

Inversion of Soil Properties with Hyperspectral Reflectance in Well-facilitied Capital Farmland Construction Areas

qiuxia zhang¹, Wenkai Liu¹, Hebing Zhang², Shou-Chen Ma², and Xinsheng Wang³

¹North China University of Water Resources and Electric Power

²Henan Polytechnic University

³Henan Institute of Science and Technology

May 19, 2022

Abstract

Well-facilitied capital farmland construction is an important measure to enhance the ability to ensure food security and accelerate new-style the modernization agriculture. Hyperspectral remote sensing can be provide data basis and technical support for realization the construction of well-facilitied capital farmland,to provide a reference for exploring the optimization of well-facilitied capital farmland construction area. Taking Xinzheng City of main grain producing areas in Henan province as the research object, using field sampling and indoor hyperspectral spectroscopy (350~2500 nm) combined, the spectral transformations such as Continuum removal(CR) are carried out after the Savitzky-Golay(SG) convolution smoothing, the best hyperspectral bands as the common index of the soil properties were selected by the correlation analysis and Fuzzy clustering maximum tree,focused on 405~431nm、781nm~831nm、1044~1087nm、1251~1410nm、1836~1898nm、2080nm~2201nm、2324~2395nm. The hyperspectral inversion model had been built by Panel date model of fixed effect variable coefficient based on the ordinary least squares estimation method (OLS), that is about the panel data of PH、SOM、AN、AP、AK、Fe、Cr、Cd、Zn、Cu、Pb of 116 samples in Xinzheng City. The results showed that: The Panel date model significantly overall, the goodness of fit is higher (in the model, $R^2=0.9991$, $F=2195.67$). The result of precision test indicates that models performed well in modeling and predicting with a good ability of quantificational prediction, with RPD values were greater than 2.5.

Inversion of Soil Properties with Hyperspectral Reflectance in Well-facilitied Capital Farmland Construction Areas

Qiuxia Zhang ¹, Wenkai Liu ¹, Hebing Zhang ², Shouchen Ma ², Xinsheng Wang ³

1. School of Public Administration, North China University of Water Resources and Electric Power, Zhengzhou, China 450046

2. School of Surveying and Landing Information Engineering, Henan Polytechnic University, Jiaozuo, China 450003

3. School of Life Science & Technology, Henan Institute of Science and Technology, Xinxiang, China 453003

Author: Qiuxia Zhang, Phd, Lecturer, the research direction is mining land reclamation and ecological reconstruction, E-mail: zhangqiuxia@ncwu.edu.cn .

Abstract: Well-facilitied capital farmland construction is an important measure to enhance the ability to ensure food security and accelerate new-style the modernization agriculture. Hyperspectral remote sensing can be provide data basis and technical support for realization the construction of well-facilitied capital farmland,to provide a reference for exploring the optimization of well-facilitied capital farmland construction area. Taking Xinzheng City of main grain producing areas in Henan

province as the research object, using field sampling and indoor hyperspectral spectroscopy (350~2500nm) combined, the spectral transformations such as Continuum removal(CR) are carried out after the Savitzky-Golay(SG) convolution smoothing, the best hyperspectral bands as the common index of the soil properties were selected by the correlation analysis and Fuzzy clustering maximum tree, focused on 405~431nm、781nm~831nm、1044~1087nm、1251~1410nm、1836~1898nm、2080nm~2201nm、2324~2395nm. The hyperspectral inversion model had been built by Panel data model of fixed effect variable coefficient based on the ordinary least squares estimation method (OLS), that is about the panel data of PH、SOM、AN、AP、AK、Fe、Cr、Cd、Zn、Cu、Pb of 116 samples in Xinzheng City. The results showed that: The Panel data model significantly overall, the goodness of fit is higher (in the model, $R^2 = 0.9991, F = 2195.67$). The result of precision test indicates that models performed well in modeling and predicting with a good ability of quantificational prediction, with RPD values were greater than 2.5.

Keywords: Well-facilitated capital farmland; Hyperspectral; Panel data model; Maximum tree clustering method; Xinzheng City

Introduction

At present, the core area of grain production has become a national strategy. Henan Province, as the most important part of it, that is in the critical period of promoting the "coordination of the four modernizations and scientific development". Agriculture is the basis for realizing modernization, so must to speed up the transformation of agricultural development mode, and improve agricultural quality, efficiency and competitiveness. Path to agricultural modernization featuring, of high output efficiency, product safety, resource conservation and environmental friendliness. However, the important bottleneck that restricts agriculture sustainable development of Henan, what is the large number of people and small amount of land, the serious shortage of reserve resources of cultivated land, and the overall low quality of cultivated land. Therefore, strengthening Well-facilitated capital farmland construction has enormous symbolic significance, for implementing the national strategy of China and promoting agricultural production.

