The GA-BPNN-Based Evaluation of Cultivated Land Quality in the PSR Framework Using Gaofen-1 Satellite Data

Shanshan Liu [1], Yiping Peng [1] (iD), Ziqing Xia [1], Yueming Hu [1,2,3,4,*], Guangxing Wang [1,5] (iD), A-Xing Zhu [1,6] (iD) and Zhenhua Liu [1,*]

[1] College of Natural Resources and Environment, South China Agricultural University, Guangzhou 510642, China; Shanl633@163.com (S.L.); pyp_scau@163.com (Y.P.); xzq@stu.scau.edu.cn (Z.X.); gxwang@siu.edu (G.W.); azhu@wisc.edu (A.-X.Z.)

[2] Guangdong Provincial Key Laboratory of Land Use and Consolidation, South China Agricultural University, Guangzhou 510642, China

[3] Guangdong Province Engineering Research Center for Land Information Technology, South China Agricultural University, Guangzhou 510642, China

[4] Key Laboratory of Construction Land Transformation, Ministry of Land and Resources, South China Agricultural University, Guangzhou 510642, China

[5] Department of Geography and Environmental Resources, College of Liberal Arts, Southern Illinois University Carbondale (SIUC), Carbondale, IL 62901, USA

[6] Department of Geography, University of Wisconsin-Madison, Madison, WI 53706, USA

* Correspondence: ymhu@scau.edu.cn (Y.H.); zhenhua@scau.edu.cn (Z.L.); Tel.: +86-020-8528-8307 (Y.H.); +86-135-3943-3693 (Z.L.)

Abstract: Rapid and efficient assessment of cultivated land quality (CLQ) using remote sensing technology is of great significance for protecting cultivated land. However, it is difficult to obtain accurate CLQ estimates using the current satellite-driven approaches in the pressure-state-response (PSR) framework, owing to the limitations of linear models and CLQ spectral indices. In order to improve the estimation accuracy of CLQ, this study used four evaluation models (the traditional linear model; partial least squares regression, PLSR; back propagation neural network, BPNN; and BPNN with genetic algorithm optimization, GA-BPNN) to evaluate CLQ for determining the accurate evaluation model. In addition, the optimal satellite-derived indicator in the land state index was selected among five vegetation indices (the normalized vegetation index, NDVI; enhanced vegetation index, EVI; modified soil-adjusted vegetation index, MSAVI; perpendicular vegetation index, PVI; and soil-adjusted vegetation index, SAVI) to improve the prediction accuracy of CLQ. This study was conducted in Conghua District of Guangzhou, Guangdong Province, China, based on Gaofen-1 (GF-1) data. The prediction accuracies from the traditional linear model, PLSR, BPNN, and GA-BPNN were compared using observations. The results demonstrated that (1) compared with other models (the traditional linear model: $R^2 = 0.14$ and RMSE = 91.53; PLSR: $R^2 = 0.33$ and RMSE = 74.58; BPNN: $R^2 = 0.50$ and RMSE = 61.75), the GA-BPNN model based on EVI in the land state index provided the most accurate estimates of CLQ, with the R^2 of 0.59 and root mean square error (RMSE) of 56.87, indicating a nonlinear relationship between CLQ and the prediction indicator; and (2) the GA-BPNN-based evaluation approach of CLQ in the PSR framework was driven to map CLQ of the study area using the GF-1 data, leading to an RMSE of 61.44 at the regional scale, implying that the GA-BPNN-based evaluation approach has the potential to map CLQ over large areas. This study provides an important reference for the high-accuracy prediction of CLQ based on remote sensing technology.

1. Introduction

Cultivated land is the most basic capital good of agricultural production [1,2]. The area of cultivated land in China covers 134.9×10^6 hm^2, accounting for 19.71% of the land area. In recent years, the rapid urbanization and industrialization in China have resulted in a strong demand for evaluating cultivated land quality (CLQ) to ensure national food security. The evaluation of CLQ represents natural conditions and the degree of anthropogenic use of cultivated land resources, involving the quantification and grading of CLQ indices [3–5]. Therefore, it is very important to quickly and effectively evaluate CLQ. However, traditional estimations of CLQ are based on field measurements, which are time-consuming and costly [6]. Traditional methods are also unable to output high precision spatial distributions of CLQ. Remote sensing technology provides a unique means for the rapid evaluation of CLQ in a time-efficient and low-cost manner. Substantial research has been conducted for the evaluation of CLQ based on remote sensing data [7–10].

The current research on remote sensing-based CLQ evaluation methods was divided into two categories: CLQ quasi-quantity evaluation and CLQ quantity evaluation. In CLQ quasi-quantity evaluation, remote sensing data were used for the inversion of traditional some indicators (e.g., soil fertility). For example, Yu (2012) constructed an evaluation indicator system with the normalized vegetation index (NDVI) extracted from MODIS, soil organic matter (SOM), and soil mass and trace elements content obtained from field sampling to evaluate CLQ [11]. Yang et al. (2012) used Landsat TM5 to invert soil texture and SOM, and combined these with traditional indicators (barrier depth and profile configuration) to evaluate CLQ [12]. However, the CLQ quasi-quantity evaluation method is not able to increase the evaluation efficiency owing to the limitations of the traditional field measurements.

With the development of the remote sensing technique, we get CLQ quantity evaluation, whereby remote sensing data were used to evaluate CLQ based on the pressure-state-response (PSR) framework in terms of resistance to pressure, current state, and land use response [7,13]. For example, Fang et al. (2008) used Slope, ratio vegetation index (RVI), NDVI, DVI, and land use degree (LUD) to establish an evaluation model for CLQ, while Liu et al. (2010) used slope, sandy area ratio in a pixel (SARP), modified soil-adjusted vegetation index (MSAVI), soil and vegetation moisture index (SVMI), and LUD to establish a linear evaluation model of CLQ [8,9]. Although these studies have demonstrated rapid CLQ evaluation processes, they estimate and map CLQ based on the PSR framework using the remote sensing technique, in which linear relationships of CLQ with spectral indices are often assumed, and there is a lack of optimizing selection of spectral indices. Thus, it is difficult to obtain accurate CLQ estimates owing to the limitations of linear models and CLQ spectral indices.

The objective of this study was to improve the evaluation accuracy of the satellite-driven approach in the PSR framework by determining the optimal evaluation models among the traditional linear model, partial least squares regression (PLSR), back propagation neural network (BPNN), and BPNN with genetic algorithm (GA-BPNN), and selecting the evaluation indicator in the land state index by analyzing the correlation coefficient between five vegetation indices (VIs) derived from Gaofen-1 (GF-1) data and soil sample fertility (SOM; and total nitrogen, TN). This study was validated using GF-1 data in the Conghua District of Guangzhou, Guangdong Province, China. It is expected that the modifications can increase the accuracy of CLQ estimates.

