

Crop distribution extraction based on Sentinel data

Yu Liu^{1,a}, Xiaoping Wang^{1,b*}, Jiaxin Qian^{1,c}

¹ Xi'an University of Science and Technology, Xi'an, 710054 China

² Xi'an Dadi Surveying And Mapping Co., Ltd, Xi'an, 710054 China ¹

Abstract: Remote sensing identification and classification of crops is the use of remote sensing for estimating crop planting area of timely and accurate monitoring of crop growth and plant diseases and insect pests in advance to make the product output to estimate the key and premise of the study using Sentinel-1 and Sentinel-2 satellite, by random forest algorithm, the traditional optical wavelengths and vegetation index The backward scattering field of red edge information and radar information in feature selection and feature classification, including winter wheat summer corn orchard woodland town water and bare land set three controls, such as the first group contains radar time characteristics, the characteristics of the second control group contains red edge long, the third group includes traditional vegetation index for phase characteristics, analyzed the different classification accuracy . The results from the confusion matrix show that the red edge band edge after index and the radar scattering information to join the crop classification accuracy is improved effectively. Sentinel optical and radar satellites with a time resolution of 5-6 days have great potential for crop monitoring research.

1 Introduction

Crop classification has important application value in crop growth, yield estimation and food security. With the advent of free sentinel data with high spatial and temporal resolution, it is possible to combine optical and radar remote sensing images with red edge information and radar information to classify crops. The identification and classification of crops by remote sensing technology is the key and premise to use remote sensing technology to estimate crop planting area, timely and accurately monitor crop growth, diseases and insect pests, and make production estimation in advance ^[1]. With the progress of science and technology, remote sensing technology is also developing continuously. More and more medium and high resolution multi-spectral remote sensing data begin to be widely used in the field of agricultural remote sensing. Sentinel-2 satellite has a high temporal resolution, and the multi-spectral imager mounted on it adds three vegetation red-edge bands, which is not available for traditional resource satellites. The red edge is a unique feature of vegetation. The addition of red edge information can improve the monitoring ability of crops. In addition, the Sentinel-1 satellite is equipped with C-band Synthetic

Aperture Radar, whose time resolution is 6 days, which can monitor the growth of crops all day and all day. C-band SAR can well reflect the canopy information of vegetation, which is beneficial to crop classification.

Due to the limitation of sensor bands, traditional remote sensing image classification can only use the blue, green, red, near-infrared and short-wave infrared of visible light. However, the spectral curves of some vegetation are very similar in these bands and it is difficult to distinguish them. The addition of red-edge information can make up for this defect ^[2-10]. However, at present, there are few studies involving red edge information and radar backscattering information in the classification, mainly due to the limitation of data sources. Red edge information can effectively reflect the growth of vegetation ^[11], and radar can monitor the surface environment all day long. Therefore, red edge information and radar information will be widely used in crop growth monitoring and grain yield estimation, etc., and sentinel data will have great potential in application.

In this context, this study will combine sentinel optical data and radar data, set three control groups of radar, red edge and traditional vegetation index, and use random forest algorithm to find the optimal features involved in

* Corresponding author: ^bemail: 18210210081@xust.stu.edu.cn,
^aemail: 18210063036@xust.stu.edu.cn, ^cemail: 873280820@qq.com

classification and conduct classification research on winter wheat and summer corn in Guanzhong area.

2 Research area and data

2.1 Introduction to the research area

The Guanzhong plain is located in the central and southern

Shaanxi province, the terrain is flat, belongs to the warm temperate zone semi-humid climate zones, rain heat over the same period, very suitable for two seasons a year of planting patterns, the main crops are winter wheat, summer corn and cotton, etc^[12]. The study area of this paper is Pucheng County in Guanzhong Plain, where the main crops are winter wheat, summer corn, cotton, pear and apple trees.

