,  $y_{i,\text{exp}}$  denotes the experimental value, and  $\bar{y}_{i,\text{pre}}$  denotes the average of the response variable.

The flow chart illustrating how the methods predict the properties of HBO rapidly through transmission NIR spectroscopy is shown in Figure 2. The same sample sets and the same characteristic wavelength ranges of NIR spectroscopy were used to construct NIR prediction modes via PCR, PLS, ANN, and KPLS.



**Figure 2.** Flow chart of the proposed integrative approach for the rapid determination of HBO.

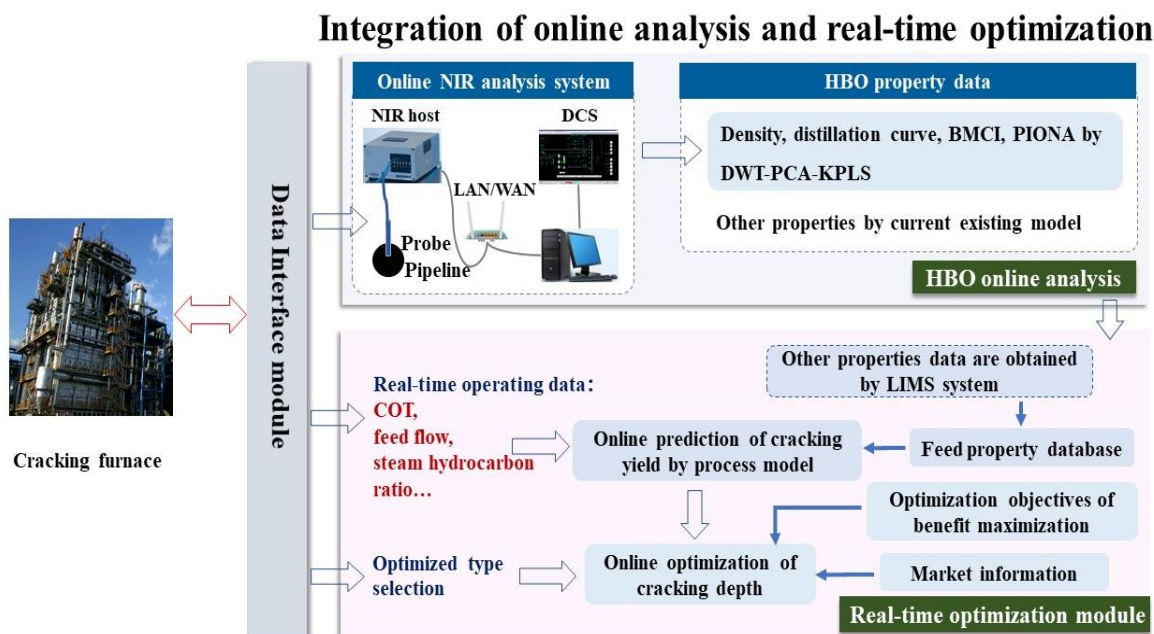
### 2.6. Real-Time Optimization of Ethylene Cracking Process Integrated with Online NIR Measurement System

A real-time optimization system was constructed to optimize the 360,000 t/year HBO ethylene cracker. The system had several functions, including optimization mode selection and setting, monitoring of cracking feed data, selection of operating conditions and setting of constraint ranges, results display of cracking depth optimization and dilution ratio (stream to feed ratio), and system status verification. The real-time optimization framework of ethylene cracking process is shown in Figure 3. It was composed of an online NIR detection system for HBO and a real-time optimization module for the ethylene cracking process. Every module constructed a security policy to ensure the safety and stability of the system.

In this case study, the online property analysis module utilizes NIR spectroscopy to provide accurate property data for HBO, including density, distillation curve,  $BMCI$ , and  $PIONA$ . These properties are obtained through practical online characterization of the



HBO, using the data interaction module. These data of properties are sent to the real-time optimization module. In this module, the yield of cracking products can be predicted by a process model of an ethylene cracking furnace with current operating conditions and feedstock properties.



**Figure 3.** Real-time optimization framework of ethylene cracking process integrated with online analysis property of HBO.

### 3. Results and Discussions

#### 3.1. Sample Statistics

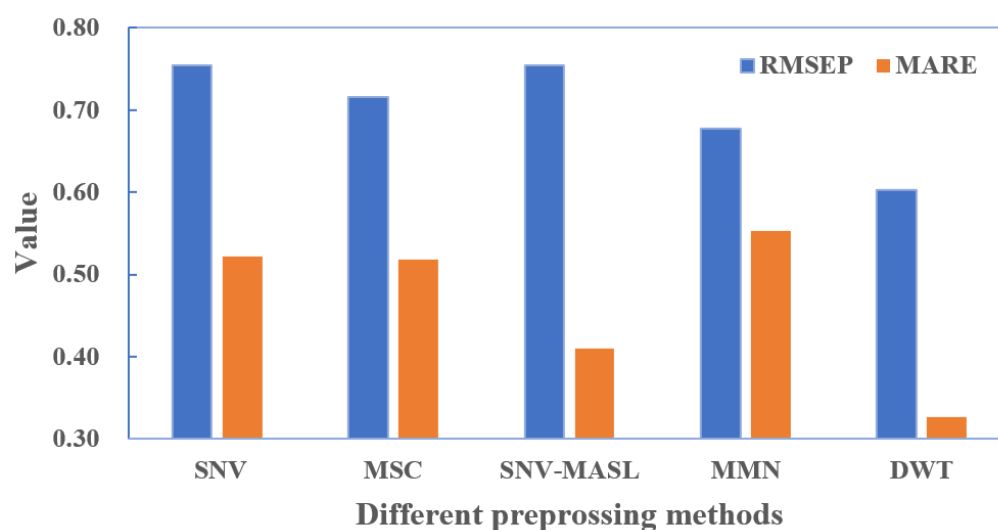
The properties of over 130 HBO samples were investigated in this study using reference methods. These samples were randomly divided into two subsets with a ratio of 6:4, which represented the subset of validation and prediction, respectively. The range of properties and the mean of the samples were calculated. Results showed that the methods had a good range and distribution. Density was 821.5–839.8 kg/m<sup>3</sup> with a mean of 832.2 kg/m<sup>3</sup> for the full set; the mean of the calibration subset and prediction subset were 832.4 kg/m<sup>3</sup> and 832.2 kg/m<sup>3</sup>, respectively. Moreover, the means of these properties for the calibration subset were close to those of the prediction subset used for external validation, which were 13.1, 26.2%, 26.8%, <0.1%, 46.1%, and 0.8% in the prediction subset. In addition, both subsets had distribution patterns that were similar and consistent with those of the full set. Owing to the low content of olefins in the HBO, these compounds were not considered in this modeling study. The characteristics of different types of samples provided are shown in the Table 1 below.

**Table 1.** Measured properties of HBO.

Properties	Range of Full Set	Mean		
		Full Set	Calibration on Subset	Prediction on Subset
Density (kg/m <sup>3</sup> )	821.5–839.8	832.2	832.4	832.2
BMCI	9.0–16.7	13.1	13.3	13.1
Paraffins	21.9–30.28	26.2	26.4	26.2
Isoparaffins	22.6–28.45	26.8	26.5	26.8
PIONA (Volume%)				
Olefins	<0.5	<0.1	<0.1	<0.1
Naphthenes	41.2–53.9	46.1	46.3	46.1
Aromatics	0–2.23	0.8	0.8	0.8

### 3.2. Spectral Features of HBO Samples and Spectroscopic Techniques

Noise elimination or reduction was performed to improve the model's performance. The OPUS software comes with the MSC, SNV, SNV-MASL, MMN, and DWT methods, which are used to eliminate or reduce noise. The DWT of the software was mainly used to deduct the influence of the instrument background or drift on the signal. MSC and SNV were applied to eliminate the effect of scattering on the spectrum due to the uneven distribution of particles and different particle sizes. MMN could be employed to eliminate the adverse effects caused by large scale differences. Therefore, the different methods used to preprocess HBO parameters (such as paraffins) are shown as in Figure 4, which were SNV, MSC, SNV-MASL, MMN, and DWT.



