

# Extraction of Refined Crop Type Over Agriculture Region of Heilongjiang

Yanmin Shuai<sup>1,2,3,4</sup>, Hao Wu<sup>1,\*</sup>, Congying Shao<sup>1</sup>, and Kexin Ning<sup>1</sup>

<sup>1</sup> Liaoning Technical University, College of Surveying and Mapping and Geographic Science, 123000 Fuxin, China

<sup>2</sup> Xinjiang Institute of Ecology and Geography, CAS, 830011 Urumqi, China

<sup>3</sup> Research Center for Ecology and Environment of Central Asia, CAS, 830011 Urumqi, China

<sup>4</sup> University of Chinese Academy of Sciences, 100049 Beijing, China

**Abstract.** The spatial distribution of fine crop types at regional scale is required by numerous research communities. The traditional approach with limited time-phases is hard to capture the signatures presented within different growth period of various crop types. With the improvement of understanding on the phenology feature of major crops and the accumulation of satellite-based observations, there is a chance to distinguish detailed crop clusters with elevated accuracy. In this work, we investigated the phenological feature of four dominant crops (soybean, wheat, maize, and paddy) in multi-spectrum space through ~800 representative crop samples within typical agriculture regions of Heilongjiang using MODIS daily surface reflectance product covering related growth period of 2005-2018. Features with the high degree of separation among land cover clusters are screened out to construct the model in identifying typical crop types in terms of weighted temporal features and classification scheme, which is applied to extract the crop map of Heilongjiang province in 2019. The results show higher accuracy achieved over main agriculture region of soybean, wheat, maize, and paddy, and reduced accuracy over field of wheat or other mixed crops at MODIS pixel scale. Our validation shows the overall accuracy of 0.9816 and kappa coefficient of 0.9702 through the comparison with ~3000 random selected ground sites. The preliminary application of the presented approach performs well via the capture on valid phenology features of major crops within dominant agriculture region of Heilongjiang, with the potential to serve the extraction of fine crop types over wide agriculture regions.

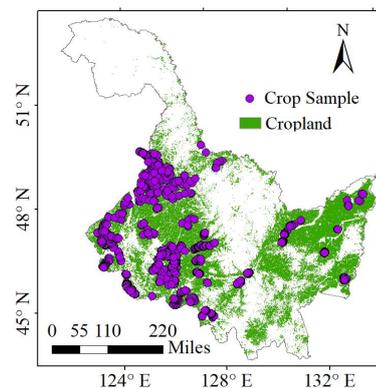
## 1 Introduction

Accurate distribution of crop types is one of important datasets for food security, agricultural management and policy development, environmental sustainability and land surface ecosystem monitoring [1-2]. With the accumulation of satellite-based observation in medium or high resolution, there is a chance to develop effective methods in retrieving crop information and distinguishing various crop types, thus provide the potential refined crop map for user communities [3]. Numerous studies demonstrated the importance of multi-temporal information in crop identification. Supervised methods have been used most widely in multi-temporal classification, like decision tree [3], random forest [4], support vector machine (SVM) [5] and other nonparametric methods [6]. These methods can process multiple time-phase images to identify various land cover clusters. However, there is still a lack of the investigation on phenology features of different crops, especially in the satellite-based thematic retrieval.

Time series classification method has great potential in crop identification, due to the ability to capture unique growth rhythm and intrinsic phenological signatures [7].

