



Crop type mapping with temporal sample migration

Shibo Zhang, Jingya Yang, Pei Leng, Yuman Ma, Hongyang Wang & Qian Song

To cite this article: Shibo Zhang, Jingya Yang, Pei Leng, Yuman Ma, Hongyang Wang & Qian Song (2023): Crop type mapping with temporal sample migration, International Journal of Remote Sensing, DOI: [10.1080/01431161.2023.2192881](https://doi.org/10.1080/01431161.2023.2192881)

To link to this article: <https://doi.org/10.1080/01431161.2023.2192881>



© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 31 Mar 2023.



Submit your article to this journal [↗](#)



Article views: 578



View related articles [↗](#)



View Crossmark data [↗](#)



Crop type mapping with temporal sample migration

Shibo Zhang, Jingya Yang, Pei Leng, Yuman Ma, Hongyang Wang and Qian Song

State Key Laboratory of Efficient Utilization of Arid and Semi-arid Arable Land in Northern China/Key Laboratory of Agricultural Remote Sensing, Ministry of Agriculture and Rural Affairs/Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing, China

ABSTRACT

Accurate and timely crop maps are crucial for monitoring agricultural production. Current supervised classification methods based on remote sensing rely heavily on ground-truth samples collected at a high cost, and the years without sampling highly limit classification accuracy. To address such a challenge, we proposed a time-migration method based on historical training samples collected in 2017, 2018 and 2020 to conduct supervised crop classification mapping in the target year (2021) with no ground samples. We chose Hailun City, Heilongjiang Province of northeastern China, as the study site; the major crops included corn, soybean, and rice. We reconstructed time series of Sentinel-2 data and selected spectro-temporal features to identify standard crop phenological curves. We calculated the similarity between reference and image spectra and designed label-matching rules to identify training samples through the dynamic time warping algorithm. We then used the historical samples to map the crop types of the target year. The results showed that the migration accuracy reached 95% for major crop. Using these samples as training data with a random forest to classify the target year, the overall accuracy reached 94.13%. The new sample time-migration method proposed in this study can efficiently migrate historical samples, greatly reducing the cost of ground-truth sampling.

ARTICLE HISTORY

Received 26 December 2022
Accepted 15 March 2023

KEYWORDS

crop mapping; historical samples; RF; time migration; Sentinel-2

1. Introduction

Reliable and timely spatial distribution information on crop types is vital for agricultural monitoring, food security, and environmental sustainability assessments (Xu et al. 2021). Traditionally, crop area information is routinely acquired and updated through communications with farmers or ground surveys. However, this procedure is prone to errors, time-consuming and expensive. In this regard, remote sensing has been acknowledged as the most effective tool for mapping crop spatial distribution and estimating crop area (Mariotto et al. 2013; Hu et al. 2018; Maxwell, Warner, and Fang 2018; Zhang et al. 2020).

CONTACT Qian Song ✉ songqian01@caas.cn 📧 State Key Laboratory of Efficient Utilization of Arid and Semi-arid Arable Land in Northern China/Key Laboratory of Agricultural Remote Sensing, Ministry of Agriculture and Rural Affairs/Institute of Agricultural Resources and Regional Planning, Chinese Academy of Agricultural Sciences, Beijing 100081, China

This article has been corrected with minor changes. These changes do not impact the academic content of the article.

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited, and is not altered, transformed, or built upon in any way. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

Many crop classification methods combine remotely sensed image data with supervised classifiers. Classical parametric supervised algorithms such as maximum likelihood, machine learning approaches such as artificial neural networks, random forests, support vector machines, decision trees and ensembles of classifiers have been applied to crop mapping based on training datasets (Zhang et al. 2019). Today, the most prominent examples of operational use of remote sensing data with machine learning for national-scale crop type mapping are the Cropland Data Layer generated by the United States Department of Agriculture (<https://nassgeodata.gmu.edu/>), the Annual Crop Inventory generated by the Agriculture and Agri-Food Canada (<https://agriculture.canada.ca/>), and the Sen2-Agri automated system (<http://www.esa-sen2agri.org/products/crop-type-map/>). These data products are used to perform assessments in various cropping systems at national scales. Large amounts of comprehensive and in-season survey-based geospatial data are input as training datasets to the classification models. Such data are also expensive and difficult to generate and update, particularly in areas where environmental and climatic conditions are especially complex (Song et al. 2017; Wen et al. 2022). However, because each crop has a specific crop calendar with well-defined planting times as well as unique seasonal growth and development rhythms, resulting in typical spectral reflectance variations within the crop-growing season, there is an opportunity for improving the utilization efficiency of historical samples to make up for missing in-season training data sets and complete annual crop classification tasks more sustainably.

Successful examples using historical samples in remote sensing mapping primarily describe land-cover categories. For instance, Huang (Huang et al. 2020) developed an automatic training sample migration method based on first all-season sample sets, and measured the spectral similarity and distance between reference and image spectral to identify the change state of training sample pixels in other years. Mao (Mao et al. 2020) mapped wetlands in China by developing a hybrid object-based and hierarchical classification approach and a new wetland classification system with the field surveys carried out from June 2014 to September 2016. However, using historical survey-based data for crop type mapping is challenged by the fact that croplands are intensively managed and may change from year to year based on farmers' decisions related to what to sow.

Substantial efforts have been made towards improving crop maps by repeatedly utilizing ground sample for multiple years without the need for year-to-year reformulation. One way to classify a crop is by using the key transition dates and corresponding values by setting thresholds derived from a vegetation index curve (You et al. 2013). For instance, Xuan (Xuan et al. 2023) identified different types of crops by designing an absolute distance index between crops based on Normalized Difference Vegetation Index (NDVI) time series, and Huang (Huang et al. 2022) also identified differences between soybean and other crops from multiple phenological indices and designed decision rules to distinguish soybean from non-soybean. Xiong (Xiong et al. 2017) has also set a large number of thresholds to extract farmland in the African region. Zhong (Zhong et al. 2016) explored the phenological characteristics of crops to create multiple indicators for decision tree classification.

However, the growth cycle of crops is not completely fixed and can be shifted by many factors, such as climate, policies, and agricultural practices (Dong et al. 2020). Several researchers have compared metrics dealing with time series, there is increasing evidence that the classic Dynamic Time Warping (DTW) measure is the best measure in most

domains (Ding et al. 2008). DTW is a measure of the similarity of two-time series of different lengths. DTW uses dynamic calculation techniques to find the minimum distance between two time series, where the series are distorted by stretching or shrinking in the time dimension. Techniques based on a DTW algorithm have been increasingly implemented in remote-sensed image analysis of the differences in curve offset (Petitjean, Inglada, and Gancarski 2012; Xue, Du, and Feng 2014; Costa et al. 2018; Gella, Bijker, and Belgiu 2021). Guan (Guan et al. 2016) constructed rice time series NDVI standard curve and successfully extracted rice acreage in Vietnam by calculating DTW similarity and setting the threshold. Maus (Maus et al. 2016) improved DTW by introducing the concept of time weighting to achieve more accurate land cover classification based on the original. François Petitjean (Petitjean, Inglada, and Gancarski 2012) demonstrated the potential and advantages of DTW in processing satellite image time series, and combined DTW and K-means for land cover classification with good results. Fan (Fan et al. 2022) also used the DTW algorithm to extract the spatial distribution of winter wheat in Hengshui City.

Nevertheless, few studies have applied the DTW method to year-to-year sample migration and whether the DTW method is capable of mapping various crop types over regions based on historical samples still needs to be tested.

