ISSN 2616-5872 Vol.5, Issue 10: 86-98, DOI: 10.25236/AJEE.2023.051012

Remote Sensing Inversion of Aboveground Biomass of Grassland in Lanzhou City Based on Machine Learning Algorithm

Wenjing Dong[1], Hua Zhang[1,*]

[1]School of Geography and Environmental Sciences, Northwest Normal University, Lanzhou, China
[*]Corresponding author: zhanghua2402@163.com

Abstract: *Grassland is the important terrestrial ecosystem, aboveground biomass is an important indicator of the productivity of grassland ecosystem, monitoring grassland is very important to the assessment of the current status of grassland growth and the conservation and development of grassland resources to Lanzhou city. In this study, the reflectance was extracted and the 9 vegetation indexes was calculated from Landsat images that combined with field sampling data of the grassland aboveground biomass in Lanzhou city in July-August 2021 to construct the two machine models that Random Forest model (RF) and eExtreme Gradient Boosting model (XGBoost) and chose the best model to invert the grass aboveground biomass in Lanzhou City from 2000 to 2023 and analyze its spatial and temporal dynamics.The results of the study show that: (1) The reflectance of other bands except b5-NIR band and nine vegetation indices were significantly correlated with grassland aboveground biomass by Pearson correlation analysis, so the remaining 14 factors except b5-NIR band were selected as model input variables. (2) Compared with the XGBoost model (R^2 of 0.78, RMSE of 37.03), the RF (R^2 of 0.89, RMSE of 23.28) model has a higher accuracy and it was more suitable for the inversion of aboveground biomass of grassland in Lanzhou city. (3) In time, the average value of aboveground biomass of grassland in Lanzhou City from 2000 to 2023 showed an increasing trend as a whole; In space, the aboveground biomass of grassland decreased firstly from southeast to northwest and then increased in Lanzhou City. The area of high value zone of grassland aboveground biomass increased, and the low value zone kept transforming to the high value zone. This study can provide theoretical reference and technical support for the estimation of aboveground biomass of grassland and the protection of grassland ecosystem in Lanzhou City.*

Keywords: *grassland aboveground biomass; Landsat remote sensing data; machine learning; vegetation index; spatial and temporal variation*

1. Introduction

Grassland ecosystem is the main type of ecosystem in arid and semi-arid areas, which can fix the surface soil, improve the soil's ability to resist wind and water erosion, and play an important role in maintaining ecological balance and regional stability[1-2]. Lanzhou City is located in the transition zone from the Xizang Plateau to the Loess Plateau, with an arid climate and a highly complex and sensitive natural environment. Grassland is the only type of vegetation in Lanzhou City that can grow by natural precipitation, and the advantage of the developed underground roots of grassland makes it able to grow rapidly under poor climate conditions, with ecological functions such as windbreak and sand fixation, climate regulation, and water conservation, etc., which plays an important role in Lanzhou City's ecological environment, carbon cycle, and the maintenance of ecological balance[3-4]. However, little research has been done on the grasslands in Lanzhou City, and there are problems such as unclear status and spatial and temporal changes of grasslands, which ignores the importance of grasslands for Lanzhou City. Therefore, the establishment of a high-precision aboveground biomass estimation model can provide technical guidance for the monitoring of aboveground biomass of grassland in Lanzhou City, and provide a basis for the ecological protection of grassland in Lanzhou City and the rational utilization of resources, which is of great significance to the sustainable development and scientific management of grassland in Lanzhou City[5-6].

The monitoring methods of aboveground biomass of grassland are divided into ground method and remote sensing inversion method. Traditional aboveground biomass survey of grassland usually obtains the measured data of regional grassland directly through cyclic sampling, mowing and other methods,

ISSN 2616-5872 Vol.5, Issue 10: 86-98, DOI: 10.25236/AJEE.2023.051012

which is highly accurate but requires a lot of manpower and material resources, and has limitations in both time and space scales[7]. With the development of remote sensing technology, a variety of remote sensing data, such as MODIS, Landsat and Sentinel-2, have become the main basis for the monitoring of aboveground biomass in grasslands[8-10]. Most of the studies have adopted the method of constructing inversion models by matching ground truth data with satellite images to estimate the aboveground biomass of grassland, which has higher accuracy and relatively less reliance on ground sampling data [11-12], and has been widely used in land use change, vegetation cover studies, and vegetation parameter estimation [13-15]. Zhang [16], Wang Ting [17], Zeng Na [18], and Cong [19] established inverse models of aboveground biomass of grassland using Landsat data to analyze the spatial and temporal changes of aboveground biomass of grassland and its response to climate change, respectively. The above studies have shown that Landsat data can enable effective estimation of aboveground biomass in grasslands. The model was constructed using traditional parametric regression methods with relatively large inversion errors. In recent years, machine learning algorithms have achieved good results in remote sensing monitoring of vegetation [20], which is based on the complete spectral set for modeling, can make full use of the spectral information, has the characteristics of nonlinearity, high prediction accuracy, and strong generalization ability, and has been widely applied in the field of remote sensing based on nonparametric machine learning algorithms for estimating the aboveground biomass of grassland [21]. Among them, Random Forest (RF) utilizes multiple datasets for integrated regression and is able to deal with the complex relationship between the predicted values caused by noise and a large amount of data [17]. The eXtreme Gradient Boosting (XGBoost) model incorporates a regular term in the cost function that controls the complexity of the model, prevents overfitting and reduces the amount of computation [22]. Li [23], Yinghui Zhao [24], and Yuxin Tan [25] compared and analyzed the performance of traditional regression methods and machine learning algorithms for estimating aboveground biomass of grassland, and found that RF and XGBoost models had better estimation results.

Based on the above research, this study selects Landsat data to extract reflectance and calculate 9 vegetation indices as predictor variables and the measured aboveground biomass of grassland in Lanzhou City in July-August 2021 to establish two machine learning models, RF and XGBoost, and uses the coefficient of determination (R^2), and root-mean-square error (RMSE), to evaluate the model accuracy, and compares the applicability of the two models to invert the aboveground biomass of grassland in Lanzhou City, and the optimal model was selected to invert the aboveground biomass of grassland in Lanzhou City from 2000 to 2023 and analyze its spatial and temporal changes, with a view to providing theoretical references for the estimation of aboveground biomass of grassland in Lanzhou City and the protection of grassland ecosystems.