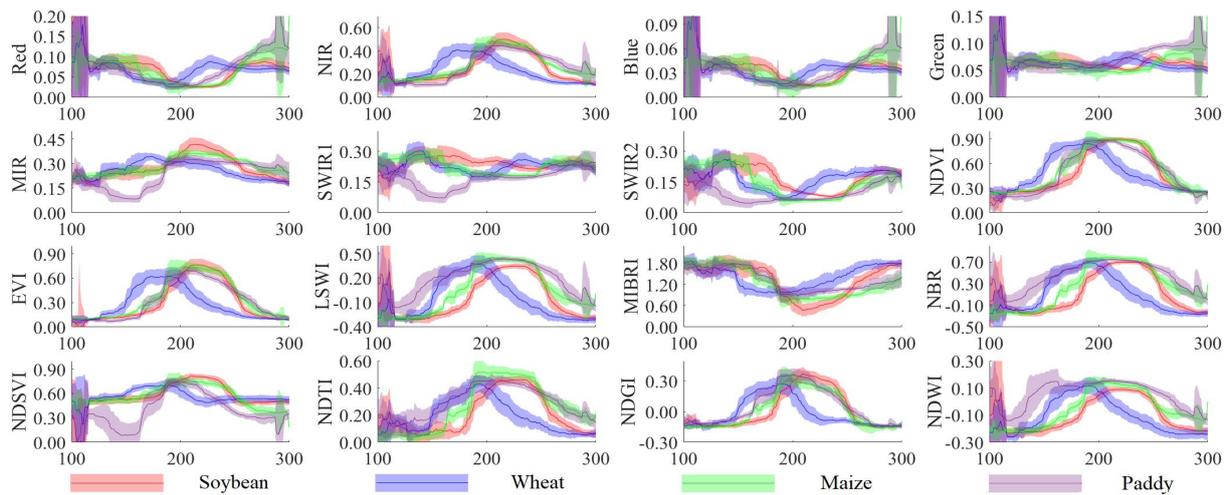
In this study, we used MODIS data to extract a variety of time series features, and established a time series feature classification model to improve the accuracy of crop identification.

## 2 Study Area and Data



**Fig. 1.** The study area and crop samples in HLJ from 2005 to 2018 (the cropland is from GlobeLand30 in 2020 [8]).

\* Corresponding author: 471920558@stu.lntu.edu.cn



**Fig. 2.** The spectral reflectance and vegetation index (VI) of crops during the growth period of DOY 100-300 (x-axis), which the solid lines represent the mean value of features, and the translucent zones represent the standard deviation of features.

Heilongjiang (HLJ) is the northernmost province of China, plays the significant role in the supply of grains and agriculture management. The total sown area is 147,700 km<sup>2</sup>, which is mainly distributed in Songnen Plain (SNP) and Sanjiang Plain (SJP). Now, percentages of plant area for soybean, maize and paddy respectively have occupied up to 40% ,13% and 12% of the total grains' amount within China. Wheat was widely planted in northern HLJ before 2010, but the current planting area has been reduced to less than 1% of the total planting area of the province [9-10]. HLJ has a continental monsoon climate. The accumulated temperature (>10°C) is 1800-2800 °C, and the frost-free period is between 100-150 days. The annual rainfall averages 300-700mm, concentrated mainly from June to August [10].

MODIS provides operational standard products at 500m in daily, 8-day, and monthly data suits. We adopted the daily MODIS surface reflectance dataset MCD43A4 (collection 6) Nadir Bidirectional Reflectance Distribution Function (BRDF)-Adjusted Reflectance (NBAR) dataset during the growth period of 2005-2019, and as well as the related quality assessment layer to screen clear observations with less aerosol contamination.

We collected 3880 representative crop samples over the main agriculture region of HLJ from 2005 to 2018, according to the ground true data in 2018 and high-resolution remote sensing images (Fig. 1). In this study, we randomly selected 20% crop (~900) samples to construct the fuzzy membership model. The remaining 80% (~3000) samples were used to verify the classification accuracy. Due to the total sown area of HLJ almost unchanged in 2019-2020 [10], the GlobalLand30 of 2020 was used to screen out the non-agricultural region to tighten this work on the identification of fine crop types. In addition, the 10-m crop maps [11] in Northeast China in 2019 was used for cross-comparison.

