An extraction method of cultivated land in karst area based on Landsat time series

Limeng Wang¹, Jiasheng Wang²,³,*

¹School of Information Science and Technology, Yunnan Normal University, 650000 Kunming, China
²The Engineering Research Center of GIS Technology in Western China Ministry of Education of China, 650000 Kunming, China
³Faculty of Geography, Yunnan Normal University, 650000 Kunming, China
* Corresponding author: jersonwang@ynnu.edu.cn

Abstract. It is of great significance to grasp the distribution of cultivated land in karst area for protecting cultivated land and maintaining regional food security. Based on Landsat 8 OLI remote sensing images, this paper took Fuyuan County of Yunnan Province as an example, used a classification method combining multi-feature construction with random forest to extract the distribution of cultivated land in Fuyuan County in 2019. The experimental results showed that the overall accuracy was 0.93, and the Kappa coefficient was 0.90. This method could reduce the interference of mixed pixels to a certain extent and had a better classification effect. The cultivated land area in Fuyuan County was 1200.42 km². The cultivated land was distributed in all townships of the whole county, and it was mainly concentrated in the northwest, central and part of the southern regions of Fuyuan County, such as Housuo Town, Zhong'an Town, Dahe Town, Yingshang Town, Zhuyuan Town, and Shibalianshan Town.

1. Introduction
Cultivated land is the foundation of agricultural development [1]. Reasonable planning and protection of cultivated land resources is very important for ensuring stable food production and maintaining stable social development [2]. Therefore, accurately grasping the distribution of cultivated land has become an important content of the study of cultivated land. Remote sensing technology provides a faster and more convenient means for the extraction of large-area cultivated land information due to its advantages of high timeliness, accuracy, wide range and reflecting dynamic changes [3]. Since cultivated land has various manifestations due to different crops, it is difficult to distinguish cultivated land from other types of land in a single-phase remote sensing image. It is necessary to use time series images to extract pheno logical characteristics and analyse the distribution of cultivated land. Remote sensing time series images have provided effective data support in surface monitoring, crop planting area extraction and yield prediction [4]. Landsat remote sensing dataset is the main data source for many long-term surface monitoring studies. For example, Liu et al. used Landsat TM/ETM remote sensing images to study the influence of cultivated land change on grain production potential in China from 1990 to 2010 [5]. Man et al. used Landsat TM/ETM+/OLI remote sensing images to analyse the temporal and spatial changes of cultivated land in Northeast China from 1990 to 2013 [6].

In the remote sensing extraction of cultivated land information, the study of cultivated land for karst area is rare. Due to the ecological environment of karst area is fragile and the cultivated land resources
are vulnerable, it is important to study the extraction method of cultivated land in karst area for the protection of cultivated land. Karst area brings great difficulties to ecological environment construction; especially serious soil erosion has caused incalculable losses to cultivated land resources. Therefore, it is more urgent to study cultivated land in karst area.

This paper took Fuyuan County, Yunnan Province as the study area, and took the multi-temporal Landsat image as the data source. We presented a cropland extraction method in karst area based on random forest classification.

2. Materials and Methods

2.1. Study area
The study area is Fuyuan County, Yunnan Province. The county is located in the east of Qujing City, Yunnan Province, with a total land area of 3251 km². The climate of the whole county belongs to the monsoon semi-humid climate in the mountains of the southern temperate zone, with sufficient sunshine, abundant precipitation in the territory, mild four seasons, and annual average temperature of about 14℃. It mainly grows rice, wheat, corn, beans and potatoes and other food crops. The terrain in this area is high in the northwest and low in the southeast, with mountains and canyons.

2.2. Research data
This study selected Landsat 8 OLI imagery surface reflectance product data in 2019 provided by the GEE cloud platform. Through GEE online programming, a total of 8 images with cloud coverage less than 1% in the study area were obtained. The selected images include spring, summer, autumn, and winter multi-season images, which is helpful to extract the coverage information of cultivated land with seasonal changes. According to the principle of randomness and uniformity, the sample data was obtained by visual interpretation of Google Earth high-definition images. The samples include 6 types of cultivated land, grassland, woodland, water body, construction land, and unused land, a total of 300.

