Determination of oil pollutants by three-dimensional fluorescence spectroscopy combined with improved pattern recognition algorithm

Pengfei Cheng¹,2, Yanping Zhu², Chuanjin Cui², Fubin Wang² and Jinyan Pan¹,2

Abstract
Petroleum refineries are one of the main sources of hazardous air pollutants, so the accurate determination of petroleum pollutants is of great significance to maintain ecological balance. In this study, three-dimensional (3D) fluorescence spectroscopy combined with pattern recognition algorithm is adopted to distinguish the composition and content of oil pollutants efficiently and accurately. Three hundred samples of kerosene, diesel, and gasoline mixed solutions with different concentrations are prepared. The principal component analysis is used to extract the optimal feature variables, and the correlation coefficient method is used to obtain eight groups of principal component features in the spectra. The dimension is selected as 8, and the principal component score is calculated, which is used as the input data of the extension neural network. Next, the pattern recognition method is improved, and the designed neural network has functions of both resolution and measurement. The results of neural network pattern recognition are used as the input of the concentration network. The first 270 samples are used as the training samples to train the network model, and the remaining 30 samples are used as test samples, which are applied to the input layer of the trained neural network. The relative fluorescence intensity, relative slope, and comprehensive background parameters are used as the input parameters, and the extension neural network is used for pattern recognition and evaluation of oil pollutants. The experimental results show that the average recognition rate of the improved pattern recognition algorithm for oil pollutants is 98.43%, and the average recovery rate of concentration is 98.67%. Further, the average time for pattern recognition is 1.53 s, while the parallel factor analysis algorithm takes 2.89 s. This suggests that the improved extension neural network is an effective and reliable pattern recognition method for the identification of mixed oil pollutants.

Keywords
Three-dimensional fluorescence spectroscopy, oil pollutants, principal component analysis, extension neural network, pattern recognition

Introduction
Oil pollution causes significant harm to the human health, environment, and local ecosystems. The main oil pollutants are the wastewater discharged from factories and offshore oil leakage, which cause severe water pollution.¹-³ Therefore, the accurate and efficient detection of the composition and content of oil pollutants is of great significance to maintain the ecological balance. Fluorescence spectroscopy is a sensitive technique with high measurement accuracy, which usually requires small concentration of experimental samples. It has been extensively applied for the detection of fluorescent substances such as oil pollutants, pesticides, food, etc.⁴-⁷ The aromatic components in pollutants such as gasoline and diesel have strong fluorescence characteristics under ultraviolet light excitation.⁸-¹¹ The three-dimensional (3D) fluorescence spectrum reflects the simultaneous change in fluorescence intensity with the excitation and emission wavelengths, and it essentially represents the

¹School of Electrical and Control Engineering, Xuzhou University of Technology, Xuzhou, China
²College of Electrical Engineering, North China University of Science and Technology, Tangshan, China

Corresponding author:
Chuanjin Cui, College of Electrical Engineering, North China University of Science and Technology, Tangshan, 063210, China.
Email: cuichuanjin@163.com
continuous distribution of energy in a two-dimensional (2D) region.\textsuperscript{12–14} Compared with the traditional 2D fluorescence spectroscopy, the 3D fluorescence spectroscopy provides more comprehensive information and has strong recognition ability, which is suitable for the detection of multi-component fluorescent substances.\textsuperscript{15–18} Traditional pattern recognition methods have poor detection accuracy and low speed. In the recent years, several scholars have extracted the characteristic parameters of fluorescence spectrum and applied statistical indicators such as origin moment and kurtosis coefficient for the spectral analysis,\textsuperscript{19–21} which can only reflect the overall characteristics of 3D fluorescence spectrum.

