

Inversion of soil properties with hyperspectral reflectance in construction areas of high-standard farmland

Zhang Qiuxia⁽¹⁾ , Liu Wenkai⁽¹⁾ , Zhang Hebing⁽²⁾ , Wang Xinsheng^{(2,3)*}  and Ma Shouchen⁽²⁾ 

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

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

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

ABSTRACT: High-standard farmland construction is an important process that can enhance food security and accelerate new-style modernization agriculture. Hyperspectral remote sensing can provide data and technical support for this type of construction to provide a reference when optimizing high-standard farmland construction areas. This study was performed in Xinzheng City, the primary grain-producing areas in Henan Province. Field sampling and indoor hyperspectral spectroscopy (350~2500 nm) were combined; spectral transformations such as continuum removal (CR) were performed after Savitzky–Golay (SG) convolution smoothing; and the best hyperspectral bands were selected as the common index of the soil properties by correlation analysis and fuzzy clustering maximum tree. A hyperspectral inversion model was built for the panel data model of the fixed effect variable coefficient based on the ordinary least squares estimation method (OLS), including panel data describing pH, organic matter, nitrogen, phosphorus, potassium, iron, chromium, cadmium, zinc, copper, and lead of 116 samples in Xinzheng City. Results show that the panel data model is of good quality overall, and the goodness of fit is higher ($\bar{R}^2 = 0.9991$, $F = 2195.67$). The precision test results indicate that the models performed well at both description and prediction, including accurate quantification, with an *RPD* above 2.5. Thus, the proposed model provides an important basis for soil information management, resource evaluation, and a reference when optimizing high-standard farmland construction processes.

Keywords: high-standard farmland, hyperspectrum, panel data model, maximum tree clustering method.

* Corresponding author:
E-mail: hpuwx@s@126.com

Received: April 19, 2023

Approved: September 26, 2023

How to cite: Qiuxia Z, Wenkai L, Hebing Z, Shouchen M, Xinsheng W. Inversion of soil properties with hyperspectral reflectance in construction areas of high-standard farmland. Rev Bras Cienc Solo. 2023;47:e0230033
<https://doi.org/10.36783/18069657rbc20230033>

Editors: José Miguel Reichert  and Marcos Gervasio Pereira 

Copyright: This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided that the original author and source are credited.



INTRODUCTION

Core areas for grain production have become a national strategy in China. Henan Province is a critical area of grain production, and an important bottleneck that restricts sustainable agricultural development in Henan is its high population density, which produces a serious shortage of reserve resources of cultivated land, and an overall low quality of cultivated land. Therefore, strengthening high-standard farmland construction is important to apply China's national strategy and promote agricultural production.

High-standard farmland refers to basic farmland with centralized contiguity, supporting facilities, high and stable yields, good ecology, and strong disaster resistance formed through rural land consolidation and construction in a certain period, which is compatible with modern agricultural production and operation modes (Ministry of Land and Resources of the People's Republic of China, 2012). To date, most studies of high-standard farmland construction include high-standard farmland construction demarcation (Zhang et al., 2018; Li et al., 2019; Dong et al., 2020), potential evaluation (Cai et al., 2019; Li et al., 2018), suitability evaluation (Tang et al., 2019; Chen et al., 2019; Zhang et al., 2020), construction sequence and mode zoning (Li et al., 2020; Wang et al., 2021; Zeng et al., 2018), project implementation and effect evaluation (Ma et al., 2018; Xiong et al., 2019; Wang et al., 2018), etc. In the standard of high-standard farmland construction, "the quality of cultivated land after completion reaches the higher level of the county", etc; therefore, in the construction of high-standard farmland the construction of field projects requires that the soil quality should reach the higher level of the region. Thus, the amount of fertilizer required should be determined soil nutrient status and soil nitrogen, phosphorus, potassium, medium and trace elements, organic matter content, soil acidification, and salinity. The soil should also be regularly monitored for other conditions, and the fertilization should be constantly adjusted based on real-world measurements (The Ministry of Agriculture of the People's Republic of China, 2012).

However, during high-standard farmland construction, in addition to more land levelling, the improvement of supporting facilities of roads, ditches and other projects strengthens the rapid acquisition and real-time monitoring of basic soil information. However, most research methods use quantitative inversion models with single soil properties, e.g., partial least squares regression (*PLSR*) (Wei et al., 2020; Bian et al., 2021; Li et al., 2021; Yin et al., 2021), back propagation neural networks (BPNN) (Yu et al., 2021), random forest (RF) (Schreiner et al., 2021; Chen et al., 2021), and support vector machine (SVM) (Shi et al., 2021; Xu et al., 2020; Selvakumar et al., 2021). Inversion models of soil properties and spectral reflectance must be built one by one, and their calculation is complex and time-consuming.

A panel data model can build a three-dimensional data model concurrently and consider various soil properties of multiple points and high spectral characteristics of the band values. Using inversion modelling, multiple soil properties are considered, model calculation is simpler, and the relationship between each model analysis of soil properties and the influence of high spectral band characteristic values on each soil property can be determined. Therefore, it is necessary to study the rapid and nondestructive testing of soil attribute information in high-standard farmland construction areas to provide technical support for the rapid acquisition and real-time monitoring of soil properties, and to support the optimization of high-standard farmland construction areas.

This study considered the high-standard farmland construction area of Xinzheng City and attempts to establish a comprehensive hyperspectral inversion model of cultivated land soil properties using a panel data model, estimate the influence of hyperspectral characteristic band values on each soil attribute, and predict the content of each soil property to provide theoretical and technical support for the rapid acquisition and real-time monitoring of soil properties in high-standard farmland construction areas.

MATERIALS AND METHODS

Overview of the researched area

Xinzheng City is located in the central part of Henan Province in China, which is a transition zone from the North China Plain and western Henan Mountains to the eastern Henan Plain; and is the core of the Central Plains economic zone under Zhengzhou City, which is located at 34° 16'~34° 39' N, 113° 30'~ 113° 54' E. Xinzheng City has a total population of 653,000 and, in 2019, had jurisdiction over 9 towns, 27 townships, 3 streets, 253 administrative villages, 921 natural villages and 24 residential areas.

According to a survey of land-use status in 2013, the total land area of Xinzheng City is 884.5915 km², and the cultivated land is 521.7641 km², accounting for 58.59 % of the total land area. The total annual grain output is 273148 Mg. According to the Integrated Land-use Planning of Xinzheng City (2010-2020), the protection index of prime farmland in Xinzheng City is 427.73 km². There are various soil types, primarily cinnamon soil, tidal soil and aeolian sand soil. The terrain is high in altitude in the west and low in the east, with shallow hills in the west, plains in the east and hills in the northwest.

Field collection of soil samples

According to the soil type, topographic characteristics and spatial variation of the study area, and considering the integrity of administrative units (towns or villages as units), sampling points were laid out in a 2 × 2 km regular grid pattern, the sampling depth was 0.00-0.30 m on the surface of the soil. A total of 154 soil samples were collected in this sampling, and invasive bodies, such as plant roots and stem residues and brick and tile fragments, were removed, as shown in figure 1. After natural air drying, grinding and passing through a 1-mm sieve, the samples were divided into four parts using the quartering method in duplicate: one part was used to determine each sample's physical and chemical properties in the laboratory, and the other part was used to determine the soil spectrum.

Soil properties measured in this study were soil pH, organic matter, nitrogen, phosphorus, potassium, iron, chromium, cadmium, zinc, copper, lead. They were measured using the regional geochemical sample analysis method (Ministry of Land and Resources of the People's Republic of China, 2012). To ensure a high-quality analysis, national geochemical standard samples were used for quality control.