Well-facilitated capital farmland refers to the basic farmland with centralized contiguity, supporting facilities, high and stable yield, good ecology and strong disaster resistance formed, through rural land consolidation and construction in a certain period of time, which is compatible with modern agricultural production and operation mode(TD/T1033-2012). At present, most of the researches on Well-facilitated capital farmland construction include Well-facilitated capital farmland construction demarcation(LI Yilong, et al., 2019; DONG Fei, et al., 2020; ZHANG Hebing, et al., 2018) potential evaluation(CAI Xiangwen. et al., 2019; LI C M, et al., 2018), suitability evaluation(TANG Feng, et al., 2019; CHEN Lin, et al., 2019; ZHANG Jing, et al., 2020), construction sequence and mode zoning(LI Long, et al., 2020; WANC Ke, et al., 2021; ZENG Ya, et al., 2020), project implementation and effect evaluation(MA Xueying, et al., 2018; XIONG Yufei, et al., 2019; WANG Xiaoqing, et al., 2018), etc. In the Standard of Well-facilitated Capital Farmland Construction(TD/T1033-2012), it is clearly proposed to "the quality of cultivated land after completion reaches the higher level of the county", etc. Therefore, the construction of Well-facilitated capital farmland requires, that the soil quality should reach the higher level of the region, while the construction of field projects. "it should be determined the amount of fertilizer, according to soil nutrient status, and the soil nitrogen, phosphorus, potassium, medium and trace elements, organic matter content, soil acidification, salinity, and should be regularly monitored other conditions, and should be constantly adjusted the fertilization formula, according to the actual situation"(NY/T 2148-2012). However, in the process of Well-facilitated capital farmland construction, in addition to the more placed on land leveling, improvement of supporting facilities of roads, ditches and other projects, while strengthen the rapid acquisition, and real-time monitoring of soil basic information. However, most research methods are quantitative inversion research for single soil properties(BIAN Zijin, et al., 2021; WEI Lifei, et al., 2020; LI zhiyuan, et al., 2021; YU Huan, et al., 2021; Schreiner Simon, et al., 2021; SHI Yuanyuan, et al., 2021; XU Xitong, et al., 2020). Inversion models of soil properties and spectral reflectance need to be built one by one, and the calculation process is complicated and time-consuming. And Panel data model can be build the three-dimensional data model at the same time, contain various soil properties of multiple points, and high spectral characteristics of the band values, By inversion

modeling get multiple soil properties at once, modeling calculation process more simple, and can according to the relationship between each model analysis of soil properties, and the influence of high spectral band characteristic values on each soil properties. Therefore, It is necessary to study the rapid and nondestructive testing of soil attribute information in Well-facilitied capital farmland construction area, so as to provide technical support for the rapid acquisition, and real-time monitoring about soil attribute, and to provide support for the optimization of well-facilitied capital farmland construction area.

This reseach taking Well-facilitied capital farmland construction area of Xinzheng City as the research object, obtained soil hyperspectral data by using ASD Field Spec3 ground object spectrometer in laboratory experiments, and combined with soil properties such as soil PH, organic matter, nitrogen, phosphorus, potassium, Fe, Cr, Cd, Cu, Zn, Pb, etc. It performed Savitzky-golay(*SG*) filtering and Continuum removal (*CR*) spectral transformation on the original spectral reflectance. And used Correlation analysis and Fuzzy clustering maximum tree method to select the common significance band of different soil attributes as the best hyperspectral characteristic band. This paper attempts to establish a comprehensive hyperspectral inversion model of cultivated land soil attributes by using Panel date model, estimate the influence of hyperspectral characteristic band values on each soil attribute, and predict the content of each soil attribute, aiming to provide theoretical and technical support for the rapid acquisition and real-time monitoring about soil attribute of Well-facilitied capital farmland construction area.

1. Materials and Methods

1.1. The overview of the researched area

Xinzheng city is located in the central part of Henan Province in China, is transition zone from the north China plain, western Henan mountain to eastern Henan plain, is the core of the Central Plains economic zone, under zhengzhou city, located in $34^{\circ}16' \sim 34^{\circ}39' N$, $113^{\circ}30' \sim 113^{\circ}54' E$, north of the provincial capital Zhengzhou, east of Zhongmu County, Weishi County, south of Changege City, Yuzhou city. It borders Xinmi city on the west, North of Zhengzhou city 38 *km*; Northeast from zhongmu county 45.6 *kilometers*, 120*kilometers* downtown Kaifeng; East to Yushi county 42.6 *kilometers*; South to Changege city 20.4 *km*, Xuchang city 40 *km*; Southwest to Yuzhou city 36.5 *km*, Pingdingshan city 84 *km*; 34.5 *kilometers* west to Xinmi urban area. It is 42*kilometers* long from north to south and 36 *kilometers* wide from east to west. It covers an area of 873 *square kilometers* and has a total population of 653,000. In 2019, It has jurisdiction over towns of 9, townships of 1, streets of 3 and administrative villages of 253, natural villages of 921 and residential areas of 24.

According to the survey of land use status in 2013, the total land area of Xinzheng city is 884.5915 km^2 , and the cultivated land is 521.7641 km^2 , accounting for 58.59% of the total land area. The total annual grain output is 273148 *t*. According to Integrated Land-use Planning of Xinzheng city (2010-2020), the protection index of prime farmland in Xinzheng City is 427.73 km^2 . Xinzheng city is warm temperate continental monsoon climate, moderate temperature, four distinct seasons; The main disastrous weather is drought, flood, wind, hail, etc. The average annual temperature is 14.2, the average annual precipitation is 676.1 *mm*, the average annual evaporation is 1476.2 *mm*, the average annual sunshine duration is 2,114.2 *h*, the average annual frost day is 67 days, the average annual total water resources is 147.73 million m^3 , and the per capita water resources are 236 m^3 . There are various soil types, mainly cinnamon soil, tidal soil and aeolian sand soil. The terrain is high in the west and low in the east, with shallow hills in the west, plains in the east and hills in the northwest.

