HYPERSPECTRAL INVERSION OF SOIL ORGANIC MATTER UNDER DIFFERENTIAL VEGETATION COVER SCENARIO BASED ON CARS-RF

PAN, Y. Q.[1] – DENG, J.[1*] – CHEN, Z. C.[2] – WANG, S. D.[2] – DENG, Q.[2] – ZHANG, X. F.[2] – JIAO, Y. H.[2]

[1]*Henan Provincial Institute of Land and Space Survey and Planning, Zhengzhou 450016, China*

[2]*School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo 454003, China*

[*]*Corresponding author*
e-mail: hndj@sohu.com

(Received 13th Jun 2025; accepted 31st Jul 2025)

Abstract. The development of coal resources has caused serious environmental problems such as soil degradation and crop damage even as promoting economic development. Accurate and efficient determination of soil organic matter content is essential to contribute to controlling soil degradation and improving land fertility in mine-grain mixed zones. Therefore, a typical ore-grain composite area was selected as the research site. The collected soil spectral data were filtered and denoised, and the second-order differential transformations were used to enhance the spectral characteristics. The competitive adaptive reweighted sampling (CARS) algorithm was used to screen the sensitive bands, and the inversion model of soil organic matter content was constructed based on partial least squares regression, random forest and XGBoost algorithms. The results show that the correlation between soil spectra and organic matter content can be improved by using SG denoising combined with second-order differential transformation to extract effective information of soil spectra. The CARS algorithm extracts important features, removes redundant spectral information, and improves modeling accuracy. In terms of inversion, the RF model built based on CARS feature extraction had the highest accuracy, and the prediction accuracy (R^2) reached 0.95 in the low vegetation coverage area. The second was XGBoost, PLSR model inversion accuracy is the lowest, in high and low vegetation coverage area R^2 is 0.79 and 0.78, respectively. The results fully demonstrate the effectiveness and feasibility of machine learning methods for retrieving hyperspectral soil organic matter content. This research can provide theoretical and scientific basis for the rapid monitoring of large-scale soil organic matter.
Keywords: *ore-grain composite area, second-order differential transformation, XGBoost, differential vegetation cover, machine learning*

Introduction

Soil plays an important role in the ecosystem (Sokol et al., 2022; Hartmann and Six, 2023). As a potential carbon sink, SOM status is also a key indicator of the restoration and maintenance of ecological functions in degraded ecosystems (Lee et al., 2023). Since the Industrial Revolution, industrial production techniques have experienced extraordinary and swift transformations (Wang et al., 2021a). The steady increase in greenhouse gases' levels has led to a global average temperature rise of 0.3-0.6℃, resulting in a cascade of environmental issues (Jones et al., 2023; Filonchyk et al., 2024). Coal mining and processing facilitate swift economic advancement while simultaneously instilling a range of ecological issues. Coal mining poses a serious threat to land resources and ecological environment (Bazaluk et al., 2023; Duo et al., 2024). Coal mining, a significant contributor to carbon emissions, is projected to devastate hundreds of square kilometers of agricultural land annually in China alone (Wang et al., 2015). Therefore,

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2305_93959412

the prediction of soil organic matter distribution in ore-grain composite area is of great significance to local agricultural management.

The traditional determination of organic matter content was mainly through chemical analysis methods. These methods exhibited high accuracy; however, they are intricate, time-intensive, and expensive, providing point-specific data that are unsuitable for large-scale development (Chenu et al., 2024; Li et al., 2024). Hyperspectral technology has been extensively utilized in soil property prediction research owing to its rapid, straightforward, non-polluting, and non-destructive characteristics (Yang et al., 2021; Ai et al., 2022). Hyperspectral remote sensing technology by combining the quantitative relationship between spectral characteristics and soil organic matter content (Yang et al., 2021; Dai et al., 2022). The results show that there is a significant correlation between soil organic matter content and spectral reflectance, especially in the wavelength range of 400~2500 nm, and some bands (such as 600, 820 and 1600 nm) have a particularly significant inversion effect on soil organic matter content (Yang et al., 2021; Yuan et al., 2024). In addition, the high-resolution nature of hyperspectral technology allows it to capture subtle changes in the soil spectrum, thus improving the inversion accuracy. Winter wheat, as an important grain crop, has a close relationship with soil fertility. The application of hyperspectral technique to soil organic matter content inversion in winter wheat farmland is also increasing gradually. Nonetheless, the model's performance and accuracy are significantly compromised by the interference from soil sample surface, the spectral testing environment, spectral noise, and issues such as information redundancy, multicollinearity, and overlapping absorption peaks in the spectral data (Wu et al., 2024; Sun et al., 2024). Prior research predominantly employed conventional transformations, including reciprocal, logarithmic, differential, and absorption peak depth techniques, to mitigate soil spectral noise. Additionally, they eliminated non-informative or redundant variables using methods, continuous projection algorithm (SPA), and competitive adaptive reweighted sampling (CARS) (Ye et al., 2008; Xia et al., 2021) to optimize the sensitive bands of organic matter, and improve the prediction accuracy and stability of the model. The CARS algorithm performs well in screening sensitive bands. Compared with the existing methods such as UVE and SPA, it can effectively enhance the accuracy and stability of the spectral preprocessing and feature band selection (Yuan et al., 2020a).

In addition, inversion models are also crucial to improve the performance of hyperspectral inversion of SOM content (Gu et al., 2019). The commonly used inversion models mainly include regression model and machine learning model. (Huang et al., 2020; Wang et al., 2024). The results show that the machine learning model performs better in the inversion of SOM content (Wang et al., 2021b, 2024a). Multiple models have been used to determine soil organic matter content. The results have shown that the accuracy of RF and SVM models is higher than the partial least squares regression models (Wang et al., 2024b). XGBoost is an ensemble learning algorithm proposed by Chen et al in 2016, which has been widely used in hyperspectral inversion of soil components (Chen and Guestrin, 2016). In terms of the performance, XGBoost, RF and SVM, in hyperspectral inversion of soil nickel content, XGBoost has had the best performance (He et al., 2024). Three integrated learning algorithms, namely XGBoost, were employed for the hyperspectral inversion of soil water content, with the accuracy ranking of the models as follows: XGBoost>RF> gradient lifting regression tree (Li et al., 2023).

The coal-grain composite region in Henan is extensive and serves as a significant production area for coal and grain in China (Bai et al., 2024). However, long-term and high-intensity coal mining has had a huge impact and distribution of soil organic matter,

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2305_93959412

thereby affecting crop productivity (Feng et al., 2019; Chen et al., 2024). Establishing a reliable dynamic monitoring method for retrieving for soil conservation and improvement of cropland quality (Paul, 2016).

This study examines the feasibility of employing hyperspectral remote sensing technologies to assess soil organic matter content in the Zhaogu mining area of Henan Province, China, we optimize a suitable spectral preprocessing method. The CARS algorithm is then used to screen the sensitive bands. The inversion models are constructed by combining RF, PLSR and XGBoost algorithms respectively, find the most efficient model. Our research has important theoretical and practical values. The quantitative inversion analysis of the soil reflection spectrum and soil organic matter content in coal mining regions can furnish theoretical insights for the establishment of a soil degradation monitoring network in these areas. At the same time, it also provides technical support for quantitative monitoring of SOM by spaceborne or aerial remote sensing hyperspectral technologies.

