



Transfer Learning for Crop classification with Cropland Data Layer data (CDL) as training samples

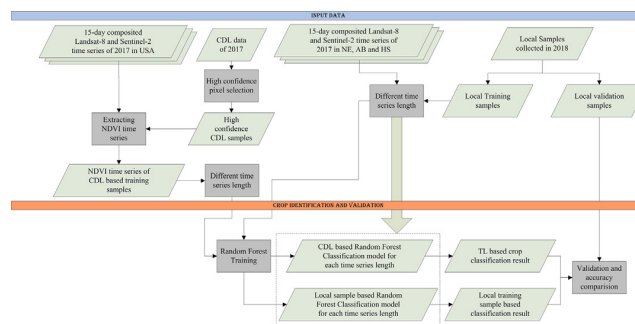
Pengyu Hao, Liping Di *, Chen Zhang, Liying Guo

Center for Spatial Information Science and Systems, George Mason University, Fairfax, Virginia, USA

HIGHLIGHTS

- TL workflow showed good potential to identify crops without local training samples.
- TL need longer time series than LO to generate proper classification accuracy.
- Operational mapping of TL depends on crop variations between training and test region.

GRAPHICAL ABSTRACT



ARTICLE INFO

Article history:

Received 14 January 2020
Received in revised form 28 March 2020
Accepted 19 April 2020
Available online 28 April 2020

Editor: Christian Herrera

Keywords:

Transfer learning
Random Forest
Cropland Data Layer (CDL)
USA
Cotton
Corn

ABSTRACT

Training samples is fundamental for crop mapping from remotely sensed images, but difficult to acquire in many regions through ground survey, causing significant challenge for crop mapping in these regions. In this paper, a transfer learning (TL) workflow is proposed to use the classification model trained in contiguous U.S.A. (CONUS) to identify crop types in other regions. The workflow is based on fact that same crop growing in different regions of world has similar temporal growth pattern. This study selected high confidence pixels across CONUS in the Cropland Data Layer (CDL) and corresponding 30-m 15-day composited NDVI time series generated from harmonized Landsat-8 and Sentinel-2 (HLS) data as training samples, trained the Random Forest (RF) classification models and then applied the models to identify crop types in three test regions, namely Hengshui in China (HS), Alberta in Canada (AB), and Nebraska in USA (NE). NDVI time series with different length were used to identify crops, the effect of time-series length on classification accuracies were then evaluated. Furthermore, local training samples in the three test regions were collected and used to identify crops (LO) for comparison. Results showed that overall classification accuracies in HS, AB and NE were 97.79%, 86.45% and 94.86%, respectively, when using TL with NDVI time series of the entire growing season for classification. However, LO could achieve higher classification accuracies earlier than TL. Because the training samples were collected across USA containing multiple growth conditions, it increased the potential that the crop growth environment in test regions could be similar to those of the training samples; but also led to situation that different crops had similar NDVI time series, which caused lower TL classification accuracy in HS at early-season. Generally, this study provides new options for crop classification in regions of training samples shortage.

© 2020 Elsevier B.V. All rights reserved.

* Corresponding author.

E-mail addresses: phao@gmu.edu (P. Hao), ldi@gmu.edu (L. Di), czhang11@masonlive.gmu.edu (C. Zhang), lguo2@gmu.edu (L. Guo).

1. Introduction

The fast growing population and climate change put great pressure on global food security (Bajzelj et al., 2014; Tilman et al., 2011), and there is an urgent need for timely and accurately mapping crop-type distribution around the world to support the crop growth monitor, yield prediction, and global food-security decision making (Gallego et al., 2008; Lobell, 2013). In recent years, satellite observations have been widely used for land cover mapping globally due to its advantage of low cost complete spatial coverage, and relatively fine spatial resolution (Chen et al., 2015; Gong et al., 2013; Gumma et al., 2017; Teluguntla et al., 2017; Xiong et al., 2017; Yu et al., 2013; Yu et al., 2014). Particularly, because each crop has its unique phenology (Wardlow and Egbert, 2008) and the phenological differences among the crops could be described by time series of remotely sensed images (Zhang et al., 2014a), the satellite time-series images have been widely used for annual crop classification (Bargiel, 2017; Skakun et al., 2017; Zhong et al., 2019).

Most existing annual crop classifications use the image time series of the entire growing season to identify crop (Loosvelt et al., 2012; Low et al., 2013; Wardlow et al., 2007; Zhong et al., 2019). This approach produces the crop distribution map after a growing season finishes (called after-season crop map hereafter). Although they have many usages in agricultural and socioeconomic decision making, the after-season crop maps cannot contribute to many agricultural and socioeconomic activities and decision making, such as crop yield forecasting (Franch et al., 2015; López-Lozano et al., 2015). To test the potential of crop mapping with remote sensing early in a growing season (called early-season here for the simplicity), several studies have been conducted in different regions by evaluating the effect of time-series length on crop classification performance (Song et al., 2017; Villa et al., 2015). For example, Skakun et al. (2017) have used MODIS NDVI time-series and growing degree days (GDD) information to separate winter wheat from spring and summer crops 6–8 weeks before the winter wheat harvest. For the corn and soybean which are more confusing phenologically than winter wheat and summer crops, Cai et al. (2018) have found that these two crops could be accurately discriminated in late July at field scale. Furthermore, Hao et al. (2015) have used MODIS data to select the optimal features for in-season crop classification in Kansas, USA, and concluded that the major crop could be accurately identified in August and longer image time series cannot further improve the classification accuracy and certainty. A similar in-season crop classification study conducted in Hebei, China has showed similar result as the major crops of the study region could be accurately identified 4–8 weeks before the their harvest (Hao et al., 2018a). These two studies indicate that the summer crops could be separated before harvest at regional scale.

Ground samples of the mapping year are important input for crop classification; particularly, Hao et al. (2018b) have suggested collecting training samples in multiple time phases to further improve the classification accuracies. But acquiring ground samples annually through field surveys is challenging because field survey is labor intensive and very expensive (Zhong et al., 2012). In addition, it is impossible for conducting field surveys in many parts of world due to political, administrative, financial, or logistic reasons. To identify crops in areas where the ground samples were not available, Wang et al. (2019) have used unsupervised learning in conjunction with aggregated crop statistics to generate 30 m resolution crop types maps in Midwestern United States. Dong et al. (2016) have proposed a phenological rule-based method which using the special phenological characters of single-season rice to identify paddy rice in Northeast Asia, but the rules can only be used to identify only specific type. Since the phenology of the same crop is more similar than that of different crops, Zhong et al. (2014) have proposed a phenology-based method to identify corn and soybean in Kansas using Landsat image time series. Furthermore, as NDVI time series can be used to describe the phenological difference among different crops, Hao et al. (2016a) have proposed a reference

NDVI time series based method (RBM), which generates reference NDVI time series from adjacent years, the reference NDVI time series can be used to identify crop types and generate training samples in the classification year (Hao et al., 2016b). All these methods are designed for applications in specific study region annually and even for a specific crop type. The spatial extendibility of models, i.e., training a model in one region and applying the trained model in other regions for crop mapping, is impossible by the design or has not been evaluated (Zhang et al., 2019c). Therefore, in region where the ground samples are not available or in shortage, the in-season crop mapping is still a big challenge.

It is a well-known fact that phenological and growing patterns of a crop are the same or very similar in the different regions of the world. Based on this fact, it is reasonable to hypothesize that a supervised classification model trained in one region can be applied to other regions to map the crops common in both the training and applied regions. Since major types of crops, e.g., corn, soybean, wheat, and rice, are distributed globally, it will make the global in-season mapping of major crops possible if the hypothesis of this study is proved to be true. The key is to generate a classification model which are robust enough to tolerant the difference in the crop growth environment and a training region where abundant ground samples are openly available. One of such regions is the USA where the annual crop mapping for the entire contiguous U.S.A. (CONUS) has been conducted since 2008 by the US Department of Agriculture (USDA) National Agricultural Statistics Service (NASS) (Boryan et al., 2011). The result Cropland Data Layer (CDL) product is freely available through the CropScape web service system (Han et al., 2014). Validation of CDL products showed that the accuracy for the major crops (i.e., Corn, Soybeans, and wheat) is around ~95% (Boryan et al., 2012).

The availability of high-accurate Cropland Data Layer (CDL) data provides alternative way for obtaining ground truth data. Several studies have used CDL as training and validation samples for crop classification (Hao et al., 2016b; Zhong et al., 2016; Zhong et al., 2014). Because the current-year CDL is not available to the public until the February/March of the next year, the in-season crop mapping by directly using CDL as ground samples is unfeasible. In order to overcome this problem, Zhang et al. (2019b) have proposed the “trusted pixel” approach, which identifies pixels with stable multiple-year crop rotation patterns in the CDL data and use the rotation patterns to predict the crop type for the pixels in the coming year. Meanwhile, the CDL data also contains quality layer, which indicates the classification confidence of each CDL pixels. Therefore, it is possible to use CDL high-confidence pixels as the ground truth for in-season crop mapping. However, the probability of using CDL data as the training data source for crop identification in other regions has not been explored. Thus, the objective of this paper are (1) to propose a transfer-learning (TL) workflow for crop classification in the ground sample shortage regions, by training a machine-learning classifier with CDL high-confidence pixels and corresponding remote sensing images and using this trained classifier to identify crop types in other study regions, (2) to estimate the effect of time series length on classification accuracies, and evaluate the potential of in-season crop identification using TL. The performances of the TL are evaluated in three representative study regions, Alberta Canada, Nebraska USA, and Hengshui (HS) China.

2. Study region and dataset

Three study regions have been selected in this study (Fig. 1a), and the choice of these sites is oriented towards having a range of major crops. The first study region (110°–112°W, 52°30′–54°N, Fig. 1b) is located in Alberta, Canada, which is a transition zone between mixed-grass prairie and montane cordillera, and cropland constitutes nearly 60% of the area (Zhang et al., 2014b). According to Fisette et al. (2013), the major crops in this study region are spring wheat and canola, which take 19% and 18% acreage of the area, respectively.

