

Article

# A Novel Hybrid Strategy for Detecting COD in Surface Water

Guiping Zhang, Qiaoling Du \*, Xinpo Lu and Yankai Wang

State Key Laboratory on Integrated Optoelectronics, College of Electronic Science and Engineering, Jilin University, 2699 Qianjin Street, Changchun 130012, China; zhanggp18@mails.jlu.edu.cn (G.Z.); luxp19@mails.jlu.edu.cn (X.L.); wangyk19@mails.jlu.edu.cn (Y.W.)

\* Correspondence: duql@jlu.edu.cn

Received: 12 November 2020; Accepted: 6 December 2020; Published: 9 December 2020



**Abstract:** The prediction of chemical oxygen demand (COD) by ultraviolet–visible absorption spectrum is a common method. Many researchers use the absorbance at the characteristic wavelength to establish COD prediction models. However, selecting the characteristic wavelength is a problem. In this paper, the extreme values of absorption spectrum change rate, was proposed as a new characteristic parameter to determine the characteristic wavelengths. On this basis, a novel hybrid strategy for detecting COD in surface water was proposed. We first proposed to combine the first derivative method with the permutation entropy method (FDPE) to determine the characteristic wavelengths. Then we used partial least square (PLS) to establish a COD prediction model. Experimental results demonstrated the linear correlation coefficient ( $R^2$ ) of the FDPE\_PLS was above 0.99 without turbidity interference. Secondly, a dual-wavelength method (DWM) was proposed to determine the turbidity values. The DWM used slopes of absorbance values at 400 nm and 600 nm to predict the turbidity values. Compared with the single-wavelength method, the DWM improves the measurement accuracy of turbidity. Finally, a new turbidity compensation method was proposed to compensate for the interference in the first derivative spectrum. After compensation, FDPE\_PLS can predict COD concentrations accurately, whose  $R^2$  was 0.99.

**Keywords:** COD; first derivative; permutation entropy; partial least squares; turbidity compensation

## 1. Introduction

Surface water is an important source of water, providing most of our basic water needs. However, surface water pollution issues have been very serious in recent years. Surface water pollution, to a great extent, damages the ecological environment and directly affects people's health [1]. The chemical oxygen demand (COD) in surface water is an important indicator of the degree of surface water pollution, which can reflect the level of oxygen-consuming organic pollutants in surface water. The determination of COD is particularly important in the analysis of water pollution [2].

Currently, the determination of COD in China mainly adopts the national standard chemical method [3]. Although the measurements of the national standard chemical method are accurate, the process is cumbersome and generally requires heating, reaction, and other steps. Other reactants are also required, which causing secondary pollution. This method also requires a long sample transfer time and reaction time, which is not convenient for field use [4]. Ultraviolet–visible (UV–Vis) absorption spectroscopy is a physical method for detecting COD concentration, whose process is simple without secondary pollution. This method can measure COD in real time [5]. Through online detection of COD concentration, the pollution status of surface water can be known in real time, which is of great significance to protect the surface water environment.

When the concentration of COD in water is detected by ultraviolet–visible absorption spectrometry, the absorbance at 254 nm is usually used to obtain the COD measurements, which is called the single-wavelength method [6]. The single-wavelength method is simple, but its stability is poor, as it is easily interfered with, and the measurement range is limited. Wang proposed a new method for selecting different calibration wavelengths based on the COD value to expand the measurement range. However, this method is essentially a single-wavelength method in different measurement ranges, and also needs a turbidity compensation algorithm to remove turbidity interference [7]. The accuracy of the turbidity compensation algorithm inevitably affected the detection accuracy of COD, so high precision was put forward for the turbidity compensation algorithm. Aiming to improve the accuracy of COD detection, the multi-wavelength method was proposed to detect COD. It is important to select appropriate characteristic wavelengths for the multi-wavelength method. Adaptive weighting algorithm (CARS), random frog, and genetic algorithm were used to determine the characteristic wavelengths of COD in aquaculture water by the multi-wavelength method [8]. These methods to extract characteristic wavelengths improve the accuracy of COD measurement in complex water samples. But each time the program runs, the distribution of selected wavelengths may change [9–11]. Therefore, these methods have problems of reliability and stability. In order to eliminate the influence of turbidity in COD detection, turbidity compensation is necessary. The absorbance at 546 nm in UV–Vis absorption spectrum was used for turbidity compensation [12]. Although the absorbance at 546 nm is simple for turbidity compensation, the accuracy of COD detection with compensation is not satisfactory. Hu et al. used the fourth derivative method to remove turbidity interference [13]. The fourth-order derivative method has the advantage of no baseline and reduces the signal strength of COD, resulting in the reduction of signal-to-noise ratio, finally affecting the accuracy of COD detection.

In this paper, a novel hybrid strategy for detecting COD in surface water was proposed. A new characteristic parameter, the extreme value of the absorption spectrum change rate, was used as a characteristic parameter to determine the characteristic wavelengths. A combined measurement method based on the first derivative method, the permutation entropy method, and the partial least squares method (FDPE\_PLS) was proposed. The first derivative method was used to extract the information of the spectral change rate in the original spectral data. The permutation entropy method was used to extract the extremes of the spectral change rate, and the partial least square method was used to establish COD prediction model. With the aim to improve the accuracy of turbidity compensation, a dual-wavelength method (DWM) and a new turbidity compensation method were proposed. Experimental results demonstrate the hybrid strategy can improve the detection accuracy of COD with turbidity interference and has stability.

This paper is organized as follows: In Section 2, we describe materials used in this article and FDPE\_PLS algorithm, DWM, and a new turbidity compensation method. We present the experimental methodology and the results in Section 3. Finally, we draw conclusions and indicate directions for future research in Section 4.

## 2. Materials and Methods

This section introduces the materials and methods used in this paper. The preparation of samples and the measurement of UV–Vis absorption spectrum are introduced. The sample was divided into a calibration set and test set by joint  $x$ – $y$  distance (SPXY). FDPE\_PLS is proposed to detect COD. DWM is proposed to measure turbidity. Finally, a new turbidity compensation method was proposed.

### 2.1. Samples Preparation and the Measurement of UV–Vis Absorbance Spectra

According to the standards for surface water environmental quality in China (GB3838-2002), the water quality of the surface water source of centralized drinking water is divided into 5 categories, and the limit values of COD are 15, 15, 20, 30, and 40 mg/L, respectively. In this paper, COD and turbidity chemical standard solutions were used to prepare water samples. The COD solution (100 mg/L) was prepared by dissolving 0.02125 g potassium hydrogen phthalate in deionized water and diluting to

250 mL. COD standard solutions whose values were 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, and 32 mg/L, and were prepared by appropriate dilution. In addition, the NTU (nephelometric turbidity unit) standard (7027-1984 ISO) was used to determine the water turbidity. Turbidity solutions of 1, 3, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 80, and 100 NTU were prepared by diluting a 400 NTU turbidity solution with deionized water. All samples were scanned on the SHIMADZU (Model No: UV-3600) UV-vis spectrophotometer interfaced to a microcomputer, with deionized water used as a blank in the spectrophotometer reference cell. The resolution of this spectrophotometer was 0.1 nm. The experimental setup is shown in Figure 1, and the optical path length in quartz cells was 10 mm. Each sample was measured 15 times, and the average was calculated as the spectral measurement of the sample. The absorption spectra of the COD samples and the turbidity samples is shown in Figures 2 and 3, respectively.

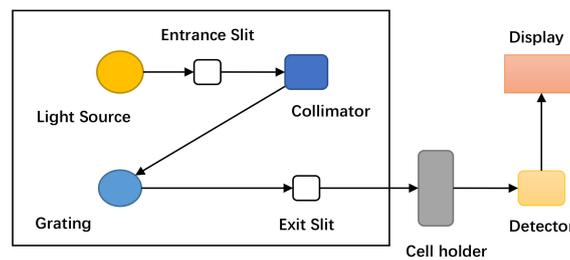


Figure 1. Experimental setup.

