Estimation of Grassland Height Based on the Random Forest Algorithm and Remote Sensing in the Tibetan Plateau

Jianpeng Yin, Qisheng Feng, Tiangang Liang, Baoping Meng, Shuxia Yang, Jinlong Gao, Jing Ge, Mengjing Hou, Jie Liu, Wei Wang, Hui Yu, and Baokang Liu

***Abstract*—Grassland height is one of the main factors used to evaluate grassland conditions. However, the retrieval of natural grassland height at the regional scale by remote sensing data and conventional statistical models will result in large errors, especially in the heterogeneous alpine grassland of the Tibetan Plateau (TP). In this article, we aimed to construct a model based on multiple variables (biogeographical, meteorological, and Moderate Resolution Imaging Spectroradiometer (MODIS) product) using a random forest (RF) algorithm to predict the spatial distribution of grassland height in the TP from 2003 to 2017. The results show the following conditions. 1) Seven variables (elevation, slope, aspect, enhanced vegetation index, reflectance in band seven of MODIS (B7), annual accumulated temperature ($\geq$0 °C), and annual precipitation) that were selected by recursive feature elimination from 11 variables have high importance in the RF model. The final model exhibits good performance, with mean R^2 and root mean squared error values of 0.51 and 6.15 cm, respectively, which were determined via 10-fold crossvalidation. 2) The mean grassland height (2003–2017) predicted by the RF model ranges from 5 to 10 cm in most areas of the TP, and the mean height is 10 cm. The grassland height in the east and southeast of the TP is significantly higher than that in other areas. 3) This article achieves a relatively accurate estimation of grassland height over a large spatial scale at 500-m spatial resolution, which plays an important role in accurately estimating aboveground biomass and evapotranspiration over grassland.**

***Index Terms*—Grassland height, random forest (RF), Tibetan Plateau (TP).**

I. Introduction

THE GRASSLAND ecosystem in the Tibetan Plateau (TP) plays a significant role in the global carbon cycle [1], global climate change [2], and the development of the pastoral economy [3]. As a typical biophysical characteristic, grassland height is used not only to accurately assess grassland aboveground biomass [3], [4], but also as an important indicator for estimating grassland degradation [5]. In addition, grassland height is an indispensable prediction parameter for the early warning and intensity estimation for snow disasters [6]. Grassland height also has an impact on the estimation of sensible heat flux and surface evapotranspiration in the TP [7] where grassland is the dominant vegetation type [8]. Therefore, accurate and dynamic monitoring of grassland height is of utmost importance.

To date, grassland height has been predicted by remote sensing data derived from various sensors, including LiDAR [9]–[11], reflected GPS signals [12], ultrasonic sensors [13], unmanned aerial vehicles [14]–[16], airborne synthetic aperture radar sensors [17], and airborne hyperspectral images [18]. These sensors provide fast- and high-accuracy monitoring of grassland height over a small scale. At the regional scale, the height retrieved in heterogeneous natural grasslands usually depends on multispectral satellite data and conventional statistical models. For example, Marsett *et al.* [19] mapped the height of herbaceous vegetation in arid grasslands at 30-m resolution based on a linear model constructed by using the near-infrared band of Landsat 5 ($R^2 = 0.85$); meanwhile, they found this method is not applicable when forbs account for more than 30% of the total vegetation cover. He *et al.* [20] retrieved the grassland height in Grasslands National Park of Canada using a regression model of leaf area index (LAI) acquired from SPOT 4 HRV images, and the accuracy of the canopy height map reached 83.1% because the measured LAI was strongly correlated with the grassland height in the study area. Chen *et al.* [7] constructed a canopy height model to predict vegetation height, which assumed that the grassland height changes throughout the growing

Manuscript received October 10, 2019; accepted November 14, 2019. Date of publication December 2, 2019; date of current version February 12, 2020. This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFC0504801; in part by the National Natural Science Foundation of China under Grants 31702175, 31672484, 41805086, and 41801191; and in part by the Program for Changjiang Scholars and Innovative Research Team in University IRT_17R50. *(Corresponding author: Tiangang Liang.)*

J. Yin, Q. Feng, T. Liang, J. Gao, J. Ge, M. Hou, and J. Liu are with the State Key Laboratory of Grassland Agro-ecosystems; Key Laboratory of Grassland Livestock Industry Innovation, Ministry of Agriculture and Rural Affairs; Engineering Research Center of Grassland Industry, Ministry of Education; College of Pastoral Agriculture Science and Technology, Lanzhou University, Lanzhou 730000, China (e-mail: yinjp17@lzu.edu.cn; fengqsh@lzu.edu.cn; tgliang@lzu.edu.cn; gaojl16@lzu.edu.cn; gej12@lzu.edu.cn; houmj17@lzu.edu.cn; liuj2018@lzu.edu.cn).

B. Meng is with the School of Geographic Sciences, Nantong University, Nantong 226007, China (e-mail: mengbp09@lzu.edu.cn).

S. Yang is with the Environmental Monitoring Central Station of Gansu, Lanzhou 730000, China (e-mail: 289865617@qq.com).

W. Wang is with the Institute of Arid Meteorology, China Meteorology Administration, Lanzhou 730020, China (e-mail: wangwei9969@163.com).

H. Yu is with the Gansu Science Institute of Soil and Water Conservation, Lanzhou 730000, China (e-mail: yuh2004@163.com).

B. Liu is with the College of Resource and Environmental Engineering, Tianshui Normal University, Tianshui 741000, China (e-mail: liubk04@qq.com).

Digital Object Identifier 10.1109/JSTARS.2019.2954696

season as a function of the normalized difference vegetation index (NDVI).

However, the above methods of retrieving grassland height at the regional scale do not consider the influence of other factors (i.e., climatic and topographic) on grassland height, and these models were applied to homogeneous grassland with small height variability. Compared to these study areas, the alpine grassland of the TP is vast, with many vegetation types and large community composition differences, which lead to increased variability in grassland height. Our previous research indicated low accuracy of alpine grassland height retrievals that were obtained by combining a simple power model and enhanced vegetation index (EVI) from Moderate Resolution Imaging Spectroradiometer (MODIS) ($R^2 = 0.2496$, RMSE = 7.02 cm), although EVI is sensitive to alpine grassland height [3]. Therefore, it must be acknowledged that conventional statistical models have limitations in accurately predicting the height of alpine grassland at the regional scale. A model for improving the prediction accuracy of alpine grassland height is urgently needed.

The random forest (RF) algorithm is a machine learning method for regression and classification [21]. RF exhibits improved performance in addressing high-dimensional regression problems [22] owing to several advantages such as avoiding overfitting, processing the nonlinearity of variables [23], and reducing calculation time [24]. For these reasons, the RF algorithm has been mainly used to construct predictive models for estimating forest height at large spatial scales in recent years [25]–[28]. Above all, the complex models constructed by the machine learning algorithm provide support for accurate prediction over large spatial scales. The combination of the RF algorithm and multispectral satellite data for the prediction of alpine grassland height is worth exploring.