2.3. Methods
This paper obtained remote sensing images covering the study area through the GEE platform, and pre-processed images through online programming of the platform API, mainly including image cloud removal, mosaic and clipping. Then, the spectral features, tasseled cap transformation and texture features were constructed and optimized. Finally, the distribution of cultivated land in the study area in 2019 was extracted by using the random forest algorithm. Meanwhile, the accuracy of the extraction results was evaluated. The technical flow was shown in Fig. 1.
2.3.1. Features construction

One of the important steps in remote sensing image classification is the selection and construction of feature variables. In this paper, spectral features, texture features and tasseled cap transformation were selected for construction. Specifically as follows:

- Spectral features. The study selected the normalized vegetation index (NDVI), enhanced vegetation index (EVI), modified soil adjustment vegetation index (MSAVI), modified normalized differential water index (MNDWI), normalized building index (NDBI), and Soil index (BSI). These indices were calculated by the GEE platform and added to the original image as independent spectral bands. The calculation formula is as follows [7-9]:

\[
\text{NDVI} = \frac{\text{NIR} - \text{Red}}{\text{NIR} + \text{Red}} 
\]

\[
\text{EVI} = \frac{2.5 \times (\text{NIR} - \text{Red})}{\text{NIR} + 6.0 \times \text{Red} - 7.5 \times \text{Blue} + 1} 
\]

\[
\text{MSAVI} = \frac{2\text{NIR} + 1 - \sqrt{(2\text{NIR} + 1)^2 - 8 \times (\text{NIR} - \text{Red})}}{\text{Red}} 
\]

\[
\text{MNDWI} = \frac{\text{Green} - \text{SWIR1}}{\text{Green} + \text{SWIR1}} 
\]

\[
\text{NDBI} = \frac{\text{MIR} - \text{NIR}}{\text{MIR} + \text{NIR}} 
\]

\[
\text{BSI} = \frac{(\text{SWIR1} + \text{Red}) - (\text{NIR} + \text{Blue})}{(\text{SWIR1} + \text{Red}) - (\text{NIR} + \text{Blue})} 
\]

- Texture features. Texture is an important feature in remote sensing image classification, especially for cultivated land classification. Gray-Level Co-occurrence Matrix (GLCM) is a commonly used
method of texture analysis [10]. Because the texture features of cultivated land are closely related to vegetation changes, this paper chose NDVI to calculate the texture features. According to the fast texture feature calculation function provided by GEE platform, six features were selected to construct the original texture features: NDVI_asm, NDVI_contrast, NDVI_corr, NDVI_ent, NDVI_idm and NDVI_var.

- Tasseled cap transformation. According to the fixed transformation matrix, the tasseled cap transformation transformed the original image projection to the direction closely related to the types and changes of ground objects, especially related to the plant growth process and soil. The transformation formula is as follows [11]:

\[
Y = cX + a
\]  

In formula (7), \(X\) is the pixel vector in the multispectral space before transformation, \(Y\) is the pixel vector in the multispectral space after transformation, \(c\) is the transformation matrix, and \(a\) is a constant.