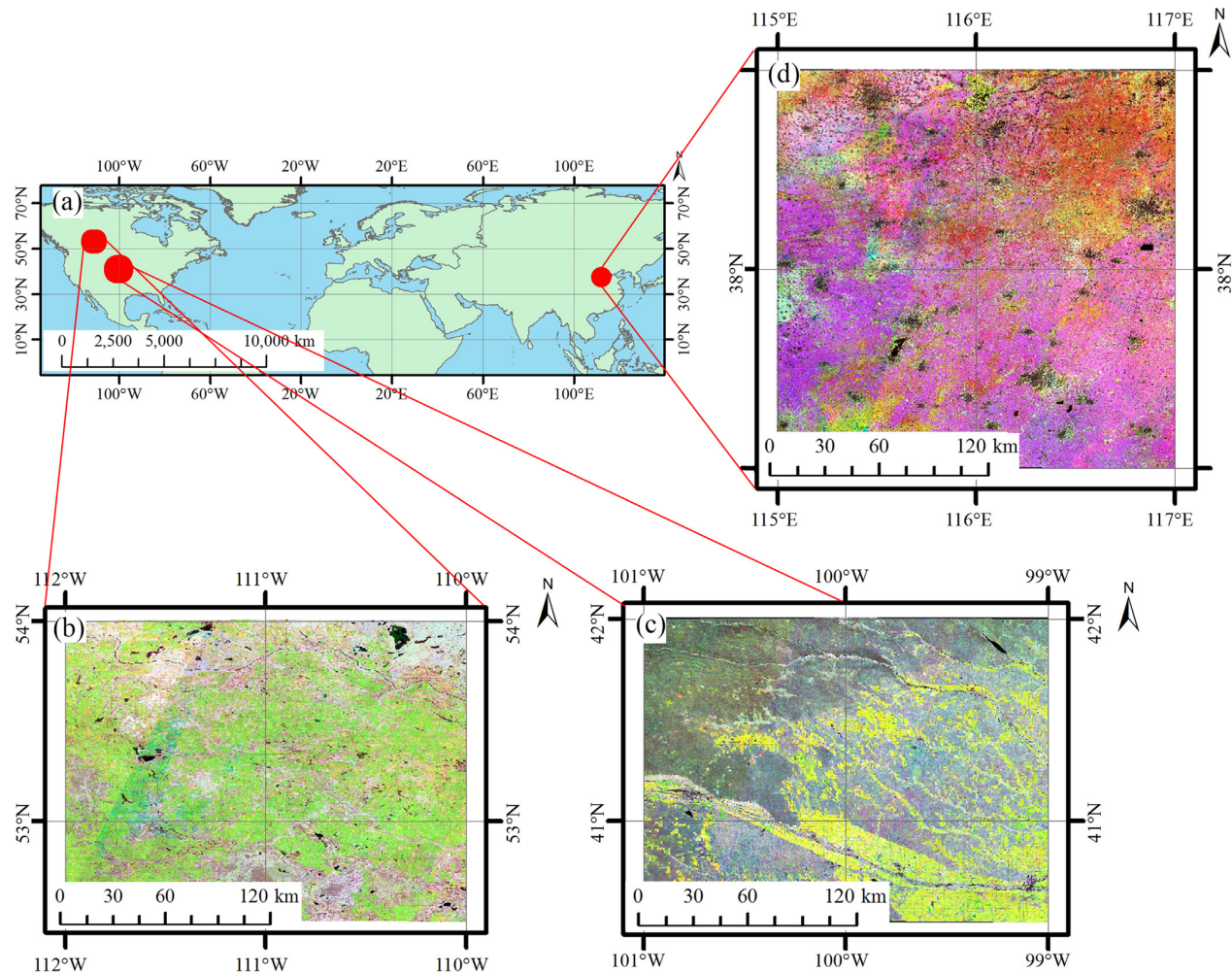


Fig. 1. Study region. (a) location of the three study regions used in this study. (b) Extent of study region AB and its 2018 NDVI image. (c) Extent of study region NE and its 2018 NDVI image. (d) Extent of study region HS and its 2018 NDVI image. The NDVI images of (b), (c) and (d) are composed by R: NDVI in September, G: NDVI in July, B: NDVI in May.

Therefore, we focus on spring wheat and canola in Alberta. The study region has a cold and temperate climate. The dry season is between October and March, monthly precipitation in the dry season are around 20 mm, and the most precipitation falls in June, with an monthly precipitation of 82 mm. The daily average temperatures in July and August are above 15 °C and below −6 °C in January (Diebel et al., 2011).

The second study region (115°–117°E, 37–39°N, Fig. 1d) is located in Hengshui (HS), China, laying in North China Plain. The acreage of farmland is 5.71×10^5 ha (approximately 70.3% of total land area) in Hengshui city (Liu et al., 2015). The study region has a warm temperate continental monsoon climate, which is characterized by hot, rainy summers and cold, dry winters. In summer, average daily temperature between June and August are above 26 °C, and average monthly precipitation in July and August are above 120 mm; in winter, daily average temperature between December and February are around 0 °C, and from November to March, the region have nearly no rainfall (Diebel et al., 2011). The major crops in the study region are winter wheat, corn and cotton, and other crops account for small proportion of the cropland. In this study, we focus on cotton, corn and winter wheat for this region. As nearly all the winter wheat field are used to sown corn after the wheat harvest, we just refer these fields as winter wheat field in this paper.

The third study region (99°–101°W, 40°30′–42°N, Fig. 1c) locates in Nebraska (NE), USA, which is a midwestern state, laying in the western

part of the Corn Belt region. It is one of the top agriculture production state in the United States with total agricultural land of 45.2 million acres, accounting for 91% of the state's total land area. Corn and soybean are two dominant crops in the state. Other crop types, such as alfalfa, sorghum and dry beans, account for a very small proportion of all croplands (Zhang et al., 2019a). In this study, we focus on the two dominant crop types (corn and soybean) in Nebraska. This study region have a the climate is cold and temperate climate, the driest month is January with 20 mm of rainfall, and the most precipitation falls in July with an average of 111 mm; and in July and August, average daily temperature are above 23 °C, and January has the lowest average daily temperature with −5.6 °C. We select NE, which is located in USA, for comparing its classification results with the other two study regions outside USA.

3. Dataset and method

3.1. Methodological overview

The overall workflow of this study is shown in Fig. 2. Firstly, we selected training samples of 2017 from the high confidence pixels (the trusted pixels) of 2017 CDL data (Section 3.2), 15-day NDVI time series of 2017 were then generated using Harmonized Landsat Sentinel-2 (HLS) data in the region where the 2017 training samples located; and

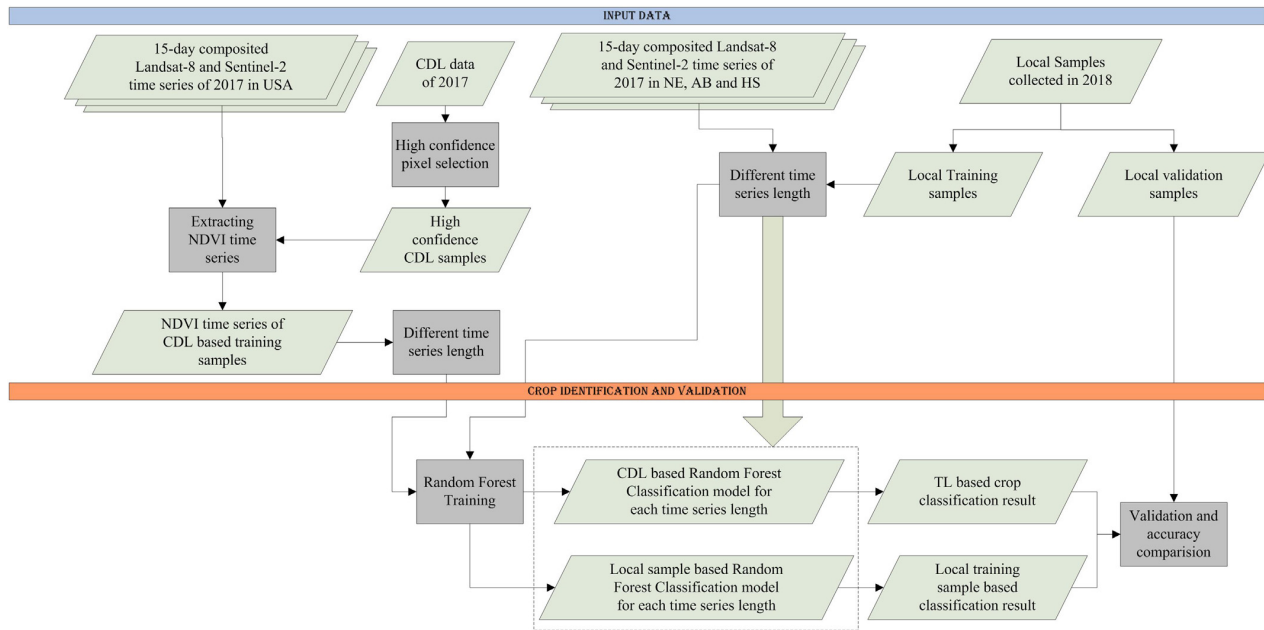


Fig. 2. General workflow of this study.

the 15-day NDVI time series of 2018 were generated using the Landsat-8 and Sentinel-2 data in the three study regions (Section 3.3). Next, NDVI time series of the 2017 training samples were extracted from corresponding 15-day composited NDVI time series data, these training NDVI time series were used to train Random Forest classifiers. We used these classifiers to identify crops in the three study regions with NDVI time series of different length (Section 3.4). Local training and validation samples were collected independently in the three study regions (Section 3.2), local training samples were used to identify crop types using Random Forest with different time-series length. Finally, all classification results were verified using the validation samples and classification accuracies were compared between classification result derived by TL and local training samples.

3.2. Training and validation sample generation

3.2.1. Training samples from CDL data

As we focused on corn, soybean, cotton, spring wheat, canola and winter wheat, we collected higher confidence pixels of these six crops from 2017 CDL data and used these samples as training samples. For winter wheat, samples of winter wheat and double cropping systems which containing winter wheat (such as wheat-corn) are combined as winter wheat samples in this study. Firstly, we generated a high confidence mask using the CDL confidence layer in 2017 and used 95% as the confidence threshold. Next, we acquired the spatial distribution of each crop with the high confidence mask. The training samples were visually selected from the high confidence distribution map of each crop. We generated 100 km * 100 km grid for the US mainland (Fig. 3a), and in each grid, we selected fields larger than 4 ha (containing 7 * 7 Landsat samples) as training target fields. One pixel in each field was selected as a training sample (Fig. 3b), and up to 30 pixels were selected as training samples for each crop type in each grid. Thus, 8448 samples were acquired for the CONUS and the number of training samples for each crop were shown in Table 1.

3.2.2. Local samples in three study regions

We collected local samples for the three study regions in 2018. The local samples for HS were collected by a ground survey, and those for NE and AB were collected from the CDL and AAFC Annual Crop

Inventory Data in 2018, respectively (Fisette et al., 2013). In HS, the fields were surveyed in June 2018, size of all surveyed crop fields were larger than 60 m*60 m to ensure that these surveyed fields can contain pure 30 m Landsat pixels. The crop types and geographic coordinates of surveyed fields were recorded, and the central points of these fields were used as local samples. In total, 1356 samples were collected in HS (Table 2). In NE, we firstly used the 2018 CDL confidence layer to generate high confidence mask using 95% confidence level as the threshold, and then generated high confidence corn and soybean maps in the study region. Next, samples were visually selected from the high confidence corn and soybean maps. This procedure is similar to the training sample selection procedure (Fig. 2b) and a total of 1312 local samples was acquired for NE. In AB, we visually selected local samples from AAFC Annual Crop Inventory Data. Because AAFC did not provide classification confidence layer, we only used crop type layer to select local samples in AB and acquired a total of 1222 samples. For the three study regions, we used 500 samples as local training samples and the other samples as local validation samples. The number of training samples for each crop were calculated based on the ratio of the local sample we collected among these crops, and the samples used as training samples were randomly selected from the local samples. The number of training and validation samples were shown in Table 2.

3.3. Landsat-8 and Sentinel-2 data

We used Harmonized Landsat Sentinel-2 (HLS) data product in this study. As Sentinel-2 MSI and Landsat-8 data have similar spectral bands, NASA's Harmonized Landsat-8 and Sentinel-2 (HLS) project co-registers the two data sources, performs Bidirectional Reflectance Distribution Function (BRDF) normalization and band pass adjustment. Therefore, HLS data provides consistent surface reflectance data at 30-m spatial resolution (Claverie et al., 2018).