2. Materials and Methods

2.1. Study Area

This study area is Conghua District of Guangzhou City, China (113 17′–114 04′ E, 23 22′–23 56′ N), located in the transition zone from the Pearl River Delta to the northern mountain area of Guangzhou.

This is a typical hilly and mountainous area and is dominated by the south subtropical monsoon climate, with an annual average temperature of 19.5–21.2 °C and an annual average precipitation of 2176.3 mm. The area of cultivated land in Guangzhou covers 13,485.99 hm^2, of which the paddy field was 11,885.16 hm^2, accounting for 87.91% of the total area. As the main crop, rice can be planted in two seasons.

2.2. Sample Data

A total of 2000 samples within Conghua District were collected at a plot-scale sampling grid according to a stratified sampling method (Figure 1c). CLQ data from the 2000 samples were obtained from the CLQ map for 2016 in the Conghua National Land Department. Further, the CLQ data calculated using gradation regulations on agriculture land quality in China (GB/T 28407-2012) were considered as standard reference values of CLQ. Out of the 2000 samples, 1500 samples in black were used to the establish evaluation models, 250 sample plots in red were employed to validate the accuracy of the estimated CLQ, and another dataset of 250 sample plots in green was used to assess the accuracy of mapping CLQ at the regional scale.
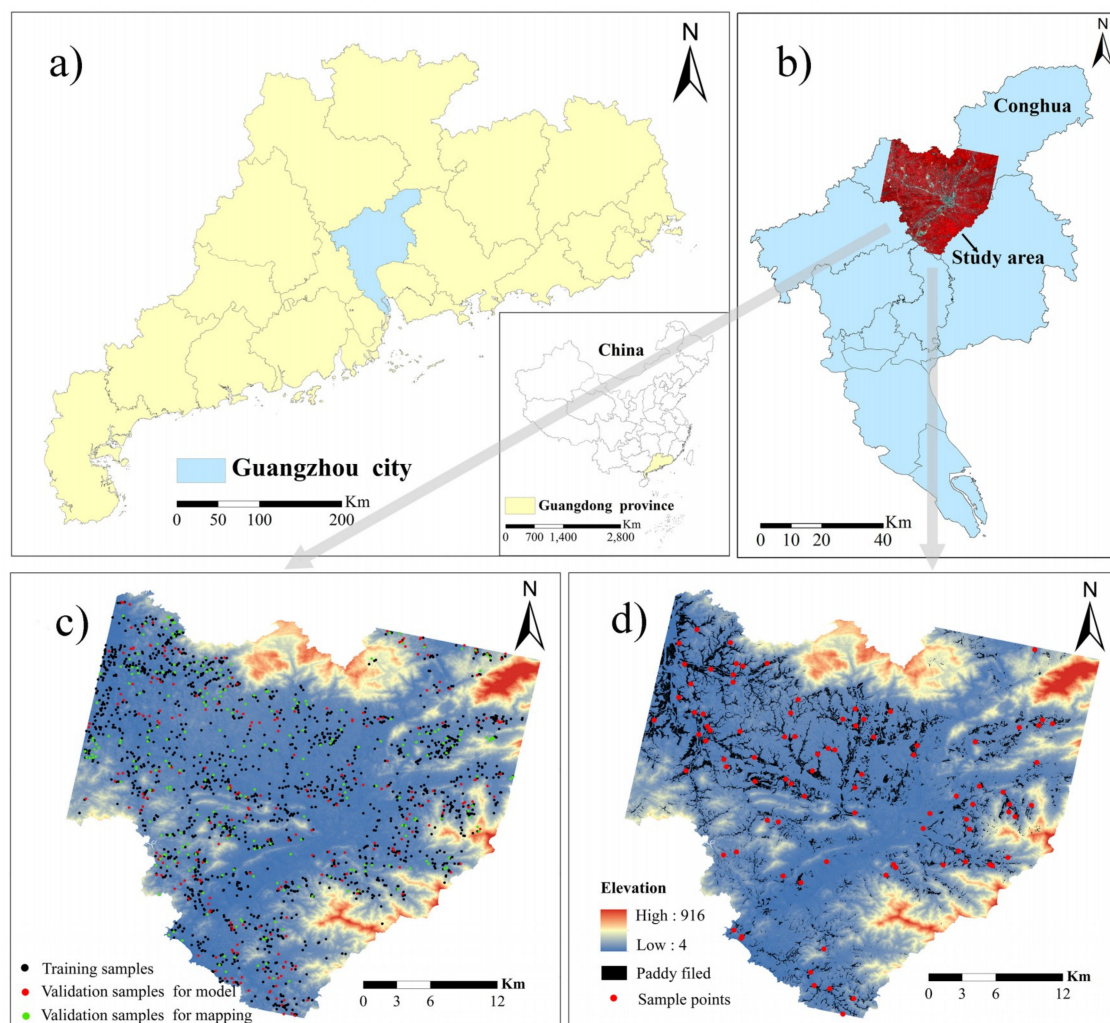

Figure 1. Study area and sampling distribution: (**a**) the study area location in Guangzhou City; (**b**) the standard pseudo-color map of the study area in Conghua District; (**c**) the spatial distribution of 2000 cultivated land quality (CLQ) samples (the training sample plots in black, the validation sample plots for model in red, and the validation sample plots for mapping in green); and (**d**) soil samples for selection of soil fertility indicator.

In order to determine the optimal soil fertility indicator, the layout of 89 soil samples was extracted from the 2000 samples based on soil type and land-use (Figure 1d). The 89 soil samples (at 0–20 cm depth) were collected via GPS positioning in May 2016. Within each of the sample plots, soil samples were collected at five points. Each soil sample was composed by mixing subsamples from five points. One of the five points was located at the plot center and the remaining four points were allocated along the diagonal lines of the plot, with an equal distance between the points. The samples were air-dried naturally in the laboratory, and gravel and animal and plant residues were removed. The soil fertility parameters of SOM and TN contents were analyzed. SOM content was determined by the potassium dichromate oxidation-external heating method, and TN was determined using the phenol disulfonic acid colorimetric method [14].

2.3. Satellite Image Data and Preprocessing

A GF-1 multi-spectral image of October 2016 with a spatial resolution of 8 m was obtained from China Center for Resources Satellite Data and Application (http://218.247.138.119:7777/DSSPlatform/index.html). A Landsat-8 thermal infrared sensor (TIRS) image and a digital elevation model (DEM) from October 2016 with a spatial resolution of 30 m were obtained from the Geospatial Data Cloud of the Computer Network Information Center, Chinese Academy of Sciences (http://www.gscloud.cn).

The radiance calibration of the GF-1 and Landsat-8 TRIS images was to convert the *DN* value of the raw image to surface spectral reflectance, which was performed based on Equation (1), expressed as follows:

$$L = Gain \times DN + Bias, \tag{1}$$

where L is the radiance, and *Gain* and *Bias* are the calibration coefficients. The values of the radiance calibration parameters are reported in Table 1. The GF-1 and Landsat-8 TRIS data were conducted using the FLAASH model. Geometric corrections and resampling were performed for the DEM and Landsat 8 TRIS imagery using ENVI 5.3. The geometric precision correction was performed on them using a quadratic polynomial calculation model. The calibration error was within 0.5 pixels.