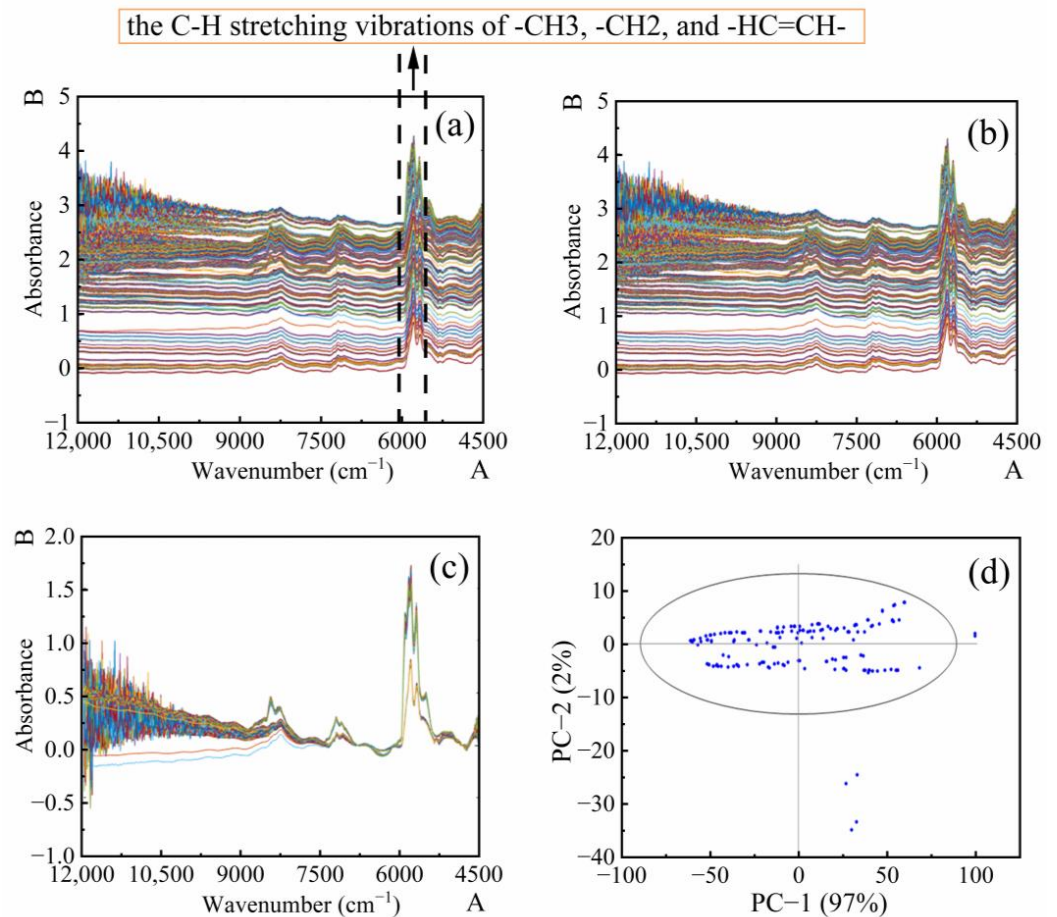
**Figure 4.** Effects of different preprocessing methods for paraffins.

As shown in Figure 4, the *RMSEP* and *MARE* values of all pretreatment methods were lower than 0.8 and 0.6, respectively. The *RMSEP* value of SNV-MASL was the highest, while the *MARE* value was relatively small. The *RMSEP* value of MMN was relatively small but the *MARE* value was particularly high. SNV and MSC had similar effects. Among the five methods, *RMSEP* and *MARE* values of DWT were the smallest; considering *RMSEP* and *MARE* comprehensively, the effect of DWT was best compared with other methods. Thus, DWT was chosen as the method for the pretreatment of NIR spectral data.

Over 4000 spectra were collected. The mean NIR spectra of the HBO samples with different properties are shown in Figure 5. The spectral curves of all samples clearly showed a similar tendency for both A and B orientations with no abnormal variation trend, suggesting accurately discriminating the properties of HBO was challenging. The line of different color represented different samples. In general, five obvious absorption peaks appeared at  $8300\text{ cm}^{-1}$ ,  $7350\text{ cm}^{-1}$ ,  $5830\text{ cm}^{-1}$ ,  $5765\text{ cm}^{-1}$ , and  $5658\text{ cm}^{-1}$ . Specifically, the wavelength at around  $5590\text{ cm}^{-1}$  and  $6055\text{ cm}^{-1}$  was the maximum absorption region, which corresponded to the first overtones of the C–H stretching vibrations of  $-\text{CH}_3$ ,  $-\text{CH}_2$ , and  $-\text{HC}=\text{CH}-$ . The peak ranging from  $7200\text{ cm}^{-1}$  to  $7500\text{ cm}^{-1}$  was the result of OH stretching overtones. The small peaks between  $4800\text{ cm}^{-1}$  and  $4550\text{ cm}^{-1}$  were attributed to the C=C and C–H stretching combination tones of unsaturated fatty acids [47].

In Figure 5a, the noise in the raw spectral data are obvious and the results of denoising absorbance data by DWT are given in Figure 5b. The spectral absorbance preprocessed by DWT with intensity and peak regions did not change, whereas the noises were removed and important data were retained. The pretreated spectra of HBO samples after eliminating baseline drift are given in Figure 5c. In the NIR spectra of HBO samples, the spectral features were enhanced and the baseline variations were weakened. The denoised spectral data were

exported for further preprocessing and development of appropriate prediction models. The important information was mainly contained between  $6780\text{ cm}^{-1}$  and  $4800\text{ cm}^{-1}$ . Therefore, this part of the spectra was considered as the characteristic variables for exploratory purposes to simplify the model.



**Figure 5.** NIR spectra of HBO samples within the range of  $12,000\text{ cm}^{-1}$ – $4500\text{ cm}^{-1}$ . (a) Raw NIR spectral data. (b) Spectra denoised by DWT. (c) Spectra after baseline correction. (d) Score plot from the PCA of the NIR spectra of the HBO samples.

