International Journal of Applied Earth Observations and Geoinformation 102 (2021) 102374

Contents lists available at ScienceDirect

International Journal of Applied Earth Observations and Geoinformation

journal homepage: www.elsevier.com/locate/jag

Rapid in-season mapping of corn and soybeans using machine-learned trusted pixels from Cropland Data Layer

Chen Zhang [a,b], Liping Di [a,b,*], Pengyu Hao [a], Zhengwei Yang [c], Li Lin [a,b], Haoteng Zhao [a,b], Liying Guo [a]

[a] Center for Spatial Information Science and Systems, George Mason University, Fairfax, VA 22030, USA
[b] Department of Geography and Geoinformation Science, George Mason University, Fairfax, VA 22030, USA
[c] Research and Development Division, U.S. Department of Agriculture National Agricultural Statistics Service, Washington, DC 20250, USA

ARTICLE INFO

Keywords:
Crop mapping
Machine learning
Land use and land cover
Cropland Data Layer
Landsat
Sentinel-2

ABSTRACT

A timely and detailed crop-specific land cover map can support many agricultural applications and decision makings. However, in-season crop mapping over a large area is still challenging due to the insufficiency of ground truth in the early stage of a growing season. To address this issue, this paper presents an efficient machine-learning workflow for the rapid in-season mapping of corn and soybeans fields without ground truth data for the current year. We use trusted pixels, a set of pixels that are predicted from the historical Cropland Data Layer (CDL) data with high confidence in the current year's crop type, to label training samples on multi-temporal satellite images for crop type classification. The entire mapping process only involves a limited number of satellite images acquired within the growing season (normally 3–4 images per scene) and no field data needs to be collected. According to the investigation on 12 states of the U.S. Corn Belt, it is found that a considerable number of trusted pixels can be identified from the historical CDL data by the trusted pixel prediction model based on artificial neural network. According to the experiment on 49 Landsat-8 scenes and 31 Sentinel-2 tiles, the in-season maps of corn and soybeans are expected to reach 85%–95% agreement with CDL as well as field data by mid-July. Once the in-season satellite imagery becomes available, the crop cover map can be rapidly created even with limited computational resources. This study provides a new perspective and detailed guidance for rapid in-season mapping of corn and soybeans, which can be potentially applied to identify more diverse crop types and scaled up to the entire United States.

1. Introduction

Remote sensing has been proven an effective and efficient Earth observation approach for land use and land cover (LULC) mapping and agricultural monitoring (Fritz et al., 2015; Hansen and Loveland, 2012; Lobell, 2013; Mulla, 2013). The crop-specific land cover map derived from the remotely sensed images can provide the fundamental information for crop yield estimation (Prasad et al., 2006; Sakamoto et al., 2014), LULC change (Lark et al., 2017; Liu et al., 2005), food security (Fischer et al., 2014), cropland diversity (Waldner et al., 2016), and many other socio-economic activities. As the most well-known annual agricultural land use map covering the conterminous United States (CONUS), the Cropland Data Layer (CDL) product of United States Department of Agriculture (USDA) National Agricultural Statistics Service (NASS) is produced using satellite imagery from the Landsat 8 OLI/TIRS sensor, the Disaster Monitoring Constellation (DMC) Deimos-1 and UK2, the ISRO ResourceSat-2 LISS-3, and the ESA Sentinel-2 sensors collected during the current growing season (USDA-NASS, 2019a). The CDL data product has been used in many agricultural studies and applications, such as crop loss assessment (Di et al., 2017), crop planting frequency modeling (Boryan et al., 2014), flood impact estimation (Shrestha et al., 2017), and crop area estimation (Song et al., 2017). Similarly, the Annual Crop Inventory (ACI) product of Agriculture and Agri-Food Canada (AAFC) is produced using satellite images from Landsat-8, Sentinel-2, and RADARSAT-2 sensors (AAFC, 2018). As a crop type digital map covering the major cropland of Canada, the ACI product has been used as reference data for the major crop types identification (Liu et al., 2016), winter wheat biomass estimation (Dong

* Corresponding author at: Center for Spatial Information Science and Systems, George Mason University, Fairfax, VA 22030, USA.
E-mail addresses: czhang11@gmu.edu (C. Zhang), ldi@gmu.edu (L. Di), phao@gmu.edu (P. Hao), Zhengwei.Yang@usda.gov (Z. Yang), llin2@gmu.edu (L. Lin), hzhao22@gmu.edu (H. Zhao), lguo2@gmu.edu (L. Guo).

https://doi.org/10.1016/j.jag.2021.102374
Received 10 February 2021; Received in revised form 17 April 2021; Accepted 18 May 2021
Available online 11 June 2021

et al., 2016), cropland extent classification (Massey et al., 2018), and measurement of landscape indicators (Ruan et al., 2019). However, both CDL and ACI are end-of-season products usually releasing to the public in the early next year although pre-public release products of CDL and ACI are available internally at the late stage of a growing season for evaluation and validation purpose, which means these data cannot meet in-season needs of users. In-season mapping aims to create a crop-specific land cover map during the growing season. For example, in early 2019, at least 1 million acres of farmland in the Midwest U.S. were damaged due to flooding (Huffstutter and Pamuk, 2019) and the production of corn and soybeans in these areas could be significantly affected. A timely and detailed crop cover map can facilitate the crop loss estimation and decision support immediately after the flood event. Since an operational field-level crop-specific mapping product for CONUS is still not publicly available within growing season, a general method for rapid mapping of major crop types is highly desired by not only the food and agricultural sectors but also researchers, insurance companies, and the LULC community.

Although many methods and algorithms can reach excellent classification performance in the study area at the early stage of a growing season (hereafter called "in-season" or "early-season" for simplicity), it is still a challenge to produce the regional-scale early-season crop cover map at the field level. One crucial issue for the challenge is the insufficiency of ground truth data in the early season. Generally, producing a high-quality regional-scale crop cover map requires a huge volume of ground truth data. However, surveying field data over a large geographic area is a time-critical activity which requires a substantial investment of human and financial resources. The production of CDL relies on massive ground truth information collected at the June Area Survey by USDA NASS (Boryan et al., 2011). The production of ACI data is based on the ground truth information by point observations from local crop insurance companies and AAFC personnel (AAFC, 2019). These surveyed data are not publicly available at all and are only available internally at the late stage of a growing season (e.g., early August) after processing and quality controls of collected ground truths. Although there are many approaches for crop type classification, it is difficult to apply to a larger area without training samples labeled by ground reference data. This challenge drives us to develop an efficient and effective approach to prepare training samples for remote sensing-based in-season crop mapping.

With the considerable demand for the timely crop cover map, crop mapping based on Moderate Resolution Spectroradiometer (MODIS) data and moderate-to-high spatial resolution data, such as Landsat data and Sentinel-2 data, has been widely studied around the world. For example, Wardlow and Egbert (2008) and Dahal et al. (2018)) explored the large-area crop mapping using MODIS time series. McNairn et al. (2014) applied the supervised decision tree classification for the early-season mapping of corn and soybeans in eastern Canada using TerraSAR-X and RADARSAT-2 data. Hao et al. (2015) and Skakun et al. (2017) investigated the capability of using MODIS data for early-season crop mapping. Vaudour et al. (2015) utilized very high spatial resolution Pleiades image for early-season mapping of crops and soil tillage operation. Tardy et al. (2017) proposed several past data fusion schemes for land cover mapping, including major winter and summer crop classes, using in-season Formosat-2 images and reference data from previous periods. Hao et al. (2018) developed an improved artificial immune network approach to identify major crops using Sentinel-1 and Sentinel-2 time series. Kussul et al. (2018) combined Sentinel-2 data with Sentinel-1A data for in-season crop mapping at regional scale in Ukraine. Phalke and Özdoğan (2018) and Johnson (2019) applied the Landsat data for the field-level cropland mapping over large area. Defourny et al. (2019) introduced the Sen2-Agri system to generate near real-time national-scale crop type maps from Sentinel-2 data using random forest and successfully demonstrated in Ukraine, Mali, and South Africa. Demarez et al. (2019) investigated the feasibility of combing Landsat-8 data and Sentinel-1 time series for the in-season classification of irrigated crops. Wang et al. (2019) used multi-temporal Sentinel-2 data to map regional land use in complex landscapes. Gao et al. (2020) presented a within-season approach to detect crop emergence during early growing season using Landsat and Sentinel-2 data.

Different from the in-season mapping, pre-season mapping aims to predict crop cover before the beginning of a growing season and normally does not involve remote sensing data directly. Modeling crop sequence is a common way to predict the crop type on cropland. Many studies have shown that the spatial information of crop planting can be predicted by modeling the long-term crop sequence using land survey data. Xiao et al. (2014) used an empirical modeling method to identify the crop sequence patterns at a regional scale. Osman et al. (2015) proposed a Markov logic-based approach of crop rotation modeling for early crop mapping. Kollas et al. (2015), Yin et al. (2017), and Giordano et al. (2020) analyzed and discussed several crop rotation models in Europe. Since gathering and processing the land survey data of large area is impractical for most researchers, the crop cover maps derived from remote sensing images, such as CDL, become an ideal data source for crop sequence modeling. From the historical CDL data, a crop rotation map can be created by extracting croplands that follow specific crop rotation such as alternate cropping patterns (Sahajpal et al., 2014; Wu and Zhang, 2019). For example, it is well-known that many farmers in the Midwestern U.S. alternatively grow corn and soybeans on the same land unit year after year because of the effect on crop yield as well as soil quality and fertility (Edwards et al., 1988; Karlen et al., 2006; Van Eerd et al., 2014). To generalize the crop-specific expert knowledge to other regions where the crop rotation patterns are different, Zhang et al. (2019a) utilized artificial neural network (ANN) to predict the spatial distribution of future crop planting before the beginning of a growing season, which has been proven effective to produce pre-season crop map and estimate crop yield for a normal year. However, one important issue of the pre-season mapping is that the prediction result only relies on prior knowledge from historical data, which could be problematic while mapping for an anomaly year with disasters or large market and policy changes. To address this issue, we assume the spectral information of different crop growth stages in remote sensing data can contribute to repair the erroneous predictions caused by accidents, artificial changes, or lack of credible prior crop sequence knowledge. Based on this idea, this paper introduces an innovative approach using machine-learned trusted pixel (hereafter called "trusted pixels"), whose crop types have been identified with high confidence by the crop sequence model automatically learned from the time series of crop-specific land cover data (e.g., CDL), to label training samples on satellite images for in-season crop type classification.

This paper has two specific objectives: (a) using ANN to automatically extract trusted pixels from historical CDL data; (b) exploring the feasibility of mapping major crops (i.e., corn and soybeans) for the U.S. Corn Belt based on multi-temporal satellite images using trusted pixels as training samples. The rest of this paper is organized as follows: Section 2 introduces the study area and data; Section 3 presents an ANN-based machine learning framework for rapid mapping of corn and soybeans; Section 4 demonstrates a group of experiments to evaluate trusted pixels and in-season crop cover maps; Section 5 discusses the advantages of the proposed method, the uncertainty of CDL data, and limitations and potential solutions of the current work; and Section 6 concludes this paper with a discussion of future works.