Laboratory test of sample spectra

This study measured soil spectral reflectance by an ASD spectrometer on treated soil samples under indoor conditions. The spectrum measurement instrument is an ASD Field Spec 3 spectrometer produced by ASD, USA. Spectral range was 350-2500 nm with a sampling interval of 1.4 nm for 350-1000 nm, a sampling interval of 2 nm for 1000-2500 nm, and a resampling interval of 1 nm. Before spectral measurement, the surface of the soil was scraped in the same direction along the edge of the soil sample vessel with a ruler and then filled with a soil sample dish placed with black rubber pad with a reflectivity of approximately 0. A halogen lamp with a power of 50 W was used as light source, and a probe with a view angle of 25° and a light incidence angle of 45° was used. The distance to the light source was 15 cm, and the distance to the probe was 15 cm. To reduce the influence of anisotropic soil sample spectra, when measuring, the sample plate was turned three times, each time with a rotation angle of approximately 90°, and the soil sample spectra were measured in four directions. Reference plate calibration was performed before and after each target spectrum acquisition and repeated five times for 20 times. View Spec Pro software was then used to obtain the average spectral reflectance based on the original reflectance's spectral value. With the band test range in the spectral data near the two ends being unstable, we removed data at 350-399 and 2401-2500 nm, which are strongly affected by external noise.

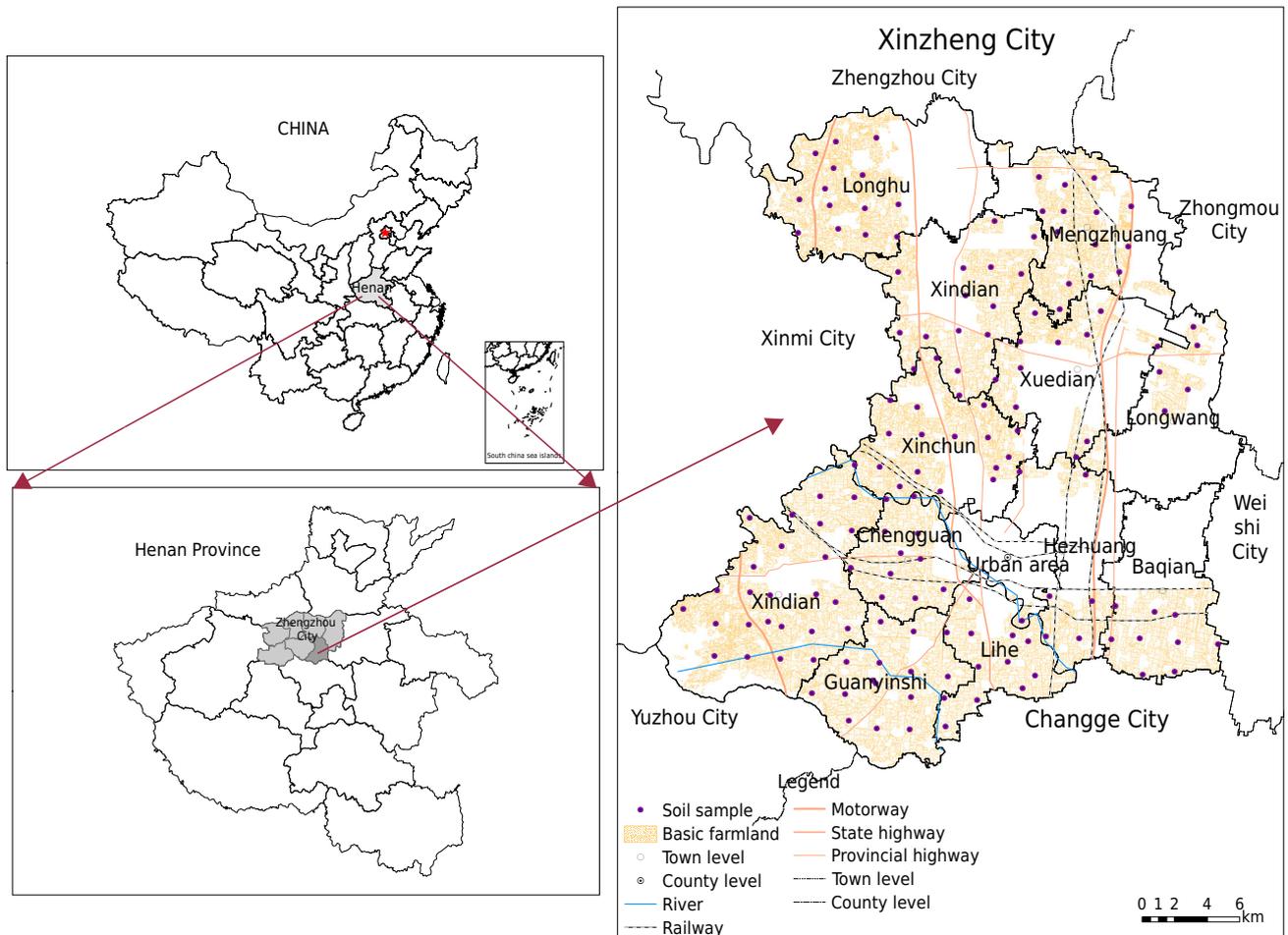


Figure 1. Soil sample distribution.

Model establishment and accuracy test

Fuzzy clustering maximum tree method

Fuzzy theory was developed on the mathematical basis of fuzzy set theory established by American cybernetics expert Professor L.A. Zadeh in 1965 and has been widely used in mathematics and many other fields (Liu et al., 2004). Fuzzy clustering number is a multitechnology that classifies objective things using the fuzzy mathematics method, establishing similarity relations according to the characteristics, similarity and affinity degree of the objective (Li et al., 1994; Wang et al., 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 the sample set matrix

Assuming that the sample set $U = \{x_1, x_2, \dots, x_n\}^T$, in which n is the number of samples, and each sample has an m dimensional vector representation (i.e., each sample has m indicators), then:

$$U = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1m} \\ x_{21} & x_{22} & \dots & x_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nm} \end{bmatrix}$$

(2) Establish the fuzzy similarity matrix

According to the given sample characteristic data, the correlation coefficient method is used to establish the fuzzy similarity matrix $R = (r_{ij})_{n \times n}$, in which r_{ij} is the similarity coefficient between different samples:

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix} \quad i, j = 1, 2, \dots, n ; \quad 0 \leq r_{ij} \leq 1$$

(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 called the largest tree, which may not be unique.

(4) Clustering

We then select the appropriate threshold λ , cut off the branches of the weight $r_{ij} < \lambda$, and obtain an unconnected graph. Each connected branch constitutes the classification of horizontal λ , and there are several branches indicating the classification of several categories.

In this study, the fuzzy similarity coefficient between the correlation coefficient curves of soil properties and spectral indices was calculated by a systematic clustering method, and a fuzzy similarity matrix was constructed to determine the similarity of the correlation coefficient between different soil properties and spectral indices. On this basis, the common hyperspectral inversion bands of different soil properties were determined by the maximum tree classification method.

Panel data model

Panel data are also called parallel data, time series and cross-section data, or pool data, and refer to taking multiple cross-sections on a time series. Sample data formed by sample observations are simultaneously selected on these cross sections (Sun et al., 2010). Using a cross section, we can perform a cross section observation that is formed by several individuals at a certain moment. Also, we can describe a time series from a longitudinal section. According to the characteristics of panel data, the hyperspectral characteristic band values of soil properties of multiple samples can be considered to be 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. Using a panel data model, a comprehensive inversion model of soil properties can also be developed without an individual inversion of each index, which reduces the tedious process of multi-index inversion (Zhang et al., 2021).