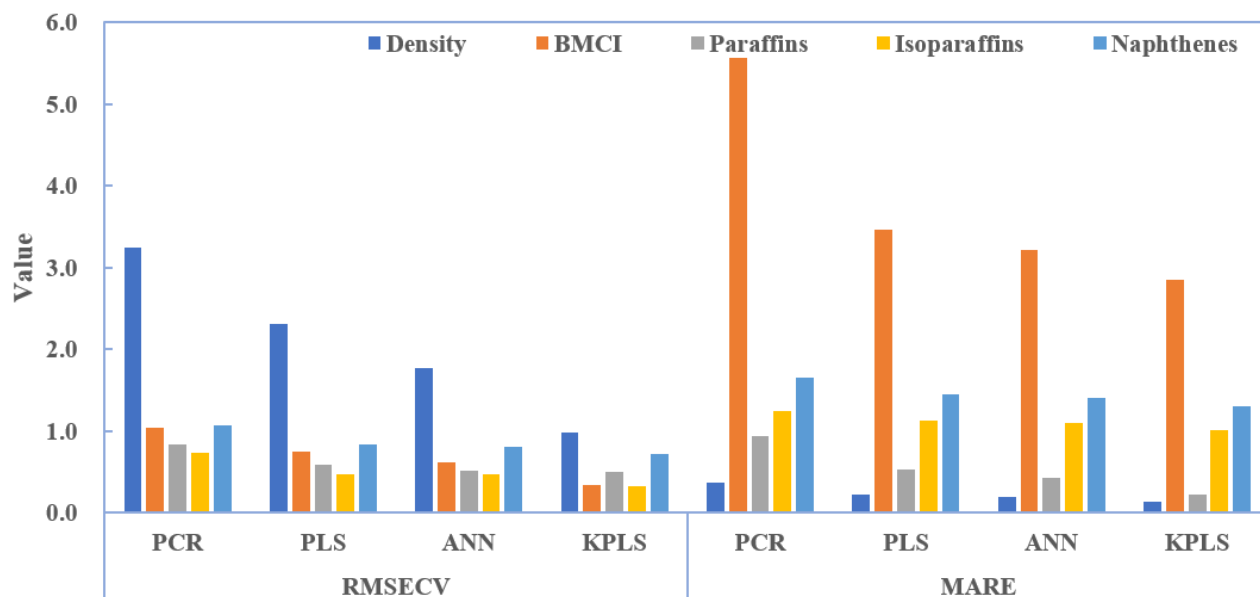
Different types of uncertainty, such as the performance of the spectrophotometer, human uncertainty during the experimental process, and sample preparation, may generate extreme data points [59]. These uncertainties will increase the variance of the spectral data and/or quality parameters being measured [60]. Therefore, extreme points in the data were detected to determine the variance of the spectral data [61].

By combination PCA with the Hotelling's T<sub>2</sub> functions to pretreat the spectral data of HBO samples, extreme data were determined by the 95% confidence limits. Figure 5d clearly shows the explained variances of data (score plot) by PC-1 and PC-2. The first and second components explained 97% and 2% of the variance in the data, respectively. A total of six data points were located outside the 95% confidence interval, i.e., outside the ellipse, so they were determined to be extreme points, of which the Hotelling's T<sub>2</sub> values were higher than 7.07.

### 3.3. NIR Model Development for Predicting HBO Properties

Based on the data of different properties obtained from NIR spectroscopy, over 4000 spectral data points were divided into groups at a ratio of 6:4 into a verification set and a prediction set. The values obtained from the KPLS models were compared with

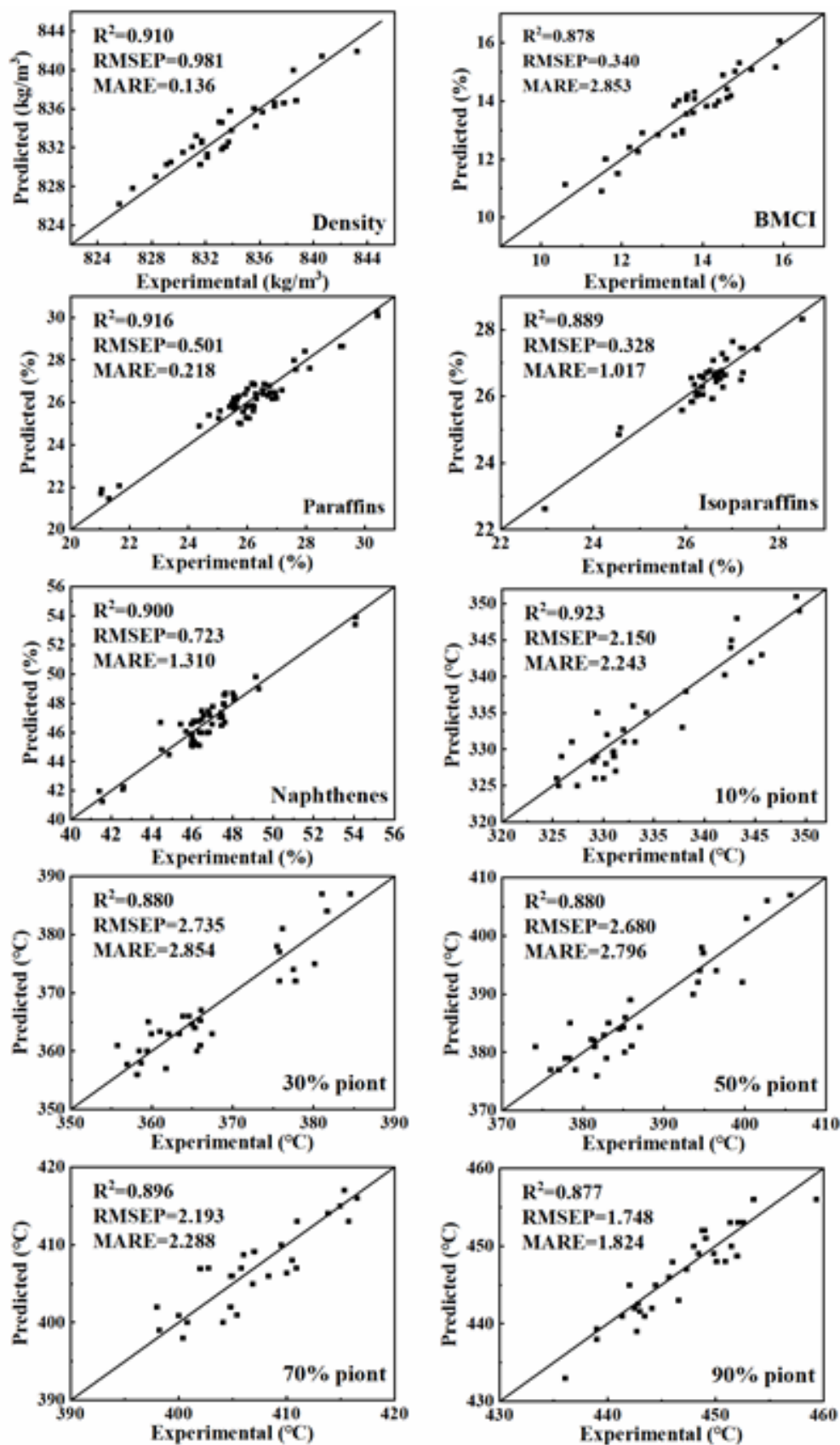
the true values to continuously narrow the gap. Online monitoring was carried in three operating periods. After selecting the preprocessing transformations on the denoised data and trying various approaches, calibration models were developed for the density, *BMCI*, and PINA content of HBO samples. The RMSECV and MARE values obtained by PCR, PLS, ANN, and KPLS are shown in Figure 6.



**Figure 6.** RMSECV and MARE of different methods (PCR, PLS, ANN, and KPLS) in the calibration subset.

As shown in Figure 6, *RMSECV* and *MARE* were basically consistent, indicating that the density, *BMCI*, and hydrocarbon composition of the HBO samples could be quickly analyzed by the four methods. KPLS consistently showed better performance and predicted results than PCR, PLS, and ANN. Its *RMSECV* and *MARE* values were smaller than those of the other models constructed by PCR, PLS, and ANN. In the KPLS model, the *RMSECV* values of density, *BMCI*, and paraffins, isoparaffins, and naphthenes were 0.981, 0.340, 0.501, 0.328, and 0.723, respectively, whereas their *MARE* values were 0.136, 2.853, 0.218, 1.017, and 1.310, respectively. Therefore, the model developed by KPLS for predicting HBO properties was assessed using the prediction set. The KPLS method also showed good prediction performance in modeling the PIONA content in gasoline. This method was used to model the PIONA content in gasoline, and achieved good prediction results [25].