2.3.2. Features optimization
A variety of original features were constructed by the above methods, including 13 spectral features, 6 texture features and 3 tasseled cap transformation, with a total of 22 features. Under a certain number of samples, too many features will cause the classification accuracy to decline. Therefore, in order to select the best feature, it is necessary to optimize feature selection. In this paper, Jeffries-Matusita (J-M) distance was chosen as the basis of feature selection. The formula is as follows [12]:

\[
JM = \sqrt{2(1-e^{-a})}
\]

\[
B = \frac{1}{8}(M_i - M_j)^T \left( \frac{V_i + V_j}{2} \right)^{-1} (M_i - M_j) + \frac{1}{2} \ln \left[ \frac{V_i + V_j}{2} \right] \left( \sqrt{\det V_i} \sqrt{\det V_j} \right]
\]  

In the above formula, \(B\) represents Babbitt distance, \(M_i\) and \(M_j\) are sample mean vectors of categories \(i\) and \(j\) respectively, \(V_i\) and \(V_j\) are corresponding sample covariance matrices. It can be concluded from the formula that the range of values is \([0, 2^{1/2}]\). 0 means that two categories are almost completely confused and cannot be distinguished on a certain feature, and \(2^{1/2}\) means that the two categories can be completely separated. The closer their values are, the greater the distinguishability between classes can be distinguished on a certain feature. Because there is overlap among different categories, the situation that J-M value is \(2^{1/2}\) does not exist in actual situations, the top two features of J-M value are reserved to participate in training classification. On the one hand, the number of features needs to be reduced, and on the other hand, considering the principle that the J-M value is closer to \(2^{1/2}\), the greater the distinguishability between categories. So, the feature with J-M value in the top two places and greater than 1 was selected as the optimized feature.

2.3.3. Random forest classification
Random forest is a classifier containing multiple decision trees. The output category results are determined by the mode of the classification results of each decision tree. In this study, 2/3 of the sample data were randomly selected as training samples, and the optimized feature combination was used as input to construct each decision tree. The key to random forest classification mainly depends on the number of decision trees, and it is extremely important to choose an appropriate decision tree value. After many experiments, this paper finally determined that the training effect was best when the number of decision trees was 220.
2.3.4. **Accuracy evaluation**
In order to verify the accuracy of the cultivated land extraction results, the most commonly used quantitative evaluation method can usually be used, that is, accuracy verification by calculating the accuracy index. By calculating the overall accuracy, user accuracy, mapping accuracy and Kappa coefficient of confusion matrix, these index results can reflect the accuracy of classification results [13]. In this paper, the overall accuracy and Kappa coefficient were used to evaluate the accuracy of cultivated land extraction.

3. **Results & Discussion**

3.1. **Feature optimization results**
Six types of objects were combined in pairs, and there were 15 combined results. There were 22 original features, including 13 spectral features, 6 texture features and 3 tasseled cap transformation features. The J-M distance corresponding to each feature in every pairwise combination was calculated, and the features whose J-M distance was greater than 1 and located in the top two places were taken as the optimization principle. Finally 14 optimized features were obtained (Table 1).

| Feature type          | Original feature       | Optimized features    |
|-----------------------|------------------------|-----------------------|
| Spectral features     | B2, B3, B4, B5, B6     | B4, B6                |
|                       | B7, NDVI, EVI          | B7, NDVI              |
|                       | MSAVI, MNDWI           | MNDWI                 |
|                       | RV1, BSI, NDBI         | NDBI, RVI, BSI, EVI   |
| Tasseled cap transformation | brightness, greenness, wetness | Brightness, Greenness, wetness |
| Texture features      | NDVI_asm, NDVI_con, NDVI_corr, NDVI_ent, NDVI_idm, NDVI_var | NDVI_con, NDVI_var |

3.2. **Accuracy verification**
After the classification was completed, this paper used the accuracy index to evaluate the accuracy. Among them, the accuracy index was calculated by the confusion matrix to obtain the overall accuracy and Kappa coefficient. After many experiments, it can be concluded that the overall accuracy of the method was 0.93, and the Kappa coefficient was 0.90, indicating that the result of the cultivated land extraction in Fuyuan County through the combination of the constructed multi-feature and the random forest classification method was reliable.