Wang et al.\textsuperscript{22} used the back propagation (BP) neural network combined with alternating trilinear decomposition (ATLD) and 3D fluorescence spectroscopy to examine the composition and content of fluorescent substances. The results showed that the BP neural network has good data compression effect, and this method can be extended for the qualitative/quantitative analysis and rapid detection of trace polycyclic aromatic hydrocarbons (PAHs) in water. Azcarate et al.\textsuperscript{23} used front-face fluorescence spectroscopy for the non-destructive evaluation of mayonnaise samples stored at two different temperatures. The results confirmed that the excitation-emission matrices (EEMs) in combination with N-way partial least square discriminant analysis (NPLS-DA) provide information related to the mayonnaise fluorescent molecular structure, facilitating the classification of samples as a function of the storage time. Wu et al.\textsuperscript{24} applied total synchronous fluorescence (TSyF) spectroscopy and convolutional neural networks (CNNs) to identify and quantify counterfeit vegetable oils. The results confirmed the feasibility of this method for vegetable oil identification. Catena et al.\textsuperscript{25} used the second-order calibration of EEMs and parallel factor analysis (PARAFAC) decomposition as an analytical approach for the detection of PAHs in food matrix (smoked tuna). The experimental results showed that the PAHs were clearly identified and quantified with decision limit (CC\textsubscript{B}) equal to 0.11 and 0.21 \textmu g/L, respectively, for benzo[a]pyrene (BaP). Lenhardt et al.\textsuperscript{26} used fluorescence spectroscopy coupled with PARAFAC and partial least squares discriminant analysis (PLS DA) for the characterization and classification of honey. The number of fluorophores present in honey, excitation and emission spectra of each fluorophore, and their relative concentration were determined using a six-component PARAFAC model. The PLS DA classification model, constructed from PARAFAC model scores, detected fake honey samples with 100% sensitivity and specificity. The honey samples were also classified using PLS DA with errors of 0.5% for linden, 10% for acacia, and nearly 20% for both sunflower and meadow mix.

In this study, three-dimensional (3D) fluorescence spectroscopy combined with pattern recognition algorithm is adopted to distinguish the composition and content of oil pollutants efficiently and accurately. Three hundred samples of kerosene, diesel, and gasoline mixed solutions with different concentrations are prepared. The first 270 samples are used as the training samples to train the network model, and the remaining 30 samples are used as test samples, which are applied to the input layer of the trained neural network. The principle component analysis is used to extract the optimal feature variables, and the correlation coefficient method is used to obtain eight groups of principal component features in the spectra. The dimension is selected as 8, and the principal component score is calculated, which is used as the input data of the extension neural network. Next, the pattern recognition method is improved, and the designed neural network has functions of both resolution and measurement. The results of neural network pattern recognition are used as the input of the concentration network. The relative fluorescence intensity, relative slope, and comprehensive background parameters are used as the input parameters, and the extension neural network is used for pattern recognition and evaluation of oil pollutants. The experimental results show that the average recognition rate of the improved pattern recognition algorithm for oil pollutants is 98.43%, and the average recovery rate of concentration is 98.67%. Further, the average time for pattern recognition is 1.53 s, while the parallel factor analysis algorithm takes 2.89 s. The experimental results show that 3D fluorescence spectroscopy combined with the extension neural network as the pattern recognition method is a reliable method for detecting oil pollutants.

**Theory**

**Principle of principal component analysis**

Principal component analysis (PCA) is a statistical method that uses orthogonal transformation to convert a set of possibly correlated variables into a set of values of linear variables, called principal components. Specifically, the fewer new variables are a linear combination of the original variables, which retain as much original statistical information as possible.\textsuperscript{27–29} From a mathematical point of view, PCA is a data dimensionality reduction technique that maps the high-dimensional data into lower dimensions.\textsuperscript{30,31}

Based on the idea of dimensionality reduction, many original variables \(x_1, x_2, \cdots, x_n\) with certain correlation are linearly combined and screened, and they are recombined to form a new set of independent comprehensive variables \(\mu_1, \mu_2, \cdots, \mu_m (m \leq p)\).\textsuperscript{32–34}

The specific steps of PCA are as follows:

1. Standardize the original data, that is, subtract each variable by the mean value and then divide it by the standard deviation to eliminate the effect of dimension.
(2) Calculate the correlation coefficient matrix \( R = (r_{ij}) \) based on the standardized data matrix \( X = (x_{ij}) \).

(3) Calculate the eigenvalues and eigenvectors of the correlation coefficient matrix \( R \) to determine the number of principal components. Firstly, the characteristic equation \( |\lambda I - R| = 0 \) is solved. Then, the eigenvectors \( e(i = 1, 2, \ldots, p) \) corresponding to these eigenvalues \( \lambda_i \) are obtained.

(4) Select the appropriate number of principal components for final analysis.

(5) The standardized fluorescence spectrum data of each sample to be tested are used to obtain the principal components, and the scores of each principal component of each sample \( F_i \) are calculated. The comprehensive score \( F = \sum_{j=1}^{m} F_i \cdot d_i \) is calculated by using the sum of variance contribution rate \( d_i \) as the weight, and each score is the quantitative descriptors of oil components with different concentrations in the sample.