2. Study area and data

This study focuses on the U.S. Corn Belt region of the Midwestern United States where agriculture has been the predominant land use class of this region in the past few decades. Corn and soybean have become the leading crops in many Corn Belt states since the late 1990s (Auch et al., 2018). As the major agriculture region over the CONUS and the largest production area of corn and soybeans in the world, the U.S. Corn

Belt is a key study area for crop mapping as well as national-scale and global-scale LULC change studies (Green et al., 2018). According to the statistics by USDA NASS, in 2018 cropland makes up over 65% of the total land in the study area and corn and soybeans cover 22.9% and 22.7% of the total cropland area, respectively. The geography and agricultural land use information for Corn Belt states are shown in Fig. 1. The study area consists of 12 states: Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, and Wisconsin. USDA NASS divided each U.S. state into several Agricultural Statistics Districts (ASDs) by geography, climate, and cropping practices. Each ASD consists of a group of contiguous counties with relatively similar agricultural characteristics and environment. This study will investigate 102 contiguous ASDs across various ecoregions, including Northern Forests, Northern Agriculture-Forest Transition Zone, Midwest Agricultural, Western Plains, Glaciated Plains, and East-Central Plains (Auch and Karstensen, 2015; Taylor et al., 2015).

The CDL time series will be used as the main reference data for crop sequence model training as well as result validation. Produced by USDA NASS, the CDL data product covers the entire CONUS at 30-meter spatial resolution from 2008 to the present and some states from 1997 to 2007. There are over 140 land use classes provided in the CDL, and the accuracy for major crop types in most areas is close to 95% according to the CDL metadata (USDA-NASS, 2019b). All CDL products can be accessed through CropScape (https://nassgeodata.gmu.edu/CropScape), a web-based geospatial information system developed and maintained by the Center for Spatial Information Science and Systems of George Mason University (Han et al., 2012; Zhang et al., 2019b).

As the two most widely accessible moderate-to-high spatial resolution data, Landsat-8 and Sentinel-2 data are explored in this study. The Landsat-8 data cover the entire Earth's surface at a 30 m resolution in a 16-day repeat cycle since 2013. The Sentinel-2 data provide the higher spatial resolution of 10–20 m with a global 5-day revisit frequency (depending on orbits and satellite, the same areas could be under different angles) since 2015. There are many ways to access Landsat data and Sentinel-2 data. The United States Geological Survey (USGS) Earth Explorer (https://earthexplorer.usgs.gov/) is the official source for downloading Landsat data. The ESA Copernicus Open Access Hub (https://scihub.copernicus.eu/) provides complete and open access to Sentinel-2 data.

Furthermore, we collected a set of field data using roadside sampling strategy within Nebraska and Iowa, which contains 128 land units (77 for corn and 51 for soybeans) surveyed in July 2018 and 900 land units (723 for corn and 177 for soybeans) surveyed in July 2019. Each land unit represents a field of certain type of crop planted with a permanent contiguous boundary. The Common Land Unit (CLU) of USDA provides the vector data set of all registered agricultural fields in the United States. However, it is not available for public. Therefore, we utilized GIS software to manually delineate polygons for the field data based on the Landsat-8 images of the same period. To ensure the surveyed land unit has a relatively permanent contiguous boundary and common land management, we verified each polygon by referring to the historical CDL data. These field data will be used as ground truth to validate the mapping results.

3. Methodology

3.1. Overall workflow

Fig. 2 shows the overall workflow of the proposed in-season crop mapping approach, which is composed of a data preparation module, a trusted pixel prediction module, a crop type classification module, and a validation module. The data preparation module converts the CDL data and satellite data into the structured form to feed the trusted pixel prediction module and classification module. The trusted pixel prediction module automatically produces trusted pixels from the historical CDL data. The crop type classification module deals with the identification of corn and soybeans from the remote sensing images. The output of the workflow is an in-season crop map. Finally, all mapping results will be validated using ground truth data and CDL data. This section presents the implementation details of each module.

3.2. Data preparation

This step aims to prepare the CDL data for the prediction of the trusted pixel as well as build the multi-temporal and multi-spectral satellite image stack for the classification model training. We processed the data used for the experiments on the Google Earth Engine (GEE) (Gorelick et al., 2017). The GEE data catalog has archived diverse standardized geospatial data sets including the complete volume of CDL, Landsat-8, and Sentinel-2 data. There are many product options for both Landsat-8 data and Sentinel-2 data. To increase the consistency between the two satellite data, we will adopt two atmospherically corrected surface reflectance data, the Landsat-8 Surface Reflectance Tier 1 product and Sentinel-2 MultiSpectral Instrument (MSI) Level-2A product. Meanwhile, we have developed the AgKit4EE toolkit on GEE (Zhang et al., 2020a) to prepare the CDL time series for the entire study area.

Spectral bands and their derived indices from satellite images can provide fundamental information for crop type classification (Arvor et al., 2011; Löw et al., 2013; Wardlow et al., 2007). This study investigates spectral features in six bands, including three visible spectrum bands (blue, green, red), one near-infrared band (NIR), and two shortwave infrared bands (SWIR-1, SWIR-2). Table 1 summarizes the information of the selected bands in Landsat-8 data and Sentinel-2 data. Additionally, we derive two commonly used indices, Normalized Difference Vegetation Index (NDVI) (Tucker, 1979) and Normalized Difference Water Index (NDWI) (Gao, 1996), from the given spectral bands. The formulas of NDVI and NDWI are shown as follows:

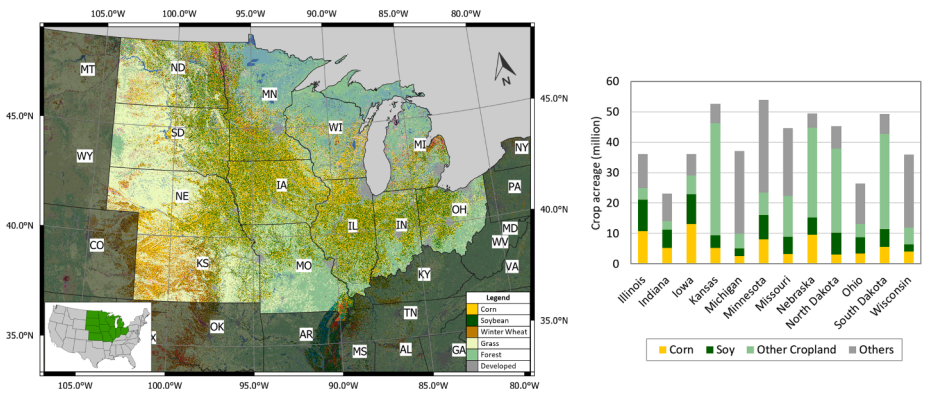


Fig. 1. Geography of the study area (data from 2018 CDL by USDA NASS).

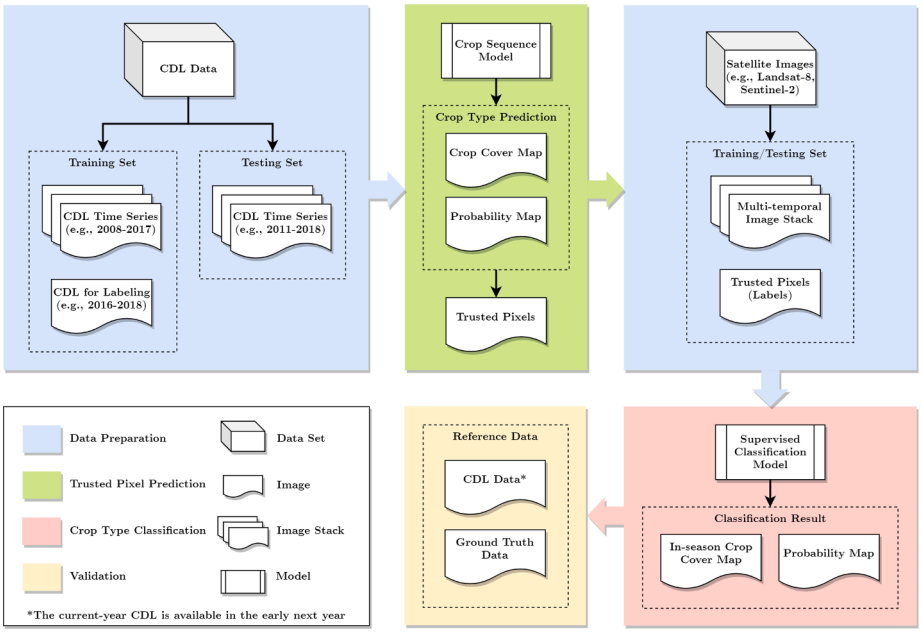


Fig. 2. Workflow for in-season crop mapping using trusted pixels and remote sensing images.

Table 1
Information of spectral bands in Landsat-8 data and Sentinel-2 data.

	Landsat-8			Sentinel-2		
Name	Band	Wavelength (micrometers)	Spatial Resolution (meters)	Band	Wavelength (micrometers)	Spatial Resolution (meters)
Blue	B2	0.452–0.512	30	B2	0.439–0.535	10
Green	B3	0.533–0.590	30	B3	0.537–0.582	10
Red	B4	0.636–0.673	30	B4	0.646–0.685	10
NIR	B5	0.851–0.879	30	B8	0.848–0.881	10
SWIR-1	B6	1.567–1.651	30	B11	1.539–1.681	20
SWIR-2	B7	2.107–2.294	30	B12	2.072–2.312	20

$$NDVI = \frac{NIR - Red}{NIR + Red}$$

$$NDWI = \frac{NIR - SWIR}{NIR + SWIR}$$

In the U.S. Corn Belt, corn is normally planted in late April through late May and soybeans are planted in May through early June (USDA-NASS, 2010). The best temporal period for satellite remote sensing of corn and soybeans is from the time of planting to the peak of the growing season in mid-July. To ensure the temporal dimension is adequately covered in the training data, we acquired all low cloud cover (< 10%) images taken between May and mid-July for each scene. After selecting spectral features, each image is composed of 8 bands (i.e., blue, green, red, NIR, SWIR-1, SWIR-2, NDVI, and NDWI). When building the image stack for crop type classification, we sort all qualified satellite images by date then stack the selected bands of each image to a 3-d array. For most Landsat scenes within the study area, there are 3–4 qualified images with at least one image per month, which means a multi-temporal image stack normally consists of 24 or 32 bands in total. Take the image stack with 24 bands as an example, the training data will contain spectral information of different crop growth stages in May (band 1–8), June (band 9–18), and July (band 19–24).

Since Landsat data is the main data source of CDL, the experiment in this paper used the Landsat-8 images as the preference data. If a Landsat image stack does not have enough low cloud cover images (less than 3 images), the Sentinel-2 data will be used to patch the missing scene. To further minimize the effects of cloudy pixels in classification, we fill the cloudy pixels with the corresponding pixels of the previous image in the image stack. Besides, the spatial resolution of the trusted pixel map and in-season crop cover map are identical with Landsat images and CDL data. To unify the spatial resolution of the satellite data, all Sentinel-2 images are resampled to 30 m using the nearest neighbor method.

3.3. Trusted pixel prediction

Trusted pixels refer to pixels predicted from the historical CDL data with high confidence in the current year's crop type. As a practical approach for discovering intricate patterns and structures in high-dimensional data, machine learning has been widely used in LULC studies. The production of trusted pixel is based on the crop sequence pattern that automatically recognized from the CDL time series. To train the crop sequence model, we integrate an ANN model with the in-season mapping workflow, which has been proven effective to predict the spatial distribution of major crop types (Zhang et al., 2019a).