1.2. Field collection of soil samples

According to the soil type, topographic characteristics and spatial variation characteristics of the study area, and taking into account the integrity of administrative units (towns or villages as units), sampling points were laid out using the $2 \text{ km} \times 2 \text{ km}$ regular grid method. Every point in a spatial database included the basic information, as its serial number, latitude and longitude coordinates, township and neighboring villages, etc. According to the map about sample point and the table of point attribute, and used GPS to accurately locate the field sampling, and the sampling depth was 0-30 *cm* on the surface of the soil, and recorded the

coordinates of the actual sampling points and detailed characteristic information of the sample site. And collected a total of 154 soil samples in this sampling, and removed the invasive body such as plant roots and stem residues and brick and tile fragments. After natural air drying, grinding and passing through a 1 mm sieve, and divided the samples into four parts by quartering method in duplicate, one was used for determination of physical and chemical properties in laboratory, the other was used for determination of soil spectrum. The main soil properties measured in this study were soil PH, SOM, AN, AP AK, Fe, Cr, Cd, Zn, Cu and Pb. And carried out the determination method according to Regional Geochemical Sample Analysis Method(DZ/T 0279-2016). In order to ensure the quality of analysis, national geochemical standard samples were used for quality control.

1. 3. Laboratory test of samples spectra

This research measured soil spectral reflectance by ASD spectrometer on treated soil samples under indoor conditions. The spectrum measurement instrument is an ASD Field Spec 3 spectrometer produced by ASD, USA. The spectral range is 350-2500nm , with sampling interval of 1.4nm for 350-1000nm , sampling interval of 2nm for 1000-2500nm , and resampling interval of 1nm . Before spectral measurement, the surface of the soil should be scraped in the same direction, along the edge of the soil sample vessel with a ruler, and then filled with soil sample dish are placed black rubber MATS of reflectivity approximately 0, halogenated lamp with power of 50W is used as light source, probe the view Angle of 25o, light incidence Angle of 45o, the distance of light source is 15cm , and the distance of probe is 15cm . To reduce the influence of anisotropy soil sample spectra, when measuring turn the sample plate 3 times , each time the rotation angle of about 90o, and obtained the soil sample spectra about four directions, reference plate calibration is performed before and after each target spectrum acquisition, repeated measurement 5 times, a total of 20 times, and used View Spec Pro software to obtain the average value of spectral reflectance as the original reflectance spectral value . Because near the two ends of the band test range (350nm and 2500nm) are unstable regions of the spectral data, removed the data of 350-399nm and 2401-2500nm ,which are greatly affected by external noise.

1.4. Model establishment and accuracy test

1.4.1. Fuzzy clustering maximum tree method

Fuzzy theory is developed on the mathematical basis of Fuzzy set theory established by American cybernetics expert Professor L.A.Zadeh in 1965, which has been widely used in mathematics and many other fields(LIU Qi, et al., 2004). Fuzzy Clustering Number is a multi-technology, which classifies objective things by using fuzzy mathematics method, establishing similarity relation according to characteristics, similarity and affinity degree of objective(LI Hongxing, et al., 1994; WANG Peizhuang,1983). Because the classification of reality is often accompanied by fuzziness, fuzzy clustering theory is more consistent with objective reality.

The basic steps of Fuzzy clustering analysis using the maximum tree method are as follows:

(1) Establish sample set matrix

Suppose that the sample set, n represents the number of samples, each sample has an m dimensional vector representation, that is, each sample has m indicators, that is

(2) Establish fuzzy similarity matrix

According to the given sample characteristic data, the correlation coefficient method is used to establish the fuzzy similarity matrix, r_{ij} is the similarity coefficient between different samples, namely

(3) Maximum tree generation

With a certain point x_i in a relatively concentrated set of classified objects, as its vertex and r_{ij} in the fuzzy similarity matrix R as its weight, it is arranged in descending order, requiring no loop (i.e. circle), until all vertices are connected, forming a special graph, namely the largest tree (the largest tree may not be unique).

(4) Clustering

Select the appropriate threshold λ , cut off the branches of the weight, get an unconnected graph, each connected branches constitute the classification of horizontal λ , there are several branches indicating the classification of several categories.

In this paper calculate the fuzzy similarity coefficient between the correlation coefficient curves of soil attributes and spectral indexes by systematic clustering method, and constructed a fuzzy similarity matrix to determine the similarity of the correlation coefficient, that between different soil attributes and spectral indexes. On this basis, determined the common hyperspectral inversion bands of different soil attributes by the maximum tree classification method.

1.4.2. Panel data model

Panel data is also called parallel data, or time series and cross section data or pool data. It refers to taking multiple cross sections on time series, sample data formed by sample observation values are simultaneously selected on these cross sections(SUN jingshui,2010). From the cross section, it is a cross section observation value formed by several individuals at a certain moment, and from the longitudinal section, it is a time series. According to the characteristics of panel data, the hyperspectral characteristic band values of soil properties of multiple samples can be regarded as the hyperspectral characteristic band values of soil properties at a sample point on the cross section, and a sequence of sample points on the vertical section. Through the construction of Panel data model, a comprehensive inversion model of soil properties can be established at the same time, without the need for individual inversion of each index, which reduces the tedious process of multi-index inversion(ZHANG Qiuxia, et al., 2017).

