



Article

Nitrogen and Phosphorus Retention Risk Assessment in a Drinking Water Source Area under Anthropogenic Activities

Yuexin Zheng ¹, Qianyang Wang ¹, Xuan Zhang ¹, Jingshan Yu ^{1,*}, Chong Li ¹, Liwen Chen ^{2,3} and Yuan Liu ¹

¹ College of Water Sciences, Beijing Normal University, Beijing 100875, China; yxzheng@mail.bnu.edu.cn (Y.Z.); 201931470001@mail.bnu.edu.cn (Q.W.); xuan@bnu.edu.cn (X.Z.); chong@mail.bnu.edu.cn (C.L.); 201921470018@mail.bnu.edu.cn (Y.L.)

² School of Geomatics and Prospecting Engineering, Jilin Jianzhu University, Changchun 130118, China; chenliwen@iga.ac.cn

³ State Key Laboratory of Earth Surface Processes and Resource Ecology, Beijing Normal University, Beijing 100875, China

* Correspondence: jingshan@bnu.edu.cn; Tel.: +86-588-07814

Abstract: Excessive nitrogen (N) and phosphorus (P) input resulting from anthropogenic activities seriously threatens the supply security of drinking water sources. Assessing nutrient input and export as well as retention risks is critical to ensuring the quality and safety of drinking water sources. Conventional balance methods for nutrient estimation rely on statistical data and a huge number of estimation coefficients, which introduces uncertainty into the model results. This study aimed to propose a convenient, reliable, and accurate nutrient prediction model to evaluate the potential nutrient retention risks of drinking water sources and reduce the uncertainty inherent in the traditional balance model. The spatial distribution of pollutants was characterized using time-series satellite images. By embedding human activity indicators, machine learning models, such as Random Forest (RF), Support Vector Machine (SVM), and Multiple Linear Regression (MLR), were constructed to estimate the input and export of nutrients. We demonstrated the proposed model's potential using a case study in the Yanghe Reservoir Basin in the North China Plain. The results indicate that the area information concerning pollution source types was effectively established based on a multi-temporal fusion method and the RF classification algorithm, and the overall classification low-end accuracy was 92%. The SVM model was found to be the best in terms of predicting nutrient input and export. The determination coefficient (R^2) and Root Mean Square Error (RMSE) of N input, P input, N export, and P export were 0.95, 0.94, 0.91, and 0.93, respectively, and 32.75, 5.18, 1.45, and 0.18, respectively. The low export ratios (2.8–3.0% and 1.1–2.2%) of N and P, the ratio of export to input, further confirmed that more than 97% and 98% of N and P, respectively, were retained in the watershed, which poses a pollution risk to the soil and the quality of drinking water sources. This nutrient prediction model is able to improve the accuracy of non-point source pollution risk assessment and provide useful information for water environment management in drinking water source regions.

Keywords: drinking water source; retention risk; remote sensing; nutrient prediction; support vector machine; random forest; human activity



Citation: Zheng, Y.; Wang, Q.; Zhang, X.; Yu, J.; Li, C.; Chen, L.; Liu, Y. Nitrogen and Phosphorus Retention Risk Assessment in a Drinking Water Source Area under Anthropogenic Activities. *Remote Sens.* **2022**, *14*, 2070. <https://doi.org/10.3390/rs14092070>

Academic Editor: Mingliang Liu

Received: 21 February 2022

Accepted: 23 April 2022

Published: 26 April 2022

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



Copyright: © 2022 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

Over the last 30 years, with the rapid economic and social development in China [1], drinking water sources have been heavily polluted by human activities [2,3]. Non-point source pollution (NPSP) is considered to be an important factor involved in the aforementioned problems in China [4]. According to the results of the first national pollution survey [5], more than half of the total nitrogen (TN) and total phosphorus (TP) discharged into the environment were produced by NPSP. The excessive input of nitrogen (N) and phosphorus (P) as a result of anthropogenic activities has caused water quality degradation

in rivers [6] and increased eutrophication [7,8]. Concurrently, the nutrient budget, meaning the surplus between input and export, retained in the watershed also pollutes the soil and aquifers [9]. Therefore, it is vital to quantify the inputs and exports of anthropogenic N and P.

The input and export of N and P are two important components when quantifying nutrient budgets [10]. The NPSP load entering the basin through leaching and runoff is subsequently exported into the river systems, which is known as the export of nutrients [10]. N and P inputs strongly affect exports [11]. The ratio of export to input is called the nutrient export ratio, and it reflects the capacity for nutrient retention and the potential risk of pollution in the riverine system [12,13]. Studies have shown that human activities can significantly affect the nutrient export ratio [14]. For example, the N export ratio in the Taihu Lake Basin has increased from 18% to 30% in the last 30 years as a result of urbanization and population growth, and the N budget has changed accordingly [13]. Hence, it is important to accurately and quantitatively estimate and predict the nutrient retention potential risk in riverine systems affected by human activities.

Model simulations are an important technical basis for the quantitative estimation and risk assessment of nutrients [15]. Physical-based models, such as the Soil and Water Assessment Tool, can simulate the physical process of pollutants by using a large number of parameters [16]. However, abundant parameters limit model application in data deficient areas [17,18]. The nutrient balance model is widely used in estimating N and P loads [19]. The majority of studies integrate the Net Anthropogenic Nitrogen Input (NANI), the Net Anthropogenic Phosphorus Input (NAPI), and the Export Coefficient Model (ECM) methods to assess nutrient input and export, and their potential risks within the watershed system [20]. For example, Lian et al. [13] first integrated the NANI and ECM models to evaluate the N load at the county level in the Taihu Lake Basin and found that urbanization and population growth are the main factors disturbing the nitrogen budget. Deng et al. [15] quantified the contribution rate of various factors to the NANI and NAPI models and constructed the management mechanism of nutrient diversity. These studies demonstrate that the nutrient balance model is a robust empirical model and is an effective predictor of N and P load [13]. Moreover, the model can be easily used to assess nutrient variation in environmental systems, depending on whether their input variables (for example, N and P fertilization, atmospheric nitrogen fixation, crop nitrogen fixation) and export variables (domestic sewage, garbage, rural excrement, urban residents, livestock, and land use) are specific [21]. However, the estimation of these nutrient balance variables is an imprecise process involving heavy calculations [22]. For instance, nutrient balance models usually rely on statistical data and abundant estimation coefficients, and the differences in nutrient estimation coefficients have a certain impact on the model results [13]. Secondly, statistical data are usually on the county level spatial scale, which poses a certain challenge in terms of accurately describing the spatial distribution of pollutants [23].

Considering the aforementioned problems, machine learning models based on time series satellite images are regarded as a useful tool [24,25]. Firstly, machine learning models have a high simulation accuracy and fast training speed [24]. This is important because it overcomes the problem whereby physical models are inadequate in data-deficient areas but also does not need to consider the influence of complex underlying surface characteristics on the estimated coefficients of the traditional model [26,27]. Secondly, remote sensing monitoring is an effective means to quantitatively evaluate pollutant exports and portray their spatial changes [28] and is often used to identify and monitor NPSP exports [20]. Human activities are the key factors affecting the input and export of nutrients [12]. Therefore, a prediction model embedded with human activity indicators is necessary for estimating the input and export of nutrients.

The major objectives of this study were: (1) to build a nutrient prediction model embedded in human activity indicators to predict the input and export of nutrients; (2) to describe the spatial distribution of pollutant input and export; (3) to evaluate the potential risk of N and P retention in the watershed as a result of human activities.

2. Materials

2.1. Study Area

The Yanghe River Basin is located in Qinhuangdao City, Hebei Province (39°N–40°N; 118°E–119°E) (Figure 1a,b) [29]. The Yanghe Reservoir is situated in the Yanghe River Basin and has a total storage capacity of 0.353 billion m³ and a controlled drainage area of 755 km² (Figure 1c) [30]. It is an important drinking water source with significant ecological and socio-economic value [31] and provides the domestic water supply for Qinhuangdao city. There are four main upstream tributaries in the Yanghe Reservoir Basin, namely, the Miwu River, Xiyang River, Dongyang River, and Maguying river (Figure 1c). This study area covers 3 counties, 7 towns, and 242 villages. The Yanghe Reservoir basin is an important grain production region, in which corn and peanuts are primarily cultivated. The basin is located in a warm temperate monsoon climate area with a mild climate type, distinct seasons, and an annual average rainfall of 750 mm [31,32]. The land-use types in the study area include cropland, forest land, bare land, water, and urban areas.

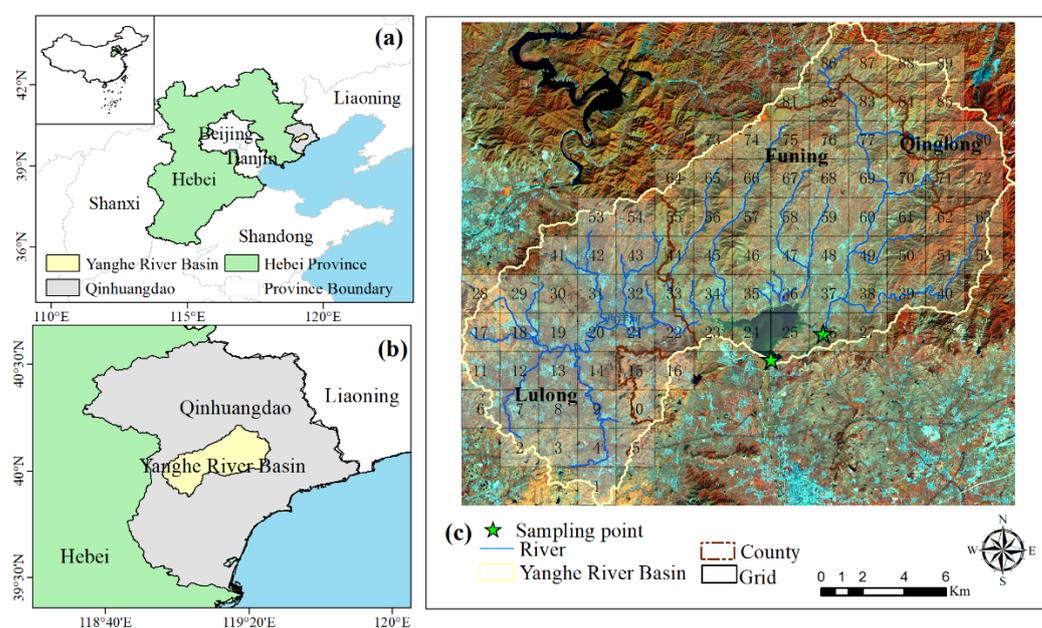


Figure 1. The location and basic characteristics of the Yanghe River Basin. (a) Geographical locations of the Yanghe River Basin in the Hebei province of China; (b) geographical location of the Yanghe River Basin in Qinhuangdao City; (c) 10 m resolution remote sensing image of the Yanghe River Basin.

According to the water quality monitoring data from Hebei Environmental Protection Administration, the main pollutants are N and P. TN and TP seriously exceed the standard; this is especially true of TN, the levels of which are categorized as being inferior to Class V (higher than 2 mg L⁻¹) [33]. Additionally, large quantities of pesticides, chemical fertilizers, livestock manure, and garbage enter the reservoir area, resulting in continuous deterioration of the water quality of the Yanghe reservoir and the intensification of eutrophication, which seriously threatens the water supply security.

2.2. Data Sources

2.2.1. Statistic Data

The county-level data between 2004 and 2015 were obtained from the annals of statistics and include N and P fertilization, population density, planting area, livestock, and poultry numbers, population numbers, and crop yields. These data were used to calculate NANI, NAPI, and ECM. The raster data of the atmospheric nitrogen deposition with a spatial resolution of 0.25° from 2004 to 2015 were obtained from the Regional Emission Inventory in Asia 2.1 [34]. Among them, the majority of relevant parameters involved in

the study were published in the bulletin of national economic and social statistics. Table 1 shows the variables used in this study and their acronyms.

