The time-series GF-1 WFV data monitoring of sugarcane using a Random Forest Algorithm in South China

Chen Chen [1], LinJiang Lou [1*], Tao Cheng[2], XinYuan Gao [1], Yu Liu [1]

[1] Land Satellite Remote Sensing Application Center, MNR, Beijing 100048, China-410208288@qq.com
[2] National Geomatics Center of China, Beijing 100830, China-

Keywords: Sugarcane, Remote Sensing, time series, Random Forest, GF-1.

Abstract

There is large distribution of sugarcane growth in south China which is play an important role of sugar industry. Remote sensing technology is used in sugarcane monitoring for large areas. However, the optical satellite data coverage is influenced by the rainy weather especially in the grand growth period of sugarcane. GF-1 WFV has widely swath 800km and short revisit time which is ideal data for this study area. In this paper, the random forest model was chosen to get a precise classification result of sugarcane based on time-series band value and 5 spectral indexes image is 89.73% and the Kappa coefficient is 0.65 which is satisfied the overall extraction of sugarcane for large area is the southern China. Furthermore, the decision tree classification was chosen as a comparative experience research.

1. Introduction

Sugarcane is an important sugar crop as well as an important bioenergy crop, which is widely planted in southern regions such as Guangxi, Yunnan, Guangdong, and Hainan in China. Monitoring sugarcane cultivation is extremely important for the estimation of sugarcane yield and the formulation of policies and production plans related to the development of the sugar industry by relevant departments(Chen et al., 2022; Linjiang et al., 2020). Remote sensing technology has been widely used in the field of agriculture, and results have been achieved in monitoring the growth of crops on a large scale(Abdel - Rahman and Ahmed, 2008; Aguiar et al., 2011; El Hajj et al., 2009; Gers; Mulianga et al., 2013; Rudorff et al., 2009; Xavier et al., 2006).

In recent years, with the development of satellite technology, the long temporal satellite images are suitable with high resolution for agriculture monitoring. Timely spatial and temporal satellites image data are widely used in sugarcane planting and growth monitoring. Mulianga forecasted sugarcane yeild using the time series the Moderate Resolution Imaging Spectro Radiometer(MODIS) data over nine years. (Mulianga et al., 2013) Landsat Thematic Mapper(TM) and Landsat Enhanced TM Plus(ETM+) time-series were used for analysis sugarcane in Brazil.(Vieira et al., 2012)

He Yajuan et al. performed sugarcane leaf area index inversion and yield estimation based on SPOT 4 and SPOT 5 data. A significant positive correlation between sugarcane Leaf Area Index (LAI) and Normalized Difference Vegetation Index (NDVI) at different fertility stages was conclude(He et al., 2013). Zeng Zhikang et al. carried out a study on crop cultivation information extraction based on China high temporal and spatial resolution satellite images, and comprehensively applied ZY-3, HJ-1, and GF-1 thematic and spectral information to obtain the time series spectral features during the critical period of crop growth. The overall classification

accuracy is 86.80%.(Zhi-Kang et al., 2017). Qin Zelin et al. used the 2 m resolution GF-1 image as a data source and extracted the crop information of sugarcane, rice and banana using the spectral, normalized vegetation index, gray value, and other features of the image, using an object-oriented classification method (Ze-Lin et al., 2017).

Sugarcane information extraction based on single time-range remote sensing images is easy to be confused with non-sugarcane crops, garden and grass cover, and the accuracy of the extraction results is low. However, based on time-series remote sensing images, it can excavate the seasonal period of sugarcane growth based on the growth pattern of sugarcane, and then differentiate it from other crop types to get the extraction of the planting range of sugarcane, which is a scientific and effective method.

The swath wide of high-resolution satellite is less than 20km such as Worldview, Pleiades and GeoEye. The distribution of sugarcane is in southern China with cloudy weather. High-resolution data is not suited the time-series monitoring in China. GF-1 was successfully launched in 2013 with two 2m resolution panchromatic/8m resolution multispectral cameras and four 16m resolution wide-swath cameras which is 800 km ideal data for time-series coverage in sugarcane growth area of China.

In this study, based on the GF-1 16m wide-swath cameras the range of sugarcane cultivation is extracted as an example to explore feasible technical methods to provide methodological references as well as other crops monitoring.

