Scholar Publishing Group
Academic Journal of Agricultural Sciences
https://doi.org/ 10.38007/AJAS.2020.010105
ISSN 2789-8415 Vol. 1, Issue 1: 50-64

Hyperspectral Prediction Model of Soil Organic Matter in Black Soil Region Based on Fuzzy Recognition

Xinxin Zhang*

Fujian Agriculture and Forestry University, Fujian, China

zhangxinxin@m.fafu.edu.cn

**corresponding author*

Keywords: Fuzzy Recognition, Organic Matter, Hyperspectral Model, Linear Regression

Abstract: Land is an important basic resource that must be relied upon, irreplaceable and non-renewable for human survival and development. The limited quantity and the scarcity of supply determine that we must make full use of land resources. The use of hyperspectral remote sensing to establish accurate soil organic matter prediction models is an inevitable requirement for rapid determination of soil organic matter, precision agriculture, and soil carbon pool estimation. Accurate understanding of soil distribution information is an important basis for soil mapping and establishment of soil databases, and has important guiding significance for land management and land use planning in China. Based on the above background, the purpose of this paper is to study the hyperspectral prediction model of soil organic matter in black soil region based on fuzzy recognition, and to establish a high-spectral estimation model of soil organic matter by linear regression and fuzzy recognition method. In this paper, the fuzzy closeness inversion model of soil organic matter content is established, and compared with the results of the linear regression model; the accuracy of the linear model is basically consistent and low. The accuracy of the fuzzy closeness inversion model is higher, and the lowest is also It reached 89.71%, and the rest were more than 90%. In addition, based on the diversity of spectral inversion factors and the dynamics of organic matter content, an interval-valued fuzzy inversion model was proposed. The prediction result is accurate for 5 samples. There is an error, and the sample deviation with error is 3.22%. This shows that the fuzzy recognition method can effectively solve the problems of nonlinearity and ambiguity, and it is feasible to use the fuzzy recognition method for soil organic matter content prediction.

1. Introduction

As one of the necessary conditions for the survival and development of human beings, soil

resources are an irreplaceable and important basic resource [1]. In recent years, with the development of the economy and the rapid growth of the population, more and more non-agricultural land has been used, resulting in excessive land reclamation and less and less land available [2-3]. Soil as the basis of agricultural production, access to continuous and reliable soil attribute information and spatial distribution information is particularly important for the realization of modern precision agriculture [4]. Therefore, it is an important task to establish a reasonable soil physical and chemical traits indicator detection system and obtain soil information quickly and accurately [5]. Therefore, how to accurately and quickly obtain soil-related information is a major challenge faced by modern science and technology workers [6]. Previous studies have found that the spectral characteristics of soil are closely related to the physical and chemical properties of soil. Soil physicochemical properties affect soil spectrum, such as soil water content, organic matter content, iron oxides, mechanical composition and texture [7].

The hyperspectral remote sensing technology developed by the imaging spectrometer in the 1980s to obtain many very narrow continuous spectra has been widely used in the prediction of soil physical and chemical indicators, opening up a new way for rapid and effective detection of soil water content. [8-10]. There are many bands of hyperspectral remote sensing, and the generated spectral curve is complete and continuous, which can accurately reflect the spectral characteristics of the ground object itself and the spectral difference between the features [11]. Therefore, the use of hyperspectral technology can achieve high-precision inversion of soil physical and chemical components quickly, easily and without damage [12]. The use of this technology has brought us unlimited convenience, which not only saves a lot of manpower and material resources, but also makes it possible to collect soil data in large areas [13]. Hyperspectral remote sensing technology has become an important means for obtaining information on farmland soil properties, and it also provides a scientific basis for evaluating and protecting land resources [14]. The theoretical study on the practical application of hyperspectral remote sensing monitoring of soil organic matter was carried out. The fuzzy identification method was used to establish a hyperspectral inversion model of soil organic matter content, and the modeling method of poor correlation between soil organic matter and spectral reflectance was discussed. The rapid, quantitative and accurate monitoring of soil organic matter has important guiding significance [15].

In this paper, the fuzzy closeness inversion model of soil organic matter content is established, and compared with the results of the linear regression model, the accuracy of the linear model is basically consistent and low. The accuracy of the fuzzy closeness inversion model is higher, and the lowest is also It reached 89.71%, and the rest were more than 90%. In addition, based on the diversity of spectral inversion factors and the dynamics of organic matter content, an interval-valued fuzzy inversion model was proposed. The prediction result is accurate for 5 samples. There is an error, and the sample deviation with error is 3.22%. This shows that the fuzzy recognition method can effectively solve the problems of nonlinearity and ambiguity, and it is feasible to use the fuzzy recognition method for soil organic matter content prediction.

The first chapter is an overview. It mainly introduces the hyperspectral prediction model of soil organic matter in black soil based on fuzzy recognition, and states the background, research significance and work done in this paper. The second chapter introduces the method of soil organic matter hyperspectral modeling, and gives a detailed overview of several modeling methods and basic fuzzy pattern recognition. The third chapter is an experiment that details the source of the experimental data and the type of data, and informs the experimental environment in which the experiment is performed and the steps that the experiment needs to perform. The fourth chapter is the experimental analysis, the experimental examples are unfolded, and the experimental content, experimental process and experimental details are introduced. The fifth chapter analyzes and summarizes, this chapter is mainly to carry out experimental analysis and summary of the

experiment.