Table 1. List of abbreviations for variables.

Variable	Abbr.
nitrogen fertilizer application	Nfer
atmospheric nitrogen deposition	Ndep
crop nitrogen fixation	Nfix
net food/feed imports of nitrogen	Nim
human nitrogen consumption	Nhum consumption
livestock nitrogen consumption	Nliv consumption
nitrogen content of livestock products	Nliv products
nitrogen content of crop products	Ncro products
phosphorus fertilizer application	Pfer
percentage of crop area in each grid	crop (%)
percentage of urban area in each grid	Urban (%)
non-food phosphorus input	Pnon
net food/feed phosphorus imports	Pim
human phosphorus consumption	Phum consumption
livestock phosphorus consumption	Pliv consumption
phosphorus content of livestock products	Pliv products
phosphorus content of crop products	Pcro products
nitrogen fertilizer used per cultivated area	Ferc N
phosphorus fertilizer used per cultivated area	Ferc P
percentage of urban land area in the total area	Urbanization (%)
percentage of forest area in each grid	Forest (%)
nitrogen and phosphorus	N and P

2.2.2. Remote Sensing Data

In the study, 94 images, taken between 2004 and 2015, were processed online using the GEE platform, specifically including multi-spectral data from the Landsat 5/8 and Sentinel 2 satellites (Table 2). As a result of the failure of the Landsat-7ETM+ Scan Lines Corrector, the data strip of the acquired image was lost, which seriously affected the use of Landsat ETM remote sensing images. Therefore, the year 2012 was not considered in this study. All of the remote sensing images were used to estimate the export and describe the spatial distribution of pollutants.

Table 2. Details of remote sensing data and satellite images used in the study.

Satellite	Sensor	Year	Spectral Bands (Numbers)	Spatial Resolution	Number of Images
Landsat5	TM	2004–2011	7	30 m	56
Landsat8	OLI	2013–2015	8	30 m	35
Sentinel-2	MSI	2015	13	10 m	3

3. Methods

The overall objective was to build a convenient, reliable, and accurate annual-scale nutrient prediction model to predict nutrient input and export, and to assess the potential risk of nutrient retention in the watershed. The framework of this study is shown in Figure 2. Firstly, we integrated traditional balance models, including the NANI, NAPI, and ECM models, and estimated the input and export of nutrients. Secondly, the Pearson correlation coefficient was used to select the human activity indicators that have an obvious impact on nutrient input and export, replacing the input variables of the traditional balance model. Subsequently, a nutrient prediction model based on the SVM, RF, and MLR algorithms was established, and the human activity indicators were used as the input data of the machine learning model to predict the input and export of nutrients. Thereafter, on the basis of

Google Earth Engine (GEE), the area information of land-use types in the Yanghe Reservoir Basin was extracted using the multi-temporal fusion method and RF algorithm, and the spatial distribution of nutrient input and export was described. Finally, on the basis of the prediction results, we evaluated the potential risk of N and P retention in the watershed.

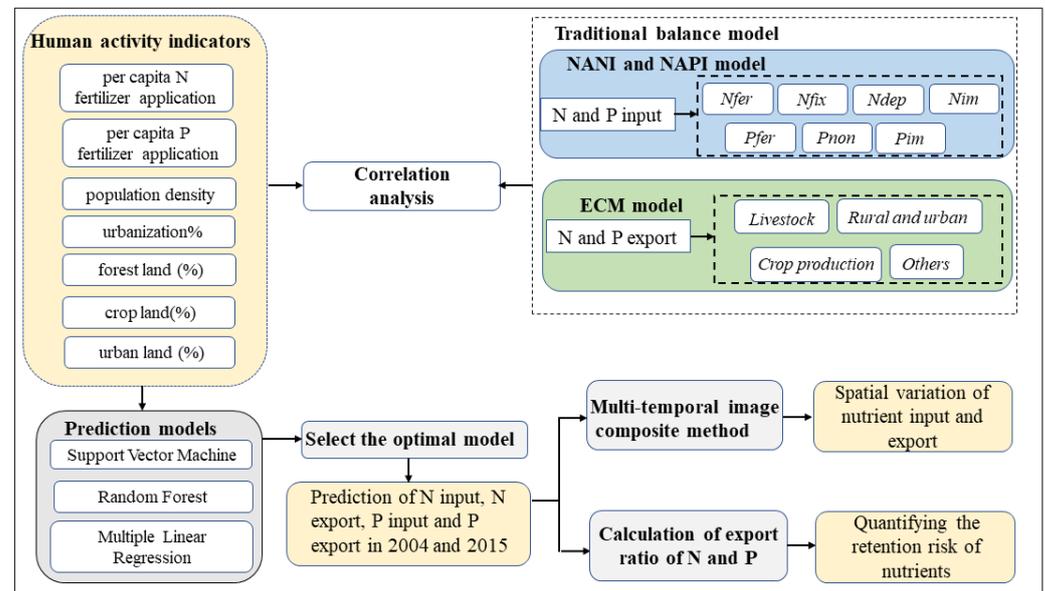


Figure 2. Flowchart of this study.

3.1. Preprocessing

The land-use types have different ground features in different periods, which has a certain impact on the amount of N and P exported to the river outlet. We classified different land-use types in the study area from 2004 to 2015 for March to October based on the multi-temporal image fusion and RF classification algorithms (see the Appendices A.1 and A.2 for specific operations).

To describe the spatial distribution of pollutants, we converted the statistical data into human activity indicators on a spatial scale based on the area information of land-use types. Finally, the county-level statistical data were rasterized to a 3 km resolution for running the machine learning algorithm. The “create fishnet” tool in the data management tools, which is a part of ArcGIS desktop software version 10.4, was used for the spatial analysis of data. By drawing fishing nets, we were able to count the number of elements occupied by the grid and then analyze the spatial distribution characteristics of the data [35,36].

3.2. Nutrient Input and Export Estimation Based on the Traditional Balance Model

Traditional balance models, including the input model (NANI and NAPI), and the export model (ECM), were used to estimate the N and P input into the watershed and exported into the river, respectively. On the one hand, the results of the traditional model were used to select the human activity indicators that have a significant impact on nutrient input and export; on the other hand, they were used as validation data for the prediction model. The calculation process of NANI are represented in Equations (1) and (2):

$$\text{NANI} = \text{Nfer} + \text{Ndep} + \text{Nfix} + \text{Nim} \quad (1)$$

$$\text{Nim} = \text{Nhum consumption} + \text{Nliv consumption} - \text{Nliv products} - \text{Ncro products} \quad (2)$$

The NANI model is mostly composed of four trails, which include nitrogen fertilization (Nfer), atmospheric nitrogen deposition (Ndep), crop nitrogen fixation (Nfix), and net food/feed imports of Nitrogen (Nim). Nfer refers to the net fertilizer amount, which is the amount of nitrogen fertilizer converted according to 100% nitrogen. For the Ndep, only

the oxidized nitrogen (NO_x) was considered since the ammonium nitrogen (NH_x) is not stable in the environment [10]. N_{fix} mostly includes symbiotic nitrogen fixation and non-symbiotic nitrogen fixation. Both the peanuts and soybeans in the Yanghe Reservoir Basin are symbiotic nitrogen fixation crops. On the basis of previous studies [37], the crop nitrogen fixation rate is shown in Appendix A Table A6. N_{im} represents the net import of food and feed, wherein the net content means the surplus or deficit between N consumption and production. N_{hum} consumption and N_{liv} consumption denote the protein consumption in human food and livestock feed, respectively. N_{liv} products represent the N content of livestock products, which mainly refers to the meat, milk, eggs, and other livestock products. The nitrogen consumption and excretion level of each species are shown in Appendix A Table A7. N_{cro} products represent the N content of agricultural crop production, which is shown in Appendix A Table A8.

The NAPI model chiefly includes phosphorus fertilizer application (P_{fer}), non-food phosphorus input (P_{non}), and net food/feed phosphorus imports (P_{im}). Among them, P_{non} primarily comes from phosphorus detergent. We calculated the NAPI and P_{im} using Equations (3) and (4), as follows:

$$\text{NAPI} = \text{Pfer} + \text{Pnon} + \text{Pim} \quad (3)$$

$$\text{Pim} = \text{Phum consumption} + \text{Pliv consumption} - \text{Pliv products} - \text{Pcro products} \quad (4)$$

where Phum consumption and Pliv consumption represent human and livestock P consumption, respectively. Pliv products and Pcro products refer to the P content of livestock products and agricultural crop production, respectively. The calculation methods of P_{fer} and P_{im} are similar to that of NANI. The units of these variables are tons P km⁻² year⁻¹.

The ECM method is a mathematical weighted equation to estimate the N and P exports from different sources to the outlet of the Yanghe Reservoir Basin from 2004 to 2015. The main pollution sources included domestic sewage, garbage, and excrement from the rural region, urban residents, livestock, and land use (cropland, forest land, urban land, and bare land). This method is often used to express the relationship between pollutants (rural and urban areas), land use types, livestock, and N and P exports [38,39]. The formula of the ECM (Equation (5)) is as follows:

$$L = \sum_{i=1} \lambda_{ij} E_{ij} A_i \quad (5)$$

where L represents the amount of nitrogen and phosphorus exports (t year⁻¹); i is the type of pollution source; j denotes the nutrient type, such as N and P; E_i represents the annual export coefficient of each pollution source (kg km⁻² year⁻¹/kg ca⁻¹ year⁻¹) in Table A9; λ_{ij} is the fraction of nutrient exports from the pollution source i to the river outlet in Appendix A Table A10; A_i is the area of land use type (km²) or the number of livestock (capita) and the population (people). The areas covered by the different land-use types from 2004 to 2015 were classified and extracted based on the multi-temporal fusion method and the RF classification algorithm. The livestock and population data of each county were derived from the local statistical yearbooks. We verified the results of the model with field monitoring data from May to October 2015. The sampling points are shown in Figure 1c, and the results are shown in Appendix A Table A11.

3.3. Prediction Model of Nutrient Input and Export

The primary purpose of machine learning model construction was to predict the N and P input into the watershed and exported into the river on an annual scale. Secondly, on the basis of the prediction results, the N and P export ratios were calculated. However, the rest of the export ratio represents the degree of potential pollution risk to soil and water quality caused by nutrient retention in the watershed.

To predict the input and export of N and P in 2004 and 2015, respectively, we set up 4 targets to obtain four predictor variables, as shown in Table 3. On the basis of the

estimation results of the traditional balance model, we used the Pearson correlation analysis to determine seven human activity indicators that have a significant impact on the input and export of N and P as the input data of the nutrient prediction model (See 4.1 for details). Those indexes include N and P fertilizer used per cultivated area (Ferc N and Ferc P (ton km²)); the percentage of urban land area in the total area (urbanization (%)); the percentage of forest area (Forest (%)), crop area (Crop (%)) and urban area (Urban (%)) in each grid; and the population density (ca km²), which is the average population per unit area of land. Before constructing the model, we performed pre-processing to remove autocorrelation, and transformed and unified the spatial scale of data, with a spatial resolution of 3 km. The time scale of model input data was 2004–2015. We determined the four variables predicted by the model, which are N and P inputs and exports. Thereafter, we divided the 980 sample points according into the training set and the testing at a ratio of 7:3. The SVM, RF, and MLR models were selected to construct the nutrient prediction model. As compared with the traditional model, machine learning can identify an optimal segmentation point, which can tackle both linear and nonlinear relationships. However, the traditional model can-not directly tackle non-linearity. When training the model, we configured and adjusted the super parameters to obtain the best performance, we determined the iterative method from the best super parameters, and we set the number of iterations to 100. Tenfold cross-validation was used to test the accuracy of the algorithm. The determination coefficient (R²) and RMSE was used to evaluate the results of the prediction model and the optimal model was selected based on this.

