Research article

Estimation of aboveground biomass of vegetation based on landsat 8 OLI images

Yanbin Zhang\textsuperscript{a,b,}\textsuperscript{*}, Ronghua Wang\textsuperscript{c}

\textsuperscript{a} Vocational and Technical College, Inner Mongolia Agricultural University, Baotou, 014109, China
\textsuperscript{b} College of Mechanical and Electrical Engineering, Inner Mongolia Agricultural University, Hohhot, 010070, China
\textsuperscript{c} Department of Mechatronics Engineering, Inner Mongolia Technical College of Mechanics and Electrics, Hohhot, 010070, China

\textbf{ARTICLE INFO}

\textbf{Keywords:}
Aboveground biomass
Measured data
Estimation model
Remote sensing inversion

\textbf{ABSTRACT}

Remote sensing estimation of aboveground biomass for desert oasis vegetation in arid area is an important means to monitor land desertification, it is of great significance to accurately evaluate the carbon sink change of desert oasis ecosystem, and maintain the stability of oasis ecosystem. The aboveground biomass information of vegetation, such as vegetation index and band factor in a delta oasis area is obtained by using Landsat 8 OLI image data; Based on the combination with the measured aboveground biomass data of vegetation, the optimal estimation model of aboveground biomass of four vegetation types (arbors, shrubs, herbs and crops) in this area is established, and the aboveground biomass of vegetation was retrieved and verified. The results showed that: (1) there was a very significant correlation between the remote sensing factors of aboveground biomass of four vegetation types and the measured aboveground biomass, and the correlation coefficient ranged from 0.711 to 0.756 (P < 0.01); (2) multiple stepwise regression (MSR) model is the optimal estimation model of aboveground biomass of arbors and shrubs, and partial least squares regression (PLSR) model is the optimal estimation model of aboveground biomass of herbs and crops, the estimation results have a good linear fitting relationship with the measured results; (3) the order of aboveground biomass of vegetation in oasis area from low to high is herbs < shrubs < arbors < crops. Among them, the aboveground biomass of grass is mainly below 280 g m\textsuperscript{-2}, the aboveground biomass of shrub is mainly 280–950 g m\textsuperscript{-2}, and the aboveground biomass of four vegetation is mainly distributed in 280–1450 g m\textsuperscript{-2}. Based on the Landsat 8 OLI image data, the remote sensing estimation model can accurately estimate the aboveground biomass of four oasis vegetation types (arbors, shrubs, herbs and crops), and reveal the spatial distribution characteristics of aboveground biomass of oasis vegetation.

1. Introduction

Vegetation aboveground biomass (AGB) can directly reflect the level of vegetation primary productivity, and the advantages and disadvantages of ecosystem structure. It is an important indicator for monitoring and evaluating land desertification and plays an irreplaceable role in maintaining and improving ecological environment \cite{1}. The aboveground biomass of vegetation is mainly used for the study of forest ecosystem productivity and community dynamic characteristics. At the same time, it is of great significance for the analysis and comparison of forest biodiversity, carbon storage and the restoration of degraded ecosystem \cite{2, 3, 4, 5}. Compared with traditional biomass measurement methods, remote sensing technology has the advantages of fast, accurate, less destructive to vegetation and macro monitoring, it has become the main method and research means of vegetation aboveground biomass estimation \cite{6, 7, 8, 9, 10}.

Vegetation index is a simple and effective measure of surface vegetation status, it has a close correlation with aboveground biomass, it is widely used in remote sensing estimation of vegetation growth status and aboveground biomass. For large-scale aboveground biomass inversion of arbors and shrubs, Ghosh et al. \cite{11} used multi-sensor collaborative data and machine learning algorithm to estimate the aboveground biomass of tropical forests in Uttar Pradesh, India. Ren Yi et al. \cite{12} constructed a biomass estimation model on arbor forest land by extracting vegetation index from Landsat 8 OLI data and combining texture features, which has a good estimation effect. With the further development of remote sensing inversion method, Qiao et al. \cite{13} used Landsat 8 OLI & GF-2 image data and introduced the weighted average height of poplar chest height section.
in the survey sample plot to construct the poplar biomass estimation model, which has high estimation accuracy. For the small area of arbor and shrub forest land, Yang et al. [14], Ding et al. [15] and Yang et al. [16] used high-resolution remote sensing images to extract spectral information, vegetation index and texture features, and combined with field investigation to estimate the forest biomass by remote sensing. In the research of remote sensing estimation of grassland biomass, MODIS vegetation index product data or Landsat images are often used to study grassland biomass in a wide range. Zhao et al. [17] constructed a remote sensing estimation model of grassland aboveground biomass in Qinghai Province based on MODIS-NDVI data and field investigation, which provides a scientific basis for grassland resource protection and utilization in Qinghai Province. Zhou et al. [18] constructed the grassland biomass estimation model in Sanjiangyuan area from 2001 to 2019 by using MODIS-NDVI data set and machine learning algorithm, it is considered that the high-precision surface model can more accurately retrieve the spatial distribution characteristics of grassland aboveground biomass. Zhang et al. [19] extracted six common vegetation indexes from Landsat 8 remote sensing data, established the remote sensing prediction model of grassland biomass on the sunny and shady slopes of ziniquan pasture on the north slope of Tianshan Mountain by using statistical analysis method, and carried out biomass spatial inversion and verification. Zhang et al. [20] accurately estimated the forage biomass in Haiyan County, Qinghai Province through different physical and chemical characteristics of vegetation using the spectral derived data of landsat8. At present, for small-scale grassland biomass estimation, UAV multispectral image, ground laser scanning technology and hyperspectral image data are often used for remote sensing monitoring [21, 22, 23]. In terms of remote sensing monitoring of crop biomass, UAV aerial photography technology combined with hyperspectral image data is often used to inverse the aboveground biomass and canopy structure parameters of crops [24, 25].