The results of the application of the KPLS model to the NIR spectral data of the HBO samples in the prediction subset are provided in Figure 7. The density, *BMCI*, PIONA content, 10% point, 30% point, 50% point, 70% point, and 90% point of the HBO samples measured by the standard test method were compared. Among them, the highest  $R^2$  value was for the 10% point, the second highest  $R^2$  value was for paraffins, and the smallest  $R^2$  value was for the 90% point, the corresponding value was 0.877. The smallest *RMSEP* was for isoparaffins, the next smallest was the *BMCI*, and the highest *RMSEP* value was for the 10% point, and the corresponding value was 2.150. The smallest *MARE* value was for density, and the next smallest value was for paraffins; the highest *MARE* value was for the 70% point, and the corresponding value was 2.854. The result of the three values of paraffins was relatively good compared to the results of other properties. The *RMSEP* and *MARE* values of the models were less than 3 and their  $R^2$  exceeded 0.87. The low *RMSEP* and *MARE* values for these properties confirmed the models' prediction performance.

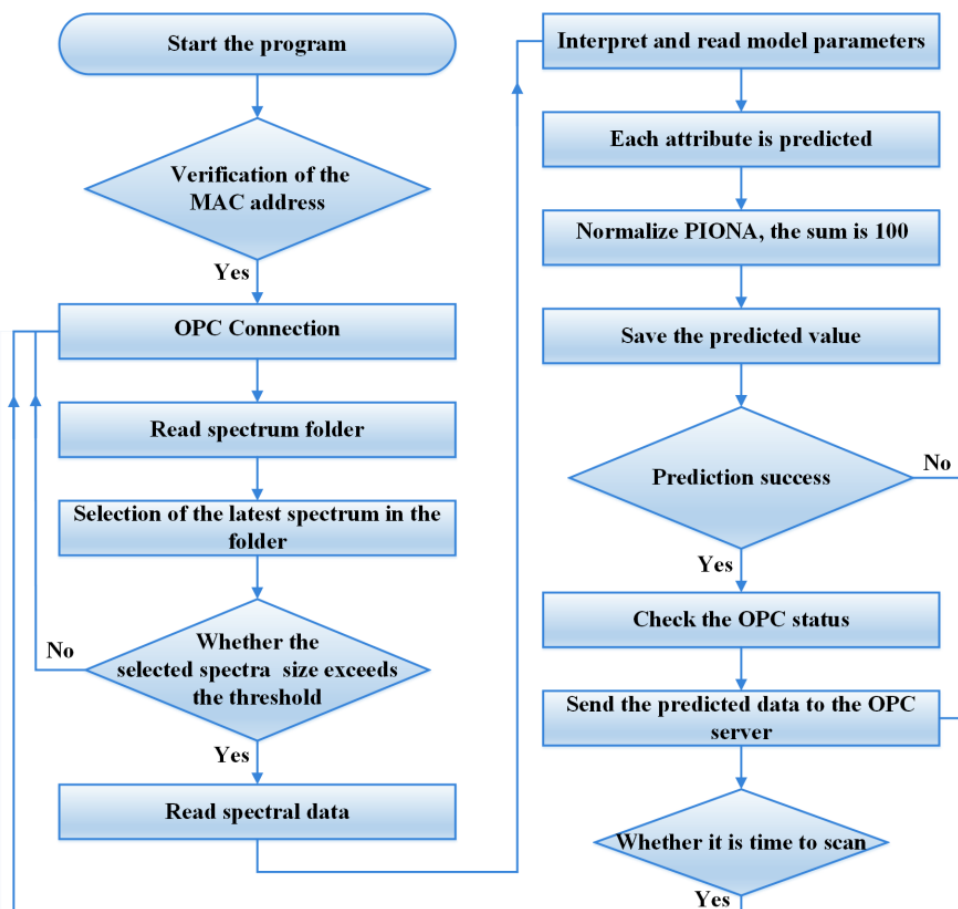


**Figure 7.** Estimation of the values of density, *BMCI*, paraffins, isoparaffins, naphthenes, 10% point, 30% point, 50% point, 70% point, and 90% point by the KPLS model according to the NIR spectra compared with the values determined by reference methods.

### 3.4. Accuracy of Online Detection of HBO Properties

#### 3.4.1. Implementation of Online NIR Detection

An online NIR system was established using an online NIR instrument for analyzing HBO properties under actual production conditions. The system is capable of reading spectra generated by spectral analysis software and using an encapsulated KPLS model for HBO NIR analysis to predict the properties of HBO samples. The model was implemented in C#. The predicted results can be transmitted to the OLE for process control (OPC) server through OPC communication and saved on the local server. A comprehensive program logic flow chart of the NIR analysis system formed herein is shown in Figure 8.

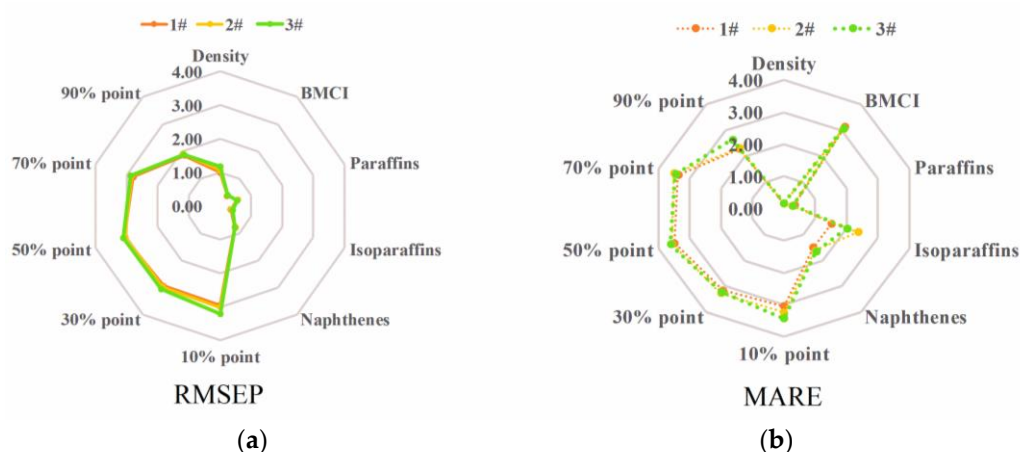


**Figure 8.** Program logic flow chart of the online NIR analysis system.

The program of the NIR analysis system determined whether the media access control (MAC) address of the current server meets the running permission. If yes, after the OPC connection, then the program would locate the latest spectrum in the folder generated by the spectroscopic analysis software based on time. If the file memory size was greater than a given threshold, then the spectrum file would be read, the present spectra pretreatment program and the NIR model parameters could be decoded and read, and the properties of the HBO samples were predicted. The predicted results were saved on the local server and transmitted to the OPC server through the OPC protocol. At this point, the whole program was completed and waited until the scan cycle finished the next round of prediction. In the long-term online prediction process, if the error was unacceptable, the spectrum pretreatment program and the quantitative model parameters were updated offline according to the method described in this paper. These updated spectroscopic preprocessing programs and NIR models were given into the online NIR system to maintain the accuracy of HBO properties online measurements.