In 2017, HLS data covering the spatial distribution of the training samples selected in this study (294 tiles in total) were utilized, For each tile, we collected all HLS data between Day of Year (DOY) 60 and 330, then calculated NDVI for each image with Red and NIR bands (Huete et al., 2002; Rouse et al., 1974), and then generated 15-day NDVI time series by selecting the maximum NDVI value within each

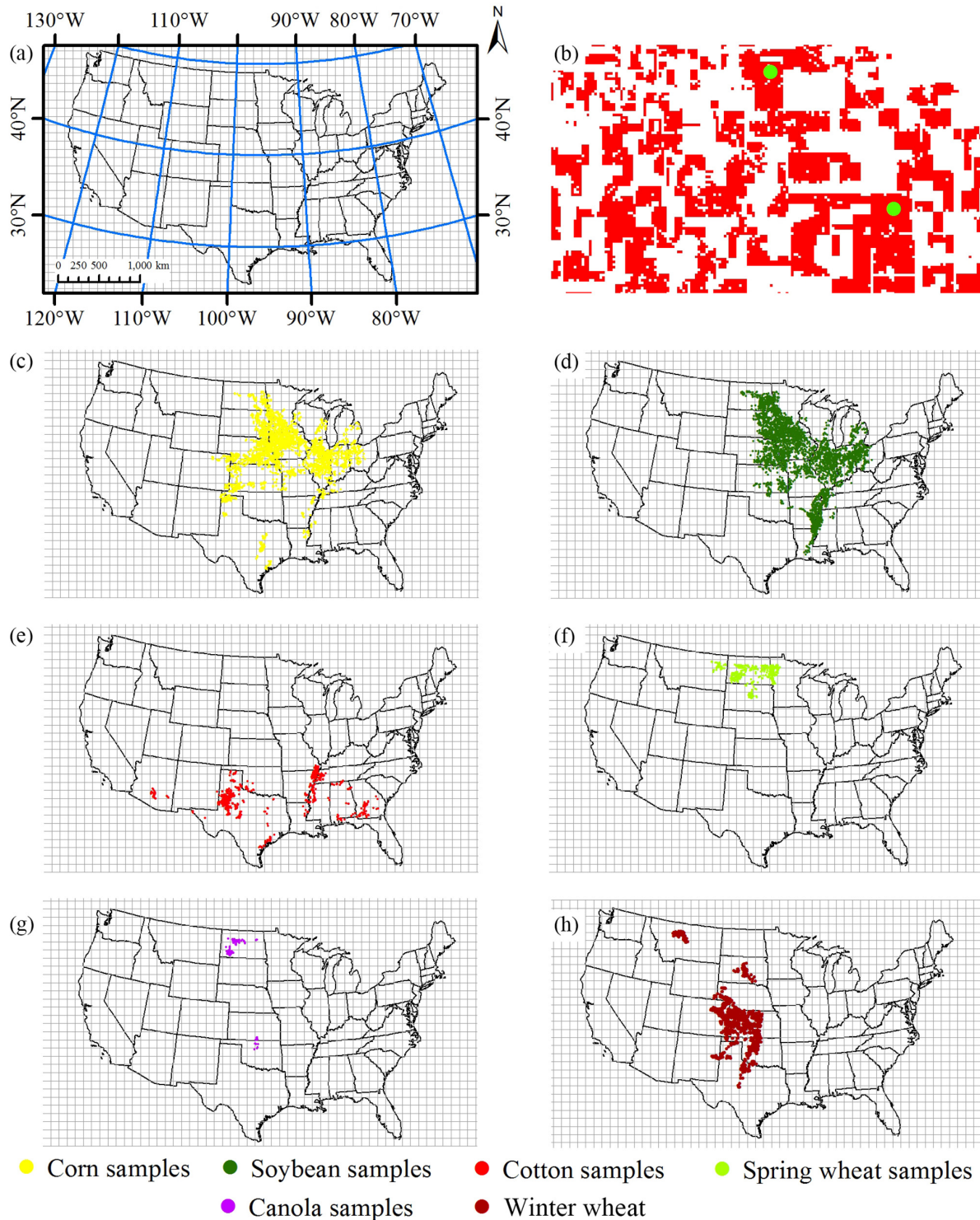


Fig. 3. General workflow of this study. (a) the 100 km × 100 km grids, (b) examples of training samples visually selected from the high confidence crop patterns, (c) spatial distribution of corn training samples, (d) spatial distribution of soybean training samples, (e) spatial distribution of cotton training samples, (f) spatial distribution of spring wheat training samples, (g) spatial distribution of canola training samples, and (h) spatial distribution of winter wheat training samples.

15-day window if there exists available value in the 15-day period. While, if there is no available value, the 15-day composited NDVI of that time period is labeled as “missing value”. We used 15-day NDVI time series in this study because our previous studies showed that

when cloud-free images cannot be acquired daily, 15-day image time series could best describe the crop phenological difference and reduce the number of “missing value”, which is the best choice for in-season crop classification (Hao et al., 2018c). Next, the missing value in the

Table 1
Number of visually selected samples and training samples.

	Corn	Soybean	Cotton	Spring wheat	Canola	Winter wheat	Total
Number of visually selected samples	2759	3135	854	450	120	1130	8448

15-day composited NDVI time series were firstly filled with a moving window method by calculating the average of the two neighboring high quality values in the time series, the gap-filled NDVI time series

were then smoothed using Savitzky-Golay (S-G) filters to further ameliorate irregular variations in the NDVI time series (Chen et al., 2004; Shen et al., 2015; Zhang et al., 2018). The equation for calculating NDVI is shown in Eq. (1).

$$NDVI = \frac{\rho(NIR) - \rho(Red)}{\rho(NIR) + \rho(Red)} \quad (1)$$

where $\rho(NIR)$ and $\rho(Red)$ donate the SR reflectance of NIR and Red band respectively, which are Band 5 and Band 4 of Landsat-8 data or B8 and B4 of Sentinel-2 data. Afterwards, the NDVI time series of all these

Table 2
Number of visually selected samples and training samples.

NE	AB		HS	
	Training	Validation	Training	Validation
Corn	311	521	Corn	72
Soybean	189	335	Corn	72
			Winter wheat	356
Total	500	856	500	722



Fig. 4. Classification accuracy with different time series length using TL and LO in Hengshui. The start point of the time series used for crop mapping is DOY 60, e.g. the 75 in X-axis denotes the time series between DOY 60 and 75 used for crop classification.

training samples were extracted from 15-day composited NDVI image time series in 2017.

In 2018, as the study region AB and NE are covered by the HLS data product (four tiles, 12UVD, 12UVE, 12UWD, 12UWE for AB and four tiles, 14TLL, 14TLM, 14TML, 14TMM for NE), we collected HLS data in these two study regions and generated 15-day time series between DOY 60 and 330, the data processing is same to the 2017 NDVI time series processing. Study region HS is not covered by HLS data product, so that we firstly generated 15-day maximum composited NDVI time series between DOY 60 and 330 of Landsat-8 Surface Reflectance (SR) dataset on Google Earth Engine (GEE) (Google, 2015), the dataset is available as "LANDSAT/LC08/C01/T1_SR" on GEE and 113 Landsat-8 images were collected. As for Sentinel-2 data, L2A product of HS in 2018 was not available, we downloaded Sentinel-2 L1C TOA reflectance from the European Space Agency (ESA) Copernicus Open Access Hub (ESA, 2016). The TOA reflectance were then atmospherically corrected using the Sen2Cor (version 2.8) processor (Main-Knorn et al., 2017). NDVI were calculated and the cloud pixels were masked using the cloud mask of the L1C product. In total, 969 images were collected and then a 15-day composited NDVI time series data were generated. For each 15-day time period, maximum NDVI of Landsat-8 and Sentinel-2 value were calculated as the composited NDVI of the 15-day time period, the NDVI image time series were gap-filled with moving window and then smoothed with S-G filter. Finally, 15-day NDVI time series of training and validation samples in the three study regions were extracted from the corresponding NDVI time series image.

3.4. Crop type classification

In all three study regions, we used the GFSAD 30 data (Teluguntla et al., 2018) to mask out non-cropland area, and then tried to identify crop types within the cropland using Random Forest (RF) classifier (Breiman, 2001). For each study region, CDL-based training samples of the major crops in the study region were used to train the RF classifier. NDVI time series of different length were used as input features and generated different RF classification models. Next, the trained RF classifier was used to identify crop types in the study regions with corresponding NDVI time series, so that we could acquire crop distribution maps generated with different time series length and further estimate the effect of time series length on crop classification accuracy. The time series length ranged from 15-day (NDVI between DOY 60 ~ 75) to 270 day (NDVI time series DOY 60 ~ 330) in all three study regions. We refer these results through transfer learning as TL in the following sections of this paper. In addition, we also used local training samples to identify crops for comparison. After applying cropland extent mask, we used the local training samples to train RF classifiers and then classified crops with NDVI time series of different length in all three study regions. These results were referred as LO in this paper.

RF is a commonly used machine learning classifier for land cover classification (Belgiu and Drăguț, 2016; Maxwell et al., 2018). The RF classifier combines multiple classification trees, during the training procedure, each tree is constructed using two-thirds of the training samples, the remaining one-third of the samples are used for test classification with "out-of-bag error" (OOB error). For the classification procedure, the classification output is determined by the majority vote of the individual classification trees. RF have the advantage of handling high dimensional data effectively. In this study, the RF was implemented using the Random Forest library for R (Breiman et al., 2013). There were two free parameters when implementing RF model; the number of trees (ntree) and the number of features to split the nodes (mtry), the two free parameters were set as 1000 and the square root of the total number of input features. In the training procedure of RF, all 2017 training samples were used, and the input features were the NDVI of the time periods included in the certain time series length. For example, when we tried to use image time series between DOY 60 and 120 to identify crop types, the input features were all 15-

day composited NDVI between DOY 60 and 120. Next, the RF models were used to identify crop types with NDVI data of corresponding time series length.

3.5. Accuracy assessment

Validation samples were used to assess the classification accuracies in both TL and LO with confusion matrix (Congalton, 1991). Three metrics, producer's accuracy (PA), user's accuracy (UA) and accuracy for an individual land-cover class (AIRC) calculated from confusion matrix, were used to evaluate the accuracy of individual classes (Congalton, 1991; Lu et al., 2014). The overall accuracy (OA) was also used to evaluate the classification accuracy of all crops in each study region. Eqs. (2)–(5) were used to calculate PA, UA, AIRC and OA.

$$PA_i = \frac{N_i}{R_i} \quad (2)$$

$$UA_i = \frac{N_i}{C_i} \quad (3)$$

$$AIRC_i = \frac{PA_i \times UA_i}{(PA_i + UA_i)/2} \quad (4)$$

$$OA = \frac{N_C}{N_A} \quad (5)$$

where PA_i and UA_i donate the PA and UA of class i , $AIRC_i$ donates AIRC of class i , N_i donates number of correctly identified validation samples of class i , R_i donates number of validation samples of class i , C_i donates number of validation samples classified as class i , and N_C and N_A donate number of correctly identified and total number of validation samples.

Besides the metrics calculated from confusion matrix, we applied McNemar's test to evaluate the pair-wise statistical significance between different classification approaches (De Wit and Clevers, 2004). McNemar's test is a non-parametric test based on the standardized normal test statistic, calculated as Eq. (6):

$$Z = \frac{f_{12} - f_{21}}{\sqrt{f_{12} + f_{21}}} \quad (6)$$

where f_{12} is the number of samples that are correctly classified by classifier 1 and incorrectly classified by classifier 2; and f_{21} is the number of samples that are correctly classified by classifier 2 and incorrectly classified by classifier 1. We defined three cases of differences in accuracy between classifier 1 and classifier 2 according to significant analysis:

- (1) No significance between classifiers 1 and 2 (N): $-1.96 < Z < 1.96$.
- (2) Positive significance (classifier 1 has higher accuracy than classifier 2) (S+): $Z > 1.96$.
- (3) Negative significance (classifier 1 has lower accuracy than classifier 2) (S-): $Z < -1.96$.

4. Result

4.1. Classification accuracies

Accuracy assessment of Hengshui (HS) study region (Fig. 4) showed that TL had good potential to identify the major crops when using NDVI time series of the entire growing season (between DOY 60 and 330) for classification, and overall classification accuracy (OA) of the study region was 97.79%, and producer's accuracy (PA) and user's accuracy (UA) were 98.26% and 88.98% for cotton, 86.67.38% and 97.85% for corn, and 97.39% and 97.22% for winter wheat. While, when the NDVI time series length was between 120 day (DOY 60–180) and 210 day (DOY 60–270), a lot of corn samples were misclassified as cotton, which led to the low corn PA and cotton UA; and after the NDVI time

series length was longer than 225 day (DOY 60–285), the majority of corn samples were correctly identified. Basically, LO results had higher classification accuracies than the TL results. For example, when NDVI time series of DOY 60–285 were used for classification, PA and UA of LO result were 98.26% and 99.12% for cotton and 98.05% and 99.12% for corn, which were significantly higher than TL derived results (PA and UA for cotton are 98.24% and 79.72%, PA and UA for corn were 72.38% and 98.70%). Wall-to-wall comparison of classification results (Fig. 5) showed that winter wheat could be accurately identified in the three sub-regions at DOY 210 time phase, TL with DOY 60–210 NDVI time series cannot distinguish cotton and corn in HS as a lot of corn pixels were misclassified as cotton. When NDVI time series of the entire growing season (DOY 60–330) were utilized, TL classification map was very similar to LO result in all the three sub-region classification maps, which indicated that TL need longer time series to generate reliable crop classification distributions in HS.