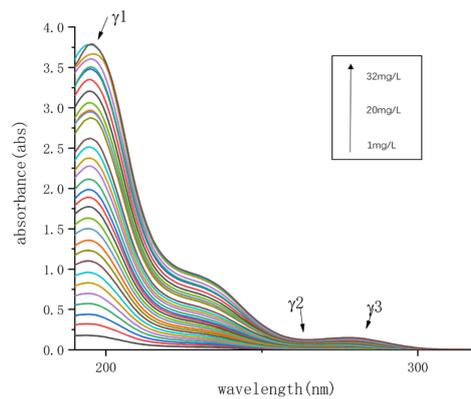


Figure 2. Full spectra of UV absorbance of COD solutions with different concentrations, where  $\gamma_1$  is the first extreme point,  $\gamma_2$  is the second extreme point,  $\gamma_3$  is the third extreme point.

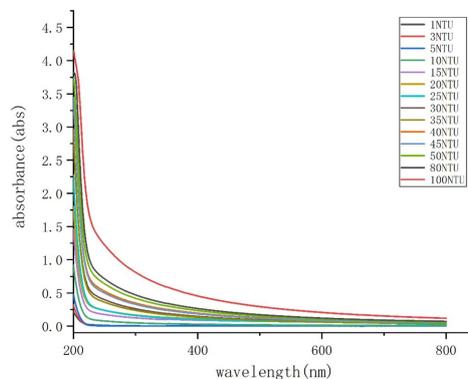


Figure 3. Full spectra of UV-Vis absorbance of turbidity solutions with different turbidity values.

## 2.2. Lambert Beer's Law

When the light passes through the solution, part of the light is absorbed by the light-absorbing substances in the solution. The energy of the light radiation is reduced. When the solution concentration is higher, the optical path of light passing through the solution is longer, and the concentration of light-absorbing material is higher, then the more light is absorbed and the less light passes through the solution [12].

The absorbance is defined as:

$$A(\lambda) = -\lg T = -\lg \frac{I_T}{I_0} = K(\lambda)CL \quad (1)$$

where  $\lambda$  is the wavelength of the incident light,  $A(\lambda)$  is the absorbance,  $T$  is the transmission degree,  $I_T$  represents the transmitted light intensity,  $I_0$  is the incident light intensity when the concentration is 0,  $L$  is the optical path,  $K(\lambda)$  is the absorption coefficient, and  $C$  is the concentration of the tested sample.

For a specific flow-through cell, its absorption optical path  $L$  remains unchanged. For specific measured wavelengths as well as specific water samples, the absorption constants  $K(\lambda)$  are constants. In this case  $A(\lambda)$  is proportional to the sample concentration. The COD concentration in water samples can be measured by detecting the absorbance of organic compounds.

## 2.3. Division of Water Samples

A suitable sample selection method of the calibration set can enhance the predictive ability of established models. The sample-set portioning based on joint  $x$ - $y$  distance (SPXY) realizes a scientific division of the samples by considering the difference of the spectrum and the concentration of the component of the sample. It has the merit of improving the prediction performance of the model effectively [14].

In this paper, the total of 32 water samples were divided into two sets by SPXY algorithm. Twenty-four samples were classified in the calibration set for modeling, and the other 8 samples were placed into the validation set for testing. The statistical characteristics of the samples were shown in Table 1.

**Table 1.** Statistical results of COD value by SPXY.

Sample Set	Samples	Mean (mg/L)	Minimum (mg/L)	Maximum (mg/L)
Calibration set	24	15.75	1	32
Validation set	8	18.75	3	31

## 2.4. Data Processing and Modeling Methods

### 2.4.1. Problem Formulation

PLS combines the advantages of principal component analysis, canonical correlation analysis and linear regression analysis, and is a widely used modeling method in spectral analysis. In order to study the statistical relationship between the dependent variable  $X = \{x_1, \dots, x_p\}$  and the independent variable  $Y = \{y_1, \dots, y_p\}$ ,  $n$  sample points were observed and the sample tables  $X$  and  $Y$  are formed. PLS extracted components  $x$  and  $y$  from  $X$  and  $Y$  respectively. When extracting  $x$  and  $y$ , the following two requirements should be satisfied:

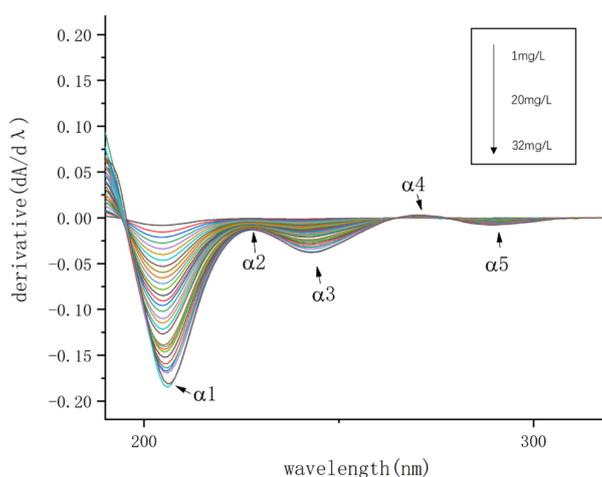
- (1)  $x$  and  $y$  should carry as much change information as possible in their respective data sheets;
- (2) The correlation between  $x$  and  $y$  reach the maximum.

Two requirements indicate that  $x$  and  $y$  should represent the table of  $X$  and  $Y$ . The independent variable  $x$  has the strongest explanatory power to the dependent variable  $y$ . From the collected spectra

(Figure 2), it can be found that the original spectra reflect comprehensive information about sample concentration. Therefore, the sampled spectral data can be used to predict the COD concentration of the solution. Researchers usually use absorbances at characteristic wavelengths as an independent variable to predict COD concentration. But the problem of characteristic wavelengths selection is difficult to solve. In order to solve this problem, the permutation entropy (PE) is used to determine the characteristic wavelengths which are used to get extreme points. The extreme points are used as independent variables of PLS to predict COD concentration.

PE can process the original spectral data and get extreme points of the original spectrum. It can be obtained that there are three extreme points  $\gamma_i (i = 1, 2, 3)$  in the spectral data of different concentrations in Figure 2. However, when PLS uses  $\gamma_i (i = 1, 2, 3)$  to predict COD concentration, the accuracy is not satisfied because the independent variable parameters are not enough.

PE can find more extremum points for each concentration by processing the first derivative data of the original spectrum. The first derivative spectrum of the original spectrum is shown in Figure 4. It can be seen from Figure 4 that there are five extreme points  $\alpha_i (i = 1, 2, 3, 4, 5)$  in the first derivative data of different concentrations, which represent the extreme values of the absorption spectrum change rate under different concentrations.  $\lambda_i (i = 1, 2, 3, 4, 5)$  is defined as the wavelength of  $\alpha_i (i = 1, 2, 3, 4, 5)$ . PE can accurately search  $\lambda_i (i = 1, 2, 3, 4, 5)$  which can be used to get  $\alpha_i (i = 1, 2, 3, 4, 5)$ . PLS uses  $\alpha_i (i = 1, 2, 3, 4, 5)$  to predict COD concentration, which improves the prediction accuracy. The comparison of the results of predicting COD by  $\gamma_i (i = 1, 2, 3)$  and  $\alpha_i (i = 1, 2, 3, 4, 5)$  will be introduced in Section 3.1.3. In this paper, the extreme values of the absorption spectrum change rate were proposed as a characteristic parameter to establish the COD prediction model.



**Figure 4.** The first derivative spectra of COD standard solution absorption spectra, where  $\alpha_1$  is the first extreme point,  $\alpha_2$  is the second extreme point,  $\alpha_3$  is the third extreme point,  $\alpha_4$  is the fourth extreme point,  $\alpha_5$  is the fifth extreme point.

