



## Article

# Water Conservation Estimation Based on Time Series NDVI in the Yellow River Basin

Yangchengsi Zhang<sup>1,2</sup>, Jiaqiang Du<sup>1,2,\*</sup>, Long Guo<sup>3</sup>, Zhilu Sheng<sup>1,2</sup>, Jinhua Wu<sup>1,2</sup> and Jing Zhang<sup>1,2,4</sup>

<sup>1</sup> State Key Laboratory of Environmental Criteria and Risk Assessment, Chinese Research Academy of Environmental Sciences, Beijing 100012, China; zhangyangchengsi@webmail.hzau.edu.cn (Y.Z.); zlsheng@mail.bnu.edu.cn (Z.S.); wujinhua18@mails.ucas.ac.cn (J.W.); zhj19@lzu.edu.cn (J.Z.)

<sup>2</sup> State Environmental Protection Key Laboratory of Regional Eco-process and Function Assessment, Chinese Research Academy of Environmental Science, Beijing 100012, China

<sup>3</sup> College of Resources and Environment, Huazhong Agricultural University, Wuhan 430070, China; guolong@mail.hzau.edu.cn

<sup>4</sup> School of Life Sciences, Lanzhou University, Lanzhou 730000, China

\* Correspondence: dujq@craes.org.cn; Tel.: +86-1381-046-0922

**Abstract:** Accurate estimation of the water conservation is of great significance for ecological red line planning. The water conservation of the Yellow River Basin has a vital influence on the development of the environment and the supply of ecological services in China. However, the existing methods used to estimate water conservation have many disadvantages, such as requiring numerous parameters, a complex calculation model, and using data that is often difficult to acquire. It is often hard to provide sufficiently precise parameters and data, resulting in a large amount of calculation time and the difficulties in the study of large scale and long time series. In this study, a time series of the Normalized Difference Vegetation Index (NDVI) was applied to estimate water conservation in two aspects using the idea of wholeness and stratification, respectively. The overall fitting results can explain nearly 30% of the water conservation by partial least squares regression and nearly 50% of it by a support vector machine. However, the results of a stratified simulation showed that water conservation and the NDVI have a certain stratified heterogeneity among different ecosystem types. The optimal fitting result was achieved in a water/wetland ecosystem with the highest coefficient of determination ( $R^2_P$ ) of 0.768 by the stratified support vector machine (SVM) model, followed by the forest and grassland ecosystem (both  $R^2_P$  of 0.698). The spatial mapping results showed that this method was most suitable for grassland ecosystem, followed by forest ecosystem. According to the results generated using the NDVI time series data, it is feasible to complete a spatial simulation of water conservation. This research can provide a reference for calculating regional or large-scale water conservation and in ecological red line planning.

**Keywords:** water conservation; NDVI time series; spatial stratified heterogeneity; digital mapping; machine learning method



**Citation:** Zhang, Y.; Du, J.; Guo, L.; Sheng, Z.; Wu, J.; Zhang, J. Water Conservation Estimation Based on Time Series NDVI in the Yellow River Basin. *Remote Sens.* **2021**, *13*, 1105. <https://doi.org/10.3390/rs13061105>

Academic Editor: Giorgio Baiamonte

Received: 19 February 2021

Accepted: 10 March 2021

Published: 14 March 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The Yellow River is often called the mother river of China. For a long period of history, the political, economic, and cultural center of China has always been in the Yellow River Basin [1]. The ecological problems of the Yellow River Basin have a profound impact on the ecological nature of China; hence, these problems have created an important ecological barrier to sound environmental management. The water cycle involves the interception, penetration, and accumulation of precipitation as the unique structure of soil interacts with water and facilitates the regulation of water flow through evapotranspiration. Water conservation mainly involves mitigating surface runoff, replenishing groundwater, slowing the seasonal fluctuations of river flow, holding back floods, addressing drought related issues, and ensuring water quality [2]. Therefore, in this paper, we define water conservation not

as the “wise use of water resources” but in the sense of how each ecosystem controls the cycling of water through the ecosystem itself.

The ecological red line is an important institutional innovation in environmental protection in China, which is the bottom line of ecological security, in essence. It refers to the strict protection of spatial boundaries and management limitation in terms of natural ecological service functions, environmental quality safety, and utilization of natural resources in order to maintain national and regional ecological security, sustainable economic and social development, and protect the health of people [2–4]. The evaluation of water conservation is one of the most crucial aspect of ecosystem services in this project. The water cycle and water conservation function of the Yellow River Basin have an impact on the ecosystem productivity, nutrient cycling, and functional regulation of water resources in the region, so these topics have attracted much attention in recent years. Many scholars worldwide have published a large number of research results in related fields [5–7]. Zhang et al. [8] reviewed the research and progress related to the concept of water conservation, summarized its main manifestations, and introduced methods used to measure water storage. Lü et al. [9] evaluated the spatial and temporal changes of water yield and water conservation in Sanjiangyuan National Park based on the InVEST model. Bai et al. [10] explored the methods used to designate an ecologic red line that has a strong support of water conservation and is based on remote sensing, meteorological and solar radiation data, and so on using a combination of GIS (Geographic Information System) technology and the regional water balance method.

However, the above studies have obvious deficiencies: a large amount of environmental variable data is usually required, and while some of the data such as evapotranspiration, the volumetric plant available water content, the plant available water content, the minimum of root restricting layer depth and vegetation rooting depth, and so on are difficult to obtain accurately. Even if they can be measured, it is difficult to obtain these data accurately over a large-scale area for a long term. Moreover, the complexity of most water conservation simulation models also makes it difficult to estimate how water can be conserved. In fact, water conservation is the result of the combined effects of several environmental factors over a long period of time, rather than the temporary state of these environmental variables. From a certain perspective, the total amount of water available in any single year should be the result of years of cumulative processes. Therefore, a time series of environmental variables is of great significance for the analysis of water conservation. Unfortunately, most studies on water conservation have not considered the influence of time on it and have not discussed the relationship between long time series environmental variables and water conservation.

The Normalized Difference Vegetation Index (NDVI) provides a good indicator that is useful for identifying long-term changes in vegetation and vegetative conditions [11–15]. A change of the NDVI can allow researchers to distinguish vegetation from other surface covers, such as cloud, water, snow, and bare soil. Vegetation intercepts and stores water, which helps in regulating water availability and conservation. Therefore, from the perspective of vegetation, a biomass may have a significant relationship with water conservation, and it could be used for the analysis and modeling of water. Aguilar et al. [16] applied the NDVI as an indicator for changes in water availability to woody vegetation. Joiner et al. [17] explored the global relationships among the NDVI, evapotranspiration (ET), and soil moisture variability on weekly timescales and found that they were significantly correlated. Fu and Burgher [11] used a regression tree analysis to investigate the NDVI dynamics of riparian vegetation and its relationship with the climate, surface water, and groundwater. However, only considering the NDVI at a certain point of time may not provide an ideal result. Meanwhile, many studies have shown that the sequential feature of a NDVI time series data could reveal more information [18–20]. Therefore, in this article, we use time series NDVI data to analyze water conservation and explore its relationship with a NDVI time series.

Wang et al. [21] and Wang et al. [22] developed a Geodetector for the stratified heterogeneity of spatial geographic phenomena to reveal its properties in different strata. Many studies have used a Geodetector to detect the spatial drivers of environmental variables such as vegetation [23–26] and achieved good results. Water conservation and the NDVI both vary among different ecosystem types. Using a Geodetector to explore the water use/conservation and NDVI among different ecosystem types can help clarify their relationships more clearly and will help researchers to build more accurate models.

Generally, this article has the following goals:

(1) Explore the relationship between the NDVI time series and water use/conservation and use this relationship for the modeling and estimation of water use.

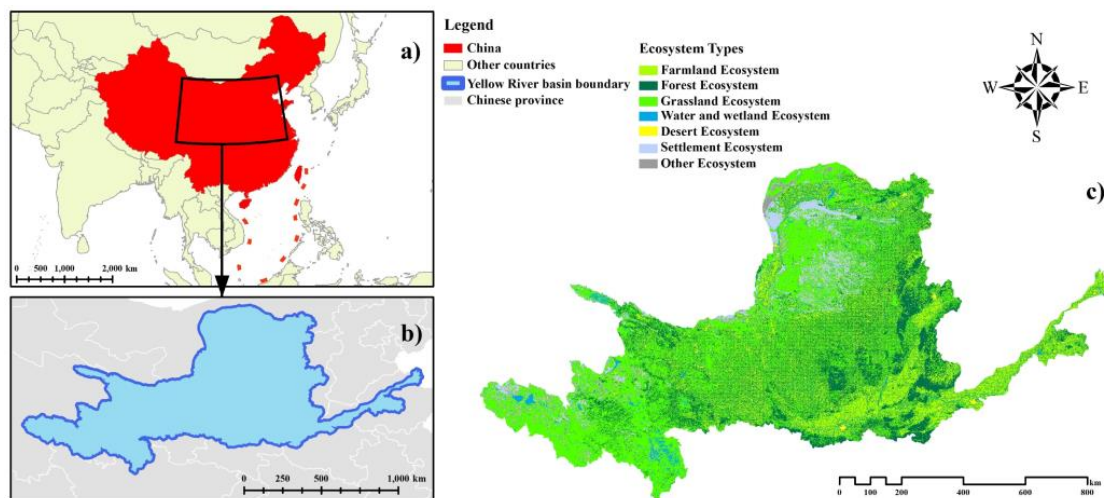
(2) Discuss the stratified heterogeneity of the NDVI and water use/conservation among different ecosystem types.

(3) Carry out a stratified modeling estimation of water use/conservation and reveal the most suitable type of this method.

## 2. Materials and Methods

### 2.1. Study Area

The Yellow River, the great river in Northern China, is about 5464-km-long with a basin area of about 752,400 km<sup>2</sup>. The basin spans the western, central, and eastern parts of China. The western river source area, with an average elevation of more than 4000 m, is composed of a series of high mountains with perennial snow cover and developed glacial landforms. The central region, with an elevation of 1000–2000 m, is a loess landform that frequently experiences serious soil erosion. Meanwhile, the eastern part is mainly composed of the Yellow River alluvial plain. The Yellow River Basin has sufficient sunshine with strong solar radiation, where annual sunshine hours are generally between 2000 and 3300 hours. The annual temperature difference in the Yellow River Basin is relatively large. Generally, the area north of 37°N has temperatures between 31–37°C, while temperatures in the area south of 37°N mostly range between 21–31°C. The annual precipitation in most areas falls between 200 and 650 mm, with more than 650 mm in the southern part of the middle, upper reaches, and lower sides of the river, and it can even reach 700–1000 mm in some areas. Water use and conservation in the Yellow River Basin have an important effect on the development of the environment and the supply of ecological services water and the landscape in China. The spatial location of this study area, Yellow River Basin, and its ecosystem types are shown in Figure 1.



**Figure 1.** The location of this study area: (a) general location of the study area within Eastern Asia, (b) location of the study area plotted over a map of the provinces and administrative areas of China, and (c) spatial distribution of the ecosystem types in the study area.

## 2.2. Data Acquisition

### 2.2.1. Evapotranspiration Data

Evapotranspiration data were acquired from the Moderate-resolution Imaging Spectroradiometer (MODIS) Global Evapotranspiration Project (MOD16) (<https://ladsweb.modaps.eosdis.nasa.gov>, accessed on 18 February 2021). This dataset is an 8-days composite product that includes five layers: total ET, average latent heat flux (LE), total potential evapotranspiration (PET), average potential latent heat flux (PLE), and evapotranspiration quality control flags (ET\_QC), which are regular land surface ET datasets for the 109.03 million km<sup>2</sup> produced at a 500-m resolution [27,28]. We downloaded all the data between 1 January 2019 and 31 December 2019, mainly using the first layer.

### 2.2.2. Precipitation Data

Precipitation data came from Global Precipitation Measurement (GPM) IMERG Final Precipitation V06 at Level 3 with a temporal resolution of 1 month and a spatial resolution 0.1° × 0.1° from the Goddard Earth Sciences Data and Information Services Center (GES DISC) (<https://disc.gsfc.nasa.gov/>, accessed on 18 February 2021) [29,30]. The unit is in mm/h, and it required a further conversion. Similarly, all data from this dataset in 2019 were downloaded in order to operate a calculation.

### 2.2.3. Normalized Difference Vegetation Index (NDVI) Time Series Data

We collected a NDVI dataset from MODIS (MOD13Q1 series) on the official website of the US National Aeronautics and Space Administration (<https://ladsweb.modaps.eosdis.nasa.gov/search/>, accessed on 18 February 2021). This dataset was provided every 16 days at a 250-m spatial resolution as a Level 3 data product. This MOD13Q1 product provides 16 layers with two primary vegetation layers: Normalized Difference Vegetation Index (NDVI) and an Enhanced Vegetation Index [31,32]. For this study, we downloaded NDVI data in the study area acquired between 2000 and 2019, including a total of 2285 images, as long as 457 issues and 20 years for the modeling analysis, which developed a long time series to predict the water conservation in the last year (2019).