This study aimed to address this knowledge gap by developing a method for migrating samples from the reference years (2017, 2018, 2020) to the target year (2021) for accurate crop classification. We implemented the temporal migration of the samples based on the similarity of the phenological curves. Three specific study objectives included: (1) Designing rules for sample migration; (2) Comparing the migration of classification results between fusion curves and single-year curves; (3) Evaluating the effect of sample size and representativeness on classification results. Our proposed method is expected to be applicable to the entire northeast of China because the similarity in cropping structure across the region.

2. Study area and data

2.1. Study area

Our study area is Hailun City, which is located in the Songnen Plain (46°58′–47°52′N, 126°14′–127°45′E) in the north-central part of Heilongjiang Province. It has a total area of 4667 km², of which 3100 km² is arable land. The topography of the study area gradually decreases from northeast to southwest, transitioning through low hills, high plains, river terraces, and river floodplains in a stepped form. The average elevation is 239 m. Hailun City has a temperate continental monsoon climate. Winters in Hailun City are long and cold, with an annual average temperature of 2.0°C. Conversely, summers are warm and short, with an annual precipitation of 500–600 mm and mainly concentrated between May and September (<https://Hailun.gov.cn>).

Heilongjiang Province is an important grain production base in China, and Hailun City is one of the representative grain-producing areas, as well as a national base county for comprehensive scientific experiments on agricultural modernization and a national base county for commercial grain. Hailun City was chosen not only because it is an important grain-producing region. It is also one of the representative regions of Heilongjiang Province. The crop cultivation structure of Hailun City is typical of one of the regions in

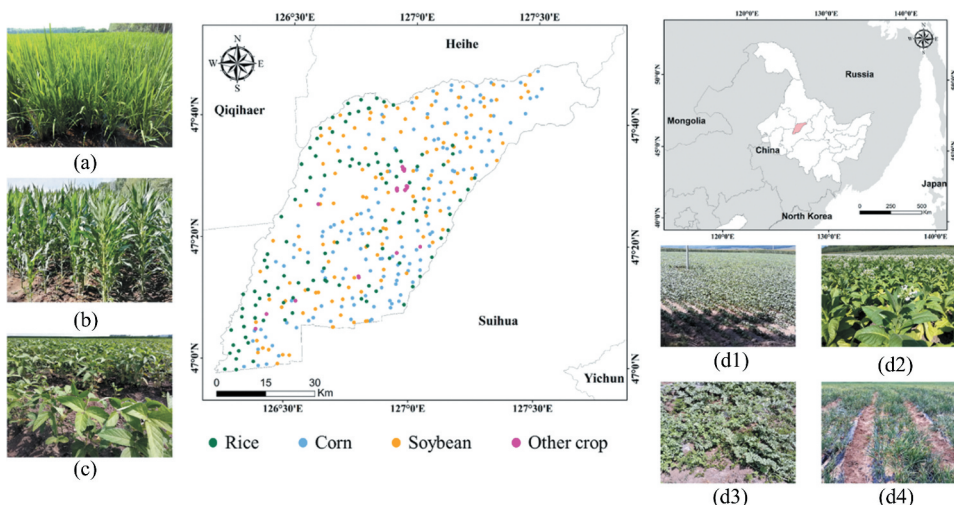


Figure 1. Study area and sample distribution in 2021. (a) rice; (b) corn; (c) soybean; (d1) string beans; (d2) lettuce; (d3) radish; (d4) onion.

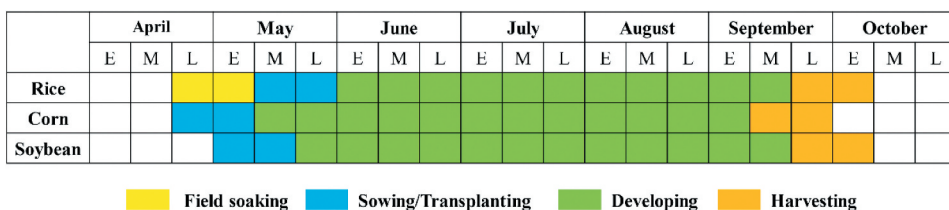


Figure 2. Crop calendar of three major crops. E, M, and L indicate early, middle, and last ten-day of a month, respectively.

Heilongjiang Province, producing mainly soybean, corn, and rice. In response to conservation tillage measures, corn and soybean crop rotations have been widely adopted in Hailun City. According to the 2020 statistics of Heilongjiang (<https://hlj.gov.cn>), the sown area of the three main crops accounted for 97.8% of the total grain sown area. The remaining other crops include wheat, peanuts and vegetables, among others (Figure 1).

Due to its climate, crops in the study area are usually grown in one season per year (Figure 2). Rice is mostly cultivated in the main season between spring and summer, approximately from April to October. Corn is mainly sown in the first half of May and harvested in the first half of October. Soybean and corn are harvested at the same time, but soybean is planted about ten days later than corn. The rotation between soybean and corn is very common in different years.

2.2. Data and processing

2.2.1. Sentinel-2 images and pre-processing

This study used Sentinel-2A/B (S2) Multi-Spectral Instrument (MSI) data. The S2 program includes two satellites, Sentinel-2A and Sentinel-2B, launched in 2015 and 2017,

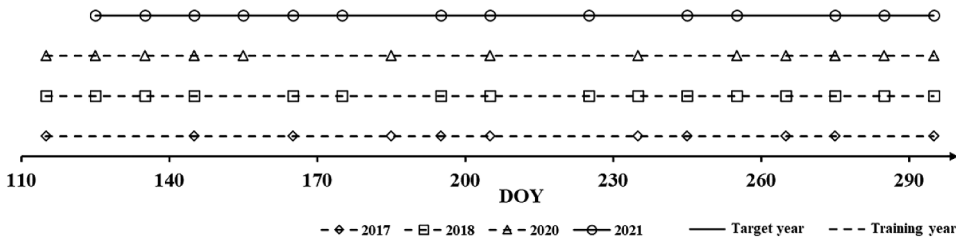


Figure 3. S2-A/B Data distribution.

respectively. The satellites acquire data in 13 bands with a spectral range of $0.4 \sim 2.4 \mu\text{m}$, covering visible light, near-infrared, and short-wave infrared bands. Together, Sentinel-2A and -2B observe earth's land surfaces with a 5-day repeat cycle and a $10 \sim 60 \text{ m}$ spatial resolution. The Google Earth Engine (GEE) platform has near real-time updates of uploaded remote-sensing datasets and processed products. The GEE platform also provides powerful computing power for the rapid processing of massive remote-sensing datasets. The S2 surface reflectance (SR) archived on the GEE is processed by radiometric and geometric corrections including ortho-rectification, atmospheric correction, and spatial registration.

To further reduce cloud contamination, we selected images with a cloud score of less than 30 and chose the time interval from late April to late October of each year (2017, 2018, 2020, 2021), covering the entire growing period of major crops (Figure 3). We used the traditional 10-day maximum composite method to reduce noise and obtain images at equal time intervals covering the different stages of the crop growth period (Liu et al., 2020).

2.2.2. Ground reference data

Ground-based field surveys were conducted in 2017, 2018, 2020, and 2021. We recorded crop types by taking pictures with our smartphones and measured sample location with a handheld Global Positioning System.

