*J.Natn.Sci.Foundation Sri Lanka 2024 52 (3):* 359 - 370
DOI: http://dx.doi.org/10.4038/jnsfsr.v52i3.11976

RESEARCH ARTICLE

## Equipment Failure Analysis

# Predicting water content in engine oil using SPA-ISSA-BP methods

**Chen Bin[1], Quan YuXuan[1], Liu Ge[2*], Li JianZhong[3], Nie GaoXing[4], Zhang YongChang[1], Shan ChenZe[1], Chen JunSen[2], Chen Rui[1], Chen Xi[1] and Zhang MengWei[2]**

[1] *School of Mechanical & Electrical, Hebei Key Laboratory of Safety Monitoring of Mining Equipment, North China Institute of Science and Technology, Hebei, 065201, China.*
[2] *School of Environmental Engineering, North China Institute of Science and Technology, Hebei, 065201, China.*
[3] *Laboratory of Environmental Biotechnology, CAS, Beijing, 100085, China.*
[4] *Faculty of Engineering, HKU, Hong Kong, 999077, China.*

Submitted: 13 December 2023; Revised: 16 July 2024; Accepted: 24 July 2024

**Abstract:** Contamination of engine oil with water can accelerate oil oxidation and decomposition, resulting in oil degradation. Therefore, the detection and prediction of water content in engine oil is crucial to ensure the long-term safe operation of engines. In this study, near-infrared spectroscopy was used to analyse engine oils with varying water contents. Spectral data were obtained, and various methods, including Savitzky-Golay (SG) smoothing, orthogonal signal correction (OSC), and successive projection algorithm (SPA) were employed to compare the performance differences of principal component regression models. An Improved Sparrow Searches Algorithm (ISSA) was then applied to optimise the Back-Propagation Neural Network (BPNN) for predicting water content in SAE 10W-30 engine oil. The results showed that the performance of the principal component regression model constructed from spectral data after SG smoothing, OSC, and the SPA feature wavelength selection had improved. The BPNN optimised by the improved Sparrow Search Algorithm accelerated the convergence speed of the model and effectively improved the prediction accuracy of the BPNN. The obtained coefficient of determination ($R^2$) was 0.99103, the root mean square error (RMSE) was 2. 2136×$10^{-4}$, and the mean absolute error (MAE) was 1. 8889×$10^{-4}$. These results provide an effective method for detecting and predicting the water content in engine oil.

**Keywords:** Engine oil, near-infrared spectroscopy, neural network, sparrow search algorithm, water content.

## INTRODUCTION

During the long-term operation of an engine, water contamination is inevitable due to factors such as environmental moisture, seal failure, and thermal decomposition products. Some moisture enters the engine through components such as piston rings and cylinder walls, and then travels into the crankcase where it mixes with the engine oil (Khamidullaevnaz 2022). Water is present in engine oil in three different states: free, emulsified, and dissolved (Abdel-Aziz *et al.,* 2016; Hensel *et al.,* 2017). These water contaminants alter the physico-chemical properties of the engine oil, accelerating the depletion of engine additives. This leads to oil degradation, and the formation of corrosive acids and water (Alimova *et al.,* 2021; Zhou & Yang, 2021). This exacerbates oil deterioration, resulting in reduced lubricity, corrosion of engine components, and accelerated formation of sludge and carbon deposits. These problems have a significant impact on the safe and stable operation of the engine. Therefore, the ability to detect and predict water content in engine oil is of paramount importance.

Extensive research has been conducted both nationally and internationally on detecting water content in engine oil, leading to the development of various methods for this purpose. Currently, the most commonly used method for detecting water content in oil is the Karl Fischer titration method (Zhen *et al.,* 2020). However, the presence of additives in engine oil can lead to falsely elevated water content results, when using this method. The dielectric constant method is also prone to errors due to insufficient mixing of the environment and samples, leading to inaccuracies in the test results (Chen & Liu,

* Corresponding author (lgncist@hotmail.com; https://orcid.org/0009-0007-7733-9036)

2020). Gas chromatography, offers high precision and speed, but it requires expensive instrumentation, involves complex procedures, and requires extensive sample pre-processing. These factors contribute to increased test time (Xie *et al.,* 2018).

Near-infrared (NIR) spectroscopy refers to a specific region of the electromagnetic spectrum with wavelengths typically ranging between 780 and 2500 nm. Within this range, the vibrational modes of various chemical bonds and molecules produce distinct absorption features. One of the most prominent is the O-H bond vibrations in water molecules, which have a significant impact on the absorption spectrum in the near-infrared region. Although the spectral signals from O-H bonds can overlap with signals from other functional groups, the use of chemometric signal processing techniques allows for the signals to be separated and resolved. This approach provides accurate water content results and can effectively replace traditional methods of determining water content.

Spectroscopic analysis has been widely used in the detection of water content in oils. Fardin-Kia *et al.* (2019) found significant differences in oils with varying water contents in the near infrared region at 5260 $cm^{-1}$. Based on this discovery, they proposed a quick, univariate Fourier transform near infrared (FT-NIR) method for detecting water content in oils.

Liu *et al.* (2020) developed a portable system for detecting water content in lubricating oil based on visible near infrared (VIS-NIR) reflectance. This system utilized reflectance spectroscopy data input into a neural network model for prediction, achieving a prediction coefficient of determination ($R^2$) of 0.98. In a related study Liu *et al.* (2023) analysed the characteristic peaks in the NIR spectra of engine oil and identified strong absorption peaks at 931 nm, 1195-1212 nm, and 1391-1430 nm. After applying orthogonal signal correction (OSC) to the spectroscopic data, they established a partial least squares regression (PLSR) model. This model enabled accurate prediction of the water content in the oil.