In this article, we collected many field-measured grassland height data and remotely sensed data from 2003 to 2017 in the TP. The primary aims of this study are to 1) develop a predictive model based on multiple variables using the RF algorithm and 2) map the spatial distribution of grassland height using a RF model over the TP.

II. Methods and Materials

A. Study Area

Known as the "third pole" of the Earth [29], the TP (26.5–39.5°N, 78.3–103.1°E) lies in southwest China [Fig. 1(a)] and has an average elevation of over 4000 m and an area of approximately 2.5 million km^2 [30]. The annual average temperature is usually below 0 °C, and the annual precipitation gradually decreases from more than 1000 mm in the southeast to less than 50 mm in the northwest [31]. Grassland is a major vegetation type in the TP, and grassland is a fragile ecosystem that accounts for more than 50% of the area [32]. The main grassland types are alpine meadow, alpine steppe, and alpine desert, which cover approximately 47%, 43%, and 9% of the total grassland area in the TP, respectively [Fig. 1(b)]. In the areas with extensive grasslands, the grassland height is variable because of variations in the climate, grazing patterns, and complex species compositions.

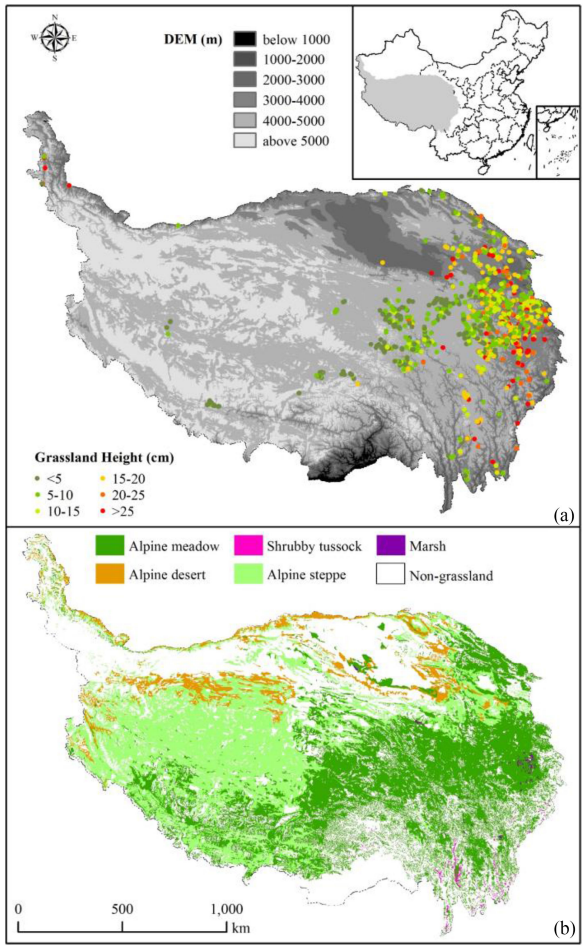


Fig. 1. Distribution of sample plots from 2003 to 2017 (a) and grassland types (b) in the TP.

B. Field Data Collection

Grassland height was measured during the peak growing season (July and August) from 2003 to 2017. To ensure spatial matching with the MODIS pixel size, the sample plots were established with an area of at least 1 km^2 over flat terrain and in representative grassland areas [Fig. 1(a)]. The distance between sample plots was more than 5 km. In each sample plot, five quadrats were uniformly set with a size of 0.5 m × 0.5 m for homogeneous grassland and 1 m × 1 m for complex grassland. In each quadrat survey, we selected more than 10 individual plants along the 2 diagonal direction of the quadrat to manually measure their natural plant height and then averaged all measured heights as the mean height of the sample plot. Meanwhile, the latitude, longitude, grassland type, and cover of each quadrat were recorded. A total of 4090 samples were collected over the past 15 years.

C. MODIS Data and Processing

MOD13A1 (version 006) data with 500-m resolution during the peak season (July and August) over 15 years (2003–2017) were downloaded from the NASA data system (https://earthdata.nasa.gov/). There are two composite images per month, and

TABLE I
ACQUIRED MOD13A1 PRODUCT

Beginning Date of Composite Image		Temporal Resolution	Spatial Resolution	Data Layers Acquired
July	August			
6/26, 7/12	7/28, 8/13	16 day	500 m	NDVI, enhanced vegetation index (EVI), red reflectance (B1), NIR reflectance (B2), blue reflectance (B3), MIR reflectance (B7)

NIR represents the near-infrared band and MIR represents the mid-infrared band; 6/26 represents the date of the first day of the 16 day's images in the product, as well as 7/12, 7/28, and 8/13.

TABLE II
EQUATIONS AND REFERENCES FOR THE INDEXES

Indexes	Formulation	Reference
Soil-adjusted vegetation index (SAVI)	$\frac{NIR-R}{NIR+R+L}(1+L)$	[35]
Modified soil-adjusted vegetation index (MSAVI)	$\frac{2NIR+1-\sqrt{(2NIR+1)^2-8(NIR-R)}}{2}$	[36]
Optimized soil-adjusted vegetation index (OSAVI)	$\frac{NIR-R}{NIR+R+0.16}$	[37]

R represents the red band, and NIR represents near-infrared band; $L = 0.5$ in the SAVI equation.

detailed descriptions are shown in Table I. The data processing included several steps as follows.

1) Transforming and registering original images in HDF format to ESRI grid format with the Albers map projection by using the MODIS Reprojection Tool (MRT).
2) Extracting B1, B2, B3, and B7 from the MODIS data layers to compute three other VIs (i.e., SAVI, MSAVI, and OSAVI) (Table II).
3) Averaging the five VIs and B7 to generate approximate monthly data for July and August in each year (flagged invalid values were not included in the processing) [33], [34].
4) Matching monthly data from the sample plots according to sampling time.

D. SRTM Data

The Shuttle Radar Topography Mission data with 90-m resolution were downloaded from http://srtm.csi.cgiar.org in GeoTIFF format to generate high-quality elevation, slope, and aspect data. Because the aspect is a circular variable, it was then sine transformed in the present study [see (1)], where aspect_T is the new aspect value after transformation. For subsequent modeling, elevation, slope, and aspect_T were resampled at the same resolution as the MOD13A1 data and extracted as variables corresponding to the sample plots.

$$\text{aspect\_}T = \sin\left(\frac{\text{aspect}\times\pi}{180}\right). \quad (1)$$

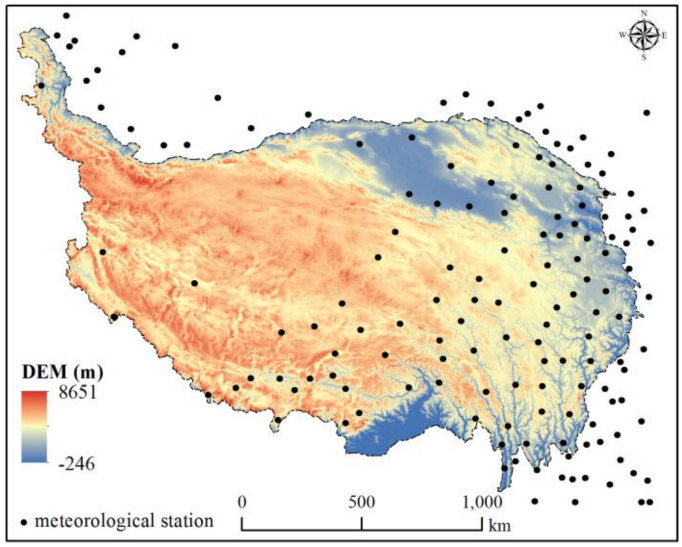


Fig. 2. Spatial distribution of meteorological stations in the TP.