The surveyed crops included rice, corn, soybean, and other crops. Since this study focused on mapping the main crops in Hailun City, various other crop types were merged into one class called 'other crop'. In addition, we selected a total of 141 samples of other classes from high-resolution Google Earth images, including buildings, woodlands, water bodies, and roads. We collected 507 samples in 2017, 655 samples in 2018, 530 samples in 2020, and 401 samples in 2021 (Table 1). In the classification task for the reference year 80% of the samples are used for training and 20% for accuracy validation. In the target year classification task, all of our ground samples collected are used for validation.

3. Methods

3.1. Overview of the sample migration method

The proposed methodology of crop type mapping with temporal sample migration consists of four parts (Figure 4). In the first part, we calculated the constructed Sentinel 2 time series data to obtain five characteristic time series, and then extracted

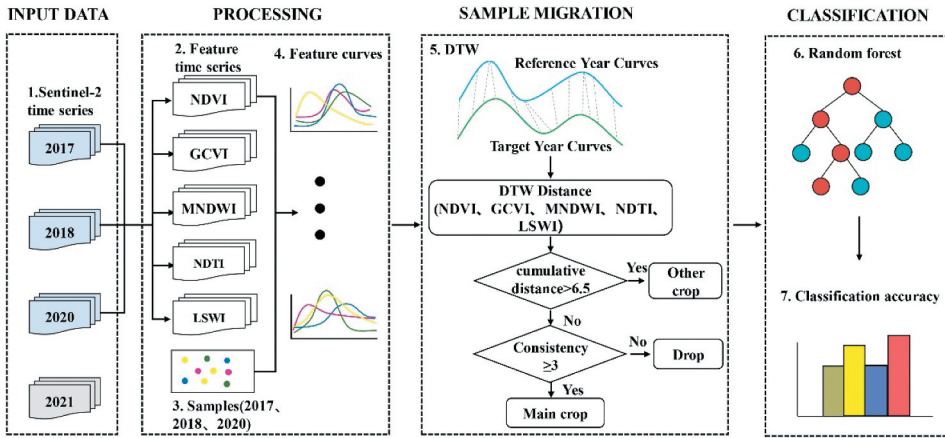


Figure 4. Sample migration and classification process.

these characteristic curves from our collected ground samples. In each year we have five characteristic curves for each crop type and we average the individual characteristic curves and use the average curve as our standard curve. Similarly we can obtain the average curve for each crop type and for each year. The feature curves of the target year were taken as the target curves, whereas the average curves represented the standard curves. The distance between them was calculated by DTW for label matching. Smaller distances were taken as an indication of the similarity between the two curves. Subsequently, we compared the crop standard curves with each target year curve, and chose the one that appears most frequently as the label for that curve. We developed a method to normalize the calculated distances and sum them to get the cumulative distance, and then set the threshold for further detailed classification. We considered the curve with a cumulative distance greater than 6.5 as other crops. On this basis consistency greater than or equal to 3 is the main crop to maintain the original label. Finally, the obtained samples are input to the random forest for classification and validation of classification accuracy evaluation.

3.2. Feature selection

Taking advantage of S2's high spatial and temporal resolution and high number of spectral bands, we calculated multiple vegetation indices using several spectral bands that are sensitive to the crop growth response and assembled a database of phenological curves. Vegetation indices also help minimize the influence of soil background, atmosphere and solar-target-sensor and are widely used for local-scale to global-scale crop growth monitoring studies (Zeng et al. 2022).

We calculated five common vegetation indices: the Green Chlorophyll Vegetation Index (GCVI), the Modified Normalized Difference Water Index (MNDWI), the Land Surface Water Index (LSWI), NDVI, and the Normalized Difference Residue Index (NDTI). Their equations are listed in Table 2.

pred, pnir, pswir1, and pswir2 are surface reflectances of Band 3 (green, 559.8 nm (S2A)/559.0 nm (S2B)), Band 4 (red, 664.5 nm (S2A)/665 nm (S2B)), Band NIR (NIR, 864.8 nm

Table 1. Ground-truth samples collected in different years.

	Rice	Corn	Soybean	Other crop	Other	Total
2017	83	116	136	31	141	507
2018	96	168	210	40	141	655
2020	104	119	125	41	141	530
2021	108	147	126	20	141	401

Table 2. Summary of five vegetation indices for crop classification.

Vegetation Index	Formula	Reference
Green Chlorophyll Vegetation Index(GCVI)	$\frac{\rho_{nir}}{\rho_{green}} - 1$	(Gitelson et al. 2003)
Modified Normalized Difference Water Index(MNDWI)	$\frac{\rho_{green} - \rho_{swir1}}{\rho_{green} + \rho_{swir1}}$	(Xu 2007)
Land Surface Water Index(LSWI)	$\frac{\rho_{nir} - \rho_{swir1}}{\rho_{nir} + \rho_{swir1}}$	(Chandrasekar et al. 2010)
Normalized Difference Vegetation Index(NDVI)	$\frac{\rho_{nir} - \rho_{red}}{\rho_{nir} + \rho_{red}}$	(Roujean and Breon 1995)
Normalized Difference Residue Index(NDTI)	$\frac{\rho_{swir1} - \rho_{swir2}}{\rho_{swir1} + \rho_{swir2}}$	(Quemada et al. 2018)

(S2A)/864 nm (S2B)), Band 11 (SWIR1, 1613.7 nm (S2A)/1610.4 nm (S2B)) and Band 12 (SWIR2, 2202.4 nm (S2A)/2185.7 nm (S2B)) of the Sentinel-2 MSI sensor.

3.3. Sample migration decision

We first extracted the time series curves of the existing samples. The next step was the calculation of the standard curve. The data were cleaned to ensure the representativeness and accuracy of the curve. We considered values greater than three times the standard deviation as outliers and removed them; subsequently, we used the mean of the sample curves as our standard gestalt curve, with all categories treated identically and obtained standard phenological curves for rice, corn and soybean for each year. We have also tried to extract the standard curves of other crops but it was difficult to find representative patterns. However, considering that our goal was to classify the major crops, we did not explore the other crops in depth.

We conducted crossover experiments in 2017, 2018 and 2020 with the migration of existing samples. We extracted phenological curves for rice, maize and soybean for each year. The phenological curves of each target year sample were compared with the standard curve of the reference year to calculate the DTW distances, deriving three distances for each feature. The smallest of these distances was then considered the label. With this treatment, we obtained fifteen distances and five labels for each sample. To proceed with label matching, we normalized the fifteen distances to the range of 0–1; then we summed them and set a threshold to distinguish other crops. Several cross-migration experiments indicated that at a threshold distance greater than 6.5, the curve should be labelled as other crops. Finally, we performed label matching for the major crops by counting the labels of each sample. A label consistency greater than or equal to 3 was considered the final label.

3.4. Scenario design

The number and spatial distribution of samples for each year vary, and the image quality for different periods in each year is uncertain. Contamination from clouds and

Table 3. Overview of the seven feature scenarios used in this study.

Feature Type	Scenario	Feature Variables
Single year	S1	All times series in 2017
Single year	S2	All times series in 2018
Single year	S3	All times series in 2020
Combined year	S4	All times series in 2017 and 2018
Combined year	S5	All times series in 2017 and 2020
Combined year	S6	All times series in 2018 and 2020
Combined year	S7	All times series in 2017, 2018 and 2020

shadows can result in some samples being unavailable, which is inevitable. To mitigate this and maximize the utilization of samples collected across all years, we explore the relationship between the quantity and migratory accuracy by designing multiple scenario models. We designed seven different sets of sample combination scenarios for different years. Table 3 lists the specific composition of these seven feature scenarios. These were classified based on the same classifier with the same validation samples for accuracy assessment. The classification accuracies generated by these feature scenarios were then used to compare and analyse the relative importance and influence mechanisms of the number and representativeness of samples on crop identification.