2. Study Area and Data

Part of the southern China was selected as the study area, covering an area of 114.50 km2, which is located within the national sugarcane production base. The GF-1 WFV time-series images of the study area were acquired, with five-time phases from April to October in 2015, respectively, containing four

---

* Corresponding author

bands of R, G, B, and NIR, with a resolution of 16 m. GF-2 with high-resolution is chosen as the sample collection.

The Rational Polynomial Coefficient (RPC) model (Liu et al., 2015) and the Digital Elevation Model (DEM) data are used to orthorectify the remote sensing images of each time-phase.

Parameter	GF1 Wide-swath camera	
Spectral band	Panchromatic image	——
	Multispectral Image	0.45 ~ 0.52 μm
		0.52 ~ 0.59 μm
		0.63 ~ 0.69 μm
		0.77 ~ 0.89 μm
GSD	Panchromatic image	——
	Multispectral Image	16 m
Swath width	800 km（Combining four cameras' observation）	
Site re-visit period without side swing	4 days	

Table 1. Parameter of GF1 satellite data

Parameter	GF-2	
Spectral Range	Panchromatic image	0.45～0.90 μm
	Multispectral Images	0.45～0.52 μm
		0.52～0.59 μm
		0.63～0.69 μm
		0.77～0.89 μm
Resolution	Panchromatic image	0.8 m
	Multispectral Images	3.2 m
Swath width	45 km	
Re-visit period	69 days	

Table 2. Parameter of GF2 satellite data
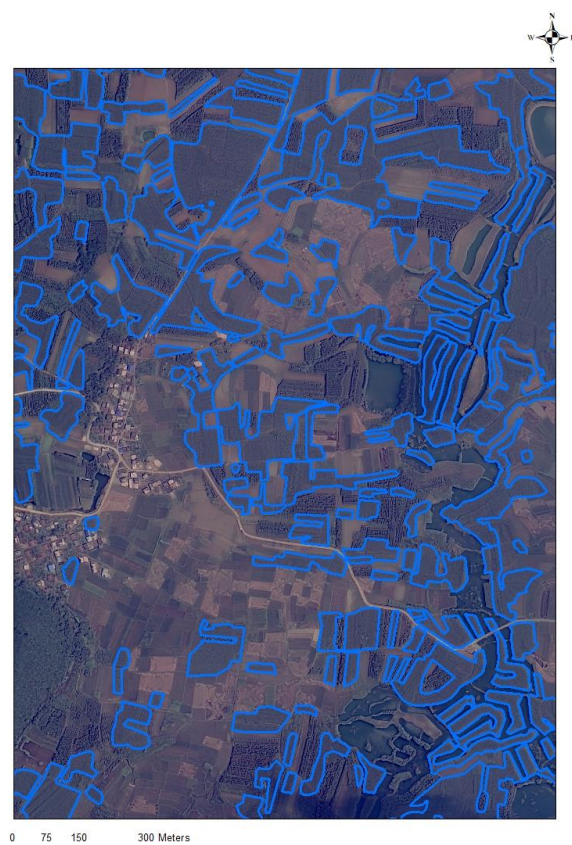

Figure 1. GF-1 image in 18th July,2015.

Figure 2. GF-2 image with sample data.

Data	Time
GF1 WFV	2015.04.03
	2015.04.19
	2015.07.18
	2015.08.24
	2015.10.21

Table 3. Acquire time of GF1 satellite data

3. Method

3.1 Date Processing

On the base of analysing and summarizing the existing crop monitoring methods, we take the extraction of sugarcane planting range as an example. In key growing period of sugarcane, the maximum difference index between sugarcane

and other crop types is used to explore the applicable methods with strong operability which is the foundation for large-scale popularization and application.

Firstly, the time series remote sensing images of GF1 and high-resolution remote sensing images of the same period was orthorectified to the World Geodetic Survey 1984(WGS84) datum. Sugarcane, non-sugarcane crops, garden land, forest and grass cover, bare ground, water area, buildings (structures) are collected on the high resolution remote sensing images as the samples for information extraction and analysis(Xulong et al., 2006). Finally, the Random Forest Model and the Decision Tree Model are chosen in this research.
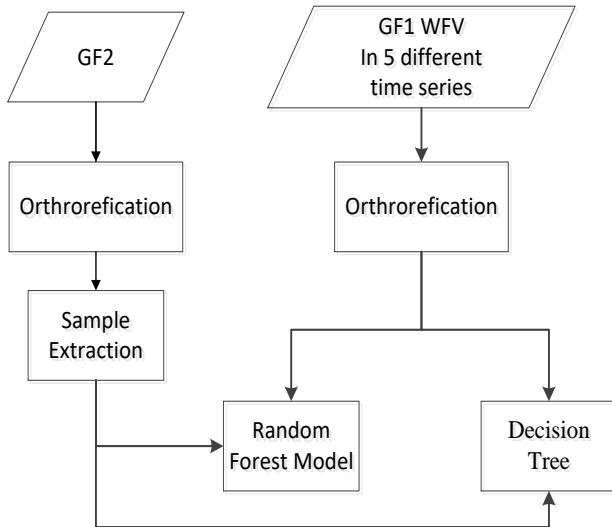

