Research on oil-water two-phase water content detection model based on near-infrared spectroscopy

Liu Ke-ke[a], Gao Guo-wang*[a], Wang Fei[a], Wu Dan[a], Wu Zhao-xue[a], Gong Yu[a]
[a]School of Electronic Engineering, Xi'an shiyou University, Xi'an, China 710065
* Corresponding author: wwgao1205@163.com

ABSTRACT

Crude oil water content is an important technical indicator in oil extraction, transportation and oil trading. Real-time online testing of crude oil water content is extremely important in estimating crude oil production and evaluating the extraction value of oil wells. At present, most of the wells at home and abroad are in the middle and late stage of development, it is difficult and inaccurate to measure under the high water content condition of crude oil, so it is necessary to adopt new detection means to improve the detection accuracy. In this paper, a study on the method of water content measurement using infrared spectroscopy was carried out. This study used S-G smoothing and normalization as the method of data pre-processing, selected the characteristic wavelengths using the continuous projection method (SPA) with a root mean square error of 4.4702, and then used partial least squares (PLS) to establish a water content detection model, and obtained a prediction root mean square error of 9.7131 and a correlation coefficient of 0.98527, which obtained a good accuracy. The feasibility of using spectroscopic detection technology to measure the water content of crude oil was demonstrated, providing a new method for oil extraction exploration and production processing.

Keywords: Crude oil water content, Near infrared spectroscopy, Continuous projection algorithm, Partial least squares, Quantitative analysis

1. INTRODUCTION

Oil is an indispensable mineral resource for economic construction and social development, and the water content of crude oil produced from most oil fields in the middle and late stages of extraction is almost all over 90% or even higher[1]. The testing of crude oil water content allows us to understand and grasp the production status of each well and each production layer, and thus develop well exploitation strategies in order to extend well life and improve recovery rates. The commonly used online measurement methods for crude oil water content include capacitance method, short-wave method, ray method and radio frequency method, etc[2]. The capacitance method is suitable for measuring crude oil with water content lower than 30%, and the stray capacitance will affect the measurement results; the short-wave method ray method is costly and easily affected by temperature and water mineralization[3]; the radio frequency method has fast response and small volume, but the measurement accuracy is low when the water content is greater than 80%.

Infrared spectroscopy is a method to determine the molecular structure of substances and identify compounds based on the different frequencies of chemical bond vibrations in molecules, which has the outstanding advantages of strong discrimination and non-destructive measurement of samples, and is now widely used in agriculture and food[6-7]. In the petrochemical field, NarveAske et al. studied and modeled the near-infrared spectra of asphaltene aggregation under high pressure; Yamate T et al[4]. used near-infrared light to detect the gas content of a single oil phase and achieved good results; Xia Bokai et al. from China University of Petroleum built a static oil water content measurement device using 960 nm LED light source[5]; Han Jian et al. from Northeast Petroleum University built a low flow rate and high water content detection device and performed flow pattern identification. Infrared spectroscopy is an indirect measurement and the accuracy of the model can be improved by collecting more spectral data of crude oil and improving the modeling algorithm. Based on these studies, the spectral data of crude oil with different water contents were collected and preprocessed by S-G (Savitzky-Golay) smoothing + normalization method, and then the feature wavelengths were extracted from the full wavelength range of 900 nm to 1700 nm using the successive projections algorithm (SPA). The extracted feature wavelengths are used as the input of Partial Least Squares regression (PLS) to build a 0-100% oil-water two-phase flow water content detection model to explore the feasibility of applying NIR spectroscopy to actual oil field extraction.

Third International Symposium on Computer Engineering and Intelligent Communications (ISCEIC 2022), edited by Xianye Ben, Proc. of SPIE Vol. 12462, 124620L · © The Authors. Published under a Creative Commons Attribution CC-BY 3.0 License · doi: 10.1117/12.2660957

2. INFRARED SPECTROSCOPY DETECTION PRINCIPLE OF CRUDE OIL WATER CONTENT

Infrared spectroscopy determines the composition by detecting the absorption intensity of a substance at a specific wavelength, in accordance with the linear superposition law of optics and the Lambert-Beer Law (Lambert-Beer Law)[8]. Lambert-Beer Law can be described as follows: when a monochromatic light is transmitted through a solution, the absorbance A is proportional to the concentration of the solution if the thickness of the solution is fixed, as shown in Equation (1).

$$A = \varepsilon cL \tag{1}$$

where L is the thickness of the solution, c is the concentration of the substance, and ε is the molar absorbance coefficient.