TABLE III
PREDICTOR VARIABLES IN THIS STUDY

Type	Variables	Description
Biogeographical parameters	Elevation	Elevation (m)
	Slope	Slope (°)
	Aspect_T	Sine of aspect
MODIS parameters	NDVI	Monthly mean NDVI
	EVI	Monthly mean EVI
	SAVI	Monthly mean SAVI
	MSAVI	Monthly mean MSAVI
	OSAVI	Monthly mean OSAVI
	B7	Monthly mean B7
Meteorological parameters	T	Annual accumulated temperature (≥ 0 °C)
	P	Annual precipitation (mm)

E. Meteorological Data

Hydrothermal conditions are strongly correlated with vegetation growth [38]. Annual accumulated temperature (≥0 °C) and annual precipitation data were collected from 151 meteorological stations from 2003 to 2017 (Fig. 2), and the data were acquired from http://data.cma.cn/.

The ANUSPLIN software (version 4.3) was developed for accurate spatial interpolation of meteorological data [39] and was widely used [40]. According to working principle of thin-plate spline, the surface of precipitation and temperature was fitted. Because topographical factors have a great impact on the spatial distribution of temperature in mountainous regions [41], elevation was considered a covariate in the spatial interpolation [42]. The final spatial distribution of meteorological data had a resolution of 500 m.

TABLE IV
DESCRIPTIVE STATISTICS OF THE MEASURED GRASSLAND HEIGHT AND CORRESPONDING MODIS EVI DURING THE PEAK GROWING SEASON FROM 2003 TO 2017

Grassland type	Samples	Height (cm)					EVI				
		Min	Max	Avg	Std	CV	Min	Max	Avg	Std	CV
Alpine meadow	3334	2.0	67.6	12.8	9.0	0.71	0.11	0.76	0.47	0.12	0.26
Alpine steppe	616	2.0	53.0	10.9	7.2	0.66	0.10	0.62	0.29	0.12	0.43
Alpine desert	99	3.0	48.0	12.3	7.1	0.57	0.11	0.39	0.20	0.06	0.30
Marsh	11	2.4	18.7	7.2	5.2	0.73	0.21	0.63	0.40	0.13	0.32
Shrubby tussock	30	4.7	34.2	13.9	7.4	0.53	0.14	0.40	0.26	0.09	0.35
All	4090	2.0	67.6	12.5	8.8	0.70	0.10	0.76	0.44	0.14	0.33

F. RF Model Construction

The RF model was established using the random Forest package (version 4.6-14) in *R* software (version 3.5.1). Because the RF algorithm is based on bootstrap resampling and the decision tree method [43], there are two key parameters that can optimize the accuracy of the model: ntree and mtry. The ntree parameter is the number of decision trees, and the accuracy of the model usually increases as ntree increases, which increases the calculation time. The mtry parameter determines the number of node splits per tree, and its default value is one-third of the number of independent variables [44]. After the optimal parameter exploration process, mtry, and ntree were set to 2 and 900, respectively.

A total of 11 predictors were initially selected as variables (Table III). To improve the accuracy and efficiency of the RF model, the variables were further selected using recursive feature elimination (RFE), which is a popular automatic method for feature selection provided by the caret package (version 6.0-81) in *R* software [45]–[47]. The process iteratively calculates the importance of all variables added into the RF model and removes the least important variables until there is one remaining variable in the model. Finally, the optimal model with the minimum number of variables will be determined according to a threshold of model performance.

G. Accuracy Evaluation

The performance of the RF model was evaluated by a 10-fold crossvalidation approach, which ensured effective generalization [48]. All data were divided equally into 10 groups before establishing the model. Each time the algorithm ran, one group was removed and set aside as the test set, and the other nine groups were used as training sets to establish the model. Therefore, each process generated a determination coefficient (R^2) and root mean squared error (RMSE) for the training set and test set. After 10 runs, the mean R^2 and RMSE values of the test set were regarded as indicators for evaluating model accuracy and stability.

III. RESULTS

A. Descriptive Statistics of Grassland Height and EVI from 2003 to 2017

Table IV shows the descriptive statistics of the measured grassland height in July and August and the corresponding EVI over the past 15 years in the TP. During the peak growing season, all observed grassland heights ranged from 2 to 67.6 cm (65.6 cm difference), and the coefficient of variation (CV) was 0.7. The corresponding MODIS EVI was 0.1–0.76 (0.66 difference), with a CV of 0.33. For the five grassland types in the study area, the average height ranged from 7.2 to 13.9 cm and the corresponding CV varied from 0.53 to 0.73. The CV of marsh was the largest among five grassland types, followed by alpine meadow, alpine steppe, alpine desert, and shrubby tussock. The CV of the EVI for the five grassland types (0.26–0.43) was far less than that of height.

TABLE V
VARIABLES SELECTED BY RFE AND THE VARIATIONS IN R^2 AND RMSE WITH THE ADDITION OF THE RESPECTIVE VARIABLE

	Variables	R^2	ΔR^2	RMSE (cm)	Δ RMSE(cm)
2	Elevation and T	0.41	–	6.76	–
3	P	0.46	+0.05	6.46	-0.30
4	B7	0.49	+0.03	6.26	-0.20
5	Slope	0.50	+0.01	6.22	-0.04
6	Aspect_T	0.51	+0.01	6.17	-0.05
7	EVI	0.51	0	6.16	-0.01

B. Variable Selection

Table V shows the selection results via RFE from the initial 11 variables. Four redundant variables (NDVI, SAVI, MSAVI, and OSAVI) are excluded, and seven variables are selected for constructing the optimal RF model. The best bivariate model includes elevation and *T* with mean R^2 and RMSE values of 0.41 and 6.76 cm, respectively. When *P* and B7 are added to the model, the R^2 and RMSE are 0.49 and 6.26 cm, respectively. With the addition of the other three variables (slope, aspect_*T,* and EVI), the R^2 increases slightly from 0.49 to 0.51, and RMSE slightly decreases from 6.26 to 6.16 cm.

The importance values of the selected variables are shown in Fig. 3. Elevation is the most important biogeographical parameter affecting grassland height, while aspect_*T* and slope are the least important biogeographical parameters. Of the MODIS parameters, B7 is more important than EVI. Meteorological parameters (*P* and *T*) are also of high importance.