2. Materials and methods

2.1 Study area

Lanzhou City is located in the central part of Gansu Province and is the capital of Gansu Province, with a geographic location of 35°34′20″-37°07′07″N, 102°35′58″-104°34′29″E. It is at the confluence of the Qinghai-Xizang Plateau, the Loess Plateau, and the Inner Mongolia Plateau, with a total area of 1.31 × 10^4 km^2, and an elevation of 1,500-3,000 m above sea level [26]. The topography of Lanzhou City is high in the west and south and low in the northeast, with the Yellow River flowing from southwest to northeast. It has a temperate continental climate with an average annual temperature of 9.8°C, an average annual precipitation of about 327 mm, mainly concentrated in June-September, and a potential annual evaporation of 1437.7 mm [27-28]. Lanzhou City is rich in water resources, the region has the Yellow River as well as the Yellow River's first-class tributaries Huangshui, Zhuanglang River and second-class tributaries Datong River and other important water source culverts [29]. The soil type is mainly gray calcareous soil [30], and the area of grassland in Lanzhou City is 7102.82 km^2, accounting for 54.2% of the total area of Lanzhou City. It is mainly distributed in Yongdeng County, Gaolan County, and Yuchong County, accounting for 90.38% of the grasslands in Lanzhou City [31]. The grassland vegetation consists mainly of arid and saline types of plants, mainly conifers *Stipa capillata Linn*, *Reaumuria songarica*, *Ajania achilloides*, *Stipa krylovii*, *Stipa brevifllora*, *Ajania przewalskii*, *Artemisia sacrorum Ledeb*, *Stipa przewalskyi*, *Peganum harmala*, *Salsola passerina*, *viridis*, *Caragana roborovskyi Kom*, *Achnatherum splendens*, *Sacrorum*, *Artemisia frigida*, *Nitraria tangutorum* etc [32].

ISSN 2616-5872 Vol.5, Issue 10: 86-98, DOI: 10.25236/AJEE.2023.051012

2.2 Data sources

2.2.1 Sampling data

The data of aboveground biomass of grassland in Lanzhou City were obtained from the grassland survey of Lanzhou Forestry Bureau Grassland Monitoring Station in July-August 2021, and a total of 111 sample plots were set up (Figure 1). In Lanzhou City, representative topographic features were selected to set up sample plots and evenly distributed, with a sample size of 1 km × 1 km. For sample plots with only herbaceous, semi-shrub and dwarf shrub plants, three sample squares were set up in each sample plot, and the distance between sample squares in the sample plots was not less than 25 m, with a sample area of 1 m × 1 m. In sample plots with shrubs and tall herbaceous plants, a sample square with a size of 10 m × 10 m was set up in the sample plot. In the sample plots with shrubs and tall herbs, 10 m × 10 m sample squares were set up in the sample plots, and the relevant data of herbs, half-shrubs and dwarf shrubs, and shrubs and tall herbs were measured in the sample squares. The herbaceous plants in the sample plots were mowed flush and bagged, and latitude, longitude, elevation, aboveground biomass, herbaceous height, and cover were recorded, and then dried naturally to a constant weight to obtain the aboveground biomass of grass dry weight, and the average of the aboveground biomass of grass dry weight of the three sampling plots within the sample plots was taken as the biomass of the samples.

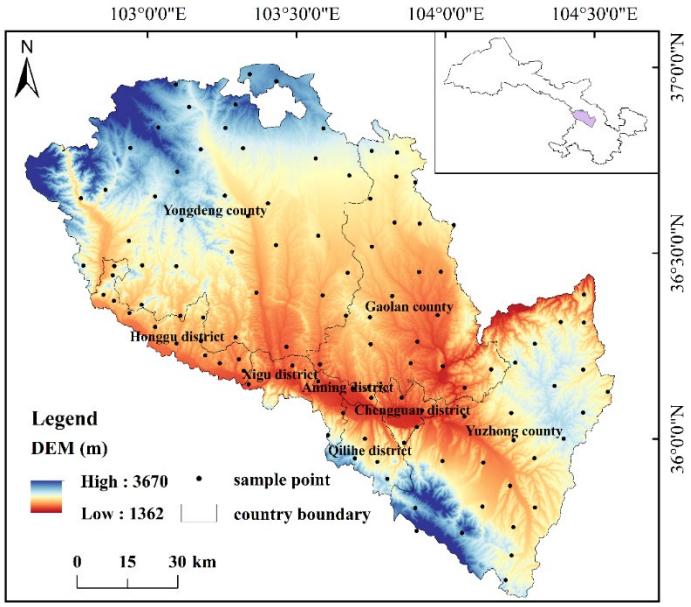

Figure 1: Spatial distribution of sampling points in the study area.

2.2.2 Remote sensing data

Remote sensing image data were selected from Landsat-5 TM, Landsat-8 OLI data, which were obtained by downloading from the United States Geological Survey (https://earthexplorer.usgs.gov).

Landsat images have a spatial resolution of 30 m and a minimum revisit period of 16 days. In order to minimize the difference between the imaging time of the image and the survey time of the ground-truthing data, the image with complete coverage of the sample plots and the imaging time closest to the sampling time was selected when downloading the image. In this study, Landsat-5 TM images and Landsat-8 OLI images with June-September coverage of the sampling sites and the whole territory and with less than 10% cloud coverage for the years 2000, 2005, 2010, 2015, 2020, and 2023 were selected. Due to image acquisition time constraints, the Landsat-5 TM sensor type was selected for images between 2000 and 2010, and the Landsat-8 OLI sensor type was selected for images between 2015 and 2023. This study download the Collection-2 Level-2 product, which has been radiometrically calibrated and atmospherically corrected, and only need to crop in ENVI 5.3 using the Lanzhou city boundary to get the pre-processed remote sensing image.

2.2.3 Other data

The Lanzhou City DEM data were obtained from the NASA Earth Science Data website (https://nasadaacs.eos.nasa.gov) at a resolution of 30m. Administrative boundaries of Lanzhou City are from the website of the State Bureau of Surveying and Mapping (http://www.sbsm.gov.cn).

ISSN 2616-5872 Vol.5, Issue 10: 86-98, DOI: 10.25236/AJEE.2023.051012

Meteorological data are from the National Xizang Plateau Scientific Data Center (http://data.tpdc.ac.cn).

2.3 Methods

2.3.1 Remote sensing variable selection

In this study, based on Landsat remote sensing data, b2-b7 reflectance data (b2- Blue band, b3-Green band, b4-Red band, b5-Near-Infrared band, b6-Short-Wave Infrared 1, and b7-Short-Wave Infrared 2) and nine vegetation indices were selected (Table 1): Normalized Difference Vegetation Index, Green Normalized Difference Vegetation Index, Difference Vegetation Index, Ratio Vegetation Index, Atmospherically Resistant Vegetation Index, Soil-Adjusted Vegetation Indices, Enhanced Vegetation Index, Modified Soil Adjusted Vegetation Index and Optimized Soil Adjusted Vegetation Index. The processed remote sensing images were loaded in ArcGIS 10.7 by bands, corresponding to the field sampling points by geographic coordinates, and the spectral reflectance of the sampling points was extracted by using the "Multi-Value Extraction to Points Tool", and then the vegetation index was calculated.