Due to the large number of sample points T and the small number of cross section N , it was determined as Fixed influence model, and Ordinary Least Squares Estimation (OLS) was selected to build the Panel data model. Then, panel data model types are determined by Analysis of Covariance, namely invariant coefficient model, variable intercept model and variable coefficient model. In order to reduce the impact of heteroscedasticity, the natural logarithm of variables was calculated on both sides of the panel data model equation, and the panel data model was obtained as:

Where:

- values of explained variables on cross section i and sample t , namely soil heavy metal element content
 - Constant term or intercept term, representing the cross section of i (influence of the individual of i)
 - Model parameter of the j explanatory variable on the i cross section
 - The value of the j explanatory variable on cross section i and sample t , namely, the reflectance of hyperspectral characteristic band of soil heavy metals
 - Random error term on cross section i and sample t
- k – Number of explanatory variables

1.4.3. Accuracy test method of inversion model

The calibration set determination coefficient R^2 and Root Mean Square Error ($RMSEC$) are used to verify the modeling accuracy. Validation set test is based on validation set determination coefficient v^2 , Root Mean Square Error ($RMSEP$) and Relative Percent Deviation (RPD), the Relative Percent Deviation is the ratio about between the standard deviation and $RMSEP$ of validation set. When $RPD > 2.5$, model has excellent predictive ability. When $2.0 < RPD \leq 2.5$, the model has good quantitative prediction ability. When $1.8 < RPD \leq 2.0$, the model has quantitative prediction ability. When $1.40 < RPD \leq 1.80$, the model has general quantitative prediction ability. When $1.00 < RPD \leq 1.40$, the model has the ability to distinguish the high value from the low value. When $RPD \leq 1.00$, the model has no predictive ability (Rossel RAV, et al., 2007). For the modeling set, the larger R^2 is, the smaller $RMSEC$ is, the higher the modeling accuracy is, and the more stable the model is. For the verification set, the larger v^2 and RPD are, the smaller $RMSEP$ is, the higher the prediction accuracy is.

2. Results and discussion

2.1. Spectral pretreatment

In the process of ASD spectrometer acquisition, acquisition and transmission of spectral signals, in addition to the spectral information of soil itself, spectrometer breeding and interference of external factors, there may be many "burr" noises in spectral curves, and the signal-to-noise ratio is reduced. In order to obtain the stable spectrum and improve the signal-to-noise ratio, it is necessary to smooth the spectral data. Savitzky-golay (*SG*) convolution smoothing method was proposed by Savitzky and Golay (Savitzky A, et al., 1964) in 1964. It is a weighted average method, that obtains smooth point data by least square fitting of the data, that to be measured in the moving window interval using polynomial method. It is a widely used smoothing method at present. In the process of *SG* filtering, need to be selected appropriate smoothing points and polynomial fitting times. The more smoothing points are taken, the smoother the spectral curve will be, but some information will be lost at the same time. Therefore, *SG* filtering smoothing based on 9-point quadratic polynomial is adopted. The transform tool used for smoothing and denoising by Unscrambler 9.7, as shown in Figure 1.

In order to better highlight the smoothing effect, the band curves of 2000-2400 *nm* were amplified (Figure.1b). By comparing the details before and after *SG* smoothing, it can be seen that *SG* smoothing can effectively remove noise, and better preserve the overall characteristics of spectral curves.

2.2. Continuum removal

In order to find the sensitive relationship between soil heavy metal content and spectral reflectance, Continuum removal (*CR*) spectral transformation after *SG* smoothing is required. To envelope as a spectral analysis method is put forward first by Clark and Rous in 1984 (Clark R N, et al., 1984), is defined as a point in a straight line connected with the wavelength change reflect or absorb protruding point of "peak value", and make the line in the "peak value" on the outside, is greater than 180° (TONG Qingxi, et al., 2006), the actual spectrum reflectance and envelope line of the corresponding band reflectance ratio, By normalizing the spectral value to 0~1 (LI Shumin, et al., 2011), the absorption and reflection characteristics of the spectral curve can be effectively highlighted and the characteristic bands can be extracted. Through proper spectral transformation, the influence of various noises can be reduced or even eliminated, the spectral sensitivity can be improved, and the prediction ability and stability of the calibration model can be improved. in the study obtained the de-envelope by constructing pop database in Envi4.8, as shown in Figure 2.

The reflectance curve of *CR* not only enhances the spectral characteristics of the original spectral curve at 1400*nm*, 1900*nm* and 2200*nm*, but also highlights the weak absorption characteristics at 410*nm*, 500*nm* and 700*nm*. It shows that the weak absorption characteristic information of the original spectral curve is enhanced, and the signal-to-noise ratio is improved by de-enveloping spectral transformation, which is helpful for the extraction of effective characteristic bands.

2.3. Selection of common spectral characteristic bands for soil properties

On the basis of soil properties significant band selection about Xinzheng well-facilitied capital farmland construction area, considering the needs of different soil property spectrum inversion, combined with the correlation coefficient curve similarity and inflection point, using the method of Fuzzy clustering tree, determine the share best band of hyperspectral inversion of soil properties about Xinzheng well-facilitied capital farmland construction area.