C. Accuracy Assessment of the RF Model

The training of the final RF model contains seven variables selected by RFE (Table V). After 10-fold crossvalidation, the mean R^2 and RMSE of the training set are 0.93 and 2.85 cm, respectively. For the test set, the mean R^2 and RMSE are 0.51 and 6.15 cm, respectively, which indicates that the RF model exhibits good performance for predicting grassland height.

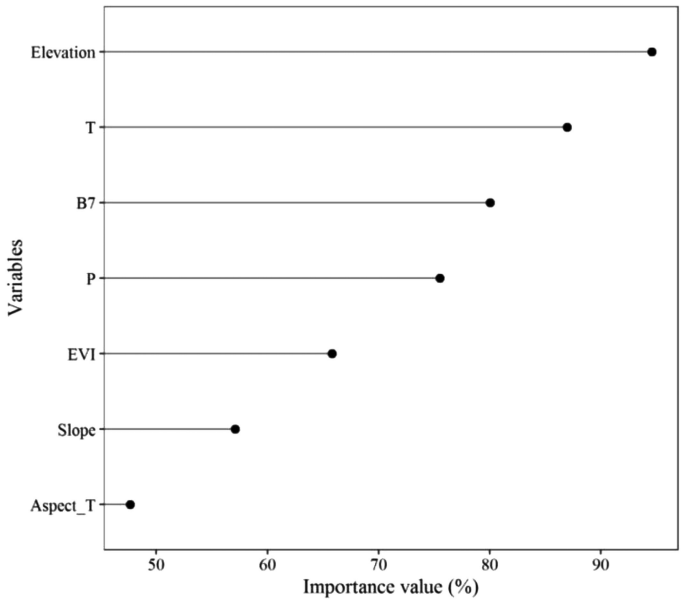


Fig. 3. Importance of selected variables in the RF model.

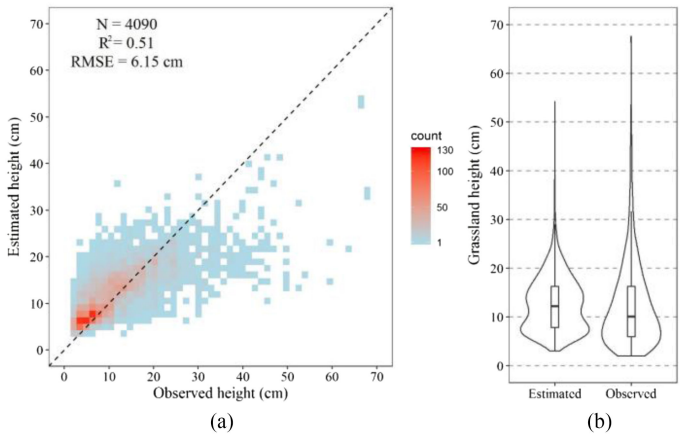


Fig. 4. Density of estimated grassland height (a) and violin plot of estimated height and observed height (b).

The comparison between the estimated height and observed height is shown in Fig. 4. Overall, the RF model underestimates grassland height in the range of high observed values (above 30 cm) [Fig. 4(a)] and overestimates in the range with low observed values (below 3 cm) because the grassland heights below 3 cm (3.6% of the overall observation values) were not predicted by the model [Fig. 4(b)]. Meanwhile, the median estimated grassland height is higher than the observed grassland height [Fig. 4(b)].

D. Spatial Distribution of Grassland Height

The spatial distribution of grassland height in August of each year (2003–2017) is predicted (shown in Fig. 5) by the optimal RF model (mentioned above). Fig. 6 shows the spatial distribution and pixel frequency of the mean grassland height

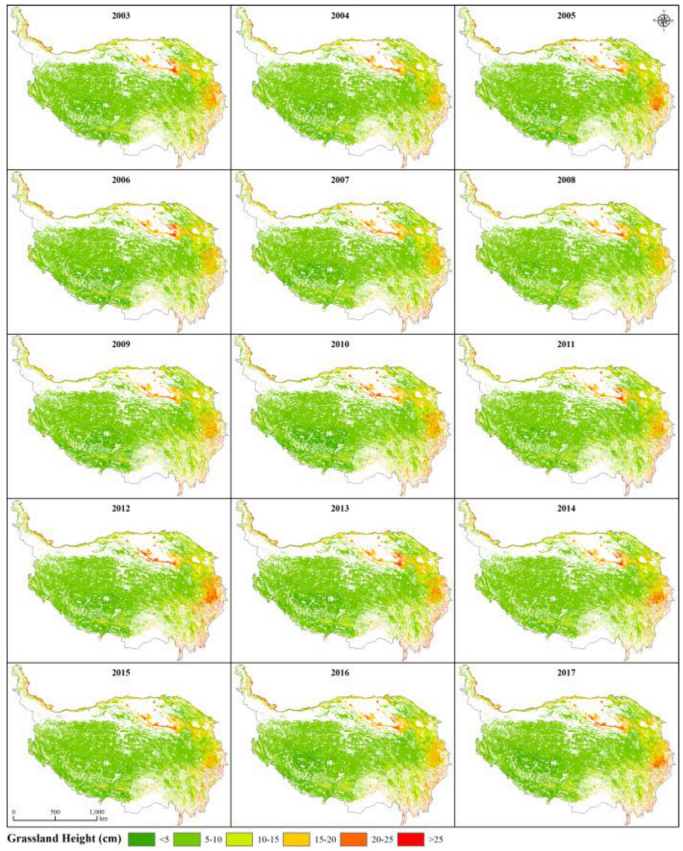


Fig. 5. Spatial distribution of the grassland height in August from 2003 to 2017 in the TP.

over the past 15 years. The predicted grassland height ranges from 3 to 44 cm, and the mean value is 10 cm. Overall, the areas with grassland heights below 5 cm account for only 2% of the total grassland area, which are distributed in a part of the southwest TP. In most regions of the study area (especially in the middle and most part of western of the TP), the grassland height is between 5 and 10 cm, which covers approximately 75% of the total grassland area. The area with grassland heights of 10–15 cm is similar to the area with grassland heights of 15–20 cm (11 vs 9%, respectively). These areas are located in a part of northeastern and some marginal areas in the north of the TP. The area with tall grassland heights (20–25 cm and above 25 cm) accounts for less than 3% of the total grassland area but is distributed in the east and southeast of the TP, where the environmental conditions for grassland growth are superior to the conditions in other regions.

Among the five main grassland types, the predicted grassland height is mostly between 5 and 10 cm in alpine meadows, alpine steppe, and alpine desert, accounting for 70, 86, and 54% of the total area for each grassland type, respectively. In the part of marsh, the predicted heights of 15–20 cm and 20–25 cm cover a large proportion (more than 68%). The predicted height in shrubby tussock is obviously higher than that in other grassland types, and the grassland with heights above 25 cm cover more than 30% of the area.