Material and methods

Study area

Zhaogu Mining area (E 113°35 '50 "-113°36' 36", N 35°24 '24 "-35°25' 11"), is selected as the research area. This area is administrated by Huixian City, Xinxiang City, Henan Province. The land use predominantly consists of agriculture, with the primary cultivation method being summer corn followed by winter wheat. The mining region intersects significantly with the agricultural land, characterizing it as a typical coal-grain composite zone. The mining area is a typical continental monsoon climate of warm temperate zone, with an average temperature of 14.1 ~ 14.9℃ and an average annual precipitation of 580~600 mm, mostly concentrated in July and August. The mining region predominantly features tidal and paddy soils, distinguished by a deep soil profile, a loose surface layer, robust nutrient availability, and effective nutrient retention, making them conducive to diverse crop cultivation. Many years of mining activities in Zhaogu mining area have led to the formation of a wide range of coal mining subsidence areas. The subsidence has resulted in waterlogged regions, significantly affecting local land resources, the ecological environment, and agricultural output (*Figure 1*).

Data collection and preprocessing

Data acquisition

In the Zhaogu Mining area, 50 samples were taken from both the low-vegetation coverage and high-vegetation coverage test areas. The collected sample data was divided into training samples and verification samples at a ratio of 7:3. One sample was chosen as a verification sample for every two samples beginning with the fifth sample, while the remaining samples served as the training samples. Some 35 samples were used for training, and another 15 samples were used for verification, across the two test areas. The original spectral data and SOM were introduced into the CARS algorithm. According to their relative importance, 32 characteristic bands were selected as potential input variables for the subsequent model.

In the study area, the NDVI values of the entire area were calculated through remote sensing images. Then, based on the distribution of NDVI values, the areas with NDVI values greater than or equal to 0.7 were defined as high vegetation coverage areas, and

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2305_93959412

the areas with NDVI values less than 0.7 were defined as low vegetation coverage areas. Utilizing NDVI distribution, the arrangement of sample points was conducted based on the distribution of the working face and the land use status in the mining area. Since a depth of 0-20cm is defined as the plough layer, it is the core distribution area of crop roots and the main area for nutrient absorption. Monitoring this layer of organic matter (SOM) can directly reflect soil fertility and agricultural production potential. In June 2023, soil sampling was conducted at a depth of 0 to 20 cm using a soil sampler with a diameter of 5.72 cm. Each sampling point was amalgamated into a single soil sample based on the five X-shaped soil samples surrounding the central point.

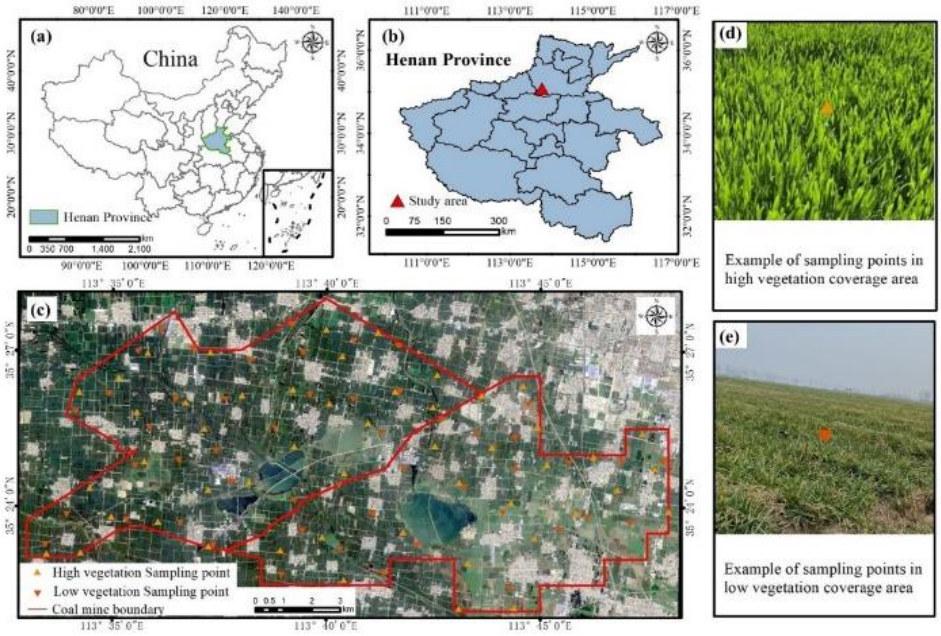

Figure 1. *Overview of the research area; (a), (b) Geographical location of the study area; (c) Sampling points; (d) Sampling sites with high vegetation cover; (e) Sampling sites with low vegetation cover*

The gathered soil samples were desiccated in a well-ventilated area of the laboratory. Upon the soil's desiccation, the dry samples were compacted and rubbed with a plexiglass rod to eliminate contaminants, including plant residues and gravel. The soil samples were divided into two parts, a part of the organic matter is measured by the potassium dichromate oxidation method. Strong oxidants are used to decompose the organic matter at high temperatures, and the content is calculated by titrating the remaining oxidant dose. While an additional portion was analyzed via laboratory spectroscopy. A portable PSR-3500 (Spectral resolution :3.5 nm) spectrometer with a wavelength spectrum of 350~2500 nm and an optical fiber with a field Angle of 25° was used to measure the spectrum of soil samples. A 50 W halogen lamp was used as the light source and a $BaSO_4$ plate was used as the calibration plate. The laboratory reflectance spectra of 100 soil samples were collected. During the measurement, to ensure that the depth of the entire sampled soil is representative, each sample was measured three times. Before each measurement, the surface was re-flattened or the sample cup was rotated 90° to reduce the error caused by uneven samples. The average of the three spectra was taken as the final spectrum of the sample.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2305_93959412

Data pre-processing

The soil spectral data was divided into two groups according to the vegetation coverage at the sampling points, and the inversion accuracies of SOM content under different crop growth conditions were compared. In order to reduce noise interference, two spectral intervals of 350-450 nm and 2400-2500 were eliminated, and the remaining 450-2400 nm of soil reflection spectra was used to invert SOM content. To effectively eliminate the spectral noise produced by soil reflection due to environmental factors and acquisition equipment, the Savitzky-Golay (SG) algorithm was employed for smoothing and denoising. This method can keep the shape and width of the signal unchanged while filtering the noise (Zhang et al., 2021). After repeatedly debugging the window size and the polynomial number, the quadratic polynomial with a window size of 15 was selected for filtering. The soil spectrum was divided into two noise regions: a severe noise region and a light noise region. The serious noise regions in high and low vegetation cover areas were 1000-1100 nm, 1800-1900 nm, 950-1050 nm, 1750-1800 nm, respectively. The light noise areas were 450~1000 nm, 1100~1800 nm, 1900~2400 nm and 50~950 nm, 1050~1750 nm, 1800~2400 nm, respectively. The soil spectral curve before and after filtering is shown in *Figure 2*. Compared with the original spectral data, the spectral quality after filtering was obviously improved. On the basis of first-order differentiation, the linear trend can also be eliminated. Therefore, we performed the second-order differential processing on the data processed by SG filter.