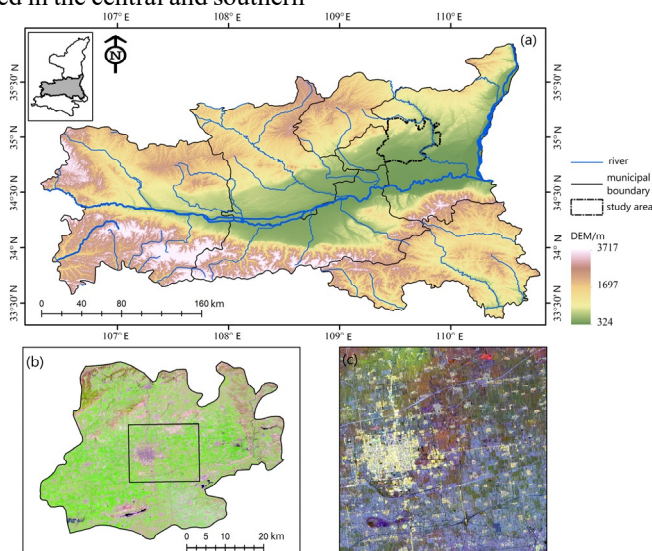


Fig.1 Illustration of the distribution of weather stations

2.2 Remote sensing data used in the research area

This article chooses the remote sensing data from NASA's official website and ESA Copernicus Open Data Center select the Level of 25 Sentinel - 1-1 Level from High resolution image product (High resolution Ground Range Detected, GRDH) and L1C class data from 14 Sentinel-2 episodes. In this paper, the Digital elevation model Satellite data center, which is derived from the Alaska by Advanced earth observation Satellite of surveying and mapping, the spatial resolution of 12.5 m.

This article chooses the Guanzhong plain region, the winter wheat and summer maize growth cycle by China's aim of crop growth and soil humidity value provides data set, as shown in table 3, the whole period of winter wheat growth cycle in October each year to June of next year, after emergence - winter - green - - such as heading to maturation stage, jointing stage and stage of summer maize growth cycle for the June to September every year, Experiencing seedling - jointing - heading - maturity and other stages. Sentinel data covers the main growing seasons of crops.

3 Method and technical process

3.1 The basic principles of random forest classification

Random Forest (RF) is a machine learning algorithm that uses multiple decision tree classifiers to parameterize models for classification, improve prediction accuracy and

prevent overfitting. They assume that different independent predictors are not correct in different regions. Random forest has many features, such as it can process thousands of input variables without deleting variables; It gives an estimate of which variables are important in the classification; With the progress of forest construction, it will produce internal unbiased estimation of generalization error. It has an efficient way to estimate missing data and remains accurate when much of it is lost; It has a method to balance the errors in unbalanced data sets of class groups; Calculate prototypes that provide information about the relationships between variables and classifications; It computes proximity between pairs of cases that can be used for clustering, locating outliers, or (by scaling) giving interesting views of the data^[13-14].

3.2 Feature selection

Spectrum, vegetation index, the traditional red edge after the vegetation index and radar scattering coefficient characteristics, among them, the traditional vegetation index including difference vegetation index (DVI), the ratio vegetation index (RVI), normalized difference vegetation index (NDVI), enhanced vegetation index (EVI), modified soil adjusted vegetation index (MSAVI), normalized water index (NDWI), improved normalized index of water (MNDWI), Red-edge vegetation index includes normalized differential Red-edge vegetation index (NDVIRE1, 2, 3), normalized differential Red-edge index (NDRE1, 2), Red-edge position index (REP), new inverted Red-edge chlorophyll index (IRECI) and Red-edge chlorophyll index (CIRD-edge), a total of 25 characteristics^[15-23]. Radar backscattering coefficient features include VV, VH, the difference between VV and

VH, and the ratio of VV to VH .

3.3 Technical process

The Sentinel-1 data is preprocessed by SARscape software, including the following steps: data import and Mosaic, geometric registration, multi-temporal scatter filtering, radiometric calibration and geographic coding. The backscattering coefficient images of VV and VH polarization with a spatial resolution of 10m were obtained and tailored according to the study area.

The Sentinel-2 data were obtained by using the SEN2COR toolbox of ESA's Snap software to obtain the L2A level products, and each band was resampled to the 10m resolution by the nearest neighbor method, and then cropped. Finally, the multitemporal surface reflectance image with spatial resolution of 10m in the study area is obtained.

The geometric registration of radar and optical remote sensing is very important in the cooperative research of radar and optical remote sensing. Due to the difficulty in obtaining accurate control points on the radar Image, the Envi5.5 Image Registration Workflow module is used to perform geometric Registration of the radar Image based on the optical Image to limit the error to 0.5 pixels.

The Sentinel-1 multitemporal radar image features and Sentinel-2 vegetation index features were calculated respectively.

Samples were evenly selected from the plots, woodlands, orchards, water bodies and buildings in the winter wheat-summer maize rotation mode in the study area, and 15 sample points were selected for each category, and the mean values of each feature were calculated. Due to the influence of cloud, rain and aerosol, the Sentinel-1 radar backscattering coefficient images were reconstructed to analyze the characteristic curves of different vegetation at different scales, different time phases and different features.