#### 2.4.2. FDPE

The derivative method is an effective method for processing spectral data. Due to the UV–Vis absorption spectroscopy, useful information can be obtained from the UV–Vis spectrum data by increasing the derivative order, while too high order would lead to a low signal intensity of the signal to be detected and great sensitivity to noise components [13]. A suitable derivative order is very important. The first derivative has almost no loss of the signal-to-noise ratio and can reflect the absorption spectrum change rate. By deriving Equation (1), Equation (2) can be obtained.

$$\frac{dA}{d\lambda} = \frac{dK(\lambda)}{d\lambda} CL \quad (2)$$

According to Equation (2),  $\frac{dK(\lambda)}{d\lambda}$  is proportional to C for a specific measurement wavelength and a specific water sample [15]. In this paper, the first derivative method was used to process the UV-Vis absorption spectrum. The first derivative spectrum of any concentration can be set as a continuous sequence  $\{x(i), i = 1, 2, \dots, N\}$ . In order to obtain  $\alpha_i (i = 1, 2, 3, 4, 5)$  in the first derivative spectrum, PE is used to search the extreme points in the sequence. The calculation process is as follows [16,17].

Let  $L (L \leq N)$  represent the window length. In the window  $\{x(i), i = 1, 2, \dots, N\}$  can be reconstructed as

$$\left\{ \begin{array}{l} X(1) = \{x(1), x(1+t), \dots, x(1+(m-1)t)\} \\ X(2) = \{x(2), x(2+t), \dots, x(2+(m-1)t)\} \\ \vdots \\ X(k) = \{x(k), x(k+t), \dots, x(k+(m-1)t)\} \\ \vdots \\ X(L-(m-1)t) = \{x(L-(m-1)t), x(L-(m-2)t), \dots, x(L)\} \end{array} \right. \quad (3)$$

In each vector  $X(k)$ ,  $x(k+(j_i-1)t) (1 \leq i \leq m)$  are arranged in increasing order. The ranking result is as follows.

$$X(k) = \{x(k+(j_1-1)t), x(k+(j_2-1)t), \dots, x(k+(j_m-1)t)\} \quad (4)$$

where  $x(k+(j_1-1)t) \leq x(k+(j_2-1)t) \leq \dots \leq x(k+(j_m-1)t)$ .

If  $x(k+(j_p-1)t) = x(k+(j_q-1)t)$  and  $j_p < j_q$ , it can be ranked as follow.

$$x(k+(j_p-1)t) \leq x(k+(j_q-1)t) \quad (5)$$

Equation (4) can be formulated as Equation (6).

$$X(k) = \{x(k+(j_1-1)t), x(k+(j_2-1)t), \dots, x(k+(j_p-1)t), x(k+(j_q-1)t), \dots, x(k+(j_m-1)t)\} \quad (6)$$

Correspondingly, each vector  $X(k)$  can generate a set of symbolic sequence,

$$S(l) = \{j_1, j_2, \dots, j_m\} \quad (7)$$

where  $l = 1, 2, 3, \dots, n (n \leq m!)$ . The probability of each symbol sequence is  $P_1, P_2, \dots, P_n$ . Then, PE of  $x(i)$  in different windows can be calculated by Equation (8).

$$H = - \sum_{l=1}^n P_l \ln P_l \quad (8)$$

$H$  represents the entropy of the first derivative spectrum in a window. When  $H = 0$ , the first derivative spectrum is regular, that is, the first derivative spectrum keeps rising or keeps falling. When  $H \neq 0$ , the trend of the first derivative spectrum changes, which means the first derivative spectrum has an extreme point. The windows where  $H$  changes from zero to non-zero can be used to find  $\alpha_i (i = 1, 2, 3, 4, 5)$  in the first derivative spectrum.

### 2.4.3. Partial Least Squares (PLS)

PLS is widely used in spectral analysis and can establish a mathematical model to predict the concentration of sample [18]. In this paper,  $\alpha_i (i = 1, 2, 3, 4, 5)$  found by PE were used to predict the concentration of COD solutions by PLS. The prediction performance of the model is based on several indices, such as the linear correlation coefficient ( $R^2$ ), root mean-squared error of calibration

(RMSEC) and root mean-squared error of validation (RMSEV). The performance indices are shown in Equations (9)–(11).

$$R^2 = 1 - \frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \tag{9}$$

$$RMSEC = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n - A - 1}} \tag{10}$$

$$RMSEV = \sqrt{\frac{1}{m} \sum_{i=1}^m (\hat{y}_i - y_i)^2} \tag{11}$$

where  $\hat{y}_i$  is the predicted value by calibration model,  $\bar{y}$  is the mean of measurements,  $y_i$  is the measurement,  $n$  is the number of calibration samples,  $m$  is the number of validation samples, and  $A$  is the number of regression factors [19,20].

#### 2.4.4. The Flow of FDPE\_PLS

In this paper, a novel hybrid strategy based on ultraviolet–visible absorption spectroscopy was proposed to detect COD in surface water. Different processing procedures were used with or without turbidity interference. The processing procedure of the hybrid strategy without turbidity interference is FDPE\_PLS. The flow chart is shown in Figure 5. The processing procedure of the hybrid strategy with turbidity interference is FDPE\_PLS combined with the turbidity compensation method. This part will be introduced in Section 2.5.3.

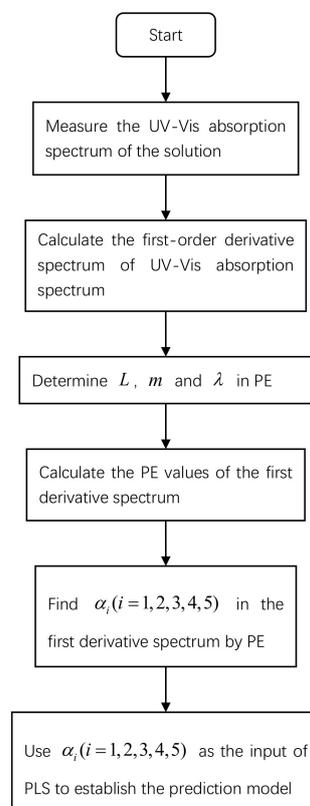


Figure 5. The flow chart of FDPE\_PLS.

## 2.5. Turbidity Compensation

### 2.5.1. Double-Wavelength Method (DWM)

Turbidity absorbs visible light, and the absorbance curve changes with the change of turbidity value, which affects the determination of COD in water by spectrophotometry. It is important to compensate for turbidity. The traditional method uses the absorbance at 546 nm to predict the turbidity value. This method is simple, but the accuracy needs to be improved. In order to improve the accuracy, a dual-wavelength method (DWM) was proposed to determine the turbidity values in this paper. Because turbidity has an absorption in the visible light band. The DWM uses the slopes of the lines connecting the absorbance values at 400 nm and 600 nm to predict the turbidity values. As shown in Figure 6, the slopes of the lines and turbidity values have a good linear relationship, which can be used to calculate the turbidity values.

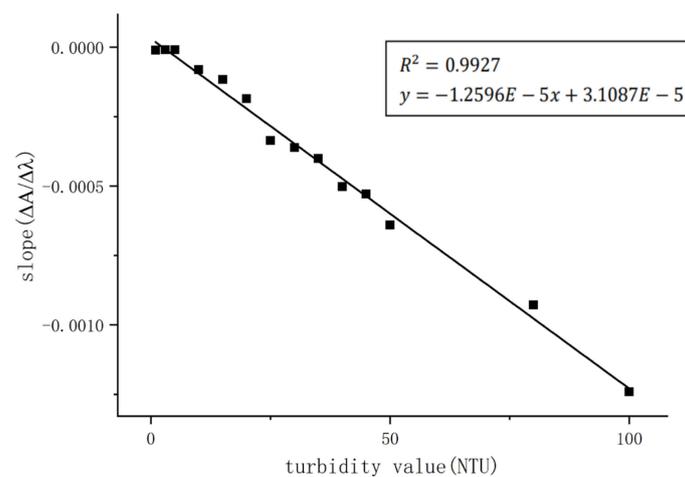


Figure 6. Fitting relationship between slopes and turbidity values.