### 3.1 Time series Vegetation Index

Vegetation Index is frequently considered as the enhanced signature of crop types than individual reflectance band, which combines multiple bands of reflectance to highlight vegetation information and suppress disturbance from non-vegetation surface clusters. We investigated phenology features of soybean, maize, paddy, and wheat through seven MODIS spectral bands and nine selected indices (shown in Table 1) that have been widely used in detecting different biophysical characteristics of vegetation, especially on the change of chlorophyll content, water content and canopy structure.

**Table 1.** The calculation formulas of the VI in this study (The B1-B7 respectively represent the reference bands 1-7 of MCD43A4)

VI	Calculated Formula of MCD43A4
Normalized Differenced Vegetation Index, NDVI [10]	$\frac{B2 - B1}{B2 + B1}$
The Enhanced Vegetation Index, EVI [5]	$2.5 \times \frac{B2 - B1}{B2 + 6 \times B1 - 7.5 \times B3 + 1}$
Land Surface Water Index, LSWI [11]	$\frac{B2 - B6}{B2 + B6}$
Mid-IR Bi-spectral Index, MIBRI [12]	$10 \times B7 - 9.8 \times B6 + 2$
Normalized Burn Ratio, NBR [12]	$\frac{B2 - B7}{B2 + B7}$
Normalized Difference Senescent Vegetation Index, NDSVI [12]	$\frac{B6 - B1}{B6 + B1}$
Normalized Difference Tillage Index, NDTI [10]	$\frac{B6 - B7}{B6 + B7}$
Normalized Difference Green Index, NDGI [10]	$\frac{B4 - B1}{B4 + B1}$
Normalized Difference Water Index, NDWI [13]	$\frac{B2 - B5}{B2 + B5}$

## 3 Method

### 3.2 Separability Index (SI)

We adopted SI to examine the separability of above selected factors in different growth period among crops. Somers defined SI as the ratio of inter-class spectral heterogeneity to intra-class spectral heterogeneity in 2010 [14]. For the spectrum of the two object types, SI increases when the intra-class spectral variance ( $\Delta_{intra}$ ) decreases and the inter-class distance ( $\Delta_{inter}$ ) increases, whereas decreases [10]. We calculated SI for each pair of crop types during DOY 100-300, and a global SI using the average method to describe the separability of time series features among all types. The formulas are as followed:

$$SI_{ij}(v, d) = \frac{\Delta_{inter}}{\Delta_{intra}} = \frac{|\mu_i - \mu_j|}{1.96 \times (\sigma_i + \sigma_j)} \quad (1)$$

$$SI_{global} = 1 / M \times \sum_{i=1}^{n-1} \sum_{j=i+1}^n SI_{ij} \quad (2)$$

where  $SI_{ij}$  is the SI of type  $i$  and type  $j$  ( $i, j = 1, 2, \dots, n$ );  $n$  is the number of classes;  $v$  refers to a band or a VI in this study;  $d$  refers DOY;  $\mu$  is the mean value of features;  $\sigma$  is the standard deviation of features;  $n$  is the number of all types;  $M$  is the number of class pairs.

### 3.3 Fuzzy Membership Model (FMM)

The FMM was used to evaluate the similarity of features between unclassified sample to target type. Fuzzy membership degree, from 0 to 1, is used to describe the degree of similarity between a sample to target class [15]. The formula is as followed:

$$s_{ik}(v, d) = e^{-\frac{(x_k - \mu)^2}{2\sigma^2}} \quad (3)$$

where  $x_k$  is the feature value of sample  $k$ ;  $s_{ik}$  means the fuzzy membership of sample  $k$  to class  $i$ ;  $v, d, \mu$  and  $\sigma$  are similar as the above formulas.