Looking at the current remote sensing estimation of vegetation aboveground biomass, it is found that most studies mainly focus on single type biomass inversion such as forest, grassland and crops, and there are relatively few studies on desert oasis vegetation with complex structure. Remote sensing estimation of aboveground biomass of desert oasis vegetation in arid area is an important means to monitor land desertification. It is of great scientific significance to accurately evaluate the carbon sink change of desert oasis ecosystem and maintain the stability of oasis ecosystem. However, due to the sparse distribution and complex and diverse types of vegetation in arid areas, it is often difficult to decompose mixed pixels [26]. At present, the investigation of aboveground biomass of desert oasis vegetation often constructs biomass estimation models according to vegetation types by extracting vegetation information and combined with multiple vegetation indexes, so as to improve the inversion accuracy of aboveground biomass of vegetation [27, 28]. Considering the significant difference in biomass of various vegetation types in the delta oasis of Weigan-Kuqa Rivers, the Landsat 8 OLI remote sensing image in July 2019 is combined with the vegetation aboveground biomass data of field survey in the same period, to extract the remote sensing factors characterizing the aboveground biomass characteristics of vegetation, and construct the conventional statistical models, Multiple stepwise regression and partial least squares regression models of aboveground biomass of arbores, shrubs, herbs and crops in the study area, the optimal estimation models of various vegetation types are obtained through verification analysis, and the spatial inversion analysis of vegetation aboveground biomass in the study area is carried out, which provides a theoretical reference for the scientific evaluation of the desert oasis ecosystem stability and carbon storage estimation.

2. Materials and methods

2.1. Overview of the study area

The study area is located in the delta oasis of Weigan-Kuqa Rivers (hereinafter referred to as Wei-Ku oasis) on the northern edge of the southern Tarim Basin. It starts from Qiulitag mountain in the north, connects the North Bank of the Tarim River in the south, borders Wensu County in the West, Luntai County and Yuli County in the East. It is a typical and complete fan-shaped plain oasis. It is administratively under the jurisdiction of Aksu region. The oasis has a temperate continental arid climate, and the average annual temperature is 11.6 °C, the annual average precipitation is 52 mm. There are many types of soil in Wei-Ku oasis, mainly tidal soil, irrigated silt soil and irrigated brown desert soil, and swamp soil, saline soil and brown calcium soil are also widely distributed in the study area [29]. The desert vegetation in this area mainly includes Tamarix ramossissima, Populus euphratica, Herculaneum strobilaceum, Halostachys caspica, Phragmites australis, Alhagi sparsiflolia and Nitraria tangutorum; the main crops are cotton (Gossypium spp.) and corn (Zea mays) [30]. Geolocation of the study area is shown in Figure 1, the research area is located in the northern part of Tarim Basin in the south of Xinjiang Uygur Autonomous Region, with a latitude of 41°25′–42°15′N and a longitude of 82°–82°30′E, appearance looks like a fan. Affected by the continental arid climate and the inter basin topographic pattern, it has developed into a typical and complete fan-shaped alluvial fan inclined plain. The climate is dry and hot in summer, dry and cold in winter, lacking precipitation all the year round, and the annual average evaporation reaches 1100 mm. However, the annual average precipitation is less than 200 mm, which belongs to a typical mid latitude arid area.

2.2. Material acquisition and processing

2.2.1. Field investigation and sampling

A 12 days ground survey was conducted in Wei-Ku oasis from mid July to late July 2019, with a total of 100 sample points. Among them, there are 69 crop sample points and 31 desert vegetation sample points. The vegetation quadrat survey is carried out centered on various points. 31 arbor quadrats (50 m × 50 m) with Tamarix as the dominant species, 112 shrub quadrats (10 m × 10 m) with yanjiemu and yansuimu as the main species, 114 herb quadrats (1 m × 1 m) with reed and camel thorn as the main species and 69 crop quadrats (10 m × 10 m) with cotton as the representative were investigated respectively, a total of 326 vegetation quadrats were investigated. In the vegetation survey, the plant height, north-south crown width, base diameter, frequency of each species, quadrat coverage and other growth parameters of each standard plant are measured and recorded in detail, and the aboveground parts of various vegetation standard plants are sampled on site. Finally, the aboveground biomass in the whole quadrat is calculated through the dry and fresh weight of the survey samples, and during field investigation,
each sample point is accurately located to obtain the longitude and latitude information of the sample point (Figure 2).

In order not to destroy the growth of vegetation, the long axis of crown width, short axis of crown width and plant height of standard plants of each typical species were measured for the arbor and shrub communities in the sample plot in the area with sparse desert vegetation. The standard plants were selected for sampling and biomass measurement of standard upright branches, and the biomass of the whole standard plant was estimated through the indirect estimation algorithm; The biomass of herbaceous community and main crop cotton was obtained by standard plant direct harvesting method. In the process of determining the biomass of the sample, in order to ensure the accuracy of the determination, the organs of the obtained plant sample are separated, weigh their fresh weight with a balance with an accuracy of 0.01 g, and then put them into kraft paper bags for sealing and storage; Take it back to the laboratory, dry it in an 80 °C constant temperature drying oven for 24 h, take it out and cool it for 20 min, and then weigh its dry weight. The moisture content is calculated by the fresh weight and dry weight of the standard branch. Finally, it is sorted in the excel table and the biomass of each organ of the standard plant and the biomass of the whole plant are calculated, so as to calculate the aboveground biomass and biomass per unit area of various vegetation.