**Pattern recognition based on extension neural network**

The standard \( BP \) network structure is used as the structure of the extension neural network model, and the expected output area is \( V^c \), which is the expected output corresponding to the class \( c \) sample. If the network training results of all kinds of samples fall within the corresponding expected output area, the classification is completed.

The learning algorithm of the extension neural network model is based on the principle of \( BP \) error. The expected output region \( V^c \) of the type \( c \) mode is a hypercube with \( (Y_{1c}^c, Y_{2c}^c, Y_{3c}^c) \) as the center and \( WD^c \) as the half side length. When \( L = 2 \), the expected output region \( V^2 \) is a square in 2D space. Assuming that the \( k \)th sample belongs to the \( c \)th class, the actual output of the network is \( (y_{1k}, y_{2k}, \ldots, y_{4k}) \), and \( V_k \) is the expected output area corresponding to the \( k \)th sample, \( V_k = V^c \). Then, the error function of the \( i \)th \( (i = 1, 2, \ldots, L) \) output unit corresponding to the \( k \)th sample is

\[
E(y_k, V_k) = \begin{cases} 
0, & y_{ki} - WD^c \leq y_{ki} \leq y_{ki} + WD^c, \\
\frac{1}{2}(y_{ki} - Y_{1c}^c + WD^c)^2, & y_{ki} < y_{ki} - WD^c, \\
\frac{1}{2}(y_{ki} - Y_{1c}^c - WD^c)^2, & y_{ki} > y_{ki} + WD^c,
\end{cases}
\]

(1)

Three-dimensional data and second-order calibration

A three-dimensional spectral image consisting of emission wavelength and excitation wavelength as \( X \)-axis, \( Y \)-axis and fluorescence intensity as \( Z \)-axis is called a three-dimensional fluorescence spectrum (EEM, Excitation-Emission Matrix). The three-dimensional fluorescence spectrum mainly reflects the shape of the fluorescence spectrum of the substance and the gradient of the fluorescence intensity from the macroscopic aspect. It is a three-dimensional characteristic curve that can simultaneously obtain the fluorescence intensity and excitation and emission wavelength changes. The fingerprint can quickly and accurately determine the fluorescence intensity corresponding to the specific excitation-emission wavelength, and Rayleigh scattering and Raman scattering can also be easily distinguished in the fingerprint. Accurately Locating the Optimum Emission Wavelength of Matter by Emission Spectra. The excitation spectra accurately determines the optimal excitation wavelength of the substance.

Three-dimensional fluorescence spectroscopy has become a hot choice for many researchers to detect petroleum pollutants due to its high sensitivity, good selectivity, rich information content and non-destructive material structure. Compared with the two-dimensional spectrum, the three-dimensional spectrum can more completely reflect all the spectral information contained in the mineral oil spectrum, which enables the three-dimensional fluorescence analysis method to better realize the qualitative and quantitative analysis of the substance.

With the development of sophisticated instruments, the understanding of the 3D data is gradually becoming mature, and the 3D data and second-order calibration method are increasingly applied for the analysis of complex chemical systems. A data matrix can be generated by a single measurement on an analytical sample, and a set of matrices can be obtained by simultaneously or sequentially measuring a number of analytical samples. Thus, a 3D data matrix can be obtained by combining multiple matrices. The second-order correction method is a method for analyzing a 3D data matrix. Tucker \(^{36} \) proposed the three-mode PCA model (called the Tucker3 model) for the processing of 3D data matrix. Its essence is to decompose the 3D data matrix \( X \) into three load matrices \( A, B, C \), and a 3D core matrix \( G \). The PARAFAC algorithm is a second-order calibration method, and the alternating least squares method is used to decompose the tri-line model. Its goal is to minimize the residual summation. \(^{36} \)

\[
\sigma = \sum_{i=1}^{J} \sum_{j=1}^{K} \sum_{k=1}^{L} e_{ijk} = \sum_{i=1}^{J} \sum_{j=1}^{K} \sum_{k=1}^{L} (x_{ijk} - \sum_{n=1}^{F} a_{ij} b_{jk} c_{nk})^2
\]

(2)

where \( \sigma \) is the residual sum of squares, and \( F \) is the number of factors selected by the parallel factor method. \( x_{ijk} \) is an element of the 3D data matrix \( X \); \( a_{ij} \) is an element of the load matrix \( A \); \( b_{jk} \) is an element of the load matrix \( B \); \( c_{nk} \) is an element of the load matrix \( C \); \( e_{ijk} \) is an element of the residual matrix \( E \).
Experiment

The Hitachi F-7000 fluorescence spectrometer has been adopted as an experimental instrument, which can quickly complete 3D spectral scanning. The voltage of photomultiplier (PMT) was 400 V, the scanning rate was 12,000 nm/min, the scanning step was 5 nm, the incident slit was 10 nm, and the exit slit was 10 nm. The excitation wavelength range was 250–400 nm, and the emission wavelength range was 270–500 nm. The starting point of the emission scanning wavelength is always 20 nm behind the excitation wavelength to fully avoid the interference of Rayleigh scattering spectrum. The algorithm is implemented on MATLAB8.0 and above.