3.5. Random forest classifier

Random forest (RF) is a machine learning algorithm integrating multiple decision trees, which uses the bootstrap resampling method to construct decision trees by randomly selecting samples from the training sample set, and combining the prediction results of all decision trees to obtain the final results by voting. With fast computing speed, robust model, and strong generalization ability, the RF algorithm is a common tool for remote sensing mapping of crops (Zhang et al. 2018; You et al. 2021). The RF classifier can take a large number of input variables and perform fast classification without being heavily influenced by ‘dimensional catastrophe’ or noise.

Another advantage of RF is that it provides a quantitative measure of each variable’s relative contribution to the classification result for the user to assess the importance of the input variables with the Gini coefficient. The Gini coefficient measures the amount of impurity in the nodes of a decision tree, which is the amount of randomness or uncertainty in the model. The higher the Gini value, the more important the feature is in making the model successful. Selecting the most representative vegetation indices to construct the phenology curves is an important step. We selected the five best-performing vegetation indices from ten commonly used vegetation indices including NDVI, MNDWI, EVI, NDPI, LSWI, NDWI, DVI, NDBI, GCVI and NDTI. The two important parameters in the random forest are *n*tree and *m*try. The general better reference value of *m*try parameter is $\sqrt{p}$, where *p* is the sum of the feature variables in the training feature set, which is also the default value in the GEE random forest classifier. We control the number of trees, and the number of trees increases gradually from 50, and when it increases to 100 the accuracy has reached its maximum without any significant improvement. Therefore the optimal *n*tree in this study is 100.

3.6. Support vector machines classifier

Support vector machines (SVM) are supervised machine learning models that analyse data for classification and regression analysis. SVM has received significant attention in the field of remote sensing image classification due to its superior performance when compared to other ML classifiers such as K-nearest neighbours (KNN), artificial neural networks, and decision trees (Sheykhmousa et al. 2020). And even Mountrakis (Mountrakis, Im, and Ogole 2011) considers SVMs to be the 'best' algorithm for classification. The most basic idea of a SVM is to find linearly divisible straight lines or hyperplanes in the sample space that separate different classes of samples. There are many such straight lines, and the support vector machine considers the best straight line to be the one that has the maximum distance after dividing the two classes of targets. SVM exhibit many unique advantages in solving small sample, nonlinear and high-dimensional pattern recognition. In the SVM machine, first we determine the kernel function type, and choose the linear kernel function among many kernel types, and the accuracy can reach the highest when the $C=1$ and gamma value is 0.01, at which time we think the accuracy of classifier reach the peak.

3.7. Classification accuracy assessment

The F1 score takes into account both precision and recall so that the two reach the highest at the same time and take a balance.

$$F1 \text{ score} = 2 \cdot (UA * PA) / (UA + PA) \quad (1)$$

where UA and PA represent the user's and the producer's accuracy, respectively (Xia et al., 2022). For each year's classification, we calculated Overall Accuracy (OA) and Kappa to evaluate the accuracy of the classification results.

$$OA = \frac{TP + TN}{TP + FP + TN + FN} \times 100\% \quad (2)$$

where TP refers to true positives, FP refers to false positives, TN refers to true negatives, and FN refers to false negatives. The fitness function is to be maximized, i.e. the higher the fitness value, the better the individual (Wen et al. 2022).

$$\text{Kappa} = \frac{P_o - P_e}{1 - P_e} \quad (3)$$

$$P_e = \frac{(TP + FN)(TP + FP) + (TN + FP)(TN + FN)}{TP + FP + TN + FN} \quad (4)$$

P_o is the sum of the number of correctly classified samples in each category divided by the total number of samples, which is the overall classification accuracy.

4. Results

4.1. Characteristic curve analysis

Multiple vegetation index curves for different years are shown in [Figure 6](#). The phenology curves of different crops in the same year varied widely, with interclass variability often greater than intraclass variability. However, the growth cycles of maize and soybean were relatively similar and harder to distinguish with several vegetation indices, with the difference between the two not as obvious as with rice. In addition, different types of crops differ in their sensitivity to individual vegetation indices. For example, NDVI was not sufficiently distinct for rice, corn, and soybean throughout the crop growing season. The transplanting period (DOY = 150) is a special stage in the rice growth cycle when the soil water content is extremely high and the MNDWI and LSWI are more sensitive to water bodies so that paddy and dryland crops can be effectively distinguished during this period. In the growth and development stage of crops, when the crop canopy cover and greenness are at their highest, vegetation indices that are more sensitive to crop canopy work best. GCVI can extract chlorophyll content in crops and effectively extract canopy characteristics using NIR and green spectral range, so it is reasonable that GCVI can distinguish these crop types better than other vegetation indices. At the maturity harvest stage of a crop, straw handling varies among the different crops harvested today with large machinery, resulting in a wide variation in surface residues after harvest. Straw residues on the surface are more sensitive at this time and play a key role in NDTI classification.

Although the shapes of the vegetation index curves of the same crop category are relatively similar across years, the date of their peak occurrence is not consistent. NDVI is the most widely used vegetation index in the field of remote sensing information extraction, which can reflect the vegetation cover status and growth state. Taking NDVI as an example, soybean peaked at DOY 203 in 2017, 208 in 2018, and 223 in 2020. Similarly, corn peaked at DOY 197, 205, and 239, and rice peaked at DOY 196, 200, and 202. Therefore, we fused the curves of three years with the DTW Barycenter Averaging (DBA) algorithm. It has been shown that the result on averaging time series can be augmented by the DBA method (Petitjean and Weber 2014). The resulting fusion curve better takes into account the different situations of multiple years and is therefore more generalizable ([Figure 5](#)).

4.2. Sample migration results for the target year

The extracted crop phenology curves for each year were used as respective reference curves, and the curves extracted from the samples in 2021 were used as target curves. We used actual 2021 labels to verify the accuracy of our method.

The migration accuracies of various crops are shown in [Figure 6](#). The single-year curves showed the most consistent performance across crops in 2017 and the worst accuracy in 2020. The 2018 curves had the highest accuracy in other crops.