As shown in Figure 1, the thickness of each solution in the oil-water mixture is L_a, L_b, respectively, the amount of substance concentration c and the molar absorption coefficient ε are measurable constants expressed as K_a, K_b, I_0 is the incident light intensity, the transmitted light intensity I_t can be expressed as shown in Equation (2).

$$I_t = I_0 e^{-(K_a L_a + K_b L_b)} \tag{2}$$

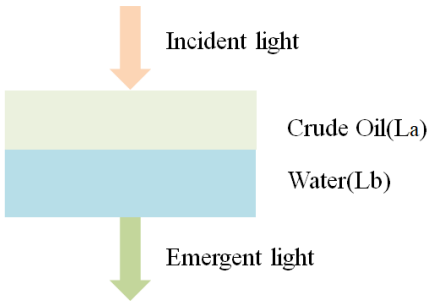

Figure 1. Absorption of crude oil into the spectrum

That is, in the oil extraction industry, as long as the light intensity of 2 or more known oil-water mixtures of different water content received bands are measured, the thickness of oil and water can be found, so as to determine the water content of crude oil, and thus carry out research on water content detection of oil-water two-phase flow.

3. EXPERIMENTAL PART

3.1 Infrared spectral acquisition of samples

Oil-water mixtures with water content ranging from 0% to 100% were used as samples, and the water content of the samples was increased in a gradient of 2.5% for a total of 41 sets of experimental samples. The study uses a tungsten halogen lamp light source with a spectral range of 250nm~2500nm to produce a full-band infrared spectrum, and the infrared spectrum is transmitted through an optical fiber into a cuvette, and the spectral information is collected using a near-infrared spectrometer and the software Uspectral-PLUS_5.4.0 after a certain ratio of oil-water mixture, and the schematic diagram of the experimental setup is shown in Figure 2. When collecting the samples, each sample was placed in the cuvette and stirred well and the experiment was repeated 20 times, and the average value was taken as the spectral data of the concentration samples after manually screening out the individual data with large differences.

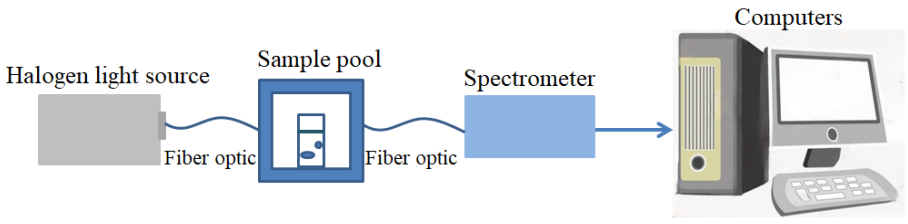

Figure 2. Experimental setup

3.2 Pre-processing of spectra

The spectral information collected using the spectrometer contains some high-frequency random noise in addition to the information about the sample's own properties, and sample inhomogeneity, baseline drift, and light scattering can also have an impact on model stability and accuracy. In this study, the original spectral data were smoothed using S-G smoothing with a window of 41 points and the adjacent averaging method, respectively. S-G smoothing is a polynomial smoothing algorithm based on the principle of least squares, also known as convolution smoothing, and the key is the solution of the matrix operator, let the filter window width be n=2m+1, and the data points within the window were fitted using a k-1th polynomial, as shown in Equation (3).

$$y = a_0 + a_1 x + a_2 x^2 + \cdots + a_{k-1} x^{k-1} \tag{3}$$

where x=(-m,-m+1,...,0,1,...m-1,m), thus forming a system of k-element linear equations with n>k, so that the fitting parameter is A, expressed in matrix as shown in Equation (4).

$$A = (X^T X)^{-1} X^T Y \tag{4}$$

The resulting filtered and smoothed spectral data values are shown in Equation (5).

$$\dot{Y} = XA = X(X^T X)^{-1} X^T XY \tag{5}$$

The adjacent averaging method removes abrupt data points by averaging one point and the data points in the neighborhood, thus filtering out certain noise, as shown in Equation (6).

$$g(x, y) = \frac{1}{M \sum f(x, y)} \tag{6}$$

where M is the number of points of the neighborhood data in the center.