Through comparative analysis of the correlation coefficient curves of 11 soil attributes and *SG-CR* transformations of Xinzheng City well-facilitied capital farmland area (Figure 3), it can be seen that the correlation coefficient curves of the same spectral transformation have similar inflection points, showing good similarity.

The soil properties corresponding to the row numbers of the fuzzy similarity matrix were PH, SOM, AN, AK, AP, Fe, Cr, Cd, Zn, Cu and Pb. The fuzzy similarity matrix of correlation coefficient curves of 11 soil attributes and *SG-CR* spectral transformation is:

Using the maximum tree classification method, $\lambda = 0.76$, PH, AP, Cr, Cd, Pb as a class, SOM, AN as a class, AK, Fe, Cu as a class, Zn as a class. According to the similarity and inflection point of correlation coefficient curves, the band of *SG-CR* spectral transformation is selected $405nm \sim 418nm \sim 781nm \sim 784nm \sim 794nm \sim 805nm \sim 807nm \sim 830nm \sim 831nm \sim 1079nm \sim 1085nm \sim 1251nm \sim 1267nm \sim 1308nm \sim 1309nm \sim 1410nm \sim 1836nm \sim 1860nm \sim 1897nm \sim 1898nm \sim 2080nm \sim 2137nm \sim 2149nm \sim 2156nm \sim 2184nm \sim 2382nm \sim 2395nm$. At the same time, according to the significant bands of soil attributes and *SG-CR* spectrum, the significant bands of 11 soil attributes passing the significance level test of $P = 0.01$ were selected, *i.e.* $406 \sim 419nm \sim 421 \sim 423nm \sim 427 \sim 431nm \sim 1044nm \sim 1062nm \sim 1087nm \sim 1887 \sim 1890nm \sim 2118nm \sim 2119nm \sim 2185 \sim 2187nm \sim 2198 \sim 2201nm \sim 2324nm \sim 2325nm$.

The significant band and the inflection point of the correlation coefficient curve were combined to determine the band $405 \sim 419nm \sim 421 \sim 423nm \sim 427 \sim 431nm \sim 781nm \sim 784nm \sim 794nm \sim 805nm \sim 807nm \sim 830nm \sim 831nm \sim 1044nm \sim 1062nm \sim 1079nm \sim 1085nm \sim 1087nm \sim 1251nm \sim 1267nm \sim 1308nm \sim 1309nm \sim 1410nm \sim 1836nm \sim 1860nm \sim 1887 \sim 1890nm \sim 1897nm \sim 1898nm \sim 2080nm \sim 2118nm \sim 2119nm \sim 2137nm \sim 2149nm \sim 2156nm \sim 2184 \sim 2187nm \sim 2198 \sim 2201nm \sim 2324nm \sim 2325nm \sim 2382nm \sim 2395nm$ as a common spectral characteristic band of soil attribute spectral inversion in well-facilitated capital farmland construction area of Xinzheng City.

2.4. Construction of panel data model

On the basis of soil types, to divide the sample set into calibration set and verification set through used Rank-KS(LIU Wei, et al., 2014) (content gradient method-Kennard-Stone) method. divided into two groups: calibration set and validation set, about the 154 samples in the study area. The calibration set included 116 samples for the construction of soil attribute inversion model, and the validation set included 38 samples for testing the prediction accuracy of the model.

Using the common spectral feature bands selected from *SG-CR* spectral transformation as independent variables of soil attribute inversion model, and constructed the panel data based on ordinary least squares estimation (*OLS*), about soil attribute content of 116 soil samples in Xinzheng city, as shown in Table 1.

The results show that the regression coefficient is significantly not 0, and the sample determination coefficient after adjustment R_v^2 is 0.9991, indicating that the goodness of fit of the model is high. A large F statistic indicates that the regression coefficient is significant and the regression model is significant as a whole.

2.5. Model accuracy check

This paper test the accuracy of soil PH, SOM, AN, AP, AK, Fe, Cr, Cd, Zn, Cu and Pb by the constructed panel data model. The results were shown in Table 2.

As can be seen from Table 2, each soil attribute in the correction set has a high coefficient R^2 , greater than 0.95. The highest is alkaline hydrolyzed nitrogen (0.998), and the corresponding root mean square error is 0.67. The lowest value was PH (0.95), and the corresponding root mean square error was 0.05, both of which were relatively low, indicating that the inversion model constructed by the panel data model could simultaneously realize the inversion of 11 soil properties, have good modeling accuracy.

According to the prediction results of the validation set, R_v^2 of PH, organic matter, alkali-hydrolyzed nitrogen, Cd and Cu were lower than the modeling set, and the rest were increased. In addition to Pb, the root mean square error was lower than that in the modeling set, R_v^2 of other soil attributes were increased. The relative errors are all greater than 2.5, indicating that the model has good quantitative prediction ability of PH, SOM, AN, AP, AK, Fe, Cr, Cd, Zn, Cu and Pb.

In order to more clearly display the modeling accuracy of Panel data of Fixed influence variable coefficient model, the soil attribute content diagram (Figure 4) and scatter diagram (Figure 5) of the measured value and the inversion value.

The comparison between Figure 4 and Figure 5 shows that, except for a few samples differences, the measured values and predicted values of most samples are concentrated near $y = x$, *i.e.* 1:1 line. The correlation

coefficients r between measured values and predicted values all pass the significance test at $P = 0.01$ level. This indicates that the panel data model with $SG-CR$ spectral transformation as independent variable has good predictive ability, and can be used to invert multiple soil properties simultaneously.