### 2.5.2. Multiple Scatter Correction (MSC)

MSC method is a commonly used data processing method, which can reduce the influence of scattering [21]. The calculation process is as follows:

- (1) Calculate the average spectrum of the sample

$$\bar{A} = \frac{\sum_{i=1}^n A_{i,j}}{n} \tag{12}$$

- (2) Perform unary linear regression

$$A_i = m_i \bar{A} + b_i \tag{13}$$

- (3) Multiple scatter correction

$$A_{i(MSC)} = \frac{(A_i - b_i)}{m_i} \tag{14}$$

### 2.5.3. Turbidity Compensation

$\alpha_i (i = 1, 2, 3, 4, 5)$  are searched in first derivative spectrum by FDPE\_PLS. Turbidity can change the amplitudes of  $\alpha_i (i = 1, 2, 3, 4, 5)$ . In order to remove the interference of turbidity, a new method called derivative compensation (DC) was proposed to combine with FDPE\_PLS for predicting the COD. We defined  $\beta_i (i = 1, 2, 3, 4, 5)$  to reflect the change of  $\alpha_i (i = 1, 2, 3, 4, 5)$  caused by turbidity. The calculation process of  $\beta_i (i = 1, 2, 3, 4, 5)$  is shown in Section 3.2.2.

$\tau_i (i = 1, 2, 3, 4, 5)$  was defined as the characteristic values after removing turbidity interference, which can be used to predict COD concentrations.  $\tau_i (i = 1, 2, 3, 4, 5)$  can be calculated by Equation (15).

$$\tau_i = \alpha_i - \beta_i (i = 1, 2, 3, 4, 5) \quad (15)$$

The specific steps of the hybrid strategy with turbidity interference are shown in Figure 7.

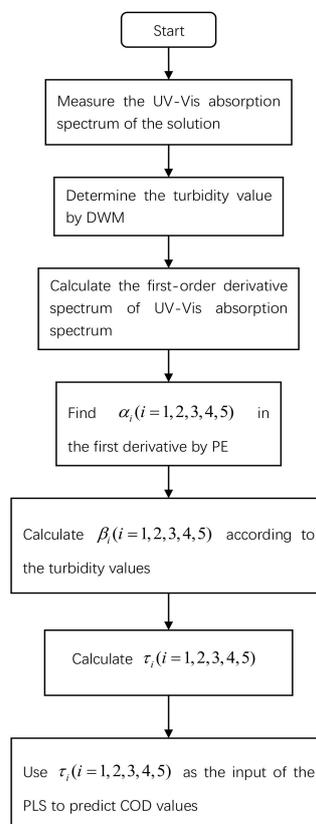


Figure 7. The flow chart of FDPE\_PLS with turbidity interference.

### 3. Results and Discussion

This section focuses on the experimental verification of the proposed method. Firstly, the experiment without turbidity interference was designed. FDPE\_PLS model was used to predict COD, and the results were compared with the PE\_PLS model. The experiment of turbidity compensation was designed, and the turbidity compensation model was established and compared with MSC. Finally, the method proposed in this paper is tested on actual water samples and compared with other common methods.

### 3.1. No Turbidity Interference

#### 3.1.1. FDPE\_PLS

##### (1) First derivative spectroscopy

In the absence of turbidity interference, the first derivative of the original spectrum was first processed. The acceleration and deceleration information of the original spectrum is added, as shown in Figure 4. Combining with PE in Section 3.1.1, (2) can accurately determine the characteristic values under the condition of reducing noise interference. It can be seen from Figure 4 that there are four local maximum points  $\alpha_1, \alpha_3, \alpha_4, \alpha_5$  and one local minimum point  $\alpha_2$  in the first derivative spectrum, which reflected four local maximum values of change rate and one local minimum value of change rate in the original spectrum. When the concentration of the solution changed, the amplitudes and wavelengths of  $\alpha_i (i = 1, 2, 3, 4, 5)$  changed accordingly.

##### (2) Feature wavelengths extraction by PE

PE can accurately locate the time and position of the sequence change. PE algorithm needs to configure parameters  $L, m, t$ . In this paper, according to reference [22] and several attempts, set  $L = 10, m = 5,$  and  $t = 1$ . After calculating the entropy values of the first derivative spectrum in the concentration range of 1 ~ 32 mg/L, Table 2 shows the windows in which entropy change from zero to non-zero.

**Table 2.** The windows in which entropy change from zero to non-zero.

Concentrations (mg/L)	Window in Which Entropy Change from Zero to Non-Zero (W)				
	1	2	3	4	5
1	7	31	42	71	89
2	8	30	43	72	88
3	7	30	43	71	88
4	8	31	43	72	87
5	8	30	43	72	88
6	8	30	43	72	88
7	8	30	43	72	87
8	8	30	43	72	88
9	8	30	43	72	88
10	8	30	43	72	88
11	8	30	44	72	88
12	8	30	43	72	88
13	8	30	43	73	88
14	8	30	44	73	88
15	8	30	44	73	88
16	8	30	44	73	88
17	8	30	44	73	88
18	8	30	44	73	88
19	8	30	44	73	88
20	8	30	45	73	89
21	8	30	45	73	89
22	8	30	45	73	90
23	8	31	45	73	90
24	8	31	45	73	91
25	8	31	45	73	91
26	8	31	45	73	92
27	9	31	45	73	92
28	9	31	45	73	92
29	9	31	46	73	92
30	9	31	46	73	93
31	9	31	46	73	92
32	9	31	46	74	93

In Table 2, it can be found that the entropy changes five times from zero to non-zero, which means that the first derivative spectrum has five extreme points. The conclusion is consistent with the actual situation in Figure 4. By determining the window where entropy changes from zero to non-zero,  $\lambda_i (i = 1, 2, 3, 4, 5)$  can be accurately found. According to the calculation of PE in reference [22],  $\lambda_i (i = 1, 2, 3, 4, 5)$  can be calculated by Equation (16).

$$\lambda_i = \lambda_0 + W + L - 3 \quad (i = 1, 2, 3, 4, 5) \tag{16}$$

where  $\lambda_0$  is starting wavelength,  $W$  is the window in which the entropy changes from zero to non-zero. The  $\lambda_i (i = 1, 2, 3, 4, 5)$  COD solutions with different concentrations are shown in Figure 8.

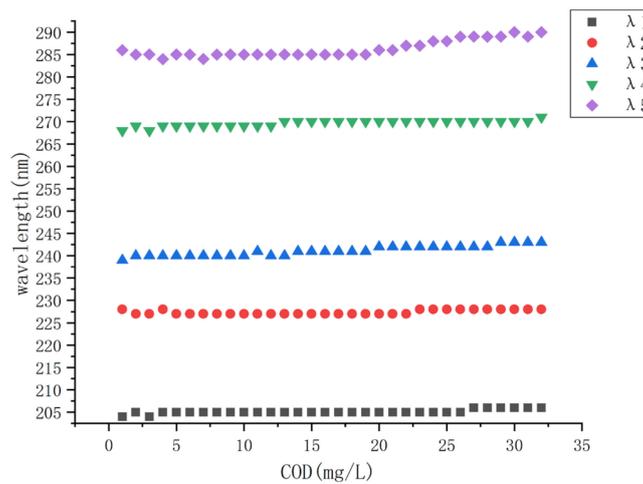


Figure 8. Found by PE in the concentration range of 1~32 mg/L.