Due to the large number of spectral wavenumbers obtained through spectroscopic analysis, which contain a significant amount of redundant data, building models directly with the original spectra can lead to problems such as reduced model generalisation ability and low prediction efficiency. Therefore, it is necessary to perform feature selection on the acquired spectral data. For example, Liu *et al.* (2023) conducted feature wavelength selection based on the regression coefficient matrix of the partial least squares regression (PLSR) model. They utilized the selected wavelengths to rebuild the PLSR model leading to a significant enhancement of the prediction accuracy of the model. In their study, Zhang *et al.* (2021) applied swarm intelligence algorithms, such as slime mould algorithm and bat algorithm, to select feature wavelengths in near-infrared spectroscopic data of water-containing soil. This approach significantly enhanced the predictive performance of the model. Currently, most researchers utilize PLSR to establish predictive models. This is a linear regression model that effectively deals with multicollinearity problems. However, its capability to fit nonlinear relationships is relatively weak (Chiappini *et al.,* 2020).

In comparison, nonlinear regression models such as support vector regression (SVR), Gaussian Process Regression (GPR), and back-propagation neural network (BPNN) have robust nonlinear fitting capabilities and can fit complex nonlinear relationships. For instance, Liu *et al.* (2021) combined SVR with PLS to achieve precise prediction of water content in crude oil. Additionally, Zhao *et al* (2010) used GPR to develop a prediction model for water content in lubricating oil based on NIR data giving superior results compared to the SVR model.

At present, SVR is facing challenges in selecting the appropriate kernel function for nonlinear problems. Additionally SVM are quite sensitive to noise in the data, which can lead to unstable model performance. On the other hand, GPR has high computational costs, with long training times, and its predictive performance decreases in high-dimensional spaces, making it unsuitable for spectral data modelling.

BPNN, however, is less affected by noise compared to SVM, has better capabilities in handling high-dimensional data than GPR, and its strong nonlinear fitting ability makes it excel in handling complex data. Specifically, BPNN can effectively learn and model complex nonlinear relationships through its multi-layer structure and nonlinear activation functions. Additionally, BPNN has adaptive learning capabilities, allowing it to automatically adjust network parameters using training data, thereby improving model prediction accuracy. However, the initial weights and biases of BPNN consist of pseudo-random numbers, which significantly affect the network's convergence speed and outcomes. Many researchers use optimisation algorithms such as genetic algorithms and particle swarm optimisation algorithms to optimise BPNN. For example, Wang *et al.* (2023) applied a genetic algorithm to optimise BPNN, achieving rapid detection of tea moisture content. Peng *et al.* (2023)

designed a BPNN with a double hidden layer optimised by a particle swarm algorithm. They developed a model that correlated multispectral data with crop moisture content, ultimately achieving accurate predictions of crop moisture content.

Currently, the presence of test noise and baseline drift during spectral acquisition adversely affects the accuracy of water content detection. Existing studies primarily use linear regression models, which are relatively ineffective in handling nonlinear relationships. Therefore, this study proposes a nonlinear modelling method that integrates various preprocessing techniques and optimisation algorithms to enhance the accuracy of water content detection in engine oil. Specifically, we will use Savitzky-Golay (S-G) convolution smoothing and orthogonal signal correction (OSC) to eliminate noise and apply the successive projection algorithm (SPA) for selecting wavelength of spectral features. Next, we will optimise the back-propagation neural network (BPNN) using the improved sparrow search algorithm (ISSA) to enhance the model's nonlinear fitting capability. Finally, we will evaluate the model's performance using $R^2$, root mean square error (RMSE), and mean absolute error (MAE) as metrics, and conduct a comparative analysis. This research aims to provide technical support for the long-term stability and safe operation of engines.

## MATERIAL AND METHODS

### Preparation of oil samples

SAE 10W-30 engine oil was chosen as the subject of the experiment. First, 10 ml of the original oil sample, free of water content, was placed in glass test tubes. Next, 10 µl of oil was removed using a pipette, and 10 µl of distilled water was added to create oil samples with a water content of 0.1%. By adjusting the volumes of oil removed and distilled water added, a total of 32 oil samples with varying water contents ranging from 0% to 0.9% were prepared. After preparation, the test tubes were placed in an ultrasonicator for 10 min to ensure thorough mixing of the oil and water. The Karl Fischer moisture meter was then used to accurately measure the water content of the oil samples, providing the final water content values of the oil samples. After the measurements, the test tubes were sealed and placed in a light-free environment to prevent photodegradation. The final water content values for the 32 oil samples are shown in Table 1.

**Table 1:** Water content of oil Samples

| Sample number | Water content | Sample number | Water content |
|---|---|---|---|
| 1 | 0 | 17 | 0.0045 |
| 2 | 0.0010 | 18 | 0.0047 |
| 3 | 0.0015 | 19 | 0.0048 |
| 4 | 0.0017 | 20 | 0.0049 |
| 5 | 0.0020 | 21 | 0.0050 |
| 6 | 0.0025 | 22 | 0.0052 |
| 7 | 0.0026 | 23 | 0.0057 |
| 8 | 0.0027 | 24 | 0.0060 |
| 9 | 0.0030 | 25 | 0.0063 |
| 10 | 0.0032 | 26 | 0.0067 |
| 11 | 0.0035 | 27 | 0.0070 |
| 12 | 0.0037 | 28 | 0.0072 |
| 13 | 0.0038 | 29 | 0.0073 |
| 14 | 0.0039 | 30 | 0.0078 |
| 15 | 0.0040 | 31 | 0.0080 |
| 16 | 0.0041 | 32 | 0.0090 |

### Experimental apparatus and spectroscopic testing method

The experiment used the ocean optics NIR_Quest near infrared spectrometer, model number 512-1.9. This spectrometer employs the Beer-Lambert's law to obtain spectral information from oil samples. The wavelength of the spectrometer ranged from 900 nm to 1700 nm, with an optical resolution of 3.1 nm, a wavelength repeatability of 0.1 nm, an InGaAs detector, and a 50W quartz halogen lamp as the light source.

The spectrometer was connected to the computer through a USB interface, and spectral data were collected using the OCEANVIEW software. Reflectance fiber optic probes were used to acquire spectra of oil samples within the 900-1700 nm range. A total of 512 spectral bands were collected, with 20 scans per sample. Each oil sample underwent three replicate experiments, and the final result was determined by averaging the values.