### 2.2.4. Other Auxiliary Data

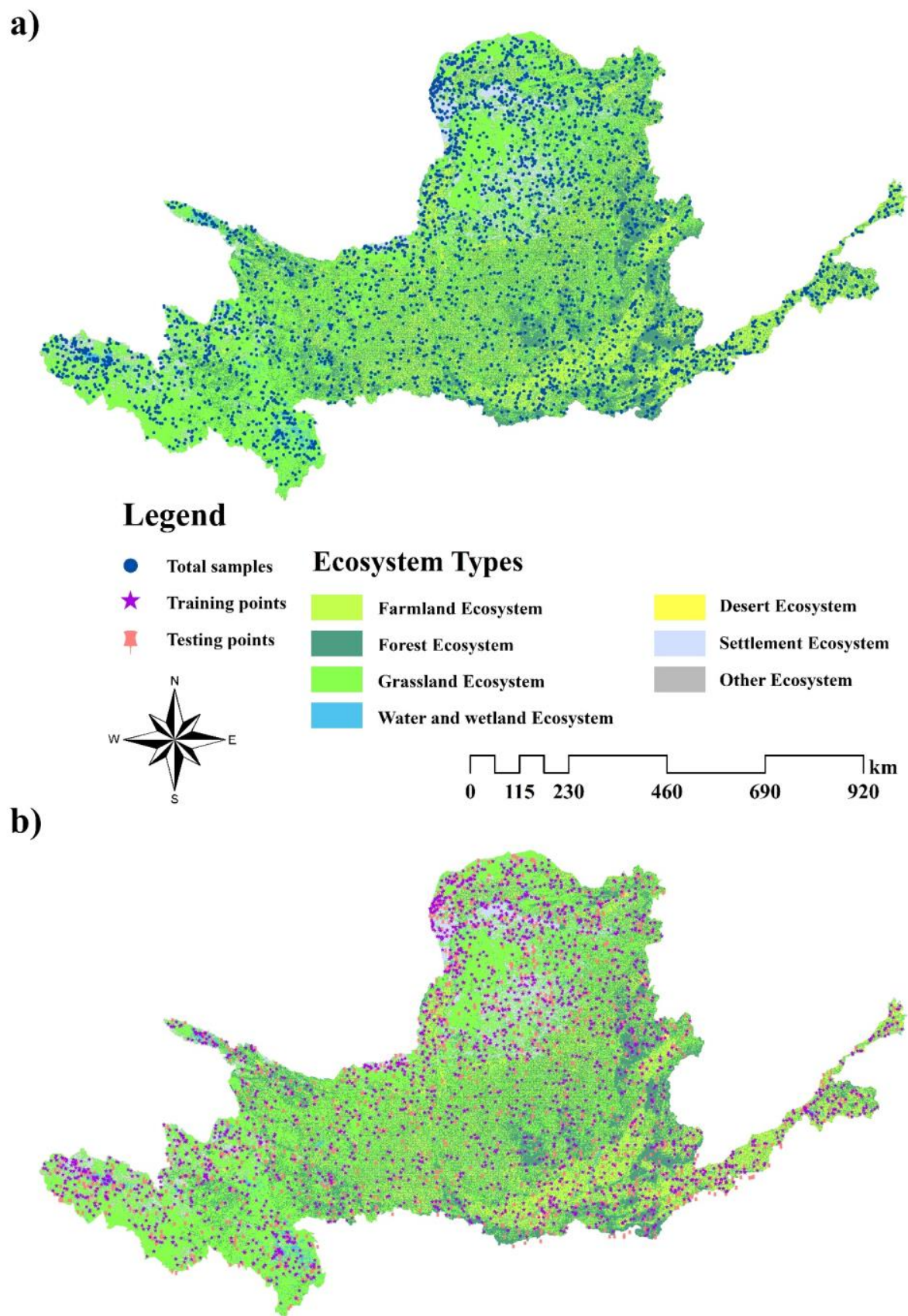
In addition to the data mentioned above, other auxiliary data were also collected. The ecosystem type dataset was downloaded from the Resource and Environment Data Cloud Platform (<http://www.resdc.cn/Default.aspx>, accessed on 18 February 2021). This dataset divides the terrestrial ecosystems of China into seven ecosystem categories: farmland, forest, grassland, water/wetland, desert, settlement, and other ecosystem types (Figure 1). In addition, the spatial range vector data and the administrative boundary data of the Yellow River Basin were also required.

## 2.3. Data Processing

All ET and NDVI data were processed using the MODIS Reprojection Tool batch process of the image mosaic and reprojection with the same spatial reference. Precipitation data (with the data format NetCDF) were processed, including defining their spatial reference and by exporting the data to a TIFF format. All the raster data were clipped to the study area.

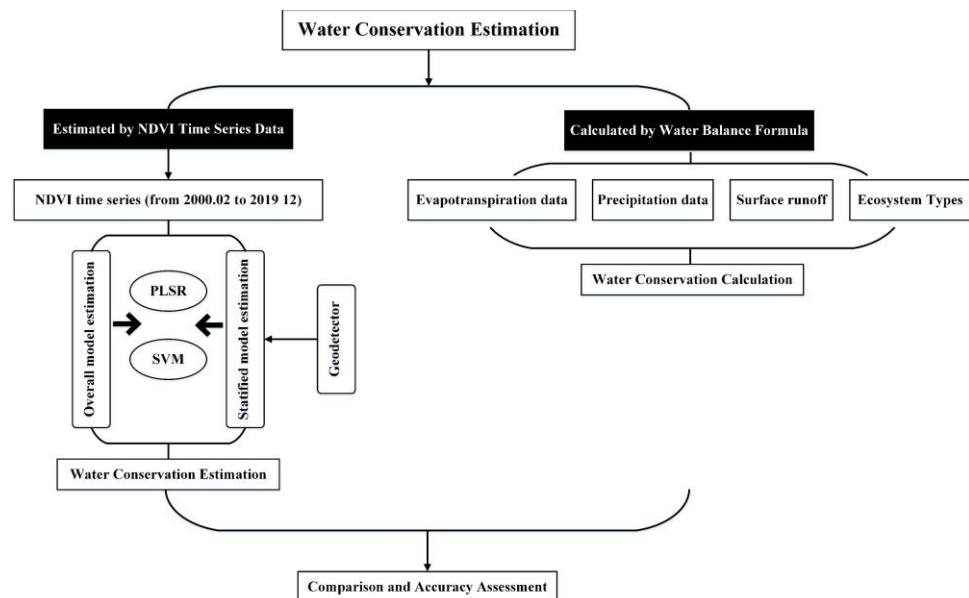
For the modeling analysis, 3500 sample points were randomly established throughout the study area (500 points for each ecosystem type; Figure 2). For the selection of the number of sample points, we used 200, 300, 500, and 1000 for each ecosystem type for various simulations. Finally, in consideration of model accuracy and calculation cost, 500 points for each ecosystem type were ultimately used here (the specific simulation results can be found in Table 1 of Appendix A).





**Figure 2.** The spatial distribution of sample points: (a) illustrates all samples and (b) illustrates the training (70%) and testing (30%) points.

The specific research process can be seen in a diagram in Figure 3.



**Figure 3.** Diagram of the research process. Note: PLSR, partial least squares regression and SVM, support vector machine.

#### 2.4. Water Conservation Calculated by A Water Balance Equation

Water conservation means that the ecosystem intercepts and accumulates precipitation as it penetrates the soil through the ecosystem's unique structural interaction with water; the ecosystem operates the regulation of water flow and the water cycle through evapotranspiration [2,33]. The existence of water conservation can help an ecosystem to retain precipitation, regulate runoff, affect rainfall, and improve water quality [6]. Guidelines for delineating red lines for ecological protection provide a recognized and relatively accurate calculation model for water conservation:

$$TQ = \sum_{i=1}^j (P_i - R_i - ET_i) \times A_i \times 10^3 \quad (1)$$

where  $TQ$  is the total content of water conservation ( $m^3$ ),  $P_i$  is the annual precipitation,  $R_i$  is the surface runoff (mm),  $ET_i$  is the evapotranspiration (mm),  $A_i$  is the ecosystem area of type  $i$ ,  $i$  is the ecosystem type  $i$  in the study area, and  $j$  is the number of ecosystem types in the study area. The surface runoff is calculated according to

$$R = P \times \alpha \quad (2)$$

where  $R$  means surface runoff (mm),  $P$  is annual precipitation, and  $\alpha$  is the surface runoff coefficient. Specific coefficient values can be seen in Table 1.

**Table 1.** Surface runoff coefficient for each ecosystem [2,34,35].

Ecosystem Type	Surface Runoff Coefficient
1. Farmland	0.43500
2. Forest	0.05912
3. Grassland	0.06842
4. Water/wetland	0.00000
5. Desert	0.48661
6. Settlement	0.64200
7. Other ecosystems	0.60000

All factors were unified into the grid data with a resolution of 250 m. The ecosystem water conservation was calculated according to the above formulas.

### 2.5. Relationship between NDVI and Water Conservation

The correlation between the water conservation and NDVI time series data, as represented by the Pearson correlation coefficient ( $r$ ), was calculated based on the Equations (3) and (4).

$$\rho(x, y) = \frac{\text{cov}(x, y)}{\sigma_x \cdot \sigma_y} \quad (3)$$

$$r = \frac{\sum_i^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i^n (x_i - \bar{x})^2} \sqrt{\sum_i^n (y_i - \bar{y})^2}} \quad (4)$$

where  $x$  is the NDVI data (from  $-1$  to  $1$ ),  $y$  is the amount water conservation ( $\text{m}^3$ ), and  $n$  the number of sequential sets.

### 2.6. Geodetector Software

We used R Geodetector software, which was developed by Wang et al. [22] to analyze the stratified heterogeneity (SH) of water conservation and the NDVI with different ecosystem types. This software can be downloaded for free from <http://www.geodetector.cn/> (accessed on 18 February 2021). Geodetector is a statistical tool used to measure the SH and to explore the determinants of SH [36]. Its tasks can be accomplished by the Geodetector  $q$ -statistic:

$$q = 1 - \frac{\sum_{h=1}^L N_h \sigma_h^2}{N \sigma^2} \quad (5)$$

where  $q$  means the degree of SH for factor  $X$  (ecosystem types) on water conservation and the NDVI, and  $N$  and  $\sigma^2$  stand for the number of ecosystem types and the variance of  $Y$  (water conservation and NDVI) in this study area, respectively.  $H$  is the number of classification or strata of factor  $Y$  or  $X$ . The value of  $q$  is strictly within  $[0, 1]$ . If  $Y$  is stratified by  $Y$  itself, then  $q = 0$  indicates that  $Y$  is absent of SH,  $q = 1$  indicates that  $Y$  is SH perfectly, and  $100\% \times q$  measures the degree of SH of  $Y$ . If  $Y$  is stratified by an explanatory variable  $X$ , then  $q = 0$  indicates that there is no coupling between  $Y$  and  $X$ ,  $q = 1$  indicates that  $Y$  is completely determined by  $X$ , and  $X$  explains  $100\% \times q$  of  $Y$ . Please note that the  $q$ -statistic measures the relationship between  $X$  and  $Y$ , both linearly and nonlinearly [21,22,36].

### 2.7. Model Establishment

A Python batch program was used to extract all the NDVI time series data values and water conservation values to the point for modeling. In order to validate the accuracy of models, the dataset was divided into training (70%) and testing (30%) data.

#### 2.7.1. Partial Least Squares Regression (PLSR) Model

Machine learning methods first proposed by Turing can effectively solve the collinearity problems existing in environmental variables, so that the estimation models can accurately represent the effects in the real world. Partial least squares regression (PLSR), as proposed by Wold et al. [37], is a multivariate statistical analysis method. It integrates the advantages of multiple linear regression, principal component, and typical correlation analysis methods. In other words, the PLSR can reduce the complete set of variable information (containing noise and repetitive information) to a smaller number of noncorrelated components (latent variables) that contain the most useful information [38] in order to effectively solve the problem of multicollinearity among dependent variables. It can discard irrelevant, redundant, and unstable information and use the most relevant  $X$  variable for regression analysis [39,40]. Thus, it enables the water conservation estimation model to become more stable and to produce a better estimate. The detailed theory were discussed by Wold et al. [37].

### 2.7.2. SVM (Support Vector Machine) Model

An SVR is a machine learning approach based on support vector machines invented by Vapnik [41]; this is a kernel-based learning method from statistical learning theory, which is suited for small input sampling sizes [42]. It is based on the computation of a linear regression function in a multidimensional feature space. The linear model created in the new space can represent a nonlinear decision boundary in the original space [43]. Its advantage is that it allows the mapping of an original nonlinear function into a higher dimensional feature space where the function can be treated as a linear function [42]. The SVMs used for regression are designed to create an optimal hyperplane that can separate class and create the widest margin between their data or that can fit data and predict with minimal empirical risk and complexity of the modeling function [44]. This method has been widely applied in regression and forecasting in various fields, such as agriculture, meteorology, and environmental monitoring studies [45–47].

### 2.8. Accuracy Validation

The validation datasets were used to analyze the estimative performance of the models for comparing the estimated results and identifying the optimal model. The accuracy of models was assessed by the root mean square error (RMSE) and coefficient of determination ( $R^2$ ) as the performance indicators. The RMSE statistic was used to evaluate the consistency between the estimated values from the estimated models and the original value calculated by the water conservation model where  $R^2$  reflects the degree of fitting between the estimated and calculated original values.