A solution of carbon tetrachloride and oil pollutants was prepared. The ratio of the oil substance and the carbon tetrachloride was 1:1000, which was used to gradually dilute 300 samples with different concentrations. The samples no. 1–270 samples were used as the training samples, and the samples no. 271–300 samples were used as the test samples. The concentrations of difficult oil pollutants tested are shown in Table 1. The realization process of oil component detection in petroleum oil in Figure 1.

The 3D fluorescence spectrum of sample no. 4, which is a solution of diesel and gasoline, is shown in Figure 2. It can be seen that although the fluorescence intensity of each oil is different, the spectra of the two mineral oils are seriously overlapped, and it is difficult to realize spectral distinction and concentration prediction by chemical methods. Further, the 3D spectrum of other samples cannot distinguish between the components, so they are not shown here.

Results and discussion

Improved pattern recognition

Firstly, the original fluorescence spectral data were standardized, and the correlation coefficient matrix was calculated. The dimension of the feature spectrum was selected as 8. Secondly, the optimal feature variables, which could reflect the complete features of fluorescence spectrum, were selected. The dimension of the feature space was compressed to reduce the amount of calculation, which was beneficial for selecting the feature with the largest amount of information and with the most significant effect on the spectrum classification. The correlation coefficients between the parameters were calculated. Finally, the principal component, which can be regarded as the characteristic spectrum of the original fluorescence spectrum of the sample, was selected. The extracted feature vectors are listed in Table 2.

The load of a variable is defined as the coefficient of the variable in the linear combination equation multiplied by the square root of the corresponding eigenvalues of the principal component, but the coefficient

Table 1. Concentration of oil pollutants to be tested.

| Sample | Diesel (mg/L) | Gasoline (mg/L) | Kerosene (mg/L) | Sample | Diesel (mg/L) | Gasoline (mg/L) | Kerosene (mg/L) |
|--------|--------------|----------------|-----------------|--------|--------------|----------------|----------------|
| 1      | 100          | 0              | 0               | 156    | 50           | 25             | 0              |
| 2      | 0            | 100            | 0               | 157    | 75           | 50             | 0              |
| 3      | 80           | 10             | 0               | 158    | 30           | 70             | 0              |
| 4      | 60           | 20             | 0               | 159    | 20           | 80             | 0              |
| 5      | 50           | 30             | 0               | 160    | 150          | 0              | 500            |
| 6      | 500          | 0              | 50              | 161    | 250          | 0              | 350            |
| 7      | 400          | 0              | 100             | 162    | 350          | 0              | 150            |
| 8      | 300          | 0              | 200             | 163    | 450          | 0              | 75             |
| 9      | 200          | 0              | 300             | 164    | 50           | 40             | 65             |
| 10     | 100          | 0              | 400             | 165    | 25           | 50             | 50             |
| ...    | ...          | ...            | ...             | ...    | ...          | ...            | ...            |
| 151    | 70           | 25             | 5               | 296    | 60           | 25             | 25             |
| 152    | 60           | 10             | 25              | 297    | 80           | 10             | 10             |
| 153    | 30           | 20             | 50              | 298    | 65           | 30             | 35             |
| 154    | 10           | 25             | 50              | 299    | 100          | 10             | 50             |
| 155    | 5            | 50             | 80              | 300    | 150          | 65             | 35             |

Figure 1. The realization process of oil component detection in petroleum oil.
itself is often called the load. The larger the load, the more the similarity of this variable with the main component. Therefore, the load can be regarded as the correlation between the variable and the principal component. A sample corresponding to a primary component is called the score by a combined calculation. The network input data is the main component score, as listed in Table 3.