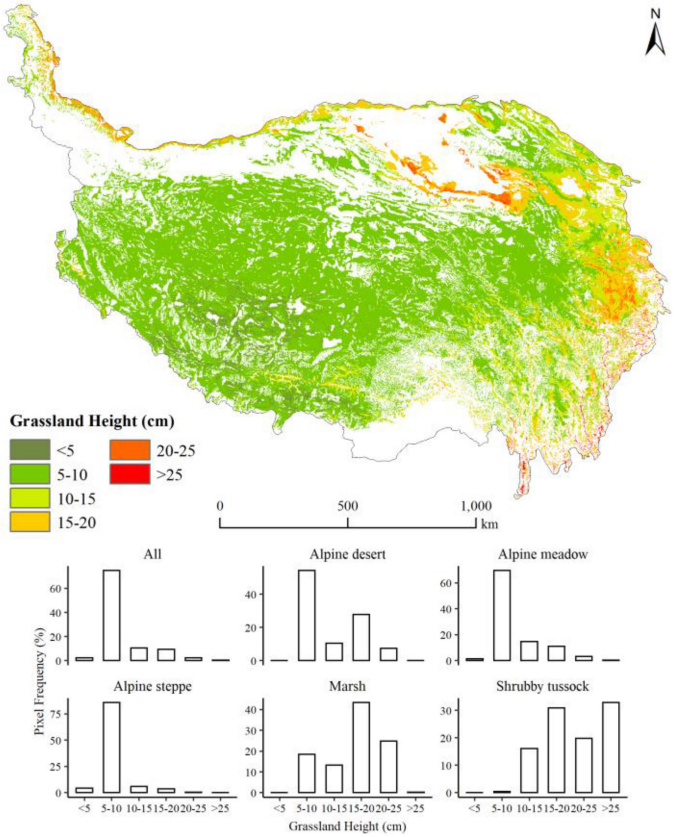


Fig. 6. Spatial distribution and pixel frequency of the mean grassland height in August from 2003 to 2017 in the TP.

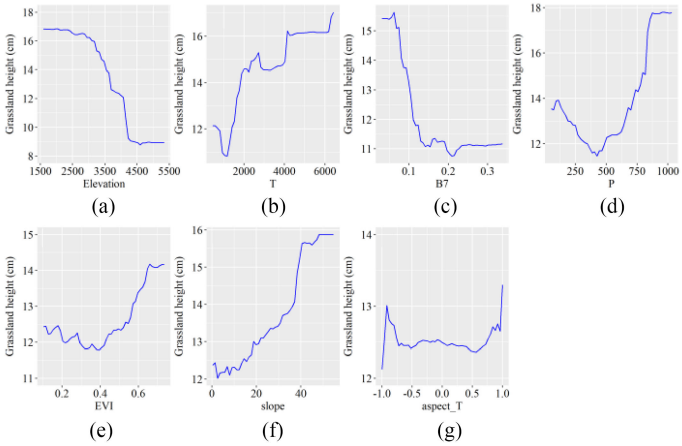


Fig. 7. Partial dependence of the variables based on the RF model.

IV. Discussion

A. Selection of Variables

Variable selection is the most critical step before modeling. Although RF model has many advantages, their accuracy is still reduced by the influence of multicollinearity among variables [49], [50]. Hence, this problem is worth considering when using RF models. In the present study, RFE was used for further variable selection, and the results showed that the final model with seven variables performed better than the model with all eleven variables. The outcome was affected by higher multicollinearity between five of the VIs (EVI, NDVI, SAVI, OSAVI, and MSAVI) among the 11 variables, and only the EVI was retained in the final RF model.

In addition, RFE also inevitably selected variables with marginal contributions. As shown in Table V, when slope, aspect_*T*, and EVI were added to the model, the R^2 increased by only 0.02, and the RMSE decreased by only 0.1 cm compared to the model with four variables. This issue, on the positive side, will increase the robustness of the model.

B. Influence of Variables on Grassland Height

The seven variables selected for predicting grassland height are comprehensible. Previous studies have declared that biogeographical parameters and meteorological parameters are important factors that influence vegetation distribution and growth [51]–[53]. The MODIS band 7 (B7) is in the shortwave infrared section of the spectrum (centered at 2130 nm) and is closely related to cellulose and lignin [54], which are indispensable elements for increasing vegetation height. EVI is the most sensitive vegetation index for predicting alpine grassland height than other MODIS VIs [3]. In addition to these variables, LAI is extremely correlated with both grassland height and aboveground biomass [20]. However, the LAI provided by the MODIS MOD15A2H product has more null values in the western and northern TP, which results in deficiencies in mapping the grassland height. Thus, we abandoned LAI in this study.

Fig. 7 shows the partial dependence of the variables, which quantifies the influence of each variable on grassland height in the RF model. Overall, the response of grassland height to each variable is nonlinear. The grassland height decreases rapidly with increasing elevation and B7 [Fig. 7(a) and (c)]. In contrast, grassland height increases with increasing slope [Fig. 7(f)]. When the annual precipitation is greater than 450 mm, its positive effect on grassland height is obvious [Fig. 7(d)]. When the annual accumulated temperature ($\geq$0 °C) is within approximately 1100 to 4200 °C, the influence on grassland height is great [Fig. 7(b)], while it has almost no influence on grassland height when it is higher than 4200 °C [Fig. 7(b)]. Aspect affects the change in grassland height by only approximately 1 cm, which reflects that it has little contribution to the accuracy of the RF model. The influence of EVI on grassland height is slightly higher than that of aspect [Fig. 7(e)].

C. Model Evaluation

Mtry and ntree are hyperparameters in the RF model. In this article, we tested several mtry and ntree values to observe the performance of the RF model via 10-fold crossvalidation. Because the final RF model used in this study contains only seven variables, mtry is set to 2–7 and ntree is set to 100–1000 at intervals of 100. Fig. 8 shows the model performance using different mtry and ntree values. However, there are not relatively large differences in the mean RMSE and R^2 values of the

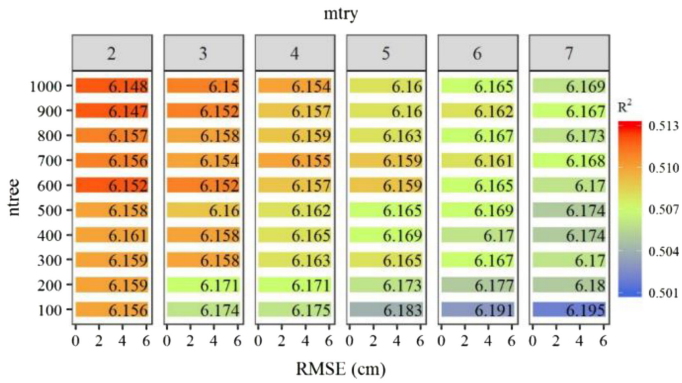


Fig. 8. Mean RMSE and R^2 of the models using different mtry and ntree values.

different models. When mtry and ntree are set to 2 and 900, respectively, the model exhibits the best performance. Fig. 8 also reflects that the RF model is insensitive to the adjustment of hyperparameters [55].