### 2.2.2. Remote sensing image acquisition and processing

The remote sensing data of the study area adopts the Landsat 8 OLI satellite image at the same time as the field survey, the imaging time is July 26, 2019. The remote sensing image completely covers the whole Wei-Ku oasis and the image quality is good, the image 1–7 bands are used in this study. Using ENVI 5.3 software to preprocess the original remote sensing image, such as radiometric calibration, atmospheric correction, pixel resampling, image clipping and mask area. Based on multi-scale segmentation of remote sensing images in the study area, combined with image spectral information, shape information and vegetation index information, the initial feature space is constructed. On the basis of field investigation, the object-oriented nearest neighbor classification method is used to extract the information of arbor, shrub, herbs and crops (Figure 2), in which arbor refers to a tree with a large height, with an independent trunk at the root, a distinct distinction between the trunk and the crown, and an upright trunk, woody plants with a height of usually six meters to tens of meters are called arbor; Shrubs refer to those relatively short trees without obvious trunks and in a clustered state, they can be generally divided into several types, such as flowers, fruits and branches; Herbs refer to plants with undeveloped xylem, few lignified cells and weak supporting capacity in their stems, they are generally short in shape, short in life and weak in stems, and most of them die at the end of the growing season; Crops refer to various plants cultivated in agriculture, mainly including food crops and cash crops. Through the actual survey sample points, the classification results of remote sensing images in the study area are evaluated, and the overall accuracy is 96.64%, kappa coefficient is 0.95, which can meet the requirements of classification accuracy.

### 2.3. Extraction of modeling factors

The band math tool of ENVI 5.3 software is used to calculate the band of the preprocessed image, and calculates 7 vegetation indexes, namely normal differential vegetation index (NDVI), differential vegetation index (DVI), ratio vegetation index (RVI) Soil adjusted vegetation index (Savi), modified soil adjusted vegetation index (MSAVI), enhanced vegetation index (EVI) and atmospheric resistant vegetation index (ARVI). The calculation formula of each vegetation index is shown in Table 1. In order to make the research more rigorous and standardized, in the selection of band combination factors, the band combination factors of band gray value, reciprocal of band gray value, 2 band ratios, 3 band combination ratios and 4 band combination ratios are selected as the modeling factors for aboveground biomass estimation. The specific calculation formula is shown in Table 2. Finally, a total of 44 variables including 7 vegetation index variables and 37 band combination factors were selected as the modeling factors of biomass estimation model.

### 2.4. Biomass estimation model

Eight conventional statistical models, multiple stepwise regression (MSR) and partial least squares regression (PLSR) were selected to construct the estimation model of aboveground biomass of vegetation in the study area (Table 3). PLSR model integrates the advantages of multiple linear regression, principal component analysis and canonical correlation analysis, and can avoid potential problems such as non normal distribution of data, uncertainty of factor structure and unrecognizable model.

### 2.5. Model evaluation

The fitting effect of the model in this study is mainly to verify and evaluate the accuracy of the aboveground biomass estimation model by calculating the determination coefficient ($R^2$, Eq. (1)), root mean square error (RMSE, Eq. (2)) and mean absolute error (MAE, Eq. (3)) [19, 23]. The closer $R^2$ value is to 1, the better the fitting degree of the model is, the smaller RMSE value and MAE value are, the higher the estimation accuracy is. The calculation formula is as follows:

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

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

| Table 1. Calculation formulas of vegetation index. |
|-----------------------------------------------|
| Vegetation index | Calculation formula |
|------------------|---------------------|
| NDVI             | $\text{NDVI} = \frac{\text{BRED} - \text{BRED}}{\text{BNIR} + \text{BRED}}$ |
| DVI              | $\text{DVI} = \frac{\text{BNIR} + \text{BRED}}{\text{BNIR} - \text{BRED}}$ |
| RVI              | $\text{RVI} = \frac{\text{BNIR}}{\text{BRED}}$ |
| SAVI             | $\text{SAVI} = \frac{(\text{BNIR} - \text{BRED})/(1 + L)}{\text{BNIR} + \text{BRED}}$ |
| MSAVI            | $\text{MSAVI} = \frac{\text{BNIR} + 1 - \sqrt{(2\text{BNIR} + 1)^2 - 8(\text{BNIR} - \text{BRED})}}{2}$ |
| EVI              | $\text{EVI} = (\frac{(\text{BNIR} - \text{BRED})}{(\text{BNIR} + 6\text{BRED} - 7.5\text{BRED} + 1)} \times 2.5)$ |
| ARVI             | $\text{ARVI} = \frac{(\text{BNIR} + (2\text{BRED} - \text{BRED}))}{\text{BNIR} + (2\text{BRED} - \text{BLUE})}$ |

Note: BNIR refers to the near-infrared band, BRED refers to the red band, BLUE refers to the blue band, and L is soil regulation coefficient. In this study, L is 0.5.

![Figure 2. Distribution of sampling points in study area.](image-url)
2.6. Data analysis

The statistical processing of field measured data is carried out by using Excel 2019 software, combined with ENVI 5.3 software to extract the remote sensing factors required for modeling, and SPSS 22.0 and Matlab 2018b statistical analysis software are used to construct the aboveground biomass estimation model of 4 vegetation types; According to the accuracy verification results of remote sensing inversion of aboveground biomass of different vegetation types, the optimal estimation model is determined, and the spatial distribution map of aboveground biomass of vegetation in the study area was made by using ArcGIS 10.2 software.