Due to the many sample points T and the few cross sections N , a fixed influence model was developed, and ordinary least squares estimation (OLS) was used to build the panel data model. Then, panel data model types were determined via an analysis of covariance, the invariant coefficient model, variable intercept model and variable coefficient model. 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 described by equation 1:

$$y_{it} = a_i + b_{1i}x_{1it} + b_{2i}x_{2it} + \cdots + b_{ki}x_{kit} + \mu_{it} \quad (i = 1, 2, \dots, N; t = 1, 2, \dots, T) \quad \text{Eq. 1}$$

in which: y_{it} is the values of explained variables on cross Section i and sample t , soil heavy metal element content; a_i is a constant or intercept term that represents the cross-section of i (influence of the individual of i); b_{ji} is the model parameter of the j th explanatory variable on the i th cross-section; x_{jit} is the value of the j th explanatory variable on cross Section i and sample t , the reflectance of the hyperspectral characteristic band of soil heavy metals; u_{it} is the random error term on cross Section i and sample t ; k is the number of explanatory variables.

Accuracy test method of the inversion model

Calibration set determination coefficient \bar{R}^2 and root mean square error (RMSEC) are used to verify the modeling accuracy. Validation set test is based on the validation set determination coefficient \bar{R}_v^2 , root mean square error (RMSEP) and relative percent deviation (RPD). The relative percent deviation is the ratio between the standard deviation and RMSEP of the validation set. When $RPD > 2.5$, the model is shown to have 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 some quantitative prediction ability. When $1.40 < RPD \leq 1.80$, the model has only general quantitative prediction ability. When $1.00 < RPD \leq 1.40$, the model has the ability to distinguish a high value from a low value. When $RPD \leq 1.00$, the model has no predictive ability (Rossel et al., 2006). To set the model's parameters, the larger \bar{R}^2 is, the smaller RMSEC is, the higher the modelling accuracy is, and the more stable the model is. For the verification set, the larger \bar{R}_v^2 and RPD are, the smaller RMSEP is, and the higher the prediction accuracy is.

RESULTS

Spectral pretreatment

During ASD spectrometer 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. To obtain a 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 et al., 2021) in 1964 and is a weighted average method that obtains smooth point data by least square fitting of the data that are to be measured in the moving window interval using a polynomial method. Convolution smoothing is widely used currently, and during SG filtering, appropriate smoothing points and polynomial fitting times must be selected. The more smoothing points that are considered, the smoother the spectral curve will be, but some information will also be lost. Therefore, SG filtering smoothing based on a 9-point quadratic polynomial is used. The transform tool used for smoothing and denoising was Unscrambler 9.7, as shown in figure 2.

To describe the smoothing effect more accurately, the band curves at 2000~2400 nm were amplified (Figure 2b). By comparing the details before and after SG smoothing, SG smoothing can effectively remove noise and better preserve the overall characteristics of spectral curves.

Continuum removal

To describe the sensitive relationship between soil heavy metal content and spectral reflectance, continuum removal (CR) spectral transformation after SG smoothing is required. A spectral analysis method was proposed first by Clark and Rous in 1984 (Clark et al., 2021) and is defined as a point in a straight line connected with the wavelength change reflect or absorb protruding point of "peak value", making the line in the "peak value" on the outside greater than 180° (Tong et al., 2006), the real spectrum reflectance

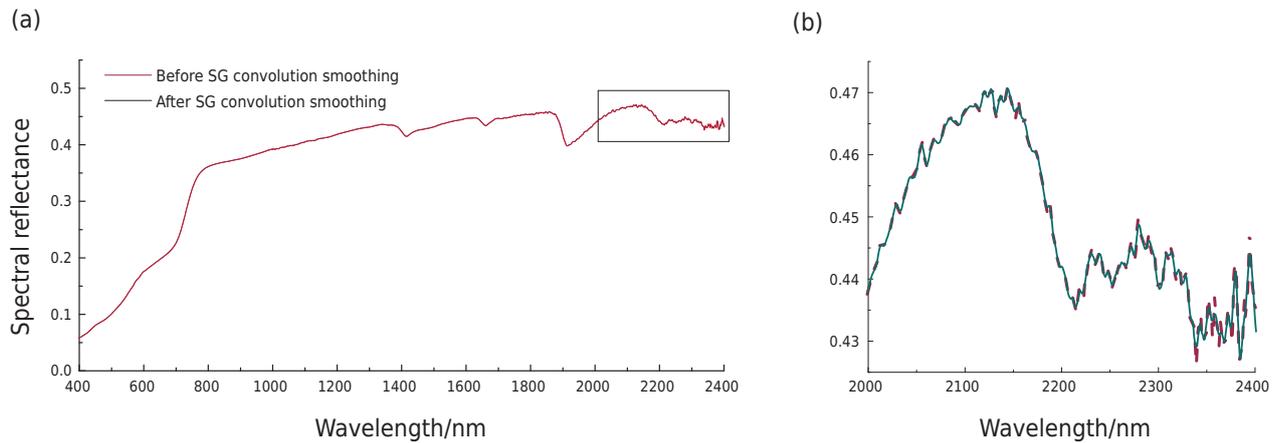


Figure 2. Spectral reflectance before and after SG b. Zoom conversely the spectral reflectance of 2000~2400 nm.

and the envelope line of the corresponding band reflectance ratio. By normalizing the spectral value to 0~1 (Li et al., 2021), the absorption and reflection characteristics of the spectral curve can be effectively highlighted, and the characteristic bands can be extracted. Via the 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. We performed the envelope removal line by constructing the Popper database in Envi 4.8, as shown in figure 3.

Reflectance curve of *CR* enhances the spectral characteristics of the original spectral curve at 1400, 1900 and 2200 nm, and also highlights the weak absorption characteristics at 410, 500 and 700 nm. These results show 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 improves the extraction of effective characteristic bands.

Selection of common spectral characteristic bands for soil properties

Based on soil property significant band selection for the Xinzheng high-standard farmland construction area, and considering the needs of different soil property spectrum inversions, combined with the correlation coefficient curve similarity and inflection point, we use the method of fuzzy clustering tree to determine the share of the best band of hyperspectral inversion of soil properties for the Xinzheng high-standard farmland construction area.

Using comparative analysis of the correlation coefficient curves of 11 soil properties and *SG-CR* transformations of Xinzheng City high-standard farmland area (Figure 4), the correlation coefficient curves of the same spectral transformation have similar inflection points, showing good similarity.

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 the correlation coefficient curves of 11 soil properties and *SG-CR* spectral transformation is:

$$R_{SG-CR} = \begin{bmatrix} 1 & 0.03 & 0.29 & 0.87 & 0.28 & 0.12 & 0.39 & 0.39 & 0.46 & 0.42 & 0.69 \\ 0.03 & 1 & 0.87 & 0.16 & 0.58 & 0.02 & 0.06 & 0.18 & 0.72 & 0.25 & 0.01 \\ 0.29 & 0.87 & 1 & 0.51 & 0.56 & 0.10 & 0.10 & 0.19 & 0.39 & 0.27 & 0.31 \\ 0.87 & 0.16 & 0.51 & 1 & 0.38 & 0.03 & 0.38 & 0.76 & 0.29 & 0.25 & 0.76 \\ 0.28 & 0.58 & 0.56 & 0.38 & 1 & 0.08 & 0.44 & 0.38 & 0.25 & 0.07 & 0.51 \\ 0.12 & 0.02 & 0.10 & 0.03 & 0.08 & 1 & 0.25 & 0.02 & 0.07 & 0.67 & 0.21 \\ 0.39 & 0.06 & 0.10 & 0.38 & 0.44 & 0.25 & 1 & 0.71 & 0.29 & 0.49 & 0.84 \\ 0.69 & 0.18 & 0.19 & 0.76 & 0.38 & 0.02 & 0.71 & 1 & 0.55 & 0.32 & 0.93 \\ 0.46 & 0.72 & 0.39 & 0.29 & 0.25 & 0.07 & 0.29 & 0.55 & 1 & 0.43 & 0.43 \\ 0.42 & 0.25 & 0.27 & 0.25 & 0.07 & 0.67 & 0.49 & 0.32 & 0.43 & 1 & 0.48 \\ 0.69 & 0.01 & 0.31 & 0.76 & 0.51 & 0.21 & 0.84 & 0.93 & 0.43 & 0.48 & 1 \end{bmatrix}$$

