Research on FT-IR spectrum data mining of diesel engine lubricating oil

Jing Li¹, Hongxiang Tian²*, Tingfeng Ming¹, Yunling Sun¹ and Shuai Zhang¹

¹College of Power Engineering, Naval University of Engineering, Wuhan, Hubei, 430033, China
²School of Mechanical and Electrical Engineering, Wuhan Donhu University, Wuhan, Hubei, 430212, China
*Corresponding author’s e-mail: hxtianwuhan@aliyun.com

Abstract. Aiming at the difficulties in quantitatively describing the pollution component information contained in FT-IR spectral data of diesel engine lubricants, wavelet transform and fractal box dimension are used to process FT-IR spectral data and get fractal dimension curve. Multiple peaks can be obtained by analyzing the curve. Two groups of lubricating oil samples with water content of 0.10%, 0.22%, 0.44% and 0.88% respectively and with ethylene glycol content of 0.10%, 0.22%, 0.44% and 0.88% respectively. For the above two groups of oil samples, the corresponding relation between peak position and contaminant ratio can be obtained, and a new method based on FT-IR prediction of contaminate ratio is feasible.

1. Introduction

Fourier infrared (FT-IR) spectroscopy monitors part of the information about contaminants in lubricants, including water, soot, ethylene glycol, fuels and incorrect oil [1]. Different contaminants correspond to different characteristic peaks and peak areas. However, there are many bands superimposed by various absorption peaks, which may be the superposition of different absorption bands of the same compound, or the superposition of different components in a mixture spectrum. The peaks can not only distinguish overlapping sub-peaks, but also have quantitative information [2]. A novel method for measuring antimony sulfide content rapidly based on Gaussian-peak fitting of Raman spectroscopy was proposed by Jiang Xi-ping [3]. To analyze the components and their within the tested material more precisely, quasi-Newton method can be applied to separate overlapping peaks of spectra plotted with discrete data by Zhu Qiang [4]. The tangent intersection point method is used to obtain the peak points of the overlapping peaks [5].

In this paper, the fractal dimension curve is obtained by processing data through wavelet transform and fractal box dimension. The relation between different contaminants proportion and peak location and peak height was found, and the fitting curve was established.

2. Data sources and algorithms

2.1 Analytical instrument and data

Infrared spectrometer was used in the experiment, and the spectrum range detected was 4000-650 cm⁻¹. The analyzer can monitor the concentration of water and ethylene glycol in lubricating oil in real time.
Using the electronic balance and the diesel oil type of Great Wall CD40-15W, two groups of oil samples were respectively provided. The water content of the first group oil samples respectively were 0.10%, 0.22%, 0.44% and 0.88%. The ethylene glycol content of the second group oil samples respectively were 0.10%, 0.22%, 0.44% and 0.88%. The above two groups of oil samples were analysed by FT-IR spectrometer.

2.2 Algorithms

(1) Scatter diagram is made with the wave number of the collected data as the abscissa and the absorbance as the ordinate. The fitting curve $f_i(x)$ of data is obtained by connecting all data points smoothly, as shown in figure 1 below. The interval between adjacent data points in the original data is not uniform, so the data should be pre-processed before the subsequent data analysis.

Figure 1. Raw data.

(2) Data points are selected at the same interval. In this paper, the data are resampled with the abscissa of 500 as the starting point and 5 as the sampling interval, and then the resampled data's ordinate is normalized to [0,1], and then the smooth connection of the resampled data to each sampling point is followed to obtain the fitted curve $f_2(x)$ after resampling. In order to facilitate data analysis in the later period, the fitting curve starts with 1.

(3) According to the formula, the continuous wavelet transform of $f_2(x)$ is performed to obtain the wavelet coefficient function $W_f(m,n)$. Where, the x-coordinate $m$ is the time shift factor. The ordinate $n$ is the scale factor of the signal, $\phi_{m,n}(t)$ is the scale after wavelet base, Using db5 wavelet generating function.

$$W_f (m,n) = \langle f(t), \phi_{m,n}(t) \rangle = \frac{1}{\sqrt{m}} \int_R f(t) \phi \left( \frac{t-n}{m} \right) dt$$