3. Results and analysis

3.1. Correlation analysis between biomass and modeling factors

Because there is a certain collinearity among the variables, through the comparison of the correlation analysis results between biomass and various remote sensing factors, the variable factor with a very significant correlation with the aboveground biomass of the four vegetation types is finally selected as the automatic variable for constructing the remote sensing estimation model (P < 0.01). As shown in Table 4, there are 15 variables in tree modeling independent variable set, 17 variables in shrub modeling independent variable set, 9 variables in herb modeling independent variable set and 20 variables in crop modeling independent variable set. Among them, the remote sensing factor most closely related to the measured aboveground biomass of arbors is B253, and the correlation coefficient is 0.755 (P < 0.01); The most closely related variable factors to the measured aboveground biomass of shrubs and herbs are vegetation index NDVI and SAVI, and the correlation coefficients are 0.720 and 0.711 (P < 0.01) respectively; The most closely related variable factor to the measured aboveground biomass of crops is 1/B3, and the correlation coefficient is 0.756 (P < 0.01), followed by 1/B4, and the correlation coefficient is 0.742 (P < 0.01).

3.2. Establishment of vegetation aboveground biomass estimation model

3.2.1. Estimation model and verification of aboveground biomass of arbors

The vegetation index and band combination factors of pixel points and the measured biomass of quadrats were extracted from the remote sensing image by using the coordinates of the center points of various squares, and the estimation model of aboveground biomass of arbors was established. The measured quadrats were randomly divided into two groups, of which 21 were used for estimation model construction and 10 for model verification. It can be seen from Table 5 that among the conventional models, the quadratic term model has the best modeling effect; Through the test of introducing and eliminating variables, the final selected modeling independent variable of multiple stepwise regression model is univariate factor (B253); The partial least squares model with remote sensing factors B253, B453, NDVI and B254 as independent variables has high estimation ability. After testing, the regression effects of the above three estimation models have reached a very significant level (P < 0.001). The determination coefficient R² of the estimation model shows quadratic term model (0.893) > MSR model (0.812) > PLSR model (0.745). The root mean square error (RMSE) from low to high is PLSR model < MSR model < quadratic term model, and the mean absolute error (MAE) is the minimum of MSR model and the maximum of quadratic term model. Comprehensive multiple index analysis shows that MSR model has high estimation accuracy and stability. By analyzing the scatter diagram composed of measured biomass and estimated biomass in the modeling set and verification set of MSR model, there is a good fitting effect between the estimated value and the measured value, so MSR model can be used as the optimal estimation model of arbor aboveground biomass in the study area (Figure 3).

3.2.2. Estimation model and verification of shrub aboveground biomass

It can be seen from Table 6 that among the conventional models constructed using the measured quadrat data of aboveground biomass of 61 shrubs combined with remote sensing factors, the conventional S-curve model with normalized vegetation index (NDVI) as the independent variable has better estimation effect; The MSR model is constructed...
Table 4. Correlation analysis of remote sensing characteristic factors and aboveground biomass of vegetation.

| Serial number | Variables | Arbors | Shrubs | Herbs | Crops | Serial number | Variables | Arbors | Shrubs | Herbs | Crops |
|---------------|-----------|--------|--------|-------|-------|---------------|-----------|--------|--------|-------|-------|
| 1             | NDVI      | 0.716* | 0.720* | 0.711* | 0.602* | 11            | B_{354}   | 0.660* | 0.617* | 0.628* | 0.673* |
| 2             | RVI       | 0.701* | 0.716* | 0.707* | 0.675* | 12            | B_{452}   | 0.657* | 0.519* | –      | 0.669* |
| 3             | SAVI      | 0.716* | 0.720* | 0.711* | 0.602* | 13            | B_{453}   | 0.737* | 0.601* | –      | 0.686* |
| 4             | ARVI      | 0.662* | 0.543* | –      | 0.589* | 14            | B_{45/2}  | 0.634* | 0.497* | –      | 0.666* |
| 5             | DVI       | 0.730* | 0.670* | 0.696* | 0.488* | 15            | B_{523}   | 0.711* | 0.660* | 0.598* | 0.679* |
| 6             | EVI       | 0.634* | 0.567* | 0.555* | 0.542* | 16            | 1/B_{3}   | –      | 0.587* | –      | 0.756* |
| 7             | 1/B_{2}   | 0.590* | 0.596* | 0.536* | 0.718* | 17            | 1/B_{3}   | –      | 0.531* | –      | 0.742* |
| 8             | B_{233}   | 0.755* | 0.701* | 0.588* | 0.682* | 18            | 1/B_{3}   | –      | –      | –      | 0.572* |
| 9             | B_{234}   | 0.650* | 0.508* | –      | 0.672* | 19            | 1/B_{3}   | –      | –      | –      | 0.660* |
| 10            | B_{332}   | 0.677* | 0.571* | –      | 0.670* | 20            | B_{34}    | –      | –      | –      | 0.557* |

Note: * indicates extremely significant correlation at the level of 0.01(bilateral), “–” means no significant correlation.

Table 5. Evaluation of different estimation models for aboveground biomass of arbors.

| Model          | Expression                           | Modeling         | Verification     | P   |
|----------------|--------------------------------------|------------------|-----------------|-----|
|                |                                      | R^2   | RMSE | MAE | R^2   | RMSE | MAE |    |
| Quadratic term | y = –1315358685 + 55681143 + 6032.2 | 0.686 | 4.517 | 0.067 | 0.893 | 9.521 | 0.118 | 0.000 |
| MSR           | y = 12444 188B_{233} + 2581.234      | 0.607 | 5.050 | 0.066 | 0.812 | 8.176 | 0.095 | 0.000 |
| PLSR          | y = –78279 023B_{233} – 3702 876B_{433} + 2641 657NDVI – 72808 192B_{234} – 2275 642 | 0.713 | 4.589 | 0.069 | 0.745 | 7.753 | 0.107 | 0.000 |

Note: “<” indicates less significant correlation than the conventional S-curve model, “#” indicates higher significance than the conventional S-curve model.