Figure 3. The flow chat of this study.

3.2 NDVI characterization of samples

Spatial overlay analysis of the samples in surface cover types with the NDVI raster results was carried out, and raster statistics were performed to statistically determine the minimum, maximum, mean, and variance of the NDVI within different tapes of samples, obtain the NDVI of seven types samples of surface cover types. The statistical results are shown in Table 4.

Type \ Time	2015 0403	2015 0419	2015 0718	2015 0824	2015 1021
Sugarcane	-0.04	-0.05	0.36	0.41	0.39
Non-sugarcane crops	0.03	-0.03	0.25	0.46	0.41
Garden land	0.13	0.15	0.44	0.46	0.46
Forest and grass cover	0.20	0.20	0.38	0.40	0.42
Bare ground	-0.07	-0.07	-0.02	-0.03	0.01
Water areas	-0.13	-0.18	-0.07	-0.12	-0.05
Buildings (structures)	-0.05	-0.08	0.02	0.02	0.06

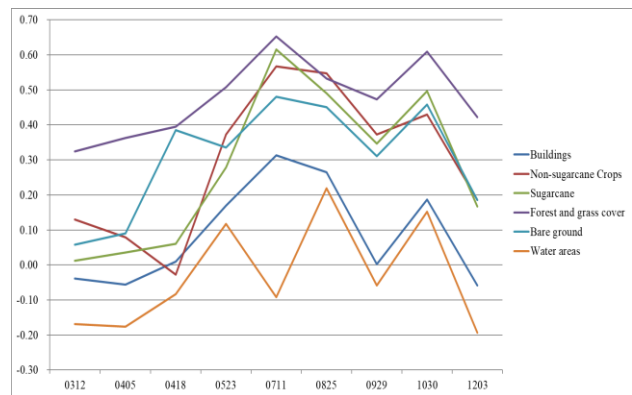Table 4. The characteristic statistical value of sample's average NDVI

Figure 4. The curve diagram of characteristic statistical value of sample's average NDVI

3.3 Random Forest Model

Random Forest (RF) is an ensemble learning method with high classification accuracy which can determined variable importance. (Breiman et al., 2001) Tree type classifiers used for regression which are combined for the predicted result. The original training data are randomly split into large numbers of nodes which is determined by the user-defined input variables. The accuracy of the classification is depended on the suitable split for a node.

RF showed the good performance on agriculture classification using satellite images in previous research which can estimate the relative importance of the input variables with large dataset and input variables.(Xulong et al., 2006; Ze-Lin et al., 2017)

In this research, the Random Forest Model was used for sugarcane classification which is suited to a precise result with data in high dimension and outlier detection(Gislason et al., 2006). The variables are given in table with four bands (Red, Green, Blue and Near Red) and 5 spectral indexes (NDVI, NDWI, SAVI, RVI and NIRV).

Variables	Description
Red	0.45-0.52 μm
Green	0.52-0.59 μm
Blue	0.63-0.69 μm
Near Infrared (NIR)	0.77-0.89 μm
NDVI	$NDVI = \dfrac{NIR - Red}{NIR + Red}$
NDWI	$NDWI = \dfrac{Green - NIR}{Green + NIR}$
SAVI	$SAVI = 1.5 * \dfrac{NIR - Red}{NIR + Red + 0.5}$
RVI	$RVI = \dfrac{NIR}{Red}$
NIRV	$NIRV = RED * \dfrac{NIR - Red}{NIR + Red}$

Table 5. Variables of the random forest

Based on the random sampling strategy, 80% of the samples are used for the training of the random forest classification model, and the maximum depth of the sub-decision tree is set to 10.

The number of sub-decision trees of the classification model is set to 300. The regression model is constructed based on the above preferred features with 80% of the samples. 20% of the samples are used to verify the classification accuracy.