These indices were calculated as follows:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (C_i - M_i)^2}, \quad (6)$$

$$R^2 = \frac{\sum_i^n (M_i - \bar{C}_i)^2}{\sum_i^n (C_i - \bar{C}_i)^2}, \quad (7)$$

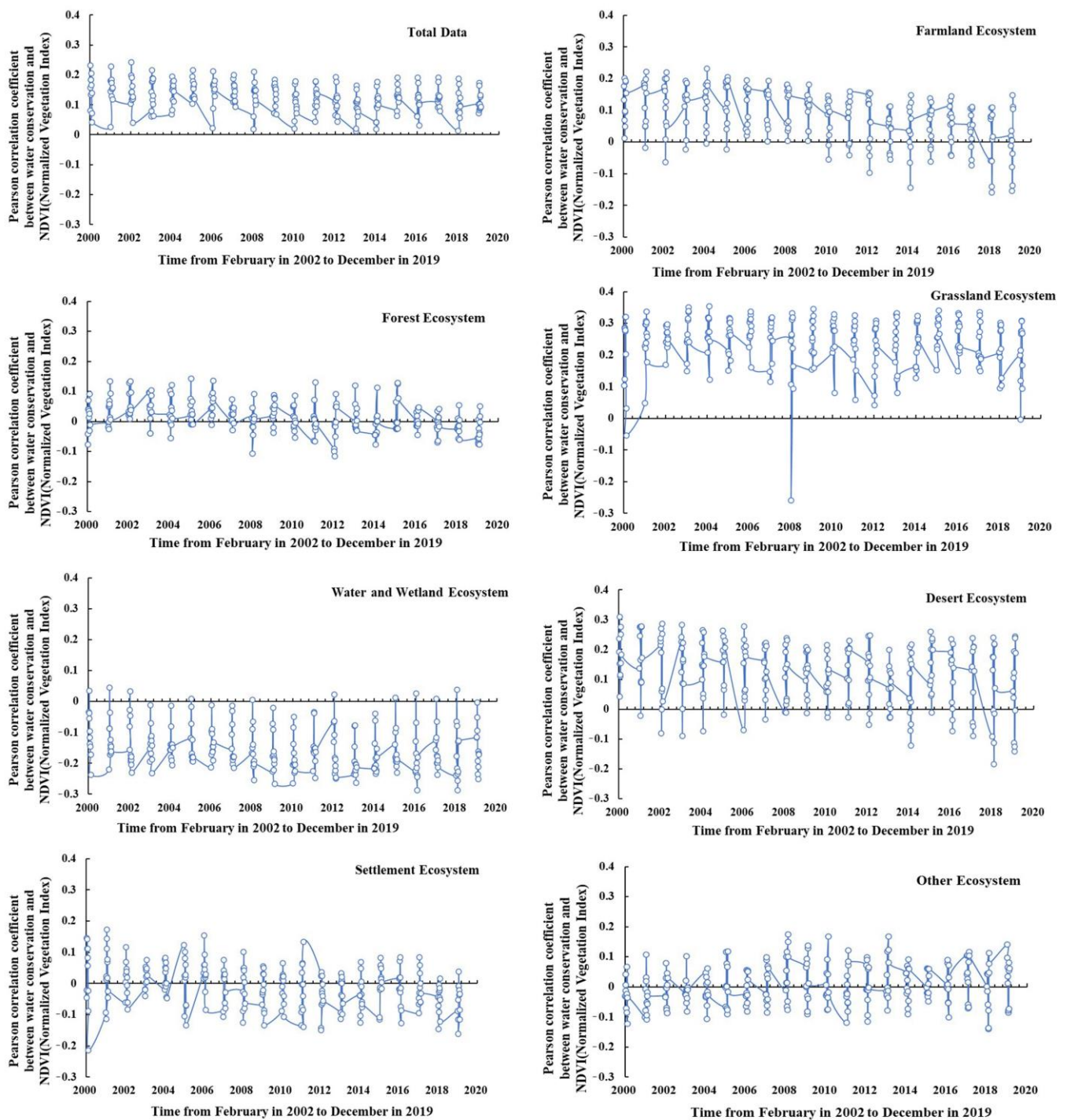
where  $C_i$  and  $M_i$  are the calculated and estimated water conservation values at site  $i$ , respectively,  $n$  is the total amount of modeling data, 3500,  $Cov$  indicates the covariance between the calculated and estimated values,  $Var$  is the variance between the calculated and estimated values, and  $\bar{C}_i$  is the average value of the calculated water conservation.

## 3. Results

### 3.1. Relationship between Water Conservation and NDVI Time Series

The correlation analysis results of the water conservation content and time series NDVI by the Pearson correlation coefficient ( $r$ ) are presented in Figure 4. Overall, the correlation between the NDVI and water conservation shows a periodic fluctuation over 20 years. For the total data, the correlation basically peaked in April (04) or May (05) (2000.04, 2001.04, 2002.04, 2003.04, 2004.04, 2005.04, 2006.04, 2007.04, 2008.04, 2009.04, 2010.05, 2011.04, 2012.04, 2013.04, 2014.05, 2015.04, 2016.04, 2017.04, 2018.04, and 2019.04), which indicated that a significant positive correlation existed between the water conservation content and NDVI time series data during these time periods. The correlation fell to its lowest value (close to 0) mostly in the winter (2000.12, 2001.01, 2002.12, 2003.02, 2004.01, 2005.01, 2006.01, 2007.12, 2008.02, 2009.12, 2010.01, 2011.01, 2012.11, 2013.01, 2014.02, 2015.02, 2016.11, 2017.12, 2018.01, and 2019.02), when the correlation between the water conservation and NDVI reached its lowest value.





**Figure 4.** The Pearson correlation coefficients between water conservation and normalized difference vegetation index (NDVI) time series data from 2000 to 2019 for all data and for seven ecosystem types.

However, when the complete set of data was stratified by ecosystem type, the results of the correlation analysis varied. The correlation coefficient after stratification still showed a periodic fluctuation over time but was different in the details. For farmland, its situation was similar to the complete set of data, but it peaked in the spring and fall, with lows in the summer. This phenomenon may be related to the properties of crops. The correlation between water conservation and the NDVI for forests was lower than that of the complete set of data. The correlation peaked in the spring and autumn in forests much the same way it did in the farmland ecosystem, while the lowest values for forests had no obvious

regularity. For the grassland ecosystem, its correlation coefficient was the highest among these stratified ecosystem types, exhibiting the most significant positive correlation between water conservation and the NDVI time series. With the exception of 2000, 2008, and 2019, almost all of the extreme values for the correlations were above zero. The highest values mostly occurred in the summer and autumn, while the lowest values occurred in the winter, spring, and, infrequently, in the autumn. The water and wetland ecosystem showed an opposite result to that of other types, where the correlation was basically less than zero, indicating that the NDVI time series is negatively correlated with water conservation. The desert ecosystem situation was similar to the one of farmland, most of which were positively correlated, with the peaks occurring in the spring and autumn and the correlation basically being insignificant in the summer. The last two ecosystems (settlements and other ecosystems) were similar, with low correlations and small insignificant fluctuations around 0.

In general, the grassland (significantly positive correlation) and water and wetland ecosystems (significantly negative correlations) had the highest correlations, followed by farmland, forest, and desert ecosystems in which the correlation was associated with the vegetation status. Not surprisingly, the correlation of the settlement and other ecosystems was the lowest.

This result indicated that a correlation exists between the NDVI time series data and water conservation, which showed different states with the stratification of the ecosystem types. In addition, it also demonstrated that the NDVI data may strongly affect the water conservation content and can be used to estimate the spatial distribution of water conservation. At the same time, the stratified estimation of water conservation based on ecosystem types should also be discussed.

### 3.2. Geodetector Analysis

The stratified heterogeneity of water conservation and the NDVI in different ecosystem types was explored by using Geodetector (Table 2, Figure 5, and Figure 6). The q-statistic values, 0.323 and 0.359, showed that there are certain differences in the water conservation and NDVI values among the different ecosystem types, respectively. They also all satisfied the significance test of  $p < 0.05$ , which demonstrated that the stratified heterogeneity of water conservation and the NDVI at this level was significant.

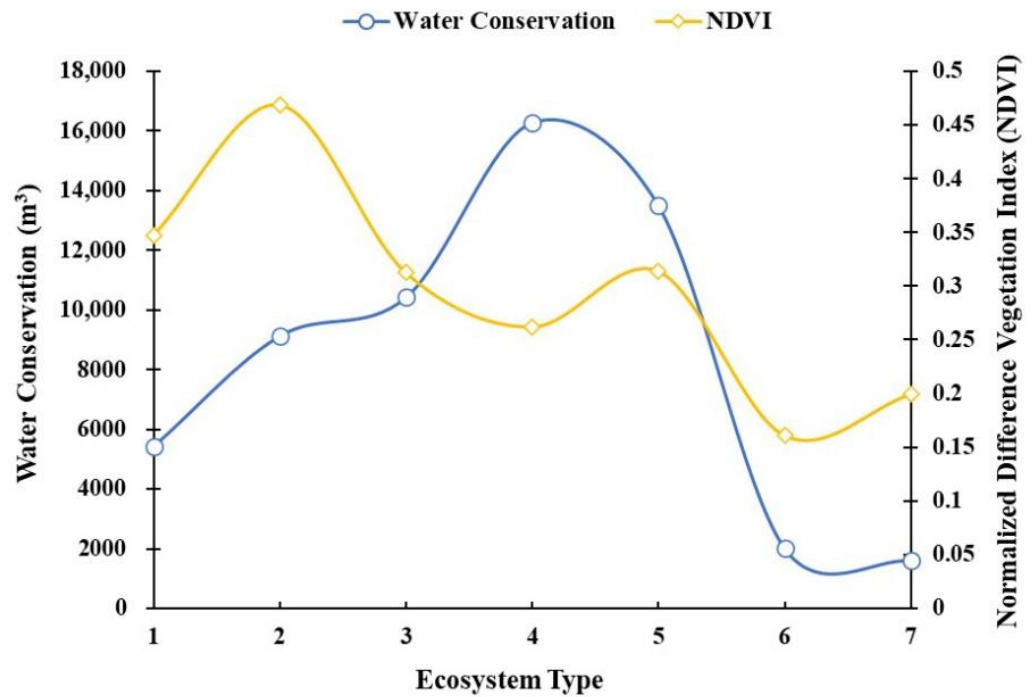
**Table 2.** Factor detector between water conservation and ecosystem type.

	q-Statistic	p-Value
Water Conservation	0.323	$2.766 \times 10^{-10}$
Mean NDVI	0.359	$9.609 \times 10^{-10}$

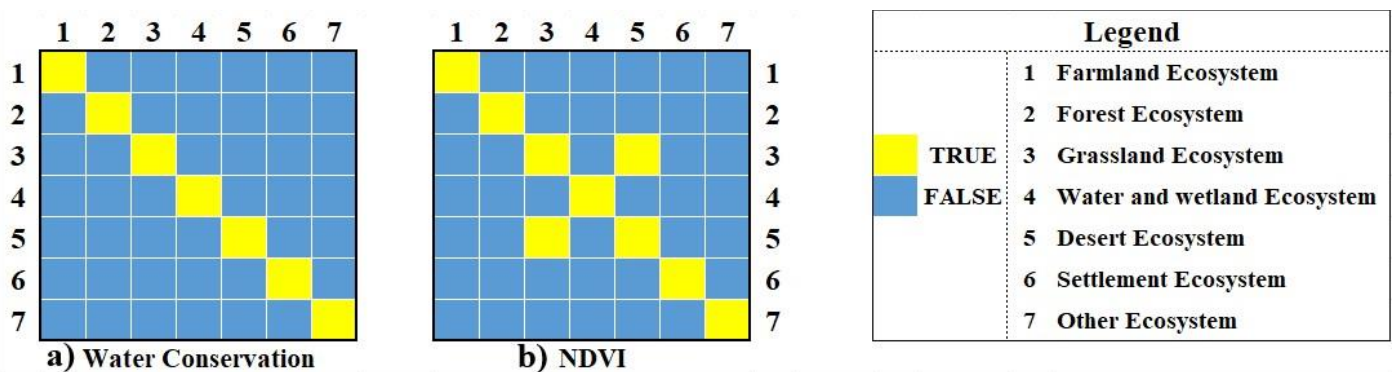
Note: NDVI, normalized difference vegetation index.

The average water conservation values were higher in the grassland, water and wetland, and desert ecosystems but lower in the farmland, settlement, and other ecosystems (Figure 5). Similarly, the same analysis provided slightly differences in the NDVI, whose mean values were higher in the farmland, forest, grassland, and desert ecosystems but were lower in the settlement and other ecosystems. From this result, it can be seen that water conservation and the NDVI in the grassland, desert, settlement, and other ecosystems varied in coordination—a high NDVI value was accompanied by having a high water conservation value, while a low NDVI value was accompanied by a low water conservation value. The other three ecosystems (farmland and forest, along with water and wetland ecosystems) had opposite values—high NDVIs contrasted with low water conservation, while low NDVIs were accompanied by high water conservation. In fact, it is not difficult to understand that the lush vegetation of farmland and forests can store water but, without adequate precipitation, would also deplete water resources. The water and wetland ecosystem has many areas of open water; therefore, the water conservation values must be very high, and the NDVI values will be reduced accordingly. However, vegetation is scarce in the desert, settlement, and other ecosystems. As long as vegetation exists, it will be able

to hold some water, which is why the water conservation values of these areas depend on the NDVI. These three ecosystems have less vegetation than the other areas, and therefore, they resulted in the lowest water conservation values. In summary, grasslands had the best water conservation in this study, and the existence of vegetation in a grassland can improve the water conservation.