**Successive projections algorithm**

The SPA is a commonly used feature selection method that has the advantage of not relying on data distribution assumptions and performs well in feature selection in high-dimensional data. It selects the most informative features by calculating the projection vector sizes between feature bands, eliminating collinearity among variables, and improving modelling accuracy while reducing model complexity (Jia *et al.,* 2023; Liu *et al.,* 2023).

In each iteration, SPA selects the feature with the maximum projection with the remaining unselected features, gradually constructing a subset of features, reducing collinearity among wavelengths, and improving the accuracy and efficiency of problem modelling.

**Sparrow search algorithm and its improved algorithm**

***Sparrow search algorithm***

The Sparrow Search Algorithm (SSA) is a heuristic optimisation algorithm inspired by the collective behaviour of sparrow populations in the natural world, utilized to solve various optimisation problems (Gharehchopogh *et al.,* 2023; Xue & Shen, 2020). The algorithm is inspired by the foraging behaviour of sparrow populations, where sparrows find food and avoid danger by cooperating and sharing information. In the algorithm, sparrows are divided into three groups: explorers, followers, and sentinels. Explorers, similar to the mutation operation in genetic algorithms, are responsible for exploring unknown regions in the solution space. Followers move towards the optimal solution based on the explorers' results, accelerating the global search process. Sentinels are randomly selected from the entire population and act as early warnings of danger; the entire population responds with anti-predatory behaviour when danger is detected.

***Improved sparrow search algorithm***

Due to problems such as uneven initial population distribution and low diversity in the later iterations, improvements are necessary for the SSA. The improvement strategies are as follows:

(1) Initialization of population using tent chaotic and inverse mapping
To ensure a uniformly distributed initial population, the Tent chaotic mapping (Li *et al.*, 2020) was used to generate N initial populations. This was followed by an inverse mapping process to generate N inverse populations. The best N of these is then retained as the initial population.

(2) Fusion of reverse learning and Cauchy mutation strategy
To expedite the convergence speed and assist the algorithm to find the optimal solution more quickly, reverse learning and Cauchy mutation have been integrated into the SSA algorithm (Ali & Pant, 2011; Xu *et al.*, 2023). The strategy to be used is selected based on the probability (rp), as shown in Equation (1).

$$r_p = -\exp(1-\frac{gen}{gen_{\max}})^{20} + 2 \quad (01)$$

(3) Greedy Rule
Finally, a greedy rule is introduced, where after updating the position of the optimal solution, its fitness is compared with the fitness of the old position to determine if the position should be updated.

**ISSA-BP Neural Network**

Combining the improved sparrow search algorithm (ISSA) with the BPNN results in the ISSA-optimised BPNN. The specific procedure is outlined as follows:

(1) Initialize the BPNN by determining the number of nodes in the input and output layers, and the number of hidden layers and nodes.
(2) Set the parameters for the ISSA and the BPNN. Use Tent chaotic mapping and inverse mapping to generate the initial population, where the D-dimensional position vector of each individual represents the initial weights and biases of the BPNN.
(3) Calculate individual fitness values and rank them.
(4) Select high-fitness individuals as explorers, the remaining individuals as followers, and update their positions. Randomly select some individuals from the whole population as sentinels and update their positions.
(5) Calculate the population fitness and rank them.
(6) Based on the selection probability (rp), update the position of the optimal solution using the selected strategy. Calculate its fitness, and then decide whether to update the position according to the greedy rule.
(7) Check if the termination condition is satisfied. If it is, output the results and exit the program. If not, repeat steps (3) through (6).

The flow chart is shown in Figure 1.