It can be seen from Figure 8 that  $\lambda_i (i = 1, 2, 3, 4, 5)$ , which are corresponded to  $\alpha_i (i = 1, 2, 3, 4, 5)$ , changes in a small range when the COD concentration changes. The variation range of  $\lambda_i (i = 1, 2, 3, 4, 5)$  is shown in Table 3.

Table 3. The variation range of  $\lambda_i (i = 1, 2, 3, 4, 5)$ .

Wavelength	Variation Range
$\lambda_1$	$204 \text{ nm} \leq \lambda_1 \leq 206 \text{ nm}$
$\lambda_2$	$227 \text{ nm} \leq \lambda_2 \leq 228 \text{ nm}$
$\lambda_3$	$239 \text{ nm} \leq \lambda_3 \leq 243 \text{ nm}$
$\lambda_4$	$268 \text{ nm} \leq \lambda_4 \leq 271 \text{ nm}$
$\lambda_5$	$284 \text{ nm} \leq \lambda_5 \leq 290 \text{ nm}$

In order to verify that  $\frac{dA}{d\lambda}$  and  $C$  are linear when  $\lambda$  changes in a small range. We established the fitting relationship between COD concentrations and  $\alpha_i (i = 1, 2, 3, 4, 5)$ . The results are shown in Figure 9 and Table 4. It can be seen from Table 4 that there are good linear relationships between COD concentrations and  $\alpha_i (i = 1, 2, 3, 4, 5)$  when  $\lambda_i (i = 1, 2, 3, 4, 5)$  changes in a small range. The selection of  $\lambda_i (i = 1, 2, 3, 4, 5)$  and  $\alpha_i (i = 1, 2, 3, 4, 5)$  are effective.

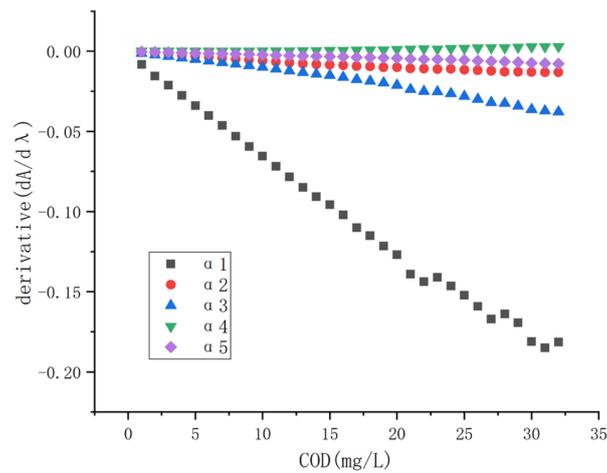


Figure 9. The linear relationship between the COD and  $\alpha_i (i = 1, 2, 3, 4, 5)$ .

Table 4. The correlation coefficient between COD concentration and  $\alpha_i (i = 1, 2, 3, 4, 5)$ .

Characteristic Points	$\alpha_1$	$\alpha_2$	$\alpha_3$	$\alpha_4$	$\alpha_5$
$R^2$	0.9942	0.9737	0.9913	0.9019	0.9927

### (3) Comparison of PE and derivative method

The original spectrum measured by the spectrometer is not smooth, and the first derivative spectrum may produce peak noise, as shown in Figure 10b. PE algorithm and the second derivative algorithm are used to calculate the spectral extremum of the first derivative in Figure 10a,b respectively. The second derivative method is used to find the extremum point of the first derivative spectrum, and the results are shown in Figure 11a,b.

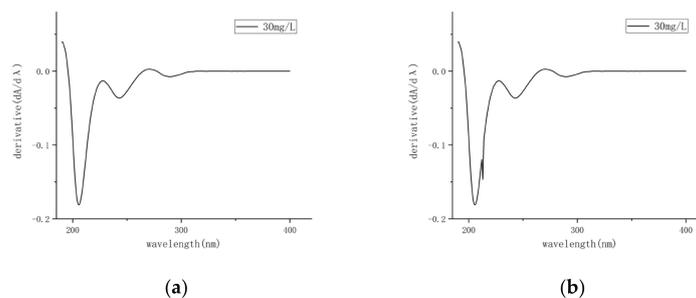


Figure 10. (a) The first derivative spectrum of COD solution after smoothing; (b) The first derivative spectrum of COD solution after unsmoothing.

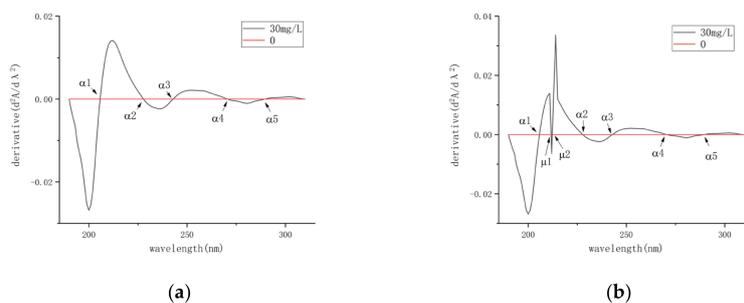


Figure 11. (a) The second derivative spectrum of COD solution after smoothing; (b) The second derivative spectrum of COD solution after unsmoothing.

It can be seen from Figure 11a that for the smoothed first derivative spectrum, the second derivative method can correctly find five extreme points  $\alpha_i (i = 1, 2, 3, 4, 5)$ . It can be seen from Figure 11b that when there is a spike noise, there is an error in the number of effective extreme points calculated by the second derivative method.  $\mu_1$  and  $\mu_2$  are interference due to noise.

Set  $L = 10, m = 5,$  and  $t = 1$ . PE algorithm was used to find the extreme points of the first derivative spectrum, the window of entropy value changes from zero to non-zero, as shown in Table 5:

**Table 5.** The windows of entropy values change from zero to non-zero.

	Window in Which Entropy Change from Zero to Non-Zero				
	1	2	3	4	5
without spike noise	9	31	46	73	93
with spike noise	9	31	46	73	93

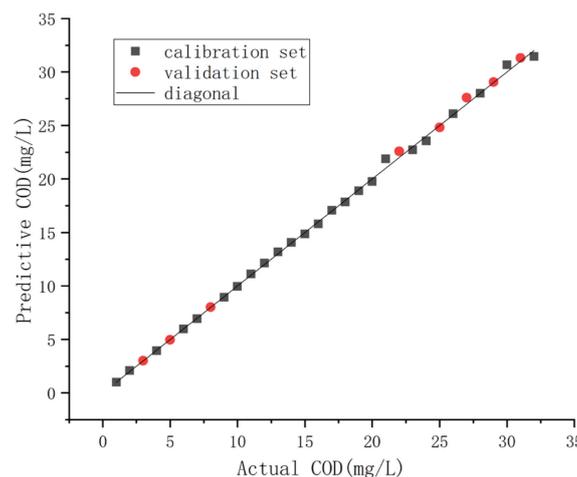
According to Equation (16) and Table 5, the effective extreme points calculated by PE algorithm are correct. PE algorithm does not need another smoothing process, which can effectively avoid the interference of peak noise and find the characteristic wavelength.

(4) PLS model

All samples were processed by SPXY and divided into a calibration set and validation set. The division result is shown in Table 1.  $\alpha_i (i = 1, 2, 3, 4, 5)$  in the calibration set were used as independent variables and COD concentrations were used as dependent variables to establish the COD prediction model. The model is shown as Equation (17).

$$C = -0.3442 - 43.8484 * \alpha_1 - 730.6449 * \alpha_2 - 212.421 * \alpha_3 + 1277.572 * \alpha_4 - 314.5272 * \alpha_5 \quad (17)$$

The fitting result of the model is shown in Figure 12, where black dots and red dots represent the fitting results of the calibration set and validation set, respectively. The performance of the fitting results is shown in Table 6.



**Figure 12.** Fitting between predictive values and actual values in the calibration set and validation set in FDPE\_PLS.