(4) In view of that characteristic of each fractal dimension calculation method, this paper use a fractal box dimension processing wavelet data to obtain a fractal dimension curve. After the wavelet coefficient function $W_f(m,n)$ is obtained, according to the calculation method of box dimension, a rectangle of certain area is selected for each point to process the one-dimensional signal $W_f(m,n)$, that is, each column in the wavelet coefficient function $W_f(m,n)$. The fractal dimension curve $D_f(n)$ is calculated by using the following formula, where $N(\varepsilon_j)$ is the minimum number of squares needed to cover the one-
The fractal dimension curve $f_3(x)$ is analyzed to obtain the position and corresponding peak values of $n$ peaks $P_1(x_1, y_1), P_2(x_2, y_2), \ldots, P_n(x_n, y_n)$.

The function between peak position and contaminant proportion is established and the correlation coefficient is calculated.

### 3. Results and discussion

The position and peak values of four wave peaks can be obtained by considering the measured wave number segment. See table 1.

| Water content | peak 1 | peak 2 | peak 3 | peak 4 |
|---------------|--------|--------|--------|--------|
| Position      | Max    | Position | Max | Position | Max |
| 0.00%         | 461    | 1.321  | 1011  | 1.18    | 1381 | 1.155 | 1926 | 1.304 |
| 0.10%         | 461    | 1.321  | 986   | 1.16    | 1371 | 1.216 | 1926 | 1.258 |
| 0.22%         | 461    | 1.321  | 991   | 1.127   | 1376 | 1.209 | 1926 | 1.237 |
| 0.44%         | 461    | 1.331  | 941   | 1.156   | 1521 | 1.238 | 1926 | 1.231 |
| 0.88%         | 456    | 1.32   | 946   | 1.161   | 1511 | 1.145 | 1926 | 1.262 |

Similarly, the spectra of glycol with different contaminant ratios were shown in figure 2. It can be seen that the peaking effect of ethylene glycol is better than that of water. As the percentage of ethylene glycol contaminates increases, the corresponding abscissa position moves to the right, that is, the wave number is large, and the peak height drops. The specific value is shown in table 2.

| ethylene glycol | peak 1 | peak 2 | peak 3 |
|-----------------|--------|--------|--------|
| Position        | Max    | Position | Max | Position | Max |
| 0.00%           | 1037   | 1.1972  | 1051  | 1.3713 | 1075 | 1.3728 |
| 0.10%           | 1037   | 1.2451  | 1051  | 1.3474 | 1076 | 1.3617 |
| 0.22%           | 1037   | 1.2465  | 1051  | 1.3158 | 1078 | 1.3681 |
| 0.44%           | 1037   | 1.2658  | 1054  | 1.2635 | 1080 | 1.2838 |
| 0.88%           | 1040   | 1.2206  | 1061  | 1.3560 | 1083 | 1.2134 |

Taking the third peak point as an example, the broken line diagram was made between the percentage content of ethylene glycol and the peak point position and peak value of the third peak point, as shown in figure 3. It can be seen that when the percentage content of ethylene glycol is greater than 0.2%, the peak value shows a certain correlation with the percentage content.
4. Conclusion

By using wavelet transform and fractal box dimension, the FT-IR spectral signal can be analyzed, and the position and maximum value of each peak can be extracted. This information is correlated with the percentage content of oil contaminants.

References

[1] ASTM E2412. (2010) Condition monitoring of in-service lubricants by trend analysis using Fourier transform infrared (FT-IR) Spectrometry.

[2] LI Y.Y., XU D.G, GUI W. (2017) A novel method for measuring antimony sulfide content based on Gaussian-peak fitting of Raman spectroscopy. Spectroscopy and spectral analysis, 37(12):3743-3748.

[3] JIANG X.P., WU F.H, XIU L.C. (2011) Application of Quasi-Newton Method for separating overlapping peaks of spectra. Journal of Zhengzhou University, 32(6): 113-116.

[4] Zhu Q., Zhang R., Yu J.C. (2017) Novel Algorithm for Decomposition of overlapped chromatographic peaks based on Function Split Iteration, Chinese Journal of Vacuum Science and Technology, 37(5): 544-547.

[5] Ye G.Y., Xu K.J. (2015) New decomposition peak methods of overlapping chromatographic peaks based on similarity principle, 36(2):439-445.