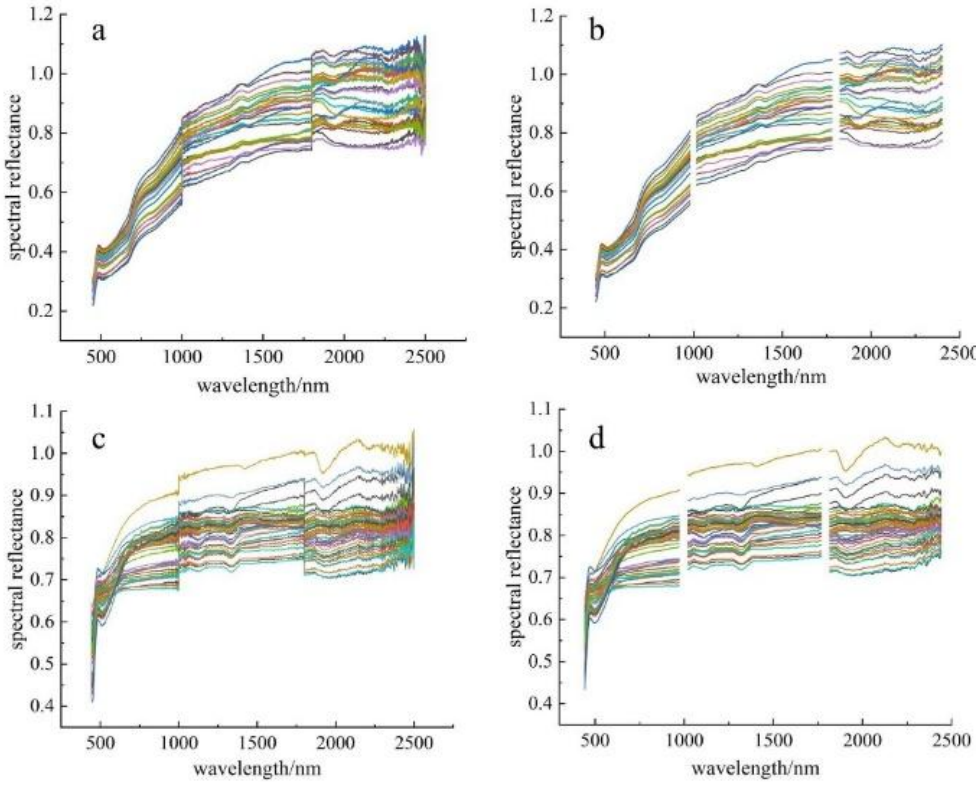

Figure 2. *Soil's Original spectrum and filtered reflectance curve (a,b) high vegetation coverage areas,(c,d)low vegetation cover areas. Annotation: Curves of different colors represent the reflectance of the soil in the vegetation-covered area varying with wavelength under different samples or conditions, reflecting the soil's ability to reflect electromagnetic waves of different wavelengths in this area*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2305_93959412

Data and processing

Figure 3 shows the schematic diagram of this research. Firstly, the collected spectral data was filtered and denoised, and the spectral features were enhanced by the second-order differential spectral changes. The CASR algorithm was used to optimize the spectral feature subset, and PLSR, RF and XGBoost were respectively used to establish the inversion models for the optimal feature combinations, so as to achieve an accurate inversion of SOM content under different vegetation coverages.

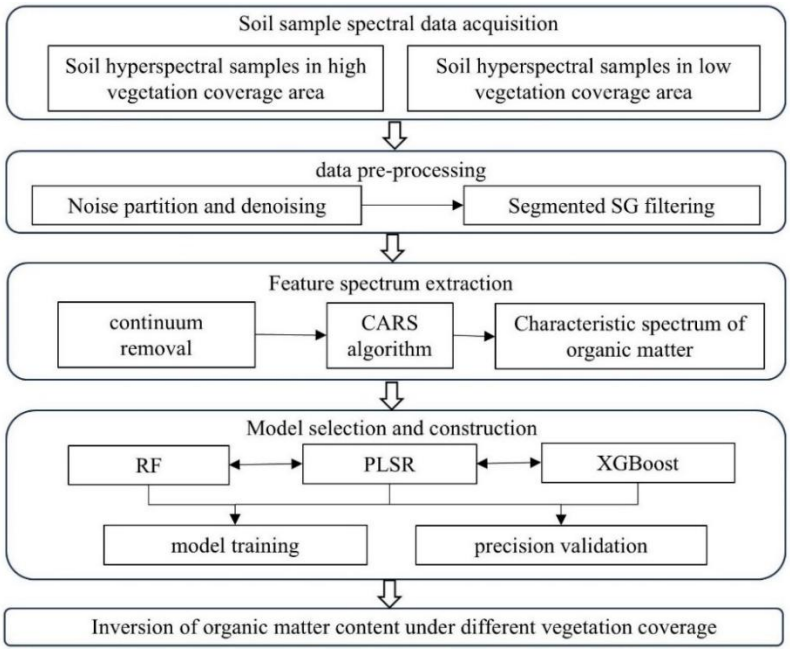

Figure 3. *Schematics diagram of this research*

CARS combines Monte Carlo (MC) and PLS regression coefficients to perform feature selection (da Silva and Wiebeck, 2018). Then, unnecessary and redundant spectral information is removed. The optimal data variable as 30 is determined by conducting Monte Carlo cross-validation modeling for variable subsets of each wavelength using the PLSR method. The algorithm is divided into four steps:

(1) Monte Carlo method is used to sample the data randomly.

(2) The attenuation index is used to evaluate the selected samples, and the samples with large absolute weights are retained. When the absolute weight value is less than the threshold, the samples are discarded, and the weight W is defined as *Eq. (1)*:

$$W = \frac{|b_i|}{\sum_{i=1}^{n} b_i} \quad \text{(Eq.1)}$$

Variable removed by CARS algorithm, set W to 0.

(3) The CARS algorithm is applied to the selection of feature bands, and the bands with a greater impact on the modeling accuracy are selected from the reflection spectrum to further reduce data redundancy.

(4) Cross-validation is employed to identify the optimal variable subset, ultimately serving as the characteristic wavelength selected by CARS.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2305_93959412

Establishment of retrieval model

Random forest

Random forest (RF) is a machine learning algorithm for classification and regression (Zhang et al., 2019). Subsequently using unselected samples for prediction in each tree. RF's arbitrary selection of features and variables mitigates the risk of overfitting in the model. In order to build the RF model, the number of variables (mtry) and the number of decision trees at the node of the binary tree in the model are adjusted. The inverse of the mean square error (MSE) is chosen as the fitness function value. In other words, the optimal model's fitness function value directly correlates with the size. The number of mtry steps is 1 at an interval of 1 to 9, and the number of decision trees is 100 at an interval of 100 to 2000.

Partial least squares method

Partial Least Squares Regression (PLSR) is capable of managing highly autocorrelated data and scenarios, demonstrating its efficacy as a multi-variable data analysis technique (Cheng and Sun, 2017). PLSR identifies the optimal function fit for a dataset by minimizing the sum of squared errors, particularly in scenarios involving significant correlations among independent variables. It can address the issue of correlation among hyperspectral data variables.

XGBoost

XGBoost algorithm is an integrated learning algorithm proposed by Chen and Guestrin (2016), involving two key components: the addition algorithm part (a strong learner is formed by the linear addition of a series of weak learners) and the forward distribution algorithm (a new learner generated in the next iteration is trained on the basis of the previous iteration). The base learner of XGBoost algorithm is the decision tree. The calculation formula is as follows *Eq. (2)*:

$$\widehat{y_i} = \sum_{p=1}^{P} f_p(W_i), \quad f_p \in F \tag{Eq.2}$$

where $\hat{y}_i$ is the inversion value of the organic matter content of the ith soil sample; f_p represents the pth decision tree;

XGBoost algorithm is (3).

$$L^{(t)} = \sum_{i=1}^{j} l\left[y_i, \hat{y}_i^{(t-1)} + f_t(W_i)\right] + \Omega(f_t) \tag{Eq.3}$$

where, $\hat{y}_i^{(t-1)}$ represents the inversion value of the organic matter content of the ith soil sample at the t-1 iteration, and l represents the loss function, which measures the error between the inverted value of the soil organic matter content $\hat{y}_i$ and the true value y_i, representing the regularization function to prevent overfitting of the model.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2305_93959412

Testing model precision

We verified the modeling performance from three aspects, where closer R^2 values to 1 indicate a better model fitting. The validation set root mean square error (RMSE) was used to assess the model's estimation ability. The model's ability to estimate is improved when the RMSE is smaller.