**Figure 5.** The average value of water conservation and the normalized difference vegetation index (NDVI) in each stratum of the ecosystem types by risk detector. Notes: Ecosystem types 1–7 correspond to farmland, forest, grassland, water and wetland, desert, settlement, and other ecosystems, respectively.



**Figure 6.** Significant differences between the means of two ecosystem types for (a) water conservation and (b) the NDVI by a risk detector. “TRUE” indicates a significant difference exists ( $t$ -test with  $p < 0.05$ ); otherwise, it is “FALSE”.

Figure 6 showed the significant differences between water conservation and mean of the NDVI values for each pair of ecosystem types. As can be seen in Figure 6, for water conservation, a significant difference exists ( $p < 0.05$ ) between each two ecosystem types. This suggested that there are some differences in water conservation between different ecosystem types, which means water conservation has obvious spatial stratified heterogeneity in the seven different ecosystem types. For what concerns the mean of the NDVI, the significant difference between different strata is not as clear as for water

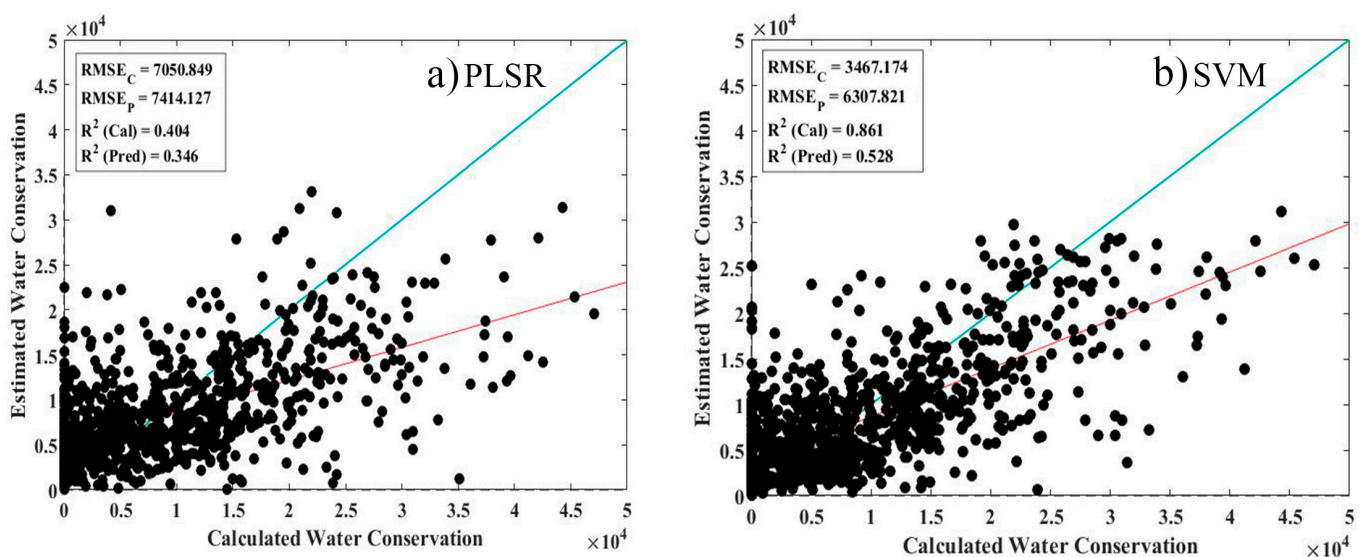
conservation. No significant difference exists between the grassland and desert ecosystems (Figure 6b). Simultaneously, although it is not possible to distinguish the grassland from the desert ecosystems in this study, based on the NDVI results by the risk detector (extensive greening has been carried out for desertification control), the effects of water conservation need to accumulate over a long period of time to make it easy to distinguish them through their difference in water conservation. This further indicated that water conservation has a cumulative effect over a long time. Despite this, the differences among the other types for NDVI were significant, indicating that NDVI has stratified heterogeneity among different ecosystem types. This result also verified the factor detector in Table 2.

In general, both water conservation and the NDVI have a certain amount of stratified heterogeneity among different ecosystem types, indicating that the water conservation and NDVI in different ecosystem types are different, which may lead to variances in their relationship. This can also be seen in Figure 4. Therefore, it is necessary to stratify the overall data based on the ecosystem type for modeling.

### 3.3. Prediction of Water Conservation by Different Predictive Models

#### 3.3.1. Overall Model Prediction Results

First, all sample points were divided into calibration (70%) and validation (30%) sets for PLSR and SVM modeling, respectively. The results are shown in Figure 7 and Table 3. The SVM model had superior calculation accuracy compared with the PLSR results. After testing the model with 30% of the data (validation data), it was found that the SVM method with  $R^2_P$  (the coefficient of determination for the predictive results) = 0.528 and  $RMSE_P$  (the root mean square error of the predictive results) = 6308 represented a higher predictive precision than the PLSR method. Simultaneously, a scatter diagram also showed that the scatter points were more aggregated around the fit line, which suggested that the SVM prediction point was closer to the calculated value of water conservation (Figure 7). However, even with this optimization, we can still see that there were many scattered points that were not well-fitted, suggesting that not all of the data was suitable for this model. It can be seen that stratified modeling of the overall data was necessary in this study.



**Figure 7.** Scatter plots of calculated and prediction values by (a) partial least squares regression (PLSR) and (b) a support vector machine (SVM) model using all the data. Note:  $RMSE_C$ , root mean square error (RMSE) of the calculated model,  $R^2_C$ , coefficient of determination for the calculated model,  $RMSE_P$ , RMSE of the predicted results, and  $R^2_P$ , coefficient of determination for the predicted results. The blue and red lines are 1:1 and the fitted lines, respectively.



**Table 3.** Accuracy assessment for the prediction models. RMSE: root mean square error and  $R^2$ : determination coefficient.

	Calculation Accuracy		Prediction Accuracy	
	RMSE	$R^2$	RMSE	$R^2$
Partial Least Squares Regression (PLSR)	7050.849	0.404	7414.127	0.346
Support Vector Machine (SVM)	3467.174	0.861	6307.821	0.528

### 3.3.2. Stratified Model Prediction Results

In order to explore the effects of stratified heterogeneity among different ecosystem types on the model accuracy, stratified PLSR and SVM models were conducted for different types. Figure 8 and Table 4 show the modeling effects of these two models for different types of ecosystem strata. Similarly, in general, the SVM fitting effect is better than the PLSR. However, for different ecosystem types, more detailed information revealed that the forest, grassland, and water and wetland ecosystems have better stratification effects than the others, and their accuracy exceeded the one of the overall fitting model. It can also be seen from the scatter plot that the prediction points of the forest, grassland, and water and wetland ecosystems are closer to the fitting line. Especially for the SVM method, the scatter points are basically concentrated along the fitting line. Unfortunately, the prediction accuracy of the stratified models in the farmland, desert, and other ecosystems were not high, and they were even lower than those of the overall data model.

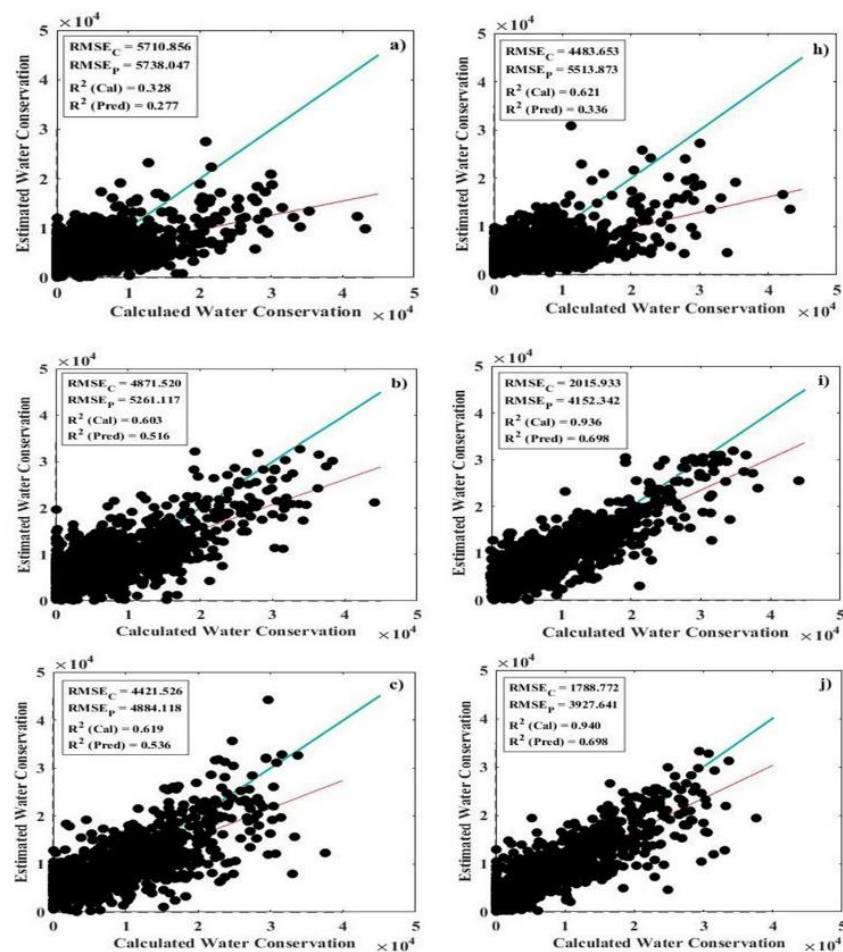
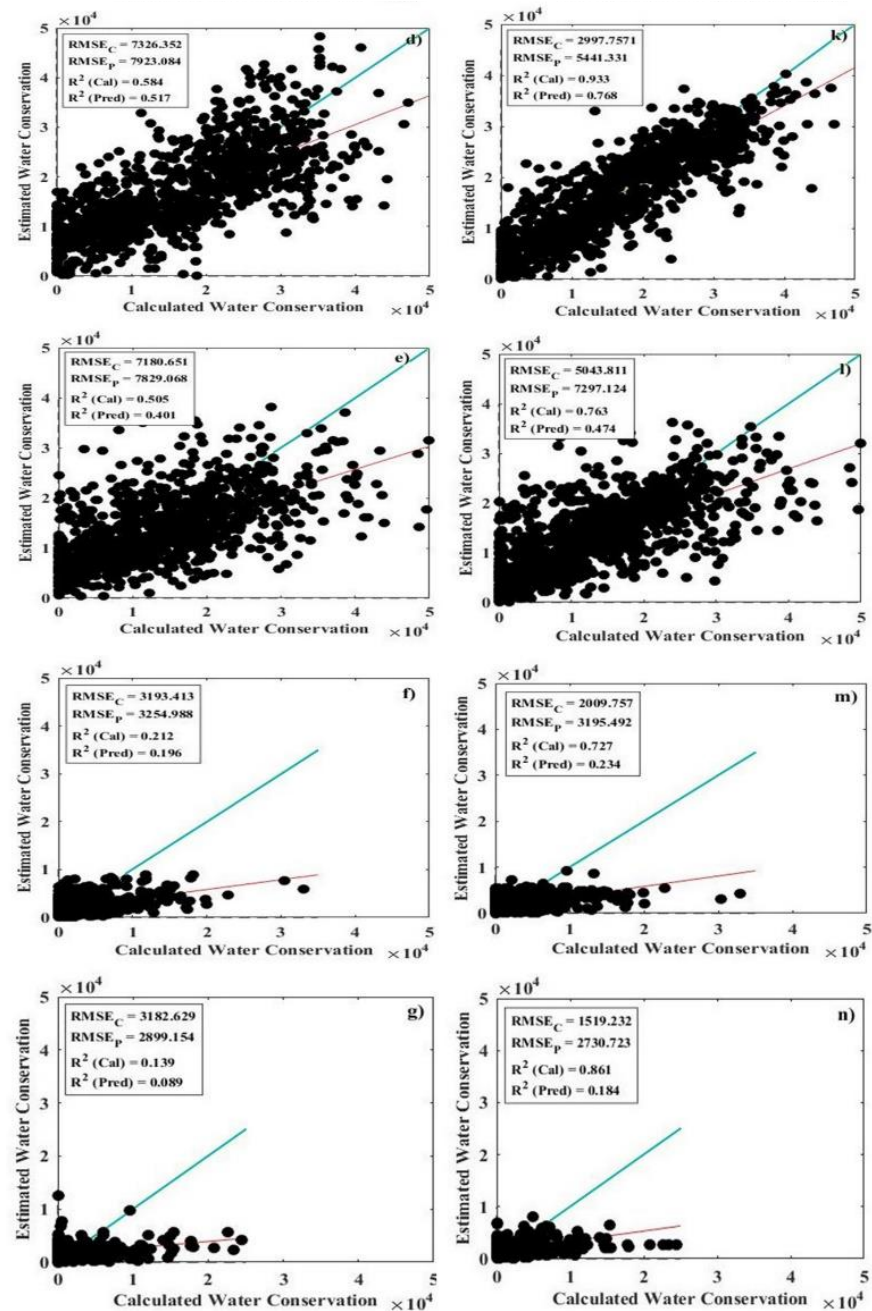


Figure 8. Cont.