Table 1: Calculation formulas of vegetation indexes.

Vegetation index	Calculation formula	Select reason
NDVI	$\frac{NIR-Red}{NIR+Red}$	Higher soil background sensitivity under medium vegetation cover
GNDVI	$\frac{NIR-Green}{NIR+Green}$	Highly sensitive to chlorophyll
DVI	$NIR-Red$	Sensitive to the effects of soil background
RVI	$\frac{NIR}{Red}$	Highly correlated with vegetation biotic factors such as leaf area index and chlorophyll content
ARVI	$\frac{NIR-Blue}{NIR+Blue}$	Reducing the dependence of vegetation indices on the atmosphere
SAVI	$\frac{(1+L)(NIR-Red)}{NIR+RED+L}$ L=0.5	Integration of soil background factors
EVI	$\frac{2.5(NIR-\mathrm{Re}d)}{NIR+6\mathrm{Re}d-7.5Blue+1}$	More accurate representation of vegetation growth
MSAVI	$\frac{2NIR+1-\sqrt{(2NIR+1)^2-8(NIR-Red)}}{2}$	Mitigation of soil impacts on vegetation
OSAVI	$\frac{NIR-Red}{NIR+Red+0.16}$	Soil effects can be better excluded

2.3.2 Variable screening

In this study, b2-b7 reflectance and nine vegetation indices were subjected to Pearson correlation analysis with the ground-based measured aboveground biomass values of grassland in SPSS 19.0, so as to select the best modeling factors for the inversion of aboveground biomass of grassland in Lanzhou City.

2.3.3 Model construction

RF is a nonparametric machine learning algorithm that employs multiple decision trees to train samples and integrate predictions, where multiple samples are constructed from the original samples after randomly extracting the data using bootstrap resampling technique, and then N decision trees are constructed by using random splitting of nodes for each resampled sample [33]. During the growth of each tree, mtry of the total feature variables are randomly selected for internal node division; finally, the predictions of the N decision trees are assembled and voting is used to decide the category of the new sample [34]. It has higher accuracy and lower root-mean-square error than other traditional statistical models, can handle high-dimensional data without feature selection, has good noise immunity and stable performance, and can avoid overfitting to some extent [35]. RF model parameters: the number of decision classification trees (ntree), the number of features split by nodes (mtry) and the minimum number of

ISSN 2616-5872 Vol.5, Issue 10: 86-98, DOI: 10.25236/AJEE.2023.051012

decision tree nodes (nodesize), in this paper, the ntree used is 100, the mtry is 4 and the nodesize is 5.

XGBoost is an integrated learning algorithm belonging to the class of gradient boosting tree algorithms, the basic principle of which is to allow a new base model to fit the deviations of the previous model, thus continuously reducing the deviations of the model [36]. The model is optimized by adjusting parameters such as the learning rate of the XGBoost training set and the max depth. In this paper, the training set learning rate is 0.1 and max depth is 10.

2.3.4 Accuracy evaluation

The model was validated using the retained sample method, where 70% of the samples were randomly selected for modeling using the R language 4.0.2 software, and the remaining 30% of the samples were used for model accuracy testing. In this study, Coefficient of Determination (R^2) and Root Mean Squared Error (RMSE) were chosen to verify the model accuracy. The closer R^2 is to 1, the greater the correlation between predicted and true values, and the better the fit. The smaller the RMSE is, the smaller the bias is, and the better the model is [36].

$$R^2 = \frac{\sum_{i=1}^{n}\left(\hat{Y}_i - y_i\right)^2}{\sum_{i=1}^{n}\left(Y_i - y_i\right)^2} \tag{1}$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}\left(Y_i - \hat{Y}_i\right)^2}{n}} \tag{2}$$

Where Y_i is the measured value of the sample, $\hat{Y}_i$ is the value of the model inversion, y_i is the mean measured value of the sample biomass, and n is the number of samples.

3. Results

3.1 Modeling Factors Selection

In order to improve the accuracy of the model, the b2-b7 reflectance and nine vegetation indices were correlated with the measured data of aboveground biomass of grassland in Lanzhou City in July-August 2021 (Figure 2). The results showed that the aboveground biomass of grass was significantly negatively correlated with b2, b3, b4, b6, and b7, and significantly positively correlated with nine vegetation indices, with correlation coefficients above 0.75, and the correlation with b5 was not significant. Therefore, 14 factors other than b5 were selected as input variables of the machine model, and grassland aboveground biomass as output variables of the model.

The grassland aboveground biomass estimation model was established based on the above 14 modeling factors, and the 14 modeling factors were ranked in terms of feature importance based on RF and XGBoost models (Figure 3). The importance of vegetation index was higher, with GNDVI, RVI, and NDVI accounting for 17.2%, 15.7%, and 13.3%, respectively, in the RF model, and NDVI, GNDVI, and OSAVI accounting for 57.1%, 31.6%, and 4.4%, respectively, in the XGBoost model, which was in line with the results of the correlation in Figure 2.

ISSN 2616-5872 Vol.5, Issue 10: 86-98, DOI: 10.25236/AJEE.2023.051012

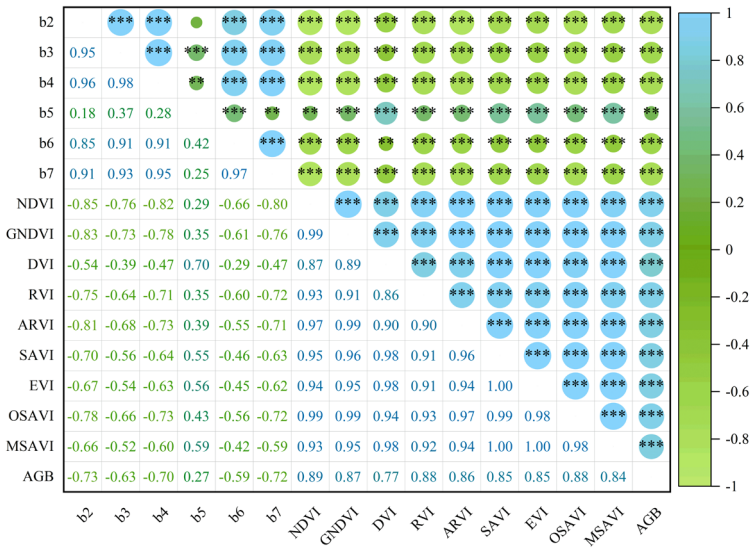

Figure 2: Correlation coefficients between the aboveground biomass and the explanatory variables.