Results

Sample set partitioning and feature extraction

Table 1 shows the statistical results of SOM content in 50 soil samples collected in the high vegetation cover area. SOM content ranges from 8.18 to 37.11 $g \cdot kg^{-1}$ for the whole sample, 5.18 to 37.02 $g \cdot kg^{-1}$ for the modeling set, and 16.50 to 32.20 $g \cdot kg^{-1}$ for the test set. The modeling set encompasses the content range of the prediction set, with an overall coefficient of variation of 1.04, indicating medium variation. Among the 50 soil samples from the low vegetation cover area in *Table 2*, the SOM content range of the samples was 7.32-45.42 $g \cdot kg^{-1}$, the sample range of the modeling set was 13.72-45.41 $g \cdot kg^{-1}$, and the sample range of the test set was 7.32-21.31 $g \cdot kg^{-1}$, with the overall coefficient of variation of 2.86.

Table 1. *The descriptive statistics of soil organic matter content in the sample set of high vegetation cover areas*

Data set	Minimum value /(g/kg)	Maximum value /(g/kg)	Average value /(g/kg)	Median /(g/kg)	Standard deviation /(g/kg)	CV
All samples	5.18	37.11	21.22	20.30	6.27	1.04
Training	5.18	37.02	14.91	18.90	6.38	2.96
Validation Sample	16.50	32.20	24.11	23.81	5.07	4.68

Table 2. *The descriptive statistics of soil organic matter content in the sample set of low vegetation cover areas*

Data set	Minimum value /(g/kg)	Maximum value /(g/kg)	Average value /(g/kg)	Median /(g/kg)	Standard deviation /(g/kg)	CV
All samples	7.32	45.42	21.10	19.85	6.93	2.86
Training	13.72	45.41	22.77	21.01	7.30	2.87
Validation Sample	7.32	21.31	17.21	18.71	3.78	4.94

As shown in *Figure 4*, "Increasing Density" refers to the density at the sample data level, the quantity or distribution density of the samples participating in the screening in the corresponding characteristic band analysis, etc. It gradually increases with the color from blue to red, that is, the red end indicates a higher "density" situation It is used to visualize the distribution differences of soil characteristic bands under different "density" features (comparison between high vegetation coverage and low vegetation coverage areas). It is obvious that the distribution differences of soil characteristic bands in the low vegetation coverage area are more significant. The original spectral data and SOM were introduced into the CARS algorithm, and 32 feature bands were screened out according

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2305_93959412

to their importance as input variables for subsequent models. Based on the selected characteristic bands in the two test areas, the common bands were statistically obtained as shown in *Figure 5*. Among them, according to the importance dividing line of 0.4, the importance of the vast majority of bands is above 0.4, and the importance of the band near 1650 nm is the greatest, reaching 0.612. The importance is the least near the 200-300 nm band, at 0.325.

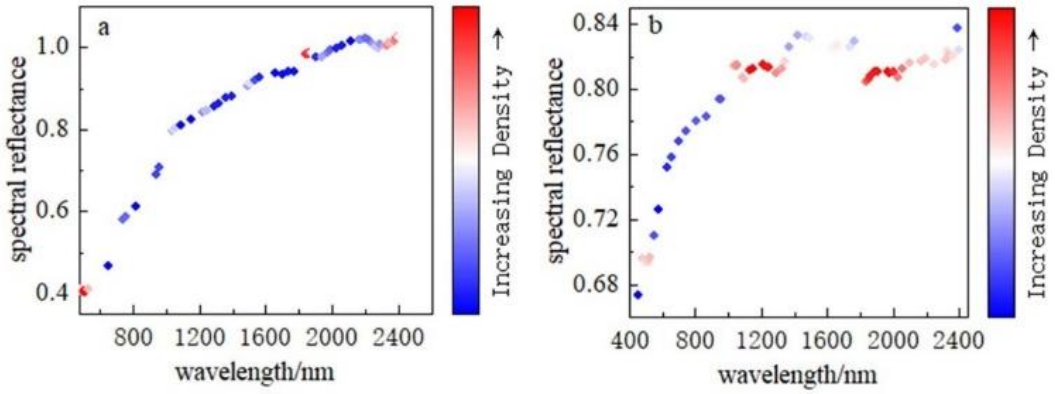

Figure 4. *Distribution of soil characteristic bands after CARS screening(a) Characteristic bands of high vegetation cover areas and (b) characteristic bands of low vegetation cover areas*

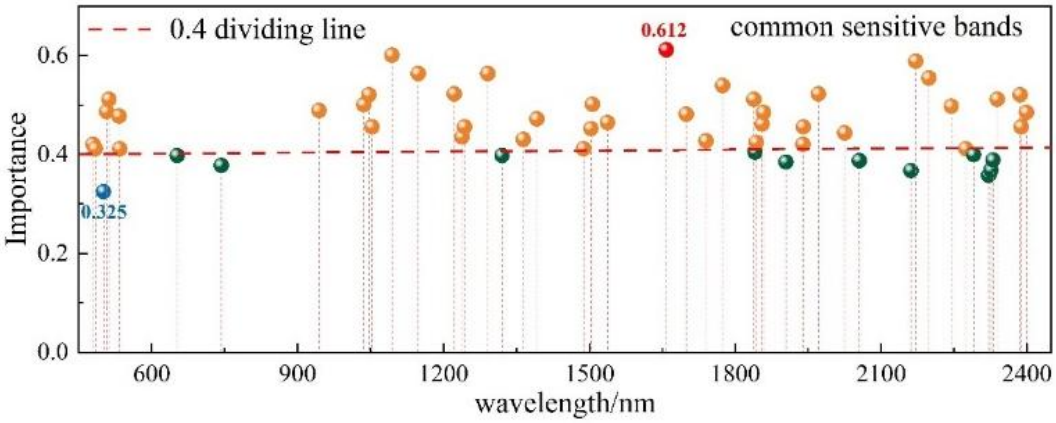

Figure 5. *Common characteristic bands of soil in different vegetation cover areas*

In this study, the 5 % area with the highest determination coefficient was used as the sensitive area of SOM. The sensitive areas of high vegetation coverage test area were mainly distributed in 407 ~ 465 nm, 652 ~ 715 nm, 988 ~ 1076 nm, 1305 ~ 1356 nm, 1743 ~ 1833 nm, 1931 ~ 1989 nm (*Figure 4a*). The sensitive areas of the low vegetation coverage test area are mainly distributed in 479 ~ 534 nm, 944 ~ 1094 nm, 1222 ~ 1291 nm, 1320 ~ 1392 nm, 1488 ~ 1537 nm, 1658 ~ 1740 nm (*Figure 4b*). On the whole, the SOM sensitive bands of soil under different vegetation coverage did not change significantly.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2305_93959412

Inversion results of soil organic matter content

Accuracy verification of the soil organic matter content in the training set

In order to verify the effectiveness of the proposed method, the feature bands of high vegetation cover areas and low vegetation cover areas in Zhaogu mining area were selected by CARS. Next, the performance of the PLSR algorithm, XGBoost algorithm and Random Forest (RF) algorithm were compared. The R^2 and RMSE in the regression model were used to evaluate the model performance. R^2 was used to measure the stability of the model. An R^2>0.8 indicated that the model was stable. RMSE was used to test the predictive ability of the model, where smaller RMSE values indicated higher accuracies. We set the learning rate to 0.1 and the maximum depth of the template to 5. We also set the number of decision trees of the RF algorithm to 1000 and the number of trees of the XGBoost algorithm to 100. The sampling times of MC was set to 40 for CARS algorithm, and 90% of the samples were extracted in each iteration. At the same time, in order to expand the range of variables selected and reduce the influence of different initial conditions on SOM content inversion, the algorithm was run five times and the optimal model was selected.

The model accuracy corresponding to each modeling method in different vegetation coverages was very similar (*Figure 6*). The accuracy measures of R^2 and RMSE of organic matter inversion by RF were 0.97 and 1.54 and 1.43, respectively. The R^2 of SOM inversion utilizing the PLSR algorithm fluctuated around 0.96, with RMSE values of 1.65 and 1.78, respectively. XGBoost algorithm showed the best inversion accuracy in the training set by an R^2 reaching 0.98. On the whole, the training accuracy of the analysis model and the R^2 of the models established by different mathematical transformations were relatively high and similar, reaching to as high as 0.95. This suggests that both the modeling set and verification set are reasonably constructed, demonstrating a satisfactory fitting without overfitting, thereby confirming the feasibility of the model for subsequent testing.