2. Proposed Method

2.1. Related Work

The purpose of the Si studied was to find out the effect of soil particle size on hyperspectral prediction of soil organic matter content in order to obtain a suitable soil sample size and reduce unnecessary work. Soil samples were screened at 10, 20, 60 and 100 mesh and then measured in the laboratory using an ASD Fieldspec Pro spectrophotometer. The soil spectral reflectance (R) is mathematically converted to the first derivative of the reflectivity (R') and the logarithm of the reciprocal of the reflectivity (Log(1 / R)). The SOM hyperspectral prediction model was established by partial least squares regression (PLSR), support vector machine (SVM) and PLSR-SVM. The results show that the soil particle size has a significant effect on the spectral reflectance of the soil. The smaller the soil particles, the higher the spectral reflectance of the soil. When the soil particle size is less than 0.25mm, it does not help to improve the accuracy of SOM prediction [16]. Qiao evaluated the potential of visible and near-infrared reflectance spectra for soil organic matter evaluation. The original spectra were processed using six pretreatment methods. The partial least squares regression method was used to establish the prediction model and the optimal spectral preprocessing method was evaluated. Correlation analysis and partial least squares regression analysis were used to determine the effective wavelength of soil organic matter. The results show that: (1) visible and near-infrared reflectance spectroscopy is an ideal method for evaluating soil organic matter; (2) different pretreatment spectra can improve its correlation with soil organic matter; Qiao researched will utilize visible and near-infrared reflectance spectroscopy Provide on-site management of agricultural inputs [17]. Quantitative monitoring of soil organic matter in cultivated land, mastering the law of spatial variation, and contributing to fertility adjustment and sustainable agricultural development. The Yu studied aims to analyze the response relationship between SOM and reflectivity of hyperspectral images of different pixel sizes, and to establish the best model for estimating SOM using imaging spectroscopy. The correlation between hyperspectral reflectance and SOM was analyzed by wavelet transform method. Then the optimal pixel size and sensitive wavelet feature scale are selected, and the inversion model of SOM is established. The results show that the wavelet transform helps to improve the correlation between soil hyperspectral characteristics and soil organic matter [18]. Wang studied the soil pH of a county in a certain province. The spectral reflectance of the ASD Field Spec HR portable terrain spectrum was measured and its spectral characteristics were analyzed. The soil pH spectrum prediction model was established by using the first-order deviation, the second-order deviation, the logarithm of the reciprocal logarithm, the first-order difference of the reciprocal logarithm and the second-order difference of the reciprocal logarithm of the original spectral reflectance of the soil. The results show that the correlation between the reflectance spectra of SNV pretreatment and soil pH is significantly improved. The soil pH optimal prediction model established by partial least squares is a prediction model based on the inverse logarithm of the spectral reflectance [19]. The video tracking of drug tablets has an important impact on the efficiency and reliability of mass production. Due to the high similarity and randomness of the targets to be searched, this topic has become a difficult point and target for drug production monitoring in recent years. Liang proposed a series of mathematical properties describing the image of reflected light intensity, and then used a set of fuzzy recognition systems and their recognition rules to classify the motion plates of the input image properties. This helps to determine their instantaneous coordinate position on a given image frame accordingly. The real-time motion of the tablet target is successfully tracked by repeating the same operation in the next frame. The accuracy and reliability of the method in the pharmaceutical

industry were verified by orthogonal tracking experiments and performance comparison [20].

2.2. Soil Organic Matter Hyperspectral Modeling Method

(1) Partial least squares regression

The Principal Component Regression we used in the past can guarantee the accuracy of the model by using only independent variables to model the noise, but when dealing with variables with less correlation, it is easy to miss the useful information and the accuracy of the model is degraded. Partial Least Squares Regression (PLSR) is a new type of multivariate statistical method. It can decompose the variables x and y together and extract the factors at the same time. It is especially suitable for the regression of multi-independent variables to multiple independent variables. mold. Hyperspectral data contains continuous spectral information, often containing thousands of reflectances, and the corresponding number of samples is relatively small. Therefore, in this paper, the partial least squares method is mainly used in the study of soil hyperspectral modeling.

The basic principle of partial least squares is to take the acquired hyperspectral data as an independent variable, and set a total of *m* independent variables $X_1, X_2, ..., X_m$ soil attribute data such as soil organic matter content, a total of *n* variables are YY. First, the independent variable and the dependent variable are normalized, then the first principal component T_1 is extracted from the independent variable, and the linear regression equation with the dependent variable is established. According to the accuracy requirement, iterative iteration can be performed until the model accuracy is satisfactory. In this way, we obtain the regression equation $T_1, T_2, ..., T_r$ of the dependent variable *Y* and *r* principal components, and then convert it into the partial least squares regression equation of the dependent variable *Y* and the independent variable *X*.

This study used a company's Unscrambler 9.7 for partial least squares modeling. Import the hyperspectral data into the software separately, set the arguments and dependent variables in the Modify.Edit set, and select the number of samples to participate in the model validation. Select Partial Least Squares in Task.Regression and Full Cross Validation in Cross-Validation. The accuracy of the model verification can be obtained in the obtained results. In summary, the first step of extracting the principal component not only ensures that more factor information is extracted while maintaining independence, and the correlation between the independent variable and the dependent variable is the largest. Since the independent variable has eliminated the multicollinearity, the results obtained at this time are more stable. Therefore, the partial least squares method can be understood as a new multivariate statistical method that integrates principal component analysis, correlation analysis and multiple linear regression.