Figure 3. Verification map of arbors optimal model in the study area.

with normalized vegetation index (NDVI) and enhanced vegetation index (EVI) as independent variables; The PLSR model constructed with 8 remote sensing factors has high accuracy. In order to further determine the regression relationship between vegetation index, band combination and measured aboveground biomass of vegetation, seven modeling factors such as NDVI, SAVI and RVI has the highest determination coefficient of its verification set are greater than that of conventional S-curve model. Although the RMSE value and MAE value of its verification set are greater than that of conventional S-curve model, they are still lower than that of PLSR model. Therefore, through comprehensive comparison, the multiple stepwise regression (MSR) model composed of normalized vegetation index (NDVI) and enhanced vegetation index (EVI) is selected as the optimal estimation model of shrub aboveground biomass in the study area, and the scatter diagram composed of measured biomass and estimated biomass is further analyzed, they have a high degree of fitting (Figure 4).

3.2.3. Estimation model of local biomass of grass

The modeling factors were extracted from remote sensing images, and combined with the aboveground biomass data of 28 measured herbaceous quadrats, the estimation model of grass aboveground biomass in the study area was constructed. According to the modeling accuracy and verification results of the estimation models (Table 7 and Figure 5), the regression effects of the three estimation models have reached a very significant level (P < 0.001). Among them, the PLSR model composed of seven modeling factors such as NDVI, SAVI and RVI has the highest estimation accuracy, and the determination coefficients of modeling and verification sets reach the highest. The modeling RMSE values are PLSR model < quadratic term model < MSR model, while MAE is PLSR model < MSR model < quadratic term model. Further, the accuracy of the grass fitting effect of the model, although the determination coefficient of MSR model is slightly lower than that of PLSR model, it is much higher than that of conventional S-curve model. Although the RMSE value and MAE value of its verification set are greater than that of conventional S-curve model, they are still lower than that of PLSR model. Therefore, through comprehensive comparison, the multiple stepwise regression (MSR) model composed of normalized vegetation index (NDVI) and enhanced vegetation index (EVI) is selected as the optimal estimation model of shrub aboveground biomass in the study area, and the scatter diagram composed of measured biomass and estimated biomass is further analyzed, they have a high degree of fitting (Figure 4).

Table 6. Effect evaluation of different estimation models for aboveground biomass of shrubs.

| Model          | Expression                           | Modeling         | Verification     | P   |
|----------------|--------------------------------------|------------------|-----------------|-----|
|                |                                      | R^2   | RMSE | MAE | R^2   | RMSE | MAE |    |
| S-curve       | y = 12.474 – e^{0.4} NDVI            | 0.560 | 0.832 | 0.689 | 0.394 | 0.826 | 0.642 | 0.000 |
| MSR           | y = 753.726NDVI – 907 902EVI + 25.979 | 0.542 | 10.525 | 6.577 | 0.637 | 5.810 | 4.125 | 0.000 |
| PLSR          | y = –121910 662NDVI + 123789 485AVI – 5723 928RVI + 762 622B_{233} – 400 296DV1 + 3537 964B_{234} + 9181 916B_{35} – 396 025EVI – 4253 029 | 0.648 | 9.233 | 5.783 | 0.721 | 6.198 | 4.910 | 0.000 |
The estimation model of aboveground biomass of crops is constructed through 44 measured quadrat data of crops and remote sensing modeling factors (Table 8). The conventional statistical model has good fitting effect with S-shaped curve, while the MSR model selects 1/B3 as the modeling independent variable. The PLSR model composed of 5 modeling independent variables has high estimation accuracy. By analyzing the fitting effects of the three models, the conventional S-curve model has the highest determination coefficient, followed by the PLSR model, and the determination coefficient of the two estimation models are greater than 0.6. The determination coefficient of MSR model is relatively low; Modeling RMSE is PLSR model, MSR model and S-curve model from small to large; In MAE, MSR model is the smallest, PLSR model is the second, and S-curve model is the largest. The above ground biomass estimation models of three crops were verified with 19 measured quadrat data. The results showed (Table 8): the regression results of the three estimation models reached a very significant level (P < 0.001). The R² of PLSR model is higher than that of conventional S-curve model and slightly lower than that of MSR model. It is verified that RMSE and MAE are also at a medium level. By comprehensively comparing the modeling and verification results, the PLSR model with high accuracy and stable prediction is selected as the optimal estimation model of aboveground biomass of crops in the study area. According to the scatter diagram analysis of the estimated biomass and measured biomass of the model (Figure 6), the estimated values are in good agreement with the measured values.