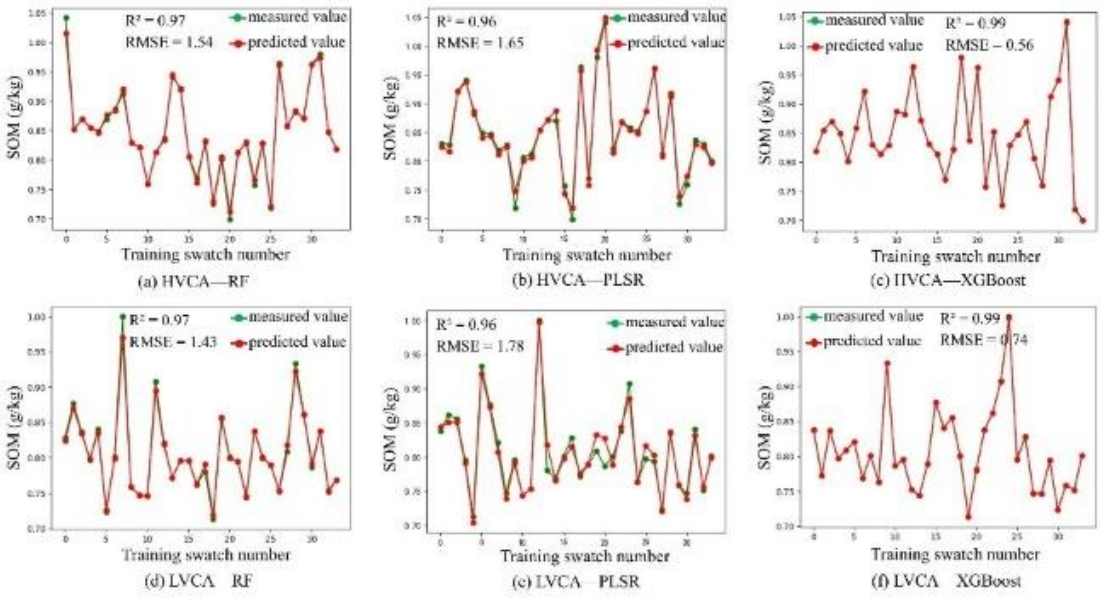

Figure 6. *Precision of soil organic matter content inversion*

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2305_93959412

Accuracy of soil organic matter content in the test set

The performance of SOM inversion models in different vegetation coverage areas in Zhaogu mining area were compared in order to select the optimal inversion models. The inversion models were established by three modeling methods, RF, PLSR and XGBoost. The two parameters of R^2 and RMSE were used for evaluation and optimal model selection.

The spectral characteristic bands selected by CARS, the inversion model. The results of the inversion model are shown in *Figure 6*. Based on the RF algorithm, the organic matter inversion in the laboratory spectra of high vegetation and low vegetation coverages in Zhaogu Mining area provided an R^2 of 0.9 and 095, both being above 0.9. The model performed relatively stable, and the R^2 of the test set and the training set were similar. RMSE reached 3.04 and 2.14, respectively, which indicated an excellent inversion of the organic matter content in the whole mining area. According to multiple cross-validations, PLSR inversion yielded R^2 of 0.79 and 0.78, and RMSE of 7.24 and 6.86, respectively. The accuracy gap was larger than RF, which may be related to the small number of samples in this experiment. The accuracy of the inversion model based on XGBoost algorithm was better than PLSR. The R^2 and RMSE of high and low vegetation cover areas were 0.85 and 0.89, and 6.21 and 5.32, respectively. Yet there was still a big gap compared with RF accuracy.

SOM content optimal inversion model in Zhaogu mining area

In order to screen the optimal inversion model of SOM content in Zhaogu mining area, the training results of different models in high vegetation cover area and low vegetation cover area were compared (as shown in *Table 3*). The performance of the three modeling methods was compared, and the optimal inversion model was selected as shown in *Figure 6*. The R^2 and RMSE exhibited varying alterations in the test set outcomes; however, the model demonstrating effective modeling retains high accuracy in the test set validation. Among them, the model established by RF algorithm had the best verifications by R^2 values of 0.90 and 0.95. The minimum RMSE values were 3.04 and 2.14, respectively, and the validation of PLSR and XGBoost models were similar, with and RMSE values ranging from 0.78 to 0.89 and 5.32 to 7.24, respectively.

Table 3. *Accuracy of the soil organic matter content inversion models*

Test Area	Model	Training set		Test set	
		R^2	RMSE	R^2	RMSE
High vegetation coverage area	RF	0.97	1.54	0.90	3.04
	PLSR	0.96	1.65	0.79	7.24
	XGBoost	0.98	1.36	0.85	6.11
Low vegetation coverage area	RF	0.97	1.43	0.95	2.14
	PLSR	0.96	1.78	0.78	6.86
	XGBoost	0.98	1.24	0.89	5.32

Furthermore, to investigate the efficacy of various modeling techniques on the test data set accuracy and the stability of SOM content, *Figure 7* (depicting the fitting to the scatter plot of SOM content test values versus predicted values) illustrates that the fitting points of the observed and predicted values for the three models are clustered around the 1:1

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2305_93959412

line, with the RF model exhibiting a higher concentration of the fitting points. The constructed RF model exhibited an excellent stability on the test set and the training set, while the other two models had a weak generalization ability on the test set. In comparison to the model's training duration and validation performance, the RF model demonstrated significant efficacy, exhibiting high predictive accuracy and stability, and was more adept at estimating SOM content. In summary, the inversion performance of RF model was significantly better than PLSR and XGBoost models, with a faster training a lower chance of overfitting. The performance of the PLSR model was marginally inferior to that of XGBoost, likely due to the limited number of input samples, as the dataset size significantly influences the efficacy of the PLSR model, which excels with larger datasets. Furthermore, the parameter adjustment process of the PLSR model is intricate, and taking into account the time expenditure and model efficacy, the RF method was deemed suitable for the inversion of SOM content spectrum in the Zhaogu mining area.

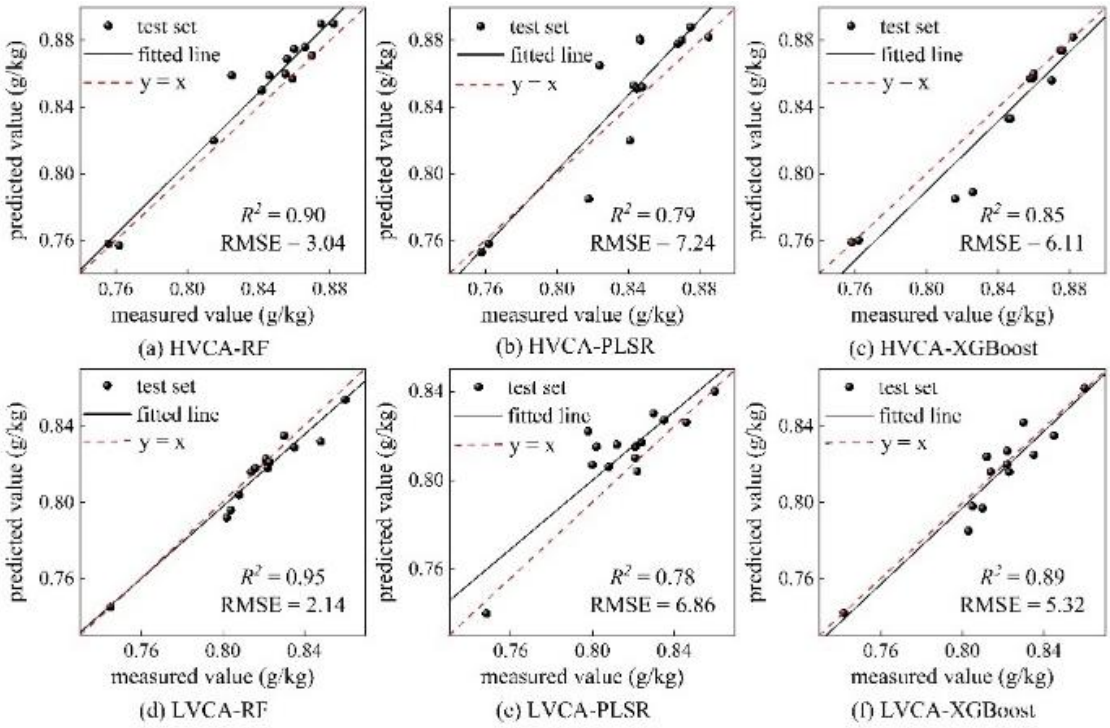

Figure 7. *Scatterplot of the results of soil spectral organic matter content inversions by different models*