(2) Multiple stepwise regression

Multi Linear Regression (MLR) is a linear regression model with multiple explanatory variables, ie two or more independent variables, used to reveal the linear relationship between variables and multiple explanatory variables. For:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p + \varepsilon \quad (1)$$

There are a total of *p* explanatory variables, namely, independent variables, and the interpreted variable *y* is explained by two parts. In this paper, the independent variable $x_1, x_2, ..., x_p$ is the soil hyperspectral reflectance, and the interpreted variable $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_p x_p$ is the measured soil organic matter content. The first part is *Y*, which represents a linear change in *y* caused by a change in p explanatory variables *x*: the second part ε is a random error indicating a linear change in *y* caused by other factors. Among them, ε and $\beta_0, \beta_1, ..., \beta_p$ are unknown

parameters in the model, which are expressed as regression constant and partial regression coefficient. If you want to expect both sides of the formula, you can get the following formula:

$$E(y)=\beta_0+\beta_1x_1+\beta_2x_2+...+\beta_px_p \quad (2)$$

That is, multiple linear regression equations, so the first problem to be solved when using multiple linear regression equations is to estimate the value of the unknown parameter $\beta_0,\beta_1,...,\beta_p$. In practical problems, we hope to select some variables that have a greater influence on the dependent variable *y* as independent variables, that is, we hope to include all the factors that have significant influence on the dependent variable *y* in the regression equation, and those that have significant influence on the regression equation. They are eliminated. Multi Stepwise Regression is a method based on this principle. It considers the effect of all independent variables on the dependent variable *y*, the degree of significance or the contribution rate, according to the ranking from large too small. Bring them into the equation. Those that are not significant to *y* may never be introduced into the regression equation, and those variables that have been introduced into the equation may lose their significance after the new variable is added, and therefore will be removed from the equation. This ensures that after the final equation is obtained, the variables in the equation are significant variables for *y*. This study used SPSSl9.0 for multiple stepwise linear regression.

(3) Orthogonal experimental design

Orthogonal Experimental Design (OED) is a method for studying and processing multi-factor experiments. For single factor or two factors, because of the few factors involved, the experimental design and analysis are relatively simple. In actual work, we often encounter many factors. If we choose a comprehensive test, the test will become very large, and often it is impossible to conduct large-scale tests due to some reasons such as test conditions and project funding. Therefore, the orthogonal experimental design is an efficient learning method for finding the optimal level combination under this multi-factor condition. At present, this method has been applied to the chemical, pharmaceutical and other industries, but it is rare in soil hyperspectral applications. In the orthogonal test design, all the factors are not combined to judge the optimal level combination, but in the whole horizontal combination, a representative horizontal combination is selected for testing, which greatly reduces the workload. E.g:

There is a three-factor, three-level test, A factor, there are three levels A1, A2, A3; B factor, there are three levels Bl, B2, B3; the same C factor has three levels C1, C2, C3. If you combine them at random, there are a total of 27 combinations. Similarly, if you increase the number of factors or the number of levels, then the number of trials will increase, the workload will increase, and in some cases it will not be completed. Orthogonal test design can be tested by orthogonal table design to find the optimal level combination. For the above three levels, the test of the three factors, if you do not consider the interaction between the factors, you can use the orthogonal table $L_9(3^4)$, using only 9 levels of combination, you can reflect the comprehensive test of 27 levels, and then find the optimal level combination.

2.3. Fuzzy Pattern Recognition

Pattern recognition is a basic intelligence possessed by human beings. It refers to the recognition of phenomena or things that are the same or similar to the known patterns from the phenomena or things to be identified. It is an important sub-discipline of artificial intelligence. Modern pattern recognition has been widely used in many fields, such as the recognition and classification of language, symbols, words, scenes and features of people (fingerprints, features.), as well as for remote sensing images and meteorological conditions., seismic information, hydrological

phenomena, and medical identification and analysis.

In nature, most of the objective things and phenomena have different degrees of ambiguity and uncertainty. The fuzzy recognition model can effectively solve the ambiguity and uncertainty of objective things and phenomena, and better establish the model. . Fuzzy pattern recognition is a combination of fuzzy mathematics recognition method. Based on fuzzy set theory, it can identify and classify objective things and phenomena with ambiguity and uncertainty more effectively than traditional modeling methods.

Fuzzy mode can be divided into H types: static pattern recognition (including (I) type, (II) type static pattern recognition) and dynamic pattern recognition. The difference between static pattern recognition and dynamic pattern recognition is whether the known pattern changes with time and conditions. If the change is changed, such as meteorological and hydrological predictions, it belongs to dynamic pattern recognition: the difference between (I) type and (II) type static pattern recognition is whether the fuzzy element or the fuzzy set is to be identified, if the object to be identified is fuzzy Meta, then (I) type static pattern recognition, if the object to be identified is a fuzzy set, then it is (II) type static pattern recognition. The basic step diagram of fuzzy pattern recognition is shown in Figure 1. The object to be identified is pre-processed first, and the features that can be identified are extracted. Then, the pattern recognition is performed according to the fuzzy pattern recognition method, and the category corresponding to the object to be clarified is output after recognition.

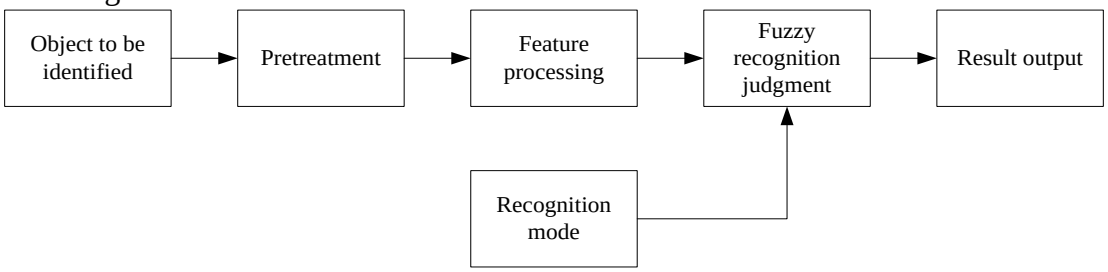

Figure 1. Fuzzy recognition step diagram

(1) Pre-processing part

This is the process of processing the information in the first paragraph. It filters out the information obtained from the object to be identified and normalizes and digitizes it.

(2) Feature Processing Section

This is the process of extracting some data from the signal that reflects the characteristics of the pattern. The patterns or objects of recognition are usually very different. They all contain their own characteristics, which are the inherent factors or preconditions that can distinguish or classify them. Therefore, we need to find various models. The important characteristics are characterized by numbers, and this process is called feature extraction.