Table 1. Radiance calibration parameter values of the Gaofen-1 (GF-1) satellite and Landsat-8 thermal infrared sensor (TIRS) images.

Satellite	Parameter Value	Bands				
		Band 1	Band 2	Band 3	Band 4	Band 6
GF-1	Gain	0.2072	0.1776	0.1770	0.1909	
	Bias	7.5348	3.9395	−1.7445	−7.2053	
Landsat-8 TRIS	Gain	1.1807	1.2098	0.9425	0.9692	17.04
	Bias	−7.3800	−7.6100	−5.9400	−6.0700	12.65

2.4. Selecting Satellite-Derived Indicators for CLQ Evaluation

CLQ evaluation indicators were obtained from the land pressure resistance index (PRI), land state index (LSI), and land use response index (LURI) based on the PSR proposed by Dumanski and Pieri [15], shown in Table 2.

Table 2. Cultivated land quality (CLQ) evaluation indicator system. TVDI, temperature vegetation drought index; VI, vegetation index; RA, road accessibility; PFD, patch fractal dimension.

Target Layer	Project Layer	Satellite-Derived Indicator Layer
	Pressure Resistance Index (PRI)	Slope
CLQ evaluation indicator system	Land State Index (LSI)	TVDI
		VIs
	Land Use Response Index (LURI)	RA
		PFD

At the satellite-derived indicator layer, each indicator was selected based on the project layer (PRI, LSI, and LURI). The PRI represents the ability to withstand environmental pressure and the probability of degradation. As PRI increases, the ability to withstand environmental pressure improves and the degradation probability falls. The environmental pressure within the study area mainly originates from soil erosion. Therefore, slope was selected as the PRI evaluation indicator (Table 2). In addition, LSI denotes the current quality and potential productivity of cultivated land, reflected by soil moisture and soil fertility. Furthermore, soil organic matter (SOM) and total nitrogen (TN) are the major determinants and indicators of soil fertility [16,17]. The temperature vegetation drought index (TVDI) is related to surface soil moisture owing to changes in thermal inertia and evaporative control (evaporation and transpiration) on net radiation partitioning (energy balance) [18]. Thus, soil moisture was replaced by TVDI in this study (Table 2), expressed as follows [19]:

$$TVDI = \frac{LST - LST_{min}}{a + bNDVI - LST_{min}},$$

(2)

where LST_{min} is the minimum surface temperature, defining the wet edge; LST is the observed surface temperature at the given pixel; NDVI is the observed normalized difference vegetation index; and a and b are parameters defining the dry edge, which is modeled as a linear fit to the data ($LST_{max} = a + b \times NDVI$), where LST_{max} is the maximum surface temperature observation for a given NDVI. More specifically, a = 48.80 and b = −33.15. TVDI is lower for wet and higher for dry conditions, with values of 1 at the dry edge and 0 at the wet edge [19].

Soil fertility (SOM and TN) is reflective of vegetation growing conditions [8,20]. Thus, remote sensing VIs (Tables 2 and 3) were selected as soil fertility indicators in the study. In order to improve the accuracy of the CLQ evaluation, the Pearson correlation was used to select the optimal satellite-derived VI as soil fertility indicators. In particular, those with the greatest correlation coefficients between soil fertility and satellite-derived VIs under the significance level of $p \leq 0.05$ were chosen. The Pearson correlation coefficient was derived, expressed as follows:

$$r_i = \frac{\sum_{n=1}^{N}\left(R_{ni} - \overline{R_i}\right)(y - \overline{y})}{\sqrt{\sum_{n=1}^{N}\left(R_{ni} - \overline{R_i}\right)^2 \sum_{n=1}^{N}(y - \overline{y})^2}},$$

(3)

where r_i is the correlation coefficient between soil fertility and a satellite-derived VI, R_{ni} is the satellite-derived VI of the nth soil sample, $\overline{R_i}$ is the mean value of the VI in the ith soil sample, y is the nth soil fertility, and $\overline{y}$ is the average value of soil fertility.

Table 3. Equations for the five VIs used in this study. NDVI, normalized vegetation index, EVI, enhanced vegetation index; SAVI, soil-adjusted vegetation index; MSAVI, modified SAVI; PVI, perpendicular vegetation index.

VIs	Algorithm Formula	Reference
NDVI	$(R_{NIR} - R_{RED})/(R_{NIR} + R_{RED})$	[21,22]
EVI	$2.5 \times (R_{NIR} - R_{RED})/(R_{NIR} + 6R_{NIR} - 7.5R_{BLUE} + 1)$	[23]
MSAVI	$\left[2R_{NIR} + 1 - \sqrt{(2R_{NIR} + 1)^2 - 8(R_{NIR} - R_{RED})}\right]/2$	[24]
SAVI	$1.5 \times (R_{NIR} - R_{RED})/(R_{NIR} + R_{RED} + 0.5)$	[25]
PVI	$(R_{NIR} - aR_{RED} - b)/\sqrt{1 + a^2}$ (a=0.9, b=0.1)	[26,27]

R_{NIR}, R_{RED}, and R_{BLUE} are the spectral reflectance of near-infrared, red, and blue bands respectively.

In the study area, land use embodied mainly in the form of land use types, road accessibility (RA), and land patch fragmentation are the farmers' response to CLQ [28–31]. As paddy fields are the only

type of cultivated land in this study, RA (Table 2) and land patch fragmentation were selected as LURIs indicators. RA was calculated according to Equation (4):

$$\text{RA} = 1 - \frac{d_i}{d}, \tag{4}$$

where d_i is the proximity distance between the ith land patch and road and d is the maximum distance among all land patches and the road. RA ϵ (0, 1), where larger RA values indicate higher plot accessibilities. The land patch fragmentation degree is reflected by the patch fractal dimension (PFD) (Table 2), expressed as follows:

$$\text{PFD} = 2 \times \ln\left(\frac{L}{4}\right) / \ln A, \tag{5}$$

where L (m) is the perimeter of the land patch and A (m^2) is the area of land patch. PFD ϵ (1, 2), with values approaching 1 indicate a greater degree of cultivated land utilization by farmers.

2.5. Modeling and Mapping Methods

In order to improve the CLQ evaluation accuracy, linear (PLSR) and nonlinear models (BPNN and GA-BPNN) were applied for CLQ predictions using satellite-derived indicators.