Discussion

Soil hyperspectrum comprehensively represents its diverse physical and chemical properties, with a significant correlation between the two (Lu et al., 2015). Various nutritional elements in the soil and the spectral absorption peaks of other physical and chemical properties exhibit overlapping characteristics, demonstrating co-frequency and double-frequency phenomena, resulting in substantial redundant wavelength information within soil hyperspectral range (Li et al., 2020; Yuan et al., 2020). Feature selection algorithm. Currently, the commonly used feature selection algorithms in soil property inversion include correlation coefficient method (CC), genetic algorithm (GA) and CARS (He et al., 2023). The iron oxide model was 0.790, derived from multiple stepwise

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2305_93959412

regression to eliminate collinear wavelengths, based on CC's optimal band. The genetic algorithm employed by predecessors for feature selection has markedly enhanced the retrieval accuracy of soil mercury content in comparison to the direct inversion of multiple linear regression and backpropagation neural networks (Wang et al., 2015). The direct application of the CC algorithm for feature selection resulted in an excessive number of input variables in the modeling process, neglecting variable collinearity, which diminished prediction accuracy. The GA algorithm is prone to premature convergence (Altarabichi et al., 2023) and may settle into a local optimum. Conversely, the CARS algorithm effectively extracts optimal feature segment combinations by systematically eliminating redundant and insignificant features based on the importance of each band (Wang et al., 2021b).

By integrating CARS algorithm and random forest (RF) model, a hyperspectral retrieval framework for soil organic matter (SOM) content in ore-grain composite area was successfully constructed. Compared to PLSR and XGBoost, RF models show significant accuracy advantages on the test set (R^2: 0.90 -- 0.95; RMSE: 2.14-3.04), thanks to their unique dual randomness mechanism: On the one hand, sample bias is reduced through Bootstrap; On the other hand, random selection of feature subsets to split nodes effectively alleviates the interference of hyperspectral data multicollinearity to the model(SabbaghGol, Saadatfar, and Khazaiepoor 2024) . This feature enables RF to remain robust even when the sample size is limited (n=100), while XGBoost performs well in the training set (R^2=0.98), but its hyperparameter sensitivity (such as learning rate, tree depth) may lead to insufficient generalization of the test set (Yang and Shami, 2020), especially in low vegetation cover areas, where complex spectral noise further magnifies this gap. Compared with the RF model (R^2=0.89) in the source region of the three rivers in the literature (Wu et al., 2021), this study introduced CARS algorithm to screen the sensitive bands, which significantly improved the inversion accuracy of the low vegetation region (R^2=0.95), indicating that the correlation between spectral characteristics of the ore-grain composite region and SOM can be enhanced through band optimization.

However, model performance is still limited by the current data size and geographic range. Although the CARS algorithm reduces redundant band interference, the small sample size may lead to bias in feature importance assessment (Button et al., 2013), especially in high vegetation cover areas, spectral interference of litter and root exudates has not been completely decoupled. In addition, the study area concentrated on a single mining area, and did not cover the SOM variation characteristics of different climatic zones or soil types. Future research can build a spatiotemporal dynamic monitoring system through multi-region joint sampling, combining UAV hyperspectral and ground sensor networks. For example, the CARS-RF model is embedded in a lightweight edge computing device to analyze the spatial heterogeneity of SOM in real time, providing decision support for land restoration and precise fertilization in mining areas.

In conclusion, this study verifies the high efficiency of CARS-RF framework for SOM inversion in ore-grain composite region, but its application potential needs to be further verified in larger scales and more complex environments. How to balance model accuracy and computational efficiency, and how to integrate multi-source data (such as liDAR, thermal infrared) to deepen the inversion mechanism will be the focus of research in the next stage.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2305_93959412

Conclusions

In this paper, the Zhaogu mining area, a typical coal-grain composite area, was selected as the research area. After processing the soil hyperspectral data by SG denoising and second-order differential spectral transformation.

The results show that:

(1) In SOM inversion, extracting the characteristic SOM spectrum for content inversion can not only enhance the inversion mechanism and reduce data redundancy, but can also significantly improve the inversion accuracy compared with the 400-2400 nm full spectrum modeling. For the soil spectral data under different vegetation coverages in the actual experiment, the inversion accuracy based on the characteristic spectral segment modeling was improved compared with the full spectrum modeling.

(2) The CARS algorithm was used to optimize the features of the spectral data, and the bands with the greatest influence on the modeling accuracy were extracted from the reflection spectrum, and the inversion model of SOM content was constructed by combining PLSR, RF and XGBoost. Inversion results based on CARS-RF showed the best prediction performance on both training and validation sets. R^2 and RMSE were obtained 0.90 and 0.95 and 3.04 and 2.14 in high and low vegetation cover areas, respectively. The results showed that the hyperspectral retrieval framework combining SG denoising, second-order differential spectral transform, CARS and RF can provide a scientific basis for the rapid monitoring of SOM content.

(3) Since soil environment is affected by many factors, such as parent material, texture, organic matter content, etc., the influence of these factors on spectral characteristics is not deeply discussed in this paper, which may lead to insufficient adaptability of the model to the complex soil environment. Although the performance comparison of PLSR, RF and XGBoost models is mentioned in this paper, it is not deeply compared with other hyperspectral inversion methods (such as BP neural network, SVM, etc.). This limits the comprehensiveness and credibility of the findings. Therefore, validation will be carried out in more mining areas and different types of soil environments in the following studies to improve the universality and applicability of the model, and the hyperspectral remote sensing data will be fused with other sensor data (such as UAV images and ground measured data, etc.) to improve the inversion accuracy and stability.

Acknowledgments. We appreciate anonymous reviewers and their valuable comments. Also, we thank Editors for the editing and comments.

REFERENCES

[1] Ai, W., Liu, S. L., Liao, H. P., Du, J. Q., Cai, Y. L., Liao, C. L., Shi, H. W., Lin, Y. D., Junaid, M., Yue, X. J., Wang, J. (2022): Application of hyperspectral imaging technology in the rapid identification of microplastics in farmland soil. – Science of The Total Environment 807: 151030. doi: https://doi.org/10.1016/j.scitotenv.2021.151030.

[2] Altarabichi, M. G., Nowaczyk, S., Pashami, S., Mashhadi, P. S. (2023): Fast Genetic Algorithm for feature selection—A qualitative approximation approach. – Expert Systems with Applications 211: 118528. doi: https://doi.org/10.1016/j.eswa.2022.118528.