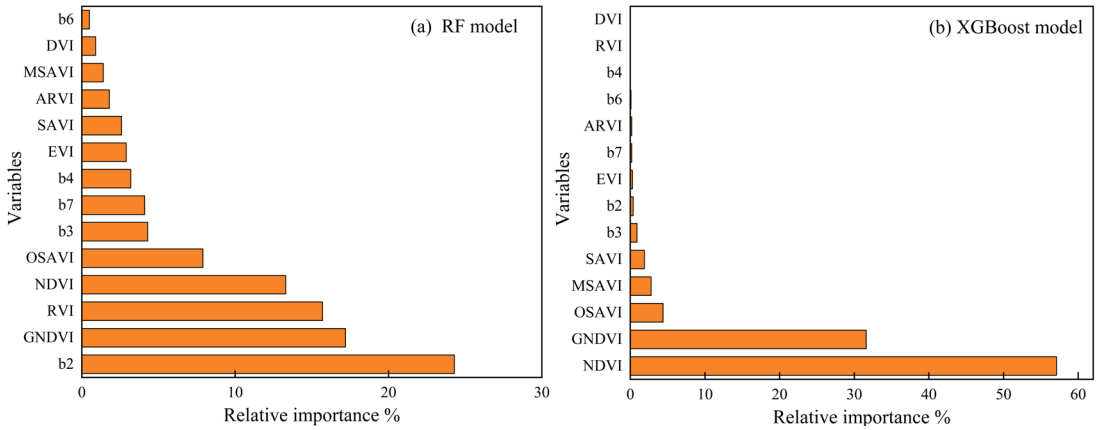

Figure 3: Importance ranking of predictor variables based on RF and XGBoost models.

3.2 Evaluation of model accuracy

The accuracy of the inversion model in this paper was measured by R^2 and RMSE (Table 2). The R^2 of the modeling samples and the validation samples of the RF were 0.89 and 0.72, and the RMSE was 23.28 and 54.38 respectively, while that of the modeling samples and the validation samples of XGBoost model were 0.78 and 0.67, and the RMSE was 37.03 and 59.09 respectively. The results showed that the RF model was better than the XGBoost model in constructing the inversion model of aboveground biomass of grassland in Lanzhou City. The results showed that the RF model had a better fitting effect than the XGBoost model and was more suitable for constructing the inverse model of aboveground biomass of grassland in Lanzhou City. In order to evaluate the accuracy of the RF model more intuitively, this study compared the measured values of aboveground biomass of grassland with the estimated values of the model, and plotted the scatter plots of the measured values and the estimated values (Figure 4). It was found that the model estimation effect was significantly higher than the discrete distribution points of the aboveground biomass of grassland in the more aggregated interval (30-300g/m^2), which further proved the applicability of the model.

Table 2: Inversion accuracy metrics of RF and XGBoost models.

Models	samples	R^2	RMSE
RF model	modeling samples	0.89	23.28
	Verification Sample	0.72	54.38
XGBoost model	modeling samples	0.78	37.03
	Verification Sample	0.67	59.09

ISSN 2616-5872 Vol.5, Issue 10: 86-98, DOI: 10.25236/AJEE.2023.051012

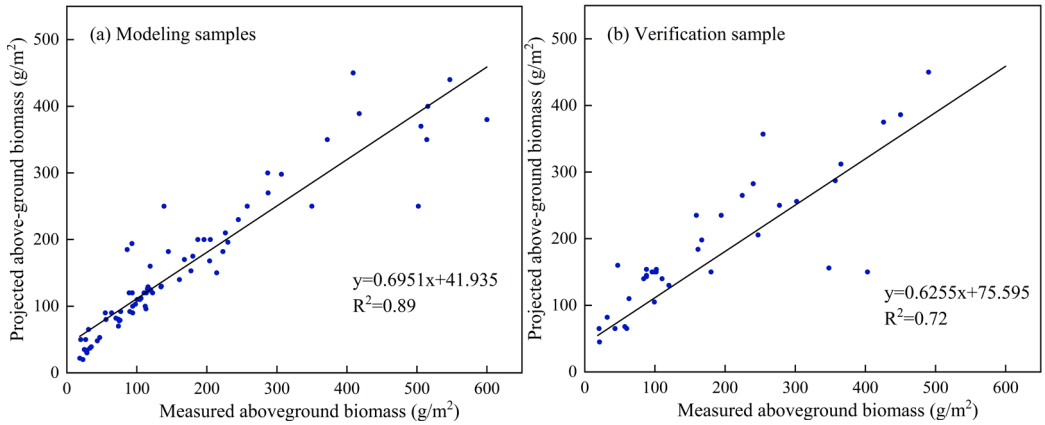

Figure 4: Accuracy verification of RF model simulation results.

3.3 Spatial and temporal changes of aboveground biomass of grassland in Lanzhou City

In this paper, the aboveground biomass of grassland in Lanzhou City was categorized into six grades: <100 g/m^2, 100-150 g/m^2, 150-200 g/m^2, 200-250 g/m^2, 250-300 g/m^2 and >300 g/m^2. The spatial and temporal changes of the aboveground biomass of grassland in Lanzhou City from 2000 to 2023 were analyzed by combining the statistics of the area and percentage of aboveground biomass of grassland of different grades (Table 3), the changes of the annual average value (Figure 5) and the spatial distribution (Figure 6).

In time, the annual mean value of grassland aboveground biomass showed an overall increasing trend from 2000 to 2023, with a decrease in the mean value of grassland aboveground biomass from 2010 to 2015 and from 2020 to 2023, which decreased by 22 and 24 g/m^2, respectively. 2005-2010 saw an increase in the mean value of grassland aboveground biomass by 18 g/m^2 . 2015-2020 Mean aboveground biomass of grassland increased more, by 33 g/m^2 (Figure 5).

In space, the high value of aboveground biomass of grassland in Lanzhou City was mainly distributed in the southeast and northwest of Lanzhou City, and the value of aboveground biomass of grassland in the central part of the city was lower, which overall showed the distribution characteristics of decreasing first and then increasing from southeast to northwest (Figure 6). The aboveground biomass of grassland in central Lanzhou City is low, such as Gaolan County, eastern Yongdeng County and northwestern Yuchong County, with an average lower than 150g/m^2; Aboveground biomass of grassland in southern Lanzhou City is moderate, about 150-250 g/m^2; Aboveground biomass of grassland was higher in northwestern and southeastern Lanzhou City, averaging higher than 250 g/m^2. In 2000, aboveground biomass of grassland was mainly in the range of 100-150 g/m^2; In 2005, aboveground biomass of grassland was mainly in the range of 0-100 g/m^2; In 2010, aboveground biomass of grassland was mainly in the range of 100-150 g/m^2 and 200-250 g/m^2; In 2015, aboveground biomass of grassland was mainly in the range of 100-150 g/m^2 and 150-200 g/m^2; In 2020, aboveground biomass of grassland was mainly in the range of 150-200 g/m^2 and 200-250 g/m^2; and in 2023, aboveground biomass of grassland was mainly in the range of 150-200 g/m^2 (Figure 6).