2.5.1. Linear Model

The traditional linear model [7–9] was constructed taking measured CLQ in 1500 training samples as the dependent variable and the corresponding evaluation indicator as independent variables.

$$Y = \sum_{i=1}^{n} \beta_i \check{X}_i, \tag{6}$$

where Y is the dependent variable (CLQ value), $\check{X}_i$ is the ith independent variable (evaluation indicator value), and β_i is the ith weight coefficient determined from expert experience. In this study, a score within the range of [0, 3000] was assigned to CLQ with reference to the GB/T 28407-2012. Therefore, all indicators or indices were normalized to a range of 0–3000.

The PLSR is a standard multivariate statistical technique that was originally developed by Herman Wold in 1966 [32]. Previous studies [33] indicate that the PLSR is highly effective in different disciplines because it allows for the analysis of data with strong correlations in the predictor variables. The PLSR used in the study is expressed as follows:

$$Y = X\beta + \varepsilon, \tag{7}$$

where Y is the dependent variable (CLQ value), X is the independent variable (evaluation indicator value), β is the coefficient matrix, and ε is the residual matrix.

2.5.2. Back Propagation Neural Network

BPNN is a multi-layer feed forward network trained by the error back propagation algorithm, and is suitable for various types of nonlinear relationship analysis methods. It consists of an input layer, an output layer, and several hidden layers [34,35]. *Trainlm* and *Purelin* were selected as the training function and the transfer function of the output layer in the BPNN, respectively. The steepest descent method and the back-propagation algorithm in the BPNN model were used to repeatedly adjust the weight and deviation of the network until the actual value and the expected output were as close as possible [36,37].

The hidden layer information acquired from the input layer was expressed as follows:

$$o_j = f_i\left(\sum \omega_{ji} o_i + \theta_j\right), \tag{8}$$

where o_i is the input layer information (the evaluation indicator); o_j is the hidden layer information; ω_{ji} is the weight of the input layer to the hidden layer; and f_i is the transfer function of the input layer to the hidden layer. In this study, the Trainlm function is chosen and θ_j is the hidden layer threshold.

The hidden layer was transferred to the output layer, expressed as follows:

$$o_k = f_j\left(\sum \omega_{kj} o_j + \theta_k\right), \tag{9}$$

where o_k is the output layer information (CLQ), f_j is the transfer function of the hidden layer to the output layer (the Purelin function was used in this study), ω_{kj} represents the weight of the hidden layer to the output layer, and θ_k is the output layer threshold.

The number of neurons in the hidden layer was determined according to the empirical formula in Equation (9):

$$n_h = 2n_i + 1, \tag{10}$$

where n_h is the number of hidden layer units and n_i is the number of input units.

If the estimated value differs greatly from the measured value, it is transferred to the error of the back propagation process. The process of reverse propagation uses the Levenberg–Marquardt algorithm from the output layer to the input layer to modify the connection weight in order to reduce the mean square error (MSE), expressed as follows:

$$\text{MSE} = \frac{1}{N} \sum (o - o_k)^2, \tag{11}$$

where o is the measured CLQ, o_k is the estimated CLQ, and N is the number of training samples.

2.5.3. Genetic Algorithm-Back Propagation Neural Network

In order to improve the prediction accuracy of the BP algorithm, GA was introduced to the BP model to optimize the weight and threshold selection of the neural network. GA has the advantages of only requiring fitting information and not tending to a local solution [38,39]. Thus, the combined GA-BPNN model was used to estimate CLQ in this study. The original BPNN weight and threshold were converted into chromosomes in GA using real-number coding. The code length was calculated using Equation (11):

$$S = i * j + j * k + j + k, \tag{12}$$

where i is the number of input layer neuron nodes, which is the number of evaluation indicators; j is the number of neurons in the hidden layer; and k is the number of neurons in the output layer. Note that, in this case, the output layer only exhibited CLQ, hence $k = 1$. This was followed by the generation of a random population of chromosomes. The BPNN was used to obtain the individual fitness value (E), which is the sum of the absolute error between the estimated and measured values of the training data, expressed as follows:

$$min\ \text{E} = \sum abs(y_k - o_k), \tag{13}$$

where y_k is the measured value of CLQ in the kth land patch and o_k is the estimated CLQ value in the kth land patch.

3. Results

3.1. Satellite-Derived Indicators of CLQ Evaluation

The satellite-derived indicators in the three project layers were calculated in the PSR framework. The PRI slope was determined using the ArcGIS 10.2 surface analysis model, based on the DEM data (Figure 2a). Next, TVDI, acting as a satellite indicator for LSI, was derived according to Equation (2). The spatial distribution of TVDI is shown in Figure 2b. In order to acquire highly-accurate CLQ

estimates, the optimal VI was selected among five VIs (NDVI, EVI, MSAVI, SAVI, and PVI) as the soil fertility indicator. In particular, the VIs with the greatest correlation coefficients with soil fertility ($p \leq 0.05$) were chosen, as shown in Table 4. It can be seen that EVI exhibited the greatest correlation coefficients with SOM and TN (0.88 and 0.90, respectively), indicating EVI as the optimal indicator for the evaluation of CLQ. The spatial distribution of EVI is presented in Figure 2c.

Table 4. Correlation coefficients between the VIs and soil fertility parameters derived from the soil samples. SOM, soil organic matter; TN, total nitrogen.

Soil Fertility Parameters	NDVI	EVI	MSAVI	SAVI	PVI
SOM (%)	0.82 **	0.88 **	0.87 **	0.84 **	0.85 **
TN (mg/kg)	0.75 **	0.90 **	0.88 **	0.78 **	0.79 **

** correlation is significant at $p < 0.01$ level.

Finally, RA and PFD, calculated according to Equations (4) and (5), respectively, and their spatial distributions are given in Figure 2d,e.
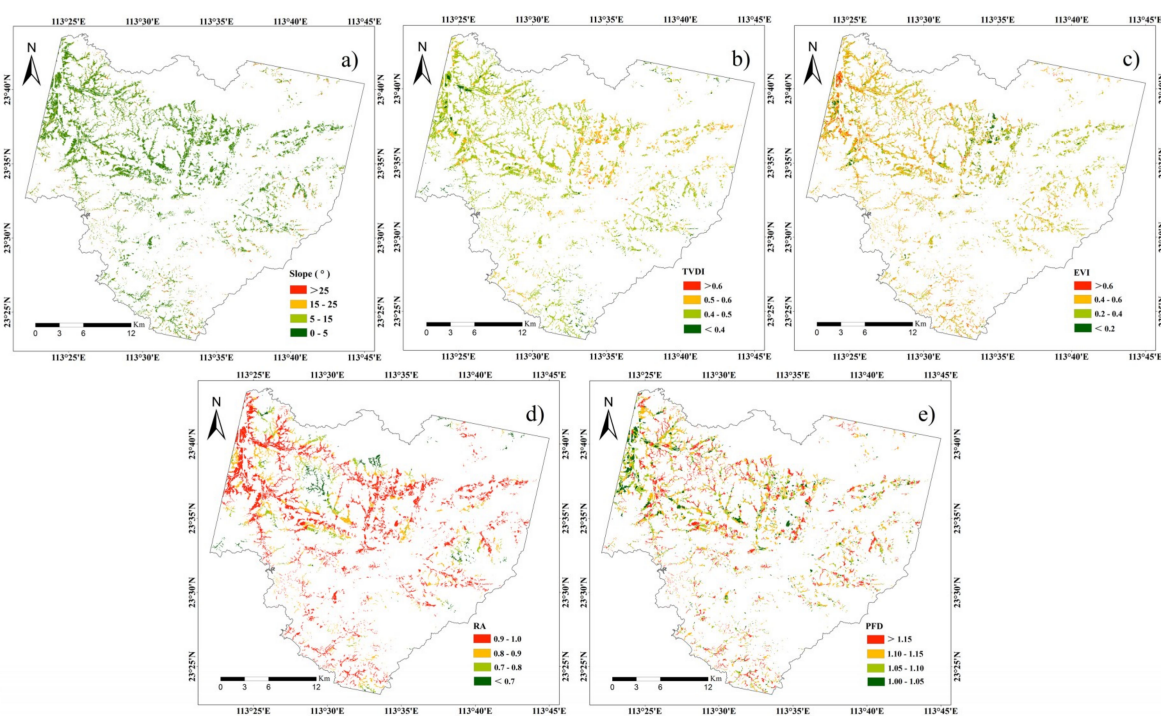