### 3.4.2. Online Detection Accuracy of Three Run Periods

The predictive performance of the prediction model on the component data was evaluated using the performance indicators described above. For online NIR prediction of petroleum properties, the absolute errors between their predicted values and laboratory analysis values were always expected to be small and within a certain range. Moreover, the stability and accuracy of prediction models were normally required in the long-term applications of the refinery production processes. In this study, the normal operation of the online NIR detection system from the start to the cleaning of the fiber optic probe was recorded as an operating cycle (about 1 month). A small number of samples (one sample every three days) were collected in every period to verify the online predictive performance of the aforementioned NIR models using RMSEP and MARE values between the online predicted values and the experimental values. Over 30 samples were obtained in three run periods. The online predictive accuracy of density, *BMCI*, paraffins, isoparaffins, and naphthenes of the HBO samples in the three run periods is shown in Figure 9.



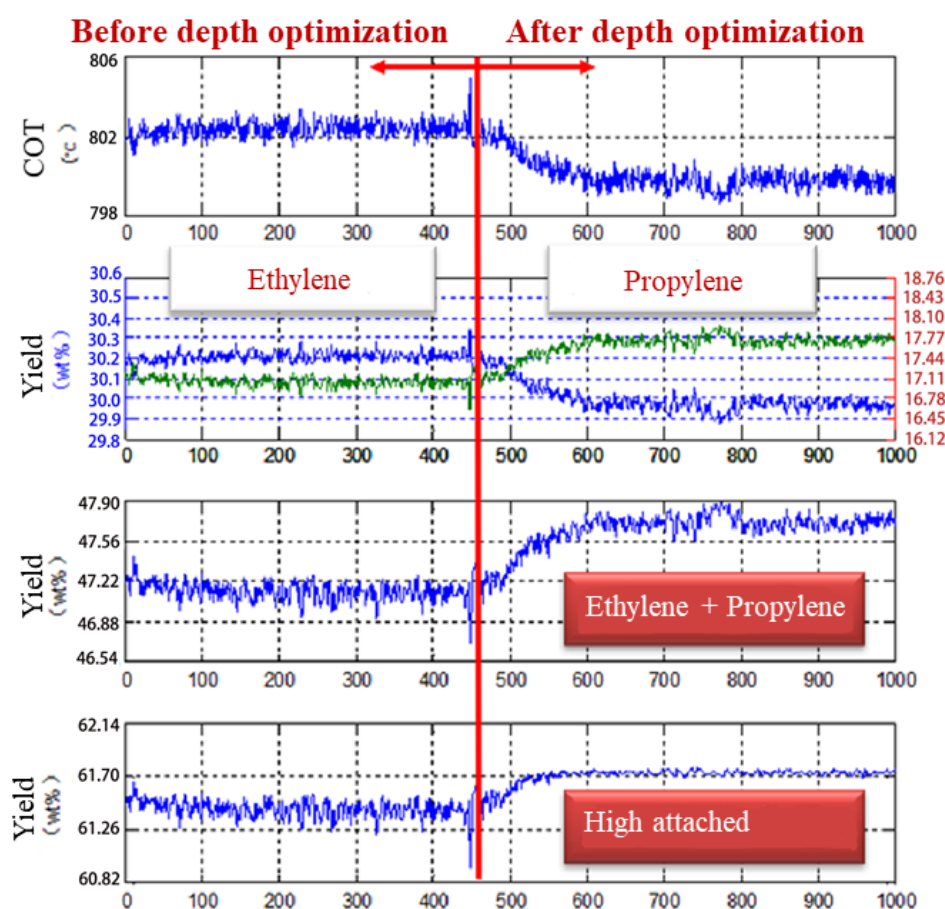
**Figure 9.** Online prediction accuracy of HBO samples in three run periods. (a) RMSEP of density, *BMCI*, paraffins, isoparaffins, naphthenes, 10% point, 30% point, 50% point, 70% point, and 90% point. (b) MARE of 10% point, 30% point, 50% point, 70% point, and 90% point.

As shown in Figure 9, the ranges of the RMSEP values of density, *BMCI*, paraffins, isoparaffins, naphthenes, 10% point, 30% point, 50% point, 70% point, and 90% point of most samples were  $-0.99$ – $1.16$   $\text{kg}/\text{m}^3$ ,  $0.37$ – $0.40$   $\text{kg}/\text{m}^3$ ,  $0.54$ – $0.58$   $\text{kg}/\text{m}^3$ ,  $0.35$ – $0.40$   $\text{kg}/\text{m}^3$ ,  $0.77$ – $0.80$   $\text{kg}/\text{m}^3$ ,  $2.95$ – $3.20$   $^{\circ}\text{C}$ ,  $2.94$ – $3.06$   $^{\circ}\text{C}$ ,  $3.05$ – $3.10$   $^{\circ}\text{C}$ ,  $2.79$ – $2.88$   $^{\circ}\text{C}$ , and  $1.85$ – $1.90$   $^{\circ}\text{C}$ , respectively. All RMSEP values of the samples' properties were less than 4 in the three run periods. The MARE values of the density, *BMCI*, paraffins, isoparaffins, naphthenes, 10% point, 30% point, 50% point, 70% point, and 90% point of these samples were all less than 4.0%. It can be seen that NIR was very effective for chemical or physical parameters. Using the *BMCI* for example, the value of the relevant *BMCI* was calculated indirectly from Equation (1). The value of the *BMCI* is correlated with the mean boiling point ( $T$ ) and density. HBO is rich in alkanes, a certain amount of cycloalkanes, and a small amount of aromatics, and its sulfur and nitrogen contents were low. Hydrocarbons contain C–H and C–C, whose molecular vibrations affect the yield of ethylene cracking. Thus, the online predictive effect of the NIR model did not significantly attenuate in the three run periods, which was acceptable for online prediction of HBO properties.

Overall, the results obtained were due to the combination of DWT with other pre-processing functions. DWT was used in denoising the raw spectral data before creating a regression model. The predictive performance of the KPLS model developed herein for the online estimation of the density; the *BMCI*; and contents of paraffins, isoparaffins, naphthenes, 10% point, 30% point, 50% point, 70% point, and 90% point of the HBO samples had relatively acceptable regression parameters.

### 3.5. Effects of Online NIR Measurement System Assisted Online Cracking Depth Optimization

The cracking depth optimization control was implemented in a 0.36 Mt/year HBO cracker (SL-IV mode). The optimization results of the COT and the yield of key products, such as ethylene, propylene, and high value-added products, are shown in Figure 10. As can be seen, after optimization, the COT control value decreased by 2.5 °C from 802.4 °C to 799.9 °C, indicating a decrease in energy consumption of the cracking furnace decreases. Ethylene yield decreased by 0.22 percentage points from 30.18% to 29.97%, whereas the yield of propylene increased by 0.74 percentage points from 16.96% to 17.70%. The sum yield of ethylene and propylene increased from 47.14% to 47.67% by 0.53 percentage points. The yield of high value-added products (H<sub>2</sub>, ethylene, propylene, butadiene, benzene) increased by 0.26 percentage points from 61.42% to 61.68%. It can be seen that the depth optimization method integrated with online NIR measurement of HBO can effectively improve the yield of the target products and bring significant economic benefits.



**Figure 10.** Effects of real-time optimization of cracking depth by combination of online analysis of ethylene cracking feed.