Table 3. Information on the four targets for the prediction of nutrient inputs and exports.

Targets	Input Variables	Predictor Variable	Method
Targets 1	Ferc N; Urbanization (%); Forest (%); Crop (%); Urban (%), Population density	N Input	SVM, RF, MLR
Targets 2	Ferc P; Urbanization (%); Forest (%); Crop (%); Urban (%), Population density	P Input	
Targets 3	Ferc N; Urbanization (%); Forest (%); Crop (%); Urban (%), Population density	N Export	
Targets 4	Ferc P; Urbanization (%); Forest (%); Crop (%); Urban (%), Population density	P Export	

3.3.1. Support Vector Machine (SVM)

SVM is a supervised learning algorithm that is mainly based on the principle of structural risk minimization and statistical learning theory [39]. To achieve the purpose of accurate classification of various types of data, the interval maximization method was used to find the maximum classification interval of the defined feature space in the data [40]. Moreover, in order to reduce the impact resulting from the limited sample data, the hyperplane analysis method was used to distance the sample data from this hyperplane [41]. When nonlinear problems are encountered, the kernel function can be used for mapping analysis [41,42]. The algorithm identifies the optimal parameter combination using the gradient descent method. The number of optimization iterations set by the program was 100. We used the sigmoid kernel function with hyperparametric penalty coefficient c , which improved the generalization ability of the model, with a value of 1. The values of hyperparameters γ and coef0 were 0.7 and 0.4, respectively [36,43].

3.3.2. Random Forest (RF)

In this study, RF was used not only for land-use classification, but also for the N and P input and export predictions. The RF method is an extension of the bagged classification tree considering the ensemble learning theory, which can improve the accuracy of models [4,44]. RF is an integrated model that uses a set of independent classification or regression trees to achieve classification or regression aims [36]. The advantage of the RF algorithm is that it is not sensitive to the noise in the training set, and is more conducive to obtaining a

robust model and avoiding overfitting conditions [45]. The randomness of RF is reflected in two aspects. The first is the randomness of samples [46]: a certain number of samples are randomly selected from the training set as the root node samples of each regression tree. The second is the randomness of features: when establishing each regression tree, a certain number of candidate features are randomly selected, and the most appropriate feature is selected as the splitting node. During parameter adjustment, the value of the `max_depth` of the decision tree was 10. When the depth of the tree reaches the maximum depth, the decision tree will stop operation no matter how many features can be branched. The `n_estimators` denote the number of decision trees, and the value is 2000. The `max_features` is one of the super parameters, which was set to 3. The maximum percentage of features considered in the decision tree was 10%. The selection of parameter values was conducted according to previous research [44–46].

3.3.3. Multiple Linear Regression (MLR)

Multiple linear regression is a widely used linear regression method [46]. A phenomenon is often associated with multiple factors, for example it is more effective and practical to predict or estimate dependent variables using the optimal combination of multiple independent variables than to predict or estimate only one independent variable [47]. To measure the error between the estimated value and the real value, a non-negative real number function is usually selected as the loss function in the linear regression model. The smaller the value of the non-negative real loss function, the smaller the error. The least-squares method is generally applicable to parameters estimation [48]. We used the MLR model to represent the relationship between the nutrient inputs, exports, and human indicators, and to explore the change characteristics of nutrients as a result of intensive human activities.

4. Results

4.1. Relationship between Human Activity Factors and Nutrient Input and Export

4.1.1. Nutrient Input and Export of Traditional Balance Model

Between 2004 and 2015, the input of N and P exhibited an upward and downward trend, respectively, (Appendix A Figure A4). Therefore, we selected the year 2004 and 2015 for the comparative analysis. Figure 3 shows the comparisons of N and P input between 2004 and 2015. According to the figure, the N and P input significantly changed in this period. The net import of food and feed in rural areas (Nim) was the main source of N inputs in 2004, and in 2015, this was 30% lower than in 2004 (Figure 3a). In addition, crop nitrogen fixation (Nfix) to N inputs accounted for the lowest contribution in those two years. The percentage of atmospheric nitrogen deposition (Ndep) of N input increased from 22% in 2004 to 27% in 2015. Another interesting finding was that although the contribution of N fertilizer applications (Nfer) to N inputs was only 15% in 2004, it was the main source of NANI in 2015. As compared with 2004, the sources of P input in 2015 decreased to varying degrees (Figure 3b). Although phosphorus fertilizer applications (Pfer) were the main source of P inputs, their contribution decreased from 61% to 52%. The contribution of non-food P inputs (Pnon) to P input was the lowest throughout, and it also decreased in 2015. The net food and feed import in rural regions (Pim) of P input in 2015 was lower than in 2004, with a decrease of 39%.

Figure 4 shows N and P exports from different pollution sources in the Yanghe River Basin. We found that livestock was the main source of N and P exports, but the contribution decreased in 2015 as compared with 2004. In addition, domestic wastes in rural and urban areas, crop production, and land-use types were also sources of N and P exports. During the study period, the N and P exports from urban life increased from 9.7% in 2004 to 10.8% in 2015 (Figure 4a,b) and from 15.4% to 17.9% (Figure 4c,d), respectively. In particular, the N and P exports from crop production increased significantly as compared with 2004, increasing by 15% and 27%, respectively. On the contrary, those from rural life decreased from 11% to 9.8% and 17% to 16%, which could be attributed to the rapid development of

urbanization. In comparison, the contribution of crop production to N exports was higher than that of P exports.

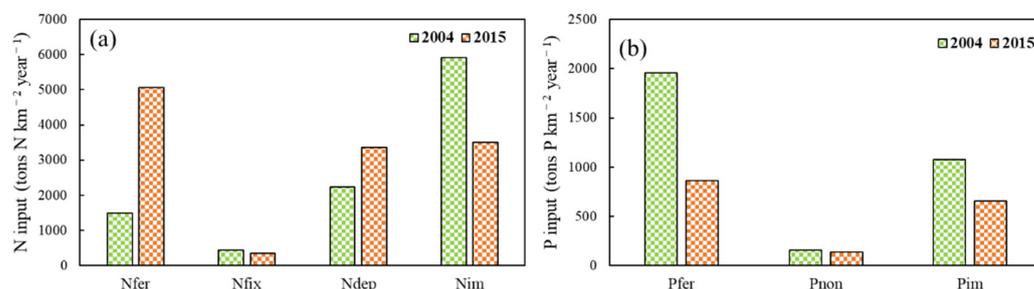


Figure 3. The main source of (a) N input and (b) P input in the Yanghe Basin in 2004 and 2015.

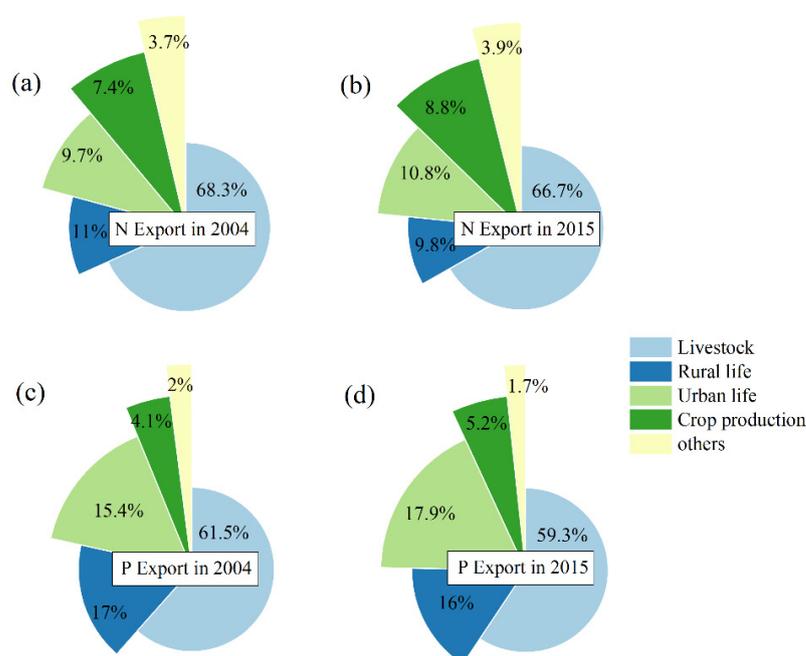


Figure 4. The main source of N exports: (a) 2004; (b) 2015; and P exports: (c) 2004; (d) 2015 in the Yanghe Basin.

4.1.2. Selecting Human Activity Indicators Based on the Balance Model

On the basis of the aforementioned results of the balance model (Figures 3 and 4), we found that N and P fertilization, net food/feed imports of N and P, livestock, crop production, and rural and urban life have obvious effects on the input and export of N and P. Among them, Ferc N, Ferc P, urbanization (%), Forest (%), Crop (%), Urban (%), and population density can be used to characterize the spatial region and dynamic changes in the above variables. We employed the Pearson correlation analysis to determine the correlations between human activity factors and nutrient inputs and exports (Figure 5). During the study period, N and P inputs exhibited a significant positive correction with human activity indicators ($p < 0.01$), including Ferc N and Ferc P, urbanization (%), Crop (%), Urban (%), and population density. Especially for urbanization (%) and population density, the high correlation coefficient (an R of 0.83 to 0.94) indicated that the densely inhabited district was the main source of N and P inputs.

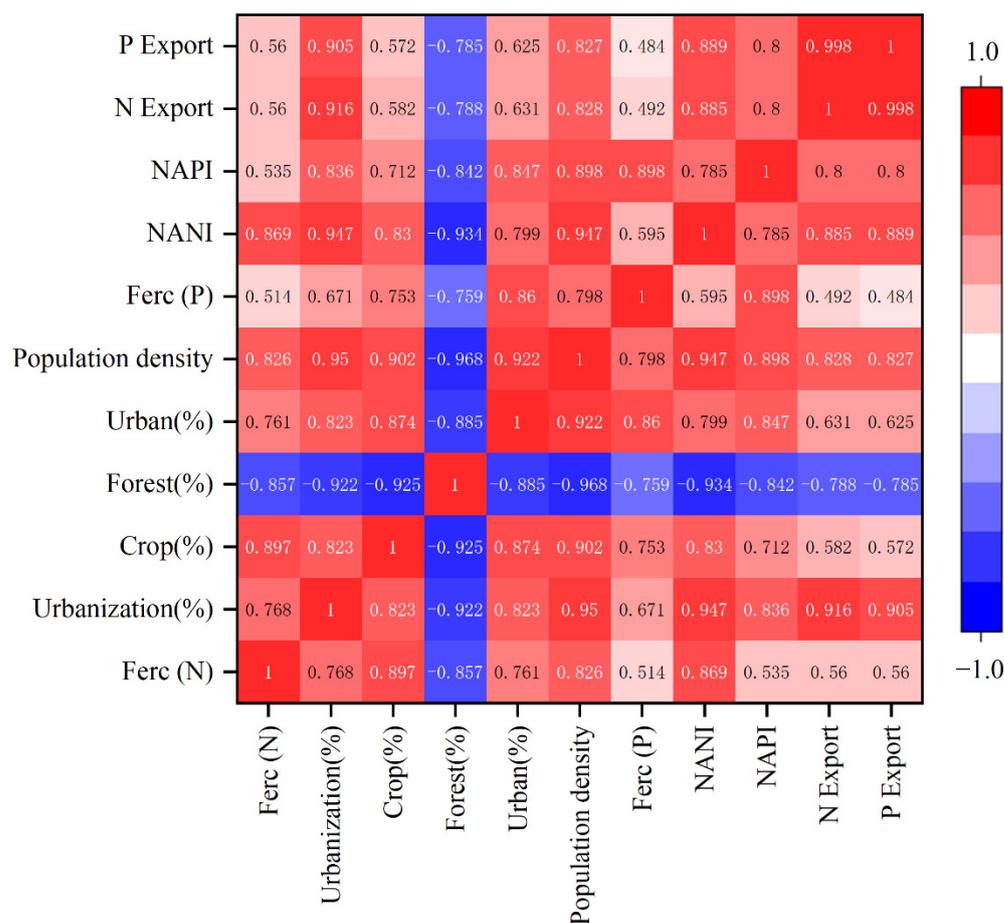


Figure 5. Pearson correlation coefficient between human indicators and NANI, NAPI, N export, P export.