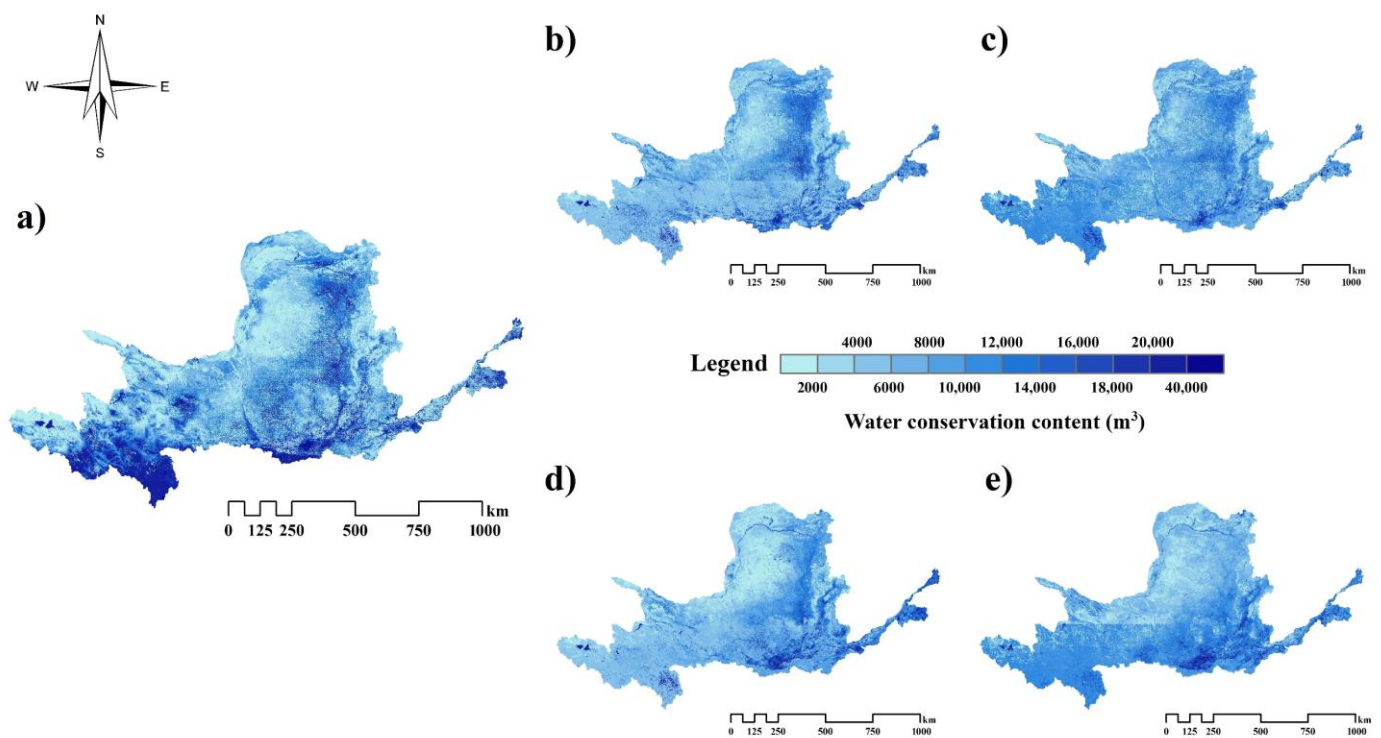
**Figure 8.** Scatter plots of the calculated and prediction values by the PLSR (partial least squares regression) and SVM (support vector machine) models using stratified data. Note: (a–g) represent the farmland, forest, grassland, water/wetland, desert, settlement, and other ecosystems, respectively by PLSR; (h–n) represent these seven ecosystems by the SVM.  $RMSE_C$ , root mean square error of the calculated model,  $R^2_C$ , coefficient of determination for the calculated model,  $RMSE_P$ , root mean square error of the predicted results, and  $R^2_P$ , coefficient of determination for the predicted results. The blue and red lines are 1:1 and the fitted lines, respectively.

**Table 4.** Accuracy assessment for the stratified prediction models.

Models	Ecosystem Type	Calculation Accuracy		Prediction Accuracy	
		RMSE	R <sup>2</sup>	RMSE	R <sup>2</sup>
Partial Least Squares Regression	Farmland	5710.856	0.328	5738.047	0.277
	Forest	4871.520	0.603	5261.117	0.516
	Grassland	4421.526	0.619	4884.118	0.536
	Water and wetland	7326.352	0.584	7923.084	0.517
	Desert	7180.651	0.505	7829.068	0.401
	Settlement	3193.413	0.212	3254.988	0.196
	Other ecosystems	3182.629	0.139	2899.154	0.089
Support Vector Machine	Farmland	4483.653	0.621	5513.873	0.336
	Forest	2015.933	0.936	4152.342	0.698
	Grassland	1788.772	0.940	3927.641	0.698
	Water and wetland	2997.757	0.933	5441.331	0.768
	Desert	5043.811	0.763	7297.124	0.474
	Settlement	2009.757	0.727	3195.492	0.234
	Other ecosystems	1519.232	0.861	2730.723	0.184

### 3.4. Digital Mapping of Water Conservation

The spatial output, in the form of digital water conservation maps, were mapped using the aforementioned models (Figure 9). Broadly speaking, the maps were somewhat similar, and all maps showed a strong spatial variation of water conservation. High predicted water conservation values occurred in the western, southern, and eastern parts of the study area. These results were basically consistent with the spatial distribution calculated by the water conservation formula, although slight differences were found in some details. First, it is possible to see that the four simulation results are not significantly different from the water conservation results calculated by the original formula (Figure 9a) in all the areas except the southwestern part of the study area. For the four simulation results, higher values were suppressed in the southwestern part, and the lower water conservation values were found there. Secondly, in other areas except the southwest corner, the spatial pattern of water conservation simulated by the stratified SVM method (Figure 9b) was the closest to the original results. As for the stratified PLSR method (Figure 9c), its results presented slightly higher water conservation values in the western part of the study area. Moreover, for what concerns the overall models (Figure 9d,e), both the SVM and PLSR had lower water conservation values in the entire study area when compared with other results. By further comparing the results of the SVM (Figure 9d) and PLSR (Figure 9e) in the overall model, it can be seen that the value in the western part of the study region was lower when using the overall SVM method, while its distribution pattern of water conservation in the eastern part was closer to the original results than that of the overall PLSR method. In summary, these simulation results can well-estimate the original distribution of water conservation in most areas, among which the hierarchical SVM method performed best.

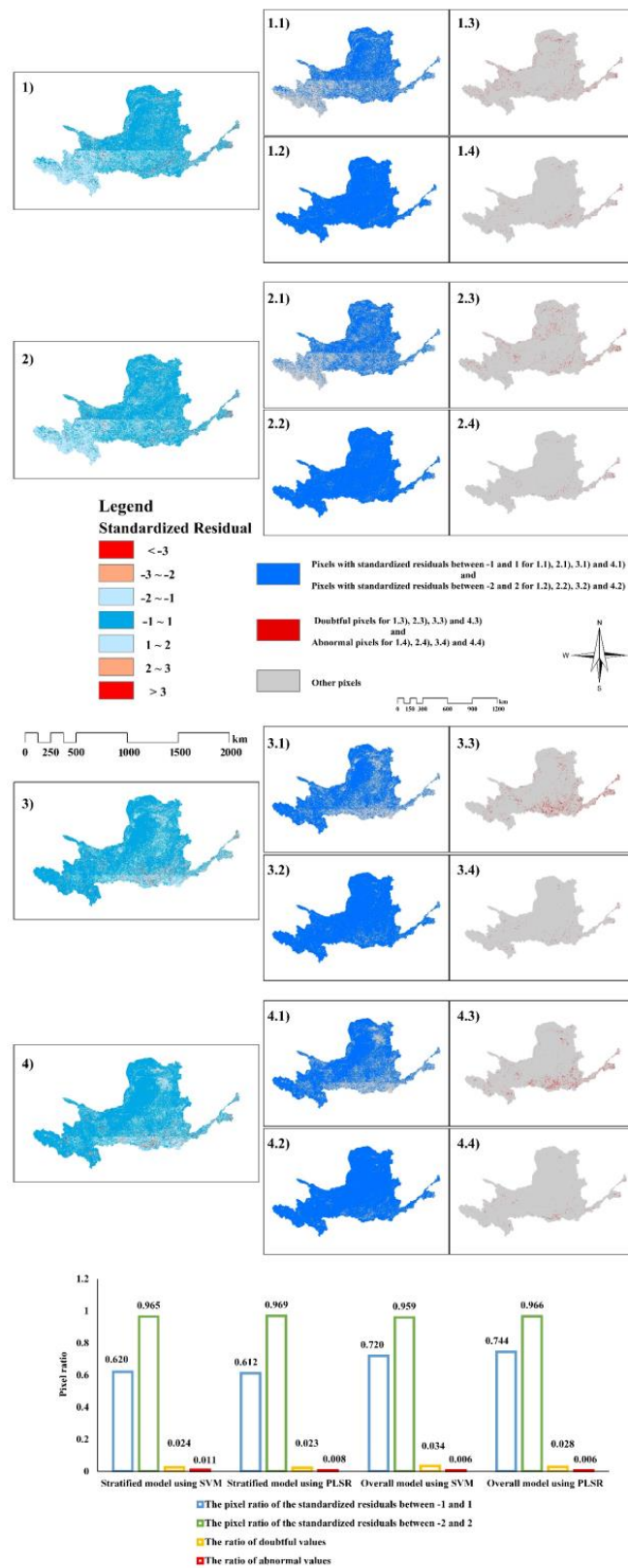


**Figure 9.** Spatial distribution of the estimated water conservation by: (a) the original water conservation formula, (b) the stratified method using a support vector machine (SVM), (c) the stratified method using partial least squares regression (PLSR), (d) the overall method using an SVM, and (e) the overall method using PLSR.

## 4. Discussion

### 4.1. Comparison of the Estimated Results of Different Models

The present study involved an overall data fitting estimation and a stratified data fitting estimation for water conservation via the PLSR and SVM models, respectively. Generally, all these methods could explain the spatial variance of water conservation to some extent. In fact, the mapping results using the stratified model was not very different from the overall model, as shown by the standardized residuals maps (Figure 10 (1, 2, 3, and 4)). They all showed that more than 95.45% of all the mapping pixels were between  $-2$  and  $2$ , indicating that the model provided good results. Nevertheless, a closer look at the details showed that the stratified model can increase the proportion of standardized residuals in the range of  $-2$  to  $2$ , while the ratio is lower with doubtful values (Figure 10 1.2, 2.2, 3.2, 4.2, and the statistical chart). However, its ratio of standardized residuals between  $-1$  and  $1$  (0.620 of SVM and 0.612 of PLSR; pixel ratio  $< 0.6827$ ) was not as high as for the overall model (0.720 of SVM and 0.744 of PLSR; pixel ratio  $> 0.6827$ ). In addition, the proportion of abnormal values was also higher than for the overall model (Figure 10 1.1, 2.1, 3.1, 4.1, and the statistical chart). Despite this, from the residual analysis of these four results, it is feasible to apply the NDVI time series data to estimate the water conservation.



**Figure 10.** The spatial distribution of the standardized residual of the four estimated results. Notes: 1, 2, 3, and 4 corresponded to the stratified support vector machine (SVM), stratified partial least squares regression (PLSR), overall SVM, and overall PLSR methods, respectively; in addition, x.1, x.2, x.3, and x.4 represent the spatial distribution of the standardized residuals between  $-1$  and  $1$ , between  $-2$  and  $2$ , doubtful pixels, and abnormal pixels, respectively, for the above four methods.

It is worth paying attention to that, after implementing SVM and PLSR for mapping, the results were also somewhat surprising. In Section 3.3.1, the overall data fitting results showed that NDVI time series data in the PLSR model can explain 0.346 of the water conservation, but the SVM can explain 0.528, which indicated that the SVM was better than the PLSR. This was also evident in the stratified model, where the fitting results of the SVM performed better than those of the PLSR in almost every ecosystem type (Table 4 and Figure 8). However, when mapping the entire study region using the SVM model, compared with the PLSR model, the SVM showed some uncertainty of estimation when using a large amount of data. Based on Figure 10 (3.3, 3.4, and the statistical chart), the pixel ratio of the SVM mapping results with the standardized residuals between  $-2$  and  $2$  in the overall method was lower than that of the PLSR, and the ratio of the doubtful and abnormal values were higher than the PLSR. Fortunately, this uncertainty was markedly reduced in the stratified method, as can be seen in Figure 10 (1.3, 2.3, and the statistical chart). This can be interpreted as a reduction of the amount of data by the stratified model during each operation, thus ensuring a better performance of the SVM.

From the perspective of the spatial distribution of the standardized residuals, it was not difficult to find that most areas fell within the interval of  $-2$  to  $2$ . When the requirements of accuracy were further improved, in the stratified model, parts of southwest were excluded from the interval of  $-1$  to  $1$ , while, in the overall model, some parts of the south and northeast were also excluded. As for the spatial distribution of the doubtful and abnormal values, all results obtained by the four methods showed an analogous and a relatively balanced pattern.