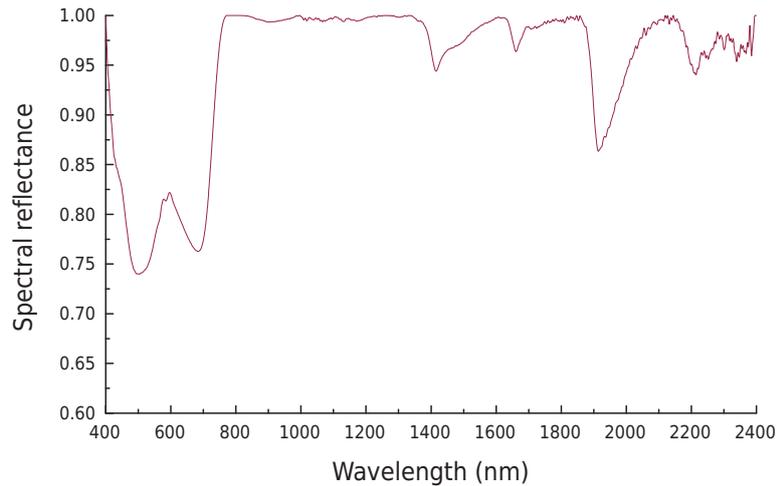


Figure 3. Reflectance spectra after continuum removal.

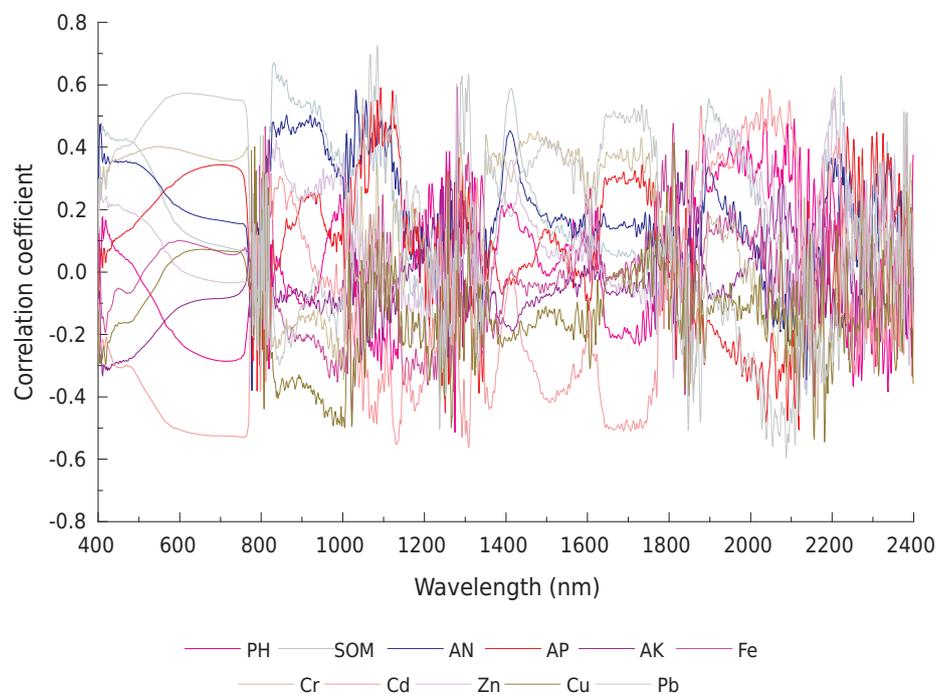


Figure 4. Comparison of the correlation coefficient between soil properties and the continuum removal reflectance spectra.

Maximum tree classification method uses $\lambda = 0.76$ and classifies pH, AP, Cr, Cd, and Pb as a class; SOM and AN as another class; AK, Fe, and Cu as another class; and Zn as a class. According to the similarity and inflection point of correlation coefficient curves, the band of SG-CR spectral transformation at 405, 418, 781, 784, 794, 805, 807, 830, 831, 1079, 1085, 1251, 1267, 1308, 1309, 1410, 1836, 1860, 1897, 1898, 2080, 2137, 2149, 2156, 2184, 2382, and 2395 nm were investigated. Concurrently, according to the significant bands of soil properties and SG-CR spectrum, the important bands of 11 soil properties passing the significance level test of $p=0.01$ were 406~419, 421~423, 427~431, 1044, 1062, 1087, 1887~1890, 2118, 2119, 2185~2187, 2198~2201, 2324, and 2325 nm.

The important band and the inflection point of the correlation coefficient curve were combined to determine that 405~419; 421~423; 427~431; 781, 784, 794, 805, 807,

830, 831, 1044, 1062, 1079, 1085, 1087, 1251, 1267, 1308, 1309, 1410, 1836, 1860, 1887~1890, 1897, 1898, 2080, 2118, 2119, 2137, 2149, 2156, 2184~2187, 2198~2201, 2324, 2325, 2382, and 2395 nm were common spectral characteristic bands of the soil properties' spectral inversions in high-standard farmland construction area of Xinzheng City.

Construction of panel data model

Based on soil types, the sample set was divided into a calibration set and a verification set using the *Rank-KS* (Liu et al., 2014) (i.e., the content gradient method-Kennard-Stone) method. The 154 samples in the study area were then divided into two groups: a calibration set and a validation set. Calibration set included 116 samples for the construction of the soil property 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 the soil property inversion model, panel data were constructed based on ordinary least squares estimation (OLS) for the soil property content of 116 soil samples in Xinzheng city, as shown in table 1.

Results show that the regression coefficient does not significantly equal 0, and the sample determination coefficient after adjustment \bar{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.

Model accuracy check

This study also tested the accuracy of soil pH, SOM, AN, AP, AK, Fe, Cr, Cd, Zn, Cu and Pb by the constructed panel data model. Results are shown in table 2. As shown in table 2, each soil property in the correction set has a high coefficient \bar{R}^2 , greater than 0.95. The highest is alkaline hydrolysed nitrogen (0.998), and the corresponding root mean square error is 0.67. The lowest pH was 0.95, and the corresponding root mean square error was 0.05, both of which were particularly low, indicating that the inversion model constructed by the panel data model could simultaneously realize the inversion of 11 soil properties and have good modelling accuracy.

According to the prediction results of the validation set, \bar{R}_V^2 of pH, organic matter, alkali-hydrolysed nitrogen, Cd and Cu were lower than the modeling set; the other metrics increased. In addition to Pb, the root mean square error was lower than that in the modeling set, and the \bar{R}_V^2 of the other soil properties increased. 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.

To more clearly demonstrate the modeling accuracy of the panel data of the fixed influence variable coefficient model, the soil property content diagram (Figure 5) and scatter diagram (Figure 6) of the measured value and the inversion value are shown.

Comparing figures 5 and 6 shows that, except for a few sample differences, the measured and predicted values of most samples are concentrated near $y = x$ (i.e., the 1:1 line). Correlation coefficients *r* between the measured values and predicted values all pass the significance test at the $p=0.01$ level, which indicates that the panel data model with SG-CR spectral transformation as an independent variable has good predictive ability and can be used to invert multiple soil properties simultaneously.