Although the RF model has high retrieval accuracy, there are still unavoidable factors that affect model accuracy. First, spatiotemporal inconsistency exists between field-measured data and satellite data. Because MODIS images are more susceptible to clouds over the TP in the summer, the approximate monthly MODIS product is generated by averaging two MODIS products over 16-day intervals, which minimizes the impacts from bad pixels and cloud pixels. Therefore, to ensure a large amount of valid data to train the RF model, we use the monthly MODIS data to match the field-measured data. Because the study area has complicated topography, the representation of the sample plots is not perfect. When grassland tends to degenerate, other surface features (i.e., bare soil and rock) in the MODIS pixel have some influence on vegetation indices (e.g., NDVI and EVI), which were used to establish the model. Hence, we will consider using high spatiotemporal resolution images to aid in future research, such as Landsat and Sentinel, whose band reflectivity is correlated with grassland height [56]. Second, the uneven distribution of the sample plots produces some errors and uncertainty in the construction of the model. Most of the sample plots are distributed in the eastern part of the TP, while the number of sample plots in the middle and western regions is limited due to limited road access and altitude restrictions. Third, due to the uneven distribution of meteorological stations on the TP, there are some errors in the spatial interpolation of meteorological data.

V. Conclusion

In this article, a large number of field-measured grassland height data from 2003 to 2017 provides powerful support for constructing a reasonable model using the RF algorithm to predict the alpine grassland height in the TP. According to previous research, we consider biogeographical parameters (elevation, slope, and aspect), meteorological parameters [annual accumulated temperature ($\geq$0 °C) (T) and annual precipitation (P)], and MODIS remote sensing parameters [NDVI, EVI, SAVI, MSAVI, OSAVI, and band 7 (B7)] as possible variables. Out of the 11 variables, the 7 variables of elevation, slope, aspect, T, P, EVI, and B7 are finally selected by RFE. Of the biogeographical parameters, elevation is more important than slope and aspect. Both T and P are of high importance for grassland height. Of the MODIS parameters, B7 is more important than EVI.

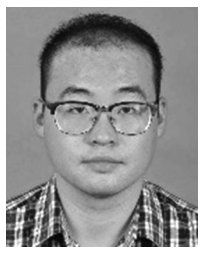
Finally, the RF model has good performance for predicting alpine grassland height (R^2 = 0.51 and RMSE = 6.15 cm). Therefore, the model was used to map the spatial distribution of the mean grassland height in August in the TP over the past 15 years. We found that the distribution of grassland height was reasonable. In most areas of the TP, the grassland height is between 5 and 10 cm, and the grassland heights in the east and southeast are significantly higher than those in other areas.

The RF algorithm cannot only construct a robust model with higher accuracy than conventional statistical models but also provide more auxiliary analysis methods for understanding the mechanisms of the process. In future studies, we will try to use more predictors (e.g., hyperspectral remote sensing and high spatiotemporal resolution images) to increase the precision of alpine grassland height inversion models.

References

[1] S. Piao *et al.*, "Impacts of climate and CO_2 changes on the vegetation growth and carbon balance of Qinghai–Tibetan grasslands over the past five decades," *Global Planet. Change.*, vol. 98–99, pp. 73–80, Dec. 2012.
[2] S. Piao *et al.*, "The carbon balance of terrestrial ecosystems in china," *Nature*, vol. 458, no. 7241, pp. 1009–1013, Apr. 2009.
[3] T. Liang *et al.*, "Multi-factor modeling of above-ground biomass in alpine grassland: A case study in the Three-River Headwaters Region, China," *Remote Sens. Environ.*, vol. 186, pp. 164–172, Dec. 2016.
[4] R. N. Lati, A. Manevich, and S. Filin, "Three-dimensional image-based modelling of linear features for plant biomass estimation," *Int. J. Remote Sens.*, vol. 34, no. 17, pp. 6135–6151, Sep. 2013.
[5] N. Guo, T. Han, J. Wang, T. Han, and B. Sun, "Ecological effects of "restoring grazing to grassland project" in Maqu county," *J. Desert Res.*, vol. 30, no. 1, pp. 154–160, Jan. 2010.
[6] W. Wang *et al.*, "Early warning of snow-caused disasters in pastoral areas on the Tibetan Plateau," *Natural Hazards Earth Syst. Sci.*, vol. 13, no. 6, pp. 1411–1425, 2010.
[7] X. Chen, Z. Su, Y. Ma, K. Yang, J. Wen, and Y. Zhang, "An improvement of roughness height parameterization of the surface energy balance system (SEBS) over the Tibetan Plateau," *J. Appl. Meteorol. Climatol.*, vol. 52, no. 3, pp. 607–622, Mar. 2010.
[8] K. Tan *et al.*, "Application of the ORCHIDEE global vegetation model to evaluate biomass and soil carbon stocks of Qinghai-Tibetan grasslands," *Global Biogeochem. Cycles*, vol. 24, Mar. 2010, Paper GB1013.
[9] P. J. Radtke, H. T. Boland, and G. Scaglia, "An evaluation of overhead laser scanning to estimate herbage removals in pasture quadrats," *Agricultural Forest Meteorol.*, vol. 150, no. 12, pp. 1523–1528, Dec. 2010.
[10] Y. Kaizu, J. M. Choi, and T. H. Kang, "Grass height and yield estimation using a three-dimensional laser scanner," *Environ. Control Biol.*, vol. 50, no. 1, pp. 41–51, Mar. 2012.
[11] G. Bareth, J. Bendig, N. Tilly, D. Hoffmeister, H. Aasen, and A. Bolten, "A comparison of UAV- and TLS-derived plant height for crop monitoring: Using polygon grids for the analysis of crop surface models (CSMs)," *Photogrammetrie Fernerkundung Geoinf.*, vol. 2, pp. 85–94, 2016.
[12] E. E. Small, K. M. Larson, and J. J. Braun, "Sensing vegetation growth with reflected GPS signals," *Geophys. Res. Lett.*, vol. 37, Jun. 2010, Paper L12401.
[13] T. Fricke, F. Richter, and M. Wachendorf, "Assessment of forage mass from grassland swards by height measurement using an ultrasonic sensor," *Comput. Electron. Agriculture*, vol. 79, no. 2, pp. 142–152, Nov. 2011.
[14] W. Van Lersel, M. Straatsma, E. Addink, and H. Middelkoop, "Monitoring height and greenness of non-woody floodplain vegetation with UAV time series," *ISPRS-J. Photogramm. Remote Sens.*, vol. 141, pp. 112–123, Jul. 2018.