[3] Bai, E., Guo, W. B., Zhang, H. B., Tan, Y., Li, X. Y., Wei, Z. Y. (2024): Degradation mechanism of cultivated land and its protection technology in the central coal-grain overlapped area of China. – Journal of Cleaner Production 468: 143075. doi: https://doi.org/10.1016/j.jclepro.2024.143075.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2305_93959412

[4] Bazaluk, O., Kuchyn, O., Saik, P., Soltabayeva, S., Brui, H., Lozynskyi, V., Cherniaiev, O. (2023): Impact of ground surface subsidence caused by underground coal mining on natural gas pipeline. – Scientific Reports 13(1): 19327. doi: 10.1038/s41598-023-46814-5.
[5] Button, K. S., Ioannidis, J. P. A., Mokrysz, C., Nosek, B. A., Flint, J., Robinson, E. S. J., Munafò, M. R. (2013): Power failure: why small sample size undermines the reliability of neuroscience. – Nature Reviews Neuroscience 14(5): 365-376. doi: 10.1038/nrn3475.
[6] Chen, T. Q., Guestrin, C. (2016): XGBoost: A Scalable Tree Boosting System. – Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining.
[7] Chen, W. Q., Xiao, W., He, T. T., Ruan, L. L., Zhao, Y. L., Hu, Z. Q. (2024): Quantify the extensive crop damage and grain losses caused by underground coal mining subsidence in eastern China. – Journal of Cleaner Production 469: 143204. doi: https://doi.org/10.1016/j.jclepro.2024.143204.
[8] Cheng, J. H., Sun, D. W. (2017): Partial Least Squares Regression (PLSR) Applied to NIR and HSI Spectral Data Modeling to Predict Chemical Properties of Fish Muscle. – Food Engineering Reviews 9(1): 36-49. doi: 10.1007/s12393-016-9147-1.
[9] Chenu, C., Rumpel, C., Védère, C., Barré, P. (2024): Methods for studying soil organic matter: nature, dynamics, spatial accessibility, and interactions with minerals. – In: Soil Microbiology, Ecology and Biochemistry (Fifth Edition). Elsevier, Chapter 13, pp. 369-406.
[10] da Silva, D. J., Wiebeck, H. (2018): CARS-PLS regression and ATR-FTIR spectroscopy for eco-friendly and fast composition analyses of LDPE/HDPE blends. – Journal of Polymer Research 25(5): 112. doi: 10.1007/s10965-018-1507-5.
[11] Dai, X., Wang, Z. K., Liu, S. X., Yao, Y. Z., Zhao, R., Xiang, T. Y., Fu, T. Z., Feng, H. P., Xiao, L. X., Yang, X. H., Wang, S. M. (2022): Hyperspectral imagery reveals large spatial variations of heavy metal content in agricultural soil - A case study of remote-sensing inversion based on Orbita Hyperspectral Satellites (OHS) imagery. – Journal of Cleaner Production 380: 134878. doi: https://doi.org/10.1016/j.jclepro.2022.134878.
[12] Duo, L. H., Wang, J. Q., Zhong, Y. P., Jiang, C. Q., Chen, Y. Y., Guo, X. F. (2024): Ecological environment quality assessment of coal mining cities based on GEE platform: A case study of Shuozhou, China. – International Journal of Coal Science & Technology 11(1): 75. doi: 10.1007/s40789-024-00723-8.
[13] Feng, Y., Wang, J. M., Bai, Z. K., Reading, L. (2019): Effects of surface coal mining and land reclamation on soil properties: A review. – Earth-Science Reviews 191: 12-25. doi: https://doi.org/10.1016/j.earscirev.2019.02.015.
[14] Filonchyk, M., Peterson, M. P., Yan, H. W., Gusev, A., Zhang, L. F., He, Y., Yang, S. W. (2024): Greenhouse gas emissions and reduction strategies for the world's largest greenhouse gas emitters. – Science of The Total Environment 944: 173895. doi: https://doi.org/10.1016/j.scitotenv.2024.173895.
[15] Gu, X. H., Wang, Y. C., Sun, Q., Yang, G. J., Zhang, C. (2019): Hyperspectral inversion of soil organic matter content in cultivated land based on wavelet transform. – Computers and Electronics in Agriculture 167: 105053. doi: https://doi.org/10.1016/j.compag.2019.105053.
[16] Hartmann, M., Six, J. (2023): Soil structure and microbiome functions in agroecosystems. – Nature Reviews Earth & Environment 4(1): 4-18. doi: 10.1038/s43017-022-00366-w.
[17] He, J. C., He, J., Liu, G., Li, W. L., Li, Z., Li, Z. (2023): Inversion analysis of soil nitrogen content using hyperspectral images with different preprocessing methods. – Ecological Informatics 78: 102381. doi: https://doi.org/10.1016/j.ecoinf.2023.102381.
[18] He, S. F., Zhou, L., Xie, H. X., Tan, S. Q. (2024): Enhancing XGBoost's accuracy in soil organic matter prediction through feature fusion. – Paddy and Water Environment 22(3): 475-489. doi: 10.1007/s10333-024-00980-y.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2305_93959412

[19] Huang, F. M., Cao, Z. S., Guo, J. F., Jiang, S-H., Li, S., Guo, Z. Z. (2020): Comparisons of heuristic, general statistical and machine learning models for landslide susceptibility prediction and mapping. – Catena 191: 104580. doi: https://doi.org/10.1016/j.catena.2020.104580.
[20] Jones, M. W., Peters, G. P., Gasser, T., Andrew, R. M., Schwingshackl, C., Gütschow, J., Houghton, R. A., Friedlingstein, P., Pongratz, J., Le Quéré, C. (2023): National contributions to climate change due to historical emissions of carbon dioxide, methane, and nitrous oxide since 1850. – Scientific Data 10(1): 155. doi: 10.1038/s41597-023-02041-1.
[21] Lee, J., Oh, Y., Lee, S. T., Seo, Y. O., Yun, J., Yang, Y., Kim, J., Zhuang, Q., Kang, H. (2023): Soil organic carbon is a key determinant of CH4 sink in global forest soils. – Nature Communications 14(1): 3110. doi: 10.1038/s41467-023-38905-8.
[22] Li, J. Q., Nie, M., Pendall, E. (2020): Soil physico-chemical properties are more important than microbial diversity and enzyme activity in controlling carbon and nitrogen stocks near Sydney, Australia. – Geoderma 366: 114201. doi: https://doi.org/10.1016/j.geoderma.2020.114201.
[23] Li, Z., Lu, T. D., Yu, K. G., Wang, J. (2023): Interpolation of GNSS Position Time Series Using GBDT, XGBoost, and RF Machine Learning Algorithms and Models Error Analysis. – Remote Sensing 15(18): 4374. doi:10.3390/rs15184374.
[24] Li, H., Ju, W. L., Song, Y. M., Cao, Y. Y., Yang, W., Li, M. Z. (2024): Soil organic matter content prediction based on two-branch convolutional neural network combining image and spectral features. – Computers and Electronics in Agriculture 217: 108561. doi: https://doi.org/10.1016/j.compag.2023.108561.
[25] Lu, Y-L., Bai, Y-L., Yang, L-P., Wang, L., Wang, Y-L., Ni, L., Zhou, L-P. (2015): Hyper-spectral characteristics and classification of farmland soil in northeast of China. – Journal of Integrative Agriculture 14(12): 2521-2528. doi: https://doi.org/10.1016/S2095-3119(15)61232-1.
[26] Paul, E. A. (2016): The nature and dynamics of soil organic matter: Plant inputs, microbial transformations, and organic matter stabilization. – Soil Biology and Biochemistry 98: 109-126. doi: https://doi.org/10.1016/j.soilbio.2016.04.001.
[27] SabbaghGol, H., Saadatfar, H., Khazaiepoor, M. (2024): Evolution of the random subset feature selection algorithm for classification problem. – Knowledge-Based Systems 285: 111352. doi: https://doi.org/10.1016/j.knosys.2023.111352.
[28] Sokol, N. W., Slessarev, E., Marschmann, G. L., Nicolas, A., Blazewicz, S. J., Brodie, E. L., Firestone, M. K., Foley, M. M., Hestrin, R., Hungate, B. A., Koch, B. J., Stone, B. W., Sullivan, M. B., Zablocki, O., Trubl, G., McFarlane, K., Stuart, R., Nuccio, E., Weber, P., Jiao, Y. Q., Zavarin, M., Kimbrel, J., Morrison, K., Adhikari, D., Bhattacharaya, A., Nico, P., Tang, J. Y., Didonato, N., Paša-Tolić, L., Greenlon, A., Sieradzki, E. T., Dijkstra, P., Schwartz, E., Sachdeva, R., Banfield, J., Pett-Ridge, J. (2022): Life and death in the soil microbiome: how ecological processes influence biogeochemistry. – Nature Reviews Microbiology 20(7): 415-430. doi: 10.1038/s41579-022-00695-z.
[29] Sun, M. Y., Liu, H. G., Li, P. F., Gong, P., Yu, X. Y., Ye, F. H., Guo, Y., Wu, Z. K. (2024): Effects of salt content and particle size on spectral reflectance and model accuracy: Estimating soil salt content in arid, saline-alkali lands. – Microchemical Journal 207: 111666. doi: https://doi.org/10.1016/j.microc.2024.111666.
[30] Wang, L., Zeng, Y., Chen, T. (2015): Back propagation neural network with adaptive differential evolution algorithm for time series forecasting. – Expert Systems with Applications 42(2): 855-863. doi: https://doi.org/10.1016/j.eswa.2014.08.018.
[31] Wang, P., Lin, C-K., Wang, Y., Liu, D. C., Song, D. J., Wu, T. (2021a): Location-specific co-benefits of carbon emissions reduction from coal-fired power plants in China. – Nature Communications 12(1): 6948. doi: 10.1038/s41467-021-27252-1.
[32] Wang, L. X., Jiang, S. Y., Jiang, S. Y. (2021b): A feature selection method via analysis of relevance, redundancy, and interaction. – Expert Systems with Applications 183: 115365. doi: https://doi.org/10.1016/j.eswa.2021.115365.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2305_93959412