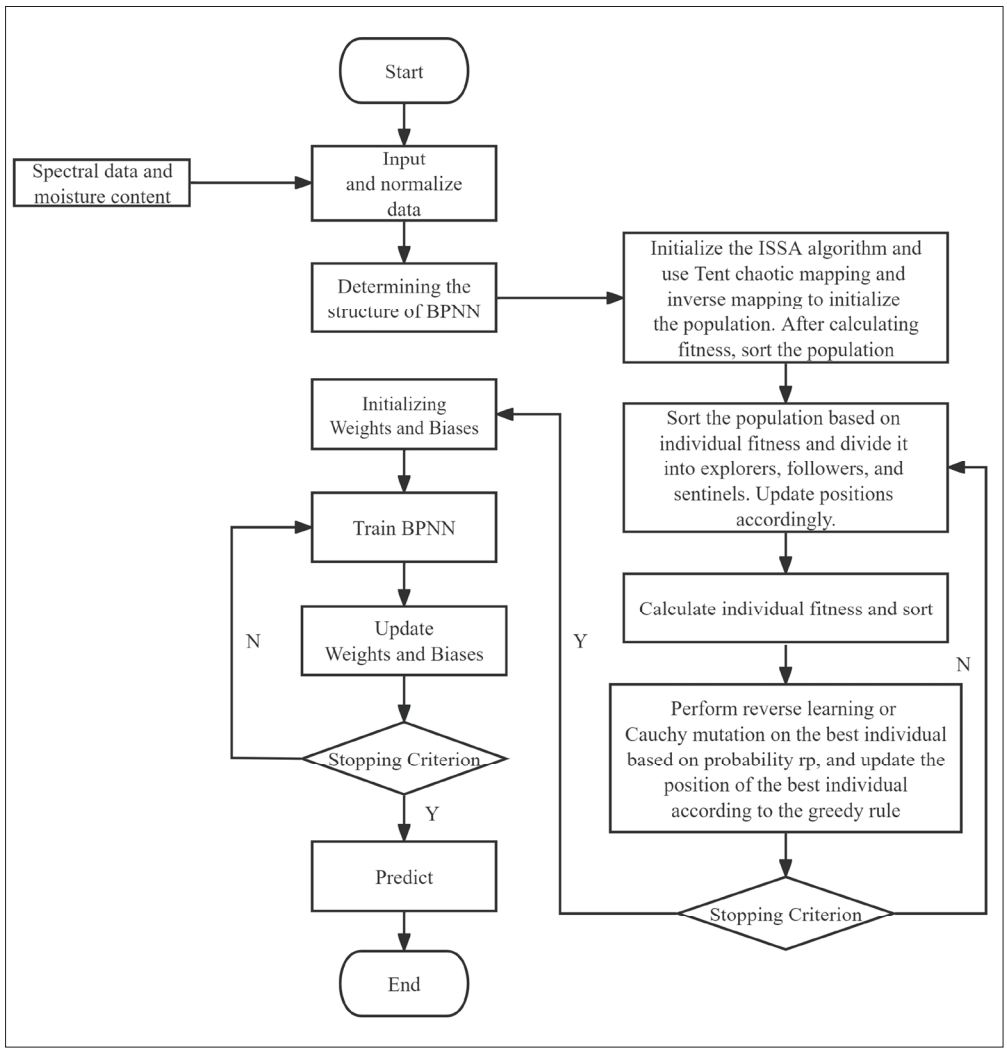


**Figure 1:** ISSA-BPNN Flowchart

## Modelling with SPA-ISSA-BPNN

### *Preprocessing of spectral data*

During the near-infrared inspection process, factors such as light leakage, fluctuations in light intensity, and sample contamination can cause interference, resulting in anomalies in the spectral data. These anomalies can significantly affect the predictive performance of the established ISSA-BPNN model. Therefore, it is necessary to remove outliers from the spectral data prior to modelling. First, principal component analysis (PCA) is applied to the spectral data. The score plot of the first two principal components is shown in Figure 2. Subsequently, Hotelling's T2 statistic ellipse with a confidence level of 95% is used to detect outliers. Sample number 32, with a moisture content of 0.9%, was found to be outside the ellipse, indicating that it is an outlier that needs to be removed.

Comparison of SVR model performance with and without removing outliers from spectral data. The performance metrics of the support vector regression (SVR) models constructed using spectral data before and after removing outliers are compared in Table 2. The validation performance of the model is significantly improved after outlier removal.

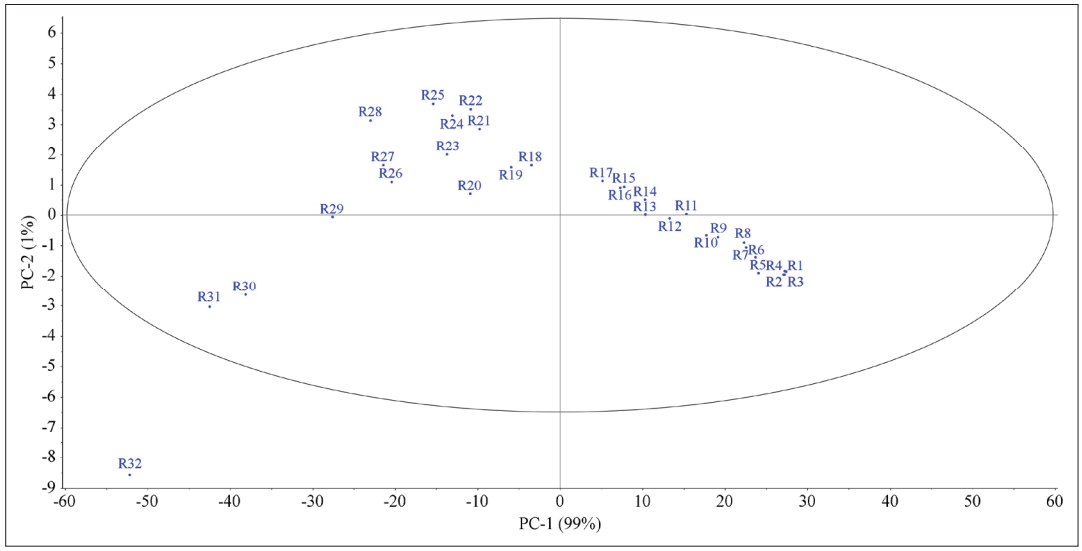


**Figure 2:** Score plot of PCA analysis

**Table 2:** Comparison of SVR Model performance metrics using spectral data before and after outlier removal

| Model | $R^2_C$ | $R_{MSE}{}^C$ | $R^2_V$ | $R_{MSE}{}^V$ |
|---|---|---|---|---|
| Before remove outliers | 0.9468 | $5.081\times10^{-4}$ | 0.9283 | $5.974\times10^{-4}$ |
| After remove outliers | 0.9454 | $4.958\times10^{-4}$ | 0.9380 | $5.314\times10^{-4}$ |

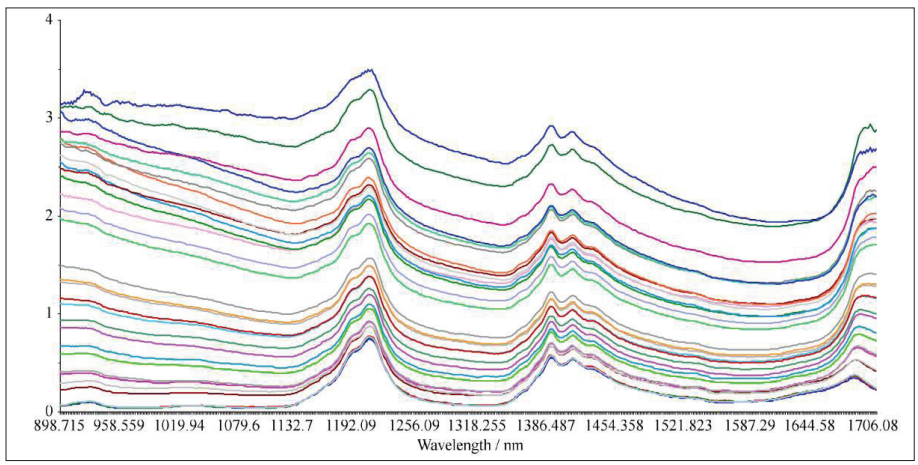


**Figure 3:** Raw spectra after outlier removal

Figure 3 shows the raw spectra of oil samples with various moisture contents after removing outliers. It is evident that the spectra of the oil samples with varying moisture contents follow a similar trend to that of the moisture-free oil, with variations only in the intensity values. However, characteristic peaks are not visible in the original spectra,

and the spectra contain considerable noise making it difficult to build the ISSA-BPNN model. To address this issue, the original spectra were subjected to polynomial smoothing using a S-G convolution with a polynomial order of 2 and a window width of 7. Subsequently, OSC was applied for further spectral processing. The processed spectral data are shown in Figure 4.