Crop types presented different identification challenges. Rice had the highest detection accuracy, with F1-scores generally above 0.91, with the highest being 0.98 for 2020. The classification accuracy of other crops was generally lower,

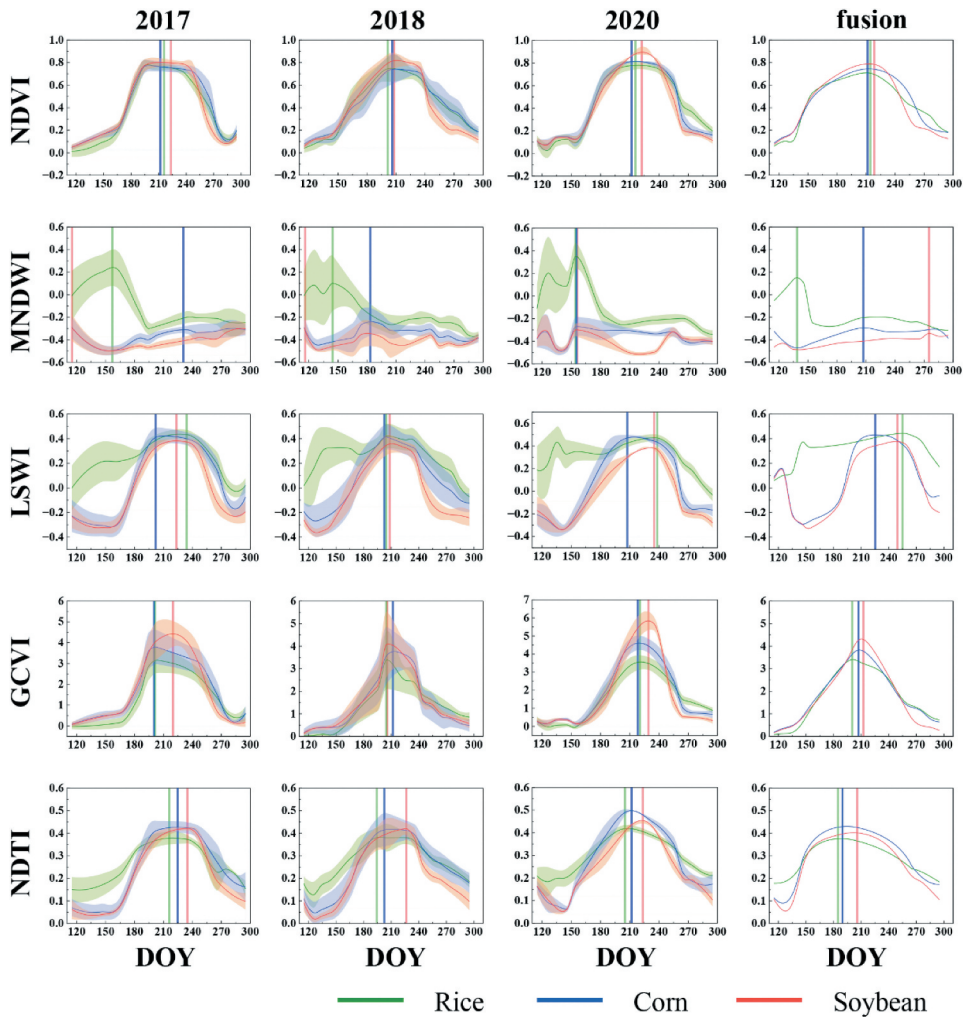


Figure 5. Vegetation index curves of different crops. Range filled color represents error bars with one positive/negative standard deviation.

between 0.65 and 0.8; the curve for 2018 had the highest accuracy in recognizing other crops, with an F1 score of 0.77. Overall, the classification accuracies of corn and soybean were relatively close, with scores for all three years above 0.85, which was relatively stable.

The migration accuracy of the fused multi-year curves obtained by linear interpolation was generally higher than that of the single-year curves and more robust. The fused curves were better matched with the target curves after linear interpolation because they maximized phenological information. Therefore, we chose the fused curve as the subsequent reference curve.

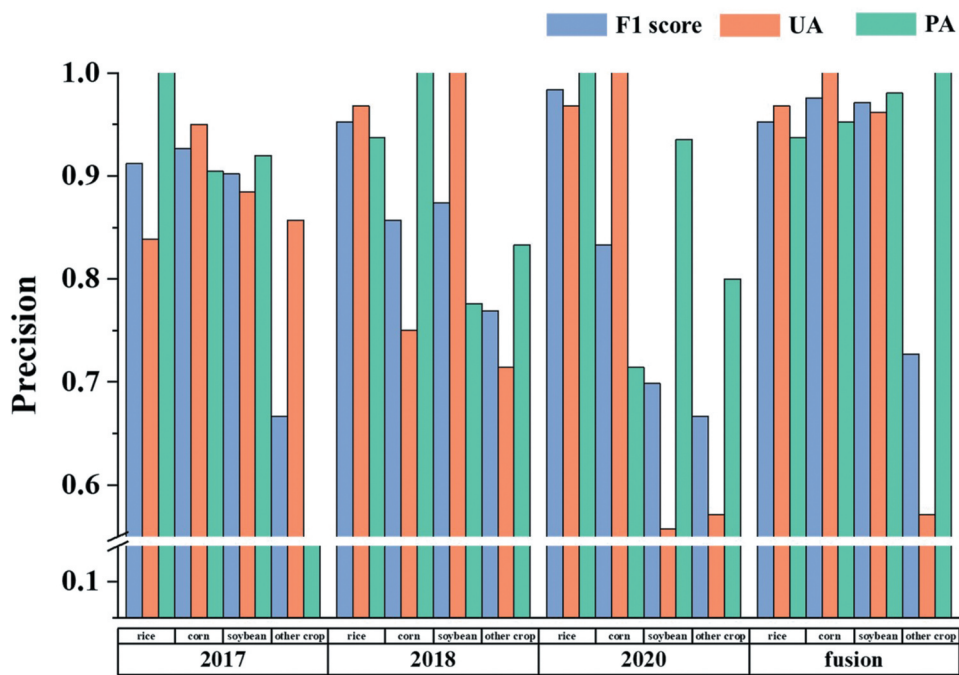


Figure 6. Sample precision of curves for different years.

4.3. Classifier comparison

We compared RF and SVM classifications for each year separately to identify the best-performing model. The results show that the classification accuracy of RF outperformed SVM in all three years of the comparison experiments (Table 4). The lowest classification accuracy for both classifiers was obtained with 2017 data and the highest classification accuracy was in 2020. The number of samples and available images was the smallest in 2017. The other two years were better than 2017 in terms of both the number of samples and the quality of images, so the classification was relatively better.

4.4. Crop classification

After identifying RF a better classifier than SVM, we migrated the existing samples to the target year (2021) for classification. Samples that had incomplete time series due to cloud contamination were discarded and the effect of sample size on the RF classifier was evaluated with different scenario combinations (Table 5). The full real sample dataset of 2021 was used to verify the classification accuracy. In the single-year comparison, the migrated samples in 2018 had the highest classification accuracy (91.11%) and also the highest number of migrated samples. The migrated samples in 2020 had the lowest classification accuracy (84.85%). Similarly, the accuracy of S5 for the 2020 and 2017 combinations was lower. Figure 7 shows that the results of scenario S3 were significantly different from the other classification results because of a serious mixing of corn and



Figure 7. Classification results of random forest for samples of different scenarios. (a) Scenario for s1; (b) Scenario for s2; (c) Scenario for s3; (d) Scenario for s4; (e) Scenario for s5; (f) Scenario for s6; (g) Scenario for s7; (1) and (2) are the details display of two representative areas.

Table 4. Overall Accuracy (OA) and Kappa coefficient of support vector machines (SVM) and random forests (RF) classification for different years.

Year	SVM		RF	
	OA(%)	Kappa	OA(%)	Kappa
2017	83.78	0.80	95.72	0.94
2018	85.71	0.83	97.21	0.96
2020	88.26	0.85	98.13	0.98

Table 5. Classification Overall Accuracy (OA) and Kappa coefficient of sample migration in different sampling scenarios.

Scenario	Number of valid samples	OA(%)	Kappa
S1	166	88.08	0.86
S2	216	91.11	0.90
S3	186	84.85	0.82
S4	382	92.19	0.91
S5	352	88.47	0.86
S6	402	92.06	0.91
S7	568	94.13	0.93

soybean, with only 52.94% user accuracy. There were only 25 training samples for soybean, 85 for corn, 55 for rice, and 21 for other crops. The number of samples for soybean was too small and out of balance with the actual data, thus leading to lower classification accuracy. When we enter all the migrated samples into the classifier, the accuracy reached 94.13% accuracy. To some extent, as the number of samples increases, the classification accuracy of RF increases.