Figure 2. Spatial distributions of satellite-derived indicators in the pressure-state-response (PSR) framework: (**a**) slope; (**b**) temperature vegetation drought index (TVDI); (**c**) enhanced vegetation index (EVI); (**d**) road accessibility (RA); and (**e**) patch fractal dimension (PFD).

From the spatial distribution of spectral indicators, the slope, PFD, and TVDI with low values mainly located in the northwest part of the study area, and EVI and RA with high values also located in this part. This indicates that, the larger the EVI and RA, the smaller the slope, PFD, and TVDI, which corresponds to the factual conditions.

3.2. The Optimal Model for CLQ Evaluation

The selected evaluation indicators (slope, TVDI, EVI, RA, and PFD) were used as independent variables and CLQ acted as the dependent variable. The CLQ predictions of the traditional linear, PLSR, BPNN, and GA-BPNN models were compared. In order to evaluate the accuracy of the prediction models for estimating CLQ, the coefficient of determination (R^2), root mean square error (RMSE)

between the estimated and observed values, and normalized root mean square error (NRMSE) [13] were calculated based on the training and validation datasets. The obtained traditional linear and PLSR evaluation model are expressed as follows:

$$\hat{Y} = \frac{1}{3} \times \widetilde{Slope} + \frac{1}{3} \times \left(\frac{\widetilde{EVI} + \widetilde{TVDI}}{2} \right) + \frac{1}{3} \times \left(\frac{\widetilde{RA} + \widetilde{PFD}}{2} \right) \quad \left(R^2 = 0.14, P = 0 \right), \tag{14}$$

$$\hat{Y} = 1480.15 - 2.47 \times slope + 569.68 \times EVI - 296.41 \times TVDI + 106.70 \times RA - 79.28 \times PFD$$
$$\left(R^2 = 0.33, P = 0 \right). \tag{15}$$

Moreover, a three-layer BPNN with a single hidden layer was used for the CLQ predictions. The number of neuron nodes in the hidden layer was fixed at 13, the number of iterations at 1000, and both the learning rate and learning objective at 0.01. In order to compare the results of the GA optimization, the network structure and parameter configuration were consistent with those of the BPNN. In the GA, the number of the maximum runs was set at 50, and the population size, crossover probability (P_c), and mutation probability (P_m) were fixed as 128, 0.9, and 0.02, respectively. A total of 1500 training samples were selected for the PLSR, BPNN, and GA-BPNN models in order to train the response relationships between CLQ and the selected evaluation indicators (Figure 3).
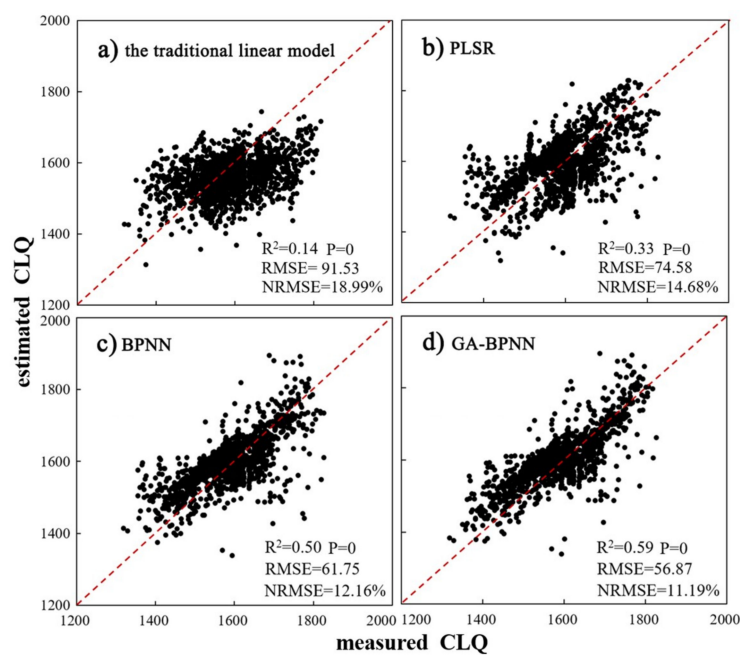

Figure 3. Scatterplots of measured versus estimated CLQ obtained by four models using the training dataset: (**a**) the traditional linear model; (**b**) partial least squares regression (PLSR); (**c**) back propagation neural network (BPNN), and (**d**) BPNN with genetic algorithm optimization (GA-BPNN). RMSE, root mean square error; NRMSE, normalized RMSE.

From the results using the training dataset, the explanatory power of CLQ was observed to vary greatly depending on the evaluation model (R^2: 0.14 for traditional linear model, 0.33 for PLSR, 0.50 for BPNN, and 0.59 for GA-BPNN). At a given CLQ, the traditional linear model estimations exhibited the greatest RSME values, and GA-BPNN exhibited the lowest. The PLSR and traditional linear models performed the worst, while the BPNN and GA-BPNN models were able to improve the overall CLQ prediction accuracy. In particular, the GA-BPNN model exhibited the highest accuracy and the strongest prediction ability, implying that there existed a significant nonlinear relationship between the evaluation indicators and CLQ.