**Table 6.** The comparison between PE\_PLS and FDPE\_PLS.

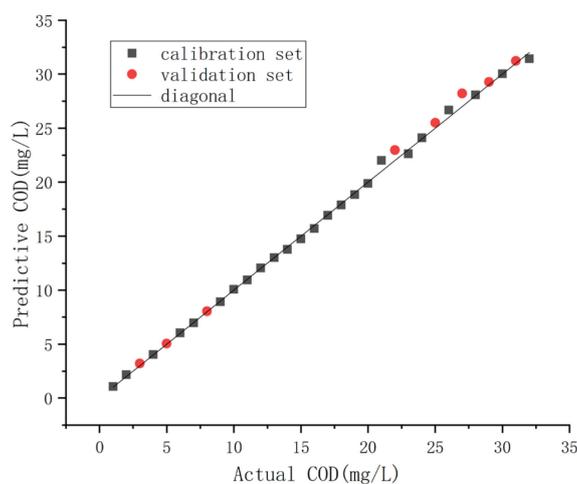
	PE_PLS		FDPE_PLS	
	Calibration Set	Validation Set	Calibration Set	Validation Set
$R^2$	0.9987	0.9988	0.9988	0.9995
RMSE	0.3219	0.4043	0.3051	0.2919
Average relative error	1.6984%	1.6107%	1.2198%	1.036%

### 3.1.2. PE\_PLS

We use the original spectrum to detect the COD concentration, and the PE algorithm was adopted to process the original spectral data directly. PE was used to find  $\gamma_i (i = 1, 2, 3)$  in the original spectrum. PLS was used to establish the prediction model. According to the division of calibration set and validation set in Table 1,  $\gamma_i (i = 1, 2, 3)$  in the calibration set were used as independent variables, and the COD concentrations were used as dependent variables to establish COD prediction model. The model is shown as Equation (18).

$$C = -0.0669 + 9.882 * \gamma_1 - 412.2077 * \gamma_2 + 300.5518 * \gamma_3 \quad (18)$$

The fitting result of the model is shown in Figure 13, where black dots and red dots represent the fitting results of the calibration set and validation set, respectively. The performance of the fitting results is shown in Table 6.



**Figure 13.** Fitting between predictive values and actual values in the calibration set and validation set in PE\_PLS.

### 3.1.3. Comparison of FDPE\_PLS and PE\_PLS

We performed two experiments without turbidity interference. The first was using FDPE\_PLS to establish the model by five extreme points in the first derivative spectrum. The second was using PE\_PLS to establish the model by three extreme points in the original spectra. The calibration and verification sets were used to verify the performance of the two COD prediction models. The performance of the two models were compared, and the result is shown in Table 6.

It can be seen from Table 6 that the performance of FDPE\_PLS is better than PE\_PLS. FDPE\_PLS has lower RMSE and average relative error, which means the prediction model established by five extreme points in the first derivative spectrum has higher accuracy. So FDPE\_PLS was chosen as the modeling method in this paper.

### 3.2. Disturbed by Turbidity

#### 3.2.1. Effectiveness of DWM for Turbidity Measurement

To verify the accuracy of the method, 10 different water samples from a city in Jilin Province were collected and measured their UV–Vis absorption spectra. The double-wavelength method proposed in this paper and the 546 nm single-wavelength method were used to predict the turbidity values of water samples, respectively. The prediction effect is shown in Table 7.

**Table 7.** Prediction results of double-wavelength method and single-wavelength method.

Processing Methods	$R^2$	Average Relative Error	Maximum Relative Error
Single-wavelength	0.98652	7.18%	22.02%
Double-wavelength	0.98787	5.52%	12.99%

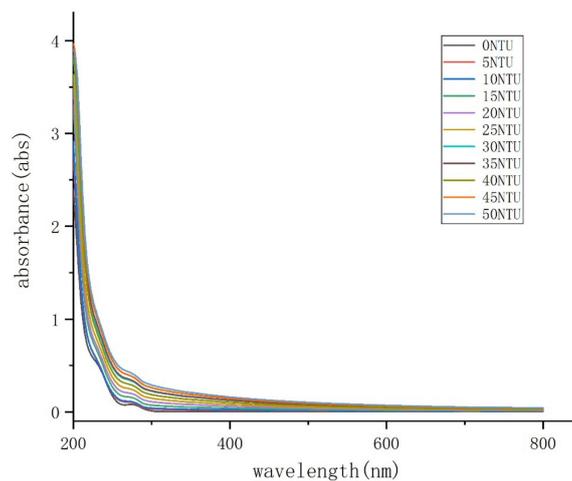
It can be seen from Table 7 that the dual-wavelength method can predict the turbidity of water samples more effectively than the single wavelength method.

#### 3.2.2. Establish Turbidity Compensation Model

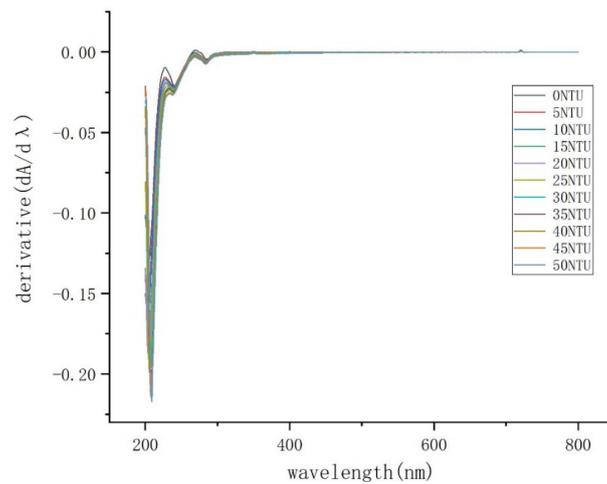
Aiming to establish the turbidity compensation model, we carried out the following experiments. We added the turbidity solutions into the COD solutions to produce ten different mixtures, as shown in Table 8. The UV–Vis absorption spectrum of these mixtures and COD solution whose concentration was 20 mg/L are shown in Figure 14. Their first derivative spectra are shown in Figure 15.

**Table 8.** The COD and turbidity values of ten samples.

Sample	1	2	3	4	5	6	7	8	9	10
COD (mg/L)	20	20	20	20	20	20	20	20	20	20
Turbidity (NTU)	5	10	15	20	25	30	35	40	45	50

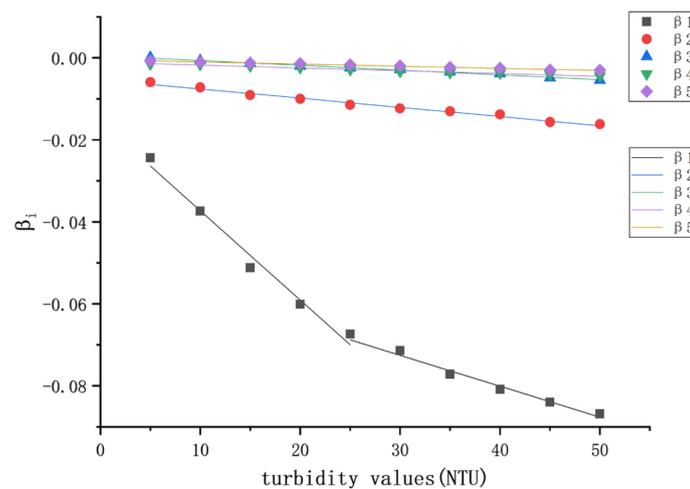


**Figure 14.** Original UV–Vis absorption spectra of solutions with different turbidity.



**Figure 15.** The first derivative of the UV-Vis absorption spectra of solutions with different turbidity.

In order to establish a turbidity compensation model for getting  $\tau_i (i = 1, 2, 3, 4, 5)$ , we had to calculate  $\beta_i (i = 1, 2, 3, 4, 5)$  first. The fitting results between  $\beta_i (i = 1, 2, 3, 4, 5)$  and turbidity values are shown in Figure 16. The turbidity values can be calculated by DWM.