5. Discussion

In this study, we proposed a historical sample-based crop classification map method for years lacking ground-truth samples. We extracted the standard curves of historical samples as reference curves and calculated their similarity to the temporal curves of corresponding image elements on the images of the target year. Then, we performed label matching according to the rules and use the labelled pixels as potential samples for the target year, and finally input them into the classifier for classification.

This method can provide a new classification solution for areas lacking samples of classification years, save on expensive sampling and labour costs, and has important implications for sustainable crop mapping. Our proposed method does not need debugging and setting thresholds for each category like the traditional DTW. Our method focuses more on extracting the most similar categories for matching, which is more convenient and faster to compute. Unlike Zhang (Zhang et al. 2019), we do not need multi-year crop classification results to extract image elements with no change in crop type. Continuous multi-year classification results are not only difficult to obtain but also not applicable in areas with frequent crop rotations. Compared with model migration, the results of training model migration are not satisfactory due to the differences between different years. Model migration methods need to consider the timing and quality of

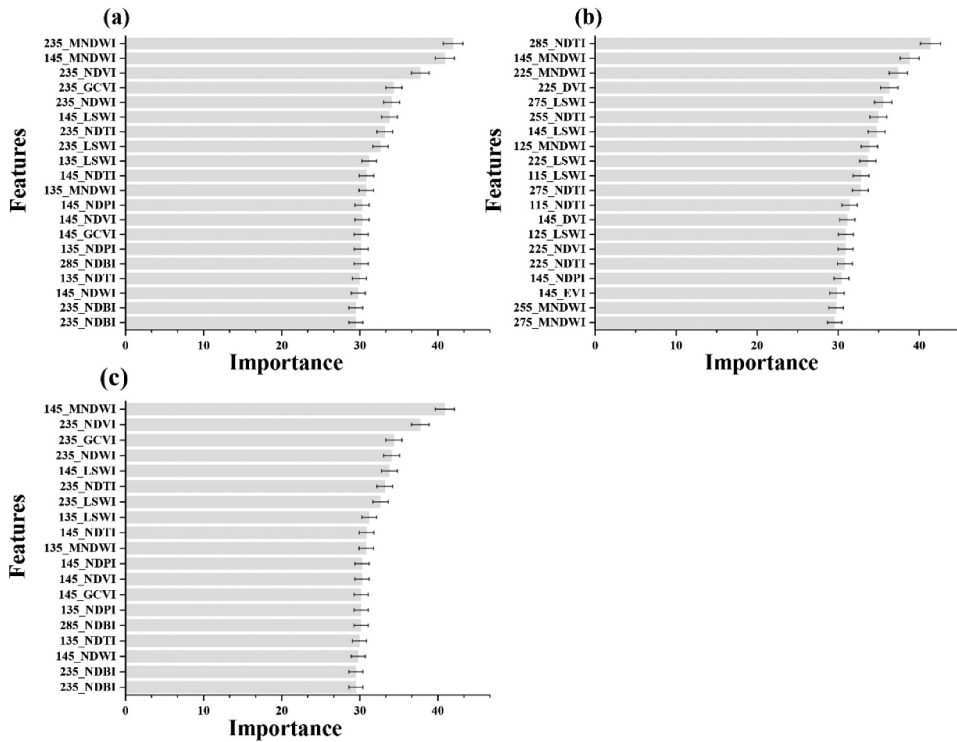


Figure 8. Top 20 features of importance in the random forest classification. (a)2017; (b)2018; (c)2020.

images between different years, which is often difficult to meet in most regions. Our proposed method, which calculates the similarity of label matching by comparing the shape of interannual phenology curves, can minimize the reliance on temporal matching.

We set up seven scenarios to explore the effect of variation in sample size on classification accuracy. The results indicate that the classification performance after migration of a single year is generally inferior to that after migration of multiple years. S2 in a single year outperforms S1 and S3, while the best performance is achieved by S7 in multiple years, with an overall accuracy of 94.13%. The worst performance is from S3, mainly due to the severely imbalanced sample type after migration in 2020, resulting in a suboptimal training of the model. Within the range of available sample numbers, the overall accuracy shows a steady upward trend as the training sample size increases, and the representativeness of the sample is also a significant factor. Overall our method can achieve a migration accuracy of more than 95% for major crop samples.

We also tried other vegetation indices to select the optimal few features. We extracted the top 20 features from the RF classification for each year and analysed these features to find that MNDWI, NDVI, LSWI, GCVI, and NDTI outperformed the other vegetation indices, occurring more frequently (Figure 8). Therefore, we used them as features for the final construction of the phenology curves. In terms of temporal characteristics, the highest frequency occurred when Julian's days (day of year, DOY) were 135, 145, 235 and 245. During DOY 130–155, rice in Heilongjiang is in the transplanting stage, when the soil water content in paddy fields is significantly higher than the soil or leaf water content in

drylands, giving the moisture-sensitive MNDWI and LSWI higher importance. During DOY 225–255, rice is in the reproductive growth stage, which is the most significant crop timing difference in the whole year.

Several limitations of our training sample migration study must be acknowledged. In addition, due to cloudy and rainy weather, many image elements lacked temporal information to calculate curve similarity. This phenomenon persisted despite the five-day return period of the S2-A/B images we used. Fewer images were available than we had previously envisioned for a sufficiently dense time series. In subsequent work, we can try additional vegetation indices and use the Time-Weighted DTW methods to control the temporal weights and improve curve similarity. Additionally, because the density of the time series is important for classification accuracy, we will try to incorporate multi-source data to fill gaps created by cloudiness. Additionally, due to the existence of unknowns crop types, we cannot guarantee the number and representativeness of the potential samples obtained by migrating the samples for a single year. For example, the lower accuracy of migrated classification in 2020 is because the training sample of soybean is too little and the model appears to be under-fitted. We can augment our sample size by combining historical samples from many years to increase the sample representativeness and minimize this phenomenon.

6. Conclusions

In this study, we proposed a new method for migrating training samples by exploiting the similarity of spectral feature profiles of historical samples. The most attractive aspect of this method is that it does not require continued field collection efforts as it makes full use of historical samples, significantly reducing the cost of continuous multi-year mapping.

The initial application of the approach in Hailun City shows encouraging success. We first selected the five best-performing vegetation indices from commonly used ones as the features for our subsequent curve construction. Subsequently, we analysed the phenology curves of different crops for three years (2017, 2018, and 2020) with different crop sensitivity to each vegetation index. We also compared the sample migration accuracy of the extracted phenology curves of a single year and the fused curves of multiple years. The sample migration was more robust for the fused curves. The comparison of two commonly used machine learning models showed the RF model outperforming SVM in the single-year classification task. The classification accuracy increases as the number of input migration samples gradually increases, and the classification accuracy of samples migrated in a single year is lower than that of scenarios where multiple years are combined.