3.3 Spectral analysis modeling methods and model evaluation

The spectral information collected in the full band contains a large amount of redundant information. In order to extract the characteristic wavelengths, the study uses the continuous projection algorithm (SPA) for dimensionality reduction. SPA uses projection analysis of vectors by projecting wavelengths onto other wavelengths, comparing the projection vector size, taking the wavelength with the largest projection vector as the wavelength to be selected, and then selecting the final characteristic wavelengths based on the correction model.

Partial least squares (PLS) is one of the most widely used methods for quantitative analysis model building in NIR spectroscopy. The PLS regression model is built by performing the principal component decomposition of the spectral matrix X_{nxp} and the detection target vector Y_{nx1}, respectively, as shown in Equation (7) (8).

$$X = TP' + E \tag{7}$$

$$Y = UQ' + F \tag{8}$$

T and P are the score matrix and load matrix of the spectral matrix X, respectively; U and Q are the score matrix and load matrix of the detection target vector Y, respectively; E and F are the residual matrices of the spectral matrix X and the detection target vector Y, respectively.

A linear combination u_i, t_i is proposed in each of X and Y for linear regression according to Equation (9).

$$u_i = t_i b_i \tag{9}$$

u_i, t_i are the i-th column of the score matrix U and T, respectively.

The algorithm is decomposition operation in an iterative manner, and in each iteration, the principal components u_i, t_i in U and T are obtained respectively to do linear regression, and if satisfactory accuracy is achieved, the iteration stops. If the accuracy is not enough, the new linear combination u_{i1}, t_{i1} is selected again from the residual matrix after selecting u_i, t_i, and the above algorithm is repeated to reach the accuracy to determine the regression equation, and the algorithm ends.

The root mean square error (RMSE) and correlation coefficient (r) are chosen as the evaluation index of model accuracy. The smaller the root mean square error of the model prediction and the closer the correlation coefficient is to 1, the higher the accuracy of the model is, and the expressions of RMSE and r are shown in equation (10)(11).

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(observed_t - predicted_t)^2} \tag{10}$$

$$r = \frac{1}{n}\sum_{k=1}^{N}\left(\frac{x_K - \mu_x}{\delta_x}\frac{y_k - \mu_y}{\delta_y}\right) \tag{11}$$

where N is the sample capacity, μ_x, δ_x are the mean and standard deviation of the spectral values x, and μ_y, δ_y are the mean and standard deviation of the water content y.

4. EXPERIMENTAL DATA ANALYSIS

4.1 Spectral pre-processing

Figure 3 shows the spectral data of the absorbance fraction of the collected crude oil samples, and it can be seen that the spectra are significantly absorbed around 900 nm, 1100 nm and 1300 nm. The study used S-G smoothing and adjacent averaging to process the collected crude oil water content spectral data, respectively, and the processing results are shown in Figure 4 and Figure 5. In order to facilitate the comprehensive comparative analysis afterwards, the spectral data were normalized to the interval [0,1] for each group of data, and the min-max normalization function was used for the operation, and the normalized spectra are shown in Figure 6.

