

# Reflectance Spectroscopy Model for Pollutant Estimation in Lubricant

Panjie Lv<sup>a</sup>, Guangjun Tian<sup>b</sup>

School of Electrical Engineering, Yanshan University, Qinhuangdao 066004, China;

<sup>a</sup>LPJ322@126.com, <sup>b</sup>t0208@126.com

**Keywords:** Reflectance spectroscopy model; Lubricants; PCA; Correlation coefficient.

**Abstract.** Principal Component Analysis combined with Correlation Coefficient is presented to choose dominant wavelengths of lubricants reflectance spectrum in the wide band of  $220\text{ nm} \sim 780\text{ nm}$ , and the index regression model is established about oil pollution concentration and spectral reflectivity. Tests show that the regression model has strong predictive power, which is suitable for the quality estimation of lubricating oil with high pollution concentration. It provides the feasibility experimental basis for further to achieve the working lubricants pollution concentration in-situ and rapid detection of spectroscopic methodology.

## Introduction

As we all know, lubricants are widely used in various types of engines, which have a very important position in today's industry. Currently oil change methods are still more of the regular oil change with significant limitations [1]. Oil on-line monitoring can timely reflect the quality of oil through the data analysis, guaranteeing oil effectiveness and timely displacement [2]. In recent years, spectroscopy technique has been widely applied with the advantages of non-destructive, real-time, speed detecting, high accuracy and so on [3-5]. In this paper, we use reflectance spectroscopy which can realize in-situ, real-time, non-destructive measurement combined with Correlation Coefficient, Principal Component Analysis(PCA) and other feature extraction methods to achieve the estimated concentration of oil pollution.

## Experimental part

**Materials and Instruments.** We chose the pure oil and full pollution oil as experimental material. They are gradually mixed by equivalent volume method. Samples obtained with different pollution concentrations successively were 50%, 25%, 12.5%, 6.25%, 3.125%, 1.5625%, along with the full pollution oil (100%) and the pure oil (0%) loaded into the vessel arrays. The spectral measurement system employs the S2000 model miniature spectrometer.

**Experiment.** The dedicated xenon lamp light source is used to excite samples in the darkroom. After adjusting, remain unchanged the distance and geometry azimuth between the detector probe and the sample, we can just conduct spectral measurement. Due to the stability of instruments, multiple measurements for each sample repeated, and the average value serves as the reflection intensity of the sample. The measured spectrum in the range of  $220\text{ nm} \sim 780\text{ nm}$  is chosen since the restriction of instrument's End Effect. The block diagram of the system is shown in Figure 1.

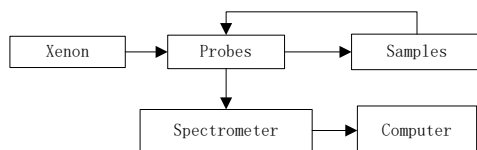


Fig. 1 Block diagram of the system

**Data Description.** Spectral data of different concentration contaminated lubricant is got through experiments. The original spectra not only contain useful reflectance information, but also contain a certain level of noise and redundant data [6]. We obtain 200 data by taking interval  $3\text{ nm}$ .

## Experimental Results and Discussion

**Calculation of Reflectivity.** We use reflectivity to represent the reflectance spectrum of the lubricants. The ratio method to calculate the spectral reflectance of the lubricants is as follows:

$$R_1(\lambda) = \frac{DN_1(\lambda)}{DN_2(\lambda)} \times R_2(\lambda). \quad (1)$$

$DN_1(\lambda)$  and  $DN_2(\lambda)$  are spectrometer output of observed target and reference target respectively;  $R_1(\lambda)$  is spectral reflectivity of the observed target;  $R_2(\lambda)$  is spectral reflectivity of reference target. Reflectivity spectra of various concentrations are obtained by formula (1), as shown in Figure 2:

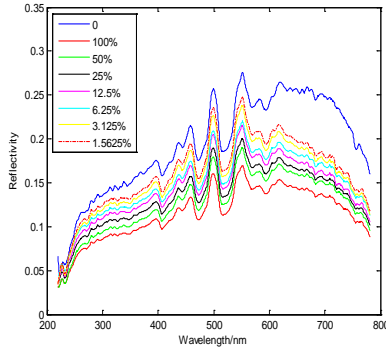


Fig.2 Reflectivity spectra

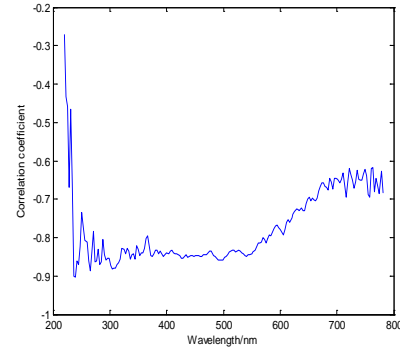


Fig. 3 Curve of correlation coefficient

### Regression Model to Select Dominant Wavelength.

**Regression Model Based on Correlation Coefficient.** We calculate the relationship degree about the two vectors of reflectivity and pollution concentration through Correlation Coefficient (CC). The correlation coefficient  $r$  is calculated by the following formula:

$$r_j = \frac{\sum_{i=1}^n (x_{i,j} - \bar{x}_j)(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_{i,j} - \bar{x}_j)^2 \sum_{i=1}^n (y_i - \bar{y})^2}}. \quad (2)$$