Table 1. Panel date models for soil property elements

Coefficient	Property										
	pH	SOM	AN	AP	AK	Fe	Cr	Cd	Zn	Cu	Pb
C	6.84										
Fixed influence coefficient	-5.53	-6.09	0.85	19.54	9.62	-3.67	-3.53	-15.13	4.28	5.17	-5.52
405 nm	33.02	153.49	-107.79	-950.01	-182.82	49.43	-104.85	345.04	-303.37	-669.61	452.59
406 nm	-52.06	-294.39	174.71	1365.99	82.18	-192.23	240.56	-510.46	512.44	1006.10	-925.44
407 nm	-9.73	470.04	122.84	-186.55	521.91	344.81	-194.45	301.41	-116.87	-296.99	1182.11
408 nm	147.11	-1014.64	-839.27	-1961.48	-1395.50	-470.80	162.11	-166.39	-680.27	-465.11	-2102.65
409 nm	-297.04	1622.78	1660.49	4376.54	2453.90	643.57	-362.05	61.41	1419.79	935.93	3447.06
410 nm	304.53	-1591.74	-1727.53	-4547.27	-2468.43	-604.22	402.19	-184.22	-1453.63	-851.17	-3515.82
411 nm	-221.69	342.55	1069.93	5300.93	2082.38	107.82	114.42	987.10	1749.87	864.23	1199.43
412 nm	115.51	1406.37	-52.31	-6417.12	-1610.87	573.50	-862.15	-2937.10	-2233.96	-1054.23	1164.39
413 nm	-91.28	-2311.87	-349.44	7068.12	1789.96	-887.37	1170.32	4031.25	2503.41	1830.10	-2076.97
414 nm	109.94	1871.51	110.26	-5971.28	-1425.11	875.18	-1112.39	-3435.36	-2237.72	-2069.80	1802.89
415 nm	-141.18	-933.01	318.63	4420.07	672.29	-809.73	883.40	2048.06	1799.34	1234.85	-751.94
416 nm	236.48	732.51	-824.70	-6222.80	-1407.90	733.01	-670.12	-489.75	-2199.77	-1352.49	447.82
417 nm	-230.27	-1260.44	696.88	7490.20	1915.12	-742.82	653.92	-214.95	2373.49	1867.73	-1689.05
418 nm	57.23	1634.95	128.70	-4784.94	-834.83	788.29	-614.26	-380.14	-1438.85	-1173.10	2693.22
419 nm	36.04	-1205.25	-417.34	1910.23	130.04	-584.12	273.62	980.86	490.75	180.83	-1697.15
421 nm	50.78	1293.33	22.74	-3842.12	-1625.98	345.70	8.75	-919.85	-980.68	-1036.70	1081.70
422 nm	-81.00	-1396.19	86.20	4329.70	2088.97	-403.48	58.49	1206.03	1262.37	1460.58	-1233.87
423 nm	35.27	601.64	-19.98	-1462.84	-836.49	261.34	-122.29	-732.65	-534.58	-605.55	674.91
427 nm	0.65	-332.37	61.11	476.10	34.15	-304.21	242.64	320.86	219.96	104.07	-990.95
428 nm	-66.93	602.87	174.01	483.18	645.61	661.20	-557.22	-1309.50	-28.98	706.19	1673.66
429 nm	56.07	-1103.54	-259.09	849.27	339.82	-603.49	339.96	1548.48	292.55	-278.98	-1643.83
430 nm	22.07	986.44	-136.86	-2324.94	-1916.10	151.24	188.65	-819.10	-525.43	-763.22	793.46
431 nm	-15.32	-257.07	136.25	600.43	967.98	60.83	-154.98	310.56	105.02	363.16	17.32
830 nm	-215.36	3698.08	1662.45	-11953.80	-6589.09	1690.79	-1747.89	-5665.86	-1234.73	-8960.59	1525.16
831 nm	409.91	-2478.08	-2288.66	5675.05	3207.34	-1406.43	1432.47	6472.64	-387.16	6078.14	-664.42
1044 nm	-14.74	-79.23	74.45	1174.69	343.61	25.45	83.35	125.30	311.06	284.02	-225.24
1062 nm	-34.18	145.10	205.03	373.67	172.78	38.56	-213.00	-128.73	133.03	-418.89	35.02
1079 nm	-22.12	-380.81	-80.90	1557.09	642.26	80.65	-7.39	-714.37	385.31	879.35	-246.60
1085 nm	79.73	224.68	-262.73	-2641.23	-742.39	73.06	-33.63	-97.15	-842.88	-584.07	747.15
1087 nm	-5.93	101.54	138.28	-87.20	-246.17	-181.76	156.95	397.00	130.79	-64.83	-204.14
1251 nm	122.05	313.16	-540.28	-4296.39	-1959.71	-74.03	-37.67	1042.33	-1077.92	-1212.65	693.75
1267 nm	-225.14	-650.78	768.55	7998.22	3819.77	460.39	-94.29	-1106.79	1669.03	3112.68	-1224.99
1308 nm	-173.69	235.55	1385.11	7567.09	3774.58	258.33	-1080.48	1069.39	1326.98	707.28	-1063.33
1309 nm	124.66	-118.70	-1060.30	-5078.62	-2990.52	-157.90	933.49	-510.43	-834.20	-512.58	1168.26
1410 nm	-22.16	-95.93	79.85	639.98	319.59	-16.05	30.57	-287.85	198.69	276.94	-136.82
1836 nm	-39.79	-80.04	171.37	1196.47	548.78	28.57	-20.32	-353.94	290.06	542.24	-172.54
1860 nm	9.30	0.65	-76.29	-458.77	-174.21	12.23	50.19	162.89	-120.31	-129.44	56.13
1887 nm	114.12	55.32	-855.26	-2247.20	-2034.64	-261.07	681.20	1329.56	-615.62	-1215.62	-28.71
1888 nm	-184.61	-262.40	1600.62	4328.27	4116.28	330.05	-948.84	-2210.04	1178.59	2627.76	110.01
1889 nm	94.94	206.79	-1009.55	-2185.98	-2202.53	154.55	299.08	620.78	-575.97	-1583.18	-300.57
1890 nm	-6.97	-26.46	169.99	-186.05	-91.81	-252.05	32.34	266.71	-33.89	132.59	152.84
1897 nm	-1.32	68.64	18.89	-843.97	13.40	1.29	-21.73	412.53	-138.16	-142.30	91.73
1898 nm	-5.83	-32.26	21.90	932.63	125.55	41.45	-38.17	-381.07	142.85	155.57	-15.16
2080 nm	14.93	15.30	-46.47	-256.17	-133.43	-17.21	-5.31	185.96	-74.60	-208.16	59.42
2118 nm	51.72	12.39	-316.73	-1354.11	-754.53	-163.41	294.91	691.10	-242.01	-740.97	8.75