### 3.4 Multi-stage Weighting Model (MWM)

The MWM was used to screen out the features with the high degree of separation among land cover clusters, and identify the crop types according to the multi-stage results. In each stage, a normalized SI matrix was used to weight the FMM, and returned membership degrees of the unclassified sample belonging to target types. The weighting formulas are as followed:

$$p_{ik}^{(m)} = \sum_v \sum_d s_{ik}(v, d) \times SI_N^{(m)}(v, d) \quad (4)$$

$$SI_N(v, d) = SI(v, d) / \sum_v \sum_d SI(v, d) \quad (5)$$

where  $p_{ik}$  means the probability of sample  $k$  belonging to class  $i$ ;  $SI_N$  is the normalized matrix of all

SI (removing the values when the feature of sample  $i$  miss observation or the value is below the threshold of  $t_1$ );  $m$  means the  $m$ -th (the  $m$  times) judgment currently in progress;  $s, v$  and  $d$  are similar as the above formulas.

Firstly, we used the global SI to weight fuzzy membership matrix and summed up the weighted matrix for identifying the sample type as the first classification stage ( $m=1$ ). The first stage returns the weighted membership degree which means the similarity degree of sample  $k$  belongs to each type. When the membership degree of class  $i$  is higher than  $t_2$  ( $t_2$  is the value of  $s$  when  $x_k - \mu = 2\sigma$ ) and higher  $t_1$  than the membership degree of other types, it is considered that sample  $k$  belongs to the type  $i$ . When the membership degree of type  $i$  is greater than  $t_3$ , the possibility of sample  $k$  belonging to type  $i$  the is marked +1, and the possibility belonging other type is marked 0. Otherwise, demanding the first stage results invalid.

Secondly, we evaluated the possibility of belonging to all paired crop types ( $m>1$ ). When the membership degree of class  $i$  is higher  $t_3$  than the membership degree of another, the possibility with the higher membership type is marked as +1, and the possibility with the lower membership type is marked as -1. Otherwise, the possibility of the judgment in  $m$ -th stage is marked as 0.

Finally, the possibilities of sample  $k$  belonging to each type are summed together. When the weighted membership degree returned in any stage is less than  $t_4$  ( $t_4$  is the value of  $s$  when  $x_k - \mu = 3\sigma$ ), the possibility of sample  $k$  belonging to this type will become 0. When there is a possibility greater than 0, take the category with the maximum possibility as sample  $k$ . If the minimal probability is less than or equal to 0, marked sample  $k$  as non-crop type. In this study, the parameters  $t_1, t_2, t_3$  and  $t_4$  in classification were set as 0.7788, 0.3000, 0.1000 and 0.3679.

## 4 Results & Discussion

### 4.1 Crop separability

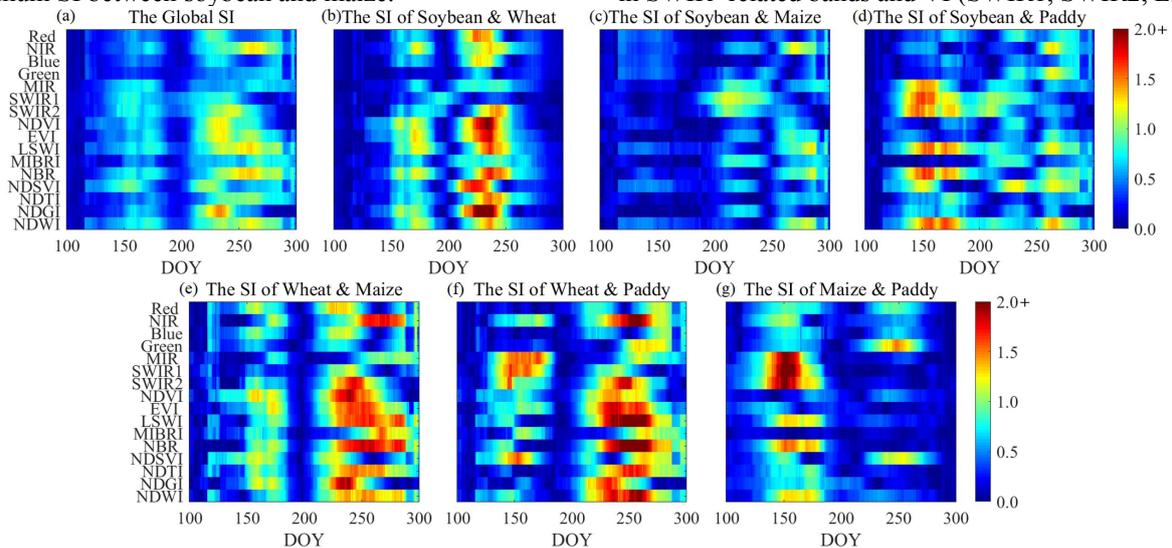
Fig.3. shows a global SI chart (a) for multiple crops and six SI charts (b-g) for each pair of crops (soybean, wheat, maize, and paddy) during the growth period of DOY 100-300 (x-axis) at seven spectral bands and nine indices (y-axis). The higher SI value means the higher separability for target crop types over 16 indices and 201 time points, and the corresponding grid becomes increasingly red. In this study, most of features with high SI values exhibit as the NIR- or SWIR- related indices, which are respectively linked with crop biomasses at the green-up and maturing stages, or the water content.