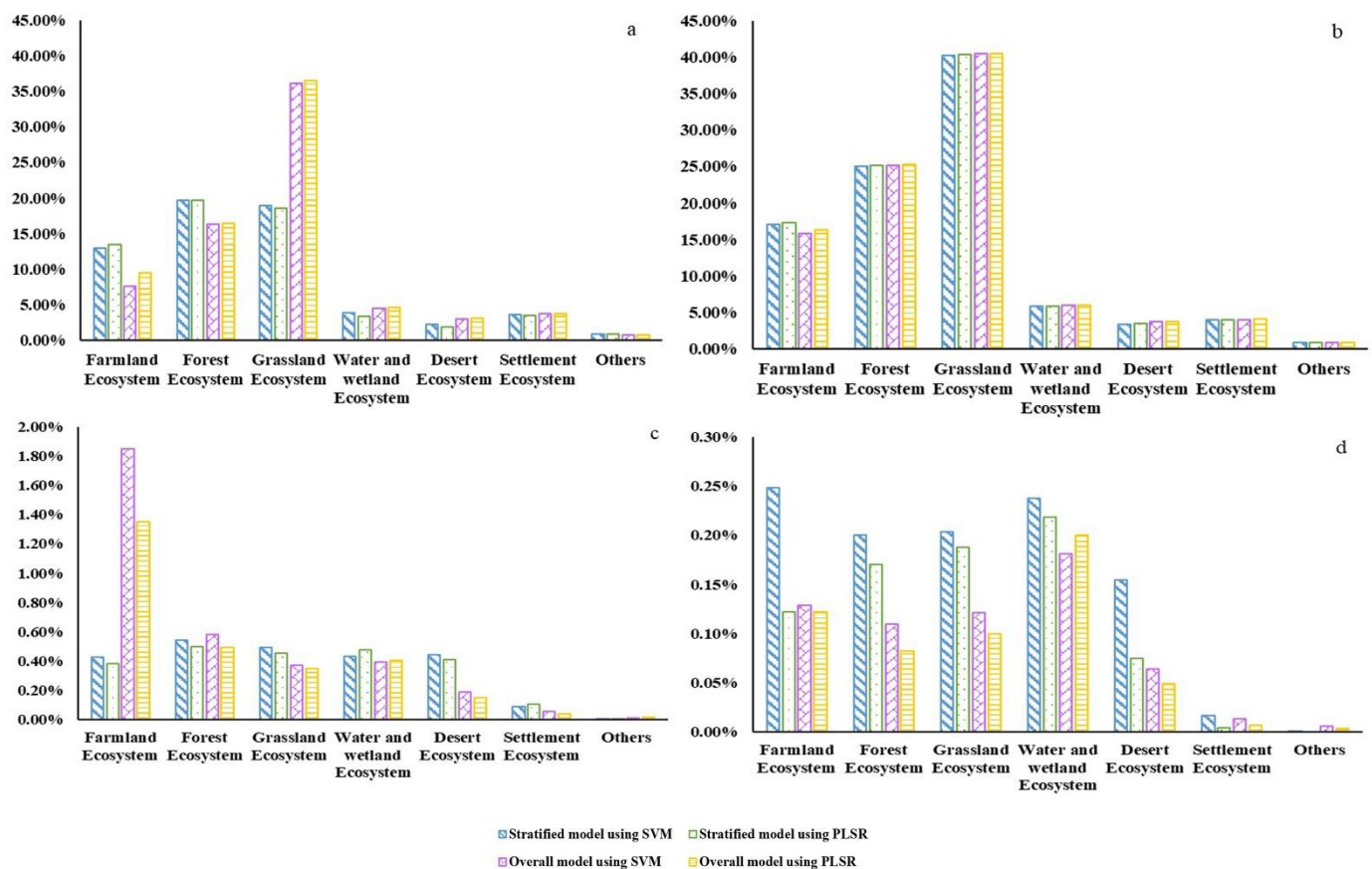
In general, a good estimate of water conservation can be achieved using the NDVI time series data with these four methods. Each of them has its own advantages. The stratified methods can make the standardized residual of the pixel values more concentrated in the interval of  $-2$  to  $2$  and reduce the percentage of the doubtful values. The overall methods are conducive to concentrating more data between  $-1$  and  $1$  and reducing the percentage of abnormal values. Then, the SVM can produce a model with excellent fitting, but it is not suitable for estimation of large areas and big data. In contrast, although the PLSR has a weak performance in fitting the model, it can withstand the test of big data operation simulation and achieve good results.

#### 4.2. Residual Interpretation of Estimated Results under Different Ecosystem Types

In this study, the overall and stratified models were both applied to verify whether the stratified heterogeneity of water conservation and the NDVI would affect the model estimation effect. The results suggest that the stratified heterogeneity did lead to the differentiation of the fitting effects among different ecosystem strata. Compared with the overall model, the stratified model was more explicit in the applicability of using the NDVI time series to estimate water conservation in different ecosystem types. Simultaneously, in order to inspect the errors generated in the mapping results by the above models among different ecosystem types, the statistics of the standardized residuals among the different ecosystem types were also analyzed. Figure 11 shows the ratio of the pixel number with the standardized residuals between  $-1$  and  $1$  (Figure 11a),  $-2$  and  $2$  (Figure 11b), doubtful values (Figure 11c), and abnormal values (Figure 11d) in the different ecosystem types to the total pixels in the entire study area. It can be discovered that the standardized residuals between  $-2$  and  $2$  were mainly derived from the farmland, forest, and grassland ecosystems; the grassland ecosystem, in particular, had the largest proportion (Figure 11b). In this subfigure (Figure 11b), the distribution of the standardized residuals ( $-2$  to  $2$ ) of both the overall and stratified models were not significantly different among each ecosystem type; however, the situation was different when the interval was reduced to  $-1$  to  $1$  (Figure 11a). In the interval of  $-1$  to  $1$  of the standardized residuals, the first three ecosystems still made major contributions in both the overall and stratified methods. Nevertheless, when compared with the stratified model, the overall model highlighted the contribution of the grassland ecosystem and inhibited the contribution of the farmland



ecosystem. Obviously, in this interval, the stratified model gave play more evenly to the contribution of the three ecosystem types having greater amounts of vegetation cover.



**Figure 11.** The ratio of the number of standardized residuals of each ecosystem type in the four intervals to the total pixels in the entire study region by the stratified support vector machine (SVM), stratified partial least squares regression (PLSR), overall SVM, and overall PLSR models, respectively: (a) using the standardized residuals in the range of  $-1$  to  $1$ , (b) using the standardized residuals in the range of  $-2$  to  $2$ , (c) showing the proportion of doubtful values, and (d) showing the proportion of abnormal values.

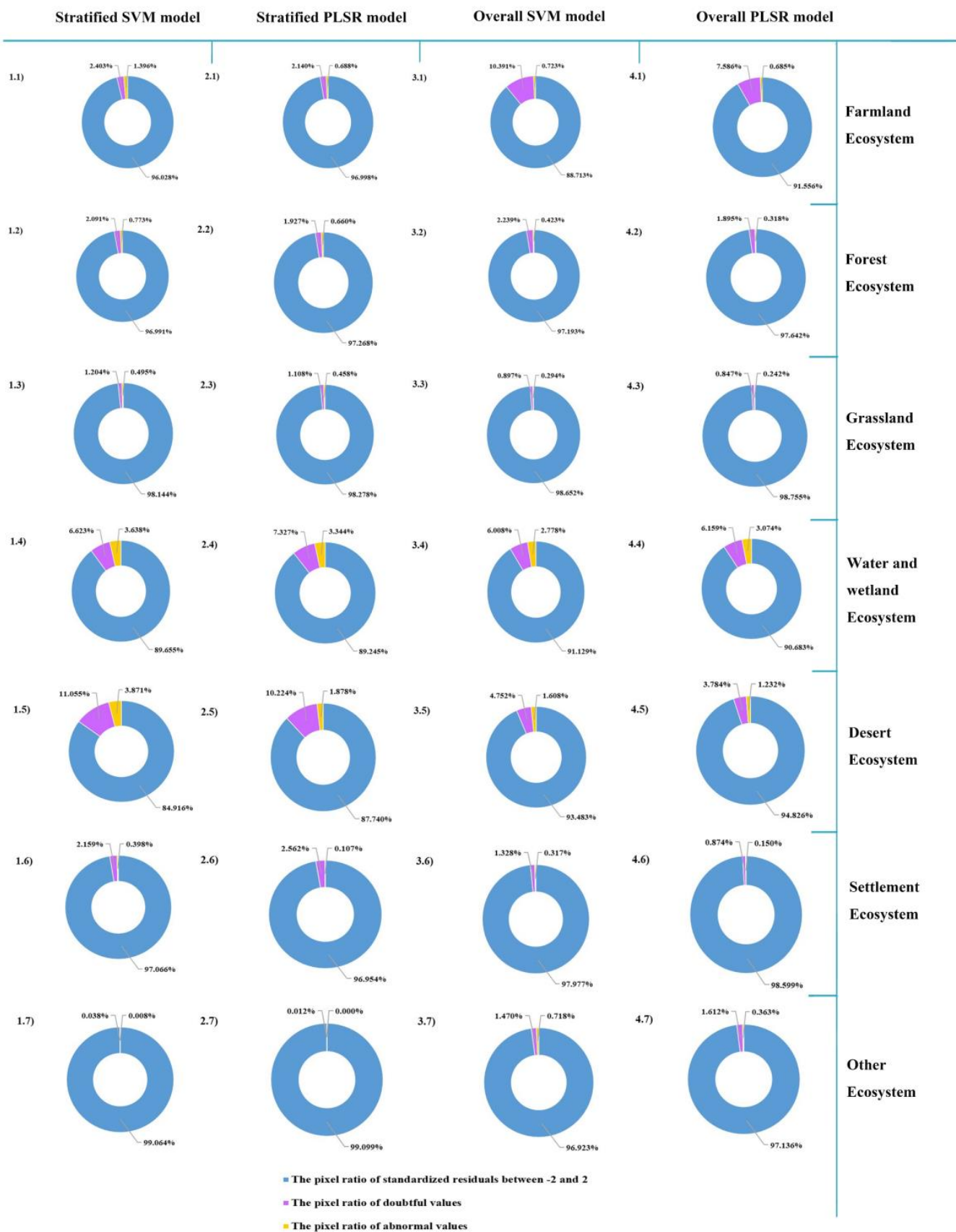
For doubtful values, a little difference was observed between the overall and stratified methods, except that the overall model resulted in more doubtful pixels in farmland. This result can also be mutually verified with the previous results (Figure 7 and Table 3), by which the overall model may not be able to give full play to the estimation ability of the NDVI time series data in farmland on water conservation. The distribution trends for the abnormal values of the two methods among the ecosystem types were generally consistent, but the proportions of the overall models were smaller. In addition, it should be mentioned that no matter which interval or which model was used, the difference between the SVM and PLSR was not very obvious.

In general, the ecosystems with better vegetation conditions, such as forests and grasslands, had better results for water conservation mapping using the NDVI time series, which is especially true for the stratified models (Figure 8 and Table 4). Surprisingly, the above models on the fitting effects performed almost as well in the water and wetland ecosystem as in these vegetation-covered areas, even sometimes exceeding their accuracy (Table 4). This result can be inferred from the correlation between the NDVI time series and water conservation in the water and wetland ecosystem, which promoted the fitting results by a relatively stronger continuous complete negative correlation (Figure 4). However, this good effect of the water/wetland ecosystem was not apparent when mapping in a large scale (Figure 11). Moreover, for the desert, settlement, and other ecosystems with poor

vegetation conditions, this approach was not suitable; sometimes, the stratified model was even less effective than the overall model. This was reflected in both fitting the estimation results and in the mapping effect. The stratified method was more suitable for individual vegetation-covered ecosystem types, while its estimation of water conservation within a large area of mixed ecosystem types was less effective than the overall method. This is also easy to understand. After all, the stratified model mainly excavates the interpretation degree of the NDVI time series data in each ecosystem for water conservation. However, in the overall model, because all the data are not stratified, the parts with a better fitting effect can improve the accuracy of the overall model, depending on the function of an optimal fitting algorithm. In addition, it is worth mentioning that, among the first three ecosystem types, grassland had the best accuracy, forest came second, and farmland was relatively worse. This may be explained by the fact that, in grassland ecosystems, the presence of vegetation helps to conserve water but not in forest ecosystems. In fact, many studies have suggested that forests may consume more water than grasslands [11,17,48]. When the vegetation in a forest is very dense, it needs to absorb a large amount of water to support its growth; however, the larger leaf area leads to more water evaporation. Therefore, despite its high NDVI values, the ability of a forest to conserve water is not as good as the one of a grassland.

Due to the great difference in the number of pixels for each ecosystem type, the results may be biased to only consider the ratio of the number of pixels in each interval of the different ecosystem types to the total number of pixels in the entire area. Based on this, the ratio of standardized residuals of the different intervals within each ecosystem were calculated (Figure 12). First, for the stratified method, the ratio of standardized residuals in the interval of  $-2$  to  $2$  within the farmland, forest, grassland, settlement, and other ecosystems were all  $> 95.45\%$ , indicating that this method is feasible for water conservation mapping in these types. Similarly, these types also passed the  $95.45\%$  test in the overall model, except for the farmland ecosystem. It can be also found from Figure 12 (3.1 and 4.1) that the proportions of doubtful and abnormal values of farmland in the overall model were significantly larger than that of the stratified model. As previously concluded, the overall model was less able to optimize the data for the estimation of water conservation in the farmland ecosystem. Secondly, for the water and wetland ecosystem and the desert ecosystem, the ratio of standardized residuals between  $-2$  and  $2$  did not reach  $95.45\%$ ; in addition, the ratios of doubtful and abnormal values within these two ecosystems were higher than in other ecosystems, suggesting this method may not be suitable for these two ecosystems. It is interesting to note that, as mentioned above, the models in the water/wetland ecosystem often achieved a good fitting effect. However, it can be deduced from the results of the standardized residual analysis (Figure 12) that this method is not recommended for mapping water conservation in a large-scale water/wetland ecosystem with a large amount of running data. Finally, as before, the results of the SVM and PLSR methods in this part of the analysis were slightly, but not obviously, different.