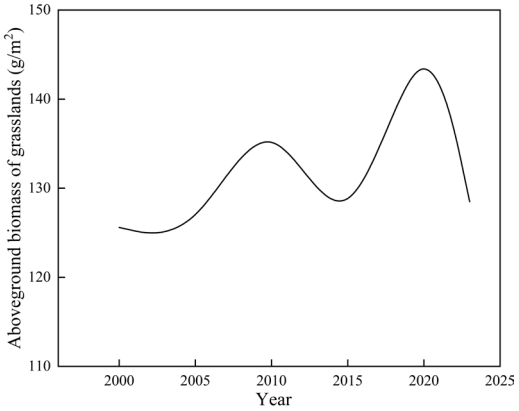

Figure 5: Inter-annual variation of the grassland aboveground bomass in Lanzhou city from 2000 to 2023

ISSN 2616-5872 Vol.5, Issue 10: 86-98, DOI: 10.25236/AJEE.2023.051012

From 2000 to 2023, the area of high value zone of aboveground biomass of grassland increased, the area of 150-200 g/m^2, 200-250 g/m^2 and 250-300 g/m^2 increased by 1410.85 km^2, 894.68 km^2 and 243.73 km^2, respectively, and the area of low value zone (100-150 g/m^2) decreased by 2906.39 km^2 (Table 2). Using the area of different classes of aboveground biomass to calculate the transfer matrix (Figure 7), the stability of different classes of aboveground biomass can be assessed. Comparing the aboveground biomass transfers of different classes in the five time periods, it can be seen that the aboveground biomass in the low aboveground biomass area (<100 g/m^2) and the high aboveground biomass area (250-300 g/m^2, >300 g/m^2) during the period of 2000-2023 is the most stable, and it is difficult to be transformed once it is formed. The most easily transformed aboveground biomass was in the 150-200 g/m^2 range. The 100-150 g/m^2 area changed the most from 2000-2005, with a major shift to 150-200 g/m^2, 200-250 g/m^2, 250-300 g/m^2 and >300 g/m^2. 2005-2010 mainly <100 g/m^2 to 250-300 g/m^2 and >300 g/m^2. 2010-2015 was dominated by the conversion of 100-150 g/m^2 to 200-250 g/m^2, 250-300 g/m^2 and >300 g/m^2, with little difference in the area of aboveground biomass converted between the classes. 2015-2020 mainly 100-150 g/m^2 to >300 g/m^2. 2020-2023 mainly all grades to 100-150 g/m^2.

Table 3: Statistics on the area (km^2) of different grassland aboveground biomass classes in Lanzhou city from 2000-2023

Year	Grassland aboveground biomass class					
	<100g/m^2	100-150 g/m^2	150-200 g/m^2	200-250 g/m^2	250-300 g/m^2	>300 g/m^2
2000	1868.6	7960.1	1191.4	1444.3	469.5	263.6
2005	5783.5	4970.9	685.9	1101.4	654.5	0.1
2010	1889.6	6653.1	2116.0	2150.0	384.6	2.8
2015	2055.2	5367.7	3433.1	1968.8	151.7	0.2
2020	599.0	4622.3	4024.7	2730.6	1047.9	170.0
2023	2435.8	5053.7	2602.3	2339.0	713.2	50.3

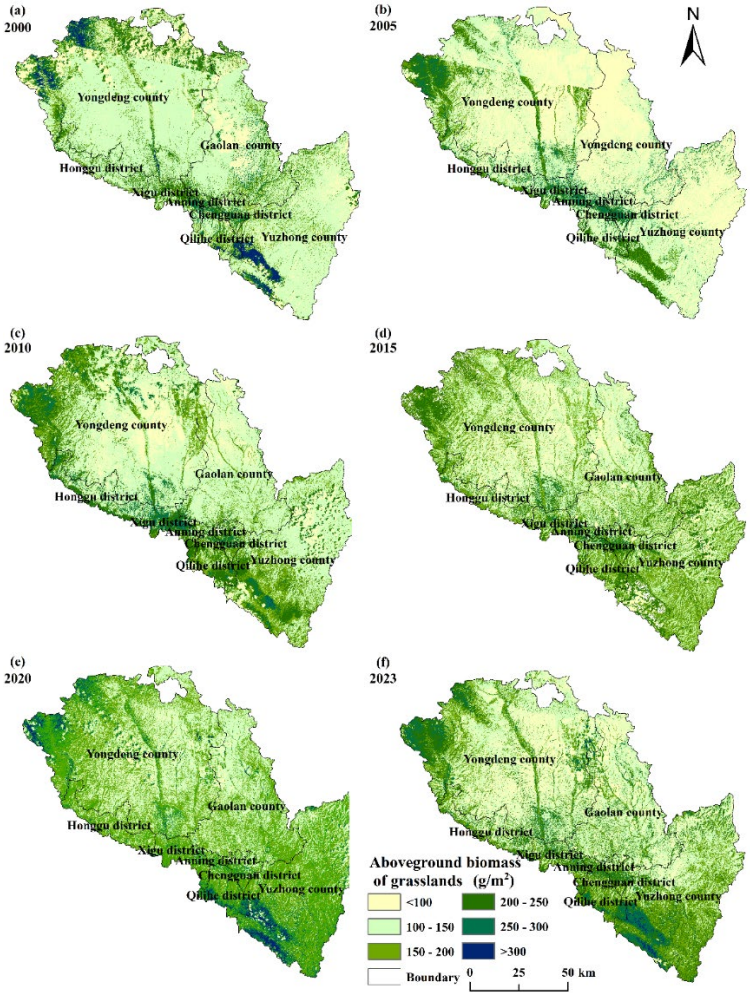

Figure 6: Spatial distribution of aboveground bomass of grassland in Lanzhou city from 2000-2023