### 3.3. Spatial distribution of aboveground biomass of vegetation

Combined with the classification results of the study area (Figure 1), ENVI 5.3 software is used for remote sensing inversion of aboveground biomass of 4 vegetation types in the study area, and the estimation results of aboveground biomass (AGB) are divided into 4 levels according to the value range. At the same time, ArcGIS 10.2 software is used to make the spatial inversion map of aboveground biomass of vegetation in the study area and statistics the results (Figure 7 and Table 9). The analysis shows that the order of aboveground biomass of vegetation in the study area from high to low is crops > shrubs > herbs. Crops have the characteristics of high planting density and high vegetation coverage, the overall level of aboveground biomass is high, which is distributed in 280–1450 g m⁻², and mainly in grade III and IV, among them, grade IV has the widest distribution area, accounting for 46.04%, but there are some spatial differences in aboveground biomass of crops under different planting and management methods. Because the arbor is tall and has obvious trunk, its aboveground biomass is relatively high compared with shrubs and herbs, mainly grade II and III, accounting for the area 175.23 and 212.98 km² respectively, while the proportion of tree biomass in grade I is the lowest, only 16.74%; The total area of shrubs in the study area is 2087.74 km², the aboveground biomass is mainly concentrated in 280–950 g m⁻², that is, the proportion in grade III can reach 99.98%, but only very low distribution in grade I and II. Shrubs are mainly distributed in the Wei-Ku oasis desert transition zone and desert area, the vegetation is mainly salt knot wood and salt spike wood. There are many branches on the ground and developed roots, they have strong ability of wind prevention, sand fixation and drought resistance, which plays an important role in preventing soil loss and regulating the surface ecological environment in the study area; The local biomass of grass is mainly below 280 g m⁻², distributed in grade I and II, and accounts for a high proportion in grade I. The main vegetation types of herbs are reed and camel thorn, with low shape and no wooden trunk, due to the interference of grazing, the grassland degradation is more serious, and the

### Table 7. Accuracy evaluation of estimation model for herbs aboveground biomass in the study area.

| Model | Expression | Modeling | Verification | P |
|-------|------------|----------|--------------|---|
|       |            | R² | RMSE | MAE | R² | RMSE | MAE |    |
| Quadratic term | Y = –33287RVI² + 5455.5RVI – 198.02 | 0.560 | 0.832 | 0.689 | 0.394 | 0.826 | 0.642 | 0.000 |
| MSR | Y = 32.65NDVI – 167.376 | 0.542 | 10.525 | 6.577 | 0.637 | 5.810 | 4.125 | 0.000 |
| PLSR | Y = –2281832.994NDVI + 2285265.973AVI – 11772.656RVI + 10756.369B3A – 302.516AVI + 4400.234B234 – 3794.417EVI – 5623.736 – 3966.025EVI – 4253.029 | 0.648 | 9.233 | 5.783 | 0.721 | 6.198 | 4.910 | 0.000 |

Figure 4. Verification of shrubs optimal model in the study area.

Figure 5. Verification of herbs optimal model in the study area.
aboveground biomass is low, which is mainly distributed in the transition zone of oasis edge and desert area.

4. Discussion

Based on the correlation analysis between the vegetation index and band factor, extracted from the oasis of Weigan Kuqa River Delta in Xinjiang and the measured vegetation aboveground biomass, 20 remote sensing factors were selected to build the aboveground biomass estimation models of trees, shrubs, herbs and crops. Among them, the multiple stepwise regression model constructed with band combination B253 and vegetation index NDVI and EVI as modeling factors has the best inversion effect ($P < 0.001$). The research further confirms the conclusion of Ye et al. [27] in the estimation of aboveground biomass of vegetation in the desert oasis transition zone on the northeast edge of Ulanbuhe desert, it shows that the aboveground biomass of desert vegetation can be accurately estimated by extracting remote sensing spectral information and linear statistical model. The research shows that the partial least squares model integrating multi band spectral information can accurately retrieve the aboveground biomass of herbs and crops in the study area, the relevant studies of Kang et al. [23] and Jia et al. [31] also believe that the partial least squares model making full use of vegetation spectral information can effectively estimate the aboveground biomass of vegetation.

With the rapid development of remote sensing information technology, the methods and models of vegetation biomass estimation are also diversifying. Ali et al. [32] and Nesha et al. [33] used nonparametric models such as machine learning and random forest to estimate vegetation biomass. Based on the conventional parametric model, the introduction of nonparametric model will become the development trend of vegetation aboveground biomass estimation. In the modeling process, the modeling factors can be further optimized, the spectral characteristics of remote sensing images can be comprehensively used, and the spatial distribution characteristics of vegetation aboveground biomass can be further explained in combination with texture information, terrain and other environmental factors [34, 35]. Therefore, in the future biomass estimation process, it is necessary to consider the in-depth mining of image spectral information and introduce more environmental factors affecting vegetation aboveground biomass, so as to improve the application scope and estimation accuracy of the inversion model. Because the growth of vegetation presents different regional characteristics with the change of regional natural conditions, whether

| Model   | Expression                                                                 | Modeling | Verification |
|---------|----------------------------------------------------------------------------|----------|--------------|
| S-curve | $Y = e^{10.115 - 1.359/1.57}$                                              | 0.623    | 0.513        |
|         |                                                                            | 22.027   | 18.648       |
|         |                                                                            | 16.058   | 14.360       |
| MSR     | $Y = 1761.667 - 178.59/1.57 + 6615.782^{1/1} - 911.79$                     | 0.571    | 0.677        |
|         |                                                                            | 19.860   | 17.638       |
|         |                                                                            | 13.008   | 13.648       |
| PLSR    | $Y = 7622.851 - 1530.927^{1/1} - 3549.965^{1/1} - 349.606 + 1530.927^{1/1}$ | 0.613    | 0.626        |
|         |                                                                            | 18.863   | 17.659       |
|         |                                                                            | 13.158   | 13.671       |

Table 8. Accuracy evaluation of crops aboveground biomass estimation model in the study area.
the vegetation aboveground biomass estimation model constructed in this study is suitable for other areas needs further investigation and verification.