4. Result and discussion

4.1 Random Forest Model Result

Using the existing results, the accuracy of the extraction results was evaluated using the confusion matrix and Kappa coefficient method, and the results are shown in Table 6.
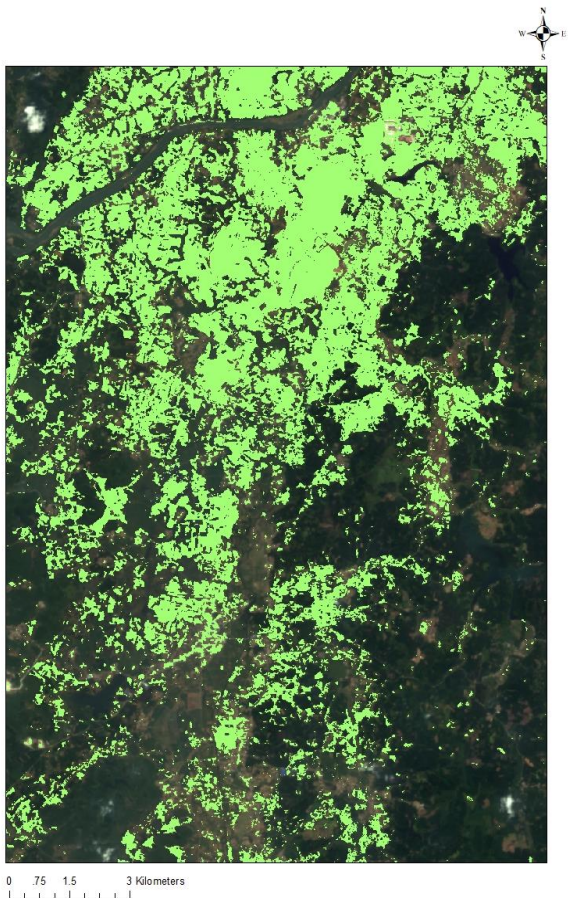

Figure 5. The extracting result of sugarcane planting range.

Land type	User's accuracy (%)	Producer's accuracy (%)
sugarcane	95.88	91.73
Other	64.41	79.17
Overall Accuracy	89.73	
Kappa Coefficient	0.65	

Table 6. The accuracy evaluation of RFM.

The overall accuracy was calculated to be 89.73% with a Kappa coefficient of 0.65. From the results of spatial overlay analysis and accuracy evaluation, the overall accuracy of the extraction results is high, which can meet the needs of mastering the overall situation of sugarcane planting in the large range.
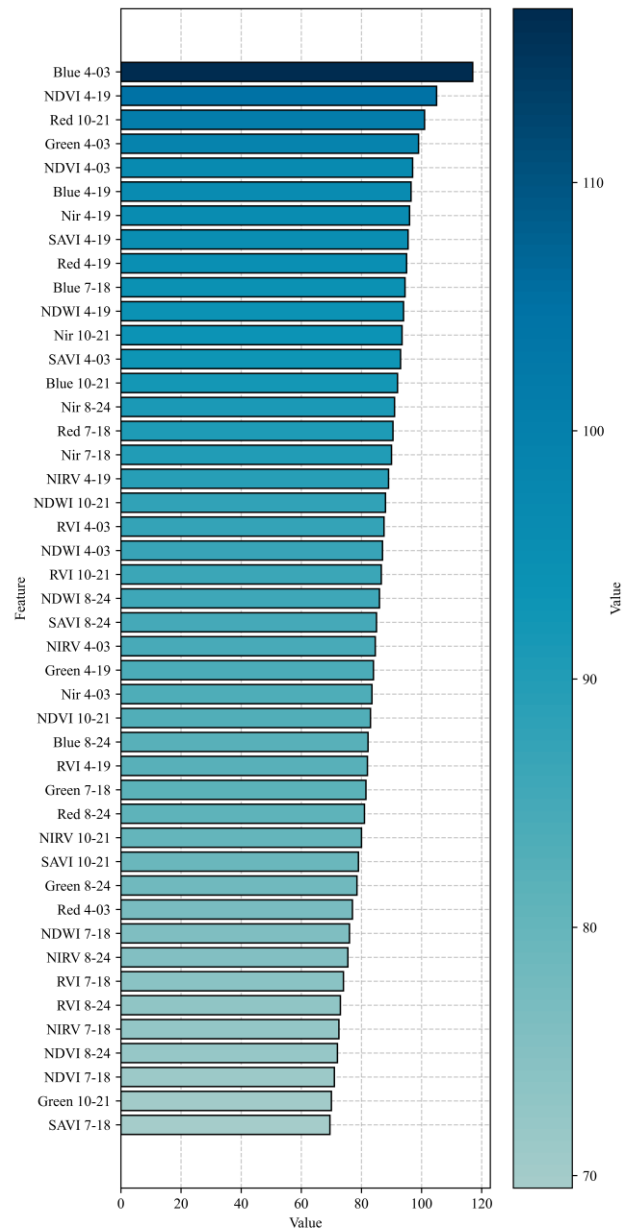