Feature extraction is very important, it includes two parts: feature extraction and feature selection. Which features are extracted directly affects the effect of classification and the accuracy of resolution. Good features should have significant differences for different types of sword patterns, or they can carry more category information. By classifying these features, you can get a lower false positive rate and a finer resolution level.

(3) Fuzzy recognition judgment

The fuzzy recognition judgment is to establish a fuzzy vector for the feature extracted from the object to be identified, and perform similarity judgment with the fuzzy set established by various standard modes. There are three main problems involved: the use of fuzzy sets to represent standard patterns; the measurement of similarities between fuzzy sets; the principle of fuzzy pattern recognition.

Create a standard set that contains all the patterns: First you need to create a collection X, $X=(x_1,x_2,...,x_m)$, and $x_i(i=1,2,...,m)$ that contain all the patterns. Then, according to the selected classification characteristics, the feature set Y, $Y=(y_1,y_2,...,y_n)$, $y_i(i=1,2,...,n)$ is established to represent each of the classification features. Finally, create a matrix between them through the pattern set and the feature set:

$$R=\begin{bmatrix} r_{11} & r_{12} & \Lambda & r_{1n} \\ r_{21} & r_{22} & \Lambda & r_{2n} \\ \Lambda & \Lambda & \Lambda & \Lambda \\ r_{m1} & r_{m2} & \Lambda & r_{mn} \end{bmatrix}=r_{ij}(m\times n) \tag{2}$$

In the formula, $0\le r_{ij}\le 1$, $1\le i\le m$, $1\le j\le n$, R represents a fuzzy relationship, and r_{ij} represents a fuzzy value that is not quantized. If x_i is larger, it indicates that the probability of y_i corresponding to the 7 feature is more ambiguous.

Metric similarity between fuzzy sets; closeness is an index to measure the closeness of two fuzzy sets. The closer the value of closeness is, the closer and more similar the two fuzzy sets are; any fuzzy set and its own closeness are 1. The common methods for calculating closeness are as follows:

$$\sigma=\frac{\sum_{i=1}^{m}(r_{ij}\wedge r_{ik})}{\sum_{i=1}^{m}(r_{ij}\vee r_{ik})} \tag{3}$$

$$\sigma_{kj}=1-\frac{1}{m}\sum_{i=1}^{m}|r_{ij}-r_{ik}| \tag{4}$$

$$\sigma_{kj}=1-\frac{1}{\sqrt{m}}\left[\sum_{i=1}^{m}(r_{ij}\wedge r_{ik})^2\right]^{1/2} \tag{5}$$

Formulas (3), (4), and (5) are the calculation formulas for maximum and minimum closeness, Hamming closeness, and Euclidean distance closeness, respectively. Where σ_{kj} represents the closeness between the *k* object to be identified and the *j* known pattern sample; r_{ij} represents the fuzzy value of the *i* feature of the *j* known pattern sample: represents the *k* object to be identified The fuzzy values of the *i* feature: $\wedge$ and $\vee$ represent the minimum and maximum values, respectively.

The principle of fuzzy pattern recognition: including the principle of maximum membership, the principle of threshold and the principle of proximity, etc. In the pattern recognition, the appropriate principle should be selected according to the specific requirements.

3. Experiments

3.1. Data Collection

Sampling a black soil area in a province. The soil samples of the cultivated layer (0~20cm) were taken along the road for 5km, and the latitude and longitude of the sample points were recorded by

GPS, a total of 675 samples. The organic matter content was measured by an elemental analyzer. The soil sample was ground, air dried, and passed through a 2 mm sieve. 675 soil samples were arranged in order of organic matter content, and 40 soil samples were randomly selected as the modeling samples at intervals $2g/kg^{-1}$. In the remaining soil samples, 24 soil samples were randomly selected as test samples. The descriptive statistics of the organic matter content of the selected soil samples are shown in Table 1.

Table 1. Descriptive statistics of the organic content of the tested soil samples

Number of samples	Minimum value ($g \cdot kg^{-1}$)	Maximum ($g \cdot kg^{-1}$)	Average value ($g \cdot kg^{-1}$)	Standard error
64	5.6	116.7	50.31	23.64

3.2. Experimental Environment

In this paper, the spectral reflectance characteristics of typical black soil visible/near-infrared bands are studied by indoor measurement of hyperspectral reflectance of soil samples in a wide range of black soil regions. Multivariate statistical analysis methods are used to analyze soil spectral reflectance data and its low-order differential and reciprocal logarithmic differentials. The normalized data was used as the independent variable, and the logarithmic transformation data of black soil organic matter was used as the dependent variable to establish a multivariate linear stepwise regression model for the determination of organic matter content in black soil with high spectral reflectance, and the rapid determination of organic matter content in black soil was realized.

3.3. Experimental Spectrometry

Spectral tests were performed on the selected 64 soil samples. Spectral measurements are taken in a dark room that controls the lighting conditions. The soil samples were placed in a sample dish with a diameter of 12 cm and a depth of 1.8 cm (optically infinitely thick soil sample depth of 1.5 cm). The surface of the soil sample was flattened with a ruler (some studies have been carried out by compacting the soil sample to soil). The purpose of the surface flattening, but the applied pressure is different, will result in different degrees of compaction between different soil samples, affecting the accuracy of the spectral measurements). The light source is a 1000 W halogen lamp, 70 cm from the surface of the soil sample, and a zenith angle of 30. Provides nearly parallel light to the soil sample to reduce the effects of shadows caused by soil roughness. The sensor probe with 80 field of view is placed vertically above 15 cm from the surface of the soil sample. The area of the probe receiving spectrum is a circle with a diameter of 2.1 cm, which is much smaller than the area of the sample dish. The probe receives the reflection spectrum of the soil. The whiteboard is used for calibration before testing. Ten spectral curves were collected for each soil sample, and the actual reflection spectrum data of the soil sample was obtained after arithmetic averaging.