Continue

Continue

Coefficient	Property										
	pH	SOM	AN	AP	AK	Fe	Cr	Cd	Zn	Cu	Pb
2119 nm	-24.67	-9.05	141.04	399.56	341.59	128.50	-178.56	-695.11	64.32	619.03	45.91
2137 nm	3.02	-27.69	-40.57	151.77	-4.72	-13.48	-8.34	119.63	47.12	-13.47	-59.03
2149 nm	-1.08	-1.61	-7.37	-19.71	-29.55	1.05	-43.72	-91.70	-32.19	-17.52	37.44
2156 nm	-6.56	-0.70	51.03	144.33	161.66	19.77	-19.76	-174.00	45.95	77.73	-44.64
2184 nm	14.69	-27.60	-85.33	-880.23	-497.16	-295.07	-143.05	690.23	-120.59	-959.91	-114.53
2185 nm	-16.47	-5.49	131.12	1622.98	781.90	473.70	505.54	-1060.03	146.02	1953.79	160.62
2186 nm	25.80	-12.20	-100.15	-1302.62	-269.90	-250.03	-589.73	66.32	-55.94	-1260.11	-299.90
2187 nm	-22.49	74.27	88.74	551.82	-85.84	49.78	230.26	416.52	12.56	146.43	273.56
2198 nm	-54.11	-15.18	415.74	1056.95	509.31	107.89	183.04	-858.43	352.04	652.05	-188.52
2199 nm	117.41	-66.14	-983.16	-2821.29	-1116.33	-372.61	-412.84	1829.88	-840.67	-1364.11	491.79
2200 nm	-65.44	246.79	758.23	1787.28	384.85	391.63	404.03	-1268.42	527.61	607.52	-241.32
2201 nm	-7.19	-197.18	-159.72	210.76	358.89	-126.55	-163.44	235.19	42.77	244.80	-89.79
2324 nm	12.60	2.50	-38.68	40.85	-92.19	-21.19	3.58	22.74	4.99	-9.95	22.85
2325 nm	-7.81	0.72	7.17	-200.31	36.49	17.74	-6.82	-13.04	-38.34	-23.70	-6.51
2382 nm	-0.30	-2.27	-7.06	19.15	-23.56	-0.05	20.66	38.74	2.14	-10.89	-12.00

$\bar{R}^2 = 0.9996$ $\bar{R}_v^2 = 0.9991$ $SSR = 1.51$
 $DW = 2.1899$ $F = 2195.67$ $Prob = 0$

Table 2. Calibration and validation for soil property elements using the panel data model

Property	Calibration set		Validation set		
	\bar{R}^2	RMSEC	\bar{R}_v^2	RMSEP	RPD
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

DISCUSSION

Hyperspectral remote sensing technology has multiple bands, high resolution and large spectral information, which can provide spatial information on soil surface conditions and properties, and evaluate and detect subtle differences in soil properties, providing conditions to determine basic soil information quickly, accurately and efficiently. Therefore, introducing hyperspectra to the acquisition of soil information in high-standard farmland construction areas is helpful for real-time monitoring of basic soil information and provides a new technique for acquiring soil information. However, most research on hyperspectral inversion of soil properties only applies single quantitative inversions with soil properties. This study found that the panel data model achieves good predictions and is thus feasible in high-standard farmland construction areas, providing technical support for the acquisition of soil property data in high-standard farmland construction areas.

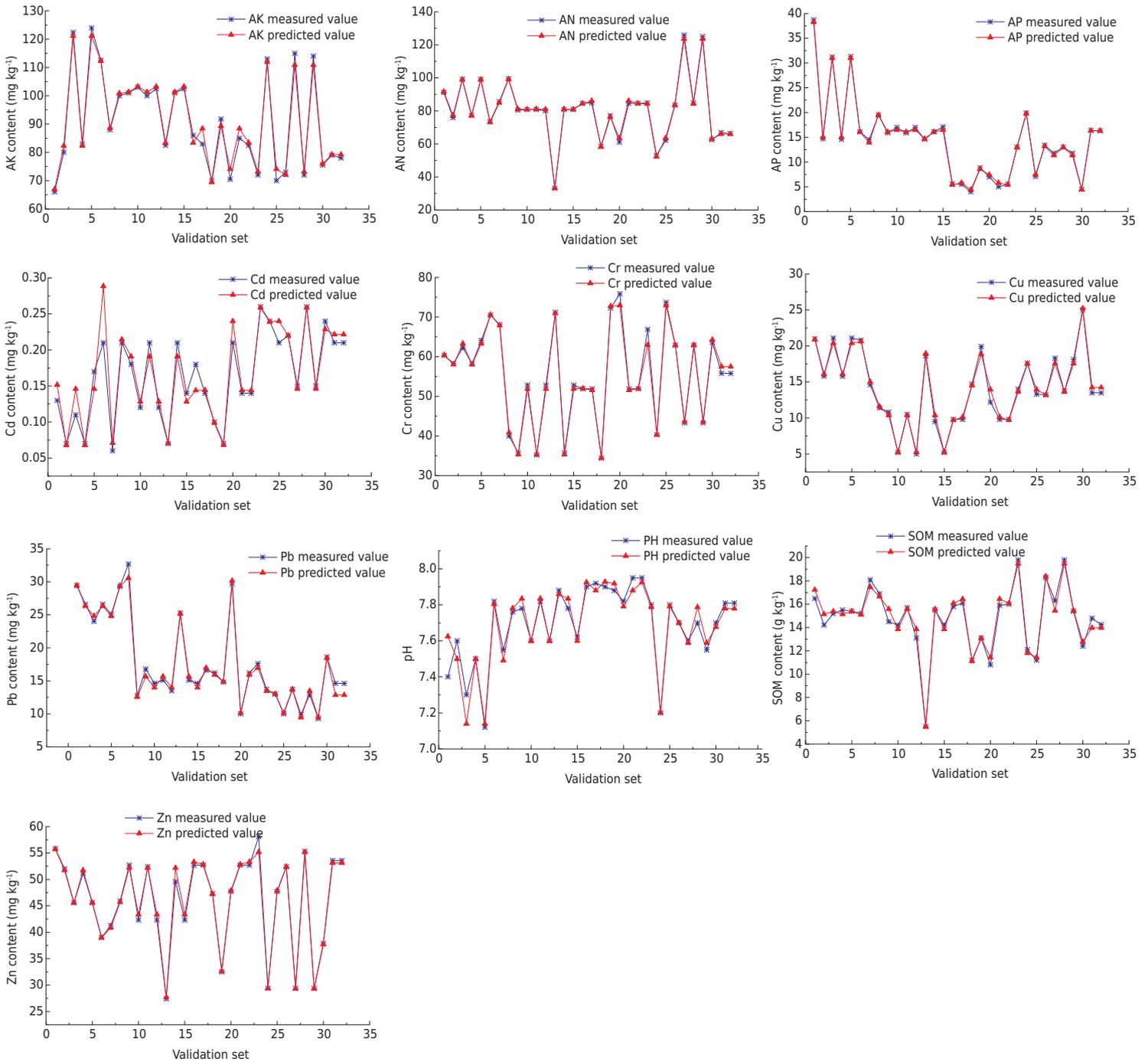


Figure 5. Inversion of soil property elements on the validation set.

By building hyperspectral inversion model of soil properties data comparing the calibration and validation set results, each soil property of calibration set has high determination coefficient, with \bar{R}^2 greater than 0.95. Also, the relative errors of the validation set are greater than 2.5. Except for a few samples have the difference from the measured values, most measured and predicted values of the samples are concentrated near $y=x$ (i.e., the 1:1 line), and the correlation coefficient r is significant at the $p=0.01$ level, indicating that the panel data model has good predictive ability when using a CR spectral transformation as the independent variable. The proposed method can thus be used for simultaneous inversion of multiple soil properties in high-standard farmland construction areas.