3. Conclusion

In construction of Well-facilitated capital farmland, in addition to emphasize on land leveling, roads, canals and other engineering facilities, and soil quality after completion reaches the higher level of the county, and the soil PH, SOM, soil nutrient, and soil pollution characteristic is indispensable basic information of cultivated land quality evaluation, it is also an important guarantee for the growing environment of crops. Therefore, the fast acquisition and real-time monitoring of soil basic information is the premise and foundation for the construction of Well-facilitated capital farmland.

Based on this, this research took the Well-facilitated capital farmland construction area in Xinzheng City as the research object, combined field sampling and indoor hyperspectral determination, conducted correlation analysis and Fuzzy clustering analysis of soil properties and $SG-CR$ transform spectrum, and selected the best common hyperspectral characteristic band. And constructed the panel data model of hyperspectral comprehensive inversion about 116 samples of 11 soil attribute, to achieve comprehensive and rapid acquisition of soil basic information. The results show that:

- (1) SG convolution smoothing of spectral reflectance can effectively remove noise, while preserving the overall characteristics of spectral curves, and improving signal-to-noise ratio. The method of spectral transformation is very important, for extracting characteristic bands and improving the prediction ability and stability of the model. By removing the envelope CR spectral transformation, the absorption and reflection characteristics of the spectral curve are effectively highlighted, the sensitivity of the spectrum is improved, and the useful information of the spectrum is enhanced.
- (2) After the correlation analysis about soil attributes and $SG-CR$ spectral transformation, and using Fuzzy clustering maximum tree method, combined with the similar inflection point of the correlation coefficient curve of soil attributes and spectral reflectance, select the common significance band of different soil attributes as the best hyperspectral characteristic band and focus on 405~431nm、781nm~831nm、1044~1087nm、1251~1410nm、1836~1898nm、2080nm~2201nm、2324~2395nm.
- (3) Based on the panel data model, constructed the comprehensive inversion model of soil properties in the well-facilitated capital farmland construction area of Xinzheng city, with the common spectral characteristic band of $SG-CR$ spectral transformation as the independent variable. The model was significant as a whole, and the goodness of fit was high ($R^2 = 0.9991, DW = 2.1899, F = 2195.67$). The coefficient R^2 of determination of 11 soil attributes were all greater than 0.95, and the root mean square root was small. From the perspective of relative analysis errors, the relative analysis errors of soil attributes of $SG-CR$ spectral transformation model are all greater than 2.5, and the measured and predicted values of most samples are concentrated near the 1:1 line, and the correlation coefficients r all pass the significance test at $P = 0.01$. It indicates that the panel data models with $SG-CR$ spectral transformation as independent variable have the ability of comprehensive inversion of soil properties, and have high precision prediction.

CRedit authorship contribution statement

All authors contributed to the study conception and design. Material preparation, data 395 collection and analysis were performed by [qixia Zhang], and [xinsheng Wang]. The first draft 396 of the manuscript was written by [qixia Zhang] and all authors commented on previous 397 versions of the manuscript. All authors read and approved the final manuscript. 398

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal 400 relationships that could have appeared to influence the work reported in this paper.

Data availability

The datasets analyzed during the current study are available from the corresponding author on reasonable request.

Funding information

This study was funded by Joint Funds of the National Natural Science Foundation of China(U21A20109), Scientific research projects in public welfare industries of Ministry of Land and Resources of China (201411022-2),PhD start-up Fund project of North China University of Water Resources and Electric Power(40934).