**Figure 16.** The linear relationship between the turbidity values and  $\beta_i (i = 1, 2, 3, 4, 5)$ .

It is clear that  $\beta_1$  caused by turbidity can be represented by two linear relationships, while the variables  $\beta_2, \beta_3, \beta_4,$  and  $\beta_5$  can be represented by one linear relationship. A series of mathematical equations were developed which describes the relationship between turbidity values and variables. The equations are shown as follows:

$$\beta_1 = \begin{cases} -0.0022x - 0.0155 & (0 < x < 25) \\ -0.0008x - 0.0499 & (25 < x < 50) \end{cases} \quad (19)$$

$$\begin{cases} \beta_2 = -0.0002x - 0.0053 \\ \beta_3 = -0.0001x + 0.0005 \\ \beta_4 = -0.0001x - 0.0011 \\ \beta_5 = -0.0001x - 0.0005 \end{cases} \quad (0 < x < 50) \quad (20)$$

where  $x$  is the turbidity value. The correlation coefficients of these linear equations are shown in Table 9.

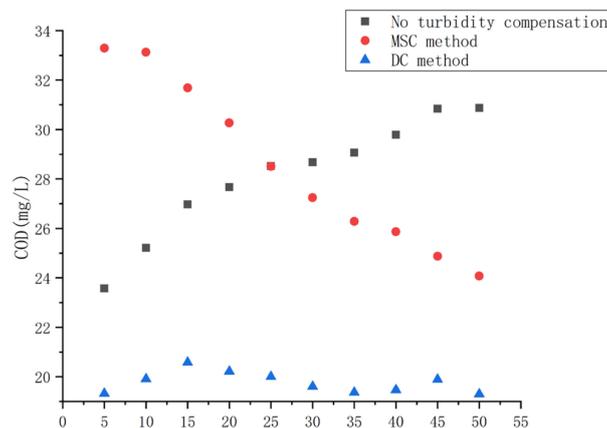
**Table 9.** The correlation coefficients of five linear equations.

	$\beta_1 (0 < x < 25)$	$\beta_1 (25 < x < 50)$	$\beta_2$	$\beta_3$	$\beta_4$	$\beta_5$
$R^2$	0.98254	0.97795	0.98756	0.9916	0.98858	0.99209

It can be seen from Table 9 that  $\beta_i (i = 1, 2, 3, 4, 5)$  and turbidity values have good linear relationships and  $\beta_i (i = 1, 2, 3, 4, 5)$  can be calculated by Equation(19) and Equation(20). From Equation (15), the turbidity compensation of the first derivative spectrum can be completed and  $\tau_i (i = 1, 2, 3, 4, 5)$  can be obtained.  $\tau_i (i = 1, 2, 3, 4, 5)$  are the characteristic values, which can be used to detect COD values with turbidity interference.

### 3.2.3. The Comparison of Turbidity Compensation Models

In order to verify the effectiveness of the derivative compensation (DC) method proposed in this paper, we processed ten mixtures, which are shown as Table 8, to remove turbidity interference by DC method and MSC method, respectively. Then FDPE\_PLS were used to predict the COD concentrations. The comparison result of MSC method and DC method is shown in Figure 17. The results show that DC is an effective method to remove turbidity interference.



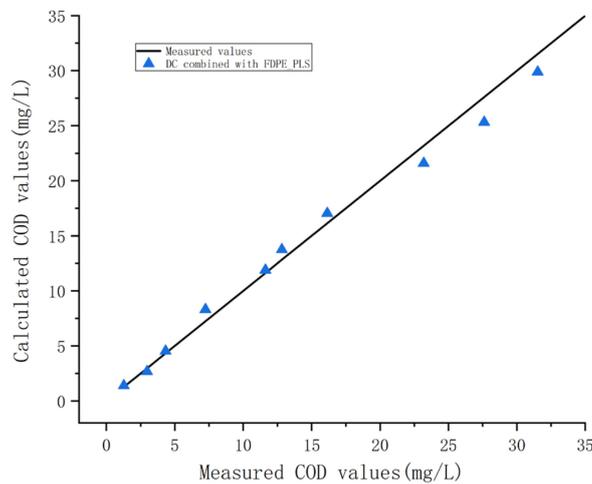
**Figure 17.** The comparison among no turbidity compensation, MSC, and DC. No turbidity compensation: the COD measurement errors become large as the turbidity values increase. MSC: the COD measurement errors become small as the turbidity values increase. DC: the COD measurement errors fluctuate with the change of turbidity value.

### 3.3. Experiments of Actual Water Sample

We have verified the accuracy of COD measurement using FDPE\_PLS with or without turbidity interference in the laboratory. In this section, the accuracy of FDPE\_PLS and DC method proposed in this paper will be verified by measuring actual water samples. Ten actual water samples collected in a city in Jilin Province were used for testing. The COD values of actual water samples were determined by chemical method in the laboratory, and the accuracy was  $\pm 5\%$ . The turbidity values were determined by turbidimeter ZD-2A with an accuracy of  $\pm 8\%$ . The COD and turbidity values of the ten water samples are shown in Table 10. After processing by DC and calculating by FDPE\_PLS, the fitting result between actual COD values and predicted COD values is shown in Figure 18. The  $R^2$  is 0.99, and the average relative error is 0.07.

**Table 10.** The COD and turbidity values of ten actual samples.

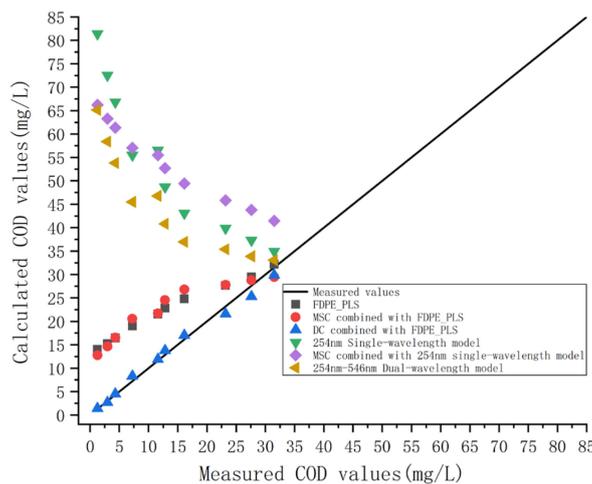
Samples	1	2	3	4	5	6	7	8	9	10
COD (mg/L)	1.28	2.97	4.33	7.24	11.62	12.83	16.14	23.19	27.61	31.52
Turbidity (NTU)	49.55	47.92	41.77	35.12	29.64	25.03	18.88	13.91	10.45	6.54



**Figure 18.** The relationship between actual COD values and predicted COD values.

### 3.4. Comparison of Several Methods

In order to compare the accuracy of the method proposed in this paper with other methods, the ten water samples were modeled by the following methods: the FDPE-PLS, MSC combined with FDPE-PLS, DC combined with FDPE-PLS, 254 nm single-wavelength model, MSC combined with 254 nm single-wavelength model and 254 nm–546 nm dual-wavelength method. The fitting results of the COD calculated by these models and the COD measurements were shown in Figure 19.



**Figure 19.** The comparison between different modeling methods.

Figure 19 shows that DC combined with FDPE-PLS has the best performance, whose  $R^2$  was 0.99 and the average relative error was 0.07. This result can prove the effectiveness of the method proposed in this paper.

#### 4. Conclusions

In this paper, a new characteristic parameter, the extreme value of the absorption spectrum change rate, is proposed for the first time as a characteristic parameter to determine the characteristic wavelength. On this basis, we proposed to use FDFE to determine the characteristic wavelength. Compared with the traditional methods, this method improved the detection accuracy and stability of the characteristic wavelength. Secondly, a dual-wavelength method was proposed for the first time, which is different from the traditional single-wavelength method. The  $R^2$  between the predicted turbidity value and the actual turbidity value was 0.98. Compared with the single-wavelength method, DWM improved the measurement accuracy of turbidity. Finally, a new turbidity compensation method was proposed. The traditional turbidity compensation method is to process the original spectral data. The new method of turbidity compensation compensates for the interference caused by turbidity in the first derivative spectrum. The experimental results showed that the FDPE\_PLS with turbidity compensation had good performance.