3.4 Using RF to determine the importance of features

Set three control group, one of the first group contains Sentinel - 1 for phase radar characteristics after the scattering coefficient, the second control include Sentinel - 2 long red edge features (three red edge band and its corresponding edge vegetation index), the third in the control group contains Sentinel - 2 long phase of traditional optical characteristics (blue, green, red, near infrared and short-wave infrared wavelengths and their corresponding vegetation index), will be three groups respectively characteristics of band combination, Through the training samples mentioned above, the importance of optical and radar characteristic variables and time-phase variables for classification was determined by using the random forest algorithm. The top 30 features in each group of importance are combined in different bands to carry out the next step of principal component analysis.

3.5 Principal component analysis of optimal feature image

Principal Component Analysis (PCA) [24] is an image processing method to remove redundant information of multi-band images, and the information of each band after Principal Component transformation is not correlated. After combining the optimal optical features and radar features selected in the previous step, the optimal optical features and radar features selected in the previous step were normalized with a confidence of 1%, and then principal component analysis was carried out, and the first six principal components of the two were retained.

The random forest algorithm was used to classify winter wheat and summer corn. The training samples were randomly sampled, with 30% as training samples and the remaining 70% as verification samples. The optimal feature images through independent principal component analysis were classified by random forest algorithm. The accuracy of the three control groups was evaluated by calculating the confusion matrix and Kappa coefficient of the classification result graph.

4 Results and analysis

4.1 Random forest classification

The RF algorithm was used to estimate the feature importance of the three characteristic variable groups. The table shows the top 30 characteristic variables in the importance ranking of each control group, in which the orange marker represents the growth cycle of winter wheat, and the green marker represents the growth cycle of summer corn. Among them, the optimal time phase of group 1 (only containing radar information) mainly concentrated in late February 2018 to late May 2018 (this stage is the winter wheat greening to maturity stage) and September 20, 2018 (this stage is the summer corn maturity stage). Group 2 contains only red edge information of the optimal phase concentration in mid-February to early April, 2018 and 2018 in late August and late September, the optimal red edge characteristics of vegetation index is NDred1, NDVIred1 and CI_{red edge}. REP index cannot effectively distinguish between crops, mainly reflects the vegetation red edge within the scope of the position of the maximum reflectivity spectrum slope, and can be seen from the diagram, no obvious difference of different vegetation red edge position, therefore, its importance is low; The optimal time phase of control group 3 (which only contained a single band and traditional vegetation index information) was similar to that of control group 2, and the optimal vegetation index was MNDWI, NDVI, NDWI and MSAVI, mainly because MNDWI and NDWI could effectively reflect the water content information of crop canopy.

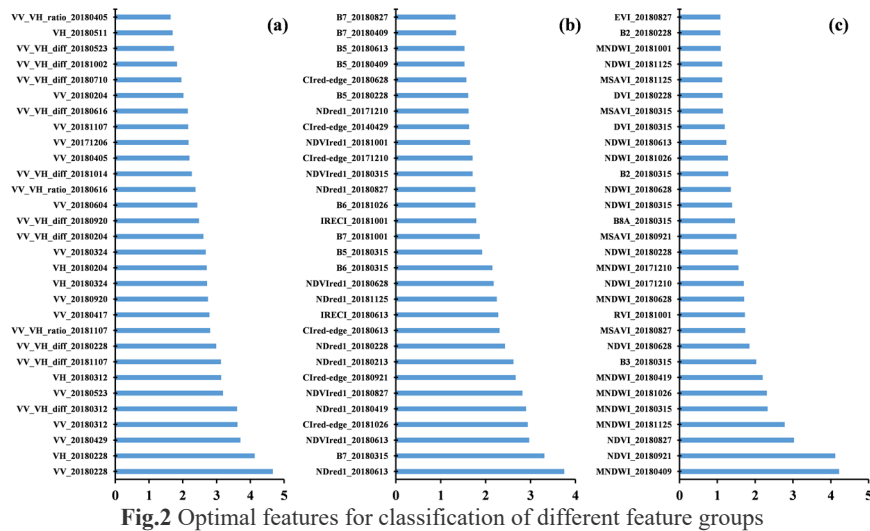


Fig.2 Optimal features for classification of different feature groups

4.2 Result analysis and precision evaluation of random forest classification

The control of different random forest classification, classification results as shown in figure 3, among them, the red edge features combination traditional vegetation index group classification results are basically identical, radar extracting feature set of flat terrain of the study area south of summer corn and fruit orchards and other control group, but the error to the northeast of the study area classified as forest land, orchard this may be due to the influence of the height in the northeast. Another set of sample point was used to calculate the confusion matrix of four classification results, and the differences of multi-temporal classification results with different features were compared. The classification accuracy of the traditional vegetation feature group is the highest, followed by the red

edge feature group and the radar feature group. In the radar feature group, the most serious misclassification is winter wheat summer corn and orchard, which may be because the difference of radar backscattering between them is not obvious in part of the growth cycle. In other part, winter wheat summer corn is misclassified into buildings, which may be caused by the phenomenon of mixed pixels. In the red-edge feature group, construction land and orchard are the most seriously misclassified, mainly because the red-edge feature group contains only three red-edge information and one near-infrared band, and lacks the most important short-wave infrared band to distinguish construction land. The classification accuracy of the three groups reached satisfactory results, and the information of winter wheat-summer maize rotation crops, orchards and woodlands in the study area could be extracted effectively. The accuracy of different feature classification groups is shown in the table.