References

- BIAN Zijin, SUN Lina, TIAN Kang, LIU Benle, et al. (2021). Estimation of Heavy Metals in Tailings and Soils Using Hyperspectral Technology: A Case Study in a Tin-Polymetallic Mining Area. *Bulletin of environmental contamination and toxicology* , 107(6): 1-10. DOI: 10.1007/S00128-021-03311-7
- CAI Xiangwen,TANG Mi,SONG Yuening. (2019). On the Construction Potential of High Standard Basic Farmland in Gongcheng Yao Autonomous County. *Scientific and Technological Management of Land and Resources* , 36(05): 56-65. DOI: 10. 3969/j.issn.1009-4210 2019 05 006
- CHEN Lin,WU Kening,FENG Zhe,YU Bing, SONG Hengfei. (2019). Suitability Evaluation of High Standard Farmland Construction from Perspective of Ecological Civilization Construction. *Soils* , 51(04): 803-812. DOI:10.13758/j.cnki.tr.2019.04.023.
- Clark R N, Roush T L. (1984). Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. *Journal of Geophysical Research:Solid Earth* (1978-012), 89(B7):6329-6340.
- DONG Fei, ZHAO Wei. (2020). Delimitation of Construction Area of High-Standard Primary Farmland-Taking Nan'an District of Chongqing City as an Example. *Research of Soil and Water conservation* , 27(02): 344-349+356. DOI: 10.13869/j.cnki.rswc.2020.02.048
- DZ/T 0279-2016 Regional Geochemical Sample Analysis Method. (2016).
- LI Shumin, LI Hong, SUN Danfeng, et al. (2011). Characteristic and Diagnostic Bands of Heavy Metals in Beijing Agricultural Soils Based on Spectroscopy.*Chinese Journal of Soil Science* ,42(3):730-735. DOI: 10.1155/2011/531540
- LI C M, SHAO J A, GUO Y, CAO F, TAN S J. (2018). Construction potential of high-standard farmland based on landform factors. *Chinese Journal of Eco-Agriculture* , 26(7): 1067-1079. DOI: 10.13930/j.cnki.cjea.171122
- LI Hongxing, WANG Peizhuang. (1994). Fuzzy Mathematics. *Beijing: National Defense Industry Pres.*
- LI Long, WANG Zhaolin,WU Dafang,LIU Yanyan. (2020). Study on Timing Sequence and Regulation Mode of High Standard Basic Farmland Construction in Village Based on TOPSIS Model. *Research of Soil and Water Conservation* , 27(03): 286-293. DOI:10.13869/j.cnki.rswc.2020.03.041.
- LI Yilong, LIAO Heping, ZHANG Yafei, ZHANG Qianqian. (2019). Construction Conditions and Development Pattern of Well-Facilitated Farmland in Town Area Based on Rural Vitalization. *Journal of Southwest University(Natural Science Edition)* ,41(02):90-99. DOI:10.13718/j.cnki.xdzk.2019.02.013.
- LI zhiyuan , DENG fan, HE jun,WEI wei. (2021). Hyperspectral Estimation Model of Heavy METAL Arsenic in Soil. *Spectroscopy and Spectral Analysis* , 41(09):2872-2878. DOI:10.3964/j.issn.1000-0539(2021)09-2872-07
- LIU Wei, ZHAO Zong, YUAN Hongfu, et al. (2014). An optimal selection method of samples of calibration set and validation set for spectral multivariate analysis. *Spectroscopy and Spectral Analysis* , 34(4): 947-951. DOI: 10. 3964/j. issn. 1000-0593 (2014) 04-0947-05

- LIU Qi, LIN Huaizhong, CHEN Chun.(2004).Application of the Maximum Tree Algorithm for Fuzzy Clustering in the Clustering of Web Page.*Application Research of Computers* , 21(11): 286-287.DOI: 10.3390/rs12203312
- MA Xueying,SHAO Jing'an, CAO Fei. (2018). Comprehensive Performance Evaluation of High Standard Farmland Construction in Mountainous Counties - A Case Study in Dianjiang Chongqing. *Journal of Natural Resources* , 33(12): 2183-2199. DOI:10.31497/zrzyxb.20171185
- NY/T 2148-2012.Criterion of High Standard Farmland.(2012).
- Rossel RAV, Taylor HJ, Mcbratney AB. (2007). Multivariate calibration of hyperspectral γ -ray energy spectra for proximal soil sensing.*European Journal of Soil Science* , 58:343-353. DOI:10.1111/j.1365-2389.2006.00859.x
- Savitzky A, Golay M J E.(1964). Smoothing and differentiation of data by simplified least squares procedures. *Analytical Chemistry* , 36(8):1627-1639.
- Schreiner Simon,Culibrk Dubravko, Bandecchi Michele, Gross Wolfgang, Middelmann Wolfgang. (2021). Soil monitoring for precision farming using hyperspectral remote sensing and soil sensors.*Automatisierungstechnik* , 69(4): 116166. DOI: 10.1515/auto-2020-0042
- SHI Yuanyuan, ZHAO Junyu, SONG Xianchong, et al. (2021). Hyperspectral band selection and modeling of soil organic matter content in a forest using the Ranger algorithm. *PloS one* , 16(6): 0253385. DOI:10.1371/journal.pone.0253385
- SUN jingshui. (2010). Econometrics learning Guide and EViews Application Guide. *Beijing: Tsinghua University Press* .
- TANG Feng, Xu Lei, ZHANG Pengtao, ZHANG Guijun, et al. (2019). Suitability evaluation and priority area delineation of high standard basic farmland construction at county level. *Transactions of the Chinese Society of Agricultural Engineering* , 35(21):242-251. DOI:10.11975/j.issn.1002-6819.2019.21.030
- TD/T1033-2012 Standard for well-facilitied capital farmland construction. (2012).
- TONG Qingxi, ZHANG Bing, Zheng Lanfen, et al. (2006). Hyperspectral remote sensing: Principle, Technology and Application. *Beijing: Higher Education Press* .
- WANG Xiaoqing, SHI Wenjiao, SUN Xiaofang, WANG Meng. (2018). Comprehensive benefit evaluation and regional differences of construction projects of well-facilitated farmland in Huang Huai Hai region. *Transactions of the Chinese Society of Agricultural Engineering* , 34(16): 238-248. DOI:10.11975/j.issn.1002-6819.2018.16.031
- WANC Ke, LI Ling, LI Peng. (2021). Study on High Standard Farmland Construction Based on Ecological Security and Food Security.*Journal of Ecology and Rural Environment* , 37 (6):706-713. DOI:10.19741/j.issn.1673-4831.2020.0781.
- WANG Peizhuang. (1983). Fuzzy Set Theory and Its Application.*Shanghai: Shanghai Science and Technology Publishing House* .
- WEI Lifei, PU Haochen, WANG Zhengxiang, YUAN Ziran, et al. (2020). Estimation of Soil Arsenic Content with Hyperspectral Remote Sensing.*Sensors* , 20(14): 4056. DOI: 10.3390/s20144056
- XU Xitong, CHEN Shengbo,Xu Zhengyuan,Yu Yan, et al. (2020). Exploring Appropriate Preprocessing Techniques for Hyperspectral Soil Organic Matter Content Estimation in Black Soil Area. *Remote Sensing* , 12(22): 3765-3765. DOI:10.3390/rs12223765
- XIONG Yufei, HE Jing. (2019). Preliminary Study on Quality Control of High Standard Farmland Construction Projects. *Academic Journal of Business & Management* , 1(1): 7-16. DOI: 10.25236/AJBM.190102