## 4. Conclusions

In this study, a novel approach combining DWT, PCA, Hotelling's T<sub>2</sub> test, and the KPLS algorithm was proposed for online prediction of HBO samples characteristics using NIR spectroscopy. The characteristic distribution of the HBO samples investigated herein had a broad range. The noise in the spectral data was removed using DWT. The spectral data of the HBO samples were pretreated by combining PCA with the Hotelling's T<sub>2</sub> functions, and extreme data were identified by the 95% confidence limits. The prediction results of the HBO calibration set samples indicated that the KPLS method consistently performed better and had better results than PCR, PLS, and ANN, as demonstrated by the distinctly different values of RMSECV and MARE. The RMSEP, MARE, and R<sup>2</sup> values for density,



*BMCI*, paraffin, isoparaffin, naphthenes, and the 10% point, 30% point, 50% point, 70% point, and 90% point obtained from the KPLS model were in good agreement with the data obtained from the standard test methods, indicating that they were suitable for prediction subset and online measurement. The NIR method proposed for the analysis of HBO chemical composition was faster than GC-MS and other standard methods. The NIR models constructed were utilized for the online measurement of HBO properties in an industrial ethylene plant. The online measurement of quality parameters of HBO through NIR spectroscopy, along with the proposed practical online characterization of HBO properties, has enhanced the real-time optimization of the ethylene cracking process. The improvement in the yield of target products and the reduction in energy consumption had promoted the economic benefits of the ethylene cracking process. Moreover, applications of integration of real-time detection and real-time optimization in plants and enterprises can boost the industrial reform from automation to digitization and intelligence.

**Author Contributions:** Conceptualization, and funding acquisition, C.F.; writing—original draft and funding acquisition, J.L.; writing—review and editing, T.X. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work was supported by National Key Research and Development Program of China (2022YFB3305903), National Natural Science Foundation of China (Key Program: 62136003), National Natural Science Foundation of China (61973124), Shanghai Committee of Science and Technology (Grant No.22DZ1101500) and Fundamental Research Funds for the Central Universities.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

### Acronyms and Parameters

ANN	artificial neural network
ASTM	American Society for Testing and Materials
<i>BMCI</i>	Bureau of Mines correlation index
DWT	discrete wavelet transform
GC	gas chromatograph
GC-MS	gas chromatograph-mass spectrometry
HBO	hydrocracking bottom oil
HCR	hydrocracking unit
LVs	latent variables
MARE	the mean absolute relative error
MLR	multiple linear regression
MMN	min-max normalization
MS	mass spectrometry
MSC	multiplicative scattering correction
NIR	near-infrared
PCA	principal component analysis
PCR	principal component regression
PC-1	the first PC
PC-2	the second PC
PCs	principal components
PLS	partial least squares regression
PIONA	paraffins, isoparaffins, naphthenes, olefins, and aromatics
RMSECV	root mean squared error of cross validations
RMSEP	the root mean squared error of prediction set
SNV	standard normalized vector
SNV-MASL	SNV-minus a straight line
WT	wavelet transform
$d_{15.6}^{15.6}$	specific gravity at 15.6 °C (g/cm <sup>3</sup> )
<b>E</b>	residual matrix

$F_{1,f,\alpha}$	critical value derived from the Fisher distribution
$i$	a given sample
$j$	dimension
$\mathbf{K}$	kernel function
$\tilde{\mathbf{K}}$	deflation of kernel function
$k$	integer value
$m$	row of samples
$N_i$	sample size
$n$	number of calibration/prediction set samples
$\mathbf{p}_i^T$	loading vector
$S_i$	standard deviation of each sample
$s_{ij}$	independent normally distributed samples
$\bar{s}_i$	mean of each sample
$S_p^2$	pooled variance
$T$	average boiling point (°C)
$\mathbf{T}$	score matrix
$\mathbf{t}_i$	score vector
$t_{f,\alpha/2}$	critical value derived from the samples' t-distribution
$\mathbf{U}$	score matrix
$\mathbf{X}$	spectral data matrix
$\mathbf{X}$	the NIR spectral peak height data matrix
$X(t)$	the function transferred by WT
$\mathbf{x}_k$	oil samples
$x_i$	NIR absorbance at the i-th wavelength
$\bar{x}_i$	mean of each sample
$\mathbf{Y}$	the corresponding property matrix
$\mathbf{y}_k$	oil samples
$\hat{y}_i$	the j-th estimated property
$y_{i,pre}$	predicted value
$y_{i,exp}$	experimental value
$\bar{y}_{i,pre}$	the average of the response variable
$\psi$	mother function
$\Phi$	a nonlinear transformation
$\sigma$	kernel parameter

## References

- Al-Jamimi, H.A.; BinMakhashen, G.M.; Deb, K.; Saleh, T.A. Multiobjective optimization and analysis of petroleum refinery catalytic processes, A review. *Fuel* **2021**, *288*, 119678. [\[CrossRef\]](#)
- Gu, H.; Li, J.; Mu, P.; Zhu, Q. Improving the Operational Efficiency of Ethylene Cracking Integrated with Refining by Molecular Management. *Ind. Eng. Chem. Res.* **2020**, *59*, 13160–13174. [\[CrossRef\]](#)
- Browning, B.; Alvarez, P.; Jansen, T.; Lacroix, M.; Geantet, C.; Tayakout-Fayolle, M. A Review of Thermal Cracking, Hydrocracking, and Slurry Phase Hydroconversion Kinetic Parameters in Lumped Models for Upgrading Heavy Oils. *Energy Fuels* **2021**, *35*, 15360–15380. [\[CrossRef\]](#)
- Liu, J.; Fang, X.; Yang, T. Novel Ebullated Bed Residue Hydrocracking Process. *Energy Fuels* **2017**, *31*, 6568–6579. [\[CrossRef\]](#)
- Alawani, N.A.; Muller, H.; Panda, S.K.; Al-Hajji, A.A.; Koseoglu, O.R. Evaluation of Polycyclic Aromatic Hydrocarbon Removal from Hydrocracking Recycle Streams. *Energy Fuels* **2020**, *34*, 179–187. [\[CrossRef\]](#)
- Upreti, N.; Pakala, H.; Surasani, V.K.; Dinda, S. Modeling of thermal cracking reaction of kerosene range hydrocarbons. *Mater. Today Proc.* **2022**, *59*, 430–436. [\[CrossRef\]](#)
- Bai, Y.; Zhang, G.; Liu, D.; Zhang, Y.; Zhao, L.; Gao, J.; Xu, C.; Meng, Q.; Gao, X. The advance in catalytic pyrolysis of naphtha technology using ZSM-5 as catalyst. *Appl. Catal. A Gen.* **2021**, *628*, 118399. [\[CrossRef\]](#)
- Fakhroleslam, M.; Sadrameli, S.M. Thermal/catalytic cracking of hydrocarbons for the production of olefins: A state-of-the-art review III, Process modeling and simulation. *Fuel* **2019**, *252*, 553–566. [\[CrossRef\]](#)
- Meng, D.; Shao, C.; Zhu, L. Two-level comprehensive energy-efficiency quantitative diagnosis scheme for ethylene-cracking furnace with multi-working-condition of fault and exception operation. *Energy* **2022**, *239*, 121835. [\[CrossRef\]](#)
- Moro, M.K.; dos Santos, F.D.; Folli, G.S.; Romão, W.; Filgueiras, P.R. A review of chemometrics models to predict crude oil properties from nuclear magnetic resonance and infrared spectroscopy. *Fuel* **2021**, *303*, 121283. [\[CrossRef\]](#)