[15] H. Zhang *et al.*, "Estimation of grassland canopy height and aboveground biomass at the quadrat scale using unmanned aerial vehicle," *Remote Sens.*, vol. 10, no. 6, Jun. 2018, Art. no. 851.
[16] J. Wijesingha, T. Moeckel, F. Hensgen, and M. Wachendorf, "Evaluation of 3D point cloud-based models for the prediction of grassland biomass," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 78, pp. 352–359, Jun. 2019.
[17] M. J. Hill, G. E. Donald, and P. J. Vickery, "Relating radar backscatter to biophysical properties of temperate perennial grassland," *Remote Sens. Environ.*, vol. 67, no. 1, pp. 15–31, Jan. 1999.
[18] A. Capolupo, L. Kooistra, C. Berendonk, L. Boccia, and J. Suomalainen, "Estimating plant traits of grasslands from UAV-acquired hyperspectral images: A comparison of statistical approaches," *ISPRS Int. Geo-Inf.*, vol. 4, no. 4, pp. 2792–2820, Dec. 2015.
[19] R. C. Marsett *et al.*, "Remote sensing for grassland management in the arid Southwest," *Rangeland Ecol. Manage.*, vol. 59, no. 5, pp. 530–540, Sep. 2006.
[20] Y. He, X. Guo, and J. F. Wilmshurst, "Reflectance measures of grassland biophysical structure," *Int. J. Remote Sens.*, vol. 30, no. 10, pp. 2509–2521, 2006.
[21] J. Xia *et al.*, "Estimates of grassland biomass and turnover time on the Tibetan Plateau," *Environ. Res. Lett.*, vol. 13, no. 1, Jan. 2018, Paper 014020.
[22] S. Karimi, A. A. Sadraddini, A. H. Nazemi, T. Xu, and A. F. Fard, "Generalizability of gene expression programming and random forest methodologies in estimating cropland and grassland leaf area index," *Comput. Electron. Agriculture*, vol. 14, pp. 232–240, Jan. 2006.
[23] M. Belgiu and L. Dragut, "Random forest in remote sensing: A review of applications and future directions," *ISPRS-J. Photogramm. Remote Sens.*, vol. 114, pp. 24–31, Apr. 2016.
[24] P. O. Gislason, J. A. Benediktsson, and J. R. Sveinsson, "Random forests for land cover classification," *Pattern Recognit. Lett.*, vol. 27, no. 4, pp. 294–300, Mar. 2006.
[25] J. M. Kellndorfer, W. S. Walker, E. Lapoint, K. Kirsch, J. Bishop, and G. Fiske, "Statistical fusion of lidar, InSAR, and optical remote sensing data for forest stand height characterization: A regional-scale method based on LVIS, SRTM, Landsat ETM+, and ancillary data sets," *J. Geophys. Res. Atmos.*, vol. 115, pp. 8–18, Mar. 2010.
[26] M. Penner, D. G. Pitt, and M. E. Woods, "Parametric vs. nonparametric LiDAR models for operational forest inventory in boreal Ontario," *Can. J. Remote Sens.*, vol. 39, no. 5, pp. 426–443, Oct. 2013.
[27] I. Fayad *et al.*, "Canopy height estimation in French Guiana with LiDAR ICESat/GLAS data using principal component analysis and random forest regressions," *Remote Sens.*, vol. 6, no. 12, pp. 11883–11914, Dec. 2014.
[28] O. S. Ahmed, S. E. Franklin, M. A. Wulder, and J. C. White, "Characterizing stand-level forest canopy cover and height using Landsat time series, samples of airborne LiDAR, and the Random Forest algorithm," *ISPRS-J. Photogramm. Remote Sens.*, vol. 101, pp. 89–101, Mar. 2015.
[29] J. Qiu, "The third pole," *Nature*, vol. 454, no. 7203, pp. 393–396, Jul. 2008.
[30] K. Yang *et al.*, "Response of hydrological cycle to recent climate changes in the Tibetan Plateau," *Climatic Change.*, vol. 109, no. 3–4, pp. 517–534, Dec. 2011.
[31] S. Piao *et al.*, "Altitude and temperature dependence of change in the spring vegetation green-up date from 1982 to 2006 in the Qinghai-Xizang Plateau," *Agricultural Forest Meteorol.*, vol. 151, no. 12, pp. 1599–1608, Dec. 2011.
[32] E. Bartholomé and A. S. Belward, "GLC2000: A new approach to global land cover mapping from Earth observation data," *Int. J. Remote Sens.*, vol. 26, no. 9, pp. 1959–1977, May 2005.
[33] D. Stojanova, P. Panov, V. Gjorgjioski, A. Kohler, and S. Dzeroski, "Estimating vegetation height and canopy cover from remotely sensed data with machine learning," *Ecol. Inform.*, vol. 5, no. 4, pp. 256–266, Jul. 2010.
[34] J. Bi *et al.*, "Amazon forests' response to droughts: A perspective from the MAIAC product," *Remote Sens.*, vol. 8, no. 4, Apr. 2016, Art. no. 356.
[35] A. R. Huete, "A soil-adjusted vegetation index (SAVI)," *Remote Sens. Environ.*, vol. 25, no. 3, pp. 295–309, Aug. 1988.
[36] J. Qi, A. Chehbouni, A. R. Huete, Y. H. Kerr, and S. Sorooshian, "A modified soil adjusted vegetation index," *Remote Sens. Environ.*, vol. 48, no. 2, pp. 119–126, May 1994.
[37] M. D. Steven, "The sensitivity of the OSAVI vegetation index to observational parameters," *Remote Sens. Environ.*, vol. 63, no. 1, pp. 49–60, Jan. 1998.
[38] W. Zhou, C. Gang, Y. Chen, S. Mu, Z. Sun, and J. Li, "Grassland coverage inter-annual variation and its coupling relation with hydrothermal factors in China during 1982–2010," *J. Geograph. Sci.*, vol. 24, no. 4, pp. 593–611, Aug. 2014.
[39] R. J. Hijmans, S. E. Cameron, J. L. Parra, P. G. Jones, and A. Jarvis, "Very high resolution interpolated climate surfaces for global land areas," *Int. J. Climatol.*, vol. 25, no. 15, pp. 1965–1978, Dec. 2005.
[40] J. Gao *et al.*, "Modeling alpine grassland forage phosphorus based on hyperspectral remote sensing and a multi-factor machine learning algorithm in the east of Tibetan Plateau, China," *ISPRS-J. Photogramm. Remote Sens.*, vol. 147, pp. 104–117, Nov. 2019.
[41] X. Zhang, J. Shao, and H. Luo, "Spatial interpolation of air temperature with ANUSPLIN in three gorges reservoir area," in *Proc. IEEE Int . Conf . Remote Sens.*, 2011, pp. 3465–3468.
[42] J. Ge *et al.*, "Modeling alpine grassland cover based on MODIS data and support vector machine regression in the headwater region of the Huanghe River, China," *Remote Sens. Environ.*, vol. 218, pp. 162–173, Sep. 2018.
[43] J. M. G. Taylor, "Random survival forests," *J. Thoracic Oncol.*, vol. 6, no. 12, pp. 1974–1975, Dec. 2011.
[44] Y. Wang *et al.*, "Prediction of aboveground grassland biomass on the Loess Plateau, China, using a random forest algorithm," *Sci Rep.*, vol. 7, Jul. 2017, Art. no. 6940.
[45] M. Kuhn, "Building predictive models in *R* using the caret package," *J. Statist. Softw.*, vol. 28, no. 5, pp. 1–26, Nov. 2008.
[46] A. Ghosh and P. K. Joshi, "A comparison of selected classification algorithms for mapping bamboo patches in lower Gangetic plains using very high resolution WorldView 2 imagery," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 26, pp. 298–311, Feb. 2014.
[47] C. W. Brungard, J. L. Boettinger, M. C. Duniway, S. A. Wills, and T. C. Edwards, "Machine learning for predicting soil classes in three semi-arid landscapes," *Geoderma*, vol. 239, pp. 68–83, Feb. 2015.
[48] J. Chou and A. D. Pham, "Nature-inspired metaheuristic optimization in least squares support vector regression for obtaining bridge scour information," *Inf. Sci.*, vol. 399, pp. 64–80, Aug. 2017.
[49] J. S. Evans and S. A. Cushman, "Gradient modeling of conifer species using random forests," *Landscape Ecol.*, vol. 24, no. 5, pp. 673–683, May 2009.
[50] M. A. Murphy, J. S. Evans, and A. Storfer, "Quantifying Bufo boreas connectivity in Yellowstone National Park with landscape genetics," *Ecology*, vol. 91, no. 1, pp. 252–261, Jan. 2010.
[51] F. P. Day and C. D. Monk, "Vegetation patterns on a southern Appalachian watershed," *Ecology*, vol. 55, no. 5, pp. 1064–1074, 1974.
[52] J. Fang, Y. Yang, W. Ma, A. Mohammat, and H. Shen, "Ecosystem carbon stocks and their changes in China's grasslands," *Sci. China-Life Sci.*, vol. 53, no. 7, pp. 757–765, Jul. 2010.
[53] S. Piao, J. Fang, L. Zhou, K. Tan, and S. Tao, "Changes in biomass carbon stocks in China's grasslands between 1982 and 1999," *Global Biogeochem. Cycle.*, vol. 21, no. 2, Apr. 2007, Paper GB2002.
[54] C. S. T. Daughtry, E. R. Hunt, P. C. Doraiswamy, and J. E. McMurtrey, "Remote sensing the spatial distribution of crop residues," *Agronomical J.*, vol. 97, no. 3, pp. 864–871, May 2005.
[55] M. Kuhn and K. Johnson, *Applied Predictive Modeling*, 1st ed. New York, NY, USA: Springer, 2013.
[56] A. Cimbelli and V. Vitale, "Grassland height assessment by satellite images," *Adv. Remote Sens.*, vol. 6, pp. 40–53, 2017.