Accuracy assessment in Alberta (Fig. 6) showed TL had potential to identify crops in Alberta. When NDVI time series of the entire growing season (DOY 60–330) were used, OA of the study region was 86.45%, PA and UA were 87.5% and 85.82% for canola, and 85.4% and 87.12% of spring wheat. In contrast, LO achieved significantly higher classification accuracies as OA was 96.55% when DOY 60–330 NDVI time series were used. In addition, both TL and LO result showed that classification accuracy increased when the data of DOY 135–150 and 195–225 were used. OA increased from 50.12% to 72.91% using TL and from 63.05% to 78.45% using LO with DOY 135–150 NDVI, and increased from 76.11% to 83.25% and from 80.30% to 94.09% using LO with DOY 195–225 NDVI. The wall-to-wall comparison of the classification results of TL and LO derived from DOY 60–225 and from DOY 60–330 NDVI time series, along with the AAFC Annual Crop Inventory data, were showed in Fig. 7. LO derived classification maps had similar crop pattern to AAFC result when DOY 60–225 NDVI time series were used in all three sub-regions, but in the TL derived results, some canola patterns were misclassified as spring maize when using DOY 60–225 NDVI time series to identify crop types. Similar to the results in HS, when using NDVI of the entire growing season (DOY 60–330) to identify crops, crop patterns of TL and LO derived results were similar.

Fig. 8 showed that when using entire time series (DOY 60–330) for classification in NE, OA of TL was 94.86%, PA and OA were 91.55% and 99.31% for corn and 98.81% and 88.39% for soybean. And accuracies of LO were higher than TL, entire time series achieved good accuracy as OA was 96.73%, PA and UA were 97.12% and 97.5% for corn and 96.12% and 95.55% for soybean. Among different time phases, images during DOY 135–180 had high contribution to discriminate the two crops. When NDVI of this time phase was used for classification, OA increased from 40.71% to 80.72% for TL and from 58.88% to 92.06% for LO. After that, OA of LO did not continue to increase with longer time series, and OA of TL increased to 92.99% at DOY 270–285 time phase. Similar to the other two study regions (HS and AB), the wall-to-wall comparison (Fig. 9) among CDL data, TL and LO results with DOY 60–330 NDVI time series showed that both TL and LO achieved good classification accuracy as the classification maps in all sub-regions are similar to CDL. For the classification result with DOY 60–180 NDVI time series, both TL and LO misclassified some corn pixels as soybean and TL had more misclassification pixels. These results indicated that the RF classifier trained with training samples of CONUS could separate crops with long NDVI time series, but cannot perform as good as the classifier trained with local training samples for early-season classification.

4.2. Statistical analysis

Fig. 10 showed that McNemar's test results of the TL and LO with different time series length. In HS, Z value was high when the time series length was longer than 90 days (DOY 60–150), which indicated that TL had significant worse classification performance than LO. When the time series length was longer than 210 days (DOY 60–270), Z-value

began to decrease, until when the entire NDVI time series were used for classification, Z-value was near to the significant threshold (1.96), which indicated that LO and TL did not have significantly difference. In AB, LO significantly outperformed TL for almost all time series lengths used for crop classification, except for the time length was between 90 days (DOY 60–150) and 120 days (DOY 60–180), which was exactly time phases which classification accuracy of TL quickly increased (Section 4.1). In NE, classification performances of TL and LO did not have significant difference when time series length was shorter than 100 days (DOY 60–165), and after time series length increased to 120 days (DOY 60–180), the Z value increased and LO significantly outperformed TL.

5. Discussion

5.1. Uncertainty of transfer learning

Although TL showed good potential to identify crops in all three study regions, there were still some misclassification caused by phenological and crop calendar difference among different regions. The monthly air temperature (acquired from ERA Interim monthly average reanalysis product (Dee et al., 2011)) and NDVI time series among cotton and corn planting regions in CONUS and HS are compared (Fig. 11). In CONUS, cotton are mainly planted in Texas, Arizona and Georgia States, where the latitude is lower than the cotton planting regions in HS; so that the air temperature in HS is slightly lower than AZ, TX and GA of USA, and growing season of cotton in HS is longer than those in CONUS. As for corn, HS corn NDVI time series are significantly lower than those in CONUS at early growing season, and corn NDVI in HS between DOY 60 and DOY 225 are quite similar to some CONUS cotton samples, which leads to the misclassification of corn samples as corn with TL (Section 4.1). While, as cotton NDVI remains high at DOY 270–285 time phase and corn NDVI is low, cotton and corn in HS could be separated by TL when using NDVI time series of the entire growing season. This indicates that although cotton and corn NDVI time series have mismatch between training and test region, the two crops could still be separated when long NDVI time series are used.

Air temperature and NDVI time series of spring wheat and canola were compared between CONUS and AB (Fig. 12). The latitude of AB is higher than that of CONUS sample collecting regions, and the air temperature of AB are slightly lower than MT, ND and SD. NDVI of spring wheat in AB is covered by the training spring wheat samples, but there are difference between canola NDVI time series in AB and ND, this difference explains the relatively low PA and UA of TL derived canola. In addition, NDVI time series of canola and spring wheat in AB are similar before DOY 180, which lead to the low accuracy of both TL and LO at early-season. More features which could detect some other characters of these two crops should be further used to improve the separability of canola and spring wheat.

Corn and soybean have phenological difference as the growing season of corn begin earlier than soybean (Fig. 13 b and c), so that the NDVI time series from DOY 150 to 180 could achieve high classification accuracy using local training samples (Fig. 8). However, as the training samples are collected across CONUS and the crop growth conditions varied among the training samples, TL derived crop maps have lower classification accuracies.

Basically, when trying to identify crops in training sample shortage region, one possible solution is to train the classification model in other region/years, and then transfer the model to the test region. During this procedure, the challenge is that the crop condition between the training and test regions are not perfectly match for the same crop, which leads to uncertainty when applying TL in training sample shortage regions. This uncertainty is common because Crop Progress and Condition report showed that in each state, same crop have different growth situations among multiple years (USDA, 2020). Wang et al. (2019) also found that crop classification accuracies are low if the

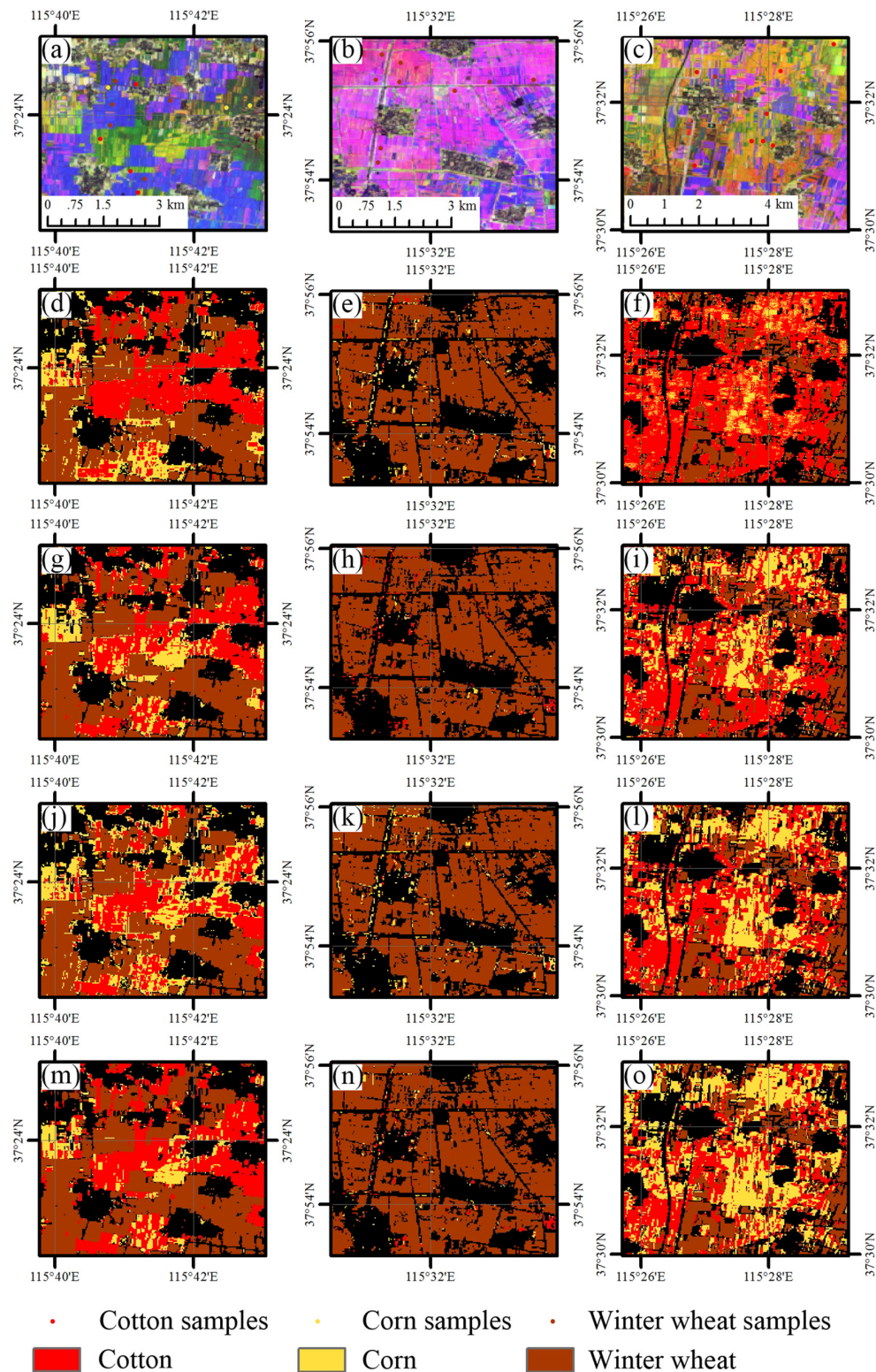


Fig. 5. Wall-to-wall comparison of TL and LO results with DOY 60–210 and DOY 60–330 monthly NDVI time series in HS. (a) (b) and (c) are the NDVI color composite images with the locations of validation samples collected in each sub-region. The NDVI images are composed by R: NDVI of DOY 270, G: NDVI of DOY 210, B: NDVI of DOY 150. (d) (e) and (f) are the classification result derived by TL using DOY 60–210 NDVI time series and (g) (h) and (i) are classification results derived by LO using DOY 60–210 NDVI time series. (j) (k) and (l) are classification results derived by TL using DOY 60–330 NDVI time series and (m) (n) and (o) are classification result derived by LO using DOY 60–330 NDVI time series.