Moreover, 250 samples were used to validate the estimation accuracy of the four models, with the results presented in Figure 4. For the validation samples, the estimated and measured values of CLQ were observed to be randomly distributed on both sides of the 1:1 line, depending on the model. More specifically, for the GA-BPNN model, the scatter points were closely located to the 1:1 line, with limited outliers. Thus, the results demonstrate that the indicators selected in this study are suitable for predicting CLQ in the study area.
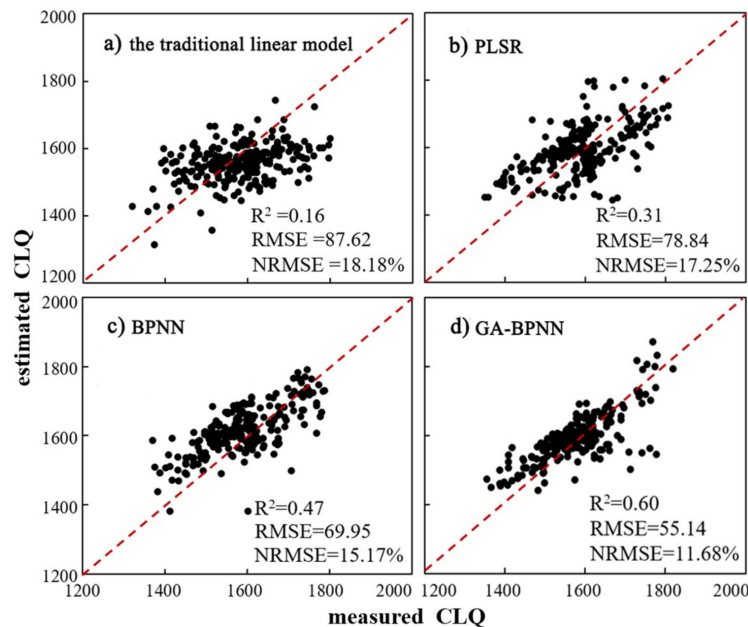
Figure 4. Scatterplots of measured versus estimated values of CLQ obtained four models using the 250-sample validation dataset for model: (**a**) the traditional linear model; (**b**) PLSR; (**c**) BPNN; and (**d**) GA-BPNN.

Furthermore, for the validation dataset, the observed prediction accuracies of the four models from high to low are GA-BPNN ($R^2 = 0.60$, RMSE = 55.41) > BPNN ($R^2 = 0.47$, RMSE = 69.95) > PLSR ($R^2 = 0.31$, RMSE = 78.84) > traditional linear model ($R^2 = 0.16$, RMSE = 87.62). This implies that the GA-BPNN model demonstrated the highest accuracy and the strongest prediction ability, consistent with the results of the training samples.

3.3. Spatial Prediction of Cultivated Land Quality at the Regional Scale

The GA-BPNN model was applied to map CLQ in Conghua District at the regional scale using the evaluation indicators. The CLQ values were then divided into three grades according to gradation regulations on agriculture land quality in China (GB/T 28407-2012). The derived spatial distribution of CLQ was presented in Figure 5.

In order to validate the mapping accuracy of CLQ using the GA-BPNN model, the CLQ estimated using the evaluation indicators was compared with the measured data from the 250 samples by calculating the R^2, RMSE, and NRMSE (Figure 6). There was an improvement in the CLQ mapping using the satellite-derived indicators based on the GA-BPNN model ($R^2 = 0.56$, RMSE = 61.44, NRMSE = 13.04%), indicating that the method proposed in this study is capable of estimating CLQ at the regional scale.
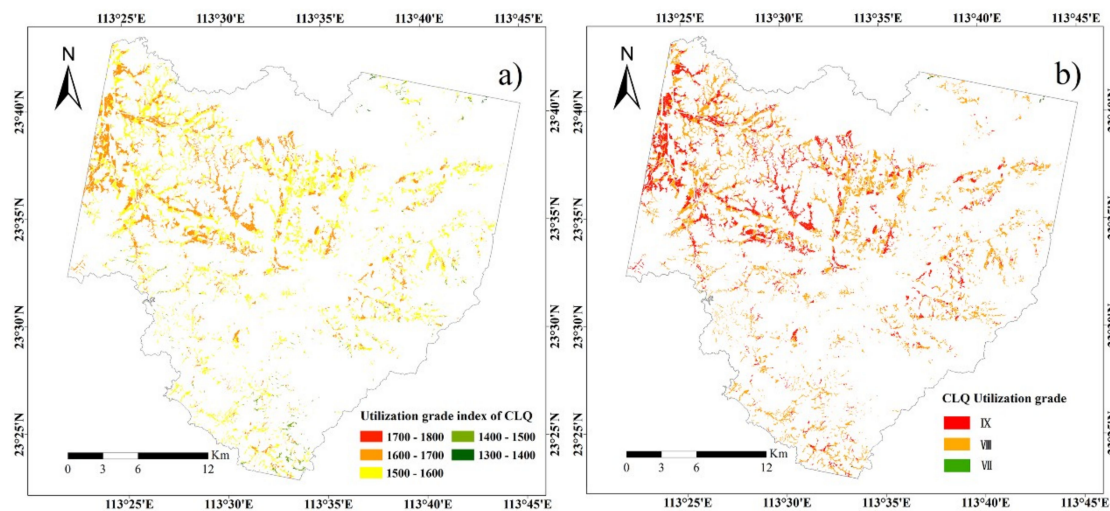
Figure 5. Spatial distributions of CLQ using GA-BPNN for the study area: (**a**) utilization grade index and (**b**) utilization grade.
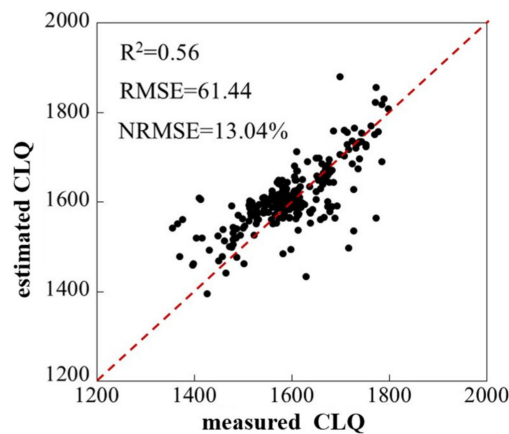

Figure 6. Measured and estimated CLQ based on the GA-BPNN model using the 250 validation sample plots for mapping.

4. Discussion

The CLQ plays a vital role in agricultural production. CLQ will change drastically over a short time when environmental change and human economic activities occur. Therefore, timely assessment and dynamic monitoring of CLQ are especially critical in the agricultural vulnerable region. Estimating CLQ using remote sensing data is a cost-efficient, yet challenging method owing to the complex nonlinear relationship between CLQ and the evaluation indices [40]. Previous studies [7–9] based on the PSR framework using remote sensing technique have generally focused on building linear models with coefficients determined from expert experience. In order to improve the CLQ evaluation accuracy from previous work, this study applied linear (traditional linear model and PLSR) and nonlinear models (BPNN and GA-BPNN) to evaluate CLQ.