The above principal component score data were input into the network as new data. The cross-validation method was used to avoid the occurrence of over-fitting in the classification process. Under the premise of having enough information, the top five characteristic parameters were selected (except concentration information). The number of input nodes of the network model was set to 5, and the number of output nodes, that is, the number of refined oil types, was set to 3. In the extension neural network, the initial weight is directly related to the training results. Under the initial weight equalization, the training samples can be trained by the network, where the loop iteration in the learning algorithm generates training errors, and the training results of the network model represent the approximate sample and the desired output. The neural network can be used for both accurate value calculation and pattern recognition. When used for pattern recognition, its output node number is related to the number of intensive points. If there are two (three) types, two (three) nodes can be used. Accordingly, the three classes can be expressed as (1,0,0), (0,1,0), and (0,0,1), that is, the expected output is (D1, D2, D3). The training results and expected output of the network model are listed in Table 4. The pattern recognition error curve is shown in Figure 3.

After training the network model with the training samples, the data of the test samples were input into the trained neural network, and the input parameters included the concentration information (relative fluorescence intensity, relative slope, comprehensive background parameters) for pattern recognition and measurement of oil pollutants. In the process of concentration measurement, the output value of the

---

**Table 2.** Characteristic feature extraction of 3D fluorescence spectrum of oil pollutants based on PCA.

| Sample | Feature vectors |
|--------|-----------------|
|        | V1   | V2   | V3   | V4   | V5   | V6   | V7   | V8   |
| 1      | 0.2365| 0.1250| 0.5540| 0.6530| 0.8784| 0.8203| 0.6356| 0.4957|
| 2      | 0.2732| 0.1475| 0.5758| 0.6320| 0.7705| 0.8042| 0.6241| 0.4537|
| 3      | 1.0000| 0.1260| 0.0002| 0.6520| 1.0000| 0.8430| 0.5475| 0.4024|
| 158    | 0.9876| 0.0035| 0.0341| 0.6805| 0.8285| 0.7473| 0.3684| 0.4837|
| 159    | 0.9350| 0.5425| 0.5630| 0.4980| 0.7674| 0.7857| 0.7894| 0.5352|
| 160    | 0.5645| 0.6630| 0.5209| 0.5350| 0.7785| 0.7432| 0.5785| 0.4347|
| 297    | 0.6754| 0.5744| 0.7735| 0.6357| 0.7752| 0.6530| 0.5647| 0.4586|
| 298    | 0.8563| 0.3862| 0.4053| 0.5745| 0.5586| 0.7358| 0.8561| 0.7794|
| 299    | 0.9450| 0.7634| 0.8096| 0.6533| 0.5429| 0.7850| 0.4857| 0.8336|
| 300    | 0.9742| 0.6823| 0.6680| 0.7750| 0.8375| 0.4972| 0.5377| 0.4792|

---

**Figure 2.** 3D fluorescence spectrum of mixed solution of 0#Diesel and 97#Gasoline.
pattern recognition network was used as the weight coefficient of the relative slope (the change in the slope value of the relationship curve between concentration and fluorescence intensity of the sample). The output results of the network model for the test sample are listed in Table 5. The statistical data of corresponding characteristics are listed in Table 6. The extension neural network was used as the pattern recognition method, and the concentration measurement process took 1.53 s.

Parafac algorithm for the detection of oil pollutants

The PARAFAC algorithm was applied for the analysis of oil pollutants. The kernel consistency diagnosis method and the residual sum of squares method were used to jointly estimate the number of factors. When the number of factors was 3, the kernel consistent coefficient decreased significantly, and the residual sum of squares also decreased. In this study, the number of factors was selected as 2. The analysis results of the PARAFAC model for the mixed solution sample are presented in Figures 4 and 5. Figure 4 shows a comparison between the theoretical and experimentally measured fluorescence excitation spectrum, and Figure 5 shows a comparison between the theoretical and experimentally measured fluorescence emission spectrum.