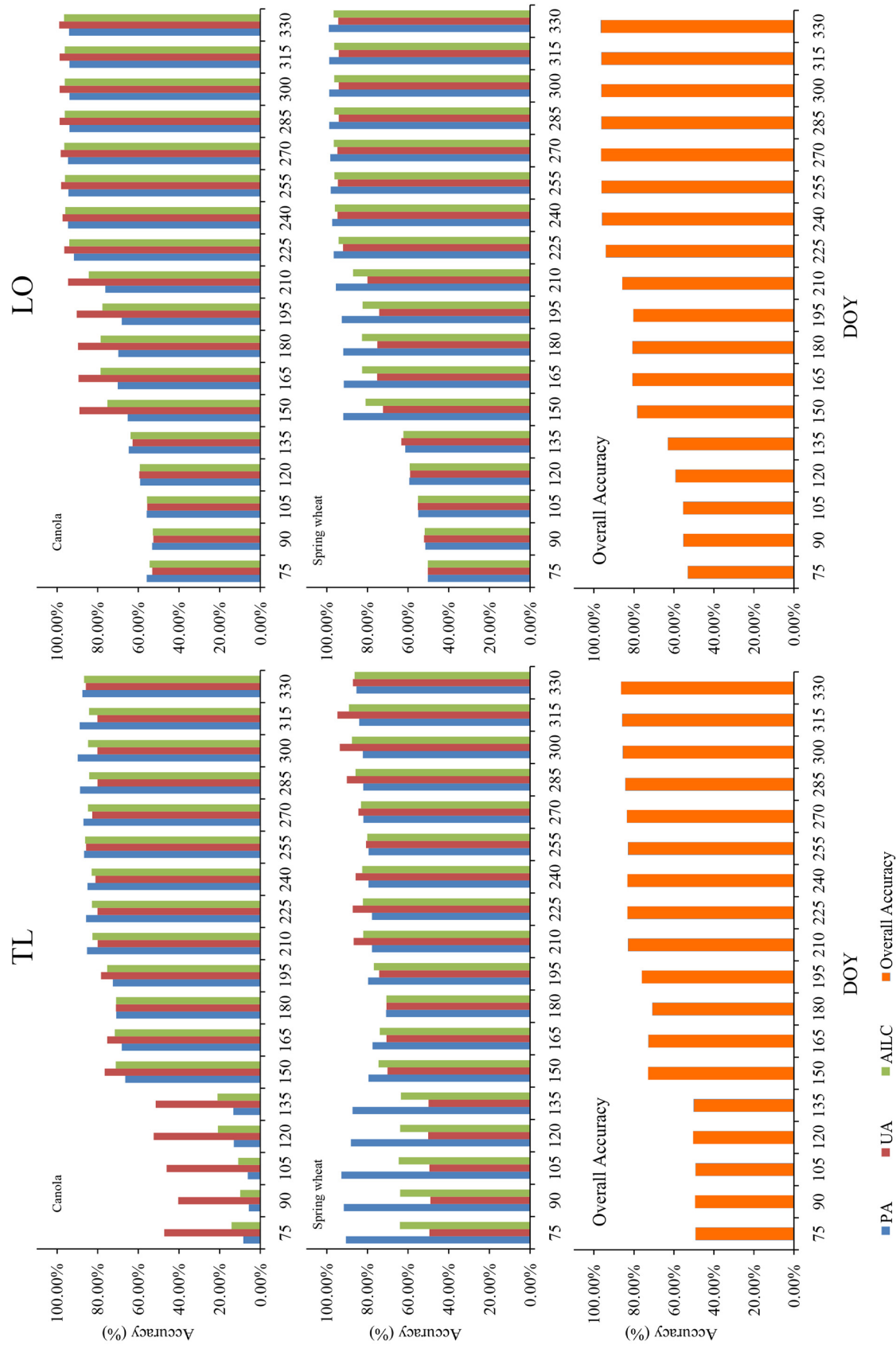


Fig. 6. Classification accuracy with different time series length using Transfer learning (TL) and local training samples (LO) in Alberta. The start point of the time series used for crop mapping is DOY 60, e.g. the 75 in X-axis denotes the time series between DOY 60 and 75 used for crop classification.

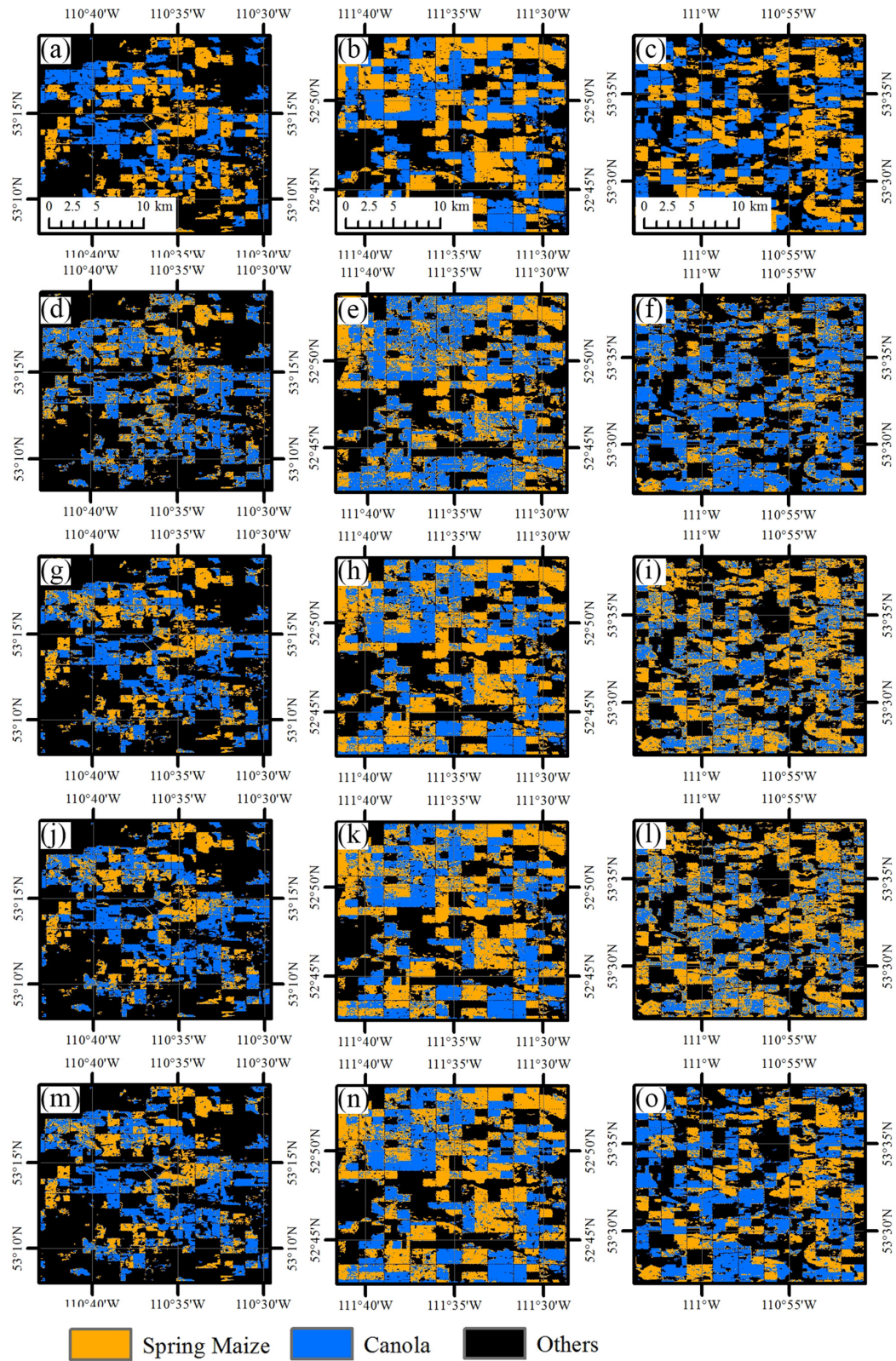


Fig. 7. Wall-to-wall comparison of TL and LO results with DOY 60–225 and DOY 60–330 image time series and AAFC Annual Crop Inventory data in Alberta. (a) (b) and (c) are the location and AAFC Annual Crop Inventory data of each sub-region; (d) (e) and (f) are the classification result derived by TL using DOY 60–225 NDVI time series; (g) (h) and (i) are classification results derived by LO using DOY 60–225 NDVI time series; (j) (k) and (l) are classification results derived by TL using DOY 60–330 NDVI time series; (m) (n) and (o) are classification result derived by LO using DOY 60–330 NDVI time series.

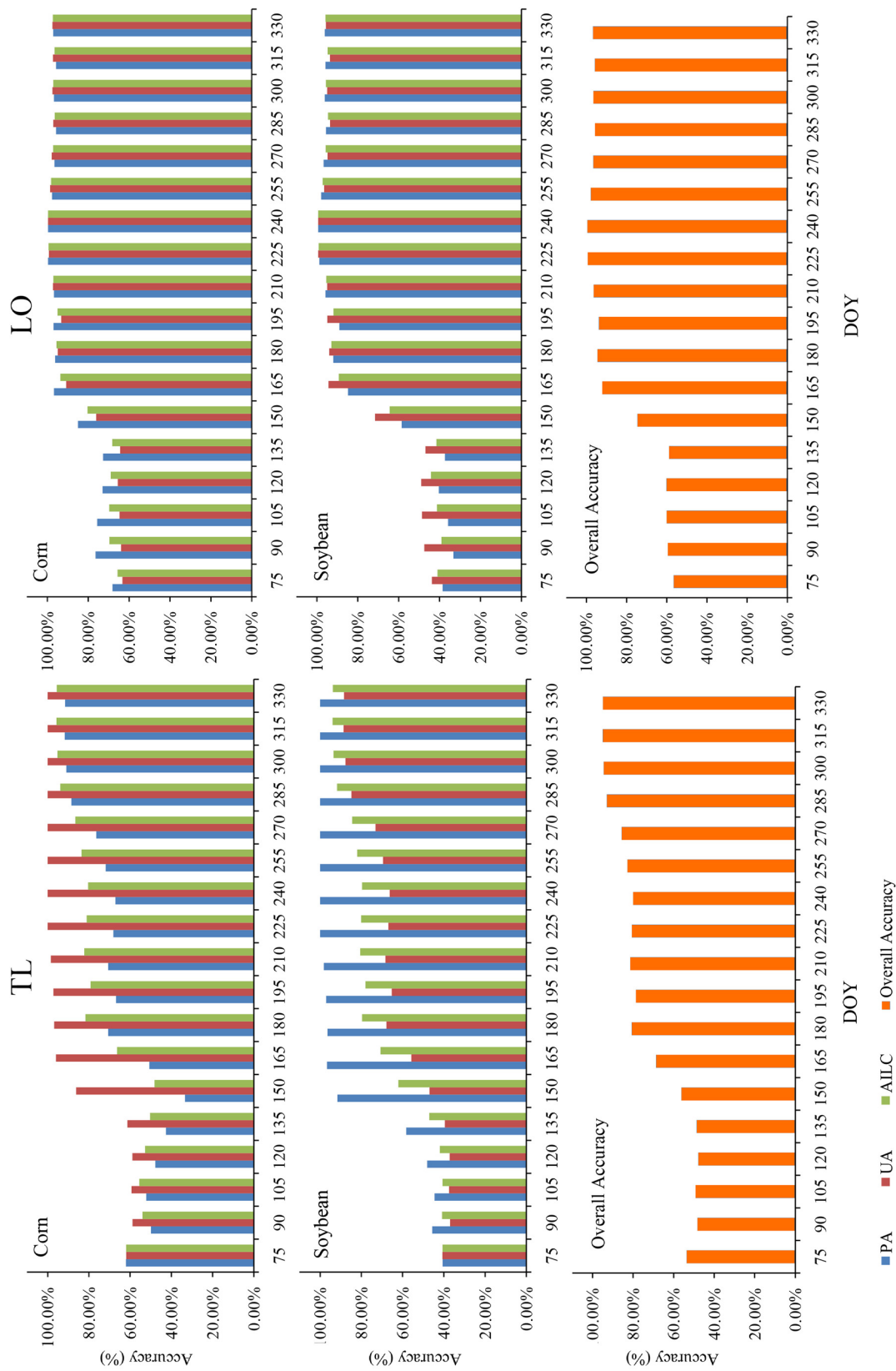


Fig. 8. Classification accuracy with different time series length using Transfer learning (TL) and local training samples (LO) in Nebraska (NE). The start point of the time series used for crop mapping is DOY 60, e.g. the 75 in X-axis denotes the time series between DOY 60 and 75 used for crop classification.