Compared with the evaluation results from the four models, BPNN and GA-BPNN models (R^2 of 0.50 and 0.59) performed better than PLSR and the traditional linear models (R^2 of 0.33 and 0.14) in accurately predicting CLQ, highlighting the evident nonlinear relationships between satellite-derived indicators and CLQ. Meanwhile, this demonstrates that the GA-BPNN model performed better on CLQ evaluation than the BPNN model, because GA was introduced to BPNN, which led to an integrated GA-BPNN method to optimize the BPNN initial input parameters (thresholds and weights) and provide the solution for the problem of being stuck in the local minima [41]. The results of the validation datasets also showed that, compared with the BPNN model without the optimization of the

input parameters, the GA-BPNN model significantly improved the estimation accuracies of CLQ by decreasing the RMSE values by 14.81.

In order to further validate the regional scale applicability of the GA-BPNN prediction model, the satellite-derived indicators were used to map CLQ over Conghua District. The comparison between the measured and estimated CLQ resulted in an RMSE of 55.14 and NRMSE of 11.68%. The result demonstrated that the satellite-driven evaluation approach in the PSR framework using the GA-BPNN model has the potential to accurately map CLQ.

It has to be pointed out that the experiment was conducted only in paddy fields. Owing to the absence of information on other types of cultivated land, we cannot currently validate the evaluation indicators and CLQ model uncertainties caused by cultivated land type. Thus, more field samples from different types of cultivated land should be collected to further validate the reliability of the indicators and the model proposed in the study. Moreover, in order to further build and validate the optimal indicators of CLQ evaluations in future studies, larger sample sizes with a CLQ range from low to high should be utilized. This will improve the accuracy of the CLQ predictions. Finally, compared with the algorithm efficiency of linear models (the traditional linear and PLSR), the BPNN and GA-BPNN algorithms have the problem of slow learning, resulting in the difficulty to satisfy practical applications, especially for large-scale training samples [42]. Thus, the addition of other prediction methods (e.g., multiple linear regression, MLR; deep learning; support vector machine, SVM; and random forest, RF) is expected to enhance the efficiency and accuracy of CLQ evaluations.

5. Conclusions

It is agreed that estimating and mapping CLQ using satellite-derived approaches based on the PSR framework is quick and effective, yet also very challenging owing to spectral indicator accuracy, modeling methods, and model transferability, which will affect the accuracy of CLQ evaluation. In order to improve the accuracy of CLQ evaluation, this study focused on determining an accurate spectral response relationship model between CLQ and satellite-derived indicators based on the PSR framework to evaluate CLQ using the traditional linear, PLSR, BPNN, and GA-BPNN algorithms. The experiment was conducted in the Conghua district of Guangzhou City. The following conclusions could be drawn: (1) comparing with other VIs (NDVI, PVI, MSAVI, and SAVI), EVI, with the greatest correlation coefficients r of 0.88 for SOM and 0.90 for TN, is the optimal indicator in the land state index; (2) the GA-BPNN model demonstrated the strongest fitting ability for CLQ (R^2 = 0.59 and RMSE = 56.87) compared with the PLSR, BPNN, and traditional linear models. This indicates a nonlinear relationship between CLQ and the prediction indicators; (3) the validation led to a relatively small NRMSE value of 13.04% for mapping CLQ, which further indicated that the GA-BPNN-based evaluation approach of CLQ in the PSR framework using GF-1 data was reliable to map CLQ at the regional scale. This study provides an important reference for high-accuracy CLQ predictions.

References

1. Tampakis, S.; Karanikola, P.; Koutroumanidis, T.; Tsitouridou, C. Protecting the productivity of cultivated lands. The viewpoints of farmers in Northern Evros. *J. Environ. Prot. Ecol.* **2010**, *11*, 601–613.
2. Xie, H.L.; Zou, J.L.; Jiang, H.L.; Zhang, N.; Choi, Y. Spatiotemporal pattern and driving forces of arable land-use intensity in China: Toward sustainable land management using emergy analysis. *Sustainability* **2014**, *6*, 3504–3520. [CrossRef]
3. Song, W.; Liu, M.L. Farmland conversion decreases regional and national land quality in China. *Land Degrad. Dev.* **2017**, *28*, 459–471. [CrossRef]
4. German, R.N.; Thompson, C.E.; Benton, T.G. Relationships among multiple aspects of agriculture's environmental impact and productivity: A meta-analysis to guide sustainable agriculture. *Biol. Rev.* **2017**, *92*, 716–738. [CrossRef] [PubMed]
5. Oliver, M.A.; Gregory, P.J. Soil, food security and human health: A review. *Eur. J. Soil Sci.* **2015**, *66*, 257–276. [CrossRef]
6. Zhu, Q.; Liao, K.H.; Xu, Y.; Yang, G.S.; Wu, S.H.; Zhou, S.L. Monitoring and prediction of soil moisture spatial? temporal variations from a hydropedological perspective: A review. *Soil Res.* **2012**, *50*, 625. [CrossRef]
7. Wang, Z.; Wang, L.M.; Xu, R.N.; Huang, H.T.; Wu, F. GIS and RS based assessment of cultivated land quality of Shandong Province. *Procedia Environ. Sci.* **2012**, *12*, 823–830. [CrossRef]
8. Liu, Y.S.; Zhang, Y.Y.; Guo, L.Y. Towards realistic assessment of cultivated land quality in an ecologically fragile environment: A satellite imagery-based approach. *Appl. Geogr.* **2010**, *30*, 271–281. [CrossRef]
9. Fang, L.N.; Song, J.P. Cultivated Land Quality Assessment Based on SPOT Multispectral Remote Sensing Image: A Case Study in Jimo City of Shandong Province. *Prog. Geogr.* **2008**, *27*, 71–78.
10. Zhang, Y.Y.; Shi, Y.X.; Wang, J.; Liu, Q.; Fang, L.N. Comprehensive evaluation of cultivated land quality based on RS and PRA: A case study in Hengshan county. *Sci. Surv. Mapp.* **2008**, *33*, 133–136.
11. Yu, X.J. *GIS and RS Supported Quantitative Evaluation of Cultivated Land Productivity in Zhaodong City*; Northeast Agricultural University: Harbing, China, 2012.
12. Yang, J.F.; Ma, J.C.; Wang, L.C. Evaluation factors for cultivated land grade identification based on multi-spectral remote sensing. *Trans. Chin. Soc. Agric. Eng.* **2012**, *28*, 230–236.
13. Ye, H.; Ma, Y.; Dong, L.M. Land Ecological Security Assessment for Bai Autonomous Prefecture of Dali Based Using PSR Model-with Data in 2009 as Case. *Energy Procedia* **2011**, *5*, 2172–2177.
14. Bao, S.D. *Soil Agricultural Chemistry Analysis*; China Agriculture Press: Beijing, China, 2000.
15. Dumanski, J.; Pieri, C. Land quality indicators: Research plan. *Agric. Ecosyst. Environ.* **2000**, *81*, 93–102. [CrossRef]
16. Al-Kaisi, M.M.; Yin, X.H.; Licht, M.A. Soil carbon and nitrogen changes as influenced by tillage and cropping systems in some Iowa soils. *Agr. Ecosyst. Environ.* **2005**, *105*, 635–647. [CrossRef]
17. Huang, B.; Sun, W.; Zhao, Y.; Zhu, J.; Yang, R.; Zou, Z.; Feng, D.; Jianping, S. Temporal and spatial variability of soil organic matter and total nitrogen in an agricultural ecosystem as affected by farming practices. *Geoderma* **2007**, *139*, 336–345. [CrossRef]
18. Sandholt, I.; Rasmussen, K.; Andersen, J. A simple interpretation of the surface temperature/vegetation index space for assessment of surface moisture status. *Remote Sens. Environ.* **2002**, *79*, 213–224. [CrossRef]
19. Li, Z.G.; Wang, Y.L.; Zhou, Q.B.; Wu, J.S.; Peng, J.; Chang, H.F. Spatiotemporal variability of land surface moisture based on vegetation and temperature characteristics in Northern Shaanxi Loess Plateau, China. *J. Arid. Environ.* **2008**, *72*, 974–985. [CrossRef]
20. Heumann, S.; Ringe, H.; Böttcher, J. Long-term net N mineralization potential as an indicator for soil fertility: Chances and constraints. *Arch. Agron. Soil Sci.* **2012**, *58*, S107–S111. [CrossRef]
21. Marti, J.; Bort, J.; Slafer, G.A.; Araus, J.L. Can wheat yield be assessed by early measurements of Normalized Difference Vegetation Index? *Ann. Appl. Biol.* **2007**, *150*, 253–257. [CrossRef]
22. Jiang, D.; Wang, N.B.; Yang, X.H.; Wang, J.H. Study on the interaction between NDVI profile and the growing status of crops. *Chin. Geogr. Sci.* **2003**, *13*, 62–65. [CrossRef]
23. Zhang, Y.J.; Alfredo, R.H.; Kamel, D.; Tomoaki, M. Development of a two-band enhanced vegetation index without a blue band. *Remote Sens. Environ.* **2008**, *112*, 3833–3845.
24. Chen, J.M. Evaluation of Vegetation Indices and a modified simple ratio for boreal applications. *Can. J. Remote Sens.* **1996**, *22*, 229–242. [CrossRef]