Where,  $x$ -Spectral reflectivity,  $\bar{x}_i = \left[ \sum_{j=1}^m x_{i,j} \right] / n$ ;  $y$ -Lubricant pollution concentration,  $\bar{y} = \left[ \sum_{i=1}^n y_i \right] / n$ .  $j=1, 2, \dots, m$ ,  $m$  is wavelength point;  $i=1, 2, \dots, n$ ,  $n$  is sample number. Matlab toolbox is used to calculate the correlation coefficient of the sample spectrum reflectivity and contaminant concentration, the correlation curve varying with wavelength is shown in Figure 3.

In Figure 3, the correlation coefficient of contaminant concentration and the reflectivity in the low frequency band is larger than the high frequency band, and the concentration is negatively correlated with reflectivity.

The maximum value of the correlation coefficient of lubricant reflectivity and pollutant concentration is corresponding to wavelength at  $240.02 \text{ nm}$ , and the Nlinfit function in the Matlab toolbox is used to establish a nonlinear regression model. The regression equation (3) of pollutant concentration  $c$  and reflectivity  $r$  is obtained. The regression curve is shown in Figure 4.

$$c = 595.4e^{-474.0r}. \quad (3)$$

**Regression Model Based on PCA.** The basic idea of PCA is to find a orthogonal coordinate system of definition of raw data the maximum variance direction. The high-dimensional data is projected onto a low dimensional space with less loss of information through the new coordinate system. The original data is represented by fewer variables[7]. Principal component(s) is selected in accordance with (accumulated) contribution rate of not less than 95%.

$$F = P \bullet T. \quad (4)$$

Matlab function Princomp is used to get the first principal component  $T_1$ , and contribution rate of  $T_1$  is 98.8330%. The first principal component score vector of each wavelength is  $P$ . According to the formula (4), the total score of the 200 wavelengths is obtained. By the PCA method to get the

maximum contribution of the wavelength is 552.48 nm. Using Matlab function Nlinfit a index regression model is established as equation (5). The regression curve is shown in Figure 5.

$$c = 2097.9e^{-148.7r} \quad (5)$$

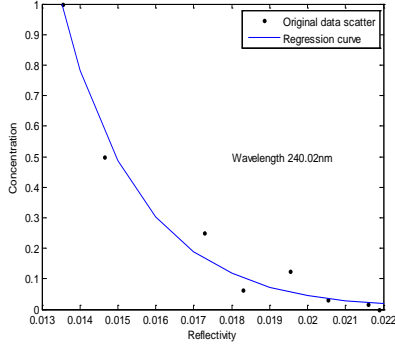


Fig. 4 Regression curve based on CC

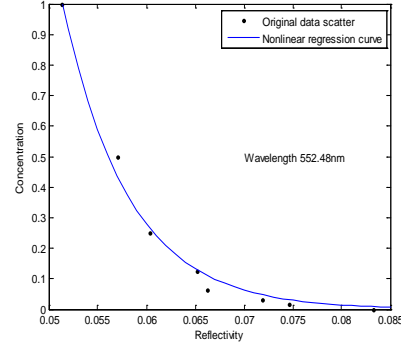


Fig. 5 Regression curve based on PCA

**Regression Model Based on PCA-CC.** The combination of PCA and CC can make the choice of dominant wavelength more rigorous and have universal significance for different materials, which is referred as PCA-CC for description convenient. Section 2.2.1 and section 2.2.2 show the two dominant wavelengths' span is large, which is enough to affect the model accuracy, and it is necessary to carry out wavelength segment analysis for taking into account the actual dominant wavelength in the supporting role. The 220 nm ~ 780 nm wavelength band is divided into three sections for analysis, and the first paragraph is 220 nm ~ 380 nm; the second is 380 nm ~ 580 nm; the third is 580 nm ~ 780 nm. These three sections are analyzed by PCA to extract characteristic wavelength, and the results are shown in Table 1.