Fig. 3. (a) shows the global SI of four crop types with highest SI value of 1.4. The global SI higher than 1.0 is mainly during DOY 210-280, especially in NIR, SWIR2 and nine VI. However, the highest SI value during 140-160 (0.8-1.0) is much lower than the value of some pairwise SI, especially to Fig.3. d and g (>1.5).

The pairwise SI between wheat and other crop types (Fig.3. b, e, f) perform highest value about 1.8-2.4, especially to soybean and maize (b, e), and almost all of

VI can effectively distinguish wheat from other crops, during DOY160-210. To reflectance features, the highest SI of reflectance features is only concentrated in low vegetation cover periods (DOY140-160), and the highest value is about 1.5, which is still much higher than the maximum SI between soybean and maize.

The SI between soybean and maize (Fig.3. c) are relatively lower than other crop pairs with maximum 1.3. NDVI is the most widely used VI in crop classification, but the highest SI value is only 0.8. The higher SI between soybean and maize (1.0-1.3) is mainly showed in SWIR- related bands and VI (SWIR1, SWIR2, LSWI,



**Fig. 3.** Estimated SI values of each pair of crops during the growth period of DOY 100-300 (x-axis) at seven spectral bands and nine indices (y- axis).

NBR, MIBRI and NDWI) during DOY200-240.

The paddy presents the remarkable separability ( $SI > 1.5$ ) to maize and soybean (Fig.3. d, g) during transplanting period (DOY130-160). For paddy and soybean (Fig.3. d), the SI of three spectral bands (MIR, SWIR1 and SWIR2) and three SWIR- related indices (LSWI, NBR and NDMI) is similar (1.5-1.7). For paddy and maize (Fig.3. g), the SI of three spectral bands (1.7-2.1) is even higher than the SI of the three SWIR- related indices (1.5-1.7).

In this part, the four crop types have showed a great possibility to distinguish each other by the time series features. There is the highest separability between wheat and other types, due to the growth and the dry leaf period of wheat are earlier 30 days approximately than other crops. Similarly, paddy is extractable, due to the spectral features during irrigation and transplanting period (DOY140-160). The SI between soybean and maize is slightly lower, but the time series features with relatively high SI last for more than 40 days, which may give helps to effectively distinguish them.