Figure 6. The important of variables in RFM.

The important of the variables in the random forest model was calculated. The blue band in 3rd of April, the NDVI in April19th and the red in 21th of October were the most important. The comparative important variables were in April which is the germination period of sugarcane in the research area.

4.2 Compare With the Decision Tree Classification

In addition, one of the traditional methods of large area classification, the decision tree classification was chosen with the same sample as a comparative experience method in this research shown in Table 7. The overall accuracy is 85.64% with a Kappa coefficient of 0.65. Compared with the traditional method, the RF model shows better improvement in the overall accuracy as an over 2% raise than the decision tree.
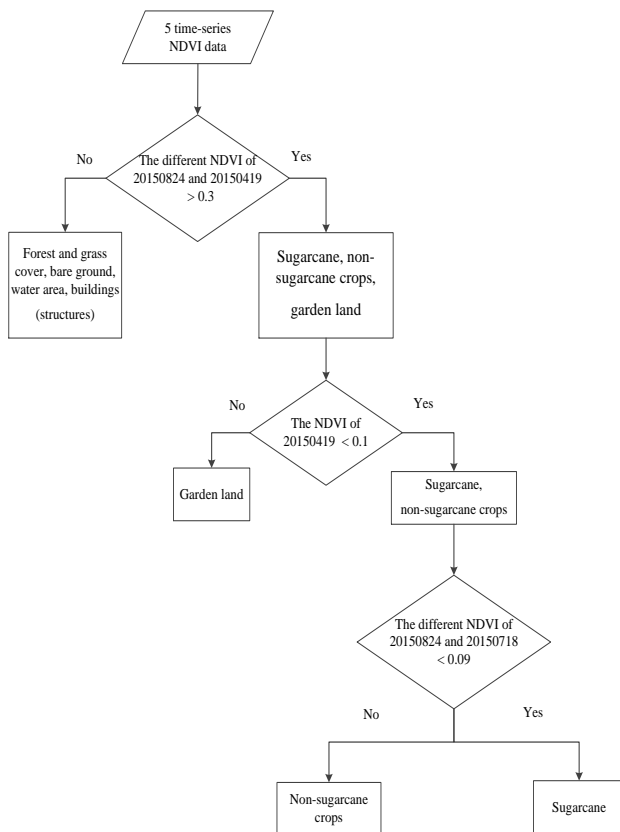
Figure 7. The flow chart of decision tree classification

Land type	User's accuracy (%)	Producer's accuracy (%)
sugarcane	0.85	0.69
Other	0.88	0.95
Overall Accuracy	87.5	
Kappa Coefficient	0.69	

Table 7. The accuracy evaluation of decision tree classification.

5. Conclusion

The combination of high spatial resolution and high temporal resolution remote sensing images to carry out crop extraction was great potential for application. It can give full play to the advantages of high temporal resolution remote sensing images to carry out time-series remote sensing monitoring and master the key seasonal period of crop growth as the key method of crop extraction; and it can also utilize high spatial resolution remote sensing images to accurately interpret crop categories and verify the accuracy of crop extraction. In this study, GF-1 WFV time-series remote sensing images combined with high-resolution remote sensing images is utilized which can satisfy the needs of crop extraction. The Random Forest model can give precise sugarcane classification in large area compared with the RF model. The satellite data in April are the important feature of the Random Forest model for sugarcane extraction in this paper.

References

Abdel- Rahman, E.M., Ahmed, F.B., 2008. The application of remote sensing techniques to sugarcane (Saccharum spp. hybrid) production: a review of the literature. International Journal of Remote Sensing 29, 3753-3767.

Aguiar, D.A., Rudorff, B.F.T., Silva, W.F., Adami, M., Mello, M.P., 2011. Remote Sensing Images in Support of Environmental Protocol: Monitoring the Sugarcane Harvest in São Paulo State, Brazil. Remote Sensing 3, 2682.