Fig. 3 illustrates the process of trusted pixel prediction. First, we convert the historical CDL time series into an image stack with crop sequence features for all pixels. Each crop sequence feature is a one-dimensional array containing the pixel-level time series of historical CDL. Second, we feed each crop sequence feature into the prediction model to predict the following year's crop type of the corresponding pixel. The ANN model for trusted pixel prediction has the fully-connected multilayer perceptron (MLP) structure, which consists of one input layer, five hidden layers, and one output layer. Each input neuron represents each crop type value of the crop sequence feature. The output layer of the neural network used SoftMax to calculate the probability value of three classes (corn, soybeans, or others). The crop type of

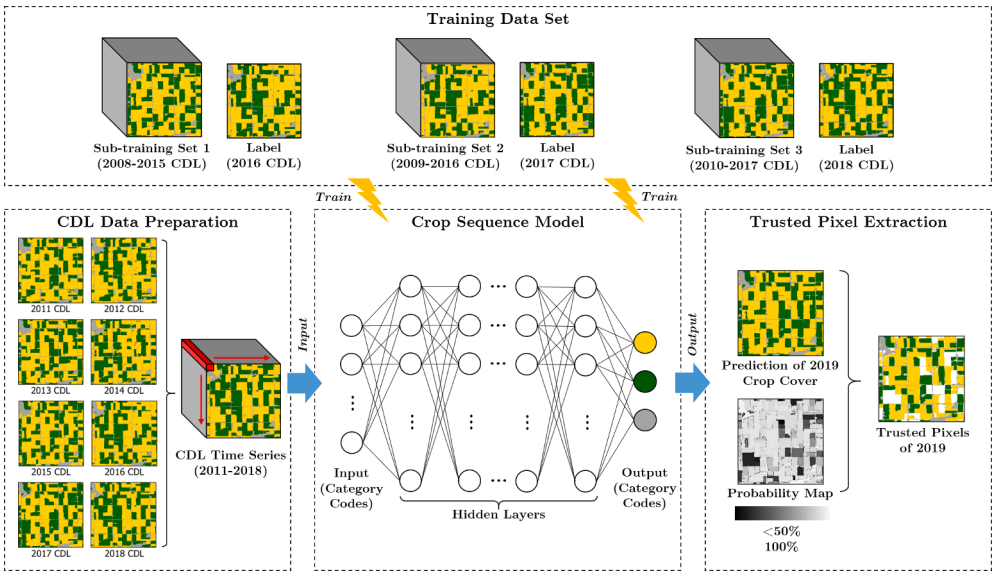


Fig. 3. Predicting trusted pixels from historical CDL time series using ANN.

the corresponding pixel will be categorized as class with the highest probability value. The final output of the prediction model is a prediction map of crop cover and its probability map. By masking the high-confident pixels (> 90%) on the prediction map, we can get a map of trusted pixels. If the sequence is similar to a regular pattern, there is a high chance that the pixel would be classified as a trusted pixel (e.g., corn 90%, soybeans 8%, others 2%). If a sequence cannot be recognized by the well-trained model, the probability of each class could be more even (e.g., corn 45%, soybeans 30%, others 25%) and it would be classified as a non-trusted pixel.

The training data set is constructed with three recursive subsets with an 8-year moving window of each. While producing trusted pixels for 2019, the ANN model is trained using sub-training set of 2010–2017 CDL labeled with 2018 CDL, 2009–2016 CDL labeled with 2017 CDL, and 2008–2015 CDL labeled with 2016 CDL. This design can efficiently extend the training data set and allow the neural network to recognize crop sequence labels for the last three consecutive years. To convert features into the readable form of neural network, the training data set will be flattened to a structured 2-D table. Each row represents a sample of a sequence of pixel-level crop type features labeled with the corresponding pixel in the label set. For example, a training sample of pixel that follows the corn-soybean rotation pattern will be represented as "1, 5, 1, 5, 1, 5, 1, 5" labeling with "1" or "5, 1, 5, 1, 5, 1, 5, 5" labeling with "5," where "1" refers to corn and "5" refers to soybeans (the full class table of CDL data is available at Google (2021)). Although the CDL data is available since 1997, we would not build the training set with a very long CDL time series because the quality of the early-year CDL varies across regions and the coverage of CDL is incomplete before 2008, which may significantly affect the accuracy of the machine-learned model.

To train a robust prediction model, the training set should provide abundant samples with diverse crop sequence features. Based on the similarity of agricultural characteristics and environment, USDA NASS divided each U.S. state into several Agricultural Statistics Districts (ASDs). To make sure the crop sequence features of prediction model as much as possible, the models need to be trained for each ASD and the trusted pixel mapping has to be done ASD by ASD. In this way, the well-trained neural network would recognize the specific crop sequence information for the corresponding ASD.

3.4. Crop type classification

Fig. 4 shows the procedure of in-season crop type classification using satellite images and trusted pixels. As described in the Section 3.2, the input data structure of the classification model is an image stack with both spectral and temporal information. The quantity of satellite images used for assembling image stack depends on the availability of cloud-free satellite images within growing season. Based on the spatial distribution of trusted pixels, the training samples are automatically labeled on the image stack. The trusted pixel-based training samples can be applied to diverse pixel-based classifiers. This study applied the MLP-

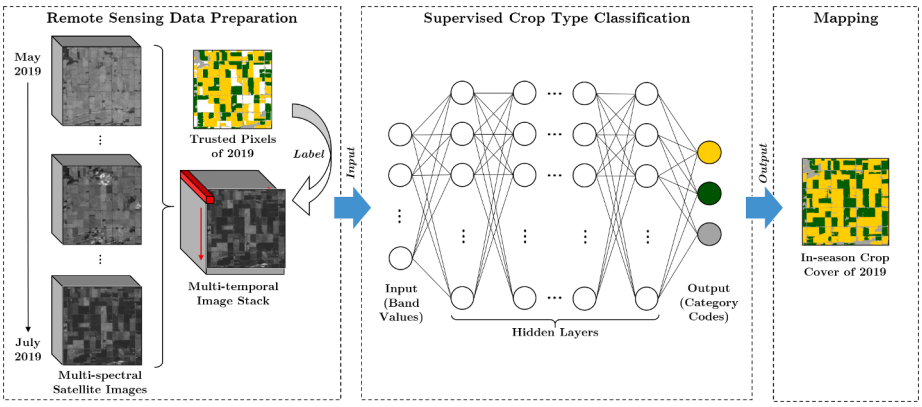


Fig. 4. In-season crop type classification using multi-temporal satellite image stack and trusted pixels.

based ANN as classifier which has a similar structure with the trusted pixel prediction model. Each input neuron represents the value in the one-dimensional band feature of the corresponding pixel. Finally, an in-season crop cover map can be generated by applying the trained classification model on the full image. The geography, season starting, and temporal collection of satellite images may significantly vary among the different scenes over large area. To test the scalability and robustness of the trusted pixel approach, we conduct crop type classification scene by scene within the entire U.S. Corn Belt.

3.5. Validation

We use overall accuracy (OA), precision, recall, and F1-score to evaluate the agreement of mapping results with reference data. These metrics are calculated by comparing all pixels of the mapping result with their corresponding pixels of the CDL map or ground truth map. The OA measures the proportion of correctly classified pixels in all pixels. The precision and recall measure the classification result of each class, which are defined as follow:

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

where *TP* represents the number of true positives, *FP* represents the number of false positives, and *FN* represents the number of false negatives. The F1-score combines precision and recall which is defined as:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

Specifically, precision measures the fraction of tuples labeled as positive which are actually positive. The recall measures the fraction of positive tuples that were detected. Take corn as an example, the precision indicates the proportion of correctly classified corn pixels in all pixels classified as corn. The higher the precision rate, the more accurate the classified corn pixels are. The recall indicates the proportion of the correct corn pixels of the classification result in all corn pixels of the reference image. The higher the recall rate, the more corn pixels are correctly classified. The result might have a high precision rate but low recall rate, or high recall rate but low precision rate. The F1-score can measure the weighted average of precision and recall. The values of the above metrics are between 0 and 1.

4. Experiments and results

4.1. Validation of trusted pixels

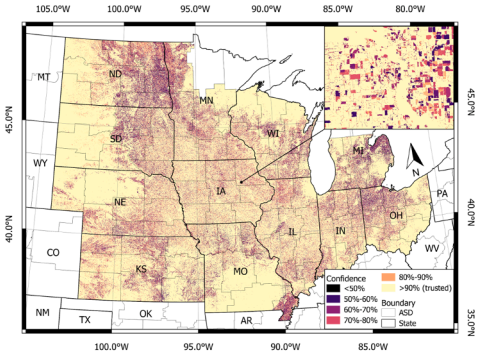
To fulfill the first objective of this study, we derived trusted pixels of 2018 by applying the crop sequence model (trained using CDL data of 2007–2016) to the CDL data of 2010–2017. Fig. 5 shows the 2018 trusted pixel map of the study area. A probability map (Fig. 5a) is created resulting from the SoftMax function where each pixel indicates the highest probability value among corn, soybeans, and others. Based on the probability map, we extracted pixels with high classification confidence from the prediction result and produced the trusted pixel map (Fig. 5b). The pixels higher than the threshold of 90% were classified as corn, soybean, or others. The pixels under the probability threshold were considered the non-trusted pixels. The result indicates that the trusted pixels of corn and soybeans are widely distributed across the U.S. Corn Belt, especially Iowa, Illinois, Indiana, Eastern Nebraska, Eastern South Dakota, Southwestern Minnesota, and Western Ohio.

This experiment used 2018 CDL data as reference data to validate the reliability of trusted pixels. Fig. 6 illustrates the spatial distribution of validation results at the ASD level. The precision rate of corn and soybeans are compared in Fig. 6a and Fig. 6b, which indicates the agreement of the trusted pixels in each class with the corresponding CDL pixels. The recall rate of corn and soybeans are compared in Fig. 6c and Fig. 6d, which reflects the percentage of correctly classified trusted pixels against the pixel number of the corresponding class in CDL. It can be found the ASDs with high-density of trusted pixels are mainly concentrated in Central Corn Belt Plains where the precision can reach 0.9 and the recall can reach 0.6. In contrast, the precision and recall of ASDs that mostly covered by non-cropland, such as Western North Dakota and Western South Dakota, are not as high as other areas.

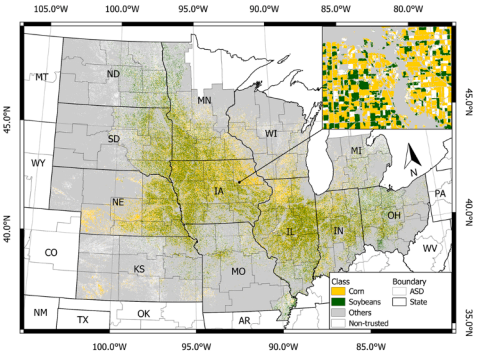
Fig. 7 displays the 9-year crop rotation patterns (8-year sequence followed by a prediction result) automatically recognized by the ANN in the above results. This example illustrates the statistics for the quantity of trusted pixels per crop rotation pattern in ASD #1750, #2770, and #3160 (Fig. 7a) and validation of the most frequent crop rotation patterns within selected regions (Fig. 7b). The overall precision of corn pixels can reach up to 0.96 while the overall precision of soybeans pixels is as high as 0.97 among selected ASDs. Such a high overall precision rate indicates the ANN is capable of recognizing common crop rotation patterns as well as patterns with minor anomaly from the historical CDL. We can see the most frequent corn-soybean rotation pattern within the selected ASDs is the alternate rotation pattern (i.e., "C-S-C-S-C-S-C-S-C" and "S-C-S-C-S-C-S-C-S"), in which the recall rate is more significant than other patterns. The monocropping of corn (i.e., "C-C-C-C-C-C-C-C-C") exists in ASD #2770 and #3160. Moreover, a variety of other crop rotation patterns were recognized, but the pixel number is much lower than the regular patterns.