[33] Wang, Z., Du, Z. P., Li, X. Y., Bao, Z. Y., Zhao, N., Yue, T. X. (2021b): Incorporation of high accuracy surface modeling into machine learning to improve soil organic matter mapping. – Ecological Indicators 129: 107975. doi: https://doi.org/10.1016/j.ecolind.2021.107975.
[34] Wang, Z. H., Cai, Y. D., Liu, D. M., Lu, J., Qiu, F., Hu, J. H., Li, Z. T., Gamage, R. P. (2024): A review of machine learning applications to geophysical logging inversion of unconventional gas reservoir parameters. – Earth-Science Reviews 258: 104969. doi: https://doi.org/10.1016/j.earscirev.2024.104969.
[35] Wang, Q. C., Shan, Y., Shi, W. B., Zhao, F. B., Li, Q., Sun, P. C., Wu, Y. P. (2024a): Assessing spatiotemporal variations of soil organic carbon and its vulnerability to climate change: a bottom-up machine learning approach. – Climate Smart Agriculture 1(2):100025. doi: https://doi.org/10.1016/j.csag.2024.100025.
[36] Wang, C. Y., Gao, B. B., Yang, K., Wang, Y. X., Sukhbaatar, C., Yin, Y., Feng, Q. L., Yao, X. C., Zhang, Z. H., Yang, J. Y. (2024b): Inversion of soil organic carbon content based on the two-point machine learning method. – Science of The Total Environment 943: 173608. doi: https://doi.org/10.1016/j.scitotenv.2024.173608.
[37] Wu, D., Jia, K. L., Zhang, X. D., Zhang, J. H., Abd El-Hamid, H. T. (2021): Remote Sensing Inversion for Simulation of Soil Salinization Based on Hyperspectral Data and Ground Analysis in Yinchuan, China. – Natural Resources Research 30(6): 4641-4656. doi: 10.1007/s11053-021-09925-2.
[38] Wu, M. J., Huang, Y. Q., Zhao, X., Jin, J., Ruan, Y. C. (2024): Effects of different spectral processing methods on soil organic matter prediction based on VNIR-SWIR spectroscopy in karst areas, Southwest China. – Journal of Soils and Sediments 24(2): 914-927. doi: 10.1007/s11368-023-03691-9.
[39] Xia, K., Xia, S. S., Shen, Q., Yang, B., Song, Q., Xu, Y. F., Zhang, S. W., Zhou, X., Zhou, Y. (2021): Moisture spectral characteristics and hyperspectral inversion of fly ash-filled reconstructed soil. – Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 253: 119590. doi: https://doi.org/10.1016/j.saa.2021.119590.
[40] Yang, L., Shami, A. (2020): On hyperparameter optimization of machine learning algorithms: Theory and practice. – Neurocomputing 415: 295-316. doi: https://doi.org/10.1016/j.neucom.2020.07.061.
[41] Yang, C. B., Feng, M. C., Song, L. F., Wang, C., Yang, W., Xie, Y. K., Jing, B. H., Xiao, L. J., Zhang, M. J., Song, X. Y., Saleem, M. (2021): Study on hyperspectral estimation model of soil organic carbon content in the wheat field under different water treatments. – Scientific Reports 11: 18582. doi: 10.1038/s41598-021-98143-0.
[42] Ye, S. F., Wang, D., Min, S. G. (2008): Successive projections algorithm combined with uninformative variable elimination for spectral variable selection. – Chemometrics and Intelligent Laboratory Systems 91(2): 194-199. doi: https://doi.org/10.1016/j.chemolab.2007.11.005.
[43] Yuan, J., Wang, X., Yan, C. X., Chen, S. B., Wang, S. R., Zhang, J. Q., Xu, Z. Y., Ju, X. P., Ding, N., Dong, Y. Z., Zhang, W. F. (2020): Wavelength Selection for Estimating Soil Organic Matter Contents Through the Radiative Transfer Model. – IEEE Access 8: 176286-176293. doi: 10.1109/ACCESS.2020.3026813.
[44] Yuan, L.-M., Mao, F., Huang, G. Z., Chen, X. J., Wu, D., Li, S. J., Zhou, X. Q., Jiang, Q. J., Lin, D. P., He, R. Y. (2020a): Models fused with successive CARS-PLS for measurement of the soluble solids content of Chinese bayberry by vis-NIRS technology. – Postharvest Biology and Technology 169: 111308. doi: https://doi.org/10.1016/j.postharvbio.2020.111308.
[45] Yuan, J., Gao, J. C., Yu, B., Yan, C. X., Ma, C. R., Xu, J. W., Liu, Y. T. (2024): Estimation of soil organic matter content based on spectral indices constructed by improved Hapke model. – Geoderma 443: 116823. doi: https://doi.org/10.1016/j.geoderma.2024.116823.
[46] Zhang, L., Huettmann, F., Liu, S. R., Sun, P. S., Yu, Z., Zhang, X. D., Mi, C. R. (2019): Classification and regression with random forests as a standard method for presence-only

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2305_93959412

data SDMs: A future conservation example using China tree species. – Ecological Informatics 52: 46-56. doi: https://doi.org/10.1016/j.ecoinf.2019.05.003.

[47] Zhang, G. S., Hao, H., Wang, Y. C., Jiang, Y., Shi, J. H., Yu, J., Cui, X. J., Li, J. S., Zhou, S., Yu, B. L. (2021): Optimized adaptive Savitzky-Golay filtering algorithm based on deep learning network for absorption spectroscopy. – Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy 263: 120187.
doi: https://doi.org/10.1016/j.saa.2021.120187.

http://www.aloki.hu ● ISSN 1589 1623 (Print) ● ISSN 1785 0037 (Online)
DOI: http://dx.doi.org/10.15666/aeer/2305_93959412