ISSN 2616-5872 Vol.5, Issue 10: 86-98, DOI: 10.25236/AJEE.2023.051012

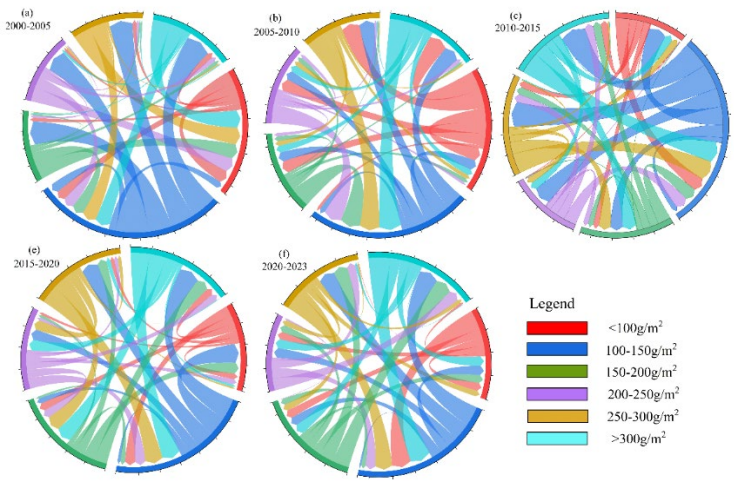

Figure 7: Transfer matrix of aboveground bomass of grassland by grade in Lanzhou city from 2000-2023

4. Discussion

At the present stage, most of the domestic and international studies on aboveground biomass of grassland are based on the vegetation index to construct the inverse model of aboveground biomass of grassland, and the vegetation index is more sensitive than the single band [36]. In this study, based on the commonly used vegetation indices such as NDVI and GNDVI, indices such as OSAVI, MSAVI, and SAVI, which have strong ability to eliminate the influence of soil, were introduced to construct the inverse model of aboveground biomass of grassland. Different vegetation indices are complementary to each other and can synthesize the status of aboveground biomass of grassland [37]. The modeling process revealed a high importance of model characteristics for RVI, NDVI, OSAVI, SAVI and EVI, which are five vegetation indices based on the near-infrared and red light bands. It has been shown that vegetation indices calculated using the red and near-infrared bands can minimize the effects of factors such as atmosphere and soil [38-40]. The effect of soil background cannot be ignored with the low aboveground biomass of grassland in most areas of Lanzhou City, which may be one of the key factors that this type of vegetation index has a significant effect on the aboveground biomass of grassland in Lanzhou City. The combination of reflectance and vegetation index to construct an inverse model of aboveground biomass of grassland can improve the model accuracy, which is consistent with the results of Song Kexin [5], Zeng Na [41], Ling [42], et al. The results showed that the RF model was more accurate and stable, which was more suitable for inverting the aboveground biomass of grassland in Lanzhou City, which was consistent with the results of Wang Xuemei [43], Zhang Peng [44], Ma [45], and Zhang Huilong [46], which showed that the accuracy of the RF model for inverting aboveground biomass was higher.

The aboveground biomass of grassland in Lanzhou City decreased and then increased from southeast to northwest, and the overall trend of aboveground biomass of grassland was increasing from 2000 to 2023. Grassland aboveground biomass decreases from 2000-2005; grassland above-ground biomass increases from 2005-2020; and grassland above-ground biomass decreases from 2020-2023. In 2002, Gansu Province comprehensively carried out the project of returning farmland to forest and grassland throughout the province [47], paying attention to the construction of the ecological environment and promoting the urbanization process while protecting the ecological environment, and the overall biomass of grassland aboveground under the support of the policy showed an increasing trend. In 2005, there was less precipitation (Figure 8), and the growth of vegetation in arid and semi-arid regions was greatly affected by moisture. The results of Wang Gongxin [48] et al. showed that precipitation and air temperature had a close relationship with the growth of grasses in arid and semi-arid regions, and that there was a significant correlation between the amount of precipitation and aboveground biomass of grasses, which could be the main reason for the low aboveground biomass of grasses in 2005. Heavy rainfall hit Lanzhou City in 2015 with hail, severely affecting vegetation growth, which may have contributed to the low aboveground biomass of grasslands in 2015. In 2022, extreme weather occurred frequently in Lanzhou City, with low precipitation and severe spring drought in spring, and the summer temperature was the highest in the same period since 1961 (Figure 8), with low precipitation days throughout the year, and the drought repeated many times, which had a great impact on the growth of vegetation, and this was the main reason for the reduction of aboveground biomass of grassland in the

ISSN 2616-5872 Vol.5, Issue 10: 86-98, DOI: 10.25236/AJEE.2023.051012

past two years. In general, better hydrothermal conditions are favorable for vegetation growth, and in arid and water-scarce areas, the increase in precipitation can be more efficiently utilized by vegetation, thus promoting vegetation growth, a conclusion that is more consistent with the findings of Yue xiao et al . The implementation of ecological conservation projects and the reduction of human activities are also important in promoting the growth of arid and semi-arid grasslands.

Based on Landsat data and field sampling data of aboveground biomass of grassland in July-August 2021, this paper constructs an inversion model of aboveground biomass of grassland using RF model and XGBoost model, compares the accuracy of the two models through R^2 and RMSE, and selects the optimal model to invert the aboveground biomass of grassland in Lanzhou City in the period of 2000-2023, and analyzes the characteristics of its spatial and temporal changes, so as to provide theoretical support for the remote monitoring of the aboveground biomass of grassland in Lanzhou City, as well as the rational use of grassland resources. The modeling factors selected in this paper are Landsat reflectance and vegetation index, although the accuracy is high, but there are still shortcomings, and in the future, it is necessary to comprehensively consider the remote sensing data with higher resolution and the factors of topography and geomorphology, land use, and soil physicochemical properties that affect the change of the aboveground biomass of grassland. In addition, the influence of human activities on the aboveground biomass of grassland should not be neglected .

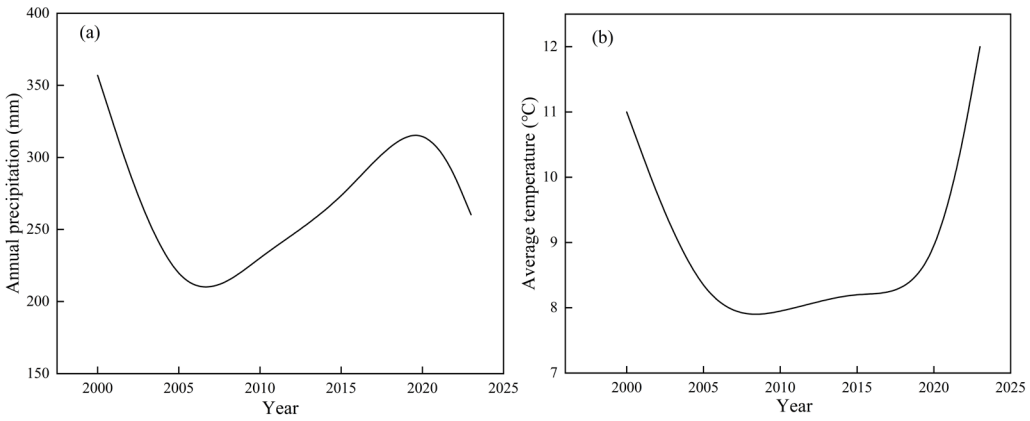

Figure 8: Changes in precipitation average temperature in the Lanzhou city from 2000-2022