YU Huan,LIU Jian,LIU Yaqiu,GU Ziping,WANG Ailing. (2021). Estimation of Soil Organic Matter Content of Cultivated Land in Hilly Regions with Hyperspectral. *Journal of Shandong Agricultural University(Natural Science Edition)* , 52(04):648-653.

ZENG Ya,ZHAO Wei,YANG Liujiào, et al. (2020). Research on Time Sequence of High-Standard Basic Farmland Construction in Peri -urban Areas: A Case Study of Nan'an District in Chongqing. *Resources and Environment in the Yangtze Basin* , 29(02):434-441.

ZHANG Hebing, ZHAO Suxia, CHEN Ningli, ZHANG Qiuxia. (2018). Project Optimization for Well-facilitated Farmland Construction Based on Coupling Coordination Degree. *Transactions of the Chinese Society for Agricultural Machinery* , 49(08):161-168. DOI:10.6041 / j. issn.1000-1298. 2018.08. 019

ZHANG Qiuxia, ZHANG Hebing, ZHANG Huijuan, et al. (2017). Hybrid Inversion Model of Heavy Metals with Hyperspectral Reflectance in Cultivated Soils of Main Grain Producing Areas. *Transactions of the Chinese Society for Agricultural Machinery* , 48(3): 148-155. DOI:10.6041/j.issn.1000-1298.2017.03.019

ZHANG Jing, SUN Lu. (2020). High Standard Farmland Construction and Soil Conservation Evaluation. *IOP Conference Series: Earth and Environmental Science* , 514(2): 022074. DOI: 10.1088/1755-1315/514/2/022074

Appendix

Figure 1 a.The spectral reflectance of before and after SG b. Zoom in contrast the spectral reflectance of 2000~2400nm

Figure 2 The reflectance spectra after Continuum removal

Figure 3 Comparison of of correlation coefficient between soil properties and the Continuum removal reflectance spectra

Figure 4 Inversion of soil properties elements on validation set

Figure 5 Scatter plot of measured values versus estimated values by Panel date model

Table 1 Panel date models for soil properties elements

Coefficien	PH
C	C
Fixed Influence Coefficient	-5.53
405nm	33.02
406nm	-52.06
407nm	-9.73
408nm	147.11
409nm	-297.04
410nm	304.53
411nm	-221.69
412nm	115.51
413nm	-91.28
414nm	109.94
415nm	-141.18
416nm	236.48
417nm	-230.27
418nm	57.23
419nm	36.04
421nm	50.78
422nm	-81.00

Coefficien	PH
423nm	35.27
427nm	0.65
428nm	-66.93
429nm	56.07
430nm	22.07
431nm	-15.32
830nm	-215.36
831nm	409.91
1044nm	-14.74
1062nm	-34.18
1079nm	-22.12
1085nm	79.73
1087nm	-5.93
1251nm	122.05
1267nm	-225.14
1308nm	-173.69
1309nm	124.66
1410nm	-22.16
1836nm	-39.79
1860nm	9.30
1887nm	114.12
1888nm	-184.61
1889nm	94.94
1890nm	-6.97
1897nm	-1.32
1898nm	-5.83
2080nm	14.93
2118nm	51.72
2119nm	-24.67
2137nm	3.02
2149nm	-1.08
2156nm	-6.56
2184nm	14.69
2185nm	-16.47
2186nm	25.80
2187nm	-22.49
2198nm	-54.11
2199nm	117.41
2200nm	-65.44
2201nm	-7.19
2324nm	12.60
2325nm	-7.81
2382nm	-0.30
$R^2= 0.9996 R_v^2= 0.9991 SSR =1.51 DW = 2.1899 F=2195.67 Prob = 0 R^2= 0.9996 R_v^2= 0.9991 SSR =1.51 DW = 2.1$	

Table 2 Calibration and validation for soil properties elements using Panel date model

Calibration set	Calibration set	Validation set	Validation set	Validation set
R^2	$RMSEC$	R_v^2	$RMSEP$	RPD

	Calibration set	Calibration set	Validation set	Validation set	Validation set
PH	0.95	0.05	0.92	0.06	3.56
SOM	0.98	0.31	0.97	0.47	5.68
AN	0.998	0.67	0.997	0.85	20.74
AP	0.989	0.96	0.998	0.31	24.64
AK	0.983	3.80	0.987	1.79	8.73
Fe	0.96	0.54	0.988	0.29	9.05
Cr	0.99	1.03	0.992	1.05	11.20
Cd	0.96	0.012	0.89	0.02	2.98
Zn	0.988	0.52	0.991	0.82	10.36
Cu	0.992	0.68	0.989	0.52	9.23
Pb	0.987	0.77	0.989	0.70	9.38