4.2. Validation of in-season crop cover map using CDL

The second objective of this study is conducting in-season crop mapping using multi-temporal satellite images and trusted pixels as training sample. In order to guarantee the image stack was assembled by at least one eligible image per month, we gathered all cloud-free

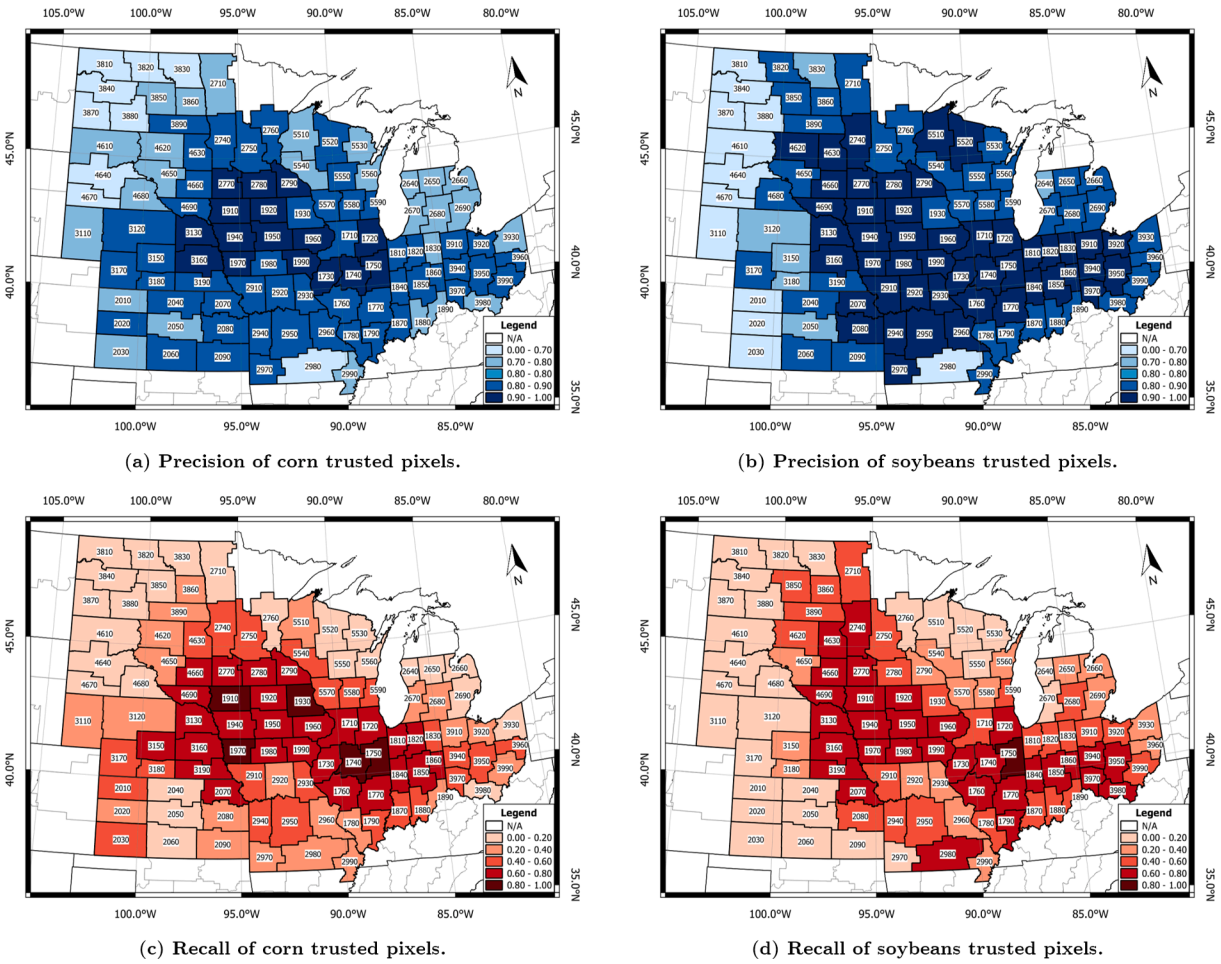


(a) Probability map (bright pixels represent trusted pixels).



(b) Trusted pixel map.

Fig. 5. Spatial distribution of 2018 trusted pixels within the study area. The probability map indicates the highest probability value resulting from the SoftMax function. The trusted pixel map includes the predicted pixels with probability higher than the threshold of 90%. A few ASDs, such as ASD #2720 and ASD #2730 (located at Northern Minnesota), and ASD #2610, ASD #2620, ASD #2630 (located at Northern Michigan), are mainly covered by non-croplands. These ASDs were not investigated.



(a) Precision of corn trusted pixels.

(b) Precision of soybeans trusted pixels.

(c) Recall of corn trusted pixels.

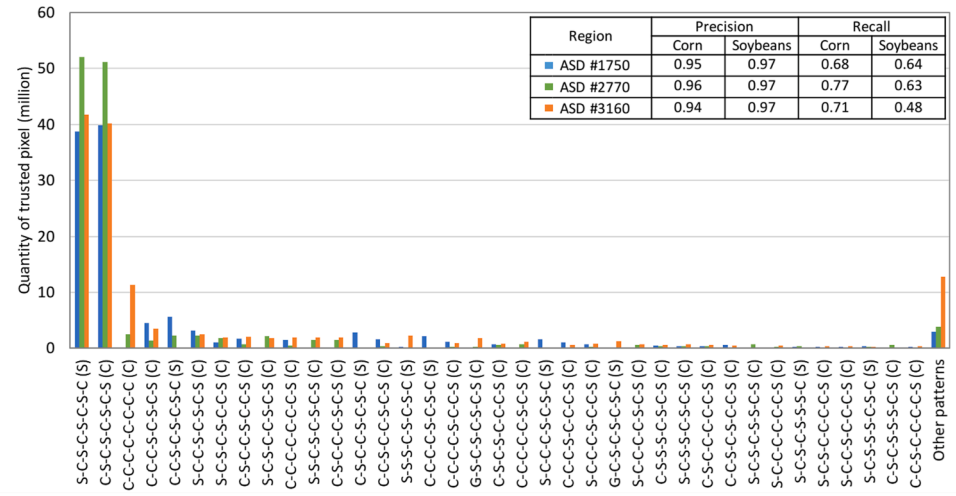
(d) Recall of soybeans trusted pixels.

Fig. 6. Spatial distribution of trusted pixels validation at ASD level. The ASDs with high-density of trusted pixels, such as Iowa, Illinois, and Indiana have a higher precision rate (> 0.9) and recall rate (> 0.6). ASDs that mostly covered by non-cropland, such as Western North Dakota (ASD #3810/3820/3840/3870/3880) and Western South Dakota (ASD #4610/4640/4670), have the lower precision (< 0.7) and recall (< 0.2).

Landsat-8 images (cloud coverage less than 10%) between May 2018 and July 2018 over the entire study area as well as Sentinel-2 images over the regions where the cloud-free Landsat-8 images were unavailable. The final mapping result covers 80 tiles, 49 of which were produced using Landsat-8 data and the rest were produced using Sentinel-2 data. The Landsat-8 data use the Worldwide Reference System-2 (WRS-2). Each scene of Landsat-8 data is 185 km*180 km and uniquely designated by a combination of Path and Row numbers. The Sentinel-2 Level-2A product is organized in ortho-rectified tiles, also called granules, of 100 km*100 km in UTM WGS84 projections. Each Sentinel-2 tile is assigned an ID, in which the first 2 digits and 1 letter correspond to the UTM zone while the two last letters correspond to a unique ID.

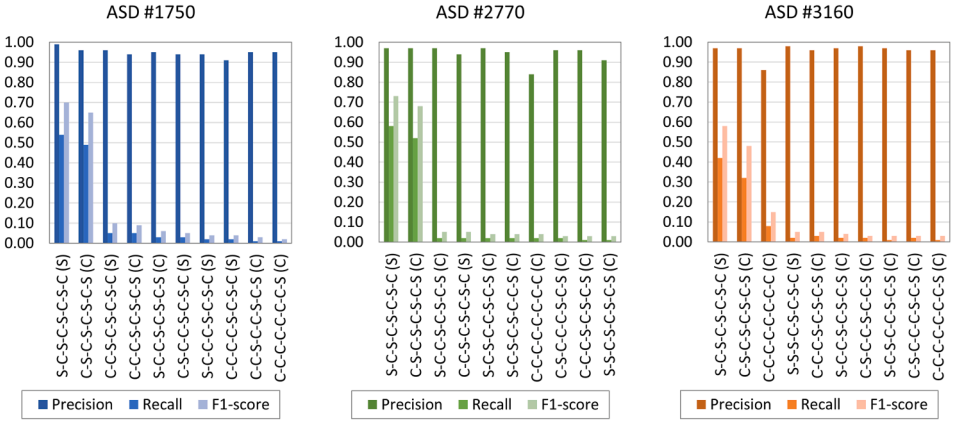
Fig. 8 demonstrates the 2018 in-season crop cover map and the detailed mapping result of six region of interests (ROI) including three Landsat-8 scenes and three Sentinel-2 tiles with the OA of 80%–85%, 85%–90%, and 90%–95%, respectively. The selected ROIs are distributed across various ecoregions in the U.S. Corn Belt. The ROI "P030R027" covers part of Lake Agassiz Plain Ecoregion and part of Northern Glaciated Plains Ecoregion. The ROI "P028R032" is on the border of Central Great Plains Ecoregion and the Western Corn Belt Plains Ecoregion. The ROI "P023R032" is in the Central Corn Belt Plains Ecoregion. The ROI "14TQQ" lies on the border of Northern Glaciated Plains Ecoregion and Western Corn Belt Plains Ecoregion. The ROI "15TWG" is in the Western Corn Belt Plains Ecoregion. The ROI "16TFM" is in Southern Michigan/Northern Indiana Drift Plains Ecoregion. To show the discrepancy of in-season map and CDL, we calculated the difference between two images and highlighted the misclassified pixels in each ROI. The result indicates the difference between CDL and in-season maps of low overall accuracy (80%–85%), such as "P030R027" and "16TFM," was relatively significant. For the in-season maps of high overall accuracy (90%–95%), such as "P023R032" and "15TWG," the crop types of nearly all land units were correctly classified. Only a few mixed pixels (mainly distributed on the boundary of land unit) were misclassified. Considering the crop cover maps were produced by July 2018 which was nearly six months earlier than the release of 2018 CDL, such difference is reasonable for the in-season mapping result.

Fig. 9 compares the overall validation results of prediction maps (including trusted pixels and non-trusted pixels), trusted pixel maps, and in-season crop cover maps for all investigated scenes. The box plots show the in-season maps can reach as high overall precision rate as the trusted pixel map. Meanwhile, the overall recall rate and overall F1-score of in-season maps are higher than both trusted pixel maps and prediction maps. Specifically, the mean/median precision values of corn and soybeans were 0.85/0.86 and 0.87/0.89. The mean/median recall rates of corn and soybeans were 0.78/0.83 and 0.78/0.83. The mean/median F1-score rate of corn and soybeans were 0.80/0.83 and 0.81/0.83. In most scenes, the precision rate of corn and soybean ranged from 0.81 to 0.90 and 0.83 to 0.92, the recall rates of corn and soybean ranged from 0.69 to 0.90 and 0.68 to 0.88, and the F1-score of corn and soybean ranged from 0.75 to 0.88 and 0.76 to 0.90. It could be reasonably inferred from the validation result that a significant number of non-trusted pixels can be correctly classified based on the spectral and temporal features from satellite images.