5. Conclusions

In this study, based on Landsat multispectral data, we extracted reflectance and calculated 9 vegetation indices, combined with the field sampling data of aboveground biomass of grassland in July-August 2021, and constructed an inversion model of aboveground biomass of grassland by using the RF model and the XGBoost model, and then compared the accuracy of the two models through R^2 and RMSE, and selected the optimal model to invert the aboveground biomass of grassland in Lanzhou City from 2000-2023 and analyzed the characteristics of its spatial and temporal variations. The main conclusions are as follows: (1) Pearson correlation analysis revealed that the reflectance of all bands except b5 band and nine vegetation indices showed significant correlation with the aboveground biomass of grassland. (2) The R^2 of the RF model was 0.89 and the RMSE was 23.28, while the R^2 of the XGBoost model was 0.78 and the RMSE was 37.03. Compared with the XGBoost model, the RF model has a higher accuracy in inverting the aboveground biomass of grasslands in Lanzhou City, and it is suitable for inverting the aboveground biomass of grasslands in Lanzhou City in a long time series. (3) Temporally, the mean value of grassland aboveground biomass showed an overall increasing trend from 2000 to 2023. Spatially, the aboveground biomass of grassland in Lanzhou City showed a trend of decreasing and then increasing from southeast to northwest. During the period 2000-2023, the low above-ground biomass of grassland (<100 g/m^2) and the high aboveground biomass of grassland (250-300 g/m^2, >300 g/m^2) are more stable and difficult to be transformed once they are formed, and the easiest to be transformed is the above-ground biomass of the interval of 150-200 g/m^2.

ISSN 2616-5872 Vol.5, Issue 10: 86-98, DOI: 10.25236/AJEE.2023.051012

Acknowledgement

Funding: This work was supported by the National Natural Science Foundation of China (grant number 41461011) and Innovation and Entrepreneurship Talent Project of Lanzhou (grant number 2019-RC-105).

References

[1] Joel C, Creed. A Synthesis of Provision and Impact in Seagrass Ecosystem Services in the Brazilian Southwest Atlantic [J]. Sustainability, (2023) 15,20.
[2] Chunjun S, Yanlong L, Tongrui Z, et al. Light grazing intensity enhances ecosystem services in semi-arid grasslands through plant trait associations. [J]. Journal of environmental management, (2023) 348, 119375-119375.
[3] Zhu W, Yicheng W, Yang L, et al. Spatiotemporal characteristics and natural forces of grassland NDVI changes in Qilian Mountains from a sub-basin perspective [J]. Ecological Indicators, (2023) 157.
[4] Chuan hua L, Yu tao W, Xiaodong W, et al. Reducing human activity promotes environmental restoration in arid and semi-arid regions: A case study in Northwest China [J]. Science of the Total Environment, (2021) 768.
[5] Wu shuaikai, Hao jie, Diao huajie, et al. Response of grassland biomass to short-term grazing intensity in the agro-pastoral zone of northern Jinn [J]. Acta Agrestia Sinica, (2023) 31.
[6] Wang Yue, Qin Rongzhu. Grassland Biomass Inversion Based on a Random Forest Algorithm and Drought Risk Assessment [J]. Remote Sensing, (2022) 14, 5745-5745.
[7] Zhang yuxin, Huang jianxi, Jin yunxiang, et al. Advances in modeling for estimating aboveground biomass in grasslands [J]. Acta Agrestia Sinica, (2022) 30, 850-858.
[8] Vahidi M ,Shafian S ,Thomas S, et al.Estimation of Bale Grazing and Sacrificed Pasture Biomass through the Integration of Sentinel Satellite Images and Machine Learning Techniques [J]. Remote Sensing, (2023) 15.
[9] Stratoulias Dimitris. Assessing the Spectral Information of Sentinel-1 and Sentinel-2 Satellites for Above-Ground Biomass Retrieval of a Tropical Forest [J]. ISPRS International Journal of Geo-Information, (2022) 11, 199-199.
[10] Ma Jiamin, Zhang Wangfei. Total and component forest aboveground biomass inversion via LiDAR-derived features and machine learning algorithms [J]. Frontiers in Plant Science, (2023) 14, 1258521-1258521.
[11] Y C, L MY, Lu Z Z. Forest aboveground biomass estimation using Landsat 8 and Sentinel-1A data with machine 1 earning algorithms [J]. Scientific Reports, (2020) 10, 9952.
[12] Ogungbuyi G M, Mohammed C, Ara I, et al. Advancing Skyborne Technologies and High-Resolution Satellites for Pasture Monitoring and Improved Management: A Review [J]. Remote Sensing, (2023) 15.
[13] Qi pengwei, Zhang xian. Characteristics of spatial and temporal changes in vegetation cover and their drivers in Chongqing from 2000 to 2019 [J]. Acta Ecologica Sinica, (2022) 42, 5427-5436.
[14] Yue xiao, Zhang liangxia, Zhou decheng, et al. Analysis of spatial and temporal evolution of ecological vulnerability and driving factors in arid and semi-arid typical ecologically fragile areas [J]. Environmental Ecology, (2023) 5, 1-9+14.
[15] Pingping J, Junhua Z, Wei H, et al. Inversion of Different Cultivated Soil Types'Salinity Using Hyperspectral Data and Machine Learning [J]. Remote Sensing, (2022) 14, 5639-5639.
[16] McDermid. A Systematic Review of the Factors Influencing the Estimation of Vegetation Aboveground Biomass Using Unmanned Aerial Systems [J]. Remote Sensing, (2020), 12, 1052.
[17] Wang ting, Zhou wei, Xiao jieyun, et al. Estimation of aboveground biomass of grassland based on remote sensing data and machine learning algorithm [J]. Journal of Glaciology and Geocryology, (2023 45, 753-762.
[18] Zeng na, Ren xiaoli, He honglin, et al. Spatio-temporal dynamics of grassland aboveground biomass in Sanjiangyuan National Park and its climatic impacts [J]. Acta Ecologica Sinica, (2023) 43, 1175-1184.
[19] Cong X, Wenjun L, Dan Z, et al. Remote Sensing-based Spatiotemporal Distribution of Grassland Aboveground Biomass and Its Response to Climate Change in the Hindu Kush Himalayan Region [J]. Chinese Geographical Science, (2022) 32.
[20] Prakash T, P. K P, Jia Y, et al. Marginal agricultural land identification in the Lower Mississippi Alluvial Valley based on remote sensing and machine learning model [J]. International Journal of Applied Earth Observation and Geoinformation, (2023) 125.
[21] Santi E, Paloscia S, Pettinato S, et al. Machine-Learning Applications for the Retrieval of Forest