Although this paper has made some progress, there are still some limitations. For example, the influence of temperature and PH was not considered, and there was no standard method for selecting the parameters of permutation entropy. These all need to be improved in future work.

**Author Contributions:** Conceptualization, G.Z. and Q.D.; data curation, G.Z.; formal analysis, G.Z.; investigation, G.Z., X.L., Y.W., and Q.D.; methodology, G.Z.; software, G.Z.; writing—original draft preparation, G.Z.; writing—review and editing, G.Z. and Q.D.; visualization, G.Z.; supervision, Q.D.; project administration, Q.D. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Acknowledgments:** This work was supported by Jilin Province Science and Technology Development Plan Project (20190303008SF), Talent Program of Hainan Medical University(XRC180010,HYPY201905,XRC180006), and Hundred-Talent Program (Hainan 2018).

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

#### References

1. Naidoo, S.; Olaniran, A.O. Treated Wastewater Effluent as a Source of Microbial Pollution of Surface Water Resources. *Int. J. Environ. Res. Public Health* **2014**, *11*, 249–270. [[CrossRef](#)] [[PubMed](#)]
2. Wang, L.L.; Liu, X.H.; Shi, X.X.; Lu, Y.R.; Qi, Y.; Wang, M.Y. Study on real-time monitoring of seawater COD by UV-Vis Spectroscopy. In *Advances in Environmental Technologies, Pts 1-6*; Zhao, J., Iranpour, R., Li, X., Jin, B., Eds.; Trans Tech Publications Ltd: Guilin, China, 2013; Volume 726–731, pp. 1534–1537.
3. Li, J.; Luo, G.; He, L.; Xu, J.; Lyu, J. Analytical Approaches for Determining Chemical Oxygen Demand in Water Bodies: A Review. *Crit. Rev. Anal. Chem.* **2018**, *48*, 47–65. [[CrossRef](#)] [[PubMed](#)]
4. Chen, B.; Wu, H.; Li, S.F.Y. Development of variable pathlength UV-vis spectroscopy combined with partial-least-squares regression for wastewater chemical oxygen demand (COD) monitoring. *Talanta* **2014**, *120*, 325–330. [[CrossRef](#)] [[PubMed](#)]
5. Li, J.; Tong, Y.; Guan, L.; Wu, S.; Li, D. A turbidity compensation method for COD measurements by UV-vis spectroscopy. *Optik* **2019**, *186*, 129–136. [[CrossRef](#)]
6. Guan, L.; Tong, Y.; Li, J.; Wu, S.; Li, D. An online surface water COD measurement method based on multi-source spectral feature-level fusion. *RSC Adv.* **2019**, *9*, 11296–11304. [[CrossRef](#)]
7. Wang, C.; Li, W.; Huang, M. High precision wide range online chemical oxygen demand measurement method based on ultraviolet absorption spectroscopy and full-spectrum data analysis. *Sens. Actuators B* **2019**, *300*, 126943. [[CrossRef](#)]
8. Wang, X.-m.; Zhang, H.-l.; Luo, W.; Liu, X.-m. Measurement of Water COD Based on UV-Vis Spectroscopy Technology. *Spectrosc. Spectr. Anal.* **2016**, *36*, 177–180.
9. Yun, Y.-H.; Wang, W.-T.; Tan, M.-L.; Liang, Y.-Z.; Li, H.-D.; Cao, D.-S.; Lu, H.-M.; Xu, Q.-S. A strategy that iteratively retains informative variables for selecting optimal variable subset in multivariate calibration. *Anal. Chim. Acta* **2014**, *807*, 36–43. [[CrossRef](#)]

10. Deng, B.-c.; Yun, Y.-h.; Liang, Y.-z.; Yi, L.-z. A novel variable selection approach that iteratively optimizes variable space using weighted binary matrix sampling. *Analyst* **2014**, *139*, 4836–4845. [[CrossRef](#)]
11. Yun, Y.-H.; Wang, W.-T.; Deng, B.-C.; Lai, G.-B.; Liu, X.-b.; Ren, D.-B.; Liang, Y.-Z.; Fan, W.; Xu, Q.-S. Using variable combination population analysis for variable selection in multivariate calibration. *Anal. Chim. Acta* **2015**, *862*, 14–23. [[CrossRef](#)]
12. Chen, J.; Liu, S.; Qi, X.; Yan, S.; Guo, Q. Study and design on chemical oxygen demand measurement based on ultraviolet absorption. *Sens. Actuators B* **2018**, *254*, 778–784. [[CrossRef](#)]
13. Hu, Y.; Zhao, D.; Qin, Y.; Wang, X. An order determination method in direct derivative absorption spectroscopy for correction of turbidity effects on COD measurements without baseline required. *Spectrochim. Acta Part A* **2020**, *226*, 117646. [[CrossRef](#)] [[PubMed](#)]
14. Tian, H.; Zhang, L.; Li, M.; Wang, Y.; Sheng, D.; Liu, J.; Wang, C. Weighted SPXY method for calibration set selection for composition analysis based on near-infrared spectroscopy. *Infrared Phys. Technol.* **2018**, *95*, 88–92. [[CrossRef](#)]
15. Wang, J.-m.; Zhang, J.-c.; Zhang, Z.-j. Rapid Determination of Nitrate Nitrogen and Nitrite Nitrogen by Second Derivative Spectrophotometry. *Spectrosc. Spectr. Anal.* **2019**, *39*, 161–165.
16. Xiao, L.; Lv, Y.; Fu, G. Fault Classification of Rotary Machinery Based on Smooth Local Subspace Projection Method and Permutation Entropy. *Appl. Sci.* **2019**, *9*, 2102. [[CrossRef](#)]
17. Huang, S.; Wang, X.; Li, C.; Kang, C. Data decomposition method combining permutation entropy and spectral substitution with ensemble empirical mode decomposition. *Measurement* **2019**, *139*, 438–453. [[CrossRef](#)]
18. De Luca, M.; Oliverio, F.; Ioele, G.; Ragno, G. Multivariate calibration techniques applied to derivative spectroscopy data for the analysis of pharmaceutical mixtures. *Chemom. Intell. Lab. Syst.* **2009**, *96*, 14–21. [[CrossRef](#)]
19. Huang, Y.; Cao, J.; Ye, S.; Duan, J.; Wu, L.; Li, Q.; Min, S.; Xiong, Y. Near-infrared spectral imaging for quantitative analysis of active component in counterfeit imidacloprid using PLS regression. *Optik* **2013**, *124*, 1644–1649. [[CrossRef](#)]
20. Zhou, K.-p.; Bi, W.-h.; Zhang, Q.-h.; Fu, X.-h.; Wu, G.-q. Influence of temperature and turbidity on water COD detection by UV absorption spectroscopy. *Optoelectron. Lett.* **2016**, *12*, 461–464. [[CrossRef](#)]
21. Tang, B.; Wei, B.; Wu, D.-c.; Mi, D.-l.; Zhao, J.-x.; Feng, P.; Jiang, S.-h.; Mao, B.-j. Experimental Research of Turbidity Influence on Water Quality Monitoring of COD in UV-Visible Spectroscopy. *Spectrosc. Spectr. Anal.* **2014**, *34*, 3020–3024.
22. Li, X.; Li, C. Application of Permutation Entropy in Feature Extraction for Near-Infrared Spectroscopy Noninvasive Blood Glucose Detection. *J. Spectrosc.* **2017**, *2017*. [[CrossRef](#)]

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).