| Sample | Score data of principal component |
|--------|-----------------------------------|
|        | Score 1 | Score 2 | Score 3 | Score 4 | Score 5 | Score 6 | Score 7 | Score 8 |
| 1      | 0.0978  | -0.3609 | -0.3164 | 0.1689  | -0.1258 | 0.3814  | 0.0120  | 0.0695  |
| 2      | 1.6253  | 0.3849  | -0.1723 | -0.0084 | 0.0164  | -0.0121 | 0.0340  | -0.0064 |
| 3      | -0.1485 | -0.2873 | -0.2382 | -0.0865 | -0.0375 | -0.0524 | 0.0371  | 0.0877  |
| 4      | -0.3062 | -0.1732 | -0.2835 | -0.0793 | -0.0392 | -0.1742 | 0.0288  | -0.0739 |
| ...    | ...     | ...     | ...     | ...     | ...     | ...     | ...     | ...     |
| 158    | -0.2917 | 0.4332  | 0.0758  | 0.1002  | -0.0302 | -0.0248 | 0.0357  | -0.0034 |
| 159    | -0.5306 | 0.5395  | 0.0528  | 0.0382  | 0.0283  | 0.0421  | 0.0242  | -0.0004 |
| 160    | -0.4126 | 0.3745  | 0.0623  | 0.0847  | -0.0121 | -0.0086 | 0.0324  | 0.0473  |
| ...    | ...     | ...     | ...     | ...     | ...     | ...     | ...     | ...     |
| 297    | -0.5842 | 0.5377  | 0.3265  | -0.0857 | 0.0379  | 0.04569 | 0.0385  | -0.0057 |
| 298    | -0.6541 | 0.4973  | 0.0875  | 0.1562  | -0.0302 | -0.0579 | 0.0430  | 0.0062  |
| 299    | 0.4726  | -0.5873 | 0.2382  | 0.0854  | -0.0415 | 0.0325  | 0.0279  | -0.0741 |
| 300    | -0.6742 | 0.3529  | 0.0774  | 0.0745  | -0.0533 | 0.04983 | 0.0382  | -0.0057 |

| Sample | Training results | Desired output |
|--------|-----------------|----------------|
|        | R1   | R2   | R3   | D1   | D2   | D3   |
| 1      | -0.0002 | 0.0001 | 0.9998 | 0    | 0    | 1    |
| 2      | 0.0000  | 0.0002 | 1.0001 | 0    | 0    | 1    |
| 3      | 0.0000  | 0.0000 | 1.0001 | 0    | 0    | 1    |
| ...    | ...   | ...   | ...   | ...  | ...  | ...  |
| 110    | 0.0001  | 0.9997 | 0.0002 | 0    | 1    | 0    |
| 111    | 0.0002  | 0.9999 | -0.0001 | 0    | 1    | 0    |
| 112    | 0.0002  | 0.9996 | -0.0002 | 0    | 1    | 0    |
| ...    | ...   | ...   | ...   | ...  | ...  | ...  |
| 268    | 1.0000  | 0.0001 | 0.0000 | 1    | 0    | 0    |
| 269    | 1.0001  | 0.0002 | 0.0000 | 1    | 0    | 0    |
| 270    | 1.0000  | 0.0001 | -0.0002 | 1    | 0    | 0    |

Figure 3. Error curve of pattern recognition.
Table 7 presents the prediction results and recovery data of test samples obtained by the PARAFAC algorithm. The experimental results show that the average pattern recognition rate of the oil pollutants based on the PARAFAC model is 93.1%. The average recovery rates of diesel and gasoline are 92.19% and 89.73%, respectively, and the average analysis time of PARAFAC model is 2.89 s. Both the concentration recovery rate and the time consumption confirm that the extension neural network is more effective than the PARAFAC model.

**Conclusion**

Combining the advantages of the data representation of PCA and the pattern recognition of extension neural network for mixed component system, the refined oil products were effectively identified and measured. The principal component analysis is used to extract the optimal feature variables, and the correlation coefficient method is used to obtain eight groups of principal component features in the spectra. The dimension is selected as 8, and the principal component score is calculated, which is used as the input data of the extension neural network. Next, the pattern recognition method is improved, and the designed neural network has functions of both resolution and measurement. The results of neural network pattern recognition are used as the input of the concentration network. The relative fluorescence intensity, relative slope, and comprehensive background parameters are used as the input parameters, and the extension neural network is used for pattern recognition and evaluation of oil pollutants. The experimental results show that the average...
The recognition rate of the improved pattern recognition algorithm for oil pollutants is 98.43%, and the average recovery rate of concentration is 98.67%. The average pattern recognition rate of the oil pollutants based on the PARAFAC model is 93.1%. The average recovery rates of diesel and gasoline are 92.19% and 89.73%, respectively.

Further, the average time for pattern recognition is 1.53 s, while the parallel factor analysis algorithm takes 2.89 s. The comparison between the theoretical and experimental characteristic fluorescence excitation and emission spectra was used to verify that the extension neural network is a very powerful tool for spectral data analysis.

In this paper, the Pattern Recognition of Extension Neural Network still has the shortcomings of easy to fall into local optimum and slow convergence in the application. In the future research, it still needs to be improved to improve the recognition accuracy and efficiency, and the recognition effect in the fields of health care and food safety is studied to expand its application fields.