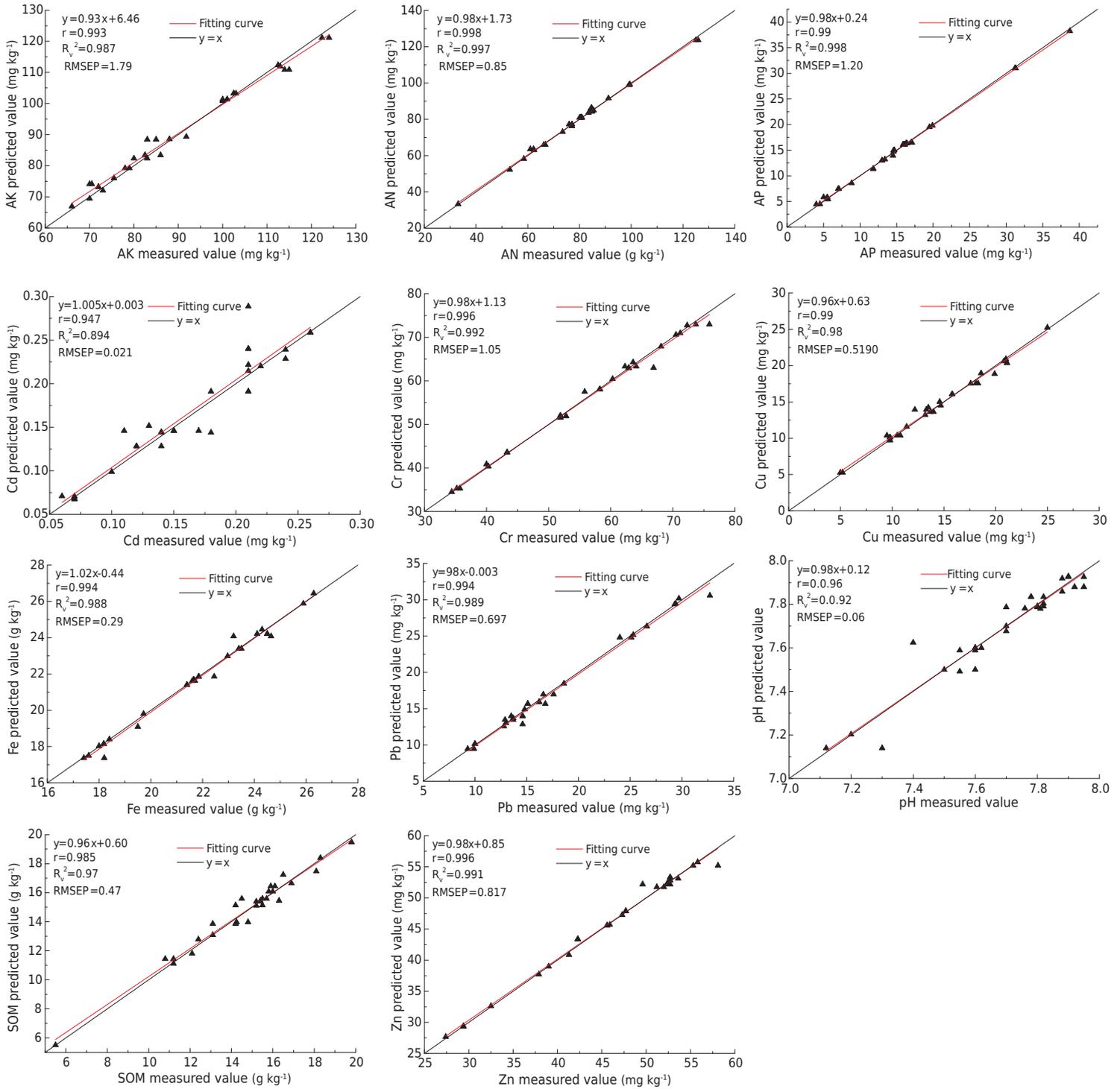


Figure 6. Inversion of soil property elements on the validation set.

Soil spectrum provides a comprehensive description of a soil sample's properties. High-standard farmland construction processes, soil organic matter and NPK nutrients, soil moisture, soil texture, oxides and other basic information are vital. Although this study uses a panel data model and concurrently solves the problem of various soil property data spectrum inversions in the high-standard farmland construction area of Xinzheng City, the universality of the proposed method still requires more verification. In the future, the applicability of this method should be tested in different regions or different soil types to

provide a data basis and technical support to achieve high-standard farmland construction. Due to the regional uniqueness of soil and the uncertainty of the field environment, whether indoor spectral data models can be applied to field and hyperspectral images should be a focus of future research.

CONCLUSION

In the construction of high-standard farmland, the soil pH, SOM, soil nutrient, and soil pollution characteristics are indispensable basic information for cultivated land quality evaluation. They are also important guarantees for the growing environment of crops. Therefore, the fast acquisition and real-time monitoring of basic soil information is the premise and foundation for constructing high-standard farmland.

The *SG* convolution smoothing of spectral reflectance can effectively remove noise, while preserving the overall characteristics of spectral curves, and improving signal-to-noise ratio. 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.

After the correlation analysis about soil properties 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 properties and spectral reflectance, select the common significance band of different soil properties as the best hyperspectral characteristic band and focus on 405~431nm, 781nm~831nm, 1044~1087nm, 1251~1410nm, 1836~1898nm, 2080nm~2201nm, 2324~2395nm.

Based on the panel data model, we constructed the comprehensive inversion model 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 relative analysis errors of soil properties 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 an independent variable have the ability of comprehensive inversion of soil properties, and have high precision prediction.

AUTHOR CONTRIBUTIONS

Conceptualization:  Liu Wenkai (equal) and  Zhang Qiuxia (equal).

Data curation:  Wang Xinsheng (equal) and  Zhang Qiuxia (equal).

Formal analysis:  Liu Wenkai (equal) and  Zhang Qiuxia (equal).

Funding acquisition:  Zhang Hebing (equal).

Investigation:  Wang Xinsheng (equal) and  Zhang Qiuxia (equal).

Methodology:  Zhang Qiuxia (equal).

Project administration:  Liu Wenkai (equal),  Zhang Hebing (equal) and  Ma Shouchen (equal).

Resources:  Zhang Hebing (equal) and  Zhang Qiuxia (equal).

Software:  Zhang Qiuxia (equal)

Supervision:  Liu Wenkai (equal) and  Zhang Hebing (equal)

Validation:  Liu Wenkai (equal)

Visualization:  Zhang Qiuxia (equal).

Writing - original draft:  Zhang Qiuxia (equal).

Writing - review & editing:  Zhang Hebing (equal),  Zhang Qiuxia (equal) and  Ma Shouchen (equal).

ACKNOWLEDGMENTS

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

DATA AVAILABILITY STATEMENT

The data presented in this study are available on request from the corresponding author.

REFERENCES

- Bian ZJ, Sun L, Tian K, Liu B, Zhang X, Mao Z, Huang B, Wu L. Estimation of heavy metals in tailings and soils using hyperspectral technology: A Case study in a tin-polymetallic mining area. *Bull Environ Contam Toxicol*. 2021;107:1022-31. <https://doi.org/10.1007/S00128-021-03311-7>
- Cai XW, Tang M, Song YN. On the construction potential of high standard basic farmland in Gongcheng Yao autonomous county. *Sci Technol Manag Land Res*. 2019;36:56-65. <https://doi.org/10.3969/j.issn.1009-4210.2019.05.006>
- Chen L, Wu KN, Feng Z, Yu B, Song HF. Suitability evaluation of high standard farmland construction from perspective of ecological civilization construction. *Soils*. 2019;51:803-12. <https://doi.org/10.13758/j.cnki.tr.2019.04.023>
- Chen S, Hu T, Luo L, He Q, Zhang S, Li M, Cui X, Li H. Rapid estimation of leaf nitrogen content in apple-trees based on canopy hyperspectral reflectance using multivariate methods. *Infrared Phys Techn*. 2020;111:103542. <https://doi.org/10.1016/j.infrared.2020.103542>
- Clark RN, Roush TL. Reflectance spectroscopy: Quantitative analysis techniques for remote sensing applications. *Journal of Geophysical Research: Solid Earth*. 1984;89:6329-40. <https://doi.org/10.1029/JB089iB07p06329>
- Dong F, Zhao W. Delimitation of construction area of high-standard primary farmland-taking Nan'an District of Chongqing City as an example. *Res Soil Water Conserv*. 2020;27:344-9. <https://doi.org/10.13869/j.cnki.rswc.2020.02.048>
- Li CM, Shao JA, Guo Y, Cao F, Tan SJ. Construction potential of high-standard farmland based on landform factors. *Chin J Eco-Agric*. 2018;26:1067-79. <https://doi.org/10.13930/j.cnki.cjea.171122>
- Li HX, Wang PZ. *Fuzzy mathematics*. Beijing: National Defense Industry Press; 1994.
- Li L, Wang ZL, Wu DF, Liu YY. Study on timing sequence and regulation mode of high standard basic farmland construction in village based on TOPSIS Model. *Res Soil Water Conserv*. 2020;27:286-93. <https://doi.org/10.13869/j.cnki.rswc.2020.03.041>
- Li SM, Li H, Sun DF, Zhou L, Bao J. Characteristic and diagnostic bands of heavy metals in Beijing agricultural soils based on spectroscopy. *Chin J Soil Sci*. 2011;42:730-5. <https://doi.org/10.1155/2011/531540>

