



Article

Spatial and Temporal Patterns of Grassland Species Diversity and Their Driving Factors in the Three Rivers Headwater Region of China from 2000 to 2021

Mingxin Yang ^{1,2,3}, Ang Chen ², Wenqiang Cao ¹, Shouxin Wang ¹, Mingyuan Xu ¹, Qiang Gu ¹, Yanhe Wang ^{1,4} and Xiuchun Yang ^{2,*}

¹ Xining Natural Resources Comprehensive Survey Center, China Geological Survey, Xining 810021, China; ymxin@bjfu.edu.cn (M.Y.); guqiang@mail.cgs.gov.cn (Q.G.)

² School of Grassland Science, Beijing Forestry University, Beijing 100081, China; chenang0226@bjfu.edu.cn

³ Key Laboratory of Coupling Process and Effect of Natural Resources Elements, Beijing 100055, China

⁴ Observation Station of Subalpine Ecology Systems in the Middle Qilian Mountains, Zhangye 734000, China

* Correspondence: yangxiuchun@bjfu.edu.cn

Abstract: Biodiversity loss will lead to a serious decline for ecosystem services, which will ultimately affect human well-being and survival. Monitoring the spatial and temporal dynamics of grassland biodiversity is essential for its conservation and sustainable development. This study integrated ground monitoring data, Landsat remote sensing, and environmental variables in the Three Rivers Headwater Region (TRHR) from 2000 to 2021. We established a reliable model for estimating grassland species diversity, analyzed the spatial and temporal patterns, trends of change, and the driving factors of changes in grassland species diversity over the past 22 years. Among models based on diverse variable selection and machine learning methods, the random forest (RF) combined stepwise regression (STEP) model was found to be the optimal model for estimating grassland species diversity in this study, which had an R^2 of 0.44 and an RMSE of 2.56 n/m² on the test set. The spatial distribution of species diversity showed a pattern of abundance in the southeast and scarcity in the northwest. Trend analysis revealed that species diversity was increasing in 80.46% of the area, whereas 16.59% of the area exhibited a decreasing trend. The analysis of driving factors indicated that the changes in species diversity were driven by both climate change and human activities over the past 22 years in the study area, of which temperature was the most significant driving factor. This study effectively monitors grassland species diversity on a large scale, thereby supporting biodiversity monitoring and grassland resource management.

Keywords: species diversity; machine learning; spatial and temporal changes; remote sensing; driving factors; the Three Rivers Headwater Region



Citation: Yang, M.; Chen, A.; Cao, W.; Wang, S.; Xu, M.; Gu, Q.; Wang, Y.; Yang, X. Spatial and Temporal Patterns of Grassland Species Diversity and Their Driving Factors in the Three Rivers Headwater Region of China from 2000 to 2021. *Remote Sens.* **2024**, *16*, 4005. <https://doi.org/10.3390/rs16214005>

Academic Editor: Nikos Koutsias

Received: 23 September 2024

Revised: 19 October 2024

Accepted: 22 October 2024

Published: 28 October 2024



Copyright: © 2024 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

Due to global climate change and its impacts on human activities, biodiversity is being lost at an unprecedented rate [1]. The loss of biodiversity will lead to a serious decline in ecosystem service functions, which will eventually affect human well-being and survival [2–4]. Grassland accounts for about 40% of the total terrestrial ecosystem area and plays an important role in human production and life [5]. Monitoring the spatial and temporal dynamics of grassland biodiversity is essential for grassland biodiversity conservation and maintaining the sustainable development of grassland ecosystem service functions [6,7].

The ground sample plot survey method is typically used for grassland species diversity monitoring; although it is the most accurate method for obtaining biodiversity monitoring data [8], it needs substantial manpower, is expensive, and requires high species identification ability. Moreover, it is impracticable in some uninhabited areas characterized

by a harsh environment [9]. In contrast, remote sensing monitoring methods, with their multiple temporal and spatial resolutions, are powerful means for monitoring patterns and trend changes in species diversity, as well as assisting in conservation efforts at large scales and over long periods [10]. The remote sensing spectral variation hypothesis posits that the greater the environmental heterogeneity of vegetation, the greater the species diversity [11,12]. Given that remotely sensed spectral heterogeneity can reflect environmental heterogeneity to a certain extent, it is often used as an indicator for monitoring the diversity of grassland vegetation species [13–16]. However, the accuracy of monitoring grassland species diversity using remote sensing spectra alone is limited at present, and the application of high-precision simulation using machine learning for monitoring grassland species diversity over long time series and large scales is not yet common. The accurate monitoring of changes in grassland species diversity requires not only remote sensing data, but also comprehensive consideration of environmental factors, such as climate, soil, and geography, all of which can affect the diversity of grassland species [17,18]. In this regard, machine learning has been widely applied in ecological monitoring in recent years owing to its ability to estimate complex relationships among multiple variables, which is more robust than traditional models and effectively improves model prediction accuracy [19,20]. Species diversity monitoring over a long time series can effectively reveal the dynamic changes in species diversity, thereby efficiently guiding biodiversity conservation. Currently, research on monitoring the dynamic changes in grassland species diversity and trend prediction is relatively scarce [21].

Studying the drivers of the spatial and temporal dynamics and distribution patterns of grassland species diversity is crucial for understanding grassland biodiversity as well as for its conservation. The dynamics of grassland species diversity are influenced not only by climatic factors, such as temperature and precipitation, but also by human activities, such as regional grazing management, population density, and ecological conservation projects [22,23]. Yao et al. [24] reported that species diversity increases significantly with precipitation but decreases with temperature in large-scale grasslands, while Zhang et al. [25] showed that the effects of long-term grazing on biodiversity and ecosystem functioning were related to the degree of drought. Numerous studies have analyzed the effects of temperature and precipitation changes on grassland species diversity [24,26]; however, relatively few studies have been conducted on the effects of human activities, and even fewer studies have been undertaken on the integrated analysis of the effects of climate change and human activities on changes in grassland species diversity, which has greatly constrained the conservation and restoration of grassland biodiversity.

The Qinghai-Tibetan Plateau, a globally important ecological barrier area, is characterized by typical alpine grassland ecosystems, with large spatial differences in vegetation growth and extreme sensitivity to climate change [27]. The TRHR, as the core area of the Tibetan Plateau, is an important water conservation area in Asia and is also one of the richest and most concentrated areas of alpine ecosystems in terms of species and genes. Accordingly, the TRHR is of significant importance for the conservation of alpine biodiversity globally, while simultaneously playing an important role in the maintenance of regional ecological functions [28]. Over recent years, the grassland ecosystem in the TRHR has been seriously degraded due to climate change and the interference of human activities. To address this, the Chinese government implemented a series of ecological protection projects and the degradation of grassland in this area has been effectively curbed [29]. Biodiversity has changed dramatically during this process, highlighting the importance of monitoring the spatial and temporal patterns of grassland species diversity in the region for the conservation of alpine ecosystems and germplasm resources. In this study, we sought to accurately analyze the spatial and temporal patterns of grassland vegetation species diversity and their driving factors in the TRHR between 2000 and 2021 based on a long-term grassland species diversity survey. The specific objectives included (1) the establishment of a reliable large-spatial-scale grassland species diversity estimation model based on machine learning; (2) analyzing the spatial and temporal patterns of grassland

species diversity in the TRHR from 2000 to 2021; and (3) revealing the impact of climate change and human activities on the spatial and temporal dynamics of grassland species diversity in the TRHR.

2. Materials and Methods

2.1. Study Area

In this work, the study area was the TRHR, which is located in the east-central part of the Tibetan Plateau. The geographic location ranges from 89°24'E to 102°41'E and 31°39'N to 36°16'N (Figure 1a), comprising a total area of approximately 395,000 km² with an average elevation of more than 4000 m; the TRHR has a harsh climate throughout the year, with an average annual temperature of between −5.4 and 6.9, and an average annual precipitation of between 392 and 764 mm. The Three River Headwaters National Park is the largest national park in China as well as the birthplace of three major rivers: the Yangtze, the Yellow, and the Lancang. It is one of the areas with the highest species diversity among high-elevation regions of the world [21]. The grasslands of the TRHR consist mainly of alpine meadows, alpine grasslands, alpine deserts, and temperate grasslands, with high species richness and diversity, and a complex community structure (Figure 1b).