11. Jurinjak Tušek, A.; Jurina, T.; Čulo, I.; Valinger, D.; Gajdoš Kljusurić, J.; Benković, M. Application of NIRs coupled with PLS and ANN modelling to predict average droplet size in oil-in-water emulsions prepared with different microfluidic devices. *Spectrochim. Acta Part A* **2022**, *270*, 120860. [[CrossRef](#)] [[PubMed](#)]
12. Wang, K.; Du, W.; Long, J. Near-Infrared Wavelength-Selection Method Based on Joint Mutual Information and Weighted Bootstrap Sampling. *IEEE Trans. Ind. Inform.* **2020**, *16*, 5884–5894. [[CrossRef](#)]
13. He, K.; Zhong, M.; Du, W. Weighted incremental minimax probability machine-based method for quality prediction in gasoline blending process. *Chemom. Intell. Lab. Syst.* **2020**, *196*, 103909. [[CrossRef](#)]
14. Zhao, N.F.; Xu, Q.S.; Tang, M.L.; Wang, H. Variable Screening for Near Infrared (NIR) Spectroscopy Data Based on Ridge Partial Least Squares Regression. *Comb. Chem. High Throughput Screen.* **2020**, *23*, 740–756. [[CrossRef](#)]
15. Du, C.; Sun, L.; Bai, H.; Liu, Y.; Yang, J.; Wang, X. Quantitative detection of azodicarbonamide in wheat flour by near-infrared spectroscopy based on two-step feature selection. *Chemom. Intell. Lab. Syst.* **2021**, *219*, 104445. [[CrossRef](#)]
16. Varghese, G.; Saeed, K.; Rutt, K.J. Determination of the oxidative stability of biodiesel fuels by near-infrared spectroscopy. *Fuel* **2021**, *290*, 120015. [[CrossRef](#)]
17. Liu, S.; Wang, S.; Hu, C.; Zhan, S.; Kong, D.; Wang, J. Rapid and accurate determination of diesel multiple properties through NIR data analysis assisted by machine learning. *Spectrochim. Acta Part A* **2022**, *277*, 121261. [[CrossRef](#)] [[PubMed](#)]
18. 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*, 118854. [[CrossRef](#)]
19. Joshi, N.B.; Mullins, O.C.; Jamaluddin, A.; Creek, J.; McFadden, J. Asphaltene Precipitation from Live Crude Oil. *Energy Fuels* **2001**, *15*, 979–986. [[CrossRef](#)]
20. Barreira, F.R.; Reis, L.G.; Nunes, R.d.C.P.; Filipakis, S.D.; Lucas, E.F. Asphaltenes Precipitation Onset, Influence of the Addition of a Second Crude Oil or Its Asphaltenes Fractions (C3I and C5I). *Energy Fuels* **2018**, *32*, 10391–10397. [[CrossRef](#)]
21. Wang, K.; He, K.; Du, W.; Long, J. Novel adaptive sample space expansion approach of NIR model for in-situ measurement of gasoline octane number in online gasoline blending processes. *Chem. Eng. Sci.* **2021**, *242*, 116672. [[CrossRef](#)]
22. Yu, H.; Du, W.; Lang, Z.Q.; Wang, K.; Long, J. A Novel Integrated Approach to Characterization of Petroleum Naphtha Properties From Near-Infrared Spectroscopy. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–13. [[CrossRef](#)]
23. Long, J.; Wang, K.; Yang, M.; Zhong, W. Rapid crude oil analysis using near-infrared reflectance spectroscopy. *Pet. Sci. Technol.* **2019**, *37*, 354–360. [[CrossRef](#)]
24. Giovenzana, V.; Beghi, R.; Tugnolo, A.; Brancadoro, L.; Guidetti, R. Comparison of two immersion probes coupled with visible/near infrared spectroscopy to assess the must infection at the grape receiving area. *Comput. Electron. Agric.* **2018**, *146*, 86–92. [[CrossRef](#)]
25. Li, J.; Chu, X. Rapid Determination of Physical and Chemical Parameters of Reformed Gasoline by Near-Infrared (NIR) Spectroscopy Combined with the Monte Carlo Virtual Spectrum Identification Method. *Energy Fuels* **2018**, *32*, 12013–12020. [[CrossRef](#)]
26. Long, J.; Jiang, S.; He, R.; Zhao, L. Diesel blending under property uncertainty: A data-driven robust optimization approach. *Fuel* **2021**, *306*, 121647. [[CrossRef](#)]
27. 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 <sup>1</sup>H NMR and NIR. *Energy Fuels* **2020**, *34*, 12173–12181. [[CrossRef](#)]
28. Zhao, N.; Ma, L.; Wang, K.; Zhang, F.; Li, M.; Liu, X.; Zhu, M.; Lu, Y.; Song, X.; Yan, H.; et al. NIR robustness model of variable selection investigation of critical quality attributes coupled with different simulate noises by prediction capability and reproducibility. *Spectrochim. Acta Part A* **2022**, *267*, 120522. [[CrossRef](#)]
29. Shi, H.; Yu, P. Comparison of grating-based near-infrared (NIR) and Fourier transform mid-infrared (ATR-FT/MIR) spectroscopy based on spectral preprocessing and wavelength selection for the determination of crude protein and moisture content in wheat. *Food Control* **2017**, *82*, 57–65. [[CrossRef](#)]
30. Dotto, A.C.; Dalmolin, R.S.D.; Grunwald, S.; ten Caten, A.; Pereira Filho, W. Two preprocessing techniques to reduce model covariables in soil property predictions by Vis-NIR spectroscopy. *Soil Tillage Res.* **2017**, *172*, 59–68. [[CrossRef](#)]
31. Chu, X.-L.; Xu, Y.-P.; Tian, S.-B.; Wang, J.; Lu, W.-Z. Rapid identification and assay of crude oils based on moving-window correlation coefficient and near infrared spectral library. *Chemom. Intell. Lab. Syst.* **2011**, *107*, 44–49. [[CrossRef](#)]
32. Cai, C.-B.; Xu, L.; Han, Q.-J.; Wu, H.-L.; Nie, J.-F.; Fu, H.-Y.; Yu, R.-Q. Combining the least correlation design, wavelet packet transform and correlation coefficient test to reduce the size of calibration set for NIR quantitative analysis in multi-component systems. *Talanta* **2010**, *81*, 799–804. [[CrossRef](#)] [[PubMed](#)]
33. Liu, H.; Zhou, W.; Ying, D.; Xiang, B.; Li, P. Generalized two-dimensional correlation NIR spectroscopy analysis of the structures on n-propanol and n-butanol. *J. Mol. Struct.* **2019**, *1181*, 436–443. [[CrossRef](#)]
34. Cruz-Tirado, J.P.; Oliveira, M.; de Jesus Filho, M.; Godoy, H.T.; Amigo, J.M.; Barbin, D.F. Shelf life estimation and kinetic degradation modeling of chia seeds (*Salvia hispanica*) using principal component analysis based on NIR-hyperspectral imaging. *Food Control* **2021**, *123*, 107777. [[CrossRef](#)]
35. Bobba, S.; Zinfolino, N.; Fissore, D. Application of Near-Infrared Spectroscopy to statistical control in freeze-drying processes. *Eur. J. Pharm. Biopharm.* **2021**, *168*, 26–37. [[CrossRef](#)] [[PubMed](#)]
36. Cui, C.; Fearn, T. Hierarchical mixture of linear regressions for multivariate spectroscopic calibration: An application for NIR calibration. *Chemom. Intell. Lab. Syst.* **2018**, *174*, 1–14. [[CrossRef](#)]

