

Automated In-Season Crop-Type Data Layer Mapping Without Ground Truth for the Conterminous United States Based on Multisource Satellite Imagery

Hui Li¹, Liping Di¹, Senior Member, IEEE, Chen Zhang¹, Member, IEEE, Li Lin¹, Member, IEEE, Liying Guo¹, Member, IEEE, Eugene G. Yu, Member, IEEE, and Zhengwei Yang¹, Member, IEEE

Abstract—Mapping nationwide in-season crop-type data is a significant and challenging task in agriculture remote sensing. The existing data product for U.S. crop-type planting, such as the Cropland Data Layer (CDL), falls short in facilitating near-real-time applications. This article designed a workflow aimed at automating the generation of in-season CDL-like products for USA. We methodically extracted trusted pixels as land cover labels from historical CDL datasets, employing Sentinel-2, Landsat 8, and Landsat-9 as sources for spectrum data, using the random forest classifier to conduct nationwide crop-type classifications. These classifications were integrated into the In-Season Crop Data Layer (ICDL) covering the entire Conterminous United States (CONUS). This approach facilitated the efficient generation of ICDLs for May, June, and July 2022, achieving satisfactory accuracy in July. Compared to Nebraska and Iowa ground truth data, ICDL achieved F1 scores of (0.911, 0.845) for corn and (0.959, 0.969) for soybean. Furthermore, ICDL's regional acreage estimates for major crops (corn, soybean, spring wheat, cotton, winter wheat, and rice) closely align with the U.S. Department of Agriculture (USDA) National Agricultural Statistics Service (NASS) figures, showing minimal variances as low as (0.01%, -0.68%, 0.19%, -4.39%, -5.78%, -1.28%). Notably, ICDL outperforms CDL in most assessments. This research consistently produces annual ICDLs from May to July that are readily accessible to the public in the iCrop system. Simultaneously, it presents an alternative technique for nationwide, in-season mapping of crop types.

Index Terms—Cropland Data Layer (CDL), In-Season Cropland Data Layer (ICDL), Landsat, Sentinel-2, trusted pixel.

I. INTRODUCTION

THE U.S. croplands play a crucial role in global sustenance by not only producing food but also serving as a vital

Manuscript received 8 March 2023; revised 3 June 2023 and 31 December 2023; accepted 29 January 2024. Date of publication 5 February 2024; date of current version 15 February 2024. The work of Liping Di was supported in part by the United States Department of Agriculture National Institute of Food and Agriculture (USDA) under Grant 2021-67021-34151; and in part by the National Science Foundation under Grant CNS-1739705, Grant 2228000, and Grant 2236137. (Corresponding author: Liping Di.)

Hui Li, Liping Di, Chen Zhang, Li Lin, Liying Guo, and Eugene G. Yu are with the Center for Spatial Information