Analyzing the proportion of standardized residuals within each ecosystem can facilitate a better realization of the feasibility and applicability of the water conservation mapping in every ecosystem using NDVI time series data. Although it was found from the model fitting accuracy (Figure 8 and Table 4) that not all of the ecosystems were suitable for this method (especially those without vegetation cover, which had low accuracy), when applied to large-scale mapping, this method showed that most ecosystem were reasonable when using this approach.



**Figure 12.** The ratio of pixel numbers of the standardized residuals of the seven ecosystem types analyzed here in each interval to the total pixel numbers in each ecosystem type.

#### 4.3. The Basic Theory of Water Conservation Estimation Using the NDVI

It is well-known that the NDVI can represent the growth status, coverage, and biomass of vegetation very well. For most vegetation-covered areas, such as grassland and forest, compared with other land types, these vegetation types can intercept precipitation and store water [8]. Thus, the presence of vegetation increases the surface roughness and reduces the surface runoff, allowing the system to preserve more water [49]. In addition, if the land surface is covered by vegetation for a long time, the physical and chemical properties of the soil will change, which is conducive to the adsorption and storage of water by soil. From this perspective, vegetation-covered areas are more capable of conserving water than areas with less vegetation. In other words, the NDVI is related to water conservation and can be used to indicate the amount of water conservation in an area to a certain extent. In response to this conjecture, the capacity of several major ecosystem types to conserve water based on NDVI time series data was estimated and verified in this study. The above results validate this idea quite well. The estimated results were fairly excellent for the areas covered by vegetation, but for the areas with scarce or no vegetation cover, some uncertainty exists about whether this idea was valid. Therefore, we suggest that in grasslands and forests with a good vegetation status, the water conservation can be directly estimated by using NDVI time series data, which avoids the need to collect large amounts of complex environmental data that are required for other methods. In these areas, the advantages of the NDVI, such as being easily accessible, having simple calculations, and providing full coverage with suitable spatial and temporal resolutions, the direct use of the NDVI time series for predicting water conservation can allow researchers to quickly obtain satisfactory and acceptable results.

In addition, it is worth mentioning that some researchers believe that water conservation is related to topography [9,50–54]. In this paper, the estimation of water conservation based on the NDVI time series data did not take topography into account. However, surprisingly, the distribution of errors was not directly related to the terrain, as seen from the standardized residuals. Therefore, in this study, from the view that topography is not the main factor influencing the error control of the water conservation function, for now, we did not consider the topography. Of course, this does not mean that the topographic factors are not important for water conservation. In fact, considering the terrain is of great significance in the relevant research related to water conservation, and the effects of the terrain area are also worth further exploring in the future.

#### 4.4. Inadequacies and Limitations

The present study has some shortcomings. For example, unfortunately, based on the digital maps (Figure 9), although our estimated results were very similar to the results of the original method, gaps still remained (such as in the southwest corner of the study area); these may be caused by data errors and need to be further explored. In addition, although we were trying to improve the prediction accuracy of these models (the accuracy did improve after several improvements), it is undeniable that none of our fitting results exceeded 0.8. There are many reasons for this. First of all, the accuracy validation we made was a relative comparison with the calculation results of the traditional water balance equation, which is likely to result in errors. Secondly, there were many uncertain factors influencing the water conservation. For example, the loss, storage, and evaporation of water all have different complex mechanisms in different land surfaces and ecosystems. Even in those good-fitting effect areas with vegetation coverage, the interaction among the soil, vegetation, and water is also very complex. Thus, from this perspective, remote sensing is limited in the prediction of water conservation. However, the standardized residual analysis showed that these methods passed the verification and qualified for spatial mapping of water conservation in the study area. The remote sensing product time series data can really solve the problem of it being difficult to collect and obtain accurate data for the complex modeling of traditional methods effectively. Especially for large-scale



regional ecological planning and management, which requires rapid mapping of the water conservation but relatively low precision, this is particularly applicable.

In general, the practical significance of our research is considered from two aspects: (1) exploring whether the NDVI time series can play an important role as an alternative factor when mapping large-scale water conservation and (2) when land surface is covered by vegetation, it can provide a method to indirectly evaluate water conservation via the vegetation index time series data. All in all, the above results demonstrated the feasibility of our method. The results of the present study will be helpful towards the scientific evaluation and quantitative management of regional water resources. However, we have to suggest that this paper only used the NDVI to prove the effect and function of the long time series of the vegetation index on water conservation, while, in fact, there may be other better vegetation indexes waiting to be applied in the future. In addition, only the PLSR and SVM were used in this study, and more effective methods should be studied in the following research to obtain better accuracy.

## 5. Conclusions

In this study, NDVI time series data were employed to model and analyze water conservation via the PLSR and SVM models by performing both overall and stratified fitting. The overall fitting results show that the NDVI in the PLSR model can explain 0.346 of water conservation, but the SVM can explain 0.528. The stratified results, considering stratified heterogeneity, showed that, for the areas with good vegetation cover (such as forest and grass ecosystems), adopting this method will achieve better results in the estimation fitting of water conservation. However, areas with severe vegetation degradation or no vegetation cover may result in instability and uncertainty with this method. From the obtained spatial digital maps of water conservation, these methods can also describe the variance in water conservation. In addition, the ability to use a long NDVI time series in the spatial simulation of water conservation was also confirmed. This will be of great significance for the scientific study of water conservation in the future.

**Author Contributions:** Conceptualization, Y.Z., J.D., and L.G.; methodology, Y.Z.; software, Y.Z.; validation, Y.Z.; formal analysis, Y.Z. and J.D.; investigation, Z.S. and J.Z.; resources, J.D. and J.W.; writing—original draft preparation, Y.Z.; writing—review and editing, Y.Z., J.D., and L.G.; project administration, J.D.; and funding acquisition, J.D. All authors have read and agreed to the published version of the manuscript.

**Funding:** This study was supported by the Special Fund for Basic Scientific Research Business of Central Public Research Institutes (Grant No.2019YSKY-017), the National Key Research and Development Program of China (Grant No. 2016YFC0500401-5) and the National Natural Science Foundation of China (Grant No. 41001055).

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The authors would like to acknowledge Xu Chengdong and Wang Jinfeng, who provided the open access software Geodetector and related theoretical guidance. In addition, we are also very grateful for the data processed and shared by Moderate-resolution Imaging Spectroradiometer (MODIS) and Global Precipitation Measurement (GPM). We thank LetPub ([www.letpub.com](http://www.letpub.com), accessed on 18 February 2021) for its linguistic assistance during the preparation of this manuscript.

**Conflicts of Interest:** The authors declare no conflict of interest.



## Appendix A

Table 1. The model estimated results of the different sample points for each ecosystem.

Sample Number for Each Ecosystem	Overall Model Accuracy				Stratified Model Accuracy				
	RMSE		R <sup>2</sup>		RMSE		R <sup>2</sup>		
	PLSR	SVM	PLSR	SVM	PLSR	SVM	PLSR	SVM	
200					6833 <sup>a</sup>	6261 <sup>a</sup>	0.188 <sup>a</sup>	0.328 <sup>a</sup>	
					5203 <sup>b</sup>	4406 <sup>b</sup>	0.485 <sup>b</sup>	0.627 <sup>b</sup>	
					4804 <sup>c</sup>	3809 <sup>c</sup>	0.511 <sup>c</sup>	0.685 <sup>c</sup>	
		7877	6576	0.237	0.462	8245 <sup>d</sup>	6036 <sup>d</sup>	0.438 <sup>d</sup>	0.690 <sup>d</sup>
					7778 <sup>e</sup>	7490 <sup>e</sup>	0.446 <sup>e</sup>	0.485 <sup>e</sup>	
					3471 <sup>f</sup>	17608 <sup>f</sup>	0.053 <sup>f</sup>	0.007 <sup>f</sup>	
					3498 <sup>g</sup>	3311 <sup>g</sup>	0.070 <sup>g</sup>	0.160 <sup>g</sup>	
					6460 <sup>a</sup>	6125 <sup>a</sup>	0.193 <sup>a</sup>	0.261 <sup>a</sup>	
					5362 <sup>b</sup>	4697 <sup>b</sup>	0.551 <sup>b</sup>	0.655 <sup>b</sup>	
					4968 <sup>c</sup>	4258 <sup>c</sup>	0.520 <sup>c</sup>	0.644 <sup>c</sup>	
300					8456 <sup>d</sup>	6240 <sup>d</sup>	0.439 <sup>d</sup>	0.687 <sup>d</sup>	
		7662	6575	0.278	0.462	7846 <sup>e</sup>	7087 <sup>e</sup>	0.443 <sup>e</sup>	0.544 <sup>e</sup>
					3412 <sup>f</sup>	3191 <sup>f</sup>	0.065 <sup>f</sup>	0.164 <sup>f</sup>	
					2995 <sup>g</sup>	2825 <sup>g</sup>	0.075 <sup>g</sup>	0.154 <sup>g</sup>	
					5738 <sup>a</sup>	5514 <sup>a</sup>	0.277 <sup>a</sup>	0.336 <sup>a</sup>	
					5261 <sup>b</sup>	4152 <sup>b</sup>	0.516 <sup>b</sup>	0.698 <sup>b</sup>	
					4884 <sup>c</sup>	3928 <sup>c</sup>	0.536 <sup>c</sup>	0.698 <sup>c</sup>	
		7414	6308	0.346	0.528	7923 <sup>d</sup>	5441 <sup>d</sup>	0.517 <sup>d</sup>	0.768 <sup>d</sup>
					7829 <sup>e</sup>	7297 <sup>e</sup>	0.401 <sup>e</sup>	0.474 <sup>e</sup>	
					3254 <sup>f</sup>	3195 <sup>f</sup>	0.196 <sup>f</sup>	0.234 <sup>f</sup>	
500					2899 <sup>g</sup>	2731 <sup>g</sup>	0.089 <sup>g</sup>	0.184 <sup>g</sup>	
					5755 <sup>a</sup>	5483 <sup>a</sup>	0.230 <sup>a</sup>	0.298 <sup>a</sup>	
					5112 <sup>b</sup>	– <sup>b</sup>	0.548 <sup>b</sup>	– <sup>b</sup>	
					4961 <sup>c</sup>	3882 <sup>c</sup>	0.546 <sup>c</sup>	0.725 <sup>c</sup>	
		7456	6089	0.367	0.576	7849 <sup>d</sup>	– <sup>d</sup>	0.543 <sup>d</sup>	– <sup>d</sup>
					7336 <sup>e</sup>	6666 <sup>e</sup>	0.488 <sup>e</sup>	0.578 <sup>e</sup>	
					3230 <sup>f</sup>	3094 <sup>f</sup>	0.143 <sup>f</sup>	0.220 <sup>f</sup>	
					2722 <sup>g</sup>	2568 <sup>g</sup>	0.109 <sup>g</sup>	0.197 <sup>g</sup>	

Note: “a”–“g” represents the farmland, forest, grassland, water/wetland, desert, settlement, and other ecosystems, respectively, in the stratified model. RMSE, root mean square error, R<sup>2</sup>, coefficient of determination, PLSR, partial least squares regression, and SVM, support vector machine. “–” means the running time exceeded the allowable time of the program, which was unable to get results due to too many sample points.

## References

1. Xi, J. Speech at the Symposium on Ecological Protection and High-Quality Development in the Yellow River Basin. *QiuShi*. **2019**, *20*, 1–2.
2. General Office of Ministry of Environmental Protection in China. *Guidelines for Setting Red Lines for Ecological Protection*; General Office of Ministry of Environmental Protection in China: Beijing, China, 2017.
3. General Office of the CPC Central Committee & the State Council. *A Guideline on an Ecological “Red Line”*; General Office of Ministry of Environmental Protection in China: Beijing, China, 2017.
4. He, P.; Gao, J.; Zhang, W.; Rao, S.; Zou, C.; Du, J.; Liu, W. China integrating conservation areas into red lines for stricter and unified management. *Land Use Policy* **2018**, *71*, 245–248. [[CrossRef](#)]
5. Yin, Y.; Wu, S.; Zhang, D.; Dai, E. Ecosystem Water Conservation Changes in Response to Climate Change in the Source Region of the Yellow River from 1981 to 2010. *Geogr. Res.* **2016**, *35*, 49–57.
6. Shihan, G.; Yang, X.; Hua, Z.; Yi, X.; Zhiyun, O. Spatial patterns of ecosystem water conservation in China and its impact factors analysis. *Acta Ecol. Sin.* **2017**, *37*, 2455–2462. [[CrossRef](#)]
7. Yu, N.; Li, C.; Xu, J.; Lu, S.; Fu, Y.; Wang, W.; Wang, Y. Water Conservation Function of Polar Plantations on Lowland in Yellow River (in Chinese). *J. Soil Water Conserv.* **2009**, *23*, 61–65.
8. Zhang, B.; Li, W.; Xie, G.; Xiang, Y. Water Conservation Function and Its Measurement Methods of Forest Ecosystem. *Chin. J. Ecol.* **2009**, *28*, 529–534.