4. Discussion

Field spectral measurement is a complex process that takes into account various factors affecting spectral data, such as temperature, humidity, illumination, light pollution, observation period, atmospheric characteristics, feature characteristics, probe selection, etc. These factors make the spectral curve always exist. Noise, in order to obtain a smooth spectral waveform, the spectral curve needs to be smoothed to remove a small amount of noise contained in the signal, improving the

reliability and accuracy of the model. The soil types and corresponding organic matter content statistics are shown in Table 2.

Table 2. Statistics on soil types and organic content

Soil type	Minimum value	Maximum	Average value
Black soil	0.743%	1.278%	0.930%
Yellow loess	0.113%	0.199%	0.156%
Mian sand	0.343%	1.716%	0.743%
Sand	0.210%	0.829%	0.530%
Paddy soil	0.114%	1.135%	0.618%
Silt	0.174%	0.174%	0.174%

The reflectance curve is obtained by removing the sludge and removing the sample with significant error, as shown in Figure 2.

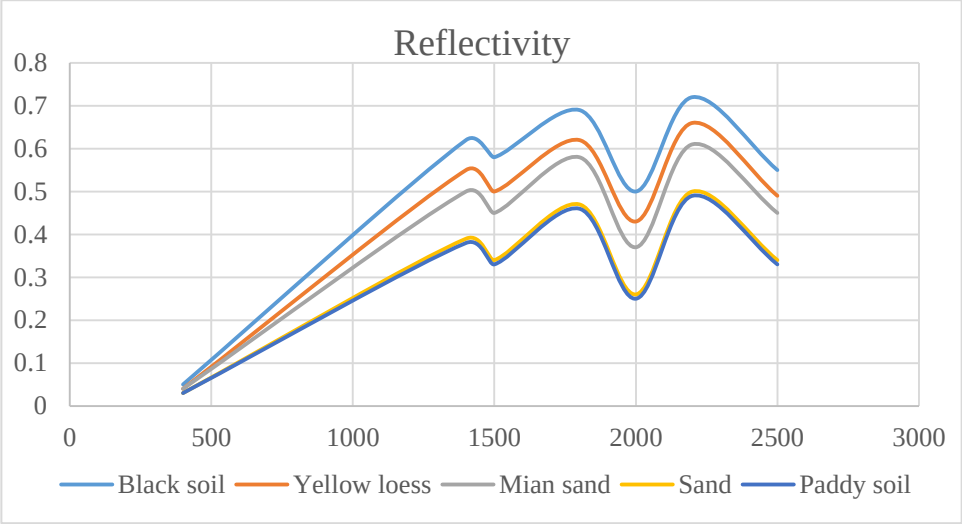

Figure 2. Reflectance curves for each soil type

It can be seen from Figure 2. that the reflectivity of the soil changes with the type of soil, and decreases in the order of mulch, paddy soil, black soil, aeolian sand and sand. Due to the different numbers of these five types of soil samples, there may be some errors. This article does not continue to conduct in-depth research, only a large number of black soils were analyzed.

4.1. Traditional Estimation Model of Soil Organic Matter Content in Black Soil Region

Regression analysis is a method to study the relationship between dependent variables and independent variables. According to the number of independent variables, regression analysis can be divided into one-way regression analysis (only one independent variable) and multiple regression analysis (two or more variables), and each analysis is divided into linear and nonlinear regression. Below, we use the linear regression to establish a hyperspectral estimation model of soil organic matter, and analyze its indicators.

A linear regression equation is an equation that reacts linearly between a dependent variable x and an independent variable y. Its expression is:

$$y_i = a + bx_i \tag{6}$$

The unary linear regression model is based on the linear regression equation. It is assumed that there are two variables x and y, where x is the independent variable and y is the dependent variable.

The expression of the linear regression model is:

$$y_i = a + bx_i + \varepsilon_i \tag{7}$$

If $\hat{a}$, $\hat{b}$ represents the estimated values of the parameters a and b, respectively, the expression of the linear regression model can be obtained according to the least squares method:

$$\hat{y} = \hat{a} + \hat{b}x \tag{8}$$

The linear regression prediction model and accuracy analysis of soil organic matter content established by logarithmic first-order differential transformation using reflectance are shown in Table 3.

Table 3. Linear model

Factor	Model equation	ρ	RMSE	R^2	Average relative error
R_{393}	$\hat{y} = -321.67x + 1.8062$	0.58	0.18	0.46	32.08
R_{443}	$\hat{y} = -507.73x + 2.6726$	0.60	0.19	0.49	33.88
R_{502}	$\hat{y} = -685.5x + 2.5837$	0.62	0.19	0.53	31.50
R_{1452}	$\hat{y} = -3504.2x + 1.403$	0.61	0.20	0.51	34.04
R_{1936}	$\hat{y} = -907.1x + 1.141$	0.58	0.19	0.47	35.35
R_{2186}	$\hat{y} = 1086.6x + 1.6478$	0.57	0.19	0.46	34.99

It can be seen from the above table that the linear estimation model of soil organic matter established at the 502m band has the highest accuracy. The correlation coefficient between the predicted value and the measured value is ρ = 0.62, the coefficient of determination is R^2 = 0.53, and the RMSE of the model is 0.19%. The error is 31.50%.