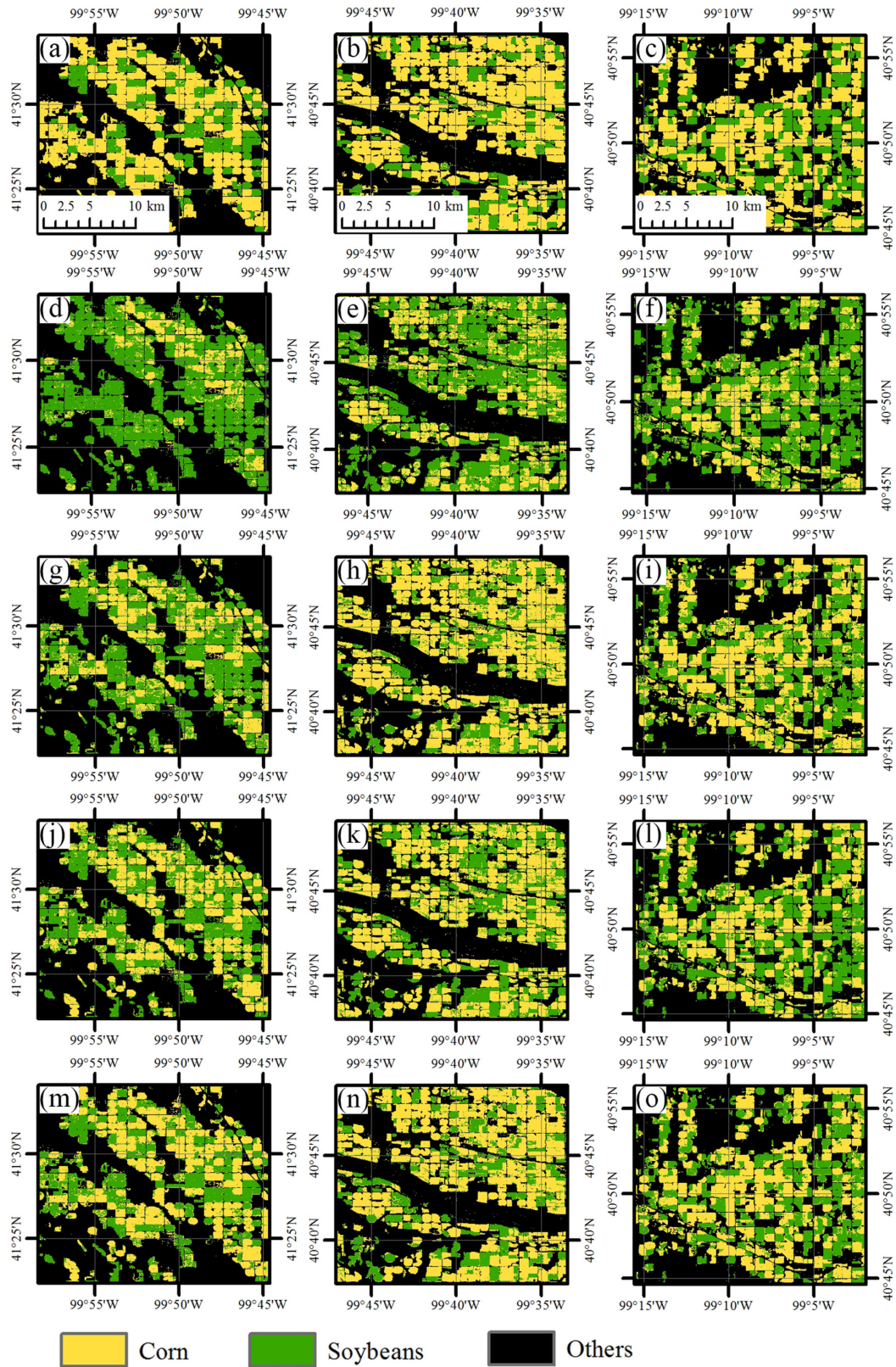


Fig. 9. Wall-to-wall comparison of TL and LO results with DOY 60–180 and 60–330 image time series and CDL data in Nebraska. (a) (b) and (c) are the location and CDL data of each sub-region; (d) (e) and (f) are the classification result derived by TL using DOY 60–180 NDVI time series; (g) (h) and (i) are classification results derived by LO using DOY 60–180 NDVI time series; (j) (k) and (l) are classification results derived by TL using 60–330 NDVI time series; (m) (n) and (o) are classification result derived by LO using DOY 60–330 NDVI time series.

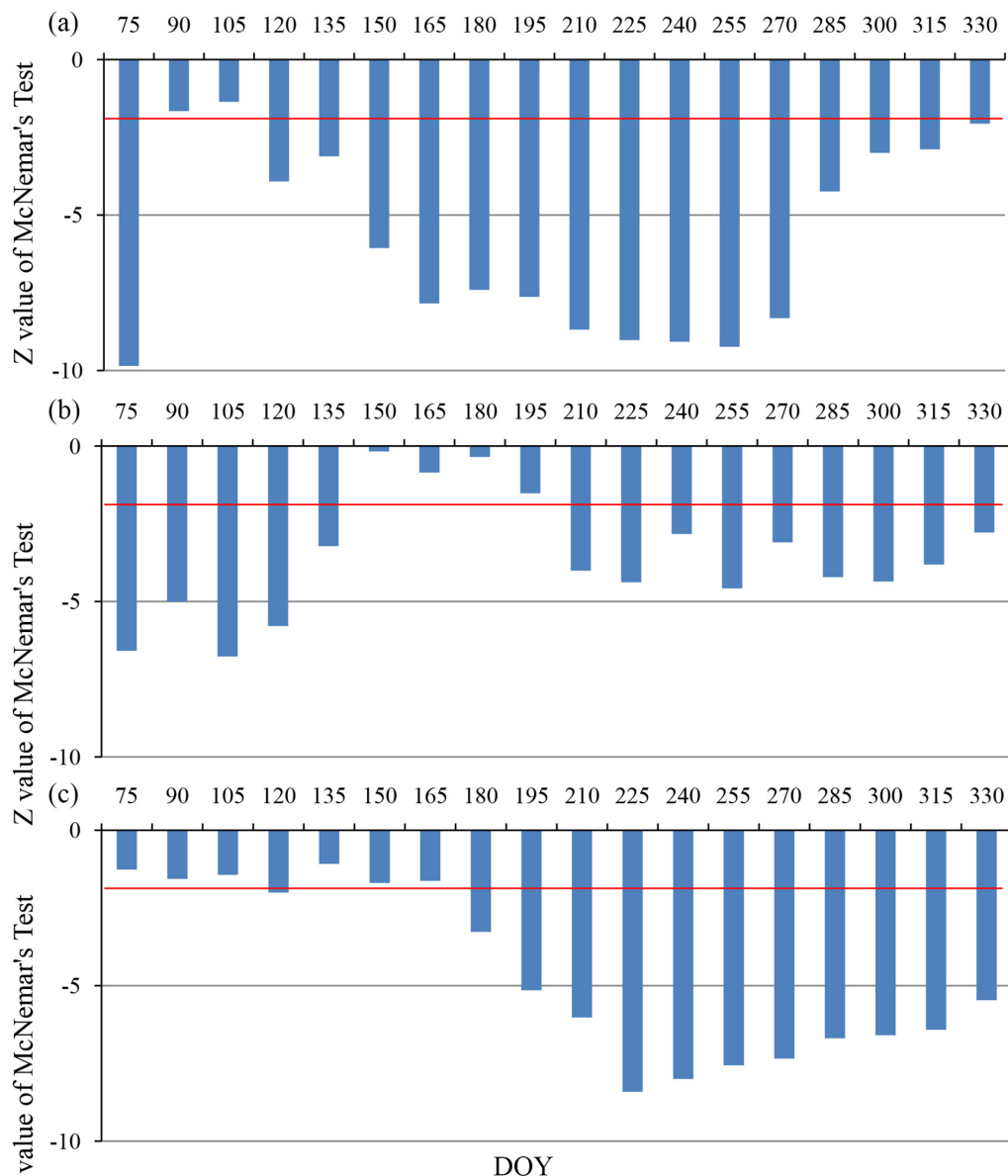


Fig. 10. McNemar's test results between TL and LO with different time series length of the three study regions. (a) McNemar result in HS, (b) McNemar result in AB, (c) McNemar result in NE; red lines in the figure donate the thresholds of positive/negative significant (-1.96 in this study). The start point of the time series used for crop mapping is DOY 60, e.g. the 75 in X-axis denotes the time series between DOY 60 and 75 used for crop classification.

GDD of training and test regions have significantly difference. Furthermore, different crops may have similar NDVI time series and lead to significantly misclassification, just like the low accuracies of cotton and corn in HS identified with short NDVI time series.

5.2. Advantage and limitation

In this study, we collected training samples from CDL data across the CONUS, used these training samples to train Random Forest classifiers, and then transfer these classifiers to identify crops in three test regions. Existing training-sample-free crop classification methods have just tried to transfer the classification model across multiple years in the same region (Hao et al., 2016a), or designed classification model for limited specific crops (Dong et al., 2016; Zhong et al., 2014); the crop growth

difference among region/years is still challenged. Compared with these existing researches, the advantage of TL in this study is that this method collect training samples from CONUS, which have a variety of growth conditions for each crop, this enlarges the potential that the crop condition in the test regions are included in the training sample set; and classification model in this study is applicable for all major crops included in the NASS CDL data. Classification accuracies in this study also showed that TL could generate crop maps with reasonable accuracies in HS and AB, although there were slightly climate differences between the training (CONUS) and test regions (HS and AB).

However, there are still some drawbacks of current TL workflow:

- (1) In this study, we collected training samples across the CONUS so that the training samples contained multiple crop growth

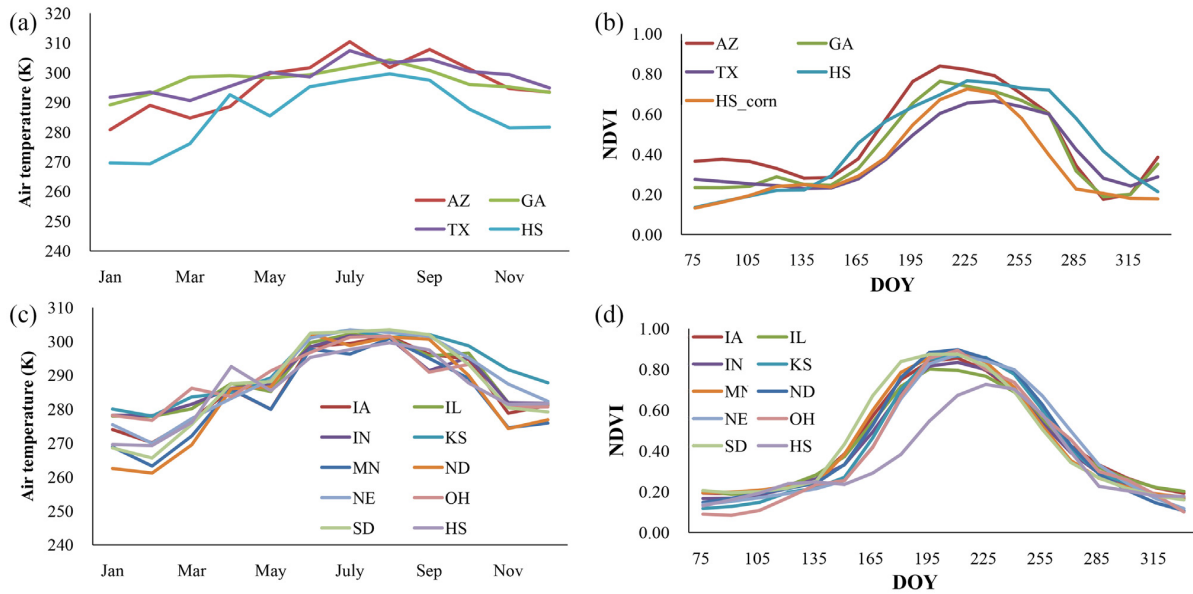


Fig. 11. Air temperature and NDVI time series of HS and CONUS. (a) average monthly air temperature of cotton planting region in CONUS and HS; (b) average cotton NDVI time series in CONUS and HS; (c) air temperature of corn planting region in CONUS and HS; (d) average corn NDVI time series time series in CONUS and HS. In subplot (a) and (c), the data source of air temperature are ERA Interim monthly average reanalysis product (Dee et al., 2011). In subplot (b) and (d), the NDVI time series are the average 15-day NDVI of samples in each state, AZ denotes Arizona, GA denotes Georgia, TX denotes Texas, IA denotes Iowa, IL denotes Illinois, IN denotes Indiana, KS denotes Kansas, MN denotes Minnesota, ND denotes North Dakota, NE denotes Nebraska, OH denotes Ohio, SD denotes South Dakota, and HS denotes the study region Hengshui in this study.

situations, increasing the possibility that the training samples contain the similar NDVI time series to those in the test regions. But this also raise a problem that same crop under different climate conditions may be confused and different crops may have similar NDVI time series, especially when time series length is short. Therefore, the TL cannot perform good at early growing season.