3.3. **The spatial distribution of cropland**
Based on feature construction and random forest classification methods, this paper extracted the distribution of cultivated land in Fuyuan County in 2019 (Fig. 2). According to the extraction results, it can be calculated that the cultivated land area in Fuyuan County was 1200.42 km². It can be seen from the figure that the cultivated land was distributed in all townships of the whole county, and it was mainly concentrated in the northwest, central and part of the southern regions of Fuyuan County, such as Housuo Town, Zhong'an Town, Dahe Town, Yingshang Town, Zhuyuan Town, and Shibalianshan Town.
4. Conclusions
Based on Landsat 8 OLI time series images, this paper constructed spectral features, tasseled cap transformation and texture features, and extracted the spatial distribution of cultivated land by using random forest classification method. The main conclusions were as follows:

- Based on the classification method combining multi-features and random forest, the overall accuracy of the method was 0.93, and the Kappa coefficient was 0.90. This method could reduce the interference of mixed pixels, and the classification effect was better.
- The cultivated land area in Fuyuan County was 1200.42 km². The cultivated land was distributed in all townships of the whole county, and it was mainly concentrated in the northwest, central and part of the southern regions of Fuyuan County, such as Housuo Town, Zhong'an Town, Dahe Town, Yingshang Town, Zhuyuan Town, and Shibaliashan Town.

Acknowledgement
This work was supported by the National Natural Science Foundation of China (No. 41961056).

References
[1] Z.Q. Fu, Y.L. Cai, et al, Correlation analysis of China's food security and cultivated land resource changes. J. Nat. Resour (04): 313-319 (2001)
[2] M.Q. Wu, X.G. Chen, Analysis of cultivated land area change and its driving force in Qinzhou city based on principal component analysis. Sci. Geogr. Sin 34 (01): 54-59 (2014)
[3] Z.Q. Pan, G.H. Liu, C.H. Zhou, Remote sensing research on the crop planting subarea of the Yellow River Delta. Geogr. Res (06): 799-806+814 (2003)
[4] Q.Y. Xu, G.J. Yang, et al, Crop planting identification based on MODISNDVI multi-year time series data. Trans. Chin. Soc. Agric. Eng 30 (11): 134-144 (2014)
[5] L. Liu, X.L. Xu, et al, Impact of Cultivated Land Change on Grain Production Potential in China from 1990 to 2010. Acta Geogr. Sin 69 (12): 1767-1778 (2014)
[6] W.D. Man, Z.M. Wang, et al, Remote sensing analysis of spatial and temporal changes of cultivated land in Northeast China from 1990 to 2013. Trans. Chin. Soc. Agric. Eng 32 (07): 1-10 (2016)
[7] S.H. Jia, Z.G. Han, et al, Research on desertification information extraction in northern Liaoning Province based on decision tree. Ecol. & Environ 20 (01): 13-18 (2011)

[8] X. Yang, S. Zhao, et al, Mapping of urban surface water bodies from Sentinel-2 imagery at 10m resolution via ndwi-based image sharpening. Remote Sens 9 (6): 596-614 (2017)

[9] K. Li, Y.A. Chen, Genetic algorithm-based urban cluster automatic threshold method by combining viirs dnb, ndvi, and ndbi to. Remote Sens 10 (2): 277-297 (2018)

[10] H. Pei, T.J. Sun, et al, Object-oriented Land Use/Cover Classification Based on Landsat 8 OLI Image Texture Features. Trans. Chin. Soc. Agric. Eng 25 (07): 22-27 (2016)

[11] Q. Liu, G. Liu, et al, Comparison of tasseled cap transformations based on the selective bands of Landsat 8 OLI reflectance images. Int. J. Remote. Sens 36 (2): 417-441 (2015)

[12] B.B. Han, S.B. Chen, et al, Classification of multi-temporal Sentinel-1 crops based on J-M distance. Sci. Tec. & Eng 21 (05): 752-766 (2019)

[13] D. Huang, S. Xu, et al, Accuracy assessment model for classification result of remote sensing image based on spatial sampling. J. Appl. Remote Sens 11 (4): 1-13 (2017)