To verify the generalizability of the method, experiments were conducted in the city of Fujin City, a typical district of the Three Rivers Plain, using the same process for their migration experiments. When the consistency was set to 3, the accuracy was only 80%, but when the consistency was increased, the improvement in accuracy was significant. When the consistency was set to 4, the accuracy reached 89%, and when the consistency was set to 5, the accuracy reached 95%. Although the accuracy is not as high as the training area when the consistency is low, the accuracy will increase as the consistency is continuously improved. When the consistency is set to the maximum, the accuracy can be nearly the same as the training area. In general, our approach is feasible and has potential

in areas of the north-east with similar cropping structures, but further exploration is needed in complex areas. To achieve better applications in Heilongjiang and even in the Northeast, future improvements can be made in two ways: adding more standard weathering curves for more crop types; expert knowledge of local agricultural practices, crop spectral monitoring, and simulation need to be taken into account to determine the curve similarity distance thresholds.

With the development of remote sensing technology, the temporal, spatial, and spectral resolution of the sensors is increasing, the data sources are becoming more and more abundant and our options are more diverse. We believe that there is still great potential for the use of historical samples with the advancement of satellite sensor

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The work was supported by the the Central Public-interest Scientific Institution Basal Research Fund [CAAS-ZDRW202107]; the Fundamental Research Funds for Central Non-profit Scientific Institution under grant [1610132021010]; the 2022 Heilongjiang Unveiled Project in Command [2021ZXJ05A0504]; Major S&T project (Innovation 2030) of China under grant [2021ZD0113704].

References

- Chandrasekar, K., M. V. R. Sesha Sai, P. S. Roy, and R. S. Dwevedi. 2010. "Land Surface Water Index (LSWI) Response to Rainfall and NDVI Using the MODIS Vegetation Index Product." *International Journal of Remote Sensing* 31 (15): 3987–4005. doi:10.1080/01431160802575653.
- Costa, W. S., L. M. G. Fonseca, T. S. Korting, H. D. N. Bendini, and R. Cartaxo Modesto De Souza. 2018. "Spatio-Temporal Segmentation Applied to Optical Remote Sensing Image Time Series." *IEEE Geoscience and Remote Sensing Letters* 15 (8): 1299–1303. doi:10.1109/lgrs.2018.2831914.
- Ding, H., G. Trajcevski, P. Scheuermann, X. Wang, and E. Keogh. 2008. "Querying and Mining of Time Series Data: Experimental Comparison of Representations and Distance Measures." *Proceedings of the VLDB Endowment* 1 (2): 1542–1552. doi:10.14778/1454159.1454226.
- Dong, Q., X. Chen, J. Chen, C. Zhang, L. Liu, X. Cao, Y. Zang, X. Zhu, and X. Cui. 2020. "Mapping Winter Wheat in North China Using Sentinel 2A/B Data: A Method Based on Phenology-Time Weighted Dynamic Time Warping." *Remote Sensing* 12 (8): 1274. doi:10.3390/rs12081274.
- Fan, L., J. Yang, X. Sun, F. Zhao, S. Liang, D. Duan, H. Chen, L. Xia, J. Sun, and P. Yang. 2022. "The Effects of Landsat Image Acquisition Date on Winter Wheat Classification in the North China Plain." *Isprs Journal of Photogrammetry and Remote Sensing* 187: 1–13. doi:10.1016/j.isprsjprs.2022.02.016.
- Gella, G. W., W. Bijker, and M. Belgiu. 2021. "Mapping Crop Types in Complex Farming Areas Using SAR Imagery with Dynamic Time Warping." *Isprs Journal of Photogrammetry and Remote Sensing* 175: 171–183. doi:10.1016/j.isprsjprs.2021.03.004.
- Gitelson, A. A., A. Viña, T. J. Arkebauer, D. C. Rundquist, G. Keydan, and B. Leavitt. 2003. "Remote Estimation of Leaf Area Index and Green Leaf Biomass in Maize Canopies." *Geophysical Research Letters* 30 (5): 1248. doi:10.1029/2002gl016450.
- Guan, X., C. Huang, G. Liu, X. Meng, and Q. Liu. 2016. "Mapping Rice Cropping Systems in Vietnam Using an NDVI-Based Time-Series Similarity Measurement Based on DTW Distance." *Remote Sensing* 8 (1): 19. doi:10.3390/rs8010019.

- Huang, Y., B. Qiu, C. Chen, X. Zhu, W. Wu, F. Jiang, D. Lin, and Y. Peng. 2022. "Automated Soybean Mapping Based on Canopy Water Content and Chlorophyll Content Using Sentinel-2 Images." *International Journal of Applied Earth Observation and Geoinformation* 109: 102801. doi:[10.1016/j.jag.2022.102801](https://doi.org/10.1016/j.jag.2022.102801).
- Huang, H., J. Wang, C. Liu, L. Liang, C. Li, and P. Gong. 2020. "The Migration of Training Samples Towards Dynamic Global Land Cover Mapping." *Isprs Journal of Photogrammetry and Remote Sensing* 161: 27–36. doi:[10.1016/j.isprsjprs.2020.01.010](https://doi.org/10.1016/j.isprsjprs.2020.01.010).
- Hu, Q., Y. Ma, B. Xu, Q. Song, H. Tang, and W. Wu. 2018. "Estimating Sub-Pixel Soybean Fraction from Time-Series MODIS Data Using an Optimized Geographically Weighted Regression Model." *Remote Sensing* 10 (4): 491. doi:[10.3390/rs10040491](https://doi.org/10.3390/rs10040491).
- Mao, D., Z. Wang, B. Du, L. Li, Y. Tian, M. Jia, Y. Zeng, K. Song, M. Jiang, and Y. Wang. 2020. "National Wetland Mapping in China: A New Product Resulting from Object-Based and Hierarchical Classification of Landsat 8 OLI Images." *Isprs Journal of Photogrammetry and Remote Sensing* 164: 11–25. doi:[10.1016/j.isprsjprs.2020.03.020](https://doi.org/10.1016/j.isprsjprs.2020.03.020).
- Mariotto, I., P. S. Thenkabail, A. Huete, E. T. Slonecker, and A. Platonov. 2013. "Hyperspectral versus Multispectral Crop-Productivity Modeling and Type Discrimination for the Hyspri Mission." *Remote Sensing of Environment* 139: 291–305. doi:[10.1016/j.rse.2013.08.002](https://doi.org/10.1016/j.rse.2013.08.002).
- Maus, V., G. Camara, R. Cartaxo, A. Sanchez, F. M. Ramos, and G. R. De Queiroz. 2016. "A Time-Weighted Dynamic Time Warping Method for Land-Use and Land-Cover Mapping." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 9 (8): 3729–3739. doi:[10.1109/jstars.2016.2517118](https://doi.org/10.1109/jstars.2016.2517118).
- Maxwell, A. E., T. A. Warner, and F. Fang. 2018. "Implementation of Machine-Learning Classification in Remote Sensing: An Applied Review." *International Journal of Remote Sensing* 39 (9): 2784–2817. doi:[10.1080/01431161.2018.1433343](https://doi.org/10.1080/01431161.2018.1433343).
- Mountrakis, G., J. Im, and C. Ogole. 2011. "Support Vector Machines in Remote Sensing: A Review." *Isprs Journal of Photogrammetry and Remote Sensing* 66 (3): 247–259. doi:[10.1016/j.isprsjprs.2010.11.001](https://doi.org/10.1016/j.isprsjprs.2010.11.001).
- Petitjean, F., J. Inglada, and P. Gancarski. 2012. "Satellite Image Time Series Analysis Under Time Warping." *IEEE Transactions on Geoscience and Remote Sensing* 50 (8): 3081–3095. doi:[10.1109/tgrs.2011.2179050](https://doi.org/10.1109/tgrs.2011.2179050).
- Petitjean, F., and J. Weber. 2014. "Efficient Satellite Image Time Series Analysis Under Time Warping." *IEEE Geoscience and Remote Sensing Letters* 11 (6): 1143–1147. doi:[10.1109/lgrs.2013.2288358](https://doi.org/10.1109/lgrs.2013.2288358).
- Quemada, M., W. D. Hively, C. S. T. Daughtry, B. T. Lamb, and J. Shermeyer. 2018. "Improved Crop Residue Cover Estimates Obtained by Coupling Spectral Indices for Residue and Moisture." *Remote Sensing of Environment* 206: 33–44. doi:[10.1016/j.rse.2017.12.012](https://doi.org/10.1016/j.rse.2017.12.012).
- Roujean, J. L., and F. M. Breon. 1995. "Estimating PAR Absorbed by Vegetation from Bidirectional Reflectance Measurements." *Remote Sensing of Environment* 51 (3): 375–384.
- Sheykhoum, M., M. Mahdianpari, H. Ghanbari, F. Mohammadimanesh, P. Ghamisi, and S. Homayouni. 2020. "Support Vector Machine versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 13: 6308–6325. doi:[10.1109/jstars.2020.3026724](https://doi.org/10.1109/jstars.2020.3026724).
- Song, X. -P., P. V. Potapov, A. Krylov, L. King, C. M. Di Bella, A. Hudson, A. Khan, B. Adusei, S. V. Stehman, and M. C. Hansen. 2017. "National-Scale Soybean Mapping and Area Estimation in the United States Using Medium Resolution Satellite Imagery and Field Survey." *Remote Sensing of Environment* 190: 383–395. doi:[10.1016/j.rse.2017.01.008](https://doi.org/10.1016/j.rse.2017.01.008).
- Wen, Y., X. Li, H. Mu, L. Zhong, H. Chen, Y. Zeng, S. Miao, et al. 2022. Mapping Corn Dynamics Using Limited but Representative Samples with Adaptive Strategies. *Isprs Journal of Photogrammetry and Remote Sensing* 190: 252–266. doi:[10.1016/j.isprsjprs.2022.06.012](https://doi.org/10.1016/j.isprsjprs.2022.06.012).
- Wen, C., M. Lu, Y. Bi, S. Zhang, B. Xue, M. Zhang, Q. Zhou, and W. Wu. 2022. "An Object-Based Genetic Programming Approach for Cropland Field Extraction." *Remote Sensing* 14 (5), 1275. doi:[10.3390/rs14051275](https://doi.org/10.3390/rs14051275).
- Xia T., Z. He, Z. Cai, C. Wang, W. Wang, J. Wang, Q. Hu and Q. Song. 2022. "Exploring the potential of Chinese GF-6 images for crop mapping in regions with complex agricultural landscapes."