9. Lü, L.; Ren, T.; Sun, G.; Defeng, Z. Spatial and Temporal Changes of Water Supply and Water Conservation Function in Sanjiangyuan National Park from 1980 to 2016. *Acta Ecol. Sin.* **2020**, *40*, 993–1003.
10. Bai, S.; Jing, L.; Li, H.; Feng, W. The Demarcation of Ecological Protection Red Line Based on Water Conservation Function. *Ecol. Environ. Sci.* **2017**, *26*, 1665–1670. [[CrossRef](#)]
11. Fu, B.; Burgher, I. Riparian vegetation NDVI dynamics and its relationship with climate, surface water and groundwater. *J. Arid. Environ.* **2015**, *113*, 59–68. [[CrossRef](#)]
12. Bao, G.; Qin, Z.; Bao, Y.; Zhou, Y.; Li, W.; Sanjiv, A. NDVI-Based Long-Term Vegetation Dynamics and Its Response to Climatic Change in the Mongolian Plateau. *Remote Sens.* **2014**, *6*, 8337–8358. [[CrossRef](#)]
13. Du, J.; Shu, J.; Yin, J.; Yuan, X.; Jiaerheng, A.; Xiong, S.; He, P.; Liu, W. Analysis on spatio-temporal trends and drivers in vegetation growth during recent decades in Xinjiang, China. *Int. J. Appl. Earth Obs. Geoinf.* **2015**, *38*, 216–228. [[CrossRef](#)]
14. Fensholt, R.; Proud, S.R. Evaluation of Earth Observation based global long term vegetation trends—Comparing GIMMS and MODIS global NDVI time series. *Remote Sens. Environ.* **2012**, *119*, 131–147. [[CrossRef](#)]
15. Petus, C.; Lewis, M.; White, D. Monitoring temporal dynamics of Great Artesian Basin wetland vegetation, Australia, using MODIS NDVI. *Ecol. Indic.* **2013**, *34*, 41–52. [[CrossRef](#)]
16. Aguilar, C.; Zinnert, J.C.; Polo, M.J.; Young, D.R. NDVI as an indicator for changes in water availability to woody vegetation. *Ecol. Indic.* **2012**, *23*, 290–300. [[CrossRef](#)]
17. Joiner, J.; Yoshida, Y.; Anderson, M.; Holmes, T.; Hain, C.; Reichle, R.; Koster, R.; Middleton, E.; Zeng, F.-W. Global relationships among traditional reflectance vegetation indices (NDVI and NDII), evapotranspiration (ET), and soil moisture variability on weekly timescales. *Remote Sens. Environ.* **2018**, *219*, 339–352. [[CrossRef](#)]
18. Lamchin, M.; Lee, W.-K.; Jeon, S.W.; Wang, S.W.; Lim, C.H.; Song, C.; Sung, M. Long-term trend and correlation between vegetation greenness and climate variables in Asia based on satellite data. *Sci. Total Environ.* **2018**, *618*, 1089–1095. [[CrossRef](#)] [[PubMed](#)]
19. Tong, X.; Wang, K.; Yue, Y.; Brandt, M.; Liu, B.; Zhang, C.; Liao, C.; Fensholt, R. Quantifying the effectiveness of ecological restoration projects on long-term vegetation dynamics in the karst regions of Southwest China. *Int. J. Appl. Earth Obs. Geoinf.* **2017**, *54*, 105–113. [[CrossRef](#)]
20. Chen, T.; De Jeu, R.; Liu, Y.; Van Der Werf, G.; Dolman, A. Using satellite based soil moisture to quantify the water driven variability in NDVI: A case study over mainland Australia. *Remote Sens. Environ.* **2014**, *140*, 330–338. [[CrossRef](#)]
21. Wang, J.F.; Li, X.H.; Christakos, G.; Liao, Y.L.; Zhang, T.; Gu, X.; Zheng, X.Y. Geographical detectors-based health risk assessment and its application in the neural tube defects study of the Heshun region, China. *Int. J. Geogr. Inf. Sci.* **2010**, *24*, 107–127. [[CrossRef](#)]
22. Wang, J.-F.; Zhang, T.-L.; Fu, B.-J. A measure of spatial stratified heterogeneity. *Ecol. Indic.* **2016**, *67*, 250–256. [[CrossRef](#)]
23. Zhu, L.; Meng, J.; Zhu, L. Applying Geodetector to disentangle the contributions of natural and anthropogenic factors to NDVI variations in the middle reaches of the Heihe River Basin. *Ecol. Indic.* **2020**, *117*, 106545. [[CrossRef](#)]
24. Zhao, R.; Zhan, L.; Yao, M.; Yang, L. A geographically weighted regression model augmented by Geodetector analysis and principal component analysis for the spatial distribution of PM2.5. *Sustain. Cities Soc.* **2020**, *56*, 102106. [[CrossRef](#)]
25. Peng, W.; Kuang, T.; Tao, S. Quantifying influences of natural factors on vegetation NDVI changes based on geographical detector in Sichuan, western China. *J. Clean. Prod.* **2019**, *233*, 353–367. [[CrossRef](#)]
26. Ma, Q.; Long, Y.; Jia, X.; Wang, H.; Li, Y. Vegetation response to climatic variation and human activities on the Ordos Plateau from 2000 to 2016. *Environ. Earth Sci.* **2019**, *78*, 709. [[CrossRef](#)]
27. Steve, R. Modis Global Evapotranspiration Project (Mod16). Available online: <https://lpdaac.usgs.gov/products/mod16a2v006/> (accessed on 18 February 2021).
28. Steven, W.R.; Qiaozhen, M.; Maosheng, Z.; Moreno, A. *User's Guide Modis Global Terrestrial Evapotranspiration (Et) Product*; University of Montana: Missoula, MT, USA, 2019.
29. George, J.H.; Bolvin, D.T.; Braithwaite, D.; Hsu, K.; Joyce, R.; Kidd, C.; Nelkin, E.J.; Sorooshian, S.; Tan, J.; Xie, P. *Algorithm Theoretical Basis Document (Atbd) Version 06 Nasa Global Precipitation Measurement (Gpm) Integrated Multi-Satellite Retrievals for Gpm (Imerg)*; National Aeronautics and Space Administration: Washington, DC, USA, 2019.
30. Huffman, G.J.; Stocker, E.F.; Bolvin, D.T.; Nelkin, E.J.; Jackson, T. GPM IMERG Final Precipitation L3 1 Month 0.1 Degree X 0.1 Degree V06. Available online: [https://disc.gsfc.nasa.gov/datasets/GPM\\_3IMERGM\\_06/summary?keywords=Hydrology](https://disc.gsfc.nasa.gov/datasets/GPM_3IMERGM_06/summary?keywords=Hydrology) (accessed on 18 February 2021).
31. Kamel, D.; Munoz, A.B.; Ramon, S.; Huete, A. *Modis Vegetation Index User's Guide*; Vegetation Index and Phenology Lab of The University of Arizona, The University of Arizona: Tucson, AZ, USA, 2015.
32. Didan, K. Modis/Terra Vegetation Indices 16-Day L3 Global 250 M Sin Grid. Available online: <https://ladsweb.modaps.eosdis.nasa.gov/search/order/1/MOD13Q1--6> (accessed on 18 February 2021).
33. Lü, Y.; Hu, J.; Sun, F.; Zhang, L.W. Water Retention and Hydrological Regulation: Harmony but Not the Same in Terrestrial Hydrological Ecosystem Services. *Acta Ecol. Sin.* **2015**, *35*, 5191–5196. [[CrossRef](#)]
34. Qian, J.; Shu, W. Mountain Sponge City Construction Monitoring Based on Surface Runoff Coefficient. *Beijing Surv. Mapp.* **2019**, *33*, 647–651.
35. Huang, S.; Chen, C. Surface Runoff Analysis on Farmland Predipitation in Huangpu River Basin. *Shanghai Environ. Sci.* **1998**, *17*, 21–23+44.

36. Wang, J.-F.; Xu, C. Geodetector: Software for Measure and Attribution of Stratified Heterogeneity (Sh). Available online: <http://www.geodetector.cn/> (accessed on 18 February 2021).
37. Wold, S.; Martens, H.; Wold, H. The multivariate calibration problem in chemistry solved by the PLS method. In *Springer Texts in Business and Economics*, 1st ed.; Kågström, B., Ruhe, A., Eds.; Springer Science and Business Media LLC: Berlin/Heidelberg, Germany, 1983; Volume 973, pp. 286–293.
38. Axelsson, C.; Skidmore, A.K.; Schlerf, M.; Fauzi, A.; Verhoef, W. Hyperspectral analysis of mangrove foliar chemistry using PLSR and support vector regression. *Int. J. Remote Sens.* **2012**, *34*, 1724–1743. [[CrossRef](#)]
39. Wold, S.; Ruhe, A.; Wold, H.; Dunn, I.W.J. The Collinearity Problem in Linear Regression. The Partial Least Squares (PLS) Approach to Generalized Inverses. *SIAM J. Sci. Stat. Comput.* **1984**, *5*, 735–743. [[CrossRef](#)]
40. Wold, S.; Sjöström, M.; Eriksson, L. PLS-regression: A basic tool of chemometrics. *Chemom. Intell. Lab. Syst.* **2001**, *58*, 109–130. [[CrossRef](#)]
41. Vapnik, V.N. *The Nature of Statistical Learning Theory*; Springer: New York, NY, USA, 1995.
42. Holtgrave, A.-K.; Förster, M.; Greifeneder, F.; Notarnicola, C.; Kleinschmit, B. Estimation of Soil Moisture in Vegetation-Covered Floodplains with Sentinel-1 SAR Data Using Support Vector Regression. *PFG J. Photogramm. Remote Sens. Geoinf. Sci.* **2018**, *86*, 85–101. [[CrossRef](#)]
43. Taghizadeh-Mehrjardi, R.; Nabiollahi, K.; Kerry, R. Digital mapping of soil organic carbon at multiple depths using different data mining techniques in Baneh region, Iran. *Geoderma* **2016**, *266*, 98–110. [[CrossRef](#)]
44. Were, K.; Bui, D.T.; Dick, Ø.B.; Singh, B.R. A comparative assessment of support vector regression, artificial neural networks, and random forests for predicting and mapping soil organic carbon stocks across an Afrotropical landscape. *Ecol. Indic.* **2015**, *52*, 394–403. [[CrossRef](#)]
45. Shrestha, N.; Shukla, S.K. Support vector machine based modeling of evapotranspiration using hydro-climatic variables in a sub-tropical environment. *Agric. For. Meteorol.* **2015**, *200*, 172–184. [[CrossRef](#)]
46. Ghorbani, M.A.; Shamshirband, S.; Haghi, D.Z.; Azani, A.; Bonakdari, H.; Ebtehaj, I. Application of firefly algorithm-based support vector machines for prediction of field capacity and permanent wilting point. *Soil Tillage Res.* **2017**, *172*, 32–38. [[CrossRef](#)]
47. Fan, J.; Wu, L.; Zhang, F.; Cai, H.; Wang, X.; Lu, X.; Xiang, Y. Evaluating the effect of air pollution on global and diffuse solar radiation prediction using support vector machine modeling based on sunshine duration and air temperature. *Renew. Sustain. Energy Rev.* **2018**, *94*, 732–747. [[CrossRef](#)]
48. Li, S.; Liang, W.; Fu, B.; Lü, Y.; Fu, S.; Wang, S.; Su, H. Vegetation changes in recent large-scale ecological restoration projects and subsequent impact on water resources in China’s Loess Plateau. *Sci. Total. Environ.* **2016**, *569–570*, 1032–1039. [[CrossRef](#)]
49. Zhou, J.; Gao, J.; Gao, Z.; Yang, W. Analyzing the Water Conservation Service Function of the Forest Ecosystem. *Acta Ecol. Sin.* **2018**, *38*, 1679–1686.
50. Wang, Y.; Zhang, C.; Liu, C.; Zen, L. Research on the Pattern and Change of Forest Water Conservation in Three-North Shelterbelt Forest Program Region, China. *Acta Ecol. Sin.* **2019**, *39*, 5847–5856. [[CrossRef](#)]
51. Dong-Qing, L.; Er-Jia, C.; Jin-Xi, Z.; Jie, G.; Ling-Ling, Y. Spatiotemporal pattern of water conservation and its influencing factors in Bailongjiang Watershed of Gansu. *J. Nat. Resour.* **2020**, *35*, 1728–1743. [[CrossRef](#)]
52. Ding, C.; Zhang, H.; Li, X.; Li, W.; Gao, Y. Quantitative Assessment of Water Conservation Function of the Natural Spruce Forest in the Central Tianshan Mountains: A Case Study of the Urumqi River Basin. *Acta Ecol. Sin.* **2019**, *37*, 3733–3743. [[CrossRef](#)]
53. Hu, J.; Lyu, Y.; Zhang, K.; Tao, Y.Z.; Li, T.; Ren, Y. The Differences of Water Conservation Function under Typical Vegetation Types in the Pailugou Catchment, Qilian Mountain, Northwest China. *Acta Ecol. Sin.* **2016**, *36*, 3338–3349. [[CrossRef](#)]
54. Hong, Y.; Yu, S.; Yang, X.; Luo, Y. Study on Spatial-Temporal Changes and Driving Factors of Water Conservation in Hengduan Mountain Region. *Geomat. Spat. Inf. Technol.* **2019**, *42*, 72–76.