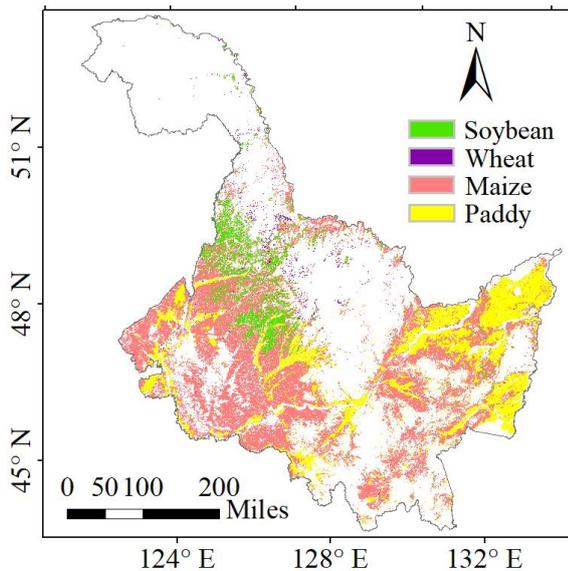
#### 4.2 Classification results

We applied the above presented method to classified 3103 random selected ground sites during 2005-2018 and the overall agriculture region of HLJ in 2019 using MODIS daily data of DOY 100-300. Table 2. shows the classification results of 3103 random selected ground samples. Each column of the matrix represents the indicates in actual crop types, and the row of the matrix represents the indicates in predicated crop types. In case of the classification results during 2005-2018, the overall accuracy is 0.9816, and the kappa coefficient is 0.9702.

Almost all samples of wheat and paddy were correctly classified. Soybean and maize samples were slightly misclassified. The user accuracy of soybean (0.9330) is lower than the producer accuracy (0.9953), means the soybean area extraction by the model may be overestimated. Relatively, the maize area extraction by the model may be underestimated. In this study, about 40% agricultural samples were correctly classified by the first classification stage in MWM, and 60% agricultural samples were correctly classified by the 2-7 classification stage in MWM.

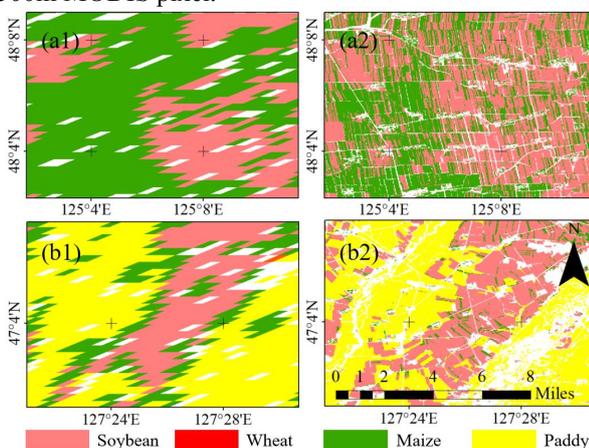
**Table 2.** The confusion matrix of classification results for samples in HLJ during2005-2018 (the Soy- means Soybean).

		Actual				ALL
		Soy-	Wheat	Maize	Paddy	
Predicted	Soy-	640	0	46	0	686
	Wheat	0	50	0	0	50
	Maize	3	0	1608	8	1619
	Paddy	0	0	0	748	748
	ALL	643	50	1654	756	3103



**Fig. 4.** The extracted crop map (500m) of HLJ in 2019

Fig.4. shows the extraction results for four crop types over the agriculture region of HLJ in 2019. Maize is the most sown crop in HLJ, followed with paddy, and they are mainly distributed in SJP and the south of SNP. Soybean is mainly distributed in the north of SNP. Wheat is the least sown crop and sporadically distributed in the north of HLJ. The 10-m crop maps in Northeast China in 2019 was used for cross-comparison, as showed in Fig.5. In the soybean dominant planting region (Fig.5. a1, a2), the extracted soybean area of MODIS-based 500-m crop map was underestimated compared to the soybean area of 10-m crop map. It has the same performance in maize dominant planting areas. For most of paddy dominant planting regions, the extracted paddy distribution in this study is highly consistent with the paddy distribution of 10m reference map. In some boundary of maize field and paddy field, numbers of MODIS pixels with mixed crop types were divided into soybean, although soybean is rarely planted in this area (Fig. 5. b1, b2). Besides, the classification results performed weak over the field of wheat or other mixed crops at MODIS pixel scale, because of the increase of fragmentation crop fields which are far smaller than 500m MODIS pixel.