- International Journal of Applied Earth Observation and Geoinformation* 107: 102702. doi:[10.1016/j.jag.2022.102702](https://doi.org/10.1016/j.jag.2022.102702).
- Xiong, J., P. S. Thenkabail, M. K. Gumma, P. Teluguntla, J. Poehnelt, R. G. Congalton, K. Yadav, and D. Thau. 2017. "Automated cropland mapping of continental Africa using Google Earth Engine cloud computing." *ISPRS Journal of Photogrammetry and Remote Sensing* 126: 225–44. doi:[10.1016/j.isprsjprs.2017.01.019](https://doi.org/10.1016/j.isprsjprs.2017.01.019).
- Xu, H. 2007. "Modification of Normalised Difference Water Index (NDWI) to Enhance Open Water Features in Remotely Sensed Imagery." *International Journal of Remote Sensing* 27 (14): 3025–3033. doi:[10.1080/01431160600589179](https://doi.org/10.1080/01431160600589179).
- Xuan, F., Y. Dong, J. Li, X. Li, W. Su, X. Huang, J. Huang, et al. 2023. Mapping Crop Type in Northeast China During 2013–2021 Using Automatic Sampling and Tile-Based Image Classification. *International Journal of Applied Earth Observation and Geoinformation* 117: 103178. doi:[10.1016/j.jag.2022.103178](https://doi.org/10.1016/j.jag.2022.103178).
- Xue, Z., P. Du, and L. Feng. 2014. "Phenology-Driven Land Cover Classification and Trend Analysis Based on Long-Term Remote Sensing Image Series." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7 (4): 1142–1156. doi:[10.1109/jstars.2013.2294956](https://doi.org/10.1109/jstars.2013.2294956).
- Xu, J., J. Yang, X. Xiong, H. Li, J. Huang, K. C. Ting, Y. Ying, and T. Lin. 2021. "Towards Interpreting Multi-Temporal Deep Learning Models in Crop Mapping." *Remote Sensing of Environment* 264: 112599. doi:[10.1016/j.rse.2021.112599](https://doi.org/10.1016/j.rse.2021.112599).
- You, N., J. Dong, J. Huang, G. Du, G. Zhang, Y. He, T. Yang, Y. Di, and X. Xiao. 2021. "The 10-M Crop Type Maps in Northeast China During 2017–2019." *Scientific Data* 8 (1): 41. doi:[10.1038/s41597-021-00827-9](https://doi.org/10.1038/s41597-021-00827-9).
- You, X., J. Meng, M. Zhang, and T. Dong. 2013. "Remote Sensing Based Detection of Crop Phenology for Agricultural Zones in China Using a New Threshold Method." *Remote Sensing* 5 (7): 3190–3211. doi:[10.3390/rs5073190](https://doi.org/10.3390/rs5073190).
- Zeng, Y., D. Hao, A. Huete, B. Dechant, J. Berry, J. M. Chen, J. Joiner, et al. 2022. "Optical Vegetation Indices for Monitoring Terrestrial Ecosystems Globally." *Nature Reviews Earth & Environment* 3 (7): 477–493. doi:[10.1038/s43017-022-00298-5](https://doi.org/10.1038/s43017-022-00298-5).
- Zhang, L., Z. Liu, D. Liu, Q. Xiong, N. Yang, T. Ren, C. Zhang, X. Zhang, and S. Li. 2019. "Crop Mapping Based on Historical Samples and New Training Samples Generation in Heilongjiang Province, China." *Sustainability* 11 (18): 5052. doi:[10.3390/su11185052](https://doi.org/10.3390/su11185052).
- Zhang, D., Y. Pan, J. Zhang, T. Hu, J. Zhao, N. Li, and Q. Chen. 2020. "A Generalized Approach Based on Convolutional Neural Networks for Large Area Cropland Mapping at Very High Resolution." *Remote Sensing of Environment* 247: 111912. doi:[10.1016/j.rse.2020.111912](https://doi.org/10.1016/j.rse.2020.111912).
- Zhang, X., B. Wu, G. Ponce-Campos, M. Zhang, S. Chang, and F. Tian. 2018. "Mapping Up-To-Date Paddy Rice Extent at 10 M Resolution in China Through the Integration of Optical and Synthetic Aperture Radar Images." *Remote Sensing* 10 (8): 1200. doi:[10.3390/rs10081200](https://doi.org/10.3390/rs10081200).
- Zhong, L., L. Hu, L. Yu, P. Gong, and G. S. Biging. 2016. "Automated Mapping of Soybean and Corn Using Phenology." *Isprs Journal of Photogrammetry and Remote Sensing* 119: 151–164. doi:[10.1016/j.isprsjprs.2016.05.014](https://doi.org/10.1016/j.isprsjprs.2016.05.014).