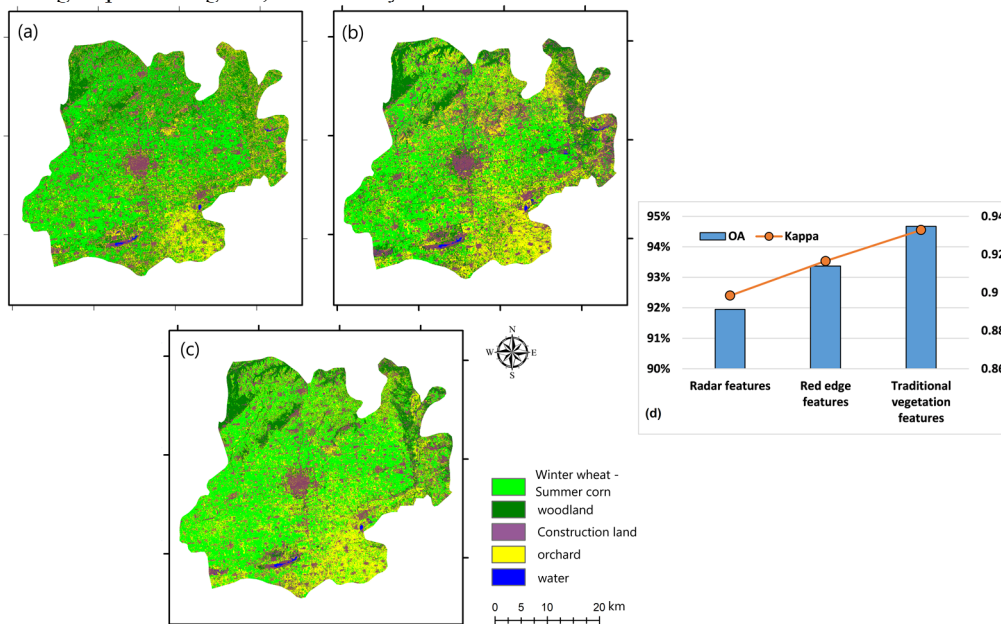


Fig. 3 Results of RF classification ((a)radar feature group, (b) red edge feature group, (c)traditional vegetation feature group (d) Classification accuracy of different feature groups)

5 Conclusion

On April 19, 2018, and this study Sentinel on September 21, 2018-2 images as winter wheat and summer corn, selection of the training sample images, setting up three

control group, using random forest algorithm to find optimal characteristics of each group involved in classification and band combination after component analysis, then the use of random forests for classification, finally analysis and accuracy evaluation. The main conclusions are as follows:

Through the random forest algorithm, the best features participating in the classification of winter wheat and summer corn were selected. Principal component analysis of the selected features could better highlight the differences of construction land, soil, crops, woodland and shadow, and enhance the information of vegetation.

Compared with traditional optical features, multitemporal radar features have certain advantages. Synthetic aperture radar (SAR) is not affected by cloud and rain and can obtain continuous crop phenology information. Although the Sentinel-2 binary satellite network can improve the temporal resolution up to five days, it is difficult to guarantee continuous and stable images during critical periods of crop growth. Combining optical and radar images with different spatial and temporal resolutions can give full play to the advantages of multivariate data, which is more conducive to the monitoring of crop growth.

Control group both can effectively extract the different features, among them, the traditional characteristics of vegetation index set is highest in the crops in the study of classification accuracy, the overall accuracy of 94.67%, red edge index set of classification accuracy is slightly lower than the traditional characteristics of vegetation index set, the precision of the radar feature set of the lowest, but accuracy is over 90%, the instructions for the red edge, radar characteristics can effectively extract the Guanzhong area of crops.