Breiman, L., Breiman, L., Cutler, R.A., 2001. Random Forests Machine Learning. Journal of Clinical Microbiology 2, 199-228.

Chen, C., Lou, L., Gao, X., Liu, Y., 2022. Comparative Analysis of Chinese High-Resolution Satellite Data for Sugarcane Classification Based on U-Net Model, in: Wang, L., Wu, Y., Gong, J. (Eds.), Proceedings of the 7th China High Resolution Earth Observation Conference (CHREOC 2020). Springer Nature Singapore, Singapore, pp. 185-192.

El Hajj, M., Bégué, A., Guillaume, S., Martiné, J.-F., 2009. Integrating SPOT-5 time series, crop growth modeling and expert knowledge for monitoring agricultural practices — The case of sugarcane harvest on Reunion Island. Remote Sensing of Environment 113, 2052-2061.

Gers, C.J., RELATING REMOTELY SENSED MULTI-TEMPORAL LANDSAT 7 ETM+ IMAGERY TO SUGARCANE CHARACTERISTICS.

Gislason, P.O., Benediktsson, J.A., Sveinsson, J.R., 2006. Random Forests for land cover classification. Pattern Recognition Letters 27, 294-300.

He, Y., Pan, X., Pei, Z., Ma, S., Shang, J., 2013. Estimation of LAI and Yield of Sugarcane Based on SPOT Remote Sensing Data. Nongye Jixie Xuebao/Transactions of the Chinese Society of Agricultural Machinery 44, 226-231.

Linjiang, L., Chen, C., Min, H., Xinyuan, G., Kun, L., Minmin, L., 2020. Sugarcane Planting Area Classification, Extraction and Accuracy Comparison Based on Chinese High-Resolution Remote Sensing Satellite Data: A Case Study of Ningming Sugarcane Demonstration Area, Chongzuo City, Guangxi. IOP Conference Series: Earth and Environmental Science 502, 012041.

Liu, J., Wang, L., Yang, L., Shao, J., Teng, F., Yang, F., Fu, C., 2015. Geometric correction of GF-1 satellite images based on block adjustment of rational polynomial model. Transactions of the Chinese Society of Agricultural Engineering 31, 146-154.

Mulianga, B., Bégué, A., Simoes, M., Todoroff, P., 2013. Forecasting Regional Sugarcane Yield Based on Time Integral and Spatial Aggregation of MODIS NDVI. Remote Sensing 5, 2184.

Rudorff, B.F.T., Adami, M., Aguiar, D.A.d., Gusso, A., Silva, W.F.d., Freitas, R.M.d., 2009. Temporal series of EVI/MODIS to identify land converted to sugarcane, 2009 IEEE International Geoscience and Remote Sensing Symposium, pp. IV-252-IV-255.

Vieira, M.A., Formaggio, A.R., Rennó, C.D., Atzberger, C., Aguiar, D.A., Mello, M.P., 2012. Object Based Image Analysis

and Data Mining applied to a remotely sensed Landsat time-series to map sugarcane over large areas. Remote Sensing of Environment 123, 553-562.

Xavier, A.C., Rudorff, B.F.T., Shimabukuro, Y.E., Berka, L.M.S., Moreira, M.A., 2006. Multi‐temporal analysis of MODIS data to classify sugarcane crop. International Journal of Remote Sensing 27, 755-768.

Xulong, L., Chunyang, H.E., Yaozhong, P., Mingchuan, Y., Jinshui, Z., 2006. Accuracy Assessment of Thematic Classification Based on Point and Cluster Sample. National Remote Sensing Bulletin 10, 366-372.

Ze-Lin, Q., Guo-Xue, X., Yu-Xiang, L.I., Zong-Bao, L., Qiu-Qun, S.U., Fu-Qian, X., Jia-Mei, Z., Xiu-Long, Z., 2017. Application of multi-temporal GF-1 image in extraction of staple crops in hilly region. Journal of Southern Agriculture.

Zhi-Kang, Z., Ze-Lin, Q., Qi-Ting, H., Guo-Xue, X., Yu-Xiang, L.I., Jian-Cheng, L., Zhou, L.U., 2017. Cropping Information Acquisition from National High Temporal-and-spatial Resolution Satellite Imaging System. Fujian Journal of Agricultural Sciences.