Region	Precision		Recall	
	Corn	Soybeans	Corn	Soybeans
ASD #1750	0.95	0.97	0.68	0.64
ASD #2770	0.96	0.97	0.77	0.63
ASD #3160	0.94	0.97	0.71	0.48

(a) Quantity of trusted pixels for all crop rotation patterns.



(b) Validation of most frequent crop rotation patterns within selected regions.

Fig. 7. Example of crop rotation patterns recognized by the ANN. ASD #1750 is located in Eastern Illinois, ASD #2770 is located in Southwestern Minnesota, ASD #3160 is located in Eastern Nebraska. For each crop rotation pattern, "C" represents corn, "S" represents soybeans, "G" represents grass.

4.3. Validation of in-season crop cover map using field data

This section includes two experiments. First, we used field data and the 2018 CDL as reference data to validate the in-season map of 2018. As shown in Fig. 10, the location of field data surveyed in 2018 is covered by the Sentinel-2 tile (14TQL) as well as two Landsat-8 scenes (P028R031, P028R032). We grouped the field data into two ROIs by state and compared them to the corresponding pixels from CDL data and in-season crop cover maps. The pixel-by-pixel validation results are summarized in Table 2. The validation result suggests the in-season map can reach the high F1 score (> 0.95) and OA ($> 95\%$) using either field data or CDL as reference data.

The second experiment validated the in-season mapping result of 2019 and derived the crop loss from the crop cover maps. Since most Landsat-8 images during the growing season of 2019 had high cloud coverage within the ROI, only the Sentinel-2 images were adopted in this experiment. The production of the 2019 crop cover map involved the following steps: (1) use CDL of 2007–2017 to train the crop sequence model; (2) generate trusted pixels by applying the well-trained model on 2011–2018 CDL time series; (3) label training samples on the multi-temporal image stack (images acquired between May 1 to June 30 of 2019) for each tile. The field data surveyed in 2019 were covered by six Sentinel-2 tiles (14TQL, 14TPM, 14TPL, 14TNM, 14TNL, 14TML). The validation results are summarized in Table 3. In early 2019, a severe flood event occurred in the Midwestern U.S., and many croplands along the Missouri River were damaged. Fig. 11 highlights the land cover change before and after the 2019 Midwest flooding and compares with the Landsat image of June 2019. It is noteworthy that our method successfully detected and mapped the crop loss which was located at the junction of Nebraska State, Iowa State, and Missouri State.

5. Discussion

5.1. Advantages of crop mapping using trusted pixels

This study presents an innovative machine learning approach for the rapid in-season mapping of corn and soybeans. Here we summarize the main advantages of the proposed method. First, the use of trusted pixel provides a new perspective on crop type classification and LULC mapping. Compared with the traditional in-season mapping methods, the proposed workflow does not require any ground truth data of the current year, which would significantly reduce the workload of gathering training samples at the early growing season, and could potentially automate the crop mapping.

Second, the trusted pixels can be used to label various moderate-to-high spatial resolution data and coupled with many kinds of pixel-based classification algorithms. With the trusted pixels, many existing LULC mapping methods, especially the machine learning algorithms that require massive training samples, can be efficiently scaled up to a large geographic area. Moreover, the trusted pixels can be potentially integrated with other supervised classification workflows, such as transfer learning-based crop type classification (Hao et al., 2020), to perform in-season mapping for regions outside the United States, especially the developing countries that are lack of historical crop cover data.

Fig. 8. Comparison of in-season mapping result and CDL data. The yellow pixels represent corn. The green pixels represent soybeans. The gray pixels represent all other land use types (e.g., water, developed area, forest, and grass). The difference between in-season map and CDL is highlighted with red pixels. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Another advantage of the proposed method is that only a few in-season satellite images for each scene are required and no historical satellite images are involved. Due to the lack of training samples, most supervised learning-based crop mapping methods cannot be directly used on the in-season satellite images. Many deep learning-based crop type classification models used hundreds even thousands of historical satellite images as training data. Some of them do reach excellent classification results, however, the training process of the classification models, especially advanced neural networks with complex structures, is extremely slow and difficult to repeat by other researchers. Since the trusted pixels are distributed in the U.S. Corn Belt, the classification model of our proposed framework can be specifically trained only using the in-season satellite images for each scene.

According to our benchmark test on the desktop with 6-core CPU and 16 GB memory, it takes 5–10 min to extract trusted pixels for each ASD and 15–20 min to produce the crop cover map for each Landsat scene. When the latest satellite images become available, an in-season crop cover map can be rapidly produced by the proposed machine learning

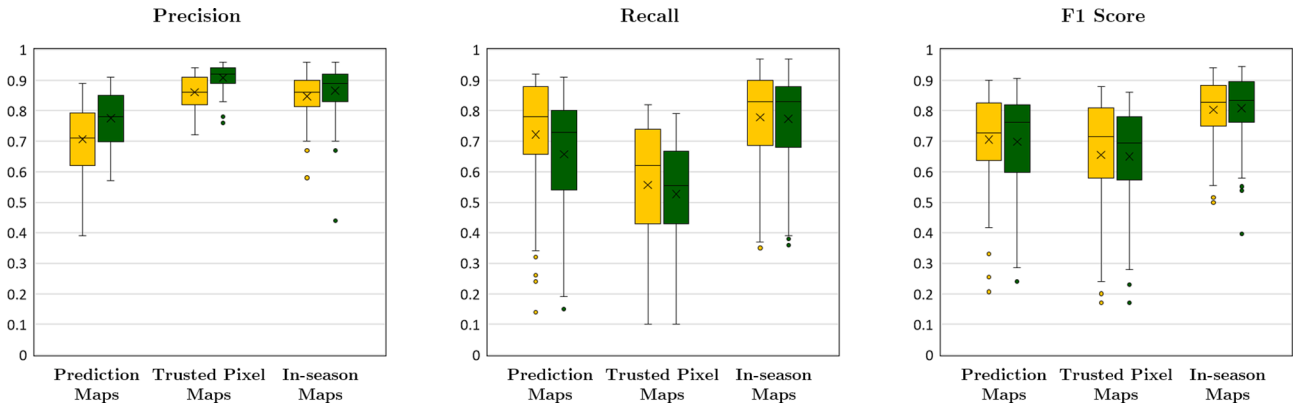


Fig. 9. Box plot of validation results for all investigated scenes. The yellow and green box represent the result of corn and soybeans, respectively. The upper and lower bounds of the box represent the first and third quartiles. The cross mark, the bar in the box, and vertical line indicate the mean, median, and minimum–maximum bound. The solid dots are outliers of each cluster. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

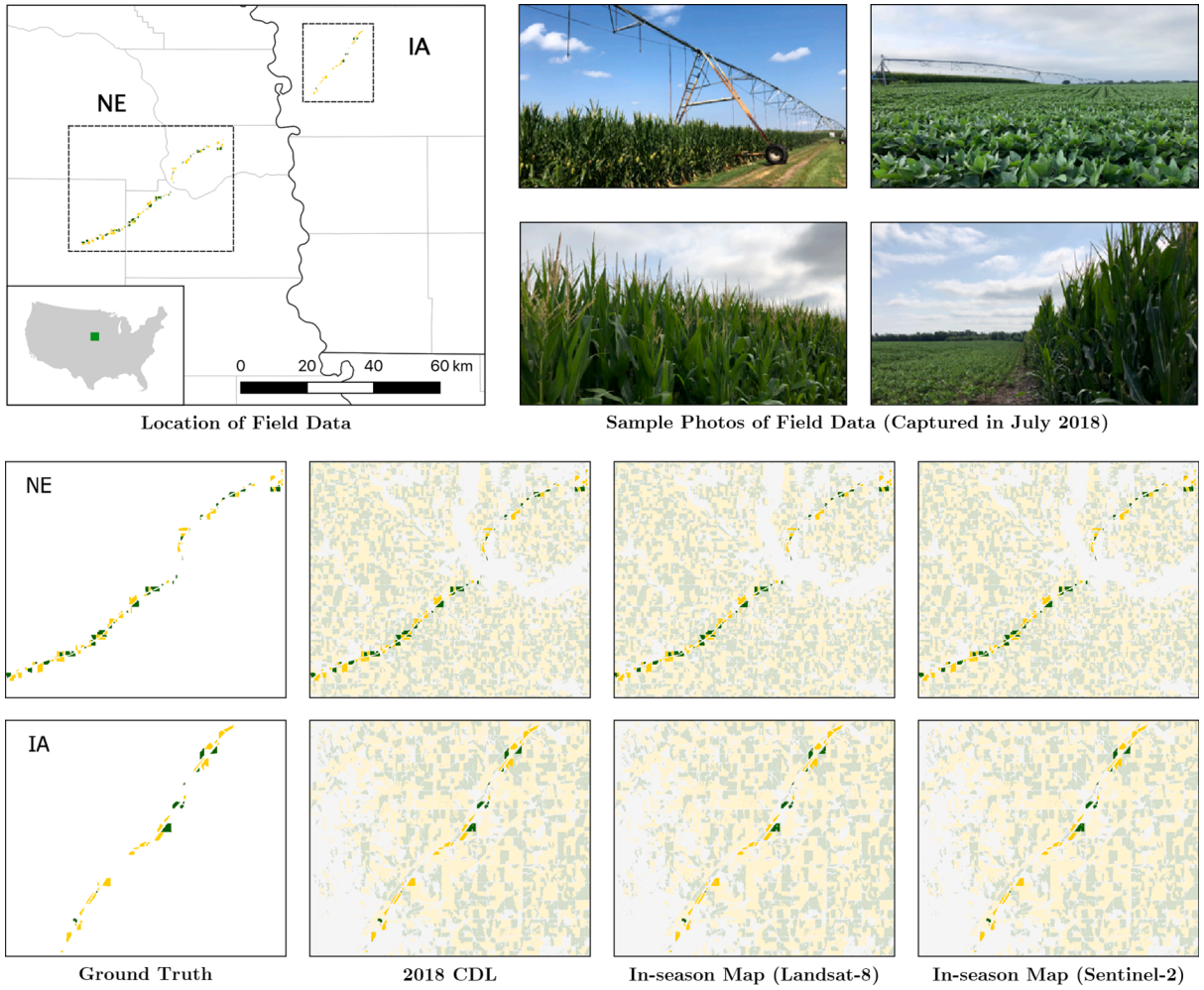


Fig. 10. Comparison of 2018 ground truth data, CDL data, and in-season crop cover map based on Landsat-8 and Sentinel-2. The corresponding pixels of ground truth data are highlighted in the CDL map and in-season crop cover map.

framework, even with the limited computational resources.

Based on the above advantages, we believe not only the agricultural sector but also other societal sectors ranging from government, academia, to industry, would be benefited from this study.

5.2. *Pattern recognition in crop mapping*

The proposed crop mapping workflow consists of two ANN models, which are separately responsible for trusted pixel prediction and the crop type classification. They have the similar MLP structure but the input layers are different depending on the structure of input data. The

Table 2
Validation of 2018 crop mapping result using field data and CDL as reference data.