Jianpeng Yin received the B.S. degree in grassland science from the College of Pastoral Agriculture Science and Technology, Lanzhou University, Lanzhou, China, in 2017. He is currently working toward the M.S. degree in grassland science with Lanzhou University.

His research interests include remote sensing and GIS of grassland.

Qisheng Feng received the B.S., M.S., and Ph.D. degrees in grassland sciences from the College of Pastoral Agriculture Science and Technology, Lanzhou University, Lanzhou, China, in 2006, 2009, and 2012, respectively.

He is currently a Lecturer and a Research Scientist with Lanzhou University. His research interests include remote sensing of grassland and climate change.

Tiangang Liang received the B.S. and M.S. degrees in geographic information system from Lanzhou University, Lanzhou, China, in 1989 and 1992, respectively, and the Ph.D. degree in ecology from the State Key Laboratory of Arid Agroecology, Lanzhou University, in 1998.

He is currently a Professor with the College of Pastoral Agriculture Science and Technology, Lanzhou University, and the Director of the Institute of Grassland Remote Sensing and Geographic Information System, Lanzhou University. His research interests include remote sensing and GIS technologies, theories, and applications on grassland ecology and environment.

Baoping Meng received the B.S. and Ph.D. degrees in grassland science from the College of Pastoral Agriculture Science and Technology, Lanzhou University, Lanzhou, China, in 2013 and 2018, respectively.

He is currently a Lecturer with the School of Geographic Sciences, Nantong University, Nantong, China. His research interests include remote sensing of grassland and climate change.

Shuxia Yang received the B.S. degree in english from WeiNan Normal University, Weinan, China, in 2008, and the Ph.D. degree in grassland science from the College of Pastoral Agriculture Science and Technology, Lanzhou University, Lanzhou, China, in 2017.

She is currently an Engineer with the Environmental Monitoring Central Station of Gansu, Lanzhou, China. Her research interest includes remote sensing of grassland.

Jinlong Gao received the B.S. degree in grassland science from Northwest A&F University, Xianyang, China, in 2014. He is currently working toward the Ph.D. degree in grassland science with the College of Pastoral Agriculture Science and Technology, Lanzhou University, Lanzhou, China.

His research interests include remote sensing inversion for nutrients of alpine grassland.

Jing Ge received the B.S. degree in grassland science from Lanzhou University, Lanzhou, China, in 2015. She is currently working toward the Ph.D. degree with the College of Pastoral Agriculture Science and Technology, Lanzhou University, Lanzhou, China.

Her research interests include modeling for grassland parameters by remote sensing.

Mengjing Hou received the B.S. degree in statistics from the Lanzhou University of Finance and Economics, Lanzhou, China, in 2017. He is currently working toward the Ph.D. degree in grassland science with the College of Pastoral Agriculture Science and Technology, Lanzhou University, Lanzhou, China.

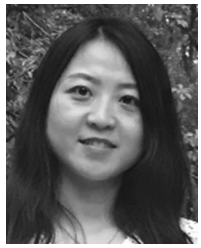
His research interests include land cover classification by remote sensing.

Jie Liu received the B.S. degree in grassland science from Lanzhou University, Lanzhou, China, in 2018. She is currently working toward the M.S. degree in grassland science with the College of Pastoral Agriculture Science and Technology, Lanzhou University, Lanzhou, China.

Her research interests include modeling for grassland parameters by remote sensing.

Wei Wang received the B.S. degree from Gansu Agricultural University, Lanzhou, China, in 2008, the M.S. degree from Lanzhou University, Lanzhou, China, in 2011, and the Ph.D. degree from the State Key Laboratory of Arid Agroecology, Lanzhou University, in 2014, all in grassland science.

He is currently an Associate Researcher with the Institute of Arid Meteorology, China Meteorological Administration, Lanzhou, China. His research interests include remote sensing of vegetation and land surface radiation.



Hui Yu received the B.S. and Ph.D. degrees in grassland science from the College of Pastoral Agriculture Science and Technology, Lanzhou University, Lanzhou, China, in 2008 and 2013, respectively.

She is currently an Engineer with Gansu Science Institute of Soil and Water Conservation, Lanzhou, China. Her research interest includes remote sensing of grassland.

Baokang Liu received the B.S. degree in computer from the Lanzhou University of Technology, Lanzhou, China, in 1995, and the M.S. and Ph.D. degrees in grassland science from the College of Pastoral Agriculture Science and Technology, Lanzhou University, Lanzhou, China, in 2007 and 2018, respectively.

He is currently an Associate Professor with the College of Resource and Environmental Engineering, Tianshui Normal University, Tianshui, China. His research interest includes remote sensing of lake, grassland, and agricultural.