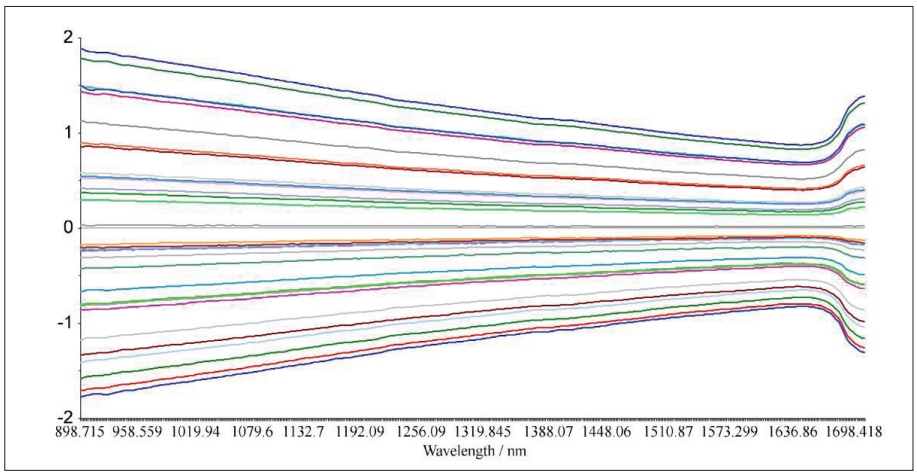


**Figure 4:** Spectral data after S-G filtering and OSC processing

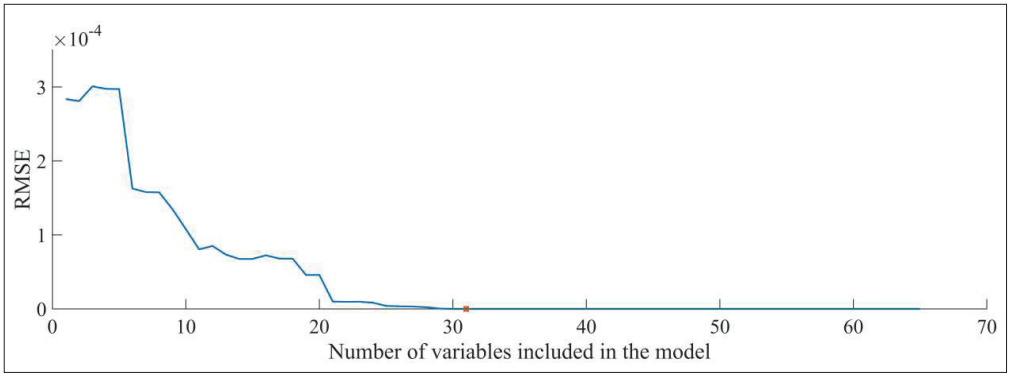


**Figure 5:** RMSE variation curve with different numbers of feature wavelengths

Due to the high volume of raw spectral data from the oil samples, which included many irrelevant spectral regions, the time for model building was prolonged, and the accuracy of predictions was compromised. Therefore, it was necessary to perform feature wavelength selection on the spectra. In this study, Spectral Projection Analysis (SPA) was utilized for selecting the spectral feature wavelengths. The wavelength extraction range was set from 25 to 50. The wavelengths selected by SPA were used to construct multiple linear regression (MLR) models. The RMSE values of the test sets for different numbers of selected feature wavelengths are shown in Figure 5. It can be observed that the RMSE value of the test set decreases to a minimum when the number of selected wavelengths is 31 and remains stable thereafter. The specific selected wavelengths are listed in Table 3. The wavelengths selected by SPA are mostly clustered around 950 nm, 1220 nm, 1420 nm, 1500-1600 nm, and 1700 nm. Among them, 950 nm corresponds to the second harmonic absorption band of the O-H bond stretching vibration, 1220 nm corresponds to the combined frequency absorption band of O-H in water, and 1420 nm

corresponds to the first harmonic absorption band of O-H stretching vibration. The range of 1500-1600 nm may correspond to the vibration absorption peak of C-H. Finally, 1700 nm corresponds to the first harmonic absorption band of O-H groups in water.

**Table 3:** Wavelengths selected by SPA algorithm

| Algorithm | Number of selected wavelengths | Wavelength | | | | |
|---|---|---|---|---|---|---|
| SPA | 31 | | 898.715 | 913.277 | 927.836 | 950.476 |
| | | 953.709 | 961.792 | 971.489 | 973.105 | 976.337 |
| | | 986.031 | 987.646 | 1000.567 | 1002.182 | 1224.119 |
| | | 1359.553 | 1381.738 | 1411.786 | 1441.76 | 1455.931 |
| | | 1518.695 | 1524.951 | 1534.327 | 1562.403 | 1579.52 |
| | | 1591.949 | 1655.376 | 1670.775 | 1689.214 | 1690.749 |
| | | 1695.351 | 1707.611 | | | |

Principal component regression (PCR) models were constructed using the data after removing outliers. Data processed by S-G convolution smoothing and OSC, and data processed by S-G convolution smoothing and OSC after SPA feature selection, are shown in Table 4. It can be observed that the models constructed using preprocessed spectral data have better parameters compared to the models constructed using original spectral data. Both R2C and R2V have improved by approximately 0.06, and RMSEC and RMSEV have decreased by approximately 50%. Moreover, the data after SPA feature selection shows a slight improvement in performance compared to the full spectrum processed with SG smoothing and OSC. This indicates that the feature wavelength selection has eliminated a significant amount of irrelevant information, further enhancing the performance of the model.