5. Conclusion

Taking the oasis vegetation in the delta oasis of Weigan-Kuqa Rivers as the research object, using remote sensing images and field measured vegetation aboveground biomass data, aboveground biomass inversion models are constructed for the four main vegetation types of arbors, shrubs, herbs and crops respectively. Through accuracy verification and comparative analysis, it is considered that, the optimal estimation model of aboveground biomass of arbors and shrubs in the study area is multiple stepwise regression model, and the estimation model of aboveground biomass of herbs and crops is partial least squares regression model. It is further verified that the 20 selected remote sensing factors have a very significant positive correlation with the measured aboveground biomass of vegetation, and the determination coefficients of the four models are all 0.6 above, the root mean square error and average absolute error are small, which meet the accuracy requirements of model estimation (P < 0.001). Through the estimation of inversion model, the order of vegetation aboveground biomass in the study area from high to low is crops > arbors > shrubs > herbs. The aboveground biomass is mainly concentrated in 280–1450 g m⁻², mainly in grade III and IV, accounting for 76.54%, with a distribution area of about 6973.82 km². The area of low-level aboveground biomass (ABG < 65 g m⁻²) accounts for about 15.02% of the total area of the study area. By analyzing the spatial distribution characteristics of aboveground biomass of vegetation in Wei-Ku oasis, it is considered that the spatial distribution pattern of aboveground biomass of vegetation is significantly different due to different land use modes, topography and hydrological conditions. According to different vegetation types, the remote sensing estimation model based on the spectral characteristics of ground objects can accurately estimate the aboveground biomass of oasis vegetation in arid areas, and carry out remote sensing quantitative inversion of its spatial distribution characteristics.

Declarations

Author contribution statement

ZHANG Yanbin: Conceived and designed the experiments; Analyzed and interpreted the data; Wrote the paper.
WANG Ronghua: Performed the experiments; Contributed reagents, materials, analysis tools or data.

Funding statement

This work was supported by Scientific research projects of colleges and universities of Inner Mongolia [NJZY21518].

Data availability statement

Data included in article/supp. material/referenced in article.

Table 9. Statistics of vegetation aboveground biomass in study area.

| Grade | AGB/(g/m²) | Arbors | | Shrub | | Herbs | | Crops |
|-------|------------|--------|------------|---------|------------|---------|--------|
|       | Area/(km²) | Proportion/(%) | Area/(km²) | Proportion/(%) | Area/(km²) | Proportion/(%) | Area/(km²) | Proportion/(%) |
| I     | 0 < AGB ≤ 65 | 78.04 | 16.74 | 0.17 | 0.01 | 1245.46 | 75.21 | 45.11 | 0.92 |
| II    | 65 < AGB ≤ 280 | 175.23 | 37.58 | 0.31 | 0.01 | 410.43 | 24.79 | 182.80 | 3.73 |
| III   | 280 < AGB ≤ 950 | 212.23 | 45.68 | 99.98 | 16.74 | 1436.62 | 29.31 |
| IV    | 950 < AGB ≤ 1450 | –– | –– | –– | –– | –– | –– | 3236.96 | 66.04 |
| Total | –– | 466.25 | 100.00 | 2087.74 | 100.00 | 1655.89 | 100.00 | 4901.49 | 100.00 |

Declaration of interest’s statement

The authors declare no conflict of interest.

Additional information

No additional information is available for this paper.

References

[1] Y. Tao, Y.M. Zhang, Multi-scale biomass estimation of desert shrubs: a case study of Haloxylon ammodendron in theGurbantunggut Desert, China, Acta Pratucult. Sin. 22 (6) (2013) 1–10.
[2] M.Y. Wu, G. Dong, Y.J. Wang, et al., Estimation of forest aboveground carbon storage in Sichuan Miyaluo Nature Reserve based on remote sensing, Acta Ecol. Sin. 40 (2) (2020) 621–628.
[3] J.R. Guan, T.Q. Shang, L.T. Yi, et al., Biomass change and community succession characteristics of dominant species in evergreen and deciduous broad-leaved mixed forests in Tianmu Mountain, Acta Ecol. Sin. 37 (20) (2017) 6761–6772.
[4] X. Zhou, X.A. Zuo, X.Y. Zhao, et al., Effect of change in semiarid sand dune habitat on aboveground plant biomass, carbon and nitrogen, Acta Pratucult. Sin. 23 (6) (2014) 36–44.
[5] K. Ren, K. Guo, J.M. Zheng, et al., Ecological benefits of different vegetation restoration modes along the Xinjin to Golmud section of Qinghai-Tibet Railway, Chin. J. Ecol. 38 (3) (2019) 627–636.
[6] S. Issa, B. Dahi, N. Saleous, et al., Carbon stock assessment of date palms using remote sensing coupled with field-based measurements in Abi Dhabi (United Arab Emirates), Int. J. Rem. Sens. 40 (19) (2019) 7561–7580.
[7] S. Issa, B. Dahi, T. Kolski, et al., A review of terrestrial carbon assessment methods using geo-spatial technologies with emphasis on arid lands, Rem. Sens. 12 (12) (2020) 2008–2014.
[8] G. Bindha, P. Rajan, E.S. Jishnu, et al., Carbon stock assessment of mangroves using remote sensing and geographic information system, Egypt. J. Rem. Sens. Space Sci. 23 (2) (2019) 1–9.
[9] D. Chen, Y.Z. Deji, Q.M. Ji, et al., Aboveground biomass estimate methods for typical grassland types in the Tibetan Plateau, Remote Sens. Land Resour. 25 (3) (2011) 43–50.
[10] L. Li, E.X. Chen, Z.Y. Li, et al., A review on forest height and above-ground biomass estimation based on synthetic aperture radar, Remote Sensing Technology and Application 31 (4) (2016) 625–633.
[11] S.M. Ghosh, M.D. Behera, Aboveground biomass estimation using multi-sensor data synergy and machine learning algorithms in a dense tropical forest, Appl. Geogr. 96 (2018) 29–40.
[12] Y. Ren, H.B. Wang, D.P. Xu, Estimation of aboveground biomass of arbor forest based on Landsat 8 image, Forest Resour. Manage. 6 (2018) 38–44.
[13] Z.N. Qiao, Q.H. Geng, Y.N. Xu, Estimating Poplar plantation biomass using satellite remote sensing data, J. Northeast For. Univ. 47 (5) (2019) 66–71.
[14] G.Q. Yang, Y. Ma, J.B. Wang, et al., AGB of Tamarix remote sensing estimation research based on GF-1 image-Take Changyi Tamarix national special marine reserves as an example, Mar. Environ. Sci. 37 (1) (2018) 78–85.
[15] Z.D. Ding, Y.J. Sun, Z. Sun, Estimation of tree biomass with GF-2, J. Beijing Normal Univ. (Nat. Sci.) 57 (1) (2021) 131–145.
[16] X.F. Yang, M. Zan, M.T. Munire, Estimation of above ground biomass of Populus euphratica forest using UAV and satellite remote sensing, Trans. Chin. Soc. Agric. Eng. 37 (1) (2021) 77–83.
[17] H.F. Zhao, X.D. Li, D. Zhang, et al., Aboveground biomass in grasslands in Qinghai Province estimated from MODIS data and its influencing factors, Acta Pratucult. Sin. 29 (12) (2020) 5–16.
[18] W. Zhou, H. Li, L. Xie, et al., Remote sensing inversion of grassland aboveground biomass based on high accuracy surface modeling, Ecol. Indicat. 121 (2021) 107215–107220.
[19] Y. Zhang, X.J. Yin, W.Q. Wang, et al., Estimation of grassland aboveground biomass using Landsat 8 OLI satellite image in the Northern hillside of Tianshan Mountain, Remote Sens. Technol. Appl. 32 (6) (2017) 1012–1021.
[20] A.W. Zhang, S. Zhang, C.F. Guo, et al., Grass biomass inversion based on Landsat 8 spectral derived data classification system, Spectrosc. Specr. Anal. 40 (1) (2020) 239–246.
[21] S.Z. Sun, C.J. Wang, X.J. Yin, et al., Estimating aboveground biomass of natural grassland based on multispectral images of unmanned aerial vehicles, J. Remote Sens. 22 (5) (2018) 848–856.