Furthermore, N and P exports were significantly and positively correlated with urbanization (%), Urban (%), and population density in the Yanghe River Basin. In other words, with the rapid development of urbanization, the population and sewage discharge increased, increasing nutrient input and export. It is worth noting that there was a negative correlation between the percentage of Forest (%) and N and P inputs and exports. The correlation coefficient between N export, P export, and NANI, NAPI in the Yanghe River Basin was higher than 0.8 ($p < 0.01$), which further suggests that the variation in N export and P export were directly affected by the variations in NANI and NAPI.

4.2. Model Performance of SVM, RF, and MLR Based on Human Activity Factors

We constructed four targets to predict the input and export of nutrients, which utilized seven human activity indicators. The input data set was from 2004 to 2015. There were 980 groups of data in total, and each group of samples contained 7 characteristic components. Among them, 686 sample points of the observed data were selected for model modeling of different prediction variables, and the remaining 294 sample points of the observed data were used to verify the accuracy of the evaluation model. The model was trained using 10 cross-validation methods. R^2 and RMSE coefficients were used to evaluate the results of the prediction model and the optimal model was selected based on this.

Table 4 shows the performances of the regression models in the training stage and the validation stage. For target 1, in order to predict N input, the R^2 and RMSE in the validation set were 0.95, 0.74, and 0.92, and 32.75, 61.09, and 49.34, respectively. Among them, the RMSE unit was consistent with the input and export of nutrients. The SVM and MLR models outperformed the RF model. For target 2, the predictive validity of SVM in

predicting P input was good, with an R^2 and RMSE of 0.94 and 5.18, respectively. In terms of N and P exports, target 3 and target 4 demonstrated that the performances of the MLR and SVM were better than that of the RF ($R^2 > 0.9$ for SVM and MLR; the RMSE exhibited a similar pattern). Those three models can be ranked as follows: SVM > MLR > RF. For the SVM model in particular, the R^2 of the validation set was above 0.91, and the value of RMSE was smaller, indicating that it almost perfectly simulated nutrient inputs and exports.

Table 4. The performance evaluation of prediction models.

Variable	Model	Training		Validation	
		R^2	RMSE	R^2	RMSE
Targets 1	SVM	0.95	36.65	0.95	32.75
	RF	0.84	40.15	0.74	61.09
	MLR	0.88	62.42	0.92	49.34
Targets 2	SVM	0.95	7.11	0.94	5.18
	RF	0.97	3.26	0.71	9.32
	MLR	0.91	11.49	0.94	9.91
Targets 3	SVM	0.93	1.21	0.91	1.45
	RF	0.72	1.78	0.77	1.79
	MLR	0.75	2.81	0.90	1.68
Targets 4	SVM	0.88	0.23	0.93	0.18
	RF	0.83	0.32	0.85	0.31
	MLR	0.77	0.41	0.93	0.24

Note: The corresponding results of SVM have been bold.

4.3. Spatial Variation of Nutrient Input and Export under Anthropogenic Activities Based on SVM

4.3.1. Variation Characteristics of Nutrient Input

Based on the evaluation results of the above machine learning model, the performance of the SVM model was better than RF and MLR. Therefore, we predicted the input and export of N and P in 2004 and 2015 based on the SVM model. The measured data and simulated data of N and P input and export in 2004 and 2015 are fitted (Appendix A Figures A5 and A6). The fitting results of N and P input show that R^2 is 0.90 to 0.96, and the R^2 of N and P export is 0.89 to 0.97.

The N inputs increased from 10,066 tons $\text{km}^{-2} \text{year}^{-1}$ in 2004 to 12,278 tons $\text{km}^{-2} \text{year}^{-1}$ in 2015. The N inputs were concentrated in the southwest of the Yanghe River Basin. Compared with 2004, N inputs increased 55% in 2015, with greater increases in the west than in the south of the Yanghe River Basin (Figure 6a,b). There was little change in the northeast. The P inputs in the Yanghe Basin decreased from 3192 tons $\text{km}^{-2} \text{year}^{-1}$ in 2004 to 1655 tons $\text{km}^{-2} \text{year}^{-1}$ in 2015 (Figure 6c,d). Moreover, the P inputs were distributed in the western region of the Yanghe River Basin. During the study period, P inputs exhibited an obvious decreasing trend, especially in the central and western regions.

4.3.2. Variation Characteristics of Nutrient Export

The spatial variation trend of N and P exports to rivers in the Yanghe River Basin in 2004 and 2015 are shown in Figure 7. The riverine export of N and P decreased from 413 tons $\text{km}^{-2} \text{year}^{-1}$ to 375 tons $\text{km}^{-2} \text{year}^{-1}$ and 61 tons $\text{km}^{-2} \text{year}^{-1}$ to 53 tons $\text{km}^{-2} \text{year}^{-1}$, respectively. N and P exports were concentrated in the southwest of the Yanghe River Basin. Among them, N and P exports from the western region were significantly higher than those from the eastern region. This was mainly due to the rapid development of urbanization in this area. As compared with 2004, the N export in the northern region exhibited a decrease in 2015, but there was a significant increment in the southern region (Figure 7a,b).

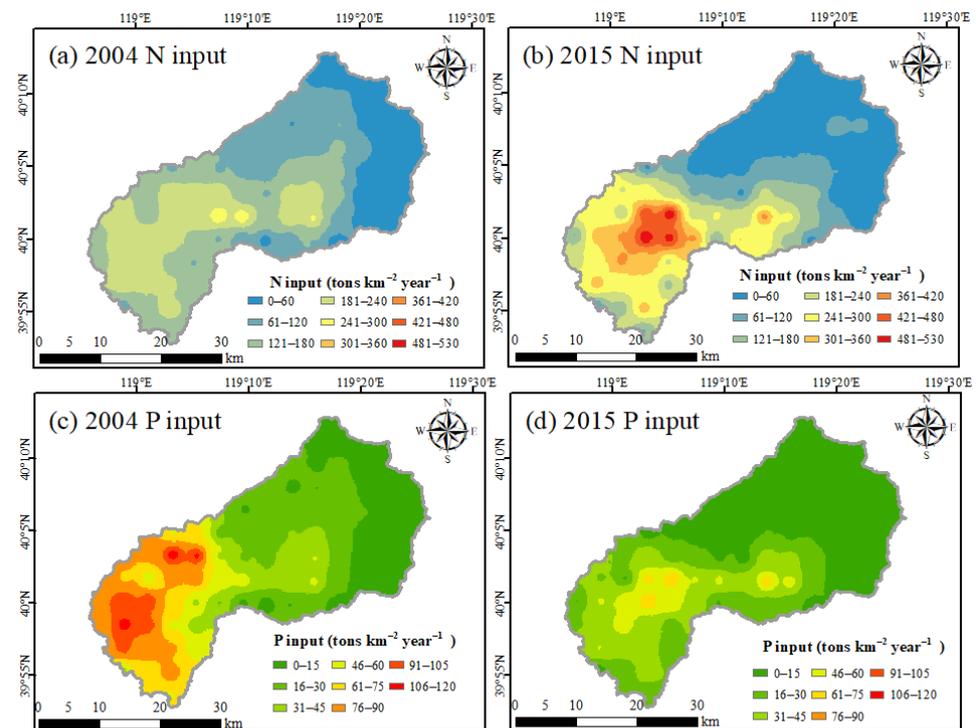


Figure 6. Spatial-changing trend for (a,b) N input and (c,d) P input in 2004 and 2015.

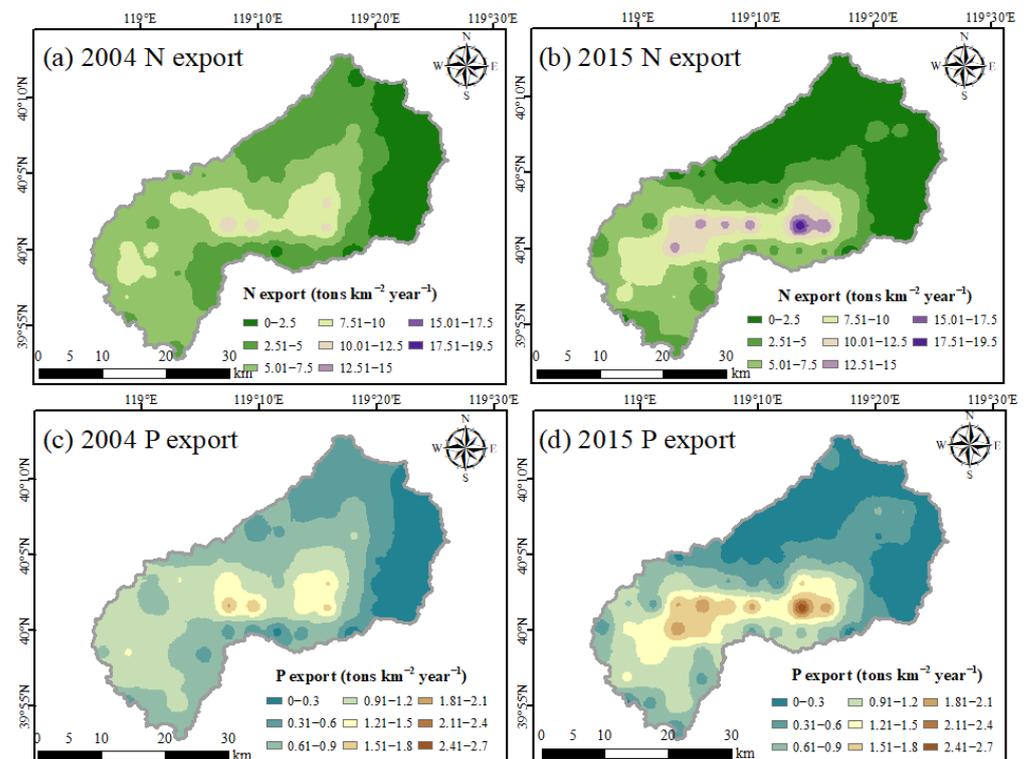


Figure 7. Spatial change trend for (a,b) N exports and (c,d) P exports in 2004 and 2015.

4.4. Quantifying the Retention Risk of Nutrients in Watersheds

Figure 8a–e show the relationship between N and P input and export in 2004 and 2015 as simulated by the SVM prediction model. Among them, the purple dot and green dot refer to the prediction results of the N and P input and export models for each grid in the study area in 2004 and 2015, respectively. We found that there was a high correlation

between the predicted values of N input, N export, and P input, and P export with an $R^2 > 0.80$. Interestingly, this finding demonstrates that the inputs and exports of N and P were closely related to the impact of human activities.