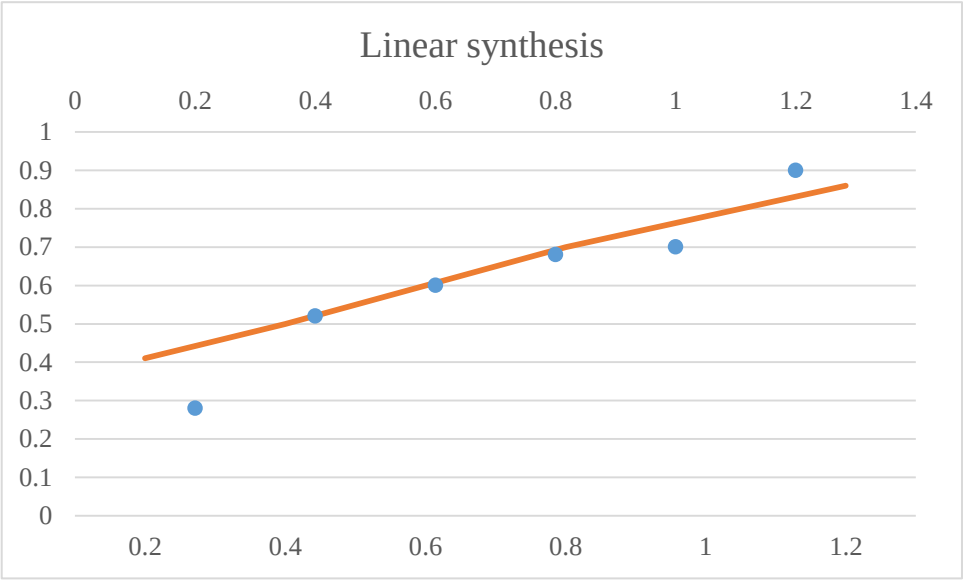

Figure 3. Predicted and actual values of a linear synthesis model

As shown in Figure 3, the correlation coefficient between the predicted value and the measured value of the linear comprehensive model is p=0.7072, the coefficient of determination R2=0.5001, the ruler distance is 0.1175%, and the average relative error is 30.78%. It can be seen that the accuracy of the model is obtained by weighting the correlation coefficient. A certain improvement, but the increase is still small, indicating that the model established by the linear regression method has low prediction accuracy.

4.2. Traditional Estimation Model Based on Fuzzy Identification of Soil Organic Matter Content in Black Soil Region

(1) Hyperspectral estimation of soil organic matter content in black soil area based on closeness

Closeness is a typical fuzzy pattern recognition method. It is an index to measure the closeness of two fuzzy sets. The larger the value of closeness, the closer and more similar the two fuzzy sets are; any fuzzy set and its own the closeness is 1. The weighted maximum and minimum method is calculated as:

$$\sigma_{kj} = \frac{\sum_{i=1}^{m} w_i \left(r_{ij} \wedge r_{ik} \right)}{\sum_{i=1}^{m} w_i \left(r_{ij} \vee r_{ik} \right)} \tag{9}$$

Table 4. Maximum and minimum method closeness prediction results

Sample number	Maximum closeness	Corresponding mode number	Predictive value(%)	Measured value(%)	Relative error(%)
65	0.902	56	0.781	0.857	-8.64
66	0.921	23	0.455	0.457	-0.42
67	0.920	48	0.719	0.686	3.29
68	0.889	64	0.849	0.958	-11.25
69	0.948	18	0.410	0.516	-19.09
70	0.550	4	0.143	0.114	22.19

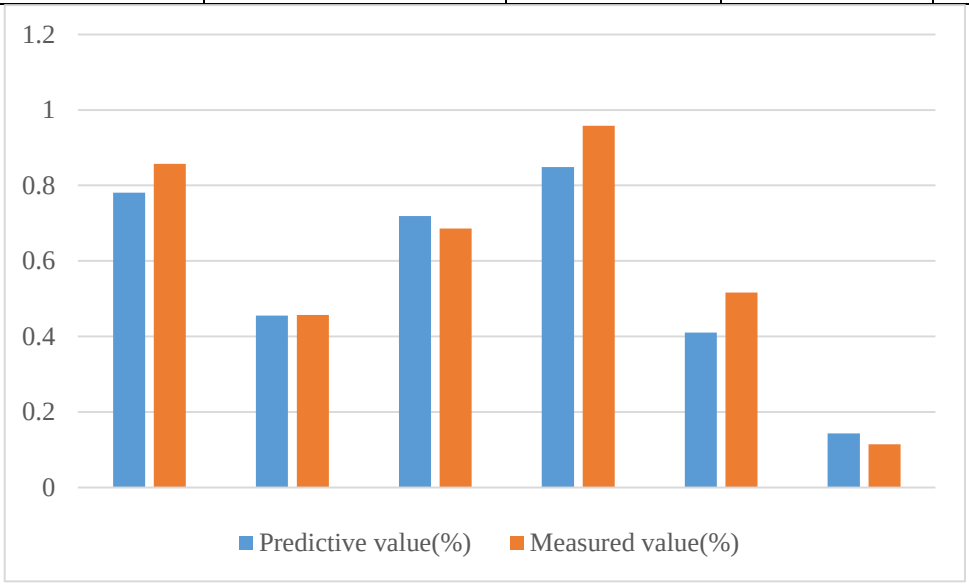

Figure 4. Predicted and measured values

As shown in Table 4 and Figure 4, the maximum and minimum method of close proximity model

has higher estimation accuracy. The average estimated error of organic matter content of the six test samples is 10.290%, and the minimum relative error is 0.427%. The maximum relative error is 23.200%,. This indicates that it is feasible to use the fuzzy closeness to estimate the soil organic matter hyperspectral.