25. Huete, A.R. A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* **1988**, *25*, 295–309. [CrossRef]

26. Schlerf, M.; Atzberger, C.; Hill, J. Remote sensing of forest biophysical variables using HyMap imaging spectrometer data. *Remote Sens. Environ.* **2005**, *95*, 177–194. [CrossRef]

27. Yoshino, K.; Ishioka, Y. Guidelines for soil conservation towards integrated basin management for sustainable development: A new approach based on the assessment of soil loss risk using remote sensing and GIS. *Paddy Water Environ.* **2005**, *3*, 235–247. [CrossRef]

28. Han, Y.; Guo, X.; Jiang, Y.F.; Rao, L.; Sun, K.; Yu, H.M. Correlation analysis between cultivated land quality and landscape pattern index in south hilly areas. *Jiangsu J. Agric. Sci.* **2018**, *34*, 1057–1065.

29. Renetzeder, C.; Schindler, S.; Peterseil, J.; Prinz, M.A.; Mücher, S.; Wrbka, T. Can we measure ecological sustainability? Landscape pattern as an indicator for naturalness and land use intensity at regional, national and European level. *Ecol. Indic.* **2010**, *10*, 39–48. [CrossRef]

30. Wang, J.; Wei, C.F.; Liu, W.P.; Zhong, S.Q.; Xing, G.X. A Method for Estimating Quality-Improvement Potential of Cultivated Land in Land Consolidation Projects in Hilly Areas. *J. Southwest Univ.* **2018**, *40*, 122–132.

31. Berg, C.N.; Blankespoor, B.; Selod, H. Roads and Rural Development in Sub-Saharan Africa. *J. Dev. Stud.* **2018**, *54*, 856–874. [CrossRef]

32. Wold, H. Nonlinear Estimation by Iterative Least Squares Procedure. In *Research Papers in Statistics*; David, F., Ed.; Wiley: Hoboken, NJ, USA, 1966; pp. 441–444.

33. Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: A basic tool of chemometrics. *Chemometr. Intell. Lab.* **2001**, *58*, 109–130. [CrossRef]

34. Lin, L.X.; Wang, Y.J.; Teng, J.Y.; Xi, X.X. Hyperspectral Analysis of Soil Total Nitrogen in Subsided Land Using the Local Correlation Maximization-Complementary Superiority (LCMCS) Method. *Sensors* **2015**, *15*, 17990–18011. [CrossRef] [PubMed]

35. Zhao, L.; Hu, Y.-M.; Zhou, W.; Liu, Z.-H.; Pan, Y.-C.; Shi, Z.; Wang, L.; Wang, G.-X. Estimation Methods for Soil Mercury Content Using Hyperspectral Remote Sensing. *Sustainability* **2018**, *10*, 2474. [CrossRef]

36. Hively, W.D.; Mccarty, G.W.; Reeves, J.B.; Lang, M.W.; Oesterling, R.A.; Delwiche, S.R. Use of Airborne Hyperspectral Imagery to Map Soil Properties in Tilled Agricultural Fields. *Appl. Environ. Soil Sci.* **2011**, *2011*, 1–13. [CrossRef]

37. Salmerón Gómez, R.; García Pérez, J.; López Martín, M.D.M.; García, C.G. Collinearity diagnostic applied in ridge estimation through the variance inflation factor. *J. Appl. Stat.* **2016**, *43*, 1831–1849. [CrossRef]

38. Saleh, S.; Ibrahim, K.; Eiteba, M.M.; Mohamed, S. Study of genetic algorithm performance through design of multi-step LC compensator for time-varying nonlinear loads. *Appl. Soft Comput.* **2016**, *48*, 535–545. [CrossRef]

39. Peng, Y.P.; Zhao, L.; Hu, Y.M.; Wang, G.X.; Wang, L.; Liu, Z.H. Prediction of Soil Nutrient Contents Using Visible and Near-Infrared Reflectance Spectroscopy. *ISPRS Int. J. Geo Inf.* **2019**, *8*, 437. [CrossRef]

40. Qie, R.Q.; Guan, X.; Yan, X.J.; Dou, S.X.; Zhao, L. Method and its application of natural quality evaluation of arable land based on self-organizing feature map neural network. *Trans. Chin. Soc. Agric. Eng.* **2014**, *30*, 298–305.

41. Khosravi, V.; Ardejani, F.D.; Yousefi, S.; Aryafar, A. Monitoring soil lead and zinc contents via combination of spectroscopy with extreme learning machine and other data mining methods. *Geoderma. Int.* **2018**, *318*, 29–41. [CrossRef]

42. Wang, W.; Yu, B. Text categorization based on combination of modified back propagation neural network and latent semantic analysis. *Neural Comput. Appl.* **2008**, *18*, 875–881. [CrossRef]