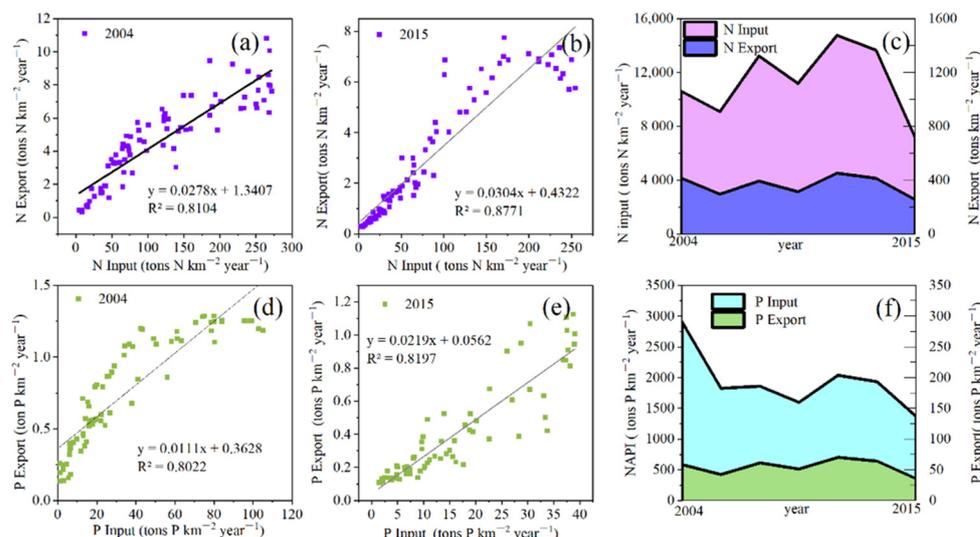


Figure 8. The relationship between N (a,b), P (d,e) inputs and exports and Nutrient retention (c,f).

Additionally, the slopes of the N input and N export fitting curves illustrate that the export ratios of riverine N were about 2.8% (0.0278) and 3.0% (0.0304) in 2004 and 2015, respectively (Figure 8a,b). Moreover, the export ratios of riverine P increased from 1.1% (0.011) in 2004 to 2.2% (0.0219) in 2015 (Figure 8d,e). As compared with 2004, the export ratios of riverine N and P increased in 2015, and the export ratios of riverine nutrients in the Yanghe River Basin were still far from the global export ratios of N (25%) and P (3%) [49]. In Figure 8c, the dark blue area represents the export of N, and the purple area represents the quantity of N retained in the watershed, reaching 97%. The same is true for Figure 8f. The green area denotes the P export, and the blue area is the retention risk of P, reaching 98%. The lower riverine nutrient export ratios were, as well as the higher input ratios were, the larger quantity of N and P inputs were retained in the Yanghe River Basin. The N and P retentions exacerbated the problem of water pollution and posed a potential risk to the water environment in the Yanghe Reservoir.

5. Discussion

5.1. The Driving Factors of N and P Inputs

On the basis of the constructed nutrient prediction model, this study quantitatively evaluated the nutrient input and export in the Yanghe Reservoir Basin. In our study, the N input increased from 10,066 tons $\text{km}^{-2} \text{year}^{-1}$ to 12,278 tons $\text{km}^{-2} \text{year}^{-1}$, and the P input decreased from 3192 tons $\text{km}^{-2} \text{year}^{-1}$ to 1655 tons $\text{km}^{-2} \text{year}^{-1}$. This result aligns with other existing studies [47,48]. We found that Nim and Nfer were the main sources of N input, and the Pfer was the main source of P input (Figure 3). The N and P inputs were concentrated in the southwest of the Yanghe Reservoir Basin (Figure 6). The cause of this phenomenon can be attributed to cropland expansion (with an increment of 71.6%) from 2004 to 2015 (Appendix A Figure A3). N and P fertilizers are widely used in the crop cultivation process. N and P loss and infiltration from the croplands directly or indirectly polluted the surrounding surface water bodies and underground confluence [50].

5.2. The Driving Factors of N and P Exports

During the study period, we found that livestock breeding was the largest pollution source in the study area; the reason was that N and P in livestock excrement were discharged into rivers, resulting in eutrophication [49]. According to survey statistics,

there are numerous breeding farms distributed along the Maguying river upstream of the Yanghe Reservoir [32]. Abundant pollutants produced by animal husbandry can easily enter the river with rainfall-runoff [51] and potentially result in a load that is far beyond the water environmental capacity of the river. Another reason is that pollutant emissions from livestock and poultry farming are positively correlated with the economy and urbanization [52]. For example, with the rapid urbanization process, the urban population in the Xiyang river increased by 20%, and the area of urban land increased by 15.5%, as compared with 2004 (Appendix A Figure A3). The pollution of livestock and poultry breeding in Yanghe Reservoir Basin has not been effectively controlled, and the pollution discharge is excessive. Consequently, we should reinforce the management of farms and improve the feces utilization rate, in order to reduce the threat of pollutants to the water source quality. In our study, crop production sources accounted for a small share of the total nutrient exports, which is different from the assessment by Lian et al. [13] regarding crop production sources for N (28%). The main reason for this difference is that the rainfall in the Yanghe Reservoir basin located in a semi-arid and semi-humid area (750 mm) is much less than that in the Taihu Lake Basin, which is located in a humid area (1115 mm). Our results are more similar to those of Zhang [12] who investigated the Baiyangdian basin, a semi-arid area with an average rainfall of 520 mm.

5.3. The Potential Risk of Nutrient Pollution Is Related to Human Activities

According to the results of the N and P input and export prediction model, we found that the export ratios of N and P during the study period were 2.8–3.0% and 1.1–2.2%, respectively, and were far lower than the global N (25%) and P (3%) export ratios [12]. Low export ratios may be caused by the terrain [52]. The Yanghe Reservoir Basin is a sub-basin of the Luanhe River Basin in the North China Plain. It is an alluvial plain with flat terrain. Moreover, alongside a runoff decrement of 50.43% in the past 30 years, there has been an increasing N and P retention risk since the intensity of the non-point source pollutant scouring effect is proportional to the runoff [53]. Furthermore, the consumption of water resources by human activities affected the drainage of rivers [54]. The water consumption of the Yanghe Reservoir Basin far exceeded its water intake, and greatly reduced the export of river nutrients. The low export ratio of N and P further indicates that a large amount of N and P inputs were retained in the basin, which also demonstrates that the potential risk of N and P pollution was inseparable from the impact of human activities. Therefore, it is suggested that in research related to nutrient input and export, human activity indicators can be used to replace the traditional statistical data to improve the accuracy and convenience of N and P input and export estimations.

5.4. Implications of Nutrient Input and Export Prediction Model

As far as we are aware, this is the first study to construct a nutrient input and export prediction model for the Yanghe Reservoir Basin. This study is important for water quality safety assessments of the Yanghe Reservoir, an important drinking water source. Embedding human activity indicators in the prediction reduced the uncertainty of the traditional balance model. Many studies illustrate that human activities significantly affect the input and export of N and P [52,53], which is consistent with our research results. On the basis of the correlation between water transparency and lake nutrients, Tyler et al. used water transparency data to replace the concentrations of TN, TP, and chlorophyll to predict the lake nutrient concentration [55]. In this study, we not only considered the linear relationship between human activities and N and P input and export, but also the non-linear relationship between them. It is worth noting that the SVM model's performance was better than that of RF or MLR. The SVM model can obtain robust results as a result of its strong generalization ability [43]. Moreover, without feature selection, SVM can process data with extremely high dimensions [56]. Therefore, the SVM model, which simplified the original method, can be applied to the prediction of nutrient input and export.

In the following, we outline various limitations of our study. Firstly, N and P inputs and exports in different regions are dominated by various human activities. Thus, it would be better to select human activity indicators according to local conditions. Secondly, climate change, which affects rainfall, may also interfere with the balance of nitrogen and phosphorus. This represents a good topic for study in the future. Thirdly, the multi-temporal fusion method and the RF classification algorithm were applied in this study to extract the area of land-use types, which has the advantage of identifying the land-use types under different time scales. However, as a result of the statistical data limitations, this study only estimated the emissions of pollutants on an annual scale, and as is well known, agricultural activities have strong seasonality. The temporal and spatial characteristics of NPSP on the seasonal or monthly scales remain to be explored, which is our future research direction. Moreover, although the proposed framework demonstrated its efficacy, the matter of data scarcity in the study area should be taken into account; therefore, comparisons with other frameworks in regions with sufficient data will need to be made in future studies.

6. Conclusions

In this study, we built a convenient, reliable, and accurate nutrient prediction model for estimating nutrient input and export and evaluating the potential nutrient retention risk for drinking water sources. The prediction model reduced the uncertainty caused by the traditional nutrient balance estimation methods, which rely on statistical data and a large number of estimation coefficients. Remote sensing images and the multi-temporal fusion method were used to select sample points, and the RF algorithm was incorporated to identify the pollution sources such as forest land, bare land, cropland, water, and urban land in the drinking water sources area. The overall classification low-end accuracy was 92%. The performance of the prediction model based on SVM was better than those based on RF and MLR. The results of the prediction model were of the same order of magnitude as those of traditional N and P estimation methods and demonstrated that the nutrient prediction model is reliable. The export ratios of N and P under human disturbance were 2.8–3.0% and 1.1–2.2%, respectively. The excessive N and P (97% to 98%) in the riverine system pose a potential risk to the drinking water sources. Next, we will introduce climate change factors, including rainfall, into our nutrient prediction model in order to assess the potential risks of the combined impact of human activities and climate change on nutrient retention.

Author Contributions: Conceptualization, Y.Z. and J.Y.; methodology, Y.Z. and C.L.; software, Q.W.; validation, Y.Z., X.Z. and L.C.; formal analysis, Y.L.; resources, J.Y.; data curation, C.L.; writing—original draft preparation, Y.Z.; writing—review and editing, Y.Z., Q.W. and X.Z.; visualization, Y.Z. and L.C.; supervision, Y.L.; project administration, Q.W.; funding acquisition, J.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Natural Science Foundation of China, grant number 51779007 and 41671018.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Appendix A

Appendix A.1. Multi-Temporal Images Fusion

To reduce the influence of cloud masks, the percentage of clouds was restricted (less than 20%) when synthesizing cloudless images. Thereafter, the cloud mask algorithm

was used to calculate the image in the specified time and space, and the minimum cloud synthetic image was reconstructed using the median synthesis method. The advantage of GEE is that it can unify the coordinate system through an embedded algorithm to ensure the geometric registration accuracy of data. According to the main pollution sources and land cover in the Yanghe Reservoir Basin [38], the land use types in the study area are mainly divided into croplands, forests, water, bare lands, and urban impervious areas. Figure A1 shows the steps involved in remote sensing images processing.

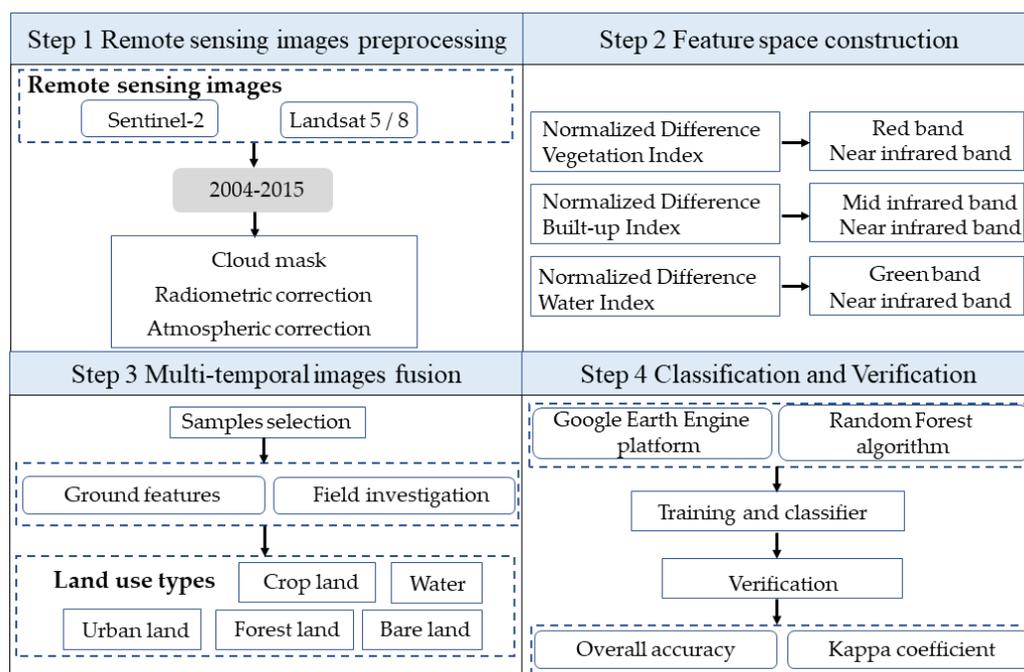


Figure A1. Flowchart of land-use type classification and extraction based on GEE.

In general, as a result of the different ground characteristics of pollution sources in various periods, multi-temporal remote sensing analysis is helpful to eliminate the errors in extracting land use information [38,39]. On the basis of the GEE platform, this study used multi-temporal remote sensing data and selected sample points in combination with the field investigation in the study area from March to October 2015 (Table A1). Moreover, three typical indexes, including NDVI (Normalized Difference Vegetation Index), NDWI (Normalized Difference Water Index), and NDBI (Normalized Difference Built-up Index), were used to construct the feature space of feature recognition in the classification task. The total number of samples obtained was 10,339 from 2004 to 2015. These were randomly assigned to the training set and the testing set at a ratio of 7:3 (Table A2).

Table A1. Principle of selecting sample points based on multi-temporal fusion.

Feature Type	Optimal Temporal Images	Sample Point Selection Standard
Forest land and crop land	March to May	The cultivated land crops have not been sown. The forest land has begun to turn green (See Appendix A Figure A2a).
Bare land and urban land	June to August	The crop land has high vegetation coverage, which is easy to distinguish between urban land and bare land (See Appendix A Figure A2b).
Water	September to October	The water body information is relatively clear and will not be affected by vegetation coverage (See Appendix A Figure A2c).

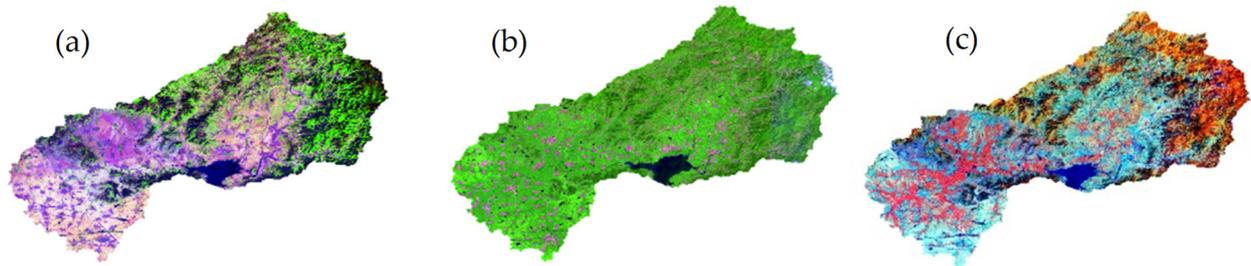


Figure A2. The feature changes for different periods. (a) March to May; (b) June to August; (c) September to October.

Table A2. The total number of sample points selected.

Type of Samples	Total Count
Urban land	2057
Forest land	2486
Crop land	2365
Water	1898
Bare land	1533

Appendix A.2. Verification of Classification Results

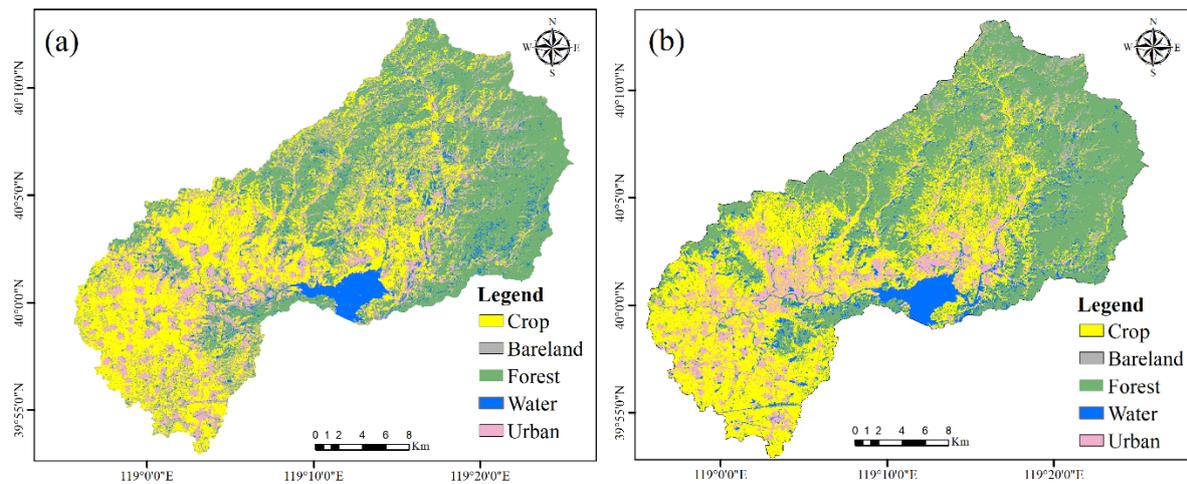
Kappa coefficient and the overall accuracy were used to evaluate the classification accuracy of land types extracted from GEE [57]. To obtain more detailed ground feature information for the Yanghe Reservoir Basin, several field investigations were carried out. The Kappa coefficient value and the overall accuracy for each year are shown in Appendix A Table A3. For example, in 2015, the Kappa coefficient and the overall accuracies were 94.0% and 95.2%, respectively, which was enough to demonstrate that the classification results are reliable (Appendix A Table A4). Figure A3 is the area variation of land use types extracted based on GEE in 2004 and 2015. The main land-use types were forests and cropland, followed by urban land in the Yanghe Reservoir Basin. Compared with 2004, under the interference of human activities, the crop land, forest land, and urban area increased in 2015, while bare land showed a decreasing trend. (Table A5). During this period, we found that a portion of crop land areas was transformed into forest land and water, which is mainly because according to the urban and rural planning of Qinhuangdao City, the study area implemented ecological protection measures, such as strengthening afforestation and a scheme for returning farmland to water.

Table A3. Classification accuracy evaluation.

Year	Overall Accuracy	Kappa Coefficient
2004	0.94	0.92
2005	0.93	0.91
2006	0.95	0.94
2007	0.93	0.91
2008	0.93	0.92
2009	0.96	0.93
2010	0.92	0.91
2011	0.94	0.91
2013	0.93	0.92
2014	0.93	0.91
2015	0.95	0.94

Table A4. Accuracy evaluation of land use types classification for 2015.

Land Use Types	Ground Real Data					Class Sum	User's (%)
	Urban Land	Forest Land	Crop Land	Water	Bare Land		
Urban land	177	4	3	0	3	187	94.65
Forest land	2	216	5	0	3	226	95.58
Crop land	2	4	205	3	1	215	95.35
Water	0	3	4	165	0	172	95.93
Bareland	4	1	3	0	131	139	94.24
Actual sum	185	228	220	168	138	939	
Producer's (%)	95.68	94.74	93.18	98.21	94.93		
Overall accuracy		95.21		Kappa		93.97	

**Figure A3.** Area variation of land-use types in (a) 2004 and (b) 2015 was extracted based on GEE.**Table A5.** Pollution source area extracted by classification in 2004 and 2015.

Type	2004 Classified Area (km ²)	2015 Classified Area (km ²)
Urban land	52.5	68.6
Forest land	275.8	321.1
Crop land	279.4	306.8
Water	49.3	40.3
Bare land	97.7	17.9

Table A6. Crop nitrogen fixation coefficient (kg ha⁻¹ year⁻¹) [58–60].

Types	Biofixation Rate
Symbiotic N fixation	
Soybeans	96
Peanuts	80
Non-symbiotic fixation	
Rice	30
Corn	15

Table A7. Animal nutrient consumption and excretion [61].

Animal Species	N Consumption and Excretion (kg N ca ⁻¹ year ⁻¹)		P Consumption and Excretion (kg P ca ⁻¹ year ⁻¹)	
	Consumption	N Excretion	Consumption	P Excretion
Pigs	16.68	11.51	3.17	4.59
Cattle	54.82	48.79	9.78	10.99
Sheep	6.85	5.75	1.06	1.26
Livestock	0.6	0.2	0.12	0.18

Note: We assumed that spoilage and inedible components caused a 10% loss of animal products available for consumption [62].

Table A8. N and P content of agricultural crop production (g kg⁻¹) [61].

Crops	Corn	Wheat	Paddy	Soybean	Peanut	Vegetable	Apple	Peach	Pear	Grape
N	14.08	17.92	11.84	56.16	19.36	2.72	0.32	0.8	0.48	0.8
P	2.44	1.88	1.1	4.65	4.65	0.3	0.12	0.13	0.13	0.12

Note: We assumed that spoilage and inedible components caused a 10% loss of animal products available for consumption [62].

Table A9. The export coefficients used in this study for different pollution sources in the Yanghe Reservoir Basin [12,58].

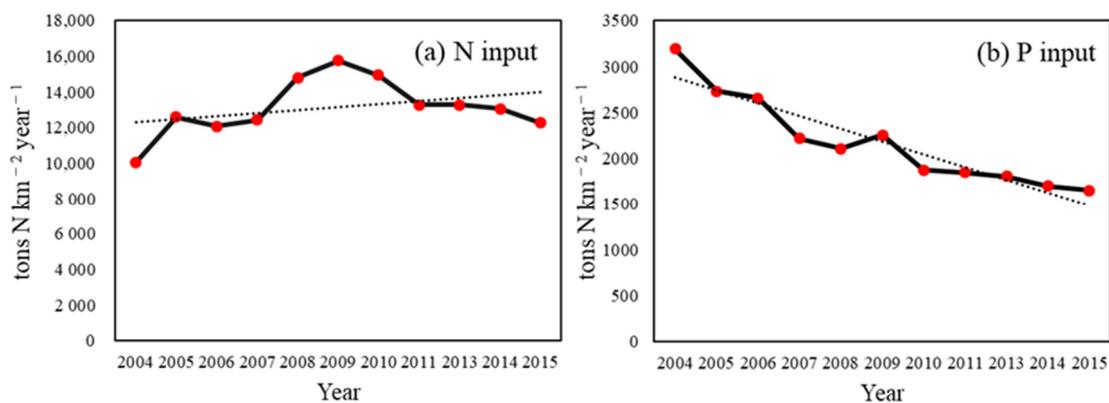
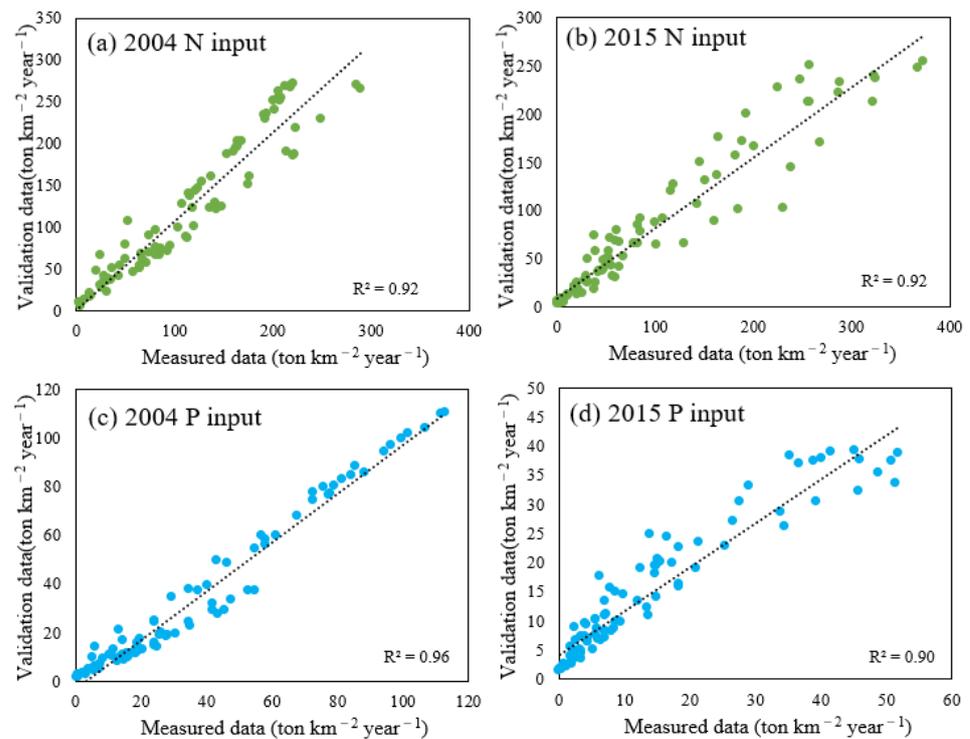
Pollution Source	Items	Export Coefficient		Unit
		TN	TP	
Land use	Crop production	1.85	0.15	tons km ⁻² yr ⁻¹
Livestock	Pigs	1.85	0.24	Kg ca ⁻¹ yr ⁻¹
	Cattle	15.33	1.28	
	Sheep	0.37	0.12	
	Poultry	0.01	0.004	
Rural life	Domestic sewage	0.34	0.04	Kg ca ⁻¹ yr ⁻¹
	Domestic garbage	1.13	0.38	
	Domestic excrement	0.11	0.02	
Urban life	Urban residents	0.85	0.2	Kg ca ⁻¹ yr ⁻¹
Others	Forest	0.3	0.16	tons km ⁻² yr ⁻¹
	Urban land	0.73	0.02	
	Others land use	0.7	0.08	

Table A10. Fraction of nutrients exported (λ_{ij}) to the rivers [63,64].

Sources	Fraction (λ_{ij})
Land use	0.07
Livestock	0.3
Domestic sewage	0.3
Domestic garbage and excrement	0.1

Table A11. Calibration of ECM model in Yanghe River watershed (755 km²) in 2015.

Land-Use Types	Areas (km ²)	Animals	Amount (Capital)	Population	Amount (People)
Crop land	306.8	Pig	15,286	Urban	238,200
Forest	321.1	Cow	878		
Urban	68.6	Sheep	5762	Rural	1,212,700
Others	58.2	Livestock	177,316		
Estimated TN load	375.69 t	Estimated TP load		53.25 t	
Watershed outlet flow	1.690×10^8 m ³ year ⁻¹	Average TN concentration (from May to October)	2.15 mg L ⁻¹	Average TP concentration (from May to October)	0.29 mg L ⁻¹
Observed TN load relative error	363.35 t	Observed TP load relative error		49.01 t	
	3.30%			8.60%	

**Figure A4.** Variation trend for N and P input from 2004 to 2015.**Figure A5.** Fitting results of measured data and simulated data of N and P input in 2004 and 2015.

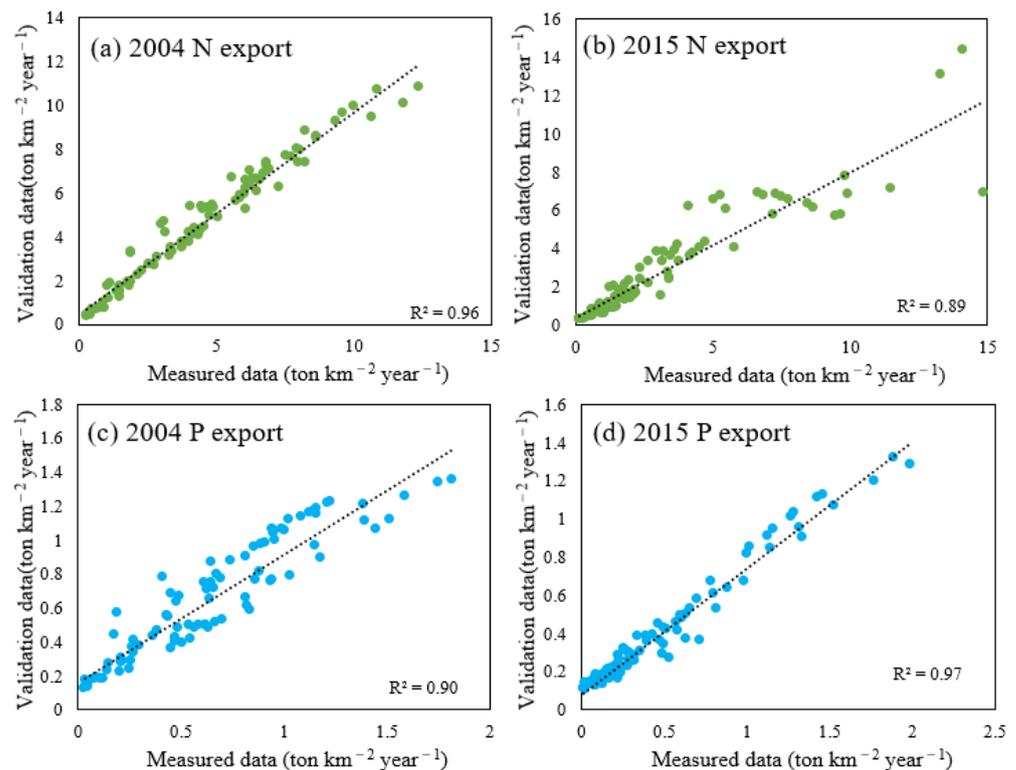


Figure A6. Fitting results of measured data and simulated data of N and P export in 2004 and 2015.

References

- Xie, C.; Huang, X.; Mu, H.; Yin, W. Impacts of Land-Use Changes on the Lakes across the Yangtze Floodplain in China. *Environ. Sci. Technol.* **2017**, *51*, 3669–3677. [[CrossRef](#)] [[PubMed](#)]
- He, J.; Charlet, L. A review of arsenic presence in China drinking water. *J. Hydrol.* **2013**, *492*, 79–88. [[CrossRef](#)]
- Xia, Y.; Ti, C.; She, D.; Yan, X. Linking river nutrient concentrations to land use and rainfall in a paddy agriculture–urban area gradient watershed in southeast China. *Sci. Total Environ.* **2016**, *566–567*, 1094–1105. [[CrossRef](#)] [[PubMed](#)]
- Ongley, E.D.; Xiaolan, Z.; Tao, Y. Current status of agricultural and rural non-point source Pollution assessment in China. *Environ. Pollut.* **2010**, *158*, 1159–1168. [[CrossRef](#)] [[PubMed](#)]
- Shi, Y.; Cheng, C.-W.; Zhu, Y.; Lei, P.; Zhou, H.-D.; Wen, T.-J. Institutional and Organizational Innovation on China Agricultural Non-point Pollution Prevention: Analysis on the 1st National Survey of Pollution Sources Bulletin. *Agric. Econ. Manag.* **2011**, *2*, 27–37.
- Ma, R.; Yang, G.; Duan, H.; Jiang, J.; Wang, S.; Feng, X.; Li, A.; Kong, F.; Xue, B.; Wu, J.; et al. China's lakes at present: Number, area and spatial distribution. *Sci. China Earth Sci.* **2011**, *54*, 283–289. [[CrossRef](#)]
- Hobbie, S.E.; Finlay, J.C.; Janke, B.D.; Nidzgorski, D.A.; Millet, D.B.; Baker, L.A. Contrasting nitrogen and phosphorus budgets in urban watersheds and implications for managing urban water pollution. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, 4177–4182. [[CrossRef](#)]
- Hollinger, E.; Cornish, P.S.; Baginska, B.; Mann, R.; Kuczera, G. Farm-scale stormwater losses of sediment and nutrients from a market garden near Sydney, Australia. *Agric. Water Manag.* **2001**, *47*, 227–241. [[CrossRef](#)]
- Leone, A.; Ripa, M.N.; Uricchio, V.; Deák, J.; Vargay, Z. Vulnerability and risk evaluation of agricultural nitrogen pollution for Hungary's main aquifer using DRASTIC and GLEAMS models. *J. Environ. Manag.* **2009**, *90*, 2969–2978. [[CrossRef](#)]
- Wang, G.; Li, J.; Sun, W.; Xue, B.; A, Y.; Liu, T. Non-point source pollution risks in a drinking water protection zone based on remote sensing data embedded within a nutrient budget model. *Water Res.* **2019**, *157*, 238–246. [[CrossRef](#)]
- Ding, X.; Liu, L. Long-Term Effects of Anthropogenic Factors on Nonpoint Source Pollution in the Upper Reaches of the Yangtze River. *Sustainability* **2019**, *11*, 2246. [[CrossRef](#)]
- Zhang, X.; Yi, Y.; Yang, Z. Nitrogen and phosphorus retention budgets of a semiarid plain basin under different human activity intensity. *Sci. Total Environ.* **2020**, *703*, 134813. [[CrossRef](#)]
- Huishu, L.; Qiuliang, L.; Xinyu, Z.; Haw, Y.; Hongyuan, W.; Limei, Z.; Hongbin, L.; Huang, J.-C.; Tianzhi, R.; Jiaogen, Z.; et al. Effects of anthropogenic activities on long-term changes of nitrogen budget in a plain river network region: A case study in the Taihu Basin. *Sci. Total Environ.* **2018**, *645*, 1212–1220. [[CrossRef](#)]
- Chen, D.; Hu, M.; Guo, Y.; Dahlgren, R.A. Influence of legacy phosphorus, land use, and climate change on anthropogenic phosphorus inputs and riverine export dynamics. *Biogeochemistry* **2015**, *123*, 99–116. [[CrossRef](#)]

15. Deng, C.; Liu, L.; Peng, D.; Li, H.; Zhao, Z.; Lyu, C.; Zhang, Z. Net anthropogenic nitrogen and phosphorus inputs in the Yangtze River economic belt: Spatiotemporal dynamics, attribution analysis, and diversity management. *J. Hydrol.* **2021**, *597*, 126221. [[CrossRef](#)]
16. Lai, G.; Luo, J.; Li, Q.; Qiu, L.; Pan, R.; Zeng, X.; Zhang, L.; Yi, F. Modification and validation of the SWAT model based on multi-plant growth mode, a case study of the Meijiang River Basin, China. *J. Hydrol.* **2020**, *585*, 124778. [[CrossRef](#)]
17. Adu, J.; Kumarasamy, M.V. Assessing Non-Point Source Pollution Models: A Review. *Pol. J. Environ. Stud.* **2018**, *27*, 1913–1922. [[CrossRef](#)]
18. Zhang, L.; Wang, Z.; Chai, J.; Fu, Y.; Wei, C.; Wang, Y. Temporal and Spatial Changes of Non-Point Source N and P and Its Decoupling from Agricultural Development in Water Source Area of Middle Route of the South-to-North Water Diversion Project. *Sustainability* **2019**, *11*, 895. [[CrossRef](#)]
19. Russell, M.J.; Weller, D.E.; Jordan, T.E.; Sigwart, K.J.; Sullivan, K.J. Net anthropogenic phosphorus inputs: Spatial and temporal variability in the Chesapeake Bay region. *Biogeochemistry* **2008**, *88*, 285–304. [[CrossRef](#)]
20. Hong, B.; Swaney, D.P.; Howarth, R.W. Estimating Net Anthropogenic Nitrogen Inputs to U.S. Watersheds: Comparison of Methodologies. *Environ. Sci. Technol.* **2013**, *47*, 5199–5207. [[CrossRef](#)]
21. Wang, X.; Feng, A.; Wang, Q.; Wu, C.; Liu, Z.; Ma, Z.; Wei, X. Spatial variability of the nutrient balance and related NPSP risk analysis for agro-ecosystems in China in 2010. *Agric. Ecosyst. Environ.* **2014**, *193*, 42–52. [[CrossRef](#)]
22. Kettering, J.; Park, J.-H.; Lindner, S.; Lee, B.; Tenhunen, J.; Kuzyakov, Y. N fluxes in an agricultural catchment under monsoon climate: A budget approach at different scales. *Agric. Ecosyst. Environ.* **2012**, *161*, 101–111. [[CrossRef](#)]
23. Oenema, O.; Kros, H.; de Vries, W. Approaches and uncertainties in nutrient budgets: Implications for nutrient management and environmental policies. *Eur. J. Agron.* **2003**, *20*, 3–16. [[CrossRef](#)]
24. Lin, Y.; Li, L.; Yu, J.; Hu, Y.; Zhang, T.; Ye, Z.; Syed, A.; Li, J. An optimized machine learning approach to water pollution variation monitoring with time-series Landsat images. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *102*, 102370. [[CrossRef](#)]
25. Kim, Y.H.; Im, J.; Ha, H.K.; Choi, J.-K.; Ha, S. Machine learning approaches to coastal water quality monitoring using GOCI satellite data. *GIScience Remote Sens.* **2014**, *51*, 158–174. [[CrossRef](#)]
26. Xiao, X.; He, J.; Huang, H.; Miller, T.R.; Christakos, G.; Reichwaldt, E.S.; Ghadouani, A.; Lin, S.; Xu, X.; Shi, J. A novel single-parameter approach for forecasting algal blooms. *Water Res.* **2017**, *108*, 222–231. [[CrossRef](#)]
27. Gebler, D.; Wiegleb, G.; Szoszkiewicz, K. Integrating river hydromorphology and water quality into ecological status modelling by artificial neural networks. *Water Res.* **2018**, *139*, 395–405. [[CrossRef](#)]
28. Worrall, F.; Burt, T.P.; Howden, N.J.K.; Whelan, M.J. The fluvial flux of nitrate from the UK terrestrial biosphere—An estimate of national-scale in-stream nitrate loss using an export coefficient model. *J. Hydrol.* **2012**, *414–415*, 31–39. [[CrossRef](#)]
29. Xu, Y.; Zhang, X.; Wang, X.; Hao, Z.; Singh, V.P.; Hao, F. Propagation from meteorological drought to hydrological drought under the impact of human activities: A case study in northern China. *J. Hydrol.* **2019**, *579*, 124147. [[CrossRef](#)]
30. Wu, L.; Zhang, X.; Hao, F.; Wu, Y.; Li, C.; Xu, Y. Evaluating the contributions of climate change and human activities to runoff in typical semi-arid area, China. *J. Hydrol.* **2020**, *590*, 125555. [[CrossRef](#)]
31. Li, D.; Bu, S.; Chen, S.; Li, Q.; Li, Y. Assessment of the impact of short-term land use/land cover changes on water resources in the Yanghe reservoir basin, China. *Water Supply* **2021**, *29*, 259. [[CrossRef](#)]
32. Li, D.; Bu, S.; Li, Q.; Chen, S.; Zhen, Z.; Fu, C. Water environment capacity estimation and pollutant reduction of Yanghe Reservoir Basin in Hebei Province, China, via 0-D water quality model. *Environ. Earth Sci.* **2021**, *80*, 380. [[CrossRef](#)]
33. Petrushevsky, N.; Manzoni, M.; Monti-Guarnieri, A. Fast Urban Land Cover Mapping Exploiting Sentinel-1 and Sentinel-2 Data. *Remote Sens.* **2022**, *14*, 36. [[CrossRef](#)]
34. Smil, V. Nitrogen in crop production: An account of global flows. *Glob. Biogeochem. Cycles* **1999**, *13*, 647–662. [[CrossRef](#)]
35. Shrestha, S.; Kazama, F.; Newham, L.T.H. A framework for estimating pollutant export coefficients from long-term in-stream water quality monitoring data. *Environ. Model. Softw.* **2008**, *23*, 182–194. [[CrossRef](#)]
36. Lu, J.; Gong, D.; Shen, Y.; Liu, M.; Chen, D. An inversed Bayesian modeling approach for estimating nitrogen export coefficients and uncertainty assessment in an agricultural watershed in eastern China. *Agric. Water Manag.* **2013**, *116*, 79–88. [[CrossRef](#)]
37. Liu, Y.; Xu, Z.; Li, C. Online semi-supervised support vector machine. *Inf. Sci.* **2018**, *439–440*, 125–141. [[CrossRef](#)]
38. Zhang, L.; Zhou, W.; Jiao, L. Wavelet support vector machine. *IEEE Trans. Syst. Man Cybern. B Cybern.* **2004**, *34*, 34–39. [[CrossRef](#)]
39. Suykens, J.A.K.; Vandewalle, J. Least squares support vector machine classifiers. *Neural Process. Lett.* **1999**, *9*, 293–300. [[CrossRef](#)]
40. Congalton, R.G. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.* **1991**, *37*, 35–46. [[CrossRef](#)]
41. Paerl, H.W.; Xu, H.; McCarthy, M.J.; Zhu, G.; Qin, B.; Li, Y.; Gardner, W.S. Controlling harmful cyanobacterial blooms in a hyper-eutrophic lake (Lake Taihu, China): The need for a dual nutrient (N & P) management strategy. *Water Res.* **2011**, *45*, 1973–1983. [[CrossRef](#)] [[PubMed](#)]
42. Prasad, R.; Deo, R.C.; Li, Y.; Maraseni, T. Weekly soil moisture forecasting with multivariate sequential, ensemble empirical mode decomposition and Boruta-random forest hybridizer algorithm approach. *CATENA* **2019**, *177*, 149–166. [[CrossRef](#)]
43. Chen, H.; Liu, H.; Chen, X.; Qiao, Y. Analysis on impacts of hydro-climatic changes and human activities on available water changes in Central Asia. *Sci. Total Environ.* **2020**, *737*, 139779. [[CrossRef](#)] [[PubMed](#)]
44. Kumar, P.P.; Bai, V.M.A.; Nair, G.G. An efficient classification framework for breast cancer using hyper parameter tuned Random Decision Forest Classifier and Bayesian Optimization. *Biomed. Signal Process. Control.* **2021**, *68*, 102682. [[CrossRef](#)]

45. Sun, D.; Xu, J.; Wen, H.; Wang, D. Assessment of landslide susceptibility mapping based on Bayesian hyperparameter optimization: A comparison between logistic regression and random forest. *Eng. Geol.* **2021**, *281*, 105972. [[CrossRef](#)]
46. Ehteram, M.; Salih, S.Q.; Yaseen, Z.M. Efficiency evaluation of reverse osmosis desalination plant using hybridized multilayer perceptron with particle swarm optimization. *Environ. Sci. Pollut. Res.* **2020**, *27*, 15278–15291. [[CrossRef](#)]
47. Xu, Y.; Zhang, X.; Hao, Z.; Hao, F.; Li, C. Projections of future meteorological droughts in China under CMIP6 from a three-dimensional perspective. *Agric. Water Manag.* **2021**, *252*, 106849. [[CrossRef](#)]
48. Sousa, S.; Martins, F.; Alvimferraz, M.; Pereira, M. Multiple linear regression and artificial neural networks based on principal components to predict ozone concentrations. *Environ. Model. Softw.* **2007**, *22*, 97–103. [[CrossRef](#)]
49. Chen, P.; Fu, C.F.; Ji, X.G.; Li, D.M. Spatial Distribution of Non-point Source Pollution Loading in Yanghe Reservoir Watershed. *J. Hydroecol.* **2018**, *39*, 58–64.
50. Wang, H.; Xu, J.; Liu, X.; Sheng, L.; Di, Z.; Li, L.; Wang, A. Study on the pollution status and control measures for the livestock and poultry breeding industry in northeastern China. *Environ. Sci. Pollut. Res.* **2018**, *25*, 4435–4445. [[CrossRef](#)]
51. Yang, Y.; Luan, W. Industrial Structure and COD Emission of Livestock and Poultry Breeding in Liaoning Province, NE China: Empirical Research on the Panel Threshold Model. *IOP Conf. Ser. Earth Environ. Sci.* **2018**, *186*, 12019. [[CrossRef](#)]
52. Huang, H.; Chen, D.; Zhang, B.; Zeng, L.; Dahlgren, R.A. Modeling and forecasting riverine dissolved inorganic nitrogen export using anthropogenic nitrogen inputs, hydroclimate, and land-use change. *J. Hydrol.* **2014**, *517*, 95–104. [[CrossRef](#)]
53. Zhang, W.; Swaney, D.P.; Hong, B.; Howarth, R.W.; Han, H.; Li, X. Net anthropogenic phosphorus inputs and riverine phosphorus fluxes in highly populated headwater watersheds in China. *Biogeochemistry* **2015**, *126*, 269–283. [[CrossRef](#)]
54. Zhang, W.; Li, H.; Li, Y. Spatio-temporal dynamics of nitrogen and phosphorus input budgets in a global hotspot of anthropogenic inputs. *Sci. Total Environ.* **2019**, *656*, 1108–1120. [[CrossRef](#)]
55. Wagner, T.; Lottig, N.R.; Bartley, M.L.; Hanks, E.M.; Schliep, E.M.; Wikle, N.B.; King, K.B.S.; McCullough, I.; Stachelek, J.; Cheruvilil, K.S.; et al. Increasing accuracy of lake nutrient predictions in thousands of lakes by leveraging water clarity data. *Limnol. Oceanogr. Lett.* **2020**, *5*, 228–235. [[CrossRef](#)]
56. Maulik, U.; Chakraborty, D. Learning with transductive SVM for semisupervised pixel classification of remote sensing imagery. *ISPRS J. Photogramm. Remote Sens.* **2013**, *77*, 66–78. [[CrossRef](#)]
57. Song, C.; Woodcock, C.E.; Seto, K.C.; Lenney, M.P.; Macomber, S.A. Classification and change detection using Landsat TM data: When and how to correct atmospheric effects? *Remote Sens. Environ.* **2001**, *75*, 230–244. [[CrossRef](#)]
58. Boyer, E.W.; Goodale, C.L.; Jaworski, N.A.; Howarth, R.W. Anthropogenic nitrogen sources and relationships to riverine nitrogen export in the northeastern U.S.A. *Biogeochemistry* **2002**, *57*, 137–169. [[CrossRef](#)]
59. Zhong, L.-P.; Cao, Y.; Li, W.-Y.; Pan, W.-P.; Xie, K.-C. Effect of the existing air pollutant control devices on mercury emission in coal-fired power plants. *J. Fuel Chem. Technol.* **2010**, *38*, 641–646. [[CrossRef](#)]
60. Yan, W.; Zhang, S.; Sun, P.; Seitzinger, S.P. How do nitrogen inputs to the Changjiang basin impact the Changjiang River nitrate: A temporal analysis for 1968–1997. *Glob. Biogeochem. Cycles* **2003**, *17*. [[CrossRef](#)]
61. Han, Y.; Yu, X.; Wang, X.; Wang, Y.; Tian, J.; Xu, L.; Wang, C. Net anthropogenic phosphorus inputs (NAPI) index application in Mainland China. *Chemosphere* **2013**, *90*, 329–337. [[CrossRef](#)]
62. Hong, B.; Swaney, D.P.; Howarth, R.W. A toolbox for calculating net anthropogenic nitrogen inputs (NANI). *Environ. Model. Softw.* **2011**, *26*, 623–633. [[CrossRef](#)]
63. Zhu, M. *Study on Agricultural NPS Loads of Haihe Basin and Assessment on Its Environmental Impact*; Chinese Academy of Agricultural Sciences: Beijing, China, 2011.
64. Shi, J.H. *Characteristics of Agricultural Nonpoint Source Pollution and Farmland Nutrients Management in Plain Areas of Baiyangdian Lake Basin*. Master's Thesis, Beijing Normal University, Beijing, China, 2012.