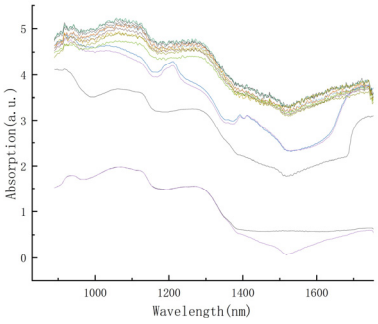

Figure 3. Absorbance diagram of the original spectrum

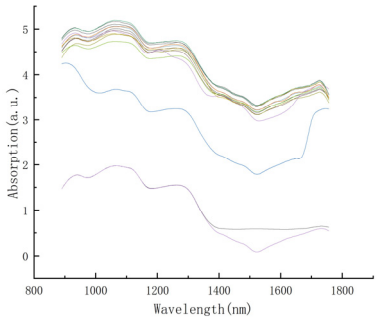

Figure 4. S-G smoothed spectrogram

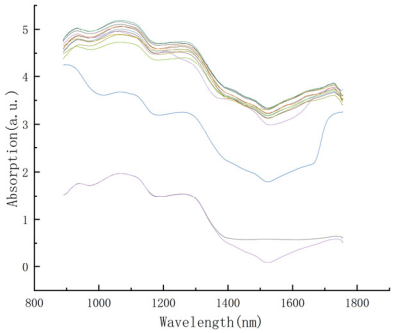

Figure 5. Adjacent averaged smoothed spectrograms

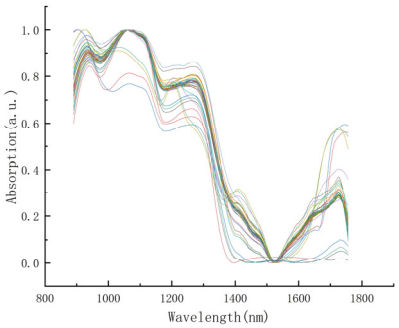

Figure 6. S-G smoothing + normalized spectrogram

Three methods of preprocessing were used: normalization, adjacent averaging, and S-G convolutional smoothing, and the preprocessed data were respectively predicted initially using PLS to compare the prediction results, so as to select a suitable preprocessing method. According to the final water content detection model, the number of SPA feature wavelengths selected, root mean square error, root mean square error of prediction, and correlation coefficient of different preprocessing methods are compared, and the specific values are shown in Table 1, and it can be found that the preprocessing method of using S-G smoothing followed by normalization to build the water content model has fewer

feature wavelengths selected and the root mean square error is the smallest, thus this paper selects the S-G smoothing + normalization In this paper, we choose S-G smoothing + normalization to preprocess the spectral data of oil-water two-phase flow.

Table 1. Effectiveness of SPA+PLS model prediction of crude oil water content based on different pretreatment methods.

Pre-processing methods	Number of characteristic wavelengths	RMSE	RMSEP	r
Unprocessed	14	6.45	12.246	0.93435
Adjacent averaging method	13	5.988	16.194	0.87596
Adjacent averaging + normalization	15	5.2171	9.5502	0.93718
S-G Smooth	11	6.9506	14.0354	0.93381
S-G smoothing + normalization	13	4.4702	9.7131	0.98527

4.2 Selection of characteristic wavelength by continuous projection method

After pre-processing the raw data to remove the noise, the regression analysis of the data is started. The continuous projection method uses the characteristic bands to express the spectral information of the full band, which greatly reduces the computational effort and improves the model efficiency. In this study, there are 41 sets of experimental samples, and 21 sets of samples are randomly selected as the sample set, 10 sets of samples as the test set, and 10 sets of samples as the prediction set, and the SPA feature wavelengths are selected for the 21 sets of sample set and 10 sets of test set spectral data after different preprocessing methods, respectively, and the number of feature wavelengths selected by SPA after the adjacent average method processing is 15, and the root mean square error is 5.2171, and the S-G The number of feature wavelengths selected after smoothing is 13, and the root mean square error is 4.4702, as shown in Figure 7 and Figure 8, that is, the number of feature wavelengths selected after S-G smoothing is less root mean square error, and the 13 feature wavelengths selected are 889.864nm, 939.28nm, 975.755nm, 1039.37nm, 1101.41nm, 1202.58nm, 1297.63nm, 1380.12nm, 1421.75nm, 1528.86nm, 1609.04nm, 1635.37nm, 1695.05nm, which only account for 2.53% of the full wavelengths.