[22] K. Xu, Y. Su, J. Liu, et al., Estimation of degraded grassland aboveground biomass using machine learning methods from terrestrial laser scanning data, Ecol. Indicat. 108 (2020) 105747–105752.

[23] X.Y. Kang, A.W. Zhang, H.Y. Pang, Estimation of grassland aboveground biomass from UAV-mounted hyperspectral image by optimized spectral reconstruction, Spectrosc. Spectr. Anal. 41 (1) (2021) 250–256.

[24] B. Li, X. Xu, L. Zhang, et al., Aboveground biomass estimation and yield prediction in potato by using UAV-based RGB and hyperspectral imaging, ISPRS J. Photogrammetry Remote Sens. 162 (2020) 161–172.

[25] H. Tao, H. Feng, L. Xu, et al., Estimation of crop growth parameters using UAV-based hyperspectral remote sensing data, Sensors 20 (5) (2020) 1296–1300.

[26] Y.Q. Liu, F. Yan, J.H. Chen, Applying Landsat 8 OLI to estimate biomass in Pisha sandstone area, Res. Soil Water Conserv. 28 (2) (2021) 135–140.

[27] J.Y. Ye, B. Wu, M.H. Liu, et al., Estimation of aboveground biomass of vegetation in the desert-oasis ecotone on the Northeastern edge of the Ulan Buh Desert, Acta Ecol. Sin. 38 (4) (2018) 1216–1225.

[28] X.M. Wang, J.J. Dong, Estimation of the aboveground biomass of desert steppes and typical steppe in Inner Mongolia using generalized linear model, Acta Agrestia Sin. 28 (6) (2020) 1711–1718.

[29] B.Z. He, J.J. Ding, B.H. Liu, et al., Spatiotemporal variation of soil salinization in Weigan-Kuqa river delta oasis, Sci. Silvae Sin. 55 (9) (2019) 185–196.

[30] Y. Huang, X.M. Wang, Analysis of species diversity of typical plant communities in oasis-desert transition zone of Tarim basin northern margin, Southwest China J. Agric. Sci. 31 (3) (2018) 605–610.

[31] X.Q. Jia, M.C. Feng, W.D. Yang, et al., Hyperspectral estimation of aboveground dry biomass of winter wheat based on the combination of vegetation indices, Chin. J. Ecol. 37 (2) (2018) 424–429.

[32] I. Ali, F. Cawkwell, E. Dwyer, et al., Satellite remote sensing of grasslands: From observation to management-A review, J. Plant Ecol. 9 (6) (2016) 649–671.

[33] M.K. Nesha, Y.A. Hussin, L. Leeuwen, et al., Modeling and mapping aboveground biomass of the restored mangroves using ALOS-2 PALSAR-2 in East Kalimantan, Indonesia, Int. J. Appl. Earth Obs. Geoinf. 91 (2020) 102158–102163.

[34] L. Sun, M. Wang, X. Fan, Spatial pattern and driving factors of biomass carbon density for natural and planted coniferous forests in mountainous terrain, Eastern Loess Plateau of China, Forest Ecosyst. 7 (1) (2020) 104–116.

[35] J.Z. Wu, F. Sun, Y. Cai, et al., Relationship between vegetation biomass and soil bulk density on unstable slopes in different climatic regions: a case study of Jiangjiagou watershed in Dongchuan District of Kunming City, Yunnan Province of Southwestern China, J. Beijing For. Univ. 42 (3) (2020) 24–35.