**Table 4:** PCR results of spectral data before and after preprocessing

| Preprocessing method | $R^2_C$ | $R_{MSE}{}^C$ | $R^2_V$ | $R_{MSE}{}^V$ |
|---|---|---|---|---|
| Original | 0.9371 | $5.128\times10^{-4}$ | 0.9368 | $6.843\times10^{-4}$ |
| SG (7,2) + OSC | 0.9855 | $2.270\times10^{-4}$ | 0.9818 | $2.435\times10^{-4}$ |
| SPA+SG (7,2) + OSC | 0.9868 | $2.234\times10^{-4}$ | 0.9880 | $2.441\times10^{-4}$ |

## Neural network modelling

To establish a model capable of predicting water content in oil and to assess its accuracy, the remaining 31 samples, after preprocessing, were randomly divided into a training set and a test set. The training set contained 24 samples, while the test set contained 7 samples. The models were constructed using the selected SPA spectra dataset from the training set and the corresponding input standard BPNN and incremental sparse spectrum analysis with ISSA-BPNN. The parameters for the BPNN and ISSA algorithms are shown in Table 5.

**Table 5:** Parameters of BPNN and ISSA algorithm

| Algorithm | Parameter name | Parameter value |
|---|---|---|
| BPNN | Training function | trainbr |
| | Maximum epochs | 400 |
| | Goal | $1\times10^{-4}$ |
| | Ratio of training set and validation set | 4:1 |
| ISSA | Pop size | 30 |
| | Maximum iterations | 80 |
| | Explore ratio | 0.4 |
| | Follower ratio | 0.6 |
| | Sentinel ratio | 0.4 |
| | Upper bound | 5 |
| | Lower bound | -5 |
| | Warning value | 0.8 |
| | Fitness function | MAE |

## RESULTS AND DISCUSSION

The fitness evolution curve during the neural network iterations is shown in Figure 6. It can be observed that the fitness value reaches its minimum after 17 iterations, ranging between 0.010 and 0.011. This result demonstrates the strong global optimisation capability of the SSA algorithm, which effectively enhances the prediction accuracy of the BPNN.

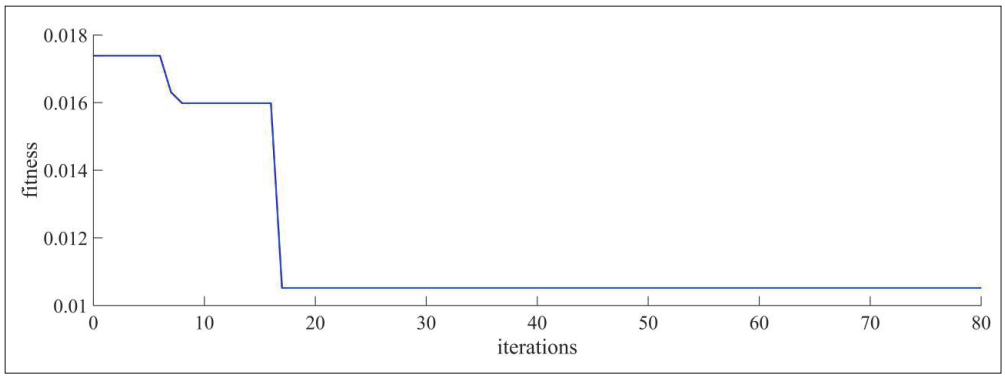


**Figure 6:** Fitness evolution curve during iterations

To visually demonstrate the significant improvement brought about by the ISSA-optimised BPNN, predictions were made on the test set using SPA-ISSA-BPNN, SPA-BPNN, and PCR models with identical training parameters. The results were then compared with the actual values, as shown in Figure 7. It is evident from the comparison that the predictive performance of the model constructed by SPA-ISSA-BPNN is significantly superior to those constructed by BPNN and PCR.

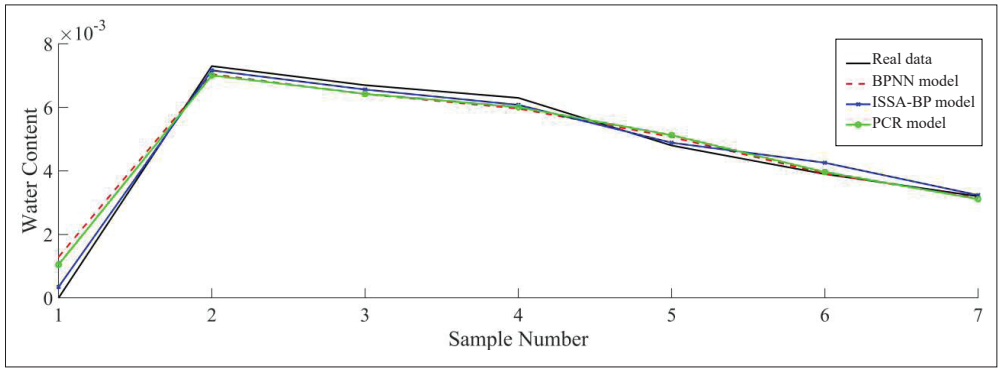


**Figure 7:** Comparison between Predicted Values and Actual Values

The residual comparison between the predicted values from the SPA-ISSA-BPNN, SPA-BPNN, and SPA-PCR models on the test set and the actual values is shown in Figure 8. It is evident that the residuals of the predicted values from the SPA-ISSA-BPNN model are significantly lower than those of the SPA-BPNN and SPA-PCR model predictions when compared to the actual values.