(2) In the test region where the climate condition is slightly different

from CONUS, NDVI of the same crop are still more similar than different crops, so that TL could correctly identify crops with acceptable accuracies, such as cotton in HS and spring wheat in AB, but if the climate and irrigation condition of the test regions are significantly different from the training regions, NDVI time series of the same crop may have larger difference, which will further decrease the classification accuracy.

(3) This study used NDVI time series because NDVI have high

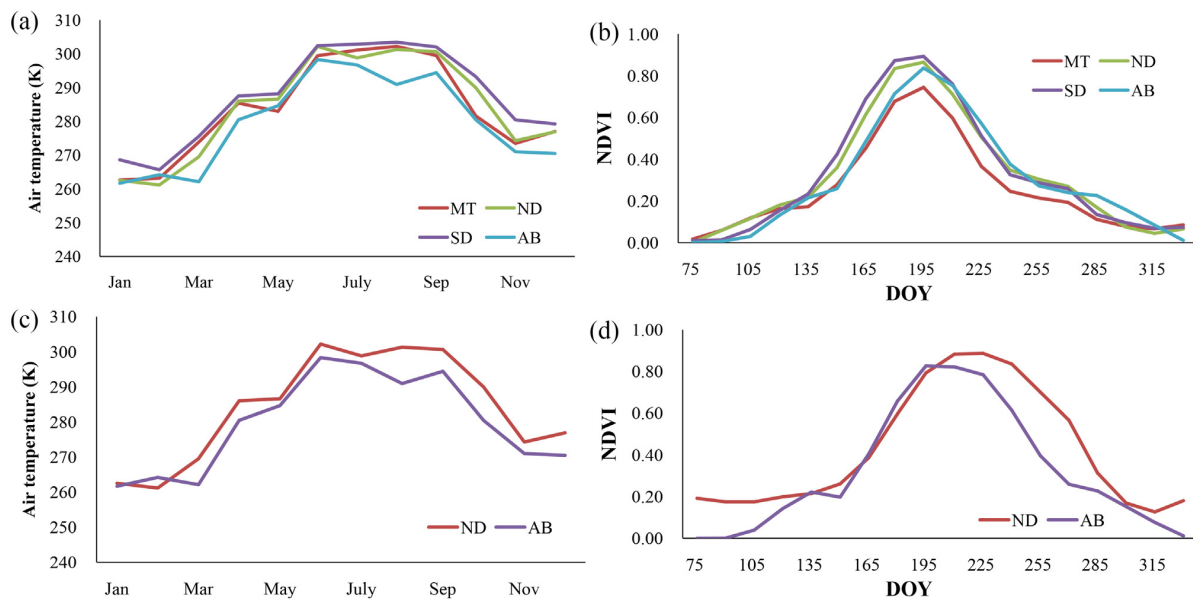


Fig. 12. Air temperature and NDVI time series of AB and CONUS. (a) average monthly air temperature of spring wheat planting region in CONUS and AB; (b) average spring wheat NDVI time series in CONUS; (c) air temperature of canola planting region in CONUS and HS; (d) average canola NDVI time series time series in CONUS and HS. In the subplot (b) and (d), the NDVI time series are the average 15-day NDVI of samples in each state, MT denotes Montana, ND denotes North Dakota, SD denotes South Dakota, and AB denotes the study region Alberta in this study.

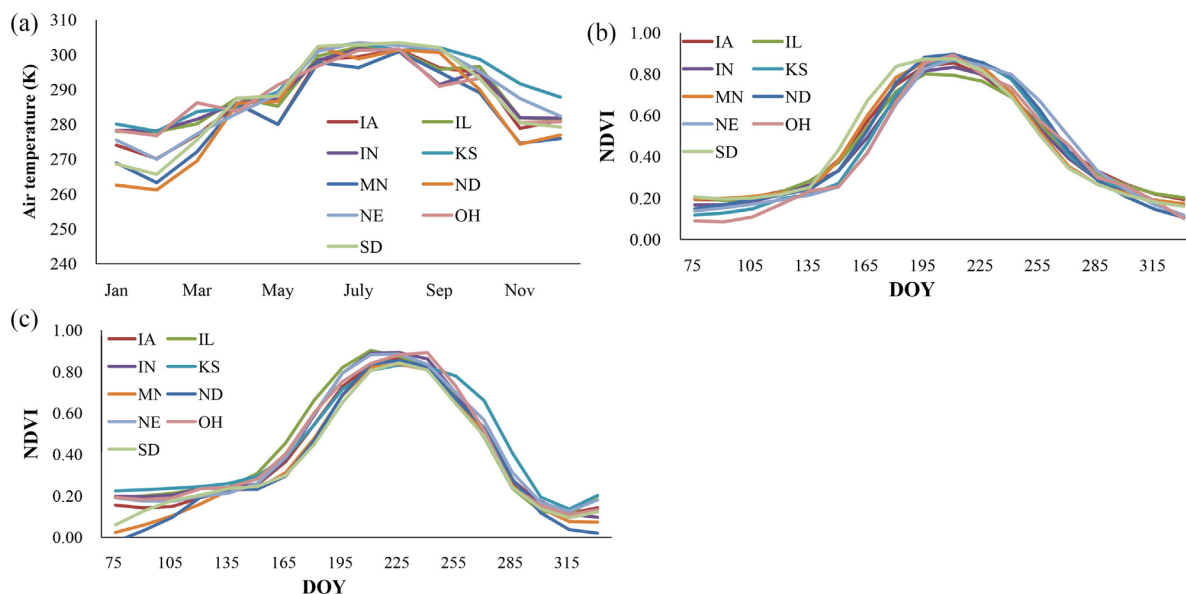


Fig. 13. Air temperature and NDVI time series of different states in CONUS. (a) average monthly air temperature of corn and soybean planting region in CONUS; (b) average corn NDVI time series in CONUS and HS; (c) average soybean NDVI time series in CONUS. In the subplot (b) and (c), the NDVI time series are the average 15-day NDVI of samples in each state, IA denotes Iowa, IL denotes Illinois, IN denotes Indiana, KS denotes Kansas, MN denotes Minnesota, ND denotes North Dakota, NE denotes Nebraska, OH denotes Ohio, SD denotes South Dakota.

contribution for crop classification (Hao et al., 2015), and NDVI is sensitive to the biomass change at green-up stage, which is more suitable for early-season crop classification (Wardlow et al., 2007). Therefore, the results could estimate the potential of TL for crop identification in training sample shortage regions. For some specific crops, such as canola, some other features with high contributions are not included in this study, which may lead to lower classification accuracies of this study.

- (4) The training samples of this study were selected from CDL data, but the remote sensing derived crop type product also contain errors. Therefore, we used the classification confidence threshold 95% to mask out the low confidence pixels and then visually selected training samples from high confidence pixels in large fields. This procedure could effectively reduce the uncertainty of the training data. For the crop classification validation, this study used ground surveyed samples to verify classification result in HS, but in AB and NE, we used CDL and AAFC Annual Crop Inventory Data for validation. The procedure for selecting validation samples in NE is similar to that of training samples, but in AB, the AAFC Annual Crop Inventory Data don't provide confidence layer, so that the accuracy assessment in AB may be slightly affected.

6. Conclusion

This paper proposed a Transfer Learning (TL) workflow, in which classification models were trained with the high confidence CDL pixels and corresponding 15-day composited NDVI time series. Training samples were collected across the CONUS to contain NDVI time series of each crop under different climate and irrigation conditions. The trained classification models were then used for crop classification in other regions. In this study, the performance of this TL workflow was tested in three test regions. The followings are the main conclusions:

- (1) Transfer learning method achieved proper classification results when using NDVI time series of the entire growing season, the OA in HS, AB and NE are 97.79%, 86.45% and 94.86%, respectively.
- (2) In HS and AB, local training samples achieved higher and earlier

classification accuracies than TL because the climate and environment is slightly different between the training region (CONUS) and the test regions, which led to the NDVI time-series mismatch (such as corn in HS) and caused misclassification in TL, but the NDVI time series of the same crop are still more similar than different crops in the most cases.

- (3) In NE, the crop growth conditions are covered by training data set, and LO still had slightly better classification performances than TL. This is because training samples collected in CONUS contained multiple crop conditions, caused confusion for TL and led to misclassification, particularly at early growing season. In contrast, NDVI of training and test samples are match with LO, so that LO performed better than TL.

Further study should focus on identifying and generating new features which are more stable to describe crop growth among different regions, so that the transfer learning workflow could be applied to regions where crop growth environment is significantly different from training regions.

CRedit authorship contribution statement

Pengyu Hao: Conceptualization, Visualization, Formal analysis, Writing - original draft. **Liping Di:** Conceptualization, Writing - review & editing, Funding acquisition. **Chen Zhang:** Data curation, Validation, Writing - original draft. **Liying Guo:** Data curation, Validation, Writing - original draft.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The CDL data used in this paper is provided by USED, NASS, we thanks them for sharing the data free. And this research is supported by the grant NSF INFEWS (grant # CNS-1739705, PI: Prof. Liping Di).

The authors thanks the reviewers for the constructive suggestions and comments.