**Acknowledgements**

Thanks to the financial support provided by the National Natural Science Foundation of China (Nos. 61771419, 21807034), and the Natural Science Foundation of Hebei Province of China (Nos. F2019209323, F2019209443, F2019209599). It also grateful to North China University of Science and Technology (Hebei, China), for providing fluorescence spectrometer instruments.

**Declaration of conflicting interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.
Funding
The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported in part by the National Natural Science Foundation of China under Grant 61771419 and 21807034, and the Natural Science Foundation of Hebei Province of China under Grant F2019209323, F2019209443, and F2019209599.

ORCID iDs
Pengfei Cheng https://orcid.org/0000-0002-4903-0871
Yanping Zhu https://orcid.org/0000-0001-7841-7998
Jinyan Pan https://orcid.org/0000-0002-7736-721X

References
1. Cui Y, Kong D, Kong L, et al. Excitation emission matrix fluorescence spectroscopy and parallel factor framework-clustering analysis for oil pollutants identification. Spectrochim Acta A: Mol Biomol Spectrosc 2021; 253(5): 119586.
2. Pourhoseini SH, Namvar-Mahboub M, Hosseini E, et al. A comparative exploration of thermal, radiative and pollutant emission characteristics of oil burner flame using palm oil biodiesel-diesel blend fuel and diesel fuel. Energy 2021; 217(4): 119338.
3. Cui Y, Kong D, Kong L, et al. A novel strategy for identifying oil pollutants based on excitation-emission matrix fluorescence spectroscopy and Zernike moments. IEEE Access 2020; 8: 17999–18006.
4. Shang F, Wang Y, Wang J, et al. Determination of three poly cyclic aromatic hydrocarbons in tea using four-way fluorescence data coupled with third-order calibration method. Microchem J 2019; 146(5): 957–964.
5. Rinot O, Borisover M, Levy GJ, et al. Fluorescence spectroscopy: A sensitive tool for identifying land-use and climatic region effects on the characteristics of water-extractable soil organic matter. Ecol Indic 2021; 121(5): 107103.
6. Al Riza DF, Kondo N, Rotich VK, et al. Cultivar and geographical origin authentication of Italian extra virgin olive oil using front-face fluorescence spectroscopy and chemometrics. Food Control 2021; 121(5): 107604.
7. Wang J, Wang S, Liu N, et al. A detection method of two carbamate pesticides residues on tomatoes utilizing excitation-emission matrix fluorescence technique. Microchem J 2021; 164(4): 105920.
8. Mirnaghi FS, Soucy N, Hollebone BP, et al. Rapid fingerprinting of spilled petroleum products using fluorescence spectroscopy coupled with parallel factor and principal component analysis. Chemosphere 2018; 208(5): 185–195.
9. Araújo KC, Barreto MC, Siqueira AS, et al. Oil spill in northeastern Brazil: application of fluorescence spectroscopy and PARAFAC in the analysis of oil-related compounds. Chemosphere 2021; 267(5): 129154.
10. Kong D, Cui Y, Kong L, et al. Classification of oil pollutants based on excitation-emission matrix fluorescence spectroscopy and two-dimensional discriminant analysis. Spectrochim Acta Mol Biomol Spectrosc 2020; 228(5): 117799.
11. Wen X, Chen G and Liu Y. Submersible High Precision PAHs Detection System for Marine Oil Spills Disaster Management[C]. In: 2018 IEEE 8th International Conference on Underwater System Technology: Theory and Applications (USYS). IEEE, 2019.
12. Sergiel I, Pohl P, Biesaga M, et al. Suitability of three-dimensional synchronous fluorescence spectroscopy for fingerprint analysis of honey samples with reference to their phenolic profiles. Food Chem 2014; 145(5): 319–326.
13. Xu J, Liu XF and Wang YT. A detection method of vegetable oils in edible blended oil based on three-dimensional fluorescence spectroscopy technique. Food Chem 2016; 212(5): 72–77.
14. Nakaya Y, Nakashima S, Moriizumi M, et al. Three dimensional excitation-emission matrix fluorescence spectroscopy of typical Japanese soil powders. Spectrochim Acta Mol Biomol Spectrosc 2020; 233(5): 118188.
15. Lee YK, Hong S and Hur J. Copper-binding properties of microplastic-derived dissolved organic matter revealed by fluorescence spectroscopy and two-dimensional correlation spectroscopy. Water Res 2021; 190(5): 116775.
16. Liu H, Saito Y, Riza DFA, et al. Rapid evaluation of quality deterioration and freshness of beef during low temperature storage using three-dimensional fluorescence spectroscopy. Food Chem 2019; 287: 369–374.
17. Carvalho DG, Ranzan L, Trierweiler LF, et al. Determination of the concentration of total phenolic compounds in aged cachaça using two-dimensional fluorescence and mid-infrared spectroscopy. Food Chem 2020; 329(5): 127142.
18. Xie W, Zhang S, Ruan L, et al. Evaluating soil dissolved organic matter extraction using three-dimensional excitation-emission matrix fluorescence spectroscopy. Pedosphere 2017; 27(5): 968–973.
19. Zhang S, Chen Z, Wen Q, et al. Assessment of maturity during co-processing of penicillin mycelial dreg via fluorescence excitation-emission matrix spectra: Characteristics of chemical and fluorescent parameters of water-extractable organic matter. Chemosphere 2016; 155: 358–366.
20. Li W, Liang C, Dong L, et al. Accumulation and characteristics of fluorescent dissolved organic matter in loess soil-based subsurface wastewater infiltration system with aeration and biochar addition. Environ Pollut 2021; 269: 116100.
21. Zhao X, Che X, Wang S, et al. Rapid detection of quinolones in water based on fluorescence spectrometry and BLLS/RBL. Acta Opt Sin 2020; 40(9): 208–214.
22. Wang Y, Yan Z and Fengkai S. Determination of poly cyclic aromatic hydrocarbons in water by BP neural network combined with ATLD and three dimensional fluorescence spectrometry. Spectroscopy and Spectral Analysis 2019; 39(11): 3420–3425.
23. Azcarate SM, Teglia CM, Karp F, et al. A novel fast quality control strategy for monitoring spoilage on mayonnaise based on modeling second-order front-face fluorescence spectroscopy data. Microchem J 2017; 133(5): 182–187.
24. Wu X, Zhao Z, Tian R, et al. Exploration of total synchronous fluorescence spectroscopy combined with pre-trained convolutional neural network in the identification and quantification of vegetable oil. Food Chem 2021; 335: 127640.
25. Catena S, Sanllorente S, Sarabia LA, et al. Unequivocal
identification and quantification of pahs content in ternary synthetic mixtures and in smoked tuna by means of excitation-emission fluorescence spectroscopy coupled with PARAFAC. *Microchem J* 2020; 154: 104561.