Reference data	ROI	Precision		Recall		F1-score		OA
		Corn	Soybean	Corn	Soybean	Corn	Soybean	
Field	NE (Landsat-8)	0.98	1.00	0.98	0.94	0.98	0.97	96.49%
data	NE (Sentinel-2)	0.99	1.00	0.98	0.97	0.99	0.99	97.76%
	IA (Landsat-8)	1.00	0.99	0.97	0.95	0.99	0.97	96.03%
	IA (Sentinel-2)	1.00	1.00	0.97	0.96	0.99	0.98	96.67%
CDL	NE (Landsat-8)	0.98	0.99	0.98	0.95	0.98	0.97	96.17%
	NE (Sentinel-2)	0.99	0.98	0.98	0.98	0.99	0.98	97.35%
	IA (Landsat-8)	0.96	0.99	0.98	0.98	0.97	0.99	95.01%
	IA (Sentinel-2)	0.96	0.98	0.98	0.99	0.97	0.99	95.19%

Table 3
Validation of 2019 crop mapping result using field data as reference data.

ROI	Precision		Recall		F1-score		OA
	Corn	Soybean	Corn	Soybean	Corn	Soybean	
14TQL	0.99	0.84	0.83	0.88	0.90	0.86	86.86%
14TPM	0.97	0.88	0.88	0.85	0.92	0.86	86.71%
14TPL	0.98	0.91	0.87	0.82	0.92	0.86	87.02%
14TNM	0.98	0.97	0.94	0.91	0.96	0.94	92.47%
14TNL	0.99	0.97	0.95	0.90	0.97	0.93	92.30%
14TML	0.96	0.92	0.90	0.81	0.93	0.86	87.81%

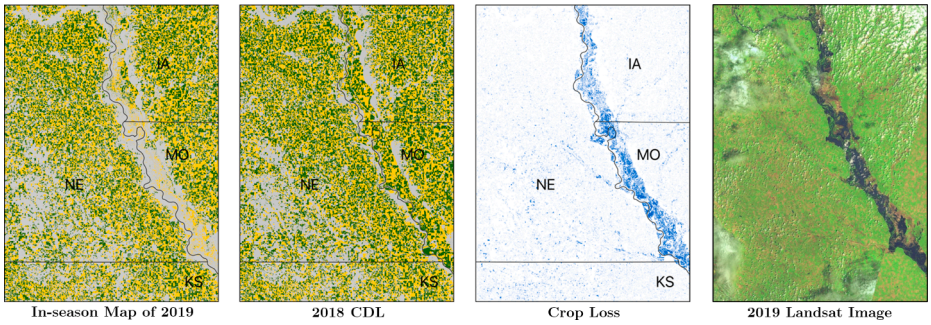


Fig. 11. Land cover change before and after the 2019 Midwest flooding. The crop loss map reflects the difference between 2019 in-season map and 2018 CDL. The Landsat image of June 2019 is composited by Band 6, Band 5, and Band 2.

trusted pixel prediction model handles CDL data while the crop type classification model deals with satellite images. Hence, the patterns learned by the two models are completely different. The trusted pixel prediction model can automatically recognize common crop rotation patterns (e.g., monocropping or alternate cropping). Some previous studies have explored the corn-soybean rotation pixels in the CDL (Sahajpal et al., 2014; Stern et al., 2012). However, they failed to investigate the crop sequence patterns that are not strictly periodic like "C-S-C-C-C-S-C-S-C" or "C-S-S-C-S-C-S-C-S" ("C" represents corn, "S" represents soybeans). The anomaly in these patterns could be due to policy change, fallow, as well as the mixed or misclassified pixel in CDL. The advantage of the trusted pixel prediction model is the neural network can automatically learn these corn-soybean rotation patterns with a minor anomaly.

The crop type classification model learns the features from the time-series profile of spectral bands and indices in the satellite images. According to the experiment result, the recall rate of in-season crop mapping result is highly correlated with the recall rate of trusted pixel while the precision rate of in-season mapping result has no obvious correlation with the precision and recall rate trusted pixel. Since the recall rate represents the fraction of the trusted pixels that are successfully predicted, it can be inferred that the classification performance could be affected by the quantity of trusted pixels. The more trusted pixels are extracted, the more spectral and temporal patterns can be recognized in the crop type classification model. On the contrary, if the volume of training set is small, for example, using pixels based on hand-specified rotation patterns as training samples, only limited spectral and temporal features would be learned by the neural network. This is also reflected in the Fig. 9 where the standard deviation is significantly higher for the recall than for the precision as the trusted pixel recall has the wide interquartile range.

The structure of the neural network used in the proposed workflow can be further improved. The neural network works as a "black box" in the process of trusted pixel prediction and crop type classification. This is a common issue for the deep learning models because of their nested non-linear structure (Samek et al., 2017). Since the development of new deep learning algorithm is not the main objective of this study, we did not pay much attention to the tuning of neural network. To achieve a better performance, the hyperparameters, such as learning rates, weights, optimizer, loss function, activation function, and batch size, need to be specifically tuned for the classification model of each scene.

5.3. Uncertainty of cropland data layer

To train the trusted pixel prediction model, we constructed the training set using CDL data from 2007. Technically, the more years included in the training set, the more features the trusted pixel prediction model can learn. This is based on the characteristics of the neural network. There are two main reasons for abandoning the early-year CDL. On the one hand, the coverage of CDL for the study area was

incomplete before 2007. On the other hand, there were many misclassified pixels in the early-year CDLs due to the cloud or lack of satellite images (Zhang et al., 2020b).

This study used CDL and field data as reference data to evaluate mapping results. From the experiment result of Section 4.2, we noticed that the agreement of in-season mapping results with ground truth is slightly higher than CDL. This difference could be primarily caused by the uncertainty in CDL data (Liu et al., 2004). It is worth noting that the current CDL data still contain a certain amount of errors, such as misclassified pixels, mixed pixels, and noisy pixels. According to the accuracy assessment of 2018 CDL by USDA NASS (USDA-NASS, 2019b), the OA of corn and soybeans in the study area is between 88% and 93%. The major crop types for a CDL state will normally have a classification accuracy of 85% to 95%. There is potential uncertainty in the field data as well. Waldner et al. (2019) discussed the biases associated with the field data collected using the roadside sampling strategy and pointed the roadside samples were less representative than random samples. Thus, the actual accuracy of the in-season map for many areas may vary with the validation result.

5.4. Limitations and potential solutions

As a recognized problem in LULC mapping, the cloud coverage in the remote sensing images could significantly affect the performance of crop type classification. Although only low cloud-covered (< 10%) satellite images were taken, it is unavoidable that some pixels in the data set contain noisy values, especially when mapping over large geographic area. A potential method to address cloud issue is to build the multi-temporal image stack using the Harmonized Landsat and Sentinel-2 (HLS) surface reflectance data set (Claverie et al., 2018). The HLS data set is composed of Landsat-8 data and Sentinel-2 data, which has been utilized in agricultural applications such as national-scale crop mapping (Griffiths et al., 2019) and identification of crop intensity (Hao et al., 2019). Another solution to tackle the cloud cover issue is combining radar data, such as Sentinel-1 data, with multispectral data, which have been proven effective to improve the crop mapping result (Ienco et al., 2019; Steinhausen et al., 2018).

There are several ways to further improve the current mapping result. As the major drawback of per-tile mapping, the tiling effect may take place in which temporal information of the image stack is significantly inconsistent between adjacent tiles. According to the experiment result, the tiling effect is ignorable in most areas but still noticeable on a few tiles. Since the availability of cloud-free images may vary from tile to tile, it is challenging to eliminate the tiling effect over the entire study area. To address this challenge, the spatial and temporal interpolation module could be integrated into the proposed workflow, which has been proven effective to improve the robustness of the crop type classification using high temporal and spatial resolution satellite images (Inglada et al., 2015). On the other hand, the pixel-based mapping may lead the salt-and-pepper effect on some land units. The CLU data could efficiently remove noises and misclassified pixels within the land unit. However, the CLU data is unavailable for public use. To make up the lack of CLU, we can apply boundary delineation methods to generate field boundaries for the object-based image analysis (Belgiu and Csillik, 2018; North et al., 2019).

The ANN model learns the crop sequence features in the historical CDL. But it is difficult to predict the crop type for land units that break the pattern in the predicting year. The traditional crop rotation patterns could be changed due to dynamic and uncertain factors, such as agricultural practices change, natural hazards, climate change effects, market situation, loss of soil fertility, water scarcity, government policy, and other socioeconomic factors. The Section 4.1 has demonstrated the capability of automatically recognizing variability in terms of crop rotation patterns using machine learning. In the next phase, we will systematically investigate new agricultural practices that are more sustainable for environmental resources based on this study. Furthermore, we will incorporate other features, such as agricultural commodity prices and weather information, into the training of prediction model.

6. Conclusion

This paper introduced an efficient and effective approach for rapid in-season field-level mapping of corn and soybean. The innovation of the proposed method is the use of trusted pixels, which compensate for the lack of ground truth data in the early growing season. By applying ANN with the historical CDL data, a considerable number of trusted pixels can be automatically identified and used to label training samples on satellite images. According to the experiments on the U.S. Corn Belt, it was found the average trusted-pixel precision of corn and soybean can reach 0.86 and 0.91 respectively. The trusted pixels were used to label training samples for classifying crop types on 49 scenes of Landsat-8 data and 31 tiles of Sentinel-2 data. Compared with the 2018 CDL, most tested scenes can achieve 85%–95% overall agreement with CDL by using images acquired during the period from early May to mid-July 2018. The validation of the 2018 mapping result against the limited numbers of ground truths in Nebraska showed an overall accuracy higher than 96%. Furthermore, the crop loss due to the Midwest flooding 2019 was successfully highlighted in the in-season crop map by the end of June.

This study has explored the feasibility of using trusted pixels for rapid in-season mapping of dominant crops (i.e., corn and soybeans) over the U.S. Corn Belt. We have successfully demonstrated trusted pixels can replace ground truth data and label satellite images for large-area crop mapping with MLP-based ANN. It can be inferred that the trusted pixel approach is also suitable for other common supervised classifiers in remote sensing image classification, such as random forest, decision tree, classification and regression tree (CART), and support vector machine (SVM). In future research, we will identify more crop types and scale up the framework to the entire CONUS. In addition, there are many subsequent studies could be conducted based on the findings of this paper. For example, the in-season crop cover map could be incorporated with the historical CDL to calculate the historical cropping patterns then predict the future crop acreage and production. We will also conduct a systematic study on the spatial and temporal trend of cropping over the U.S. Corn Belt in the future.

CRediT authorship contribution statement

Chen Zhang: Conceptualization, Methodology, Formal analysis, Writing - original draft. **Liping Di:** Conceptualization, Methodology, Supervision, Funding acquisition, Writing - review & editing. **Pengyu Hao:** Methodology, Investigation. **Zhengwei Yang:** Supervision, Resources. **Li Lin:** Data curation, Validation. **Haoteng Zhao:** Data curation, Validation. **Liying Guo:** Investigation, Validation.

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.

Acknowledgments

This research is supported by a grant from National Science Foundation INFEWS program (Grant #: CNS-1739705, PI: Dr. Liping Di). The authors would like to thank three anonymous reviewers for their valuable comments and suggestions.