37. Reda, R.; Saffaj, T.; Derrouz, H.; Itqiq, S.E.; Bouzida, I.; Saidi, O.; Lakssir, B.; El Hadrami, E.M. Comparing CalReg performance with other multivariate methods for estimating selected soil properties from Moroccan agricultural regions using NIR spectroscopy. *Chemom. Intell. Lab. Syst.* **2021**, *211*, 104277. [CrossRef]
38. Franzoi, M.; Ghetti, M.; Di Monte, L.; De Marchi, M. Investigation of weight loss in mozzarella cheese using NIR predicted chemical composition and multivariate analysis. *J. Food Compos. Anal.* **2021**, *102*, 104002. [CrossRef]
39. Lei, T.; Sun, D.-W. A novel NIR spectral calibration method: Sparse coefficients wavelength selection and regression (SCWR). *Anal. Chim. Acta* **2020**, *1110*, 169–180. [CrossRef]
40. Fung Shek, C.; Betenbaugh, M. Taking the pulse of bioprocesses: At-line and in-line monitoring of mammalian cell cultures. *Curr. Opin. Biotechnol.* **2021**, *71*, 191–197. [CrossRef]
41. Labbé, N.; Lee, S.-H.; Cho, H.-W.; Jeong, M.K.; André, N. Enhanced discrimination and calibration of biomass NIR spectral data using non-linear kernel methods. *Bioresour. Technol.* **2008**, *99*, 8445–8452. [CrossRef] [PubMed]
42. Binetti, G.; Del Coco, L.; Ragone, R.; Zelasco, S.; Perri, E.; Montemurro, C.; Valentini, R.; Naso, D.; Fanizzi, F.P.; Schena, F.P. Cultivar classification of Apulian olive oils: Use of artificial neural networks for comparing NMR, NIR and merceological data. *Food Chem.* **2017**, *219*, 131–138. [CrossRef] [PubMed]
43. Pan, H.; Li, X.; Shen, Y.; Wu, X.; Ju, F.; Li, Y.; Wu, G.; Ai, B.; Xu, B.; Ling, H. Novel design of lubricant-type vacuum distillation process for lube base oils production from hydrocracking tail oil. *Chin. J. Chem. Eng.* **2022**, *45*, 121–132. [CrossRef]
44. Campos, M.I.; Antolin, G.; Debán, L.; Pardo, R. Assessing the influence of temperature on NIRS prediction models for the determination of sodium content in dry-cured ham slices. *Food Chem.* **2018**, *257*, 237–242. [CrossRef] [PubMed]
45. Zhong, Y.; Gu, J.; Su, Y.; Zhao, L.; Zhou, Y.; Peng, J. Real-time screening of hepatotoxins in natural medicine by peroxy-nitrite responsive lanthanide-based NIR-II luminescent probes. *Chem. Eng. J.* **2022**, *433*, 133263. [CrossRef]
46. Standard Test Method for Hydrocarbon Types Analysis of Gas-Oil Saturates Fractions by High Ionizing Voltage Mass Spectrometry. Available online: <https://docin.com/p-1271811601.html> (accessed on 27 August 2015).
47. Zhang, K.; Tan, Z.; Chen, C.; Sun, X.S.; Wang, D. Rapid Prediction of Camelina Seed Oil Content Using Near-Infrared Spectroscopy. *Energy Fuels* **2017**, *31*, 5629–5634. [CrossRef]
48. Abasi, S.; Minaei, S.; Jamshidi, B.; Fathi, D.; Khoshtaghaza, M.H. Rapid measurement of apple quality parameters using wavelet de-noising transform with Vis/NIR analysis. *Sci. Hortic.* **2019**, *252*, 7–13. [CrossRef]
49. Jamei, M.; Ahmadianfar, I.; Chu, X.; Yaseen, Z.M. Prediction of surface water total dissolved solids using hybridized wavelet-multigene genetic programming: New approach. *J. Hydrol.* **2020**, *589*, 125335. [CrossRef]
50. Zucatelli, P.J.; Nascimento, E.G.S.; Santos, A.Á.B.; Arce, A.M.G.; Moreira, D.M. An investigation on deep learning and wavelet transform to nowcast wind power and wind power ramp: A case study in Brazil and Uruguay. *Energy* **2021**, *230*, 120842. [CrossRef]
51. Long, J.; Li, T.; Yang, M.; Hu, G.; Zhong, W. Hybrid Strategy Integrating Variable Selection and a Neural Network for Fluid Catalytic Cracking Modeling. *Ind. Eng. Chem. Res.* **2019**, *58*, 247–258. [CrossRef]
52. Hotelling, H. Analysis of a complex of statistical variables into principal components. *J. Educ. Psychol.* **1933**, *24*, 417. [CrossRef]
53. Glazko, G.V.; Emmert-Streib, F. Unite and conquer: Univariate and multivariate approaches for finding differentially expressed gene sets. *Bioinformatics* **2009**, *25*, 2348–2354. [CrossRef] [PubMed]
54. Sadeghi, Z.; Mashhadi-Hossainali, M. Application of the T2-Hotelling test for investigating ionospheric anomalies before large earthquakes. *J. Atmos. Sol. Terr. Phys.* **2019**, *185*, 7–21. [CrossRef]
55. Hosani, E.A.; Meribout, M.; Al-Durra, A.; Al-Wahedi, K.; Teniou, S. A New Optical-Based Device for Online Black Powder Detection in Gas Pipelines. *IEEE Trans. Instrum. Meas.* **2014**, *63*, 2238–2252. [CrossRef]
56. Haghi, R.K.; Yang, J.; Tohidi, B. Fourier Transform Near-Infrared (FTNIR) Spectroscopy and Partial Least-Squares (PLS) Algorithm for Monitoring Compositional Changes in Hydrocarbon Gases under In Situ Pressure. *Energy Fuels* **2017**, *31*, 10245–10259. [CrossRef]
57. Zhu, Q.; Zhao, Z.; Liu, F. Developing new products with kernel partial least squares model inversion. *Comput. Chem. Eng.* **2021**, *155*, 107537. [CrossRef]
58. Rohani, A.; Taki, M.; Abdollahpour, M. A novel soft computing model (Gaussian process regression with K-fold cross validation) for daily and monthly solar radiation forecasting (Part I). *Renew. Energy* **2018**, *115*, 411–422. [CrossRef]
59. Toscano, G.; Rinnan, Å.; Pizzi, A.; Mancini, M. The Use of Near-Infrared (NIR) Spectroscopy and Principal Component Analysis (PCA) To Discriminate Bark and Wood of the Most Common Species of the Pellet Sector. *Energy Fuels* **2017**, *31*, 2814–2821. [CrossRef]
60. Heidari, P.; Rezaei, M.; Sahebi, M.; Khadivi, A. Phenotypic variability of *Pyrus boissieriana* Buhse: Implications for conservation and breeding. *Sci. Hortic.* **2019**, *247*, 1–8. [CrossRef]
61. Jamshidi, B.; Mohajerani, E.; Farazmand, H.; Mahmoudi, A.; Hemmati, A. Pattern recognition-based optical technique for non-destructive detection of *Ectomyelois ceratoniae* infestation in pomegranates during hidden activity of the larvae. *Spectrochim. Acta Part A* **2019**, *206*, 552–557. [CrossRef]

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.