References :

1. Ding Y W. Extraction of winter wheat planting area based on MODIS image data in northern Anhui Province [D]. Anhui University,2019.
2. Zhang X W, Chen Y S, Meng Q, Wang X. Crop phenology extraction based on time series MODIS NDVI [J]. Chinese Agricultural Science Bulletin,2018,34(20):158-164.
3. Xu Q Y, Yang G J, Long H L, Wang C C, Li X C ,Huang D C . Crop planting identification based on MODISNDVI multi-year time series data [J]. Transactions of the Chinese Society of Agricultural Engineering,2014,30(11):134-144.
4. Jin Z T, Li W G, Jing Y S. Research on the Appropriate Scale of Winter Wheat Planting Area Extraction Based on Image Fusion [J]. Jiangsu Journal of Agricultural Sciences,2015,31(06):1312-1317.
5. Ji X S, Li X, Wan Z F, Yao X Zhu Y, Cheng T. Classification of cotton and jujube trees in Alar city of Xinjiang based on high spatial resolution satellite image [J]. Agricultural Sciences in China,2019,52(06):997-1008.
6. Yang H W, Fang J Y, Zhao D. Feature selection for fine classification of crops by remote sensing based on improved separation threshold method [J]. Journal of Applied Sciences,2019,37(01):51-63.
7. Liu Z H, Liu L, Guo H, Cheng P. The extraction of spring wheat based on GF1-NDVI time series image was studied [J]. Beijing surveying and mapping,2018,32(06):643-646.
8. Zhao Z, Yan L. 16 - band image spectral feature classification of Worldview satellite [J]. Remote sensing information,2019,34(01):36-43.
9. Bi K Y, Niu Z, Huang N, Kang J, Pei J. Vegetation identification based on Sentinel-2a time series data and object-oriented decision tree method [J]. Geography and Geographic Information Science,2017,33(05):16-20+27+127.
10. Wang L J, Guo Y, He J, Wang L M, Zhang X W, Liu T. A Sentinel-2a Image Crop Extraction Method Based on Decision Tree and SVM [J]. Transactions of the Chinese Society for Agricultural Machinery,2018,49(09):146-153.
11. Fang C Y, Wang L, Xu H Q. Comparative study of red edge index of different vegetation in health discrimination of urban grassland [J]. Journal of Geo-information Science,2017,19(10):1382-1392.
12. Zhang P, Ma X Y, Zhang L, Xing X G. Land use classification in Guanzhong region based on different vegetation indices and classification regression trees [J]. Soil and Water Conservation Research,2018,25(03):310-316.
13. Belgiu M, Drăguț L. Random forest in remote sensing: A review of applications and future directions[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2016, 114: 24-31.
14. Gislason P O, Benediktsson J A, Sveinsson J R. Random forest classification of multisource remote sensing and geographic data[C]. 2004 IEEE International Geoscience and Remote Sensing Symposium, 2004, 2: 1049-1052.
15. Tucker C J. Red and photographic infrared linear combinations for monitoring vegetation[J]. Remote Sensing and Environment, 1979, 8(2): 127-150.
16. Major D J, Frédéric Baret, Guyot G. A ratio vegetation index adjusted for soil brightness[J]. International Journal of Remote Sensing, 1990, 11(5): 727-740.
17. Schell J A, Deering D. Monitoring vegetation systems in the Great Plains with ERTS[J]. NASA Special Publication, 1973, 351: 309.
18. Gitelson A A, Merzlyak M N. Remote sensing of chlorophyll concentration in higher plant leaves[J]. Advances in Space Research, 1998, 22(5): 689-692.
19. Qi J, Chehbouni A, Huete A R, et al. A modified soil adjusted vegetation index[J]. Remote sensing of environment, 1994, 48(2): 119-126.
20. Christos E and Alexandre N. Red-Edge Normalized Difference Vegetation Index (NDVI705) from Sentinel-2 imagery to assess post-fire regeneration [J]. Remote Sensing Applications: Society and

- Environment, 2020, 17: 100283.
21. Frampton W J, Dash J, Watmough G, et al. Evaluating the capabilities of Sentinel-2 for quantitative estimation of biophysical variables in vegetation[J]. ISPRS journal of photogrammetry and remote sensing, 2013, 82: 83-92.
 22. Haboudane D, Miller J R, Pattey E, et al. Hyperspectral vegetation indices and novel algorithms for predicting green LAI of crop canopies: Modeling and validation in the context of precision agriculture[J]. Remote Sensing of Environment, 2004, 90(3): 337-352.
 23. Gao B C. NDWI-A normalized difference water index for remote sensing of vegetation liquid water from space[J]. Remote sensing of environment, 1996, 58(3): 257-266.
 24. Wang Z M, Du B, Zhang L P, et al. Hyperspectral image classification based on texture feature and morphological feature fusion [J]. Journal of photons, 2014, 43(08): 122-129.