References

AAFC, 2018. Annual Crop Inventory 2018. https://open.canada.ca/data/en/dataset/1f2ad87e-6103-4ead-bdd5-147c33fa11e6.
AAFC, 2019. ISO 19131 Annual Crop Inventory - Data Product Specifications.

Arvor, D., Jonathan, M., Meirelles, M.S.P., Dubreuil, V., Durieux, L., 2011. Classification of MODIS EVI time series for crop mapping in the state of Mato Grosso, Brazil. Int. J. Remote Sens. 32, 7847–7871. https://doi.org/10.1080/01431161.2010.531783.

Auch, R.F., Karstensen, K.A., 2015. Status and trends of land change in the Midwest-South Central United States - 1973 to 2000. USGS Numbered Series No. 1794-C. U.S. Geological Survey, Reston, VA.

Auch, R.F., Xian, G., Laingen, C.R., Sayler, K.L., Reker, R.R., 2018. Human drivers, biophysical changes, and climatic variation affecting contemporary cropping proportions in the northern prairie of the U.S. J. Land Use Sci. 13, 32–58. https://doi.org/10.1080/1747423X.2017.1413433.

Belgiu, M., Csillik, O., 2018. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. Remote Sens. Environ. 204, 509–523. https://doi.org/10.1016/j.rse.2017.10.005.

Boryan, C., Yang, Z., Di, L., Hunt, K., 2014. A new automatic stratification method for U. S. agricultural area sampling frame construction based on the cropland data layer. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 7, 4317–4327. https://doi.org/10.1109/JSTARS.2014.2322584.

Boryan, C., Yang, Z., Mueller, R., Craig, M., 2011. Monitoring US agriculture: The US department of agriculture, national agricultural statistics service, cropland data layer program. Geocarto Int. 26, 341–358. https://doi.org/10.1080/10106049.2011.562309.

Claverie, M., Ju, J., Masek, J.G., Dungan, J.L., Vermote, E.F., Roger, J.C., Skakun, S.V., Justice, C., 2018. The harmonized landsat and sentinel-2 surface reflectance data set. Remote Sens. Environ. 219, 145–161. https://doi.org/10.1016/j.rse.2018.09.002.

Dahal, D., Wylie, B., Howard, D., 2018. Rapid crop cover mapping for the conterminous United States. Sci. Rep. 8, 8631. https://doi.org/10.1038/s41598-018-26284-w.

Defourny, P., Bontemps, S., Bellemans, N., Cara, C., Dedieu, G., Guzzonato, E., Hagolle, O., Inglada, J., Nicola, L., Rabaute, T., Savinaud, M., Udroiu, C., Valero, S., Bégué, A., Dejoux, J.F., El Harti, A., Ezzahar, J., Kussul, N., Labbassi, K., Lebourgeois, V., Miao, Z., Newby, T., Nyamugama, A., Salh, N., Shelestov, A., Simonneaux, V., Traore, P.S., Traore, S.S., Koetz, B., 2019. Near real-time agriculture monitoring at national scale at parcel resolution: Performance assessment of the Sen2-Agri automated system in various cropping systems around the world. Remote Sens. Environ. 221, 551–568. https://doi.org/10.1016/j.rse.2018.11.007.

Demarez, V., Helen, F., Marais-Sicre, C., Baup, F., 2019. In-season mapping of irrigated crops using landsat 8 and sentinel-1 time series. Remote Sens. 11, 118. https://doi.org/10.3390/rs11020118.

Di, L., Yu, E.G., Kang, L., Shrestha, R., Bai, Y., 2017. RF-CLASS: A remote-sensing-based flood crop loss assessment cyber-service system for supporting crop statistics and insurance decision-making. J. Integrative Agriculture 16, 408–423. https://doi.org/10.1016/S2095-3119(16)61499-5.

Dong, T., Liu, J., Qian, B., Zhao, T., Jing, Q., Geng, X., Wang, J., Huffman, T., Shang, J., 2016. Estimating winter wheat biomass by assimilating leaf area index derived from fusion of Landsat-8 and MODIS data. Int. J. Appl. Earth Obs. Geoinf. 49, 63–74. https://doi.org/10.1016/j.jag.2016.02.001.

Edwards, J.H., Thurlow, D.L., Eason, J.T., 1988. Influence of tillage and crop rotation on yields of corn, soybean, and wheat. Agron. J. 80, 76–80. https://doi.org/10.2134/agronj1988.00021962008000010018x.

Fischer, T., Byerlee, D., Edmeades, G., 2014. Crop yields and global food security: Will yield increase continue to feed the world?

Fritz, S., See, L., McCallum, I., You, L., Bun, A., Moltchanova, E., Duerauer, M., Albrecht, F., Schill, C., Perger, C., Havlik, P., Mosnier, A., Thornton, P., Wood-Sichra, U., Herrero, M., Becker-Reshef, I., Justice, C., Hansen, M., Gong, P., Aziz, S. A., Cipriani, A., Cumani, R., Cecchi, G., Conchedda, G., Ferreira, S., Gomez, A., Haffani, M., Kayitakire, F., Malanding, J., Mueller, R., Newby, T., Nonguierma, A., Olusegun, A., Ortner, S., Rajak, D.R., Rocha, J., Schepaschenko, D., Schepaschenko, M., Terekhov, A., Tiangwa, A., Vancutsem, C., Vintrou, E., Wenbin, W., van der Velde, M., Dunwoody, A., Kraxner, F., Obersteiner, M., 2015. Mapping global cropland and field size. Glob. Change Biol. 21, 1980–1992. https://doi.org/10.1111/gcb.12838.

Gao, B.C., 1996. NDWI A normalized difference water index for remote sensing of vegetation liquid water from space. Remote Sens. Environ. 58, 257–266.

Gao, F., Anderson, M., Daughtry, C., Karnieli, A., Hively, D., Kustas, W., 2020. A within-season approach for detecting early growth stages in corn and soybean using high temporal and spatial resolution imagery. Remote Sens. Environ. 242, 111752 https://doi.org/10.1016/j.rse.2020.111752.

Giordano, S., Bailly, S., Landrieu, L., Chehata, N., 2020. Improved crop classification with rotation knowledge using sentinel-1 and -2 time series. Photogramm. Eng. Remote Sens. 86, 431–441. https://doi.org/10.14358/PERS.86.7.431.

Google, 2021. USDA NASS Cropland Data Layers. Earth Engine Data Catalog.

Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens. Environ. 202, 18–27. https://doi.org/10.1016/j.rse.2017.06.031.

Green, T.R., Kipka, H., David, O., McMaster, G.S., 2018. Where is the USA Corn Belt, and how is it changing? Sci. Total Environ. 618, 1613–1618. https://doi.org/10.1016/j.scitotenv.2017.09.325.

Griffiths, P., Nendel, C., Hostert, P., 2019. Intra-annual reflectance composites from Sentinel-2 and Landsat for national-scale crop and land cover mapping. Remote Sens. Environ. 220, 135–151. https://doi.org/10.1016/j.rse.2018.10.031.

Han, W., Yang, Z., Di, L., Mueller, R., 2012. CropScape: A Web service based application for exploring and disseminating US conterminous geospatial cropland data products for decision support. Comput. Electron. Agric. 84, 111–123. https://doi.org/10.1016/j.compag.2012.03.005.

Hansen, M.C., Loveland, T.R., 2012. A review of large area monitoring of land cover change using Landsat data. Remote Sens. Environ., Landsat Legacy Special issue 122, 66–74. https://doi.org/10.1016/j.rse.2011.08.024.

Hao, P., Di, L., Zhang, C., Guo, L., 2020. Transfer learning for crop classification with cropland data layer data (CDL) as training samples. Sci. Total Environ. 733, 138869 https://doi.org/10.1016/j.scitotenv.2020.138869.

Hao, P., Tang, H., Chen, Z., Liu, Z., 2018. Early-season crop mapping using improved artificial immune network (IAIN) and Sentinel data. PeerJ 6, e5431. https://doi.org/10.7717/peerj.5431.

Hao, P., Tang, H., Chen, Z., Yu, L., Wu, M., 2019. High resolution crop intensity mapping using harmonized Landsat-8 and Sentinel-2 data. J. Integrative Agriculture 18, 2883–2897. https://doi.org/10.1016/S2095-3119(19)62599-2.

Hao, P., Zhan, Y., 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. https://doi.org/10.3390/rs70505347.

Huffstutter, P.J., Pamuk, H., 2019. 1 Million Acres of Midwest Farmlands Flooded as Corn Planting Deadlines Approach. Insurance J. https://www.insurancejournal.com/news/midwest/2019/04/01/522389.htm.

Ienco, D., Interdonato, R., Gaetano, R., Ho Tong Minh, D., 2019. Combining sentinel-1 and sentinel-2 satellite image time series for land cover mapping via a multi-source deep learning architecture. ISPRS J. Photogramm. Remote Sens. 158, 11–22. https://doi.org/10.1016/j.isprsjprs.2019.09.016.

Inglada, J., Arias, M., Tardy, B., Hagolle, O., Valero, S., Morin, D., Dedieu, G., Sepulcre, G., Bontemps, S., Defourny, P., Koetz, B., Inglada, J., Arias, M., Tardy, B., Hagolle, O., Valero, S., Morin, D., Dedieu, G., Sepulcre, G., Bontemps, S., Defourny, P., Koetz, B., 2015. Assessment of an operational system for crop type map production using high temporal and spatial resolution satellite optical imagery. Remote Sens. 7, 12356–12379. https://doi.org/10.3390/rs70912356.

Johnson, D.M., 2019. Using the Landsat archive to map crop cover history across the United States. Remote Sens. Environ. 232, 111286 https://doi.org/10.1016/j.rse.2019.111286.

Karlen, D.L., Hurley, E.G., Andrews, S.S., Cambardella, C.A., Meek, D.W., Duffy, M.D., Mallarino, A.P., 2006. Crop rotation effects on soil quality at three northern corn/soybean belt locations. Agron. J. 98, 484–495. https://doi.org/10.2134/agronj2005.0098.

Kollas, C., Kersebaum, K.C., Nendel, C., Manevski, K., Müller, C., Palosuo, T., Armas-Herrera, C.M., Beaudoin, N., Bindi, M., Charfeddine, M., Conradt, T., Constantin, J., Eitzinger, J., Ewert, F., Ferrise, R., Gaiser, T., de Cortazar-Atauri, I.G., Giglio, L., Hlavinka, P., Hoffmann, H., Hoffmann, M.P., Launay, M., Manderscheid, R., Mary, B., Mirschel, W., Moriondo, M., Olesen, J.E., Öztürk, I., Pacholski, A., Ripoche-Wachter, D., Roggero, P.P., Roncossek, S., Rötter, R.P., Ruget, F., Sharif, B., Trnka, M., Ventrella, D., Waha, K., Wegehenkel, M., Weigel, H.J., Wu, L., 2015. Crop rotation modelling model intercomparison. Eur. J. Agron. 70, 98–111. https://doi.org/10.1016/j.eja.2015.06.007.

Kussul, N., Mykola, L., Shelestov, A., Skakun, S., 2018. Crop inventory at regional scale in Ukraine: Developing in season and end of season crop maps with multi-temporal optical and SAR satellite imagery. Eur. J. Remote Sens. 51, 627–636. https://doi.org/10.1080/22797254.2018.1454265.