**Fig. 5.** Comparison between MODIS-based 500-m crop map (a1, b1) and the sentinel-based 10-m crop map (a2, b2) in case of two typical crop regions of HLJ in 2019.

## 5 Conclusions

We developed a physiological feature determined classification model at daily intervals during the growing season, and we classified four major crops (soybean, wheat, maize, and paddy) in case study of agriculture region of HLJ. The results show higher accuracy achieved over main agriculture region of soybean, wheat, maize, and paddy, and reduced accuracy over field of wheat or other mixed crops at MODIS pixel scale. Compared with the traditional multi temporal classification method, our model potentially improves the degree of automatic extraction of fine crop categories in the large region scale, due to the classification mechanism is interpretable and flexible. In the following study, we are going to continuously contribute efforts to improve the model and the identification of the misclassified targets.

This work was supported by the National Natural Science Foundation of China (No. 42071351), the National Key Research and Development Program of China (No. 2020YFA0608501, No. 2017YFB0504204), the Liaoning Revitalization Talents Program (No. XLYC1802027), talent recruitment program of CAS (No. Y938091, No. Y674141001), Project supported discipline innovation team of Liaoning Technical University (No. LNTU20TD-23).

## References

1. B. Wu, J. Meng and Q. Li, China crop watch system with remote sensing. *Journal of Remote Sensing*, **8** 481-497. (2004)
2. V.S. Konduri, J. Kumar, J.W.W. Hargrove, et al. Mapping crops within the growing season across the United States. *Remote Sensing of Environment*, **251** 112048. (2020)
3. Z. Chen, J. Ren, H Tang, et al. Progress and perspectives on agricultural remote sensing research and applications in China. *Journal of Remote Sensing*, **20(5)**: 748-767 (2016)
4. L. Zhong, P. Gong and G. Biging, Efficient corn and soybean mapping with temporal extend ability: A multi-year experiment using Landsat imagery. *Remote Sensing of Environment*, **140** 1-13 (2014)
5. H. Thi, T Nguyen, L.V. Nguyen, et al. Mapping Maize Cropping Patterns in Dak Lak, Vietnam Through MODIS EVI Time Series. *Agronomy*, **10**: 1-16. (2020)
6. Q. Hu, W. Wu, Y. Shao, et al. Advances in remote sensing extraction of crop planting structure. *Scientia Agricultura Sinica.*, **48** 1900-1914. (2015)
7. Q. Xu, G. Yang, H. Long, et al. Crop information identification based on MODIS NDVI time-series data. *Transactions of the Chinese Society of Agricultural Engineering*, **30** 134-144. (2014)

8. C. Jun, Y. Ban and S. Li, China: Open access to Earth land-cover map. *Nature*, **514**: 434-434. (2014)
9. National Bureau of Statistics. Statistics of main crop sown areas in each province of China. (2021). <http://www.stats.gov.cn/tjsj/>
10. Q. Hu, Research on crop remote sensing recognition method based on time series MODIS images. Chinese Academy of Agricultural Sciences. (2018)
11. N. You, J. Dong, J. Huang, G. Du, et al. The 10-m crop type maps in Northeast China during 2017–2019. *Scientific Data*, **8** 1-11. (2021)
12. L. Gu, Y. Shuai, C. Shao, et al. Angle Effect on Typical Optical Remote Sensing Indices in Vegetation Monitoring. *Remote. Sensing*, **13**: 1699. (2021)
13. B. Gao, NDWI—A normalized difference water index for remote sensing of vegetation liquid water from space. *Remote Sensing of Environment*, **58**: 257-266. (1996)
14. B. Somers, S Delalieux, W.W. Verstraeten, et al. An automated wave band selection technique for optimized hyperspectral mixture analysis. *International Journal of Remote Sensing*. **31** 5549-5568. (2010)
15. Z. Peng and Y. Sun, *Fuzzy Mathematics and its Applications*. Wuhan University Press, Wuhan. (2007)