- Li YY, Liao HP, Zhang YF, Zhang QQ. Construction conditions and development pattern of well-facilitated farmland in town area based on rural vitalization. *J Southwest University*. 2019;41:90-9. <https://doi.org/10.13718/j.cnki.xdzk.2019.02.013>
- Li ZY, Deng F, He J, Wei W. Hyperspectral estimation model of heavy metal arsenic in soil. *Spectrosc Spect Anal*. 2021;41:2872-8. [https://doi.org/10.3964/j.issn.1000-0539\(2021\)09-2872-07](https://doi.org/10.3964/j.issn.1000-0539(2021)09-2872-07)
- Liu Q, Lin H Z, Chen C. Application of the maximum tree algorithm for fuzzy clustering in the clustering of Web Page. *Appl Res Comput*. 2004;21:286-7. <https://doi.org/10.3390/rs12203312>
- Liu W, Zhao Z, Yuan HF, Song CF, Li XY. An optimal selection method of samples of calibration set and validation set for spectral multivariate analysis. *Spectrosc Spect Anal*. 2014;34:947-51. [https://doi.org/10.3964/j.issn.1000-0593\(2014\)04-0947-05](https://doi.org/10.3964/j.issn.1000-0593(2014)04-0947-05)
- Ma X, Shao JA, Cao F. Comprehensive Performance evaluation of high standard farmland construction in mountainous counties - A case study in Dianjiang Chongqing. *J Nat Res*. 2018;33:2183-99. <https://doi.org/10.31497/zrzyxb.20171185>
- Ministry of Agriculture of the People's Republic of China. NY/T 2148-2012, Criterion of High Standard Farmland. 2012.
- Ministry of Land and Resources of the People's Republic of China. TD/T1033-2012 Standard for well-facilitated capital farmland construction. 2012.
- Ministry of Land and Resources of the People's Republic of China. DZ/T 0279-2016, Regional Geochemical Sample Analysis Method. 2016.
- Rossel RAV, Taylor HJ, Mcbratney AB. Multivariate calibration of hyperspectral γ -ray energy spectra for proximal soil sensing. *Eur J Soil Sci*. 2007;58:343-53. <https://doi.org/10.1111/j.1365-2389.2006.00859.x>
- Savitzky A, Golay MJE. Smoothing and differentiation of data by simplified least squares procedures. *Anal Chem*. 1964;36:1627-39. <https://doi.org/10.1021/ac60214a047>
- Schreiner S, Culibrk D, Bandecchi M, Gross W, Middelmann W. Soil monitoring for precision farming using hyperspectral remote sensing and soil sensors. *At-Autom*. 2021;69:325-35. <https://doi.org/10.1515/auto-2020-0042>
- Selvakumar A, Balasundaram A. Machine learning based classification of diseased mango leaves. *Int J Recent Innov Trends Comput Commun*. 2022;10:38-44. <https://doi.org/10.17762/ijritcc.v10i7.5563>
- Shi YY, Zhao J, Song X, Qin Z, Wu L, Wang H, Tang J. Hyperspectral band selection and modeling of soil organic matter content in a forest using the Ranger algorithm. *PLoS One*. 2021;16:e0253385. <https://doi.org/10.1371/journal.pone.0253385>
- Sun JS. *Econometrics learning Guide and EVIEWS - Application Guide*. Beijing: Tsinghua University Press; 2010.
- Tang F, Xu L, Zhang P, Zhang G, Fu M, Zhang J. Suitability evaluation and priority area delineation of high standard basic farmland construction at county level. *Trans Chin Soc Agric Eng*. 2019;35:242-51. <https://doi.org/10.11975/j.issn.1002-6819.2019.21.030>
- Tong QX, Zhang B, Zheng LF. *Hyperspectral remote sensing: Principle, technology and application*. Beijing: Higher Education Press; 2006. <https://doi.org/10.11873/j.issn.1004-0323.2015.6.1095>
- Wang K, Li L, Li P. Study on high standard farmland construction based on ecological security and food security. *J Ecol Rural Environ*. 2021;37:706-13. <https://doi.org/10.19741/j.issn.1673-4831.2020.0781>
- Wang PZ. *Fuzzy set theory and its application*. Shanghai: Shanghai Science and Technology Publishing House; 1983.
- Wang XQ, Shi WJ, Sun XF, Wang M. Comprehensive benefit evaluation and regional differences of construction projects of well-facilitated farmland in Huang Huai Hai region. *Trans Chin Soc Agric Eng*. 2018;34:238-48. <https://doi.org/10.11975/j.issn.1002-6819.2018.16.031>

- Wei L, Pu H, Wang Z, Yuan Z, Yan X, Cao L. Estimation of soil arsenic content with hyperspectral remote sensing. *Sensors*. 2020;20:4056. <https://doi.org/10.3390/s20144056>
- Xiong YF, He J. Preliminary Study on Quality Control of High Standard Farmland Construction Projects. *Acad J Bus Manag*. 2019;1:7-16. <https://doi.org/10.25236/AJBM.190102>
- Xu X, Chen S, Xu Z, Yu Y, Zhang S, Dai R. Exploring appropriate preprocessing techniques for hyperspectral soil organic matter content estimation in black soil area. *Remote Sens*. 2020;12:3765. <https://doi.org/10.3390/rs12223765>
- Yin F, Wu M, Liu L, Zhu Y, Feng J, Yin D, Yin C. Predicting the abundance of copper in soil using reflectance spectroscopy and GF5 hyperspectral imagery. *Int J Appl Earth Obs*. 2021;102:102420. <https://doi.org/10.1016/j.jag.2021.102420>
- Yu H, Liu J, Liu YQ, Gu ZP, Wang AL. Estimation of soil organic matter content of cultivated land in Hilly Regions with hyperspectral. *J Shandong Agric Univ*. 2021;52:648-53. <https://doi.org/10.3969/j.issn.1000-2324.2021.04.021>
- Zeng Y, Zhao W, Yang LJ. Research on time sequence of high-standard basic farmland construction in peri-urban areas: a case study of Nan'an district in Chongqing. *Res Environ Yangtze Basin*. 2020;29:434-41. <https://doi.org/10.11870 / cjlyzyyhj202002016>
- Zhang HB, Zhao SX, Chen NL Zhang QX. Project optimization for well-facilitated farmland construction based on coupling coordination degree. *Trans Chin Soc Agric Mach*. 2018;49:161-8. <https://doi.org/10.6041 / j. issn.1000-1298. 2018.08. 019>
- Zhang J, Sun L. High standard farmland construction and soil conservation evaluation. *IOP Conf Ser: Earth Environ Sci*. 2020;514:022074. <https://doi.org/10.1088/1755-1315/514/2/022074>
- Zhang Q, Zhang H, Zhang H, Wang XS, Liu WK. Hybrid inversion model of heavy metals with hyperspectral reflectance in cultivated soils of main grain producing areas. *Trans Chin Soc Agric Mach*. 2017;48:148-55. <https://doi.org/10.6041/j.issn.1000-1298.2017.03.019>