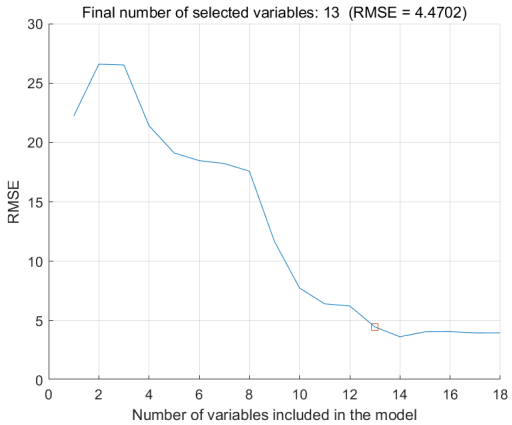

Figure 7. SPA error analysis graph after S-G smoothing

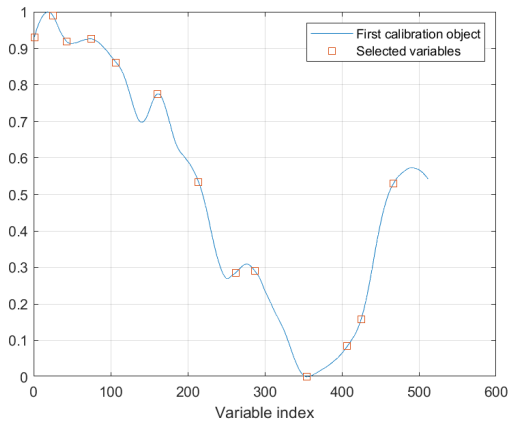

Figure 8. S-G smoothing feature band selection diagram

4.3 Partial least squares model validation

The PLS method can be used for multivariate quantitative correction of spectra, and is widely used in the field of spectroscopic research. In this study, there are 41 experimental samples, including 21 sample sets, 10 test sets, and 10 prediction sets. The characteristic wavelength spectral data selected by SPA were used as the independent variables and the crude oil water content was used as the dependent variable for PLS modeling. A quantitative analysis model for oil-water two-phase flow water content was established, and the prediction results are shown in Figure 9, which shows the correlation plots between the reference values of 10 sets of prediction sets and the predicted values generated by PLS, and the root mean square error of prediction is 9.7131, and the correlation coefficient is 0.98527 tends to be 1, obtaining a satisfactory prediction accuracy.

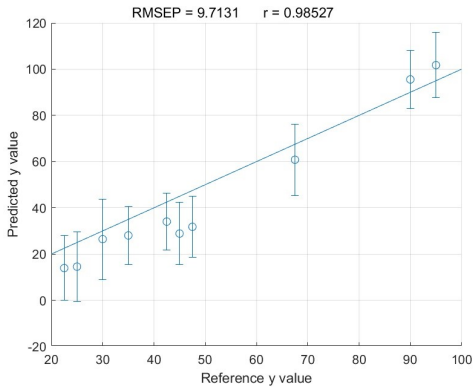

Figure 9. Water content PLS quantitative analysis model results

5. CONCLUSION

The absorbance data of the full-band infrared spectra of 900-1700 nm for oil-water two-phase flow samples with different water contents were collected by building an experimental setup, S-G smoothing with a window point of 41 and normalization were used for pre-processing, and 13 feature bands were selected using SPA with root mean square error RMSE=4.4702, and then partial least squares was performed to establish the water content detection model, and finally the predicted root mean square error RMSEP=9.7131 and correlation coefficient r=0.98527 were obtained with relatively small errors and the model accuracy was high. This study verified the feasibility of infrared spectroscopy detection technology for crude oil water content detection by testing oil-water mixtures in the range of 0% to 100% water content, and provided an experimental basis and technical theory for the application of infrared spectroscopy detection technology to the actual crude oil water content detection. The research after IR spectroscopy can also establish a more perfect database by expanding the number of samples and using new algorithms to further improve the detection accuracy and provide a new way for logging exploration technology, which has great development prospects.

ACKNOWLEDGMENT

This project is supported by the Innovation and Practical Ability Cultivation Program for Postgraduates of Xi 'an Shiyou University (YCS21113132), Natural Science Foundation of Shaanxi Province (2022GY-435).

REFERENCES

[1] Gorbunov, A. and Evstegneev, N., "A device for water content measuring of oil," Journal of Physics: Conference Series, 929(1) (2017).
[2] Douglas, R.K., Nawar, S., Cipullo, S., Alamar, M.C., Coulon, F. and Mouazen, A.M., "Evaluation of vis-NIR reflectance spectroscopy sensitivity to weathering for enhanced assessment of oil contaminated soils," Science of the Total Environment, 626 (2018).
[3] Hidajat, K. and Chong, S.M., "Quality Characterisation of Crude Oils by Partial Least Square Calibration of NIR Spectral Profiles," Journal of Near Infrared Spectroscopy, 8(1) (2000).
[4] Hu, X.T., "Study on water content detection of crude oil based on NIR technology," China Academy of Metrology, 37-42 (2015).
[5] Li, C., Hua X. and Ben W.L., "Optimization Design of Electromagnetic Conductance Sensor for the Water Content of Crude Oil inside Pipeline," Applied Mechanics and Materials, 1245(55-57) (2011).
[6] Zhu, Y., Zou L. and Tan, T.L., "A clustering group lasso method for quantification of adulteration in black cumin seed oil using Fourier transform infrared spectroscopy," Chemometrics and Intelligent Laboratory Systems, 220 (2022).
[7] Han, J., Li, Y.Z., Cao, Z.M., Liu, Q. and Mou, H.W., "Water Content Prediction for High Water-Cut Crude Oil Based on SPA-PLS Using Near Infrared Spectroscopy," Spectroscopy and Spectral Analysis, 3452-3458 (2019).

[8] Li, Y.Z., "Study on water content detection method for low flow rate and high water content crude oil based on NIR," Northeastern Petroleum University, 12-18 (2019).