Table 1 First three wavelengths feature based on the PCA

Method	220~380 (nm)	380~580 (nm)	580~780 (nm)
PCA	377.90	552.48	617.38
	375.14	549.84	620.28
	372.37	546.87	614.47

From Table 1, it is found that the three dominant wavelengths are near, so an intermediate wavelength of each segment is selected, which are 375.14 nm, 549.84 nm, 617.38 nm. Nonlinear regression models are established for the three wavelengths. Correlation coefficient at each of three wavelengths and the root mean square error are shown in Table 2.

Table 2 Correlation coefficient and root mean square error of dominant wavelength

Wavelength bands(nm)	Dominant wavelength(nm)	Correlation Coefficient	Root mean square error
220~380	375.14	-0.8498	0.0141
380~580	549.84	-0.8420	0.0328
580~780	617.38	-0.7591	0.0466

From Table 2, we find the 375.14 nm wavelength having the best modeling effect, so the dominant wavelength is found at 375.14 nm. By using Matlab tools, an nonlinear regression model is established as equation (6). Its regression curve is shown in Figure 6.

$$c = 5213.8e^{-280.0r} \quad (6)$$

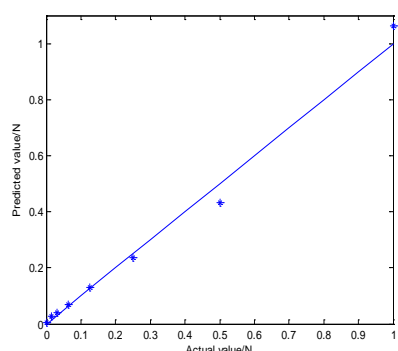
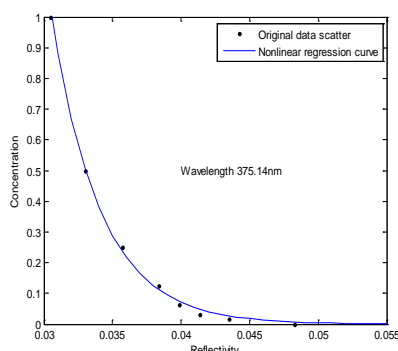


Fig. 6 Regression curve based on PCA - CC

Fig. 7 Prediction renderings based on PCA-CC

**Prediction Performance.** We test the prediction set by the above model and get the results shown in Table 3. RMSEP1 is the root mean square error of correction set; RMSEP2 is the root mean square error of prediction set.

Table 3 Comparison of prediction effect of different models

Method	Wavelength(nm)	RMSEP1	RMSEP2
CC	240.02	0.0502	0.1529
PCA	552.48	0.0318	0.0396
PCA -CC	375.14	0.0141	0.0337

Seen from Table 3, the model using PCA-CC to extract characteristic wavelength 375.14 *nm* has better prediction performance than using other methods. Its prediction curve is shown in Figure 7, seen from which, when the concentration is greater than 6%, the index regression model work well in prediction and estimation of pollution concentration of lubricant oil.

## Conclusion

The spectral reflectance experimental results show the higher contamination of the lubricants, the lower intensity of the reflected spectrum. PCA combined with Correlation Coefficient is used to select advantage wavelengths, and an index relationship model between spectral reflectance and pollution concentration is established at the wavelength 375.14 *nm*. This model has higher ability for lubricants quality estimation of higher pollution concentration. This result has a certain reference value for spectroscopic methodology of working lubricants online monitoring.

## Acknowledgements

This work was supported by the National Natural Science Foundation of China (Grant No. 20577038) and the Natural Science Foundation of Hebei Province (Grant No. D2004000195).

## References

- [1] Z.Y. Zhao, A.G. OuYang, The lubricant quality near-infrared and raman spectroscopy testing methods, East China Jiaotong University. (2012)
- [2] X.G. Han, Discussion the equipment fault diagnosis of oil monitoring, China Petroleum and Chemical Standard and Quality. 5 (2010) 278.
- [3] Gebarin Sabrin Khaled F J, Determining proper oil and filter change intervals: Can onboard automotive sensors help, Practicing Oil Analysis Magazine. 5 (2004) 17-23.
- [4] S.Z. Yang, J.Q. Hu, Developing directions of lubricants monitoring technology, Synthetic Lubricants. 3 (2011) 18-20.

- [5] D.B. Hou, J. Zhang, Water Quality Analysis by UV-Vis Spectroscopy: A Review of Methodology and Application, Spectroscopy and Spectral Analysis, 33 (2013) 1839-1844.
- [6] J. Wang, B.C. Yin, Research on data gathering restoration and compression based on sparse representation, Beijing industry university(2015)
- [7] Y.F. Yuan, Z.H. Tao. Identification of Cortex Phellodendri by Fourier-Transform Infrared Spectroscopy and Principal Component Analysis, Spectroscopy and Spectral Analysis, 31 (2011) 1258-1261.