References

- Bajzelj, B., Richards, K.S., Allwood, J.M., Smith, P., Dennis, J.S., Curmi, E., et al., 2014. Importance of food-demand management for climate mitigation. *Nat. Clim. Chang.* 4, 924–929.
- Bargiel, D., 2017. A new method for crop classification combining time series of radar images and crop phenology information. *Remote Sens. Environ.* 198, 369–383.
- Belgiu, M., Drăguț, L., 2016. Random forest in remote sensing: a review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* 114, 24–31.
- Boryan, C., Yang, Z.W., Mueller, R., Craig, M., 2011. Monitoring US agriculture: the US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program. *Geocarto International* 26, 341–358.
- Boryan, C., Yang, Z.W., Di, L.P., 2012. *Deriving 2011 Cultivated Land Cover Data Sets Using USDA National Agricultural Statistics Service Historic Cropland Data Layers*. 2012 IEEE International Geoscience and Remote Sensing Symposium, pp. 6297–6300.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32.
- Breiman, L., Cutler, A., Liaw, A., Wiener, M., 2013. *Breiman and Cutler's Random Forests for Classification and Regression*.
- Cai, Y.P., Guan, K.Y., Peng, J., Wang, S.W., Seifert, C., Wardlow, B., et al., 2018. A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach. *Remote Sens. Environ.* 210, 35–47.
- Chen, J., Jonsson, P., Tamura, M., Gu, Z.H., Matsuhashita, B., Eklundh, L., 2004. A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky-Golay filter. *Remote Sens. Environ.* 91, 332–344.
- Chen, J., Chen, J., Liao, A., Cao, X., Chen, L., Chen, X., et al., 2015. Global land cover mapping at 30 m resolution: a POK-based operational approach. *ISPRS Journal of Photogrammetry & Remote Sensing* 103, 7–27.
- Claverie, M., Ju, J., Masek, J.G., Dungan, J.L., Vermote, E.F., Roger, J.-C., et al., 2018. The harmonized Landsat and Sentinel-2 surface reflectance data set. *Remote Sens. Environ.* 219, 145–161.
- Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sens. Environ.* 37, 35–46.
- De Wit, A.J.W., Clevers, J.G.P.W., 2004. Efficiency and accuracy of per-field classification for operational crop mapping. *Int. J. Remote Sens.* 25, 4091–4112.
- Dee, D.P., Uppala, S.M., Simmons, A.J., Berrisford, P., Poli, P., Kobayashi, S., et al., 2011. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* 137, 553–597.
- Diebel, J., Nords, J., Kretschmer, O., 2011. *Weather Spark*. p. 2010.
- Dong, J., Xiao, X., Menarguez, M.A., Zhang, G., Qin, Y., Thau, D., et al., 2016. Mapping paddy rice planting area in northeastern Asia with Landsat 8 images, phenology-based algorithm and Google Earth Engine. *Remote Sens. Environ.* 185, 142–154.
- ESA, 2016. *The Copernicus Open Access Hub*. p. 2020.
- Fisette, T., Rollin, P., Aly, Z., Campbell, L., Daneshfar, B., Filyer, P., et al., 2013. *AAFC Annual Crop Inventory*. 2013 Second International Conference on Agro-Geoinformatics. Ieee, New York, pp. 269–273.
- Franch, B., Vermote, E.F., Becker-Reshef, I., Claverie, M., Huang, J., Zhang, J., et al., 2015. Improving the timeliness of winter wheat production forecast in the United States of America, Ukraine and China using MODIS data and NCAR growing degree day information. *Remote Sens. Environ.* 161, 131–148.
- Gallego, J., Craig, M., Michaelsen, J., Bossyns, B., 2008. Best practices for crop area estimation with remote sensing. https://www.earthobservations.org/documents/cop/ag_gams/GEOS%20best%20practices%20area%20estimation%20final.pdf.
- Gong, P., Wang, J., Yu, L., Zhao, Y.C., Zhao, Y.Y., Liang, L., et al., 2013. Finer resolution observation and monitoring of global land cover: first mapping results with Landsat TM and ETM+ data. *Int. J. Remote Sens.* 34, 2607–2654.
- Google, 2015. *Google Earth Engine*.
- Gumma, M.K., Thenkabail, P.S., Teluguntla, P., Oliphant, A.J., Xiong, J., Congalton, R.G., Yadav, K., Phalke, A., Smith, C., 2017. *NASA Making Earth System Data Records for Use in Research Environments (MEaSUREs) Global Food Security-support Analysis Data (GFSAD) @ 30-m for South Asia, Afghanistan and Iran: Cropland Extent Product (GFSAD30SAAFGRICE)*. DAAC NELP.
- Han, W.G., Yang, Z.W., Di, L.P., Zhang, B., Peng, C.M., 2014. Enhancing agricultural geospatial data dissemination and applications using geospatial web services. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7, 4539–4547.
- Hao, P., Zhan, Y.L., Wang, L., Niu, Z., Shakir, M., 2015. Feature selection of time series MODIS data for early crop classification using random Forest: a case study in Kansas, USA. *Remote Sens.* 7, 5347–5369.
- Hao, P., Wang, L., Zhan, Y., Niu, Z., 2016a. Using moderate-resolution temporal NDVI profiles for high-resolution crop mapping in years of absent ground reference data: a case study of bole and Manas Counties in Xinjiang, China. *ISPRS Int. J. Geo Inf.* 5, 67 In this issue.
- Hao, P., Wang, L., Zhan, Y., Wang, C., Niu, Z., Wu, M., 2016b. Crop classification using crop knowledge of the previous year: case study in Southwest Kansas, USA. *European Journal of Remote Sensing* 49, 1061–1077 In this issue.
- Hao, P., Tang, H., Chen, Z., Liu, Z., 2018a. Early-season crop mapping using improved artificial immune network (IAIN) and Sentinel data. *PeerJ* 6.
- Hao, P., Tang, H., Chen, Z., Yu, L., Wu, M., 2018b. A sampling workflow based on unsupervised clusters and multi-temporal sample interpretation (UCMT) for cropland mapping. *Remote Sens. Lett.* 9, 952–961. <https://doi.org/10.1080/2150704X.2018.1500045> In this issue.
- Hao, P., Wu, M., Niu, Z., Wang, L., Zhan, Y., 2018c. Estimation of different data compositions for early-season crop type classification. *PeerJ* 6.
- Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X., Ferreira, L.G., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sens. Environ.* 83, 195–213.
- Liu, H., Guo, H., Yang, L., Wu, L., Li, F., Li, S., et al., 2015. Occurrence and formation of high fluoride groundwater in the Hengshui area of the North China Plain. *Environ. Earth Sci.* 74, 2329–2340.
- Lobell, D.B., 2013. The use of satellite data for crop yield gap analysis. *Field Crop Res.* 143, 56–64.
- Loosvelt, L., Peters, J., Skriver, H., De Baets, B., Verhoest, N.E.C., 2012. Impact of reducing polarimetric SAR input on the uncertainty of crop classifications based on the random forests algorithm. *IEEE Trans. Geosci. Remote Sens.* 50, 4185–4200.
- López-Lozano, R., Duveiller, G., Seguí, L., Meroni, M., García-Condado, S., Hooker, J., et al., 2015. Towards regional grain yield forecasting with 1km-resolution EO biophysical products: strengths and limitations at pan-European level. *Agric. For. Meteorol.* 206, 12–32.
- Low, F., Michel, U., Dech, S., Conrad, C., 2013. Impact of feature selection on the accuracy and spatial uncertainty of per-field crop classification using support vector machines. *ISPRS J. Photogramm. Remote Sens.* 85, 102–119.
- Lu, D.S., Li, G.Y., Moran, E., Dutra, L., Batistella, M., 2014. The roles of textural images in improving land-cover classification in the Brazilian Amazon. *Int. J. Remote Sens.* 35, 8188–8207.
- Main-Knorn, M., Pflug, B., Louis, J., Debaecker, V., Müller-Wilm, U., Gascon, F., 2017. *Sen2Cor for Sentinel-2*. vol 10427. SPIE.
- Maxwell, A.E., Warner, T.A., Fang, F., 2018. Implementation of machine-learning classification in remote sensing: an applied review. *Int. J. Remote Sens.* 39, 2784–2817.
- Rouse, J.W., Haas, R.H., Schell, J.A., Deering, D.W., Harlan, J.C., 1974. Monitoring the Vernal Advancements and Retrogradation of Natural Vegetation. *NASA/GSFC*, pp. 1–137.
- Shen, H., Li, X., Cheng, Q., Zeng, C., Yang, G., Li, H., et al., 2015. Missing information reconstruction of remote sensing data: a technical review. *IEEE Geoscience and Remote Sensing Magazine* 3, 61–85.
- Skakun, S., Franch, B., Vermote, E., Roger, J.-C., Becker-Reshef, I., Justice, C., et al., 2017. Early season large-area winter crop mapping using MODIS NDVI data, growing degree days information and a Gaussian mixture model. *Remote Sens. Environ.* 195, 244–258.
- Song, Q., Hu, Q., Zhou, Q., Hovis, C., Xiang, M., Tang, H., et al., 2017. In-season crop mapping with GF-1/WFV data by combining object-based image analysis and random Forest. *Remote Sens.* 9, 1184.
- Teluguntla, P., Thenkabail, P.S., Xiong, J., Gumma, M.K., Congalton, R.G., Oliphant, A.J., Sankey, T., Poehnell, J., Yadav, K., Massey, R., Phalke, A., Smith, C., 2017. *NASA Making Earth System Data Records for Use in Research Environments (MEaSUREs) Global Food Security-support Analysis Data (GFSAD) @ 30-m for Australia, New Zealand, China, and Mongolia: Cropland Extent Product (GFSAD30AUNZCNMOCE)*. DAAC NELP.
- Teluguntla, P., Thenkabail, P.S., Oliphant, A., Xiong, J., Gumma, M.K., Congalton, R.G., et al., 2018. A 30-m landsat-derived cropland extent product of Australia and China using random forest machine learning algorithm on Google Earth Engine cloud computing platform. *ISPRS J. Photogramm. Remote Sens.* 144, 325–340.
- Tilman, D., Balzer, C., Hill, J., Befort, B.L., 2011. Global food demand and the sustainable intensification of agriculture. *Proc. Natl. Acad. Sci. U. S. A.* 108, 20260–20264.
- USDA N, 2020. *Crop Progress and Condition*. p. 2019.
- Villa, P., Stroppiana, D., Fontanelli, G., Azar, R., Brivio, P., 2015. In-season mapping of crop type with optical and X-band SAR data: a classification tree approach using synoptic seasonal features. *Remote Sens.* 7, 12859–12886.
- Wang, S., Azzari, G., Lobell, D.B., 2019. Crop type mapping without field-level labels: random forest transfer and unsupervised clustering techniques. *Remote Sens. Environ.* 222, 303–317.
- Wardlow, B.D., Egbert, S.L., 2008. Large-area crop mapping using time-series MODIS 250 m NDVI data: an assessment for the U.S. Central Great Plains. *Remote Sens. Environ.* 112, 1096–1116.
- Wardlow, B.D., Egbert, S.L., Kastens, J.H., 2007. Analysis of time-series MODIS 250 m vegetation index data for crop classification in the US Central Great Plains. *Remote Sens. Environ.* 108, 290–310.
- Xiong, J., Thenkabail, P.S., Tilton, J.C., Gumma, M.K., Teluguntla, P., Congalton, R.G., Yadav, K., Dungan, J., Oliphant, A.J., Poehnell, J., Smith, C., Massey, R., 2017. *NASA Making Earth System Data Records for Use in Research Environments (MEaSUREs) Global Food Security-support Analysis Data (GFSAD) @ 30-m for Africa: Cropland Extent Product (GFSAD30AFCE)*. DAAC NELP.
- Yu, L., Wang, J., Clinton, N., Xin, Q., Zhong, L., Chen, Y., et al., 2013. FROM-GC: 30 m global cropland extent derived through multisource data integration. *International Journal of Digital Earth* 6, 521–533.
- Yu, L., Wang, J., Li, X., Li, C., Zhao, Y., Gong, P., 2014. A multi-resolution global land cover dataset through multisource data aggregation. *Science China-Earth Sciences* 57, 2317–2329.
- Zhang, J.H., Feng, L.L., Yao, F.M., 2014a. Improved maize cultivated area estimation over a large scale combining MODIS-EVI time series data and crop phenological information. *ISPRS J. Photogramm. Remote Sens.* 94, 102–113.
- Zhang, X., Huffman, T., Liu, J., Liu, H., 2014b. Soil capability as a predictor of cropland change in Alberta, Canada from 1988 to 2010. *Soil Use Manag.* 30, 403–413.
- Zhang, X., Liu, L., Liu, Y., Jayavelu, S., Wang, J., Moon, M., et al., 2018. Generation and evaluation of the VIIRS land surface phenology product. *Remote Sens. Environ.* 216, 212–229.
- Zhang, C., Di, L., Lin, L., Guo, L., 2019a. Extracting trusted pixels from historical cropland data layer using crop rotation patterns: a case study in Nebraska, USA. 2019 8th International Conference on Agro-Geoinformatics (Agro-Geoinformatics), pp. 1–6.

- Zhang, C., Di, L., Lin, L., Guo, L., 2019b. Machine-learned prediction of annual crop planting in the U.S. Corn Belt based on historical crop planting maps. *Comput. Electron. Agric.* 166, 104989. <https://www.sciencedirect.com/science/article/pii/S0168169919309482>.
- Zhang, L., Liu, Z., Liu, D., Xiong, Q., Yang, N., Ren, T., et al., 2019c. Crop mapping based on historical samples and new training samples generation in Heilongjiang Province, China. *Sustainability* 11, 5052.
- Zhong, L.H., Gong, P., Biging, G.S., 2012. Phenology-based crop classification algorithm and its implications on agricultural water use assessments in California's Central Valley. *Photogramm. Eng. Remote. Sens.* 78, 799–813.
- Zhong, L.H., Gong, P., Biging, G.S., 2014. Efficient corn and soybean mapping with temporal extendability: a multi-year experiment using Landsat imagery. *Remote Sens. Environ.* 140, 1–13.
- Zhong, L., Hu, L., Yu, L., Gong, P., Biging, G.S., 2016. Automated mapping of soybean and corn using phenology. *ISPRS J. Photogramm. Remote Sens.* 119, 151–164.
- Zhong, L., Hu, L., Zhou, H., 2019. Deep learning based multi-temporal crop classification. *Remote Sens. Environ.* 221, 430–443.