Lark, T.J., Mueller, R.M., Johnson, D.M., Gibbs, H.K., 2017. Measuring land-use and land-cover change using the U.S. Department of agriculture's cropland data layer: Cautions and recommendations. Int. J. Appl. Earth Obs. Geoinf. 62, 224–235. https://doi.org/10.1016/j.jag.2017.06.007.

Liu, J., Huffman, T., Shang, J., Qian, B., Dong, T., Zhang, Y., 2016. Identifying major crop types in eastern Canada using a fuzzy decision tree classifier and phenological indicators derived from time series MODIS data. Can. J. Remote Sens. 42, 259–273. https://doi.org/10.1080/07038992.2016.1171133.

Liu, J., Liu, M., Tian, H., Zhuang, D., Zhang, Z., Zhang, W., Tang, X., Deng, X., 2005. Spatial and temporal patterns of China's cropland during 1990: An analysis based on Landsat TM data. Remote Sens. Environ. 98, 442–456. https://doi.org/10.1016/j.rse.2005.08.012.

Liu, W., Gopal, S., Woodcock, C.E., 2004. Uncertainty and confidence in land cover classification using a hybrid classifier approach. Photogramm. Eng. Remote Sens. 70, 963–971. https://doi.org/10.14358/PERS.70.8.963.

Lobell, D.B., 2013. The use of satellite data for crop yield gap analysis. Field crops research, crop yield gap analysis rationale. Methods and Applications 143, 56–64. https://doi.org/10.1016/j.fcr.2012.08.008.

Löw, 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. https://doi.org/10.1016/J.ISPRSJPRS.2013.08.007.

Massey, R., Sankey, T.T., Yadav, K., Congalton, R.G., Tilton, J.C., 2018. Integrating cloud-based workflows in continental-scale cropland extent classification. Remote Sens. Environ. 219, 162–179. https://doi.org/10.1016/j.rse.2018.10.013.

McNairn, H., Kross, A., Lapen, D., Caves, R., Shang, J., 2014. Early season monitoring of corn and soybeans with TerraSAR-X and RADARSAT-2. Int. J. Appl. Earth Obs. Geoinf. 28, 252–259. https://doi.org/10.1016/j.jag.2013.12.015.

Mulla, D.J., 2013. Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. Biosyst. Eng., Special Issue: Sens. Technol. Sustain. Agriculture 114, 358–371. https://doi.org/10.1016/j.biosystemseng.2012.08.009.

North, H.C., Pairman, D., Belliss, S.E., 2019. Boundary delineation of agricultural fields in multitemporal satellite imagery. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 12, 237–251. https://doi.org/10.1109/JSTARS.2018.2884513.

Osman, J., Inglada, J., Dejoux, J.F., 2015. Assessment of a markov logic model of crop rotations for early crop mapping. Comput. Electron. Agric. 113, 234–243. https://doi.org/10.1016/j.compag.2015.02.015.

Phalke, A.R., Özdoğan, M., 2018. Large area cropland extent mapping with Landsat data and a generalized classifier. Remote Sens. Environ. 219, 180–195. https://doi.org/10.1016/j.rse.2018.09.025.

Prasad, A.K., Chai, L., Singh, R.P., Kafatos, M., 2006. Crop yield estimation model for Iowa using remote sensing and surface parameters. Int. J. Appl. Earth Obs. Geoinf. 8, 26–33. https://doi.org/10.1016/j.jag.2005.06.002.

Ruan, X., Huang, J., Williams, D.A.R., Harker, K.J., Gergel, S.E., 2019. High spatial resolution landscape indicators show promise in explaining water quality in urban streams. Ecol. Ind. 103, 321–330. https://doi.org/10.1016/j.ecolind.2019.03.013.

Sahajpal, R., Zhang, X., Izaurralde, R.C., Gelfand, I., Hurtt, G.C., 2014. Identifying representative crop rotation patterns and grassland loss in the US Western Corn Belt. Comput. Electron. Agric. 108, 173–182. https://doi.org/10.1016/j.compag.2014.08.005.

Sakamoto, T., Gitelson, A.A., Arkebauer, T.J., 2014. Near real-time prediction of U.S. Corn yields based on time-series MODIS data. Remote Sens. Environ. 147, 219–231. https://doi.org/10.1016/j.rse.2014.03.008.

Samek, W., Wiegand, T., Müller, K.R., 2017. Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models. arXiv: 1708.08296 [cs, stat].

Shrestha, R., Di, L., Yu, E.G., Kang, L., Shao, Y., Bai, Y., 2017. Regression model to estimate flood impact on corn yield using MODIS NDVI and USDA cropland data layer. J. Integrative Agriculture 16, 398–407. https://doi.org/10.1016/S2095-3119(16)61502-2.

Skakun, S., Franch, B., Vermote, E., Roger, J.C., Becker-Reshef, I., Justice, C., Kussul, N., 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. https://doi.org/10.1016/j.rse.2017.04.026.

Song, X.P., Potapov, P.V., Krylov, A., King, L., Di Bella, C.M., Hudson, A., Khan, A., Adusei, B., Stehman, S.V., Hansen, M.C., 2017. National-scale soybean mapping and area estimation in the United States using medium resolution satellite imagery and field survey. Remote Sens. Environ. 190, 383–395. https://doi.org/10.1016/J.RSE.2017.01.008.

Steinhausen, M.J., Wagner, P.D., Narasimhan, B., Waske, B., 2018. Combining Sentinel-1 and Sentinel-2 data for improved land use and land cover mapping of monsoon regions. Int. J. Appl. Earth Obs. Geoinf. 73, 595–604. https://doi.org/10.1016/j.jag.2018.08.011.

Stern, A.J., Doraiswamy, P.C., Raymond Hunt, E., 2012. Changes of crop rotation in Iowa determined from the United States Department of Agriculture, National Agricultural Statistics Service cropland data layer product. J. Appl. Remote Sens. 6, 063590 https://doi.org/10.1117/1.JRS.6.063590.

Tardy, B., Inglada, J., Michel, J., 2017. Fusion approaches for land cover map production using high resolution image time series without reference data of the corresponding period. Remote Sens. 9, 1151. https://doi.org/10.3390/rs9111151.

Taylor, J., Acevedo, W., Auch, R.F., Drummond, M.A., 2015. Status and trends of land change in the Great Plains of the United States - 1973 to 2000. USGS Numbered Series No. 1794-B. U.S. Geological Survey, Reston, VA.

Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens. Environ. 8, 127–150. https://doi.org/10.1016/0034-4257(79)90013-0.

USDA-NASS, 2010. Usual Planting and Harvesting Dates for U.S. Field Crops.

USDA-NASS, 2019a. Cropland Data Layer Releases. https://www.nass.usda.gov/Research_and_Science/Cropland/metadata/meta.php.

USDA-NASS, 2019b. Cropland Data Layer Metadata. https://www.nass.usda.gov/Research_and_Science/Cropland/Release/.

Van Eerd, L.L., Congreves, K.A., Hayes, A., Verhallen, A., Hooker, D.C., 2014. Long-term tillage and crop rotation effects on soil quality, organic carbon, and total nitrogen. Can. J. Soil Sci. 94, 303–315. https://doi.org/10.4141/cjss2013-093.

Vaudour, E., Noirot-Cosson, P.E., Membrive, O., 2015. Early-season mapping of crops and cultural operations using very high spatial resolution Pléiades images. Int. J. Appl. Earth Obs. Geoinf. 42, 128–141. https://doi.org/10.1016/j.jag.2015.06.003.

Waldner, F., Abelleyra, D.D., Verón, S.R., Zhang, M., Wu, B., Plotnikov, D., Bartalev, S., Lavreniuk, M., Skakun, S., Kussul, N., Maire, G.L., Dupuy, S., Jarvis, I., Defourny, P., 2016. Towards a set of agrosystem-specific cropland mapping methods to address the global cropland diversity. Int. J. Remote Sens. 37, 3196–3231. https://doi.org/10.1080/01431161.2016.1194545.

Waldner, F., Bellemans, N., Hochman, Z., Newby, T., de Abelleyra, D., Verón, S.R., Bartalev, S., Lavreniuk, M., Kussul, N., Maire, G.L., Simoes, M., Skakun, S., Defourny, P., 2019. Roadside collection of training data for cropland mapping is viable when environmental and management gradients are surveyed. Int. J. Appl. Earth Obs. Geoinf. 80, 82–93. https://doi.org/10.1016/j.jag.2019.01.002.

Wang, M., Liu, Z., Ali Baig, M.H., Wang, Y., Li, Y., Chen, Y., 2019. Mapping sugarcane in complex landscapes by integrating multi-temporal Sentinel-2 images and machine learning algorithms. Land Use Policy 88, 104190. https://doi.org/10.1016/j.landusepol.2019.104190.

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. https://doi.org/10.1016/J.RSE.2007.07.019.

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 U.S. Central Great Plains. Remote Sens. Environ. 108, 290–310. https://doi.org/10.1016/J.RSE.2006.11.021.

Wu, X., Zhang, X., 2019. An efficient pixel clustering-based method for mining spatial sequential patterns from serial remote sensing images. Comput. Geosci. 124, 128–139. https://doi.org/10.1016/j.cageo.2019.01.005.

Xiao, Y., Mignolet, C., Mari, J.F., Benoît, M., 2014. Modeling the spatial distribution of crop sequences at a large regional scale using land-cover survey data: A case from France. Comput. Electron. Agric. 102, 51–63. https://doi.org/10.1016/j.compag.2014.01.010.

Yin, X., Kersebaum, K.C., Kollas, C., Baby, S., Beaudoin, N., Manevski, K., Palosuo, T., Nendel, C., Wu, L., Hoffmann, M., Hoffmann, H., Sharif, B., Armas-Herrera, C.M., Bindi, M., Charfeddine, M., Conradt, T., Constantin, J., Ewert, F., Ferrise, R., Gaiser, T., de Cortazar-Atauri, I.G., Giglio, L., Hlavinka, P., Lana, M., Launay, M., Louarn, G., Manderscheid, R., Mary, B., Mirschel, W., Moriondo, M., Öztürk, I., Pacholski, A., Ripoche-Wachter, D., Rötter, R.P., Ruget, F., Trnka, M., Ventrella, D., Weigel, H.J., Olesen, J.E., 2017. Multi-model uncertainty analysis in predicting grain N for crop rotations in Europe. Eur. J. Agron. 84, 152–165. https://doi.org/10.1016/j.eja.2016.12.009.

Zhang, C., Di, L., Lin, L., Guo, L., 2019a. 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://doi.org/10.1016/j.compag.2019.104989.

Zhang, C., Di, L., Yang, Z., Lin, L., Hao, P., 2020a. AgKit4EE: A toolkit for agricultural land use modeling of the conterminous United States based on Google Earth Engine. Environ. Modell. Software 129, 104694. https://doi.org/10.1016/j.envsoft.2020.104694.

Zhang, C., Di, L., Yang, Z., Lin, L., Yu, E.G., Yu, Z., Rahman, M.S., Zhao, H., 2019b. Cloud Environment for Disseminating NASS Cropland Data Layer, in: 2019 8th International Conference on Agro-Geoinformatics (Agro-Geoinformatics). pp. 1–5. https://doi.org/10.1109/Agro-Geoinformatics.2019.8820465.

Zhang, C., Yang, Z., Di, L., Lin, L., Hao, P., 2020b. Refinement of Cropland Data Layer Using Machine Learning, in: ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. Baltimore, Maryland, USA, pp. 161–164. https://doi.org/10.5194/isprs-archives-XLII-3-W11-161-2020.