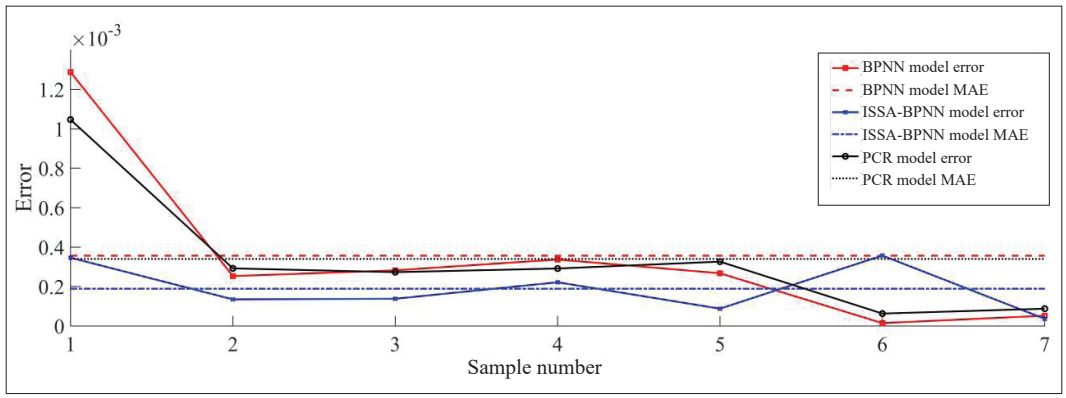


**Figure 8:** Comparison of residuals and mean absolute errors of predicted values

**Table 6:** Comparison of model performance metrics

| Model | $R^2$ | $R_{MSE}$ | MAE |
|---|---|---|---|
| SPA-ISSA-BP | 0.9910 | $2.2136\times10^{-4}$ | $1.8889\times10^{-4}$ |
| SPA-BP | 0.9479 | $5.3335\times10^{-4}$ | $3.5635\times10^{-4}$ |
| SPA-PCR | 0.9619 | $4.5656\times10^{-4}$ | $3.4009\times10^{-4}$ |

Using the coefficient of determination ($R^2$), root mean square error (RMSE), and mean absolute error (MAE) as evaluation metrics, the comparison of the prediction performance metrics for the SPA-ISSA-BPNN, SPA-BPNN, and SPA-PCR models is presented in Table 6. It can be observed that the SPA-ISSA-BP model reduced the $R_{MSE}$ and MAE by $2.272\times10^{-5}$ and $1.6746\times10^{-4}$, respectively, compared to the SPA-BP model, with an increase in $R^2$ of 0.0431. The SPA-ISSA-BP model decreased the RMSE and MAE by $2.3520\times10^{-5}$ and $1.5120\times10^{-4}$, respectively, compared to the SPA-PCR model, with an increase in $R^2$ by 0.0291. In conclusion, the initial weights and biases of the BPNN are random, which often causes the network to get stuck in local minima during training, resulting in high residual errors and low prediction accuracy. However, the ISSA-

optimised BPNN can effectively explore the solution space, and avoid getting stuck in local minima, and thus increasing the likelihood of the network converging to the global optimum. PCR, being a linear regression model, may not fully capture the nonlinear relationship between the spectral data and the component to be measured. Therefore, the residual errors of the ISSA-optimised BPNN model are smaller than those of the non-optimised BPNN model and the PCR model, indicating the higher prediction accuracy. This meets the practical accuracy requirements for the detection of water content in engine oil.

## CONCLUSION

In this study, near-infrared spectroscopy was used for the quantitative analysis of 31 samples with water contents ranging from 0% to 0.9%. The acquired spectral data were subjected to outlier removal, followed by S-G convolution and OSC processing. Subsequently, the SPA algorithm was used to select the feature wavelengths. The processed data were divided into training and test sets. A SPA-ISSA-BPNN model based on the selected feature wavelengths was established using the training set, and its predictive performance was compared with that of a regular BPNN model. The results showed that the performance metrics of the PCR models constructed using spectral data processed with SG, OSC, and SG, OSC, SPA were superior to those of the PCR model constructed using raw spectral data. Among these, the model built using spectral data processed with SG, OSC, and SPA showed better predictive performance than the model built using data processed with SG, OSC, thus reducing the complexity of the model. SPA-BP and SPA-ISSA-BPNN network models were developed using data processed with SG, OSC, and SPA. The performance metrics of the SPA-ISSA-BP model were superior, with an $R^2$ of 0.99103, RMSE of $2.2136\times10^{-4}$, and MAE of $1.8889\times10^{-4}$.

In conclusion, the developed SPA-ISSA-BP engine oil water content prediction model utilizes SPA for feature wavelength selection, reducing the number of wavelengths required for modelling and simplifying model complexity. Additionally, the ISSA algorithm optimises the weights and biases of the BPNN network, further improving the model's accuracy and robustness. The constructed SPA-ISSA-BP engine oil water content prediction model ultimately achieves rapid detection of the water content in SAE 10W-30 engine oil. Although this study specifically focuses on SAE 10W-30 engine oil, the proposed method has broad applicability and potential. It provides a reference for monitoring the water content in other types of engine oil, which is crucial for ensuring the long-term stability and safe operation of engines.

**Acknowledgments**

The authors are grateful to National Natural Science Foundation of China (Grant number:51375516) and Open Project of Key Laboratory of Environmental Biotechnology (Grant number:kf2021004) for financial support.

## REFERENCES

Abdel-Aziz, M. H., El-Ashtoukhy, E. S. Z., Zoromba, M. S., & Bassyouni, M. (2016). Oil-in-water emulsion breaking by electrocoagulation in a modified electrochemical cell. *International journal of electrochemical science, 11*(11), 9634-9643. https://doi.org/10.20964/2016.11.53

Ali, M., & Pant, M. (2011). Improving the performance of differential evolution algorithm using Cauchy mutation. *Soft Computing, 15*(5), 991-1007. https://doi.org/10.1007/s00500-010-0655-2

Alimova, Z., Akhmatjanov, R., Kholikova, N., & Karimova, K. (2021). Ways to improve the anticorrosive properties of motor oils used in vehicles. *E3S Web of Conferences, 264*, Article 05004. https://doi.org/10.1051/e3sconf/202126405004

Chen, B., & Liu, G. (2020). Research progress in on-line monitoring methods of micro-water content in transformer oil. *High Voltage Engineering*, *46*(4), 1405-1416. https://doi.org/10.13336/j.1003-6520.hve.20200430034