(2) Fuzzy identification of soil organic matter content based on normal fuzzy set

If a membership function of *n* types on the identified entire object *U* is known. For the variable *x*, its membership is:

$$A(x)=e^{\left[-\left(\frac{x-a}{b}\right)^2\right]} \tag{10}$$

Where *a* is the mean and $b^2=2\sigma^2$, σ^2 is the corresponding variance. Approximate value based on Taylor series expansion

$$A(x)=\begin{cases}1-\left(\dfrac{x-a}{b}\right)^2\\0\end{cases} \tag{11}$$

If there are *n* types of *m* indicators, the membership function of the *i* type on the *j* indicator is

$$A_{ij}(x)=\begin{cases}0 & x\le a_{ij}^{(1)}-b_{ij}\\1-\left(\dfrac{x-a_{ij}^{(1)}}{b_{ij}}\right)^2 & a_{ij}^{(1)}-b_{ij}<x<a_{ij}^{(1)}\\1 & a_{ij}^{(1)}<a_{ij}^{(2)}\\1-\left(\dfrac{x-a_{ij}^{(2)}}{b_{ij}}\right)^2 & a_{ij}^{(2)}<x<a_{ij}^{(2)}+b_{ij}\\0 & a_{ij}^{(2)}+b_{ij}<x\end{cases} \tag{12}$$

Where, $a_{ij}^{(1)}$ and $a_{ij}^{(2)}$ are the minimum and maximum values of the *j* index of the *i* element, respectively, and $b_{ij}^2=2\sigma_{ij}^2$, σ_{ij} is the standard deviation of the *j* index of the *i* element. The calculation method is as follows: firstly, calculate the membership degree of each indicator of the normal fuzzy set of each identified sample and the known category, and then take the minimum value among the membership degrees among the indicators, and finally take the maximum value among the membership degrees among the categories. The data is then input into the DPs data processing system for fuzzy pattern recognition, and the recognition results are shown in Table 5.

Table 5. Identification results

Sample number to be identified	Maximum value for various types of closeness	Measured value	Closest to the category number
65	0.8546	0.857	8
66	0.9886	0.457	3
67	1	0.686	5
68	0.6204	0.848	8
69	0.9785	0.416	3
70	1	0.125	1

The maximum value and measured value of various types of closeness are shown in Figure 5.

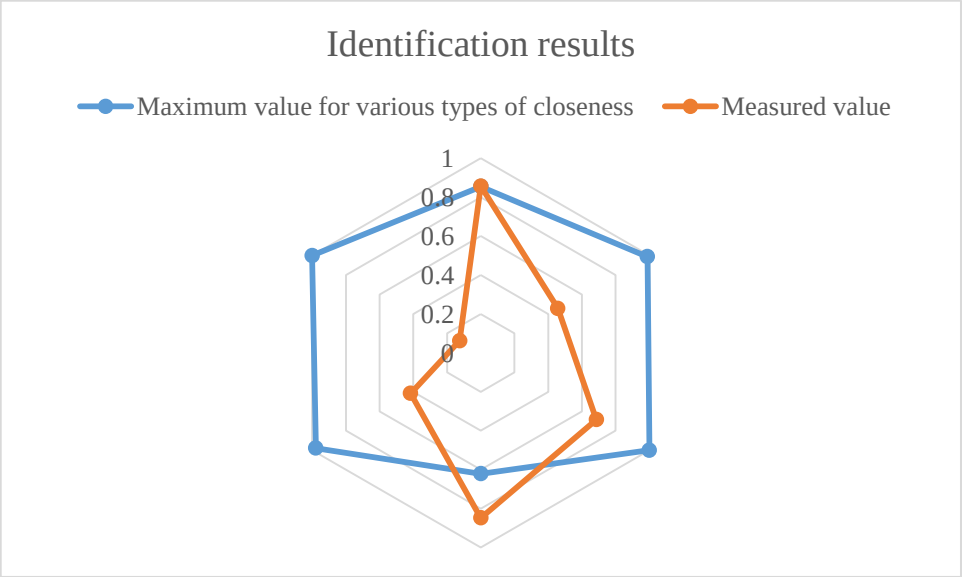

Figure 5. Maximum and measured values for various types of closeness

It can be seen from the results in the above table that the closest sample of the five samples is correct, one has errors, and the sample No. 69 should be classified into the fourth category, but the third category is identified, and the range of the third category is 0.44. % is 0.5l%, and the measured value of sample 69 is 0.527. There is a certain error between the two, but it is not large, and the error is 3.22%. Since the organic matter content is constantly changing and uncertain, there will be certain errors in the prediction with certain point values. The prediction results are not very accurate, and the organic matter can be classified by using the maximum membership degree for fuzzy pattern recognition after classifying the organic matter content. In the corresponding category, this method can make full use of the inherent information of the data, and the operation is simpler and more convenient.

5. Conclusion

In this paper, 64 soil samples from a black soil region of a province were studied. The outdoor reflectance spectra and organic matter content data were obtained for the collected samples. Under the pretreatment conditions of reflectance spectral data, different methods were used to analyze the spectra. The data were mathematically transformed, and the spectral reflectance characteristics of soil organic matter content were analyzed. The characteristic factors were selected according to the single correlation analysis method. Finally, the fuzzy identification prediction model of soil organic matter content was established, and other commonly used models were established to compare and illustrate the effectiveness of the proposed model.

Through comparison of examples, we can find that the average relative error modeled by the fuzzy closeness method is only one higher than 10%, reaching 10.29%, and the others are less than 10%. The interval inversion is used for category identification, and 5 are accurate. Identification, 1 error, the error of the sample deviation is 3.22%, the error is also small, which shows that the identification of organic matter content by interval inversion is very effective, compared with the point value prediction will reflect the organic matter content more effectively. Dynamic changes. In summary, the fuzzy recognition method effectively solves the problems of nonlinearity and ambiguity, and it is feasible to use the fuzzy recognition method for soil organic matter content prediction. In order to solve the randomness and ambiguity of experimental data, a fuzzy closeness

inversion model of soil organic matter content was established. According to the diversity of spectral inversion factors and the dynamics of organic matter content, the interval value fuzzy inversion model was further proposed. By comparing with traditional modeling methods, the fuzzy recognition model has certain potential for high-spectral inversion of soil organic matter content.