ISSN 2616-5872 Vol.5, Issue 10: 86-98, DOI: 10.25236/AJEE.2023.051012

Biomass from Airborne P-Band SAR Data [J]. Remote Sensing, (2020) 12, 804-804.
[22] Paula H N, José R P, Salgado B B, et al. A practical approach for soil unit weight estimation using artificial neural networks [J]. Journal of South American Earth Sciences, (2023) 131.
[23] Li Yingchang, Li Chao, Li Mingyang, et al. Influence of variable selectionand forest type on forest aboveground biomass esimation using machine learning algorithms [J]. Forests, (2019) 10, 1073.
[24] Hamada Y, Zumpf R C, Quinn J J, et al. Estimating Field-Level Perennial Bioenergy Grass Biomass Yields Using the Normalized Difference Red-Edge Index and Linear Regression Analysis for Central Virginia, USA [J].Energies, (2023) 16.
[25] Tan yuxin, Tian yichao, Huang zhuomei, et al. Inversion of aboveground biomass in mangrove forests of the Maowei Sea, Beibu Bay, China - Based on the XG Boost machine learning algorithm [J]. Acta Ecologica Sinica, (2023) 43, 4674-4688.
[26] Wancheng Z, Runping Z, Tian T, et al. Correction: Acute effects of air pollution on type II diabetes mellitus hospitalization in Lanzhou, China [J]. Environmental geochemistry and health, (2023) 45, 9947-9947.
[27] Duo L, Wang J, Zhang F, et al. Assessing the Spatiotemporal Evolution and Drivers of Ecological Environment Quality Using an Enhanced Remote Sensing Ecological Index in Lanzhou City, China [J]. Remote Sensing, (2023) 15.
[28] Xiao K, Xin C, Chen N, et al. Landscape Ecological Risk Assessment during 2000–2020 Based on Land Use and Land Cover Changes in Lanzhou City [J]. Academic Journal of Environment Earth Science, (2023) 5.
[29] Chai J, Zhang Z, Chen L, et al. Analysis of the Spatial and Temporal Evolution Characteristics and Driving Forces of the Surface Thermal Environment in Lanzhou City [J]. Sustainability, (2023) 15.
[30] Niu yalin, Li kongming, Wang xueyan, et al. Ecological stoichiometric characteristics and functional traits of dogbane flower leaves in response to slope orientation [J]. Chinese Journal of Ecology, (2020) 39, 1946-1955.
[31] Office of the Leading Group of the Third National Land Survey of Lanzhou City, Department of Nature Resources of Lanzhou, Lanzhou Municipal Bureau of Statistics [N]. Lanzhou Daily News, (2022) 2.
[32] Li zhengping, Wang wanpeng, Zhu gong, et al. Patterns and applications of vegetation restoration in arid and semi-arid areas of Lanzhou [J]. Pratacultural Science, (2004) 12, 124-127.
[33] A. H Z. High-resolution forest canopy cover estimation in ecodiverse landscape using machine learning and Google Earth Engine: Validity and reliability assessment [J]. Remote Sensing Applications: Society and Environment, (2023) 33.
[34] Sémanou K D, Abudu R K, Rikiatu H, et al. Analysis of the future potential impact of environmental and climate changes on wildfire spread in Ghana's ecological zones using a Random Forest (RF) machine learning approach [J]. Remote Sensing Applications: Society and Environment, (2023) 33.
[35] Xi bo X, Xiaoguang W, Pei jie Y, et al. Strategy for mapping soil salt contents during the bare soil period through a satellite image: Optimal calibration set combined with random forest [J]. Catena, (2023) 223.
[36] HansPeter P. An adjusted coefficient of determination (R^2) for generalized linear mixed models in one go [J]. Biometrical journal. Biometrische Zeitschrift, (2023) 65.
[37] Gathot W, Syamsu M R, Muhammad K, et al. Comparison of Mangrove Index (MI) and Normalized Difference Vegetation Index (NDVI) for the detection of degraded mangroves in Alas Purwo Banyuwangi and Segara Anakan Cilacap Indonesia [J]. Ecological Engineering, (2023) 197.
[38] Darmawan S, Sari D K. Developme ent of aboveground mangroveforests' biomass dataset for Southeast Asia based on ALOS-PALSAR 25m mosaic [J]. Remote Sensing, (2019) 13, 044519.
[39] Liang boming, Liu xin, Hao yuanyuan, et al. Extraction of vegetation biomass in desert areas based on five vegetation indices [J]. Arid Zone Research, (2023) 40, 647-654.
[40] Márcia, Gilson, André, et al. Aboveground biomass estimation in dry forest in northeastern Brazil using metrics extracted from sentinel-2 data: Comparing parametric and non-parametric estimation methods [J]. Advances in Space Research, (2023) 72.
[41] Zeng na, Ren xiaoli, He honglin, et al. Spatio-temporal dynamics of grassland aboveground biomass in Sanjiangyuan National Park and its climatic impacts [J]. Acta Ecologica Sinica, (2023) 43, 1175-1184.
[42] Ling xin B, Quan L, Song Q, et al. Grassland Biomass Inversion Based on a Random Forest Algorithm and Drought Risk Assessment [J]. Remote Sensing, (2022) 14.
[43] Wang xuemei, Yang xuefeng, Zhao feng, et al. Estimation of aboveground biomass in arid zone oasis based on machine learning algorithm [J]. Ecology and Environmental Sciences, (2023) 32, 1007-1015.
[44] Lv hao, Wang xing, Song naiping, et al. Soil moisture dynamics of four typical herbaceous plant communities in desert grasslands in response to extreme drought precipitation [J/OL]. Journal of Soil

ISSN 2616-5872 Vol.5, Issue 10: 86-98, DOI: 10.25236/AJEE.2023.051012

and Water Conservation, (2023) 1-8.
[45] Ma Y, Zhang L, Im J, et al. Novel Features of Canopy Height Distribution for Aboveground Biomass Estimation Using Machine Learning: A Case Study in Natural Secondary Forests [J]. Remote Sensing, (2023) 15.
[46] Zhang huilong, Yang xiuchun, Yang dong, et al. Spatial and temporal changes and trend prediction of grassland vegetation cover in Inner Mongolia from 2000 to 2020 [J]. Acta Prataculturae Sinica, (2023) 32, 1-13.
[47] Ma weikun, Zhu li. Gansu comprehensively launches the project of returning farmland to forest and grassland [N]. China Nationalities Daily (newspaper), (2002) 1.
[48] Wang gongxin, Jing changqing, Dong ping, et al. Estimation of biomass of desert grassland in Xinjiang and study of influencing factors [J]. Acta Agrestia Sinica, (2022) 30, 1862-1872.