26. Lenhardt L, Bro R, Zeković I, et al. Fluorescence spectroscopy coupled with PARAFAC and PLS DA for characterization and classification of honey. *Food Chem* 2015; 175: 284–291.

27. Ouyang Y. Evaluation of river water quality monitoring stations by principal component analysis. *Water Res* 2005; 39(12): 2621–2635.

28. Lucas L and Jauzein M. Use of principal component analysis to profile temporal and spatial variations of chlorinated solvent concentration in groundwater. *Environ Pollut* 2008; 151(1): 205–212.

29. Olsen RL, Chappell RW and Loftis JC. Water quality sample collection, data treatment and results presentation for principal components analysis – literature review and Illinois River watershed case study. *Water Res* 2012; 46(9): 3110–3122.

30. Akinnuwesi BA, Macaulay BO and Aribisala BS. Breast cancer risk assessment and early diagnosis using principal component analysis and support vector machine techniques. *Inform Med Unlocked* 2020; 21(5): 100459.

31. Bellandi G, Weijers S, Gori R, et al. Towards an online mitigation strategy for N2O emissions through principal components analysis and clustering techniques. *J Environ Manag* 2020; 261: 110219.

32. Farrugia J, Griffin S, Valdramidis VP, et al. Principal component analysis of hyperspectral data for early detection of mould in cheeselets. *Current Research in Food Science* 2021; 4: 18–27.

33. Li WM, Zhu HP, Ding LY, et al. Structural damage recognition by grouped data based on Principal Component Analysis theory. *Autom Constr* 2012; 22: 258–270.

34. El-Saeid RH, Abdel-Salam Z, Pagnotta S, et al. Classification of sedimentary and igneous rocks by laser induced breakdown spectroscopy and nanoparticle-enhanced laser induced breakdown spectroscopy combined with principal component analysis and graph theory. *Spectrochim Acta Part B* 2019; 158(5): 105622.

35. Tucker LR. Some mathematical notes on three-mode factor analysis. *Psychometrika* 1966; 31(3): 279–311.

36. Wang J, Wang S, Shang F, et al. Missing data recovery combined with Parallel factor analysis model for eliminating Rayleigh scattering in the process of detecting pesticide mixture. *Spectrochim Acta A: Mol Biol Spectrosc* 2020; 232: 118187.