In the soil spectral classification, due to the influence of the soil parent material in different regions, the soil spectral characteristics of the same soils will also be different, so it is also necessary to consider the soil or even the soil level. This paper mainly attempts to classify the characteristics of soil spectral curves by extracting them, and answers the question of how to solve the influence of different soil type reflectance spectra on soil organic matter prediction models. However, in order to obtain higher classification accuracy, it is necessary to consider auxiliary materials such as soil parent material, geographical location, altitude, etc., and the current research area is only soil in the black soil area. Therefore, the subsequent work needs to first determine the physical and chemical parameters of different soil types. Taking the above-mentioned influencing factors into consideration, and expanding the sample type and the scope of the study area, and improving the accuracy of spectral classification, provides theoretical support for large-scale soil rapid classification.

Funding

This article is not supported by any foundation.

Data Availability

Data sharing is not applicable to this article as no new data were created or analysed in this study.

Conflict of Interest

The author states that this article has no conflict of interest.

References

[1] Fu C B, Xiong H G, Tian A H. Study on the Effect of Fractional Derivative on the Hyperspectral Data of Soil Organic Matter Content in Arid Region. Journal of Spectroscopy, 2019, 2019(2):1-11. https://doi.org/10.1155/2019/7159317

[2] Pascucci S, Casa R, Belviso C, et al. Estimation of soil organic carbon from airborne hyperspectral thermal infrared data: a case study. European Journal of Soil Science, 2015, 65(6):865-875. https://doi.org/10.1111/ejss.12203

[3] Hong-Qi W U, Fan Y M, Jing H E, et al. Response of Soil Hyperspectral Characteristics of Different Particle Sizes to Soil Organic Matter. Acta Agrestia Sinica, 2014, 22(2):266-270.

[4] Haifeng W, Zhitao Z, Karnieli A, et al. Hyperspectral estimation of desert soil organic matter content based on gray correlation-ridge regression model. Transactions of the Chinese Society of Agricultural Engineering, 2018, 34(14):124-131.

[5] Lei Y U, Hong Y S, Zhou Y, et al. Inversion of Soil Organic Matter Content Using Hyperspectral Data Based on Continuous Wavelet Transformation. Spectroscopy & Spectral Analysis, 2016, 36(5):1428-1433. https://doi.org/10.1016/j.jag.2016.01.005

[6] Vaudour E, Gilliot J M, Bel L, et al. Regional prediction of soil organic carbon content over temperate croplands using visible near-infrared airborne hyperspectral imagery and synchronous field spectra. International Journal of Applied Earth Observation &

Geoinformation, 2016, 49(July 2016):24-38.

[7] Hou Y, Tiyip T, Sawut M, et al. Estimation model of desert soil organic matter content using hyperspectral data. Nongye Gongcheng Xuebao/transactions of the Chinese Society of Agricultural Engineering, 2014, 30(16):113-120.

[8] Wu M Z, Li X M, Sha J M. Spectral Inversion Models for Prediction of Total Chromium Content in Subtropical Soil. Spectroscopy & Spectral Analysis, 2014, 34(6):1660-1666.

[9] Wang J, Tashpolat Tiyip, Ding J, et al. Estimation of desert soil organic carbon content based on hyperspectral data preprocessing with fractional differential. Nongye Gongcheng Xuebao/transactions of the Chinese Society of Agricultural Engineering, 2016, 32(21):161-169.

[10] Xiao Z Z, Yi L I, Feng H. Hyperspectral Models and Forcasting of Physico-Chemical Properties for Salinized Soils in Northwest China. Spectroscopy & Spectral Analysis, 2016, 36(5):1615-1622.

[11] Prakash N, Singh Y. Fuzzy Support Vector Machines for Face Recognition: A Review. Maropoulos P G, 2015, 131(3):24-26. https://doi.org/10.5120/ijca2015907224

[12] Fajardo M, Mcbratney A, Whelan B. Fuzzy clustering of Vis–NIR spectra for the objective recognition of soil morphological horizons in soil profiles. Geoderma, 2016, 263:244-253. https://doi.org/10.1016/j.geoderma.2015.05.010

[13] Liang Z, Liu X, Tan S, et al. Video tracking for high-similarity drug tablets based on reflective lightness intensity and fuzzy recognition system. Proceedings of the Institution of Mechanical Engineers Part H Journal of Engineering in Medicine, 2016, 230(3):211-229. https://doi.org/10.1177/0954411916629174

[14] Gorbachev S, Syryamkin V. Adaptive Neuro-Fuzzy Recognition Technology Intersecting Objects. Applied Mechanics & Materials, 2015, 756:683-688.

[15] Guo W, Liu Y, Xu X, et al. Application of variable fuzzy pattern recognition method on landslide stability assessment. Transactions of the Chinese Society of Agricultural Engineering, 2015, 31(8):176-182.

[16] Si H, Yao Y, Wang D, et al. Hyperspectral prediction of soil organic matter contents under different soil moisture contents. Transactions of the Chinese Society of Agricultural Engineering, 2015, 31(9):114-120.

[17] Qiao X X, Wang C, Feng M C, et al. Hyperspectral estimation of soil organic matter based on different spectral preprocessing techniques. Spectroscopy Letters, 2017, 50(3):156-163. https://doi.org/10.1080/00387010.2017.1297958

[18] Yu L, Hong Y, Geng L, et al. Hyperspectral estimation of soil organic matter content based on partial least squares regression. Transactions of the Chinese Society of Agricultural Engineering, 2015, 31(14):103-109.

[19] Wang X, Meng J. Mapping soil organic matter content in field using HJ-1 satellite image. Transactions of the Chinese Society of Agricultural Engineering, 2014, 30(8):101-108.

[20] Liang Z, Liu X, Tan S, et al. Video tracking for high-similarity drug tablets based on reflective lightness intensity and fuzzy recognition system. Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine, 2016, 230(3):211-229. https://doi.org/10.1177/0954411916629174