Chiappini, F. A., Teglia, C. M., Forno, Á. G., Goicoechea, & H. C. (2020). Modelling of bioprocess non-linear fluorescence data for at-line prediction of etanercept based on artificial neural networks optimised by response surface methodology. *Talanta, 210*, Article 120664. https://doi.org/10.1016/j.talanta.2019.120664

Fardin-Kia, A. R., Karunathilaka, S. R., Yakes, B. J., Lee, K., Ellsworth, Z., Brückner, L., & Mossoba, M. M. (2019). A rapid, univariate ft-nir procedure to determine moisture concentration in olive oil. *Journal of Oleo Science, 68*(11), 1105-1112. https://doi.org/10.5650/jos.ess19118

Gharehchopogh, F. S., Namazi, M., Ebrahimi, L., & Abdollahzadeh, B. (2023). Advances in sparrow search algorithm: a comprehensive survey. *Archives of Computational Methods in Engineering, 30*(1), 427-455. https://doi.org/10.1007/s11831-022-09804-w

Hensel, J. K., Carpenter, A. P., Ciszewski, R. K., Schabes, B. K., Kittredge, C. T., Moore, F. G., & Richmond, G. L. (2017). Molecular characterization of water and surfactant aot at nanoemulsion surfaces. *Proceedings of the National Academy of Sciences, 114*(51), 13351-13356. https://doi.org/10.1073/pnas.1700099114

Jia, Z., Wang, Z., Li, X., Qiu, H., Hou, G., & Fan, P. (2023). Marine sediment particle size classification based on the fusion of principal component analysis and continuous projection algorithm. *Spectroscopy and Spectral Analysis, 43*(10), 3075-3080. https://doi.org/10.3964/j.issn.1000-0593(2023)10-3075-06

Khamidullaevna, A. Z., & Bahtiyor H T. (2022). Studies of anticorrosive properties motor oils and ways to improve. *European International Journal of Multidisciplinary Research and Management Studies, 2*(06), 6-12. https://doi.org/10.55640/eijmrms-02-06-02

Li, Y., Han, M., & Guo, Q. (2020). Modified whale optimization algorithm based on tent chaotic mapping and its application in structural optimization. *KSCE Journal of Civil Engineering, 24*(12), 3703-3713. https://doi.org/10.1007/s12205-020-0504-5

Liu, C., Tang, X., Yu, T., Wang, T., Lu, Z., & Yu, W. (2020). Measurement of moisture content in lubricating oils of high-speed rail gearbox by vis-nir spectroscopy. *Optik, 224*, Article 165694. https://doi.org/10.1016/j.ijleo.2020.165694

Liu,G ., Chen, B., Shang, Z. X., & Quan, Y. X. (2023). Near infrared spectroscopy analysis of moisture in engine oil. *Spectroscopy and Spectral Analysis, 43*(2), 449-454. https://doi.org/10.3964/j.issn.1000-0593(2023)02-0449-06

Liu, H. M., Liu, Y. J., Zhong, Z. C., Song, Y., Li, Z., & Xu, Y. (2021). Detection and analysis of water content of crude oil by near infrared spectroscopy. *Spectroscopy and Spectral Analysis, 41*(2), 505-510. https://doi.org/10.3964/j.issn.1000-0593(2021)02-0505-06

Liu, K., Gao, G., Wang, F., Wu, D., Wu, Z., & Gong, Y.,(2023). Research on oil-water two-phase water content detection model based on near-infrared spectroscopy, *Third International Symposium on Computer Engineering and Intelligent Communications,* 12462, Article 124620L. https://doi.org/10.1117/12.2660957

Peng, Y., He, M., Zheng, Z.,& He, Y. (2023). Enhanced neural network for rapid identification of crop water and nitrogen content using multispectral imaging. *Agronomy, 13*(10), Article 2464. https://doi.org/10.3390/agronomy13102464

Wang, F., Xie, B., Lü, E., Zeng, Z., Mei, S., Ma, C., & Guo, J. (2023). Design of a moisture content detection system for yinghong no. 9 tea leaves based on machine vision. *Applied Sciences, 13*(3), Article 1806. https://doi.org/10.3390/app13031806

Xie, W., Gong, Y., & Yu, K. (2018). Efficient quantification of water content in edible oils by headspace gas chromatography with vapour phase calibration. *Journal of the Science of Food and Agriculture, 98*(8), 3208-3212. https://doi.org/10.1002/jsfa.8795

Xu, Z., Su, Y., Yang, F., & Zhang, M. (2023). A whale optimization algorithm with distributed collaboration and reverse learning ability. *Computers, Materials and Continua, 75*(3), 5965-5986. https://doi.org/10.32604/cmc.2023.037611

Xue, J., & Shen, B. (2020). A novel swarm intelligence optimization approach: sparrow search algorithm. *Systems Science and Control Engineering, 8*(1), 22-34. https://doi.org/10.1080/21642583.2019.1708830

Zhang, D., Liu, J., Yang, L., Cui, T., He, X., Yu, T., & N. O. Kheiry, A. (2021). Application of swarm intelligence algorithms to the characteristic wavelength selection of soil moisture content. *International Journal of Agricultural and Biological Engineering, 14*(6), 153-161. https://doi.org/10.25165/j.ijabe.20211406.6629

Zhao, Y., Xu, X., Jiang, L. L., Zhang, Y., Tan, L. H., & He, Y. (2010). Determination of water content in automobile lubricant using near-infrared spectroscopy improved by machine learning analysis. *International Conference on Information Science and Applications*, 2010, 1-8. https://doi.org/10.1109/ICISA.2010.5480392

Zhen, Z., Wang, H., Yue, Y., Li, D., & Li, J. (2020). Determination of water content of crude oil by azeotropic distillation karl fischer coulometric titration. *Analytical and Bioanalytical Chemistry*, *412*(12). https://doi.org/10.1007/s00216-020-02714-5

Zhou, F., & Yang, K. (2021). The influence of water content on the acid number of diesel engine lubricant in electrochemical measurement. *IOP Conference Series. Materials Science and Engineering, 1043*(5), 52054. https://doi.org/10.1088/1757-899X/1043/5/052054