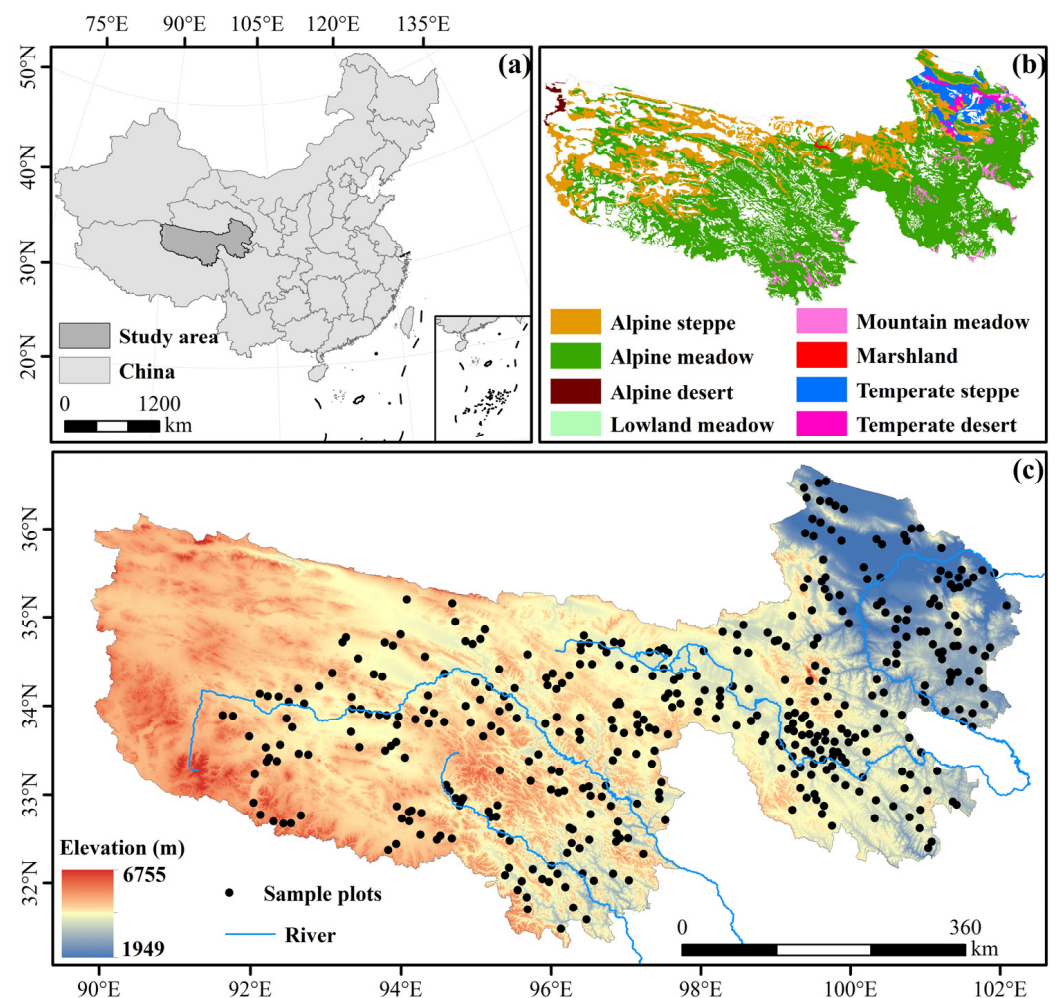


Figure 1. Overview of the study area. (a) Locations of the study area; (b) grassland types of the study area; (c) elevation of the study area and the distribution of sample plots.

2.2. Data Sources and Processing

2.2.1. Grassland Field Data

Species richness is one of the indices of species diversity, so we collected and screened data from 2136 ground survey plots in the study area between 2000 and 2021 (Figure 1c). The plot survey information mainly included latitude and longitude, grassland type, species richness, and biomass of each square, and survey sampling was conducted in the grassland growing season of July–August in each year. Data sources included two parts—the first part comprised continuous monitoring data of the Qinghai Provincial Grassland Station from 2005 to 2018, totaling 1839 sample plots; the second part consisted of data relating to the field survey of the project area in 2021, totaling 297 sample plots (Table S1). Both parts of the data were collected based on the *Grassland Resources Survey Technical Regulation* issued by the Ministry of Agriculture to ensure the standardization of field survey data collection. For the field survey, representative survey sample plots were set up in areas with relatively uniform grass vegetation and the same grassland type. According to the evenness of vegetation communities and species uniformity within the sample plots, 3–5 sample squares (1×1 m) were appropriately set up within each sample plot and the indexes were investigated following their localization with GPS. In each sample square, species richness was recorded as the number of all species occurring within the square. The average value of species richness of all the sample squares within each sample plot was used as the species diversity index for that plot.

2.2.2. Remotely Sensed Vegetation Index and Preprocessing

The remotely sensed vegetation index was invoked through the Google Earth Engine (GEE) platform from Landsat 5/7/8 data with a spatial resolution of 30 m. Landsat TM5 was used in 2000–2011, Landsat ETM+7 in 2012, and Landsat OLI8 in 2013–2021. Six vegetation indices—enhanced vegetation index (EVI), ratio vegetation index (RVI), soil-adjusted vegetation index (SAVI), normalized difference vegetation index (NDVI), green normalized vegetation index (GNDVI) and kernel normalized difference vegetation index (KNDVI)—were calculated and downloaded from the remote sensing data relating to the vegetation growing season in the study area from May to October using GEE (Table S2). The vegetation indices were downloaded using UTM projection and WGS_84 datum in tif format. The spatially and temporally matched vegetation indices were extracted in ArcGIS software, and the mean value of the vegetation indices of all the squares in each sample plot was taken as the vegetation index of that plot.

2.2.3. Data for Other Variables

A series of meteorological, soil, geographic, and human activity data from the study area over 22 years were collected and downloaded to support the modeling of species diversity and the analysis of drivers of species diversity change. The sources and overview of the data are listed below.

Climate data were obtained from the National Geoscience Data Center (<http://www.geodata.cn>, accessed on 19 April 2022) and month-by-month data for precipitation, temperature, and potential evapotranspiration from 2000 to 2021 were downloaded using an application at a spatial resolution of 1 km. The data were downloaded in nc format and were then converted to tif format in ArcGIS; the average values of the 12 months for each year were calculated by raster as mean annual precipitation (MAP), mean annual temperature (MAT), and mean annual potential evapotranspiration (MET). The May to October averages were used as growing season mean precipitation (GMAP), growing season mean temperature (GMAT), and growing season mean potential evapotranspiration (GMET).

Soil attribute data were obtained from the basic attribute dataset of China's high-resolution National Soil Information Grid [30], which consists of 90 m spatial resolution data in tif format, and includes the following 12 factors: pH, soil organic carbon (SOC) content, total nitrogen (TN) content, total phosphorus (TP) content, total potassium (TK) content, bulkiness (SBD), gravel content (CF), sand (SND) content, powder (SLT) content,

clay (CLY) content, soil texture type (CLS), and soil thickness (THK) at six soil depths—0~5, 5~15, 15~30, 30~60, 60~100, and 100~200 cm.

The digital elevation model (DEM) data were downloaded from ASTER GDEM at 30 m spatial resolution through the Earth Remote Sensing Data Analysis Center (<https://www.geodata.cn>, accessed on 16 March 2023); elevation (ELE), slope (SLOPE), and slope direction (ASPECT) data were extracted separately using ArcGIS software.

To keep the spatial resolution and coordinate system of all the variable data of the study consistent, we resampled the preprocessed meteorologic and soil data to 30 m, unified the projection and spatial reference coordinate system to UTM and WGS_84, and then extracted the spatial and temporal matches of the 22 above-mentioned variable values into the sample plot data. In addition, the grassland category data within the downloaded LUCC2020 surface classification data were extracted with the appropriate amount of the study area boundary to obtain the grassland cover data for the grassland boundary mask in the study area.

To reveal which human activities most influence regional grassland species diversity, three proxy variables—population density (POP), gross domestic product (GDP) per capita, and livestock number (LSK)—were collected. The POP data were obtained from WorldPop (<http://www.WorldPop.org/>, accessed on 11 September 2023), which has a spatial resolution of 1 km and a time resolution from 2000 to 2020. The GDP data were obtained from the China GDP Spatial Distribution Kilometer Grid dataset (<http://www.resdc.cn/>, accessed on 23 October 2023), which has a spatial resolution of 1 km and a time resolution from 1995 to 2019. The LSK data were derived from the Tibetan Plateau Livestock Raster Dataset from 1985 to 2015, which has a spatial resolution of 10 km [31].

2.3. Construction and Assessment of Grassland Species Diversity Models

2.3.1. Variable Selection

We considered as many spectral and environmental variables affecting species diversity as possible in the model construction; however, studies have shown that more variables do not necessarily lead to higher model accuracy, and the use of appropriate variable selection methods can not only remove the multicollinearity problem among variables and improve model accuracy but can also reduce data redundancy and improve the efficiency of model computation [32]. Therefore, in this study, four commonly used variable selection methods, Genetic Algorithm (GA), Recursive Elimination of Features (REF), Stepwise Regression (STEP), and Lasso Regression (LASSO), were used to screen the preliminary constructed features of 27 variables. The four variable selection methods are described in the Supplementary Material. The performance of the four methods was evaluated by modeling the results obtained with each method and comparing the accuracy of the models. All the variable selection methods were implemented in RStudio.

2.3.2. Construction and Assessment of Machine Learning Models

In this study, species diversity estimated in the ground survey was used as the dependent variable, and 27 variables consisting of the vegetation index and other environmental variables were used as the independent variables to construct a sample dataset containing the ground species diversity, remotely sensed vegetation index, and environmental variables. This dataset was modeled using a machine learning algorithm, in which 70% of the sample data was used as the model training data, and the remaining 30% was used as the model testing data. Using the training data for model fitting, hyperparameter tuning was performed using Bayesian optimization with 10-fold cross-validation; the resulting hyperparameters corresponding to the model with the smallest error were selected and the final model was obtained by training again on the whole model training data. The accuracy of the final model was evaluated using the test data.

Different machine learning methods perform differently in different studies and with different datasets. In this study, to achieve the optimal model simulation, we selected four machine learning methods, Extreme Gradient Boosting (XGboost), Random Forest (RF),

Knott's Nest (KNN), and Support Vector Machines (SVM), which are commonly used and have superior performance, to compare model training and accuracy. The model with the best robustness was selected as the machine learning model for the inverse mapping of species diversity. The four machine learning methods are described in the Supplementary Material. Model training, hyperparameter optimization, and accuracy computation using the four machine learning methods were implemented in RStudio.

For accuracy evaluation, the coefficient of determination (R^2), root mean square error (RMSE) and mean absolute error (MAE) were used as constant quantities. These metrics were calculated using the following formulae:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}} \quad (1)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (2)$$

$$\text{MAE} = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (3)$$

where n is the number of sample plots, y_i is the model predicted value for the i th sample plot, $\bar{y}_i$ is the measured value for the i th sample plot, and $\hat{y}_i$ is the average of the measured values.

2.4. The Spatiotemporal Dynamics of Grassland Species Diversity

2.4.1. Spatial Distribution

The constructed optimal species diversity model together with the spatial dataset of independent variables for each year was used to perform model inversion to obtain the spatial distribution dataset of grassland species diversity in the study area between 2000 and 2021. Additionally, the data for these 22 years were mosaicked in ArcGIS to obtain the average spatial distribution of grassland species diversity in the study area, and the spatial distribution pattern of species diversity can be discerned from the spatial distribution data.

2.4.2. Trend in Spatiotemporal Changes

The change in species diversity in the time series was analyzed image-by-image using the Theil-Sen Median (Sen's slope) method, which is a robust nonparametric statistical method for trend calculation [33]. When the slope < -0.0005 , the diversity of grassland species is decreasing; when $-0.0005 < \text{slope} < 0.0005$, grassland species diversity is stable; when the slope ≥ 0.0005 , grassland species diversity is increasing. The significance of changes in species diversity was further tested using the Mann-Kendall (MK) method, which is a nonparametric test suitable for detecting changes in time series [34]. $|Z| > 1.96$ indicated that the Z-value calculated from the significance test passed the 95% confidence test, meaning that the change was significant, while $|Z| < 1.96$ indicated that the change did not pass the 95% confidence test, which meant that it was not significant.

The combined Sen slope + MK approach classified the trends in species diversity into five classes (Table S3). Species diversity trends were all calculated in RStudio using the trend package, and classification statistics were completed in ArcGIS.

2.4.3. Future Trends

The Hurst index can effectively predict future trends relative to past time series data. In this study, the Hurst index was evaluated based on the re-scaled extreme deviation (R/S) analysis method, and the Hurst index of the time series ranged from 0–1. When Hurst > 0.5 , this indicates that the time series has continuity, i.e., the future trend is in line with the past trend of the time series; when Hurst < 0.5 , the time series has opposite continuity, i.e., the future trend is opposite to the past time series trend; when Hurst = 0.5, the time series has uncertainty, i.e., it is not possible to predict the future trend of change with the past time

series trend. In this study, Hurst and the slope of species diversity calculations for 2000 to 2021 were superimposed and analyzed to predict the future trend of species diversity; the trend was defined as a continuous increase when $Hurst > 0.5$ and the slope ≥ 0.0005 , i.e., future species diversity continued to show an increasing trend; when $Hurst > 0.5$ and the slope < -0.0005 , the trend was defined as a continuous decrease, i.e., species diversity continued to decrease in the future; when $Hurst < 0.5$ and the slope ≥ 0.0005 , the trend was defined as anti-continuously increasing, i.e., species diversity has been increasing in the past but will decrease in the future; when $Hurst < 0.5$ and the slope < -0.0005 , the trend was defined as anti-continuously decreasing, i.e., species diversity has been in a decreasing trend in the past, and will increase in the future.

2.5. Detection of the Driving Factors of Grassland Species Diversity Dynamics

The geographical detector model is a method used for geospatial analysis to detect the influencing mechanism of potential variables from the perspective of spatial heterogeneity. The optimal parameters-based geographical detector (OPGD) model was developed to address issues related to spatial data discretization and spatial stratification discontinuity number, among others, associated with the geographical detector model, so as to improve parameter optimization and the accuracy of spatial analysis [35]. Factor detectors are mainly used to analyze the effect of latent variables on the explained factors, and there are many modules, such as single factor detection and interaction detection. In this study, we used the slope of species diversity from 2000 to 2021 as a response variable, and the slopes of change in climate change variables and human variables from 2000 to 2021 as latent variables to analyze the impacts on grassland species diversity dynamics using one-way probes and interaction probes, respectively. The climate change variables included mean annual precipitation (MAP), mean annual temperature (MAT), and mean annual potential evapotranspiration (MET) (see Section 2.2.3), while the human variables comprised population density (POP), gross domestic product (GDP) per capita, and the number of livestock (LSK) within the study area. Finally, slope of change calculations were performed for the six variables over 22 years, i.e., SMAP, SMAT, SMET, SPOP, SGDP, and SLSK, and their spatial resolution was standardized to 10 km. They were then transformed into categorical variables and subjected to q-value calculations and analyses using the GD package in RStudio.

3. Results

3.1. Variable Selection and Model Accuracy Evaluation

Four variable selection methods—GA, REF, STEP, and LASSO—were used to screen 27 variable characteristics (Table 1). The results showed that all four methods streamlined the 27 variables; LASSO selected 24 variables, REF selected 21 variables, STEP selected 17 variables, and GA selected 12 variables. In addition, the results of all four variable selections included vegetation indices, climate, soil, and geographic variables, and seven variables—KNDVI, MAT, MAP, MET, THK, SBD, and ELE—were selected by all four methods.

Twenty species diversity estimation models were constructed using the results of the four variable selections above and all 27 variables (ALL) with the four machine learning methods—XGboost, RF, KNN, and SVM—respectively (Table 2). Regarding the performance of the four models, the R^2 of the model training set ranged between 0.35 and 0.88, while that of the model testing set ranged from 0.34 to 0.44; the RF model outperformed the other three in terms of both training and testing, followed by XGboost, SVM, and KNN. Regarding model accuracy, the models constructed based on the variables selected by the STEP and REF methods had the best accuracy, followed by LASSO; the model constructed using the variables selected by GA displayed the lowest accuracy. Additionally, the accuracy of the model constructed based on the GA method was lower than that of the model constructed with ALL, indicating that the GA variable selection method did not have the effect of optimizing the model's accuracy in this study; meanwhile, the STEP,

REF, and LASSO methods relatively streamlined the variable multiple covariance and data redundancy problems of the model, while simultaneously improving the model's accuracy.

Table 1. The results of variable selection.

Method	Variables Selected	Number
GA	KNDVI MAT MAP GMAP MET CF SND THK TK SBD ELE SLOPE	12
REF	EVI NDVI GNDVI KNDVI RVI SAVI MAT GMAT MAP GMAP MET GMET THK SBD TN TK SOC TP SND PH ELE	21
STEP	EVI KNDVI RVI SAVI MAT MAP MET GMET CF CLY SLT SND SOC THK TN SBD ELE	17
LASSO	EVI GNDVI KNDVI RVI SAVI MAT GMAT MAP GMAP MET GMET CF CLS CLY SLT SOC THK TK TN TP SBD ELE ASP SLOP	24

Table 2. Model construction and accuracy evaluation.

Model Method and Variable Selection Method	Training Set			Test Set		
	RMSE	R ²	MAE	RMSE	R ²	MAE
XGboost-ALL	2.25	0.58	1.79	2.63	0.40	2.12
XGboost-GA	2.03	0.67	1.58	2.66	0.39	2.11
XGboost-REF	2.38	0.53	1.90	2.62	0.41	2.11
XGboost-STEP	2.38	0.52	1.88	2.63	0.40	2.12
XGboost-LASSO	2.30	0.56	1.83	2.66	0.39	2.13
RF-ALL	1.55	0.84	1.22	2.57	0.43	2.06
RF-GA	1.38	0.88	1.09	2.61	0.41	2.10
RF-REF	1.39	0.88	1.10	2.58	0.42	2.08
RF-STEP	1.49	0.85	1.19	2.56	0.44	2.06
RF-LASSO	1.61	0.83	1.28	2.59	0.42	2.07
KNN-ALL	2.56	0.44	2.06	2.71	0.36	2.18
KNN-GA	2.54	0.45	2.04	2.76	0.34	2.21
KNN-REF	2.43	0.49	1.93	2.71	0.37	2.16
KNN-STEP	2.52	0.46	2.00	2.66	0.39	2.13
KNN-LASSO	2.55	0.44	2.04	2.70	0.37	2.17
SVM-ALL	2.69	0.38	2.13	2.63	0.40	2.10
SVM-GA	2.76	0.35	2.19	2.71	0.37	2.17
SVM-REF	2.71	0.37	2.14	2.62	0.41	2.09
SVM-STEP	2.70	0.38	2.14	2.61	0.41	2.08
SVM-LASSO	2.70	0.38	2.13	2.62	0.41	2.09

A comprehensive comparison of the models constructed using the different variable selection methods combined with machine learning indicated that the RF-STEP method had the highest accuracy in this study (Figure S1), with an R² of 0.85, an RMSE of 1.49 n/m², and an MAE of 1.19 n/m² in the model training set, and an R² of 0.44, an RMSE of 2.56 n/m², and an MAE of 2.06 n/m² in the model testing set. Accordingly, the model constructed with the RF-STEP method was considered the best for estimating grassland species diversity in this study.

3.2. Patterns of the Spatial Distribution of the Average Grassland Species Diversity from 2000 to 2021

Based on the selected optimal model, the grassland species diversity in the study area from 2000 to 2021 was inverted, and the spatial distribution was mapped (Figure S2). Regarding spatial distribution, species diversity showed a pattern of richness in the south-east and scarcity in the northwest. Furthermore, while the overall spatial distribution characteristics were similar between years, there were also local annual differences, and the species diversity in the east-central region showed an increasing trend with time.

The interannual dynamics of species diversity in the study area over 22 years were analyzed by counting the annual mean values of species diversity for each year (Figure 2b). For the annual time series, the species diversity in the study area generally showed a fluctuating but modest increase (an average of 1 n/m^2) over 22 years, with the lowest average species diversity of 7.44 n/m^2 being recorded in 2008, while the highest (8.31 n/m^2) was noted in 2020. From 2000 to 2008, species diversity showed insignificant interannual fluctuations, with a slope of -0.002 while, from 2008 onwards, species diversity showed a significantly increasing trend, with a slope of 0.036.

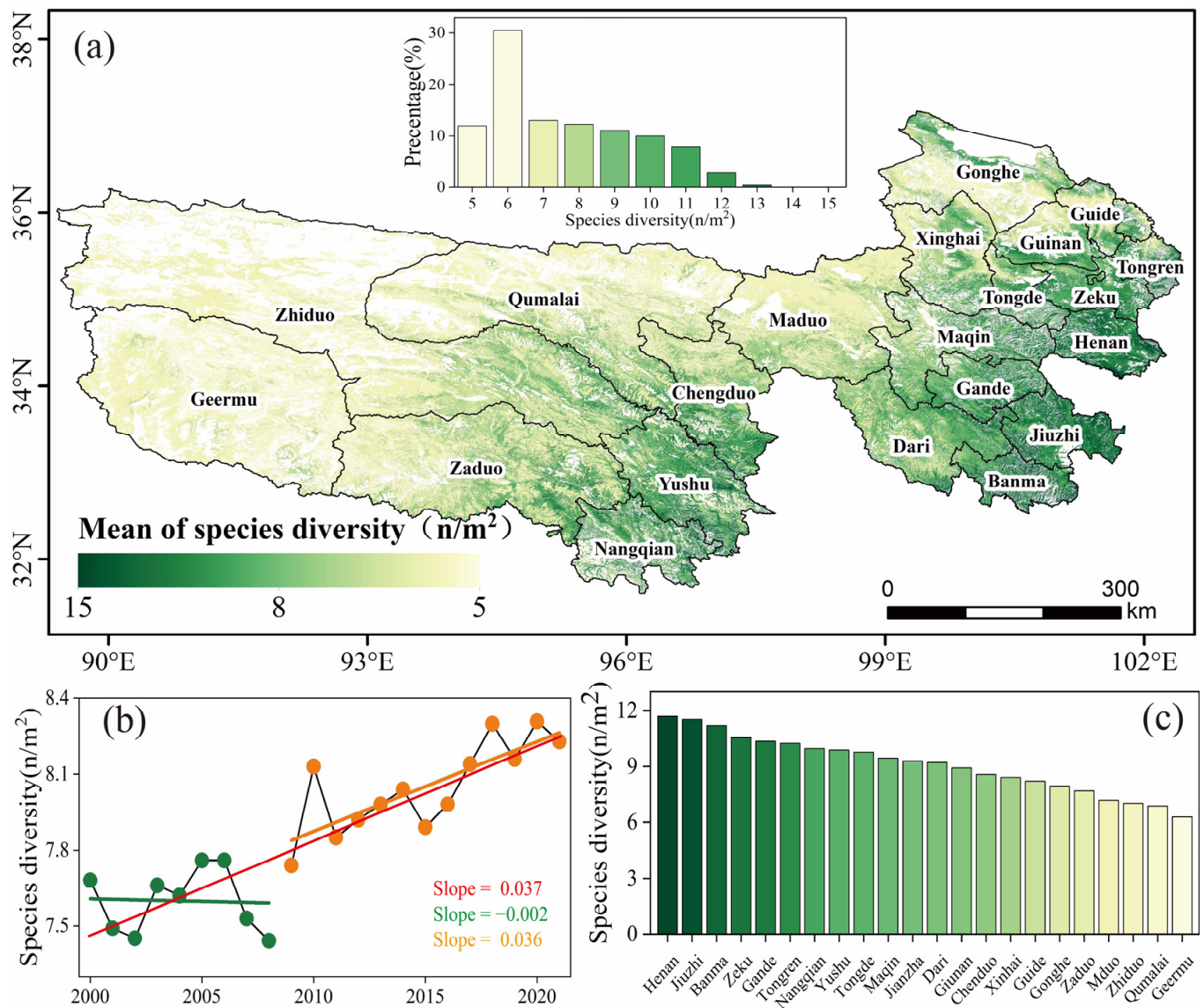


Figure 2. Spatial distribution of species diversity. (a) Spatial distribution of average species diversity from 2000–2021; (b) inter-annual variation in species diversity; (c) species diversity by county.

In addition, we mosaicked the mean species diversity between 2000 and 2021 to obtain the average spatial distribution of species diversity over the past 22 years (Figure 2a). The results showed that the areas with rich species diversity were mainly distributed in Henan and Zeku counties in Huangnan Prefecture; Guinan and Tongde counties in Hainan Prefecture; and Maqin, Gande, Jiuzhi, and Banma counties in Guoluo Prefecture. The areas with sparse species diversity were mainly distributed in Geermu, Qumalai, Zhiduo, Zaduo, etc. (Figure 2c). Quantitative analysis indicated that the average species richness in the

study area over the past 22 years was 8.08 n/m^2 , with a minimum value of 5.05 n/m^2 , a maximum value of 15.29 n/m^2 , and a root mean square error of 1.98 n/m^2 .

3.3. Spatial and Temporal Trends in Grassland Species Diversity Between 2000 and 2021

The spatial and temporal trends of grassland species diversity in the study area from 2000 to 2021 were obtained using Sen's slope analysis and the MK trend test of the 22-year species diversity image-by-image element (Figure 3). The results showed that species diversity in the study area showed an increasing trend over the 22 years in most areas, accounting for 80.46% of the total area, of which the significantly increasing areas were mainly distributed in Xinghai, Guinan, Guide, Tongde in Hainan Prefecture, Dari in Guoluo Prefecture, and localized areas of Zhiduo and Zado in Yushu Prefecture, accounting for 27.86%; slightly increasing areas covered the whole study area, accounting for 52.60% of the total area. Over the past 22 years, while species diversity increased in most regions, there were also localized areas showing a decreasing trend in species diversity, accounting for 16.59% of the total study area; the slight decreasing area was relatively small, accounting for 0.93% of the total, while the rest (15.66%) displayed a significantly decreasing trend. Areas showing a decreasing trend were mainly distributed in localized areas of Gande, Bama, Jiuzhi, and Maduo counties in Guoluo Prefecture; Yushu, Chenduo, Qumalai, and Zhiduo counties in Yushu Prefecture; and Geermu City. In addition, 2.95% of the area in the western part of the study area showed unchanging grassland species diversity over the last 22 years.

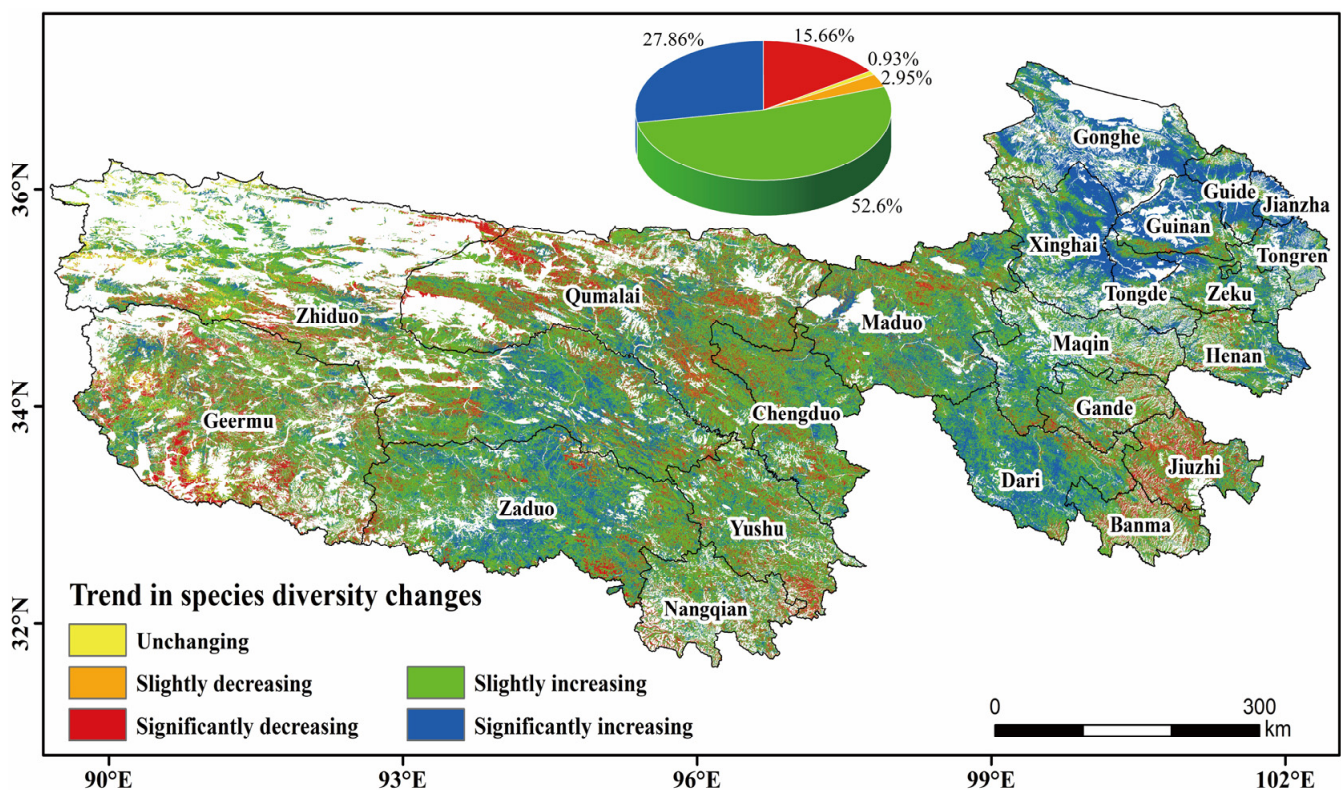


Figure 3. Trend in grassland species diversity from 2000 to 2021.

3.4. Future Trends in Grassland Species Diversity

Here, we evaluated the Hurst index on the time series of species diversity from 2000 to 2021 in the study area based on a rescaled extreme deviation (R/S) analysis and combined values with the species diversity slope for the 22 years to predict future trends (Figure 4). In terms of future trends in the spatial distribution pattern, 60.10% of the areas in the study area will still maintain the trend of continuously increasing species diversity in the future, and basically cover the whole area in terms of spatial distribution; 20.36% of the areas

showed an anti-continuously increasing in species diversity, i.e., the previous increase will show a decreasing trend in the future, and the distribution area also basically covers the whole area; 8.04% of the regions will maintain a trend of continuously decreasing species diversity in the future, mainly in the localized areas of Banma and Jiuzhi counties in Guoluo Prefecture and Yushu and Qumalai counties in Yushu Prefecture; 8.56% of the regions showed an anti-continuously decreasing trend, i.e., previous decreases will show an increasing trend in the future, with the relevant areas being primarily distributed adjacent to the areas of continuous decreases, which indicates that species diversity is in a state of gradual recovery in these regions. In addition 2.94% of the area future changes in species diversity were not detected in this study. Overall, the future changes in species diversity in the study area exhibit a positive trend, i.e., species diversity is expected to increase in most areas; however, there are also localized areas where species diversity will continue to decline, and these must be targeted in future conservation efforts.

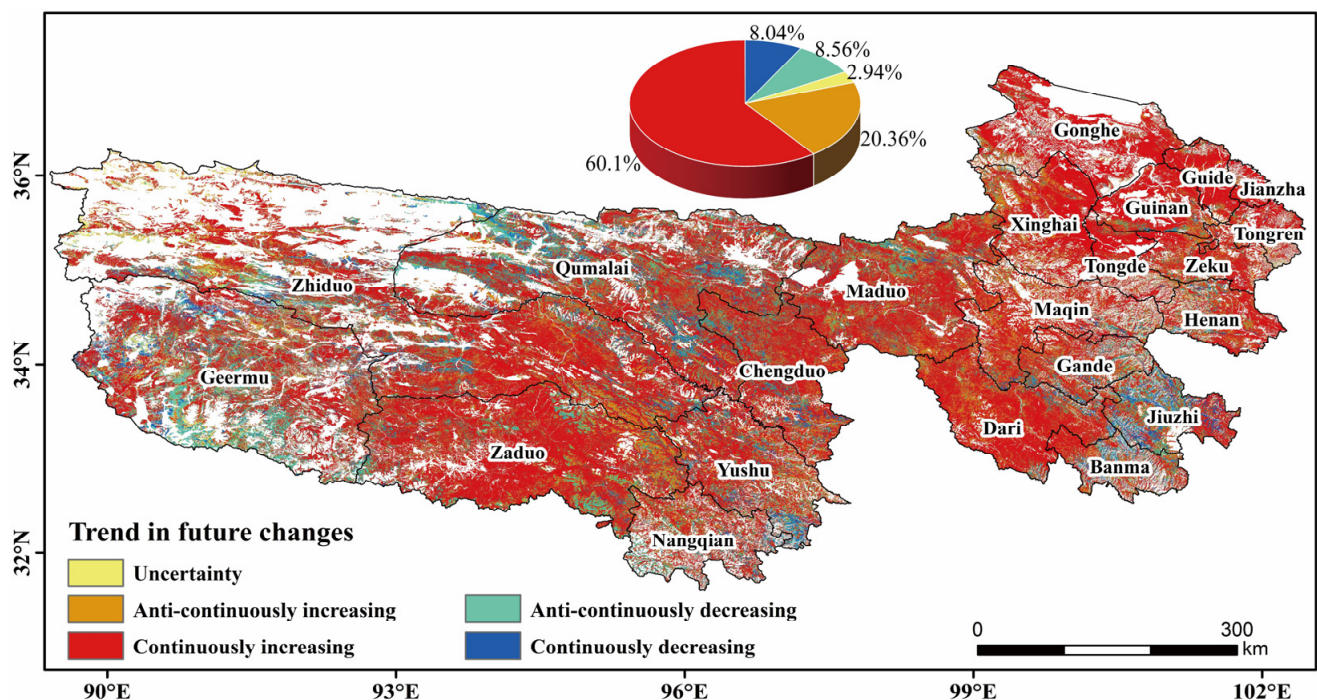


Figure 4. Trend in future changes in grassland species diversity.

3.5. The Drivers of Spatial and Temporal Dynamics of Grassland Species Diversity

Climate change and human activities as drivers of changes in species diversity in the study area over 22 years were analyzed using the OPGD model. Single-factor detection analysis yielded q -values for six latent variables, all of which passed the significance test (Figure 5a). The obtained q -values indicated that the change in species diversity in the study area was mainly driven by changes in regional temperatures, with a q -value of 0.299 for SMAT, followed changes in regional GDP per capita, with a q -value of 0.216 for SGDP, with the weakest driver being SMET, with a q -value of 0.016; the overall ranking of the six latent variables as drivers of the change in species diversity in the study area was SMAT > SGDP > SLSK > SPOP > SMAP > SMET. Among the three latent variables related to human activities, SGDP was the main driver of the change in species diversity, followed by SLSK, and, finally, SPOP. Among the three latent variables related to climate change, SMAT was the main driver, followed by SMAP, and then SMET.

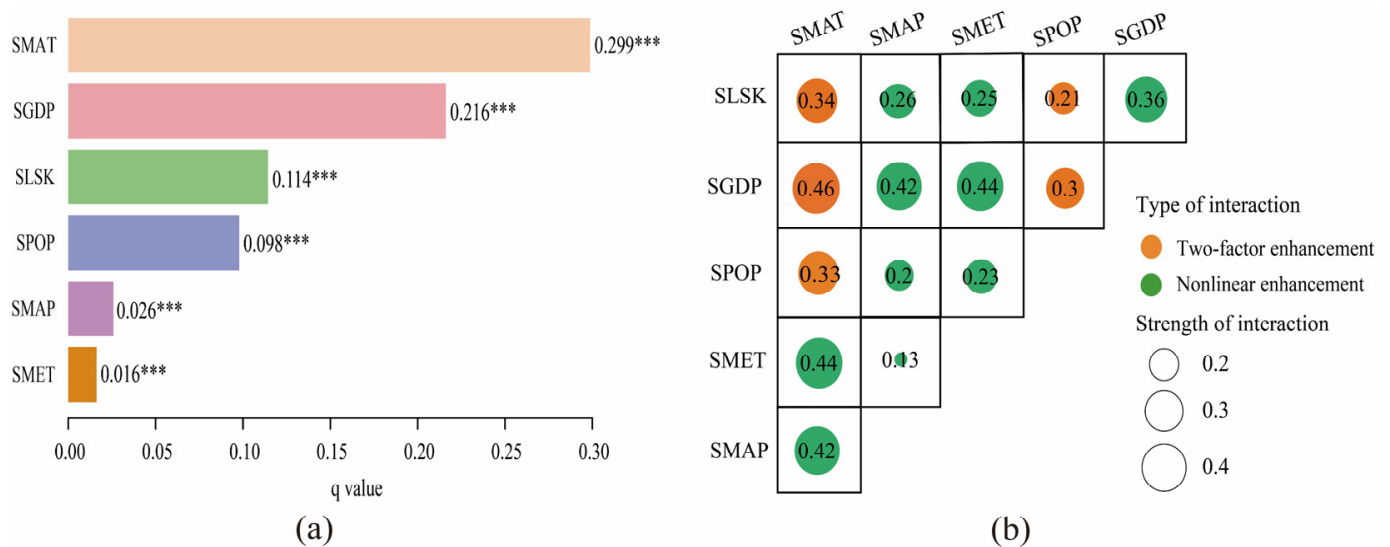


Figure 5. The OPGD model analysis. (a) Single-factor detection, *** indicates $p < 0.001$; (b) interaction detection.

Interaction probing yielded both two-factor enhancement and nonlinear enhancement results (Figure 5b). $SMAT \cap SLSK$, $SMAT \cap SGDP$, $SMAT \cap SPOP$, $SPOP \cap SLSK$, and $SPOP \cap SGDP$ exhibited two-factor enhancement, whereas the other interaction types displayed nonlinear enhancement. Regarding the strength of the interaction, the driving force of the interaction was significantly stronger than the driving force of the single factor for the change in species diversity in the study area; $SMAT \cap SGDP$ exhibited the strongest interaction, with a q-value of 0.46, followed by $SMAT \cap SMET$ and $SMET \cap SGDP$, with a q-value of 0.44. With a q-value of 0.13, $SMAP \cap SMET$ showed the weakest interaction, despite the interaction strengthening the driving force.

4. Discussion

4.1. Analysis of Variable Selection Results for Different Modeling Approaches

In this study, four variable selection methods—GA, REF, STEP, and LASSO—were used to screen the characteristics of 27 variables. The results showed that all four methods streamlined the 27 variables, thus eliminating the problem of multiple covariance among the variables, and effectively improving the amount of data and computation in later modeling. From the results of variable selection, the most frequently selected of the six vegetation indices was KNDVI, which was selected by all four methods, while the least frequently selected was NDVI, which was selected only by the REF method, followed by GNDVI. The NDVI is the most widely used vegetation index for observing surface vegetation and has been frequently used for monitoring vegetation species diversity in the past [13,15,36]. However, NDVI has a saturation effect in vegetation monitoring, and, over recent years, KNDVI has been shown to be more robust than NDVI for dealing with saturation and bias, and also displays greater robustness across spatial and temporal scales [37], and our results also suggest that KNDVI outperforms NDVI as well as other vegetation indices in characterizing vegetation species diversity. Among the six climatic variables, temperature, precipitation, and evapotranspiration were selected likely because the growth and distribution of vegetation are mainly limited by hydrothermal conditions [38,39]. The twelve soil attribute variables that stood out as selected were mainly THK, TN, SOC, and SBD, which were closely related [40,41]. Among the three geographic variables, ELE was selected by all four methods. ELE dominates hydrothermal conditions and is one of the most important drivers of the distribution of vegetation species diversity; meanwhile, ASPECT and SLOPE, although affecting the distribution of vegetation species, are likely to have less of an impact than ELE at large scales [42].

4.2. Analysis of Species Diversity Modeling and Accuracy Evaluation

In this study, four machine learning methods combined with different variable selection techniques were used to construct 20 grassland species diversity estimation models. The results of the evaluation of accuracy of each model indicated that, irrespective of which machine learning model was used, the GA method of variable selection led to the greatest reduction in the number of variables, but did not significantly improve model accuracy, whereas model accuracy was effectively improved when the other three variable selection methods were used. This finding also reflects that fewer variables do not necessarily lead to higher model accuracy and that studies need to combine different modeling methods and variables to optimize model accuracy. In line with the results of recent studies, the model constructed by the STEP method has the highest accuracy, followed by LASSO, and the model constructed by the GA method has poorer accuracy [13,21,32]. In addition, overall, the accuracy of the models using the RF machine learning method was slightly better than that observed when the XGboost, KNN, and SVM methods were used. Compared with the other methods, the RF method is more widely used in machine learning models for big data and has better model accuracy due to its strong robustness [32,43]. Although the XGboost-based model was slightly less accurate than the RF-based one in this study, its performance in machine learning is also significant [44]. Jia et al. [45] estimated grassland biomass using the same four machine learning methods as those used in this study and reported that the use of XGboost provided the best accuracy. The KNN algorithm is used to improve model accuracy by identifying the closest neighboring K parameters; accordingly, given the spatial heterogeneity of the grassland on a large scale in the study area, the model was stochastic, which, on the one hand, made it easy for the model to fall into the local minima while, on the other, rendered the model more sensitive to parameter variations. Thus, the KNN model showed only average predictive ability in this study. The SVM model generally performs better in dealing with classification problems with small samples and may be less effective in regions of high spatial heterogeneity at large scales, as was the case in this study [46]. Overall, the accuracy of machine learning models is affected by factors such as model algorithms, explanatory variables, and sample size [20].

In this study, we used Landsat remote sensing data to construct a model of grassland species diversity in the TRHR from 2000 to 2021. The optimal model precision R^2 obtained through precision evaluation was 0.44. Sabatini et al. [39] used a machine learning model to characterize the global vegetation species diversity from 1885 to 2015 and obtained an R^2 value of 0.49, which is similar to that of this study. Meanwhile, Fauvel et al. [13] and Madonsela et al. [14] estimated grassland species diversity using Landsat and Sentinel 2 remote sensing data, respectively, and obtained R^2 values of 0.27 and 0.34, respectively, indicating that the accuracy of their models was slightly lower than that recorded in this study. Wang et al. [47] and Zhao et al. [48] estimated grassland species diversity at high resolution using UAVs, handheld spectrometers, and other equipment, and achieved R^2 values of 0.47 and 0.73, respectively, which were significantly better than those of this study in terms of accuracy. A comparison of the accuracy of different studies illustrated that large-spatial-scale, long-time-series, and high-resolution models inevitably have lower accuracy than small-spatial-scale and high-resolution annual studies due owing to the large heterogeneity in time and space [6,49,50]. Given that different spatial resolutions represent different area sizes, and there is a species-area relationship in species diversity [51], the estimation of species diversity from remote sensing data combined with ground samples should focus more on scale matching between the resolution and the sample plots. In addition, remotely sensed spectral diversity has some limitations in the estimation of species diversity in natural grasslands. On the one hand, remotely sensed spectral diversity represents the spatial heterogeneity of grasslands, whereas leaf or canopy heterogeneity in the vertical structure of grassland vegetation is difficult to characterize [7,52]. On the other hand, remotely sensed spectral heterogeneity has a saturation effect in characterizing vegetation species diversity. Studies have shown that, when there are more than 14 species, diversity shows a convergence trend in remotely sensed spectral diversity, which leads to

underestimation in areas with high species richness [48]. The model of the present study is also not well-simulated in areas with more than 15 species.

4.3. Analysis of Spatial and Temporal Patterns of Species Diversity

Overall, species diversity changes over the past 22 years showed an increasing trend, which may be mainly due to two aspects. First, in the context of global climate change, the climate of the TRHR in recent years has undergone significant warming and humidification [27], which is conducive to the growth of vegetation and the succession of species in the TRHR [53]. Secondly, to tackle the problem of grassland degradation in the Three River Headwaters National Nature Reserve, the State has implemented several ecological protection and restoration projects since the beginning of the twenty-first century, including the management of the Black Soil Bank, the return of farmland to forests and grassland, and a reduction in the number of livestock, and these measures have yielded good results [29]. In this study, we found that there was little inter-annual fluctuation in species diversity between 2000 and 2008; however, species diversity increased significantly after 2008, which may be attributed to the implementation of Phase I and Phase II ecological protection and construction projects in the TRHR from 2005 to 2020, which have effectively contributed to the increase in species diversity. Therefore, the increasing trend in grassland species diversity in the study area based on the above two drivers is an expected result. In addition, we found that there was spatial heterogeneity in grassland species diversity in the study area over the past 22 years. The areas with significant increases in species diversity are mainly located in the Republican Basin, Dari, and Zaduo, which may be due to the increase in precipitation and temperature in the region in recent years, coupled with the implementation of a more stringent ecological protection policy [54]. The areas with a significant reduction in species diversity are mainly found in the local areas of Jiuzhi, Bama, Yushu, and Qumalai, which may be related to the characteristics of the local vegetation communities, climatic characteristics, and grassland utilization. Importantly, future conservation projects should pay special attention to the areas in which species diversity is declining.

In this study, we predicted future trends in species diversity by combining past trends in species diversity with Hurst index analysis from 2000 to 2021. Regarding the Hurst index, despite its prediction of future changes in species diversity, the time scale for predicting the future was not clear, while the prediction of the future was based on the past climate and human activities environmental context, but unfortunately, future climate change and human activities disturbances cannot remain unchanged from the past, so this is a limitation of this method in predicting the future. When predicting future trends, we found that nearly 70% of the study area showed an increasing trend in species diversity; nevertheless, 8% of the total study area will maintain the current trend of decreasing species diversity in the future while, in 20% of the area, species diversity shifted from an increasing trend in the past to a decreasing trend. This suggests that, overall, 28% of the area may lose species diversity in the future; accordingly, in the next step of biodiversity conservation, attention should focus on locally degraded areas such as Goluo and Yushu Prefectures, and on prioritizing the implementation of conservation policies for these areas.

4.4. Analysis of the Driving Factors of Changes in Species Diversity

Climate change and human activities are the most important drivers of species diversity change in grassland vegetation [1,26]. Our analysis indicated that temperature change has been the main driver of species diversity changes in the TRHR over the past 22 years. Similarly, Sabatini et al. [39] reported that changes in global vegetation species diversity in a long time series were mainly limited by temperature. Importantly, the Three Rivers Headwaters is located in an alpine region, where temperature most directly determines vegetation growth and succession [55]. Among human activity-related variables, changes in GDP per capita exerted the greatest effect on regional species diversity, possibly because GDP per capita represents the strength of regional human economic activity, and a stronger

economy will inevitably have an impact on a region's resources and environment [56]. In this study, we showed that the interaction between climate change and human activities constituted a greater driving force for change in regional grassland species diversity compared with either variable alone. Shang et al. [57] showed that climate change is the main driving force for vegetation changes in northwest China at multiple spatial and temporal levels. This may be due to the fact that, on the one hand, most of the TRHR is uninhabited, which renders its regional vegetation more sensitive to the impacts of human activities while, on the other hand, climate change is a slow process and human activities, guided by regional policies (e.g., grazing bans, ecological protection, migration and relocation, land cover changes), display large yearly differences, and stronger differential changes in human activities may exert greater effects on vegetation species diversity. Consequently, there is also a need to focus on regional policies for human activities in future biodiversity conservation planning.

5. Conclusions

In this study, grassland ground monitoring data were combined with satellite remote sensing data and environmental variables to construct multiple machine learning models to establish a reliable grassland species diversity estimation model, analyze the spatial and temporal patterns of grassland species diversity and future trends for grassland species diversity change in the TRHR from 2000 to 2021, and reveal the effects of climate change and human activities on the changes of species diversity in the study area. Based on the results of this study, the following three conclusions can be drawn:

- (1) The model constructed based on the STEP variable selection method combined with the RF machine learning method could accurately assess grassland species diversity in the TRHR. Through the comprehensive modeling of multiple variable selection and multiple machine learning models in addition to accuracy evaluation, we concluded that the model constructed based on the RF-STEP method had the highest accuracy of all the models tested, with an R^2 of 0.44, an RMSE of 2.56 n/m², and an MAE of 2.06 n/m² for the model test set.
- (2) The effectiveness of the conservation of the diversity of grassland vegetation species in the TRHR from 2000 to 2021 has been good, with species diversity showing an increasing trend in most areas; however, some areas still exhibited a decreasing trend in the past and are predicted to do so again in the future, an aspect that needs to be emphasized in the future conservation of biodiversity in the TRHR.
- (3) This study revealed that climate change and human activities are both key drivers of changes in grassland vegetation species diversity, with temperature change being the most significant driver of changes in grassland species diversity in the TRHR.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/rs16214005/s1>, 1. Introduction to the four variable selection methods; 2. Introduction to four machine learning methods; Table S1. Descriptive statistics of species diversity in ground survey sample plots from 2005–2021; Table S2. Calculation formula of vegetation index; Table S3. Grading of trends of species diversity changes; Figure S1. Optimal model training set and test set fitting; Figure S2. Spatial distribution of grassland species diversity in the Three River Headwaters Region during 2000–2021. Refs. [21,37,44,58–66] are cited in Supplementary Materials.

Author Contributions: Conceptualization, M.Y.; Formal analysis, M.Y. and A.C.; Funding acquisition, M.Y. and Y.W.; Investigation, M.Y., W.C., M.X. and Q.G.; Methodology, M.Y. and S.W.; Supervision, X.Y.; Visualization, M.Y. and X.Y.; Writing—original draft, M.Y.; Writing—review and editing, M.Y., A.C. and X.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Science and Technology Innovation Fund of Command Centre of Integrated Natural Resources Survey [Grant No. KC20220018] and Geological Survey of China (GSC) project [Grant No. DD20242555; DD20230094].

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Su, Y.; Guo, Q.; Guan, H.; Hu, T.; Jin, S.; Wang, Z.; Liu, L.; Jiang, L.; Guo, K.; Xie, Z.; et al. Human-climate coupled changes in vegetation community complexity of China since 1980s. *Earth's Future* **2022**, *10*, e2021EF002553. [\[CrossRef\]](#)
2. Cardinale, B.J.; Duffy, J.E.; Gonzalez, A.; Hooper, D.U.; Perrings, C.; Venail, P.; Narwani, A.; Mace, G.M.; Tilman, D.; Wardle, D.A.; et al. Biodiversity loss and its impact on humanity. *Nature* **2012**, *486*, 59–67. [\[CrossRef\]](#)
3. Pennekamp, F.; Pontarp, M.; Tabi, A.; Altermatt, F.; Alther, R.; Choffat, Y.; Fronhofer, E.A.; Ganesanandamoorthy, P.; Garnier, A.; Griffiths, J.I.; et al. Biodiversity increases and decreases ecosystem stability. *Nature* **2018**, *563*, 109–112. [\[CrossRef\]](#) [\[PubMed\]](#)
4. Pörtner, H.O.; Scholes, R.J.; Arneth, A.; Barnes, D.K.A.; Burrows, M.T.; Diamond, S.E.; Duarte, C.M.; Kiessling, W.; Leadley, P.; Managi, S.; et al. Overcoming the coupled climate and biodiversity crises and their societal impacts. *Science* **2023**, *380*, eabl4881. [\[CrossRef\]](#)
5. Sun, Y.; Yuan, Y.; Luo, Y.; Ji, W.; Bian, Q.; Zhu, Z.; Wang, J.; Qin, Y.; He, X.Z.; Li, M.; et al. An improved method for monitoring multiscale plant species diversity of alpine grassland using UAV: A case study in the Source Region of the Yellow River, China. *Front. Plant Sci.* **2022**, *13*, 905715. [\[CrossRef\]](#) [\[PubMed\]](#)
6. Wagg, C.; Roscher, C.; Weigelt, A.; Vogel, A.; Ebeling, A.; de Luca, E.; Roeder, A.; Kleinspehn, C.; Temperton, V.M.; Meyer, S.T.; et al. Biodiversity–stability relationships strengthen over time in a long-term grassland experiment. *Nat. Commun.* **2022**, *13*, 7752. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Thornley, R.H.; Gerard, F.F.; White, K.; Verhoef, A. Prediction of Grassland Biodiversity Using Measures of Spectral Variance: A Meta-Analytical Review. *Remote Sens.* **2023**, *15*, 668. [\[CrossRef\]](#)
8. Divíšek, J.; Chytrý, M. High-resolution and large-extent mapping of plant species richness using vegetation-plot databases. *Ecol. Indic.* **2018**, *89*, 840–851. [\[CrossRef\]](#)
9. Levrel, H.; Fontaine, B.; Henry, P.; Jiguet, F.; Julliard, R.; Kerbiriou, C.; Couvet, D. Balancing state and volunteer investment in biodiversity monitoring for the implementation of CBD indicators: A French example. *Ecol. Econ.* **2010**, *69*, 1580–1586. [\[CrossRef\]](#)
10. Reddy, C.S.; Kurian, A.; Srivastava, G.; Singhal, J.; Varghese, A.O.; Padalia, H.; Ayyappan, N.; Rajashekar, G.; Jha, C.S.; Rao, P.V.N. Remote sensing enabled essential biodiversity variables for biodiversity assessment and monitoring: Technological advancement and potentials. *Biodivers. Conserv.* **2021**, *30*, 1–14. [\[CrossRef\]](#)
11. Rocchini, D.; Balkenhol, N.; Carter, G.A.; Foody, G.M.; Gillespie, T.W.; He, K.S.; Kark, S.; Levin, N.; Lucas, K.; Luoto, M.; et al. Remotely sensed spectral heterogeneity as a proxy of species diversity: Recent advances and open challenges. *Ecol. Inform.* **2010**, *5*, 318–329. [\[CrossRef\]](#)
12. Stein, A.; Gerstner, K.; Kreft, H. Environmental heterogeneity as a universal driver of species richness across taxa, biomes and spatial scales. *Ecol. Lett.* **2014**, *17*, 866–880. [\[CrossRef\]](#) [\[PubMed\]](#)
13. Fauvel, M.; Lopes, M.; Dubo, T.; Rivers-Moore, J.; Frison, P.; Gross, N.; Ouin, A. Prediction of plant diversity in grasslands using Sentinel-1 and -2 satellite image time series. *Remote Sens. Environ.* **2020**, *237*, 111536. [\[CrossRef\]](#)
14. Madonsela, S.; Cho, M.A.; Ramoelo, A.; Mutanga, O. Remote sensing of species diversity using Landsat 8 spectral variables. *ISPRS J. Photogramm.* **2017**, *133*, 116–127. [\[CrossRef\]](#)
15. Wang, R.; Gamon, J.A. Remote sensing of terrestrial plant biodiversity. *Remote Sens. Environ.* **2019**, *231*, 111218. [\[CrossRef\]](#)
16. Perrone, M.; Di Febbraro, M.; Conti, L.; Divíšek, J.; Chytrý, M.; Keil, P.; Carranza, M.L.; Rocchini, D.; Torresani, M.; Moudrý, V.; et al. The relationship between spectral and plant diversity: Disentangling the influence of metrics and habitat types at the landscape scale. *Remote Sens. Environ.* **2023**, *293*, 113591. [\[CrossRef\]](#)
17. Daleo, P.; Alberti, J.; Chaneton, E.J.; Iribarne, O.; Tognetti, P.M.; Bakker, J.D.; Borer, E.T.; Bruschetti, M.; MacDougall, A.S.; Pascual, J.; et al. Environmental heterogeneity modulates the effect of plant diversity on the spatial variability of grassland biomass. *Nat. Commun.* **2023**, *14*, 1809. [\[CrossRef\]](#)
18. Coelho, M.T.P.; Barreto, E.; Rangel, T.F.; Diniz-Filho, J.A.F.; Wüest, R.O.; Bach, W.; Skeels, A.; McFadden, I.R.; Roberts, D.W.; Pellissier, L.; et al. The geography of climate and the global patterns of species diversity. *Nature* **2023**, *622*, 537–544. [\[CrossRef\]](#)
19. Morais, T.G.; Teixeira, R.F.M.; Figueiredo, M.; Domingos, T. The use of machine learning methods to estimate aboveground biomass of grasslands: A review. *Ecol. Indic.* **2021**, *130*, 108081. [\[CrossRef\]](#)
20. Yu, Q.; Ji, W.; Prihodko, L.; Ross, C.W.; Anchang, J.Y.; Hanan, N.P. Study becomes insight: Ecological learning from machine learning. *Methods Ecol. Evol.* **2021**, *12*, 2117–2128. [\[CrossRef\]](#)
21. Yang, M.; Chen, A.; Zhang, M.; Gu, Q.; Wang, Y.; Guo, J.; Yang, D.; Zhao, Y.; Huang, Q.; Ma, L.; et al. Relationship between plant species diversity and aboveground biomass in alpine grasslands on the Qinghai–Tibet Plateau: Spatial patterns and the factors driving them. *Front. Ecol. Evol.* **2023**, *11*, 1138884. [\[CrossRef\]](#)
22. Li, M.; Zhang, X.; Niu, B.; He, Y.; Wang, X.; Wu, J. Changes in plant species richness distribution in Tibetan alpine grasslands under different precipitation scenarios. *Glob. Ecol. Conserv.* **2020**, *21*, e848. [\[CrossRef\]](#)
23. Tallwin, J.R.B.; Rook, A.J.; Rutter, S.M. Impact of grazing management on biodiversity of grasslands. *Anim. Sci.* **2005**, *81*, 193–198. [\[CrossRef\]](#)

24. Yao, Z.; Xin, Y.; Yang, L.; Zhao, L.; Ali, A. Precipitation and temperature regulate species diversity, plant coverage and aboveground biomass through opposing mechanisms in large-scale grasslands. *Front. Plant Sci.* **2022**, *13*, 999636. [[CrossRef](#)] [[PubMed](#)]
25. Zhang, M.; Delgado-Baquerizo, M.; Li, G.; Isbell, F.; Wang, Y.; Hautier, Y.; Wang, Y.; Xiao, Y.; Cai, J.; Pan, X.; et al. Experimental impacts of grazing on grassland biodiversity and function are explained by aridity. *Nat. Commun.* **2023**, *14*, 5040. [[CrossRef](#)] [[PubMed](#)]
26. Harrison, S.P.; Gornish, E.S.; Copeland, S. Climate-driven diversity loss in a grassland community. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 8672–8677. [[CrossRef](#)]
27. Piao, S.; Zhang, X.; Wang, T.; Liang, E.; Wang, S.; Zhu, J.; Niu, B. Response of Qinghai-Tibet Plateau ecosystems to climate change and its feedbacks. *Chin. Sci. Bull.* **2019**, *64*, 2842–2855. [[CrossRef](#)]
28. Mi, X.; Feng, G.; Hu, Y.; Zhang, J.; Ma, K. The global significance of biodiversity science in China: An overview. *Natl. Sci. Rev.* **2021**, *36*, 384–398. [[CrossRef](#)] [[PubMed](#)]
29. Zhang, Y.; Zhao, X.; Gong, J.; Luo, F.; Pan, Y. Effectiveness and driving mechanism of ecological restoration efforts in China from 2009 to 2019. *Sci. Total Environ.* **2024**, *910*, 168676. [[CrossRef](#)]
30. Liu, F.; Wu, H.; Zhao, Y.; Li, D.; Yang, J.; Song, X.; Shi, Z.; Zhu, A.; Zhang, G. Mapping high resolution National Soil Information Grids of China. *Sci. Bull.* **2022**, *67*, 328–340. [[CrossRef](#)]
31. Meng, N.; Wang, L.; Qi, W.; Dai, X.; Li, Z.; Yang, Y.; Li, R.; Ma, J.; Zheng, H. A high-resolution gridded grazing dataset of grassland ecosystem on the Qinghai-Tibet Plateau in 1982–2015. *Sci. Data* **2023**, *10*, 68. [[CrossRef](#)]
32. Ge, J.; Hou, M.; Liang, T.; Feng, Q.; Meng, X.; Liu, J.; Bao, X.; Gao, H. Spatiotemporal dynamics of grassland aboveground biomass and its driving factors in North China over the past 20 years. *Sci. Total Environ.* **2022**, *826*, 154226. [[CrossRef](#)] [[PubMed](#)]
33. Jiang, W.; Yuan, L.; Wang, W.; Cao, R.; Zhang, Y.; Shen, W. Spatio-temporal analysis of vegetation variation in the Yellow River Basin. *Ecol. Indic.* **2015**, *51*, 117–126. [[CrossRef](#)]
34. Yue, S.; Pilon, P.; Cavadias, G. Power of the Mann±Kendall and Spearman’s rho tests for detecting monotonic trends in hydrological series. *J. Hydrol.* **2001**, *259*, 254–271. [[CrossRef](#)]
35. Song, Y.; Wang, J.; Ge, Y.; Xu, C. An optimal parameters-based geographical detector model enhances geographic characteristics of explanatory variables for spatial heterogeneity analysis: Cases with different types of spatial data. *GIScience Remote Sens.* **2020**, *57*, 593–610. [[CrossRef](#)]
36. Oindo, B.O.; Skidmore, A.K. Interannual variability of NDVI and species richness in Kenya. *Int. J. Remote Sens.* **2002**, *23*, 285–298. [[CrossRef](#)]
37. Camps-Valls, G.; Campos-Taberner, M.; Moreno-Martínez, Á.; Walther, S.; Duveiller, G.; Cescatti, A.; Mahecha, M.D.; Muñoz-Marí, J.; García-Haro, F.J.; Guanter, L.; et al. A unified vegetation index for quantifying the terrestrial biosphere. *Sci. Adv.* **2021**, *7*, eabc7447. [[CrossRef](#)]
38. Peng, Y.; Feng, J.; Sang, W.; Axmacher, J.C. Geographical divergence of species richness and local homogenization of plant assemblages due to climate change in grasslands. *Biodivers. Conserv.* **2022**, *31*, 797–810. [[CrossRef](#)]
39. Sabatini, F.M.; Jiménez-Alfaro, B.; Jandt, U.; Chytrý, M.; Field, R.; Kessler, M.; Lenoir, J.; Schrod, F.; Wiser, S.K.; Arfin Khan, M.A.S.; et al. Global patterns of vascular plant alpha diversity. *Nat. Commun.* **2022**, *13*, 4683. [[CrossRef](#)]
40. Gholizadeh, H.; Gamon, J.A.; Zygielbaum, A.I.; Wang, R.; Schweiger, A.K.; Cavender-Bares, J. Remote sensing of biodiversity: Soil correction and data dimension reduction methods improve assessment of α -diversity (species richness) in prairie ecosystems. *Remote Sens. Environ.* **2018**, *206*, 240–253. [[CrossRef](#)]
41. López-Angulo, J.; Pescador, D.S.; Sánchez, A.M.; Luzuriaga, A.L.; Cavieres, L.A.; Escudero, A. Impacts of climate, soil and biotic interactions on the interplay of the different facets of alpine plant diversity. *Sci. Total Environ.* **2020**, *698*, 133960. [[CrossRef](#)] [[PubMed](#)]
42. Tsiftsis, S.; Štípková, Z.; Kindlmann, P. Role of way of life, latitude, elevation and climate on the richness and distribution of orchid species. *Biodivers. Conserv.* **2018**, *28*, 75–96. [[CrossRef](#)]
43. Cheng, C.; He, N.; Li, M.; Xu, L.; Cai, W.; Li, X.; Zhao, W.; Li, C.; Sun, O.J. Plant species richness on the Tibetan Plateau: Patterns and determinants. *Ecography* **2023**, *2023*, e06265. [[CrossRef](#)]
44. Bentéjac, C.; Csörgő, A.; Martínez-Muñoz, G. A comparative analysis of gradient boosting algorithms. *Artif. Intell. Rev.* **2021**, *54*, 1937–1967. [[CrossRef](#)]
45. Jia, Z.; Zhang, Z.; Cheng, Y.; Buhebaoyin; Borjigin, S.; Quan, Z. Grassland biomass spatiotemporal patterns and response to climate change in eastern Inner Mongolia based on XGBoost model estimates. *Ecol. Indic.* **2024**, *158*, 111554. [[CrossRef](#)]
46. Mountrakis, G.; Im, J.; Ogole, C. Support vector machines in remote sensing: A review. *ISPRS J. Photogramm.* **2011**, *66*, 247–259. [[CrossRef](#)]
47. Wang, R.; Gamon, J.; Montgomery, R.; Townsend, P.; Zygielbaum, A.; Bitan, K.; Tilman, D.; Cavender-Bares, J. Seasonal Variation in the NDVI-Species Richness Relationship in a Prairie Grassland Experiment (Cedar Creek). *Remote Sens.* **2016**, *8*, 128. [[CrossRef](#)]
48. Zhao, Y.; Sun, Y.; Chen, W.; Zhao, Y.; Liu, X.; Bai, Y. The Potential of Mapping Grassland Plant Diversity with the Links among Spectral Diversity, Functional Trait Diversity, and Species Diversity. *Remote Sens.* **2021**, *13*, 3034. [[CrossRef](#)]
49. Meyer, H.; Pebesma, E. Machine learning-based global maps of ecological variables and the challenge of assessing them. *Nat. Commun.* **2022**, *13*, 2208. [[CrossRef](#)]

50. Muro, J.; Linstädter, A.; Magdon, P.; Wöllauer, S.; Männer, F.A.; Schwarz, L.; Ghazaryan, G.; Schultz, J.; Malenovský, Z.; Dubovyk, O. Predicting plant biomass and species richness in temperate grasslands across regions, time, and land management with remote sensing and deep learning. *Remote Sens. Environ.* **2022**, *282*, 113262. [[CrossRef](#)]
51. Dong, S.; Tang, L.; Wang, X.; Liu, Y.; Liu, S.; Liu, Q.; Wu, Y.; Li, Y.; Su, X.; Zhao, C. Determination of plant diversity in alpine grasslands on the Tibetan Plateau minimum sample area. *Biodivers. Sci.* **2013**, *21*, 651–657. [[CrossRef](#)]
52. Imran, H.A.; Gianelle, D.; Scotton, M.; Rocchini, D.; Dalponte, M.; Macolino, S.; Sakowska, K.; Pornaro, C.; Vescovo, L. Potential and Limitations of Grasslands α -Diversity Prediction Using Fine-Scale Hyperspectral Imagery. *Remote Sens.* **2021**, *13*, 2649. [[CrossRef](#)]
53. Huang, K.; Zhang, Y.; Zhu, J.; Liu, Y.; Zu, J.; Zhang, J. The Influences of Climate Change and Human Activities on Vegetation Dynamics in the Qinghai-Tibet Plateau. *Remote Sens.* **2016**, *8*, 876. [[CrossRef](#)]
54. Ma, R.; Xiao, F.; Dong, Z.; Nan, W.; Hu, G.; Cheng, H. Dynamics of land desertification and its drivers in Gonghe County, Qinghai Province, 2000–2020. *J. Arid. Land Resour. Environ.* **2022**, *36*, 139–148. [[CrossRef](#)]
55. Ma, L.; Zhang, Z.; Shi, G.; Su, H.; Qin, R.; Chang, T.; Wei, J.; Zhou, C.; Hu, X.; Shao, X.; et al. Warming changed the relationship between species diversity and primary productivity of alpine meadow on the Tibetan Plateau. *Ecol. Indic.* **2022**, *145*, 109691. [[CrossRef](#)]
56. Wei, X.; Wei, W.; Liu, C. Spatial and temporal changes of grassland vegetation and its relationship with human activities over the past 40 years in Qinghai Province. *J. Ecol.* **2021**, *40*, 2541–2552. [[CrossRef](#)]
57. Shang, J.; Zhang, Y.; Peng, Y.; Huang, Y.; Zhu, L.; Wu, Z.; Wang, J.; Cui, Y. Climate change drives NDVI variations at multiple spatiotemporal levels rather than human disturbance in Northwest China. *Environ. Sci. Pollut. R.* **2022**, *29*, 13782–13796. [[CrossRef](#)]
58. Andersen, C.M.; Bro, R. Variable selection in regression—A tutorial. *J. Chemom.* **2010**, *24*, 728–737. [[CrossRef](#)]
59. Duan, K.-B.; Rajapakse, J.; Wang, H.; Azuaje, F. Multiple SVM-RFE for Gene Selection in Cancer Classification with Expression Data. *IEEE Trans. NanoBiosci.* **2005**, *4*, 228–234. [[CrossRef](#)]
60. Smith, G. Step away from stepwise. *J. Big Data* **2018**, *5*, 32. [[CrossRef](#)]
61. Tibshirani, R. Regression shrinkage via the lasso. *J. R. Stat. Soc. Ser. B* **1996**, *58*, 267–288. [[CrossRef](#)]
62. Deng, Z.; Zhu, X.; Cheng, D.; Zong, M.; Zhang, S. Efficient kNN classification algorithm for big data. *Neurocomputing* **2016**, *195*, 143–148. [[CrossRef](#)]
63. Tucker, C.J. Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sens. Environ.* **1979**, *8*, 127–150. [[CrossRef](#)]
64. Gitelson, A.A.; Kaufman, Y.J.; Merzlyak, M.N. Use of a green channel in remote sensing of global vegetation from EOS-MODIS. *Remote. Sens. Environ.* **1996**, *58*, 289–298. [[CrossRef](#)]
65. Huete, A.; Didan, K.; Miura, T.; Rodriguez, E.P.; Gao, X.; Ferreira, L.G. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* **2002**, *83*, 195–213. [[CrossRef](#)]
66. Huete, A.R. A soil-adjusted vegetation index (SAVI). *Remote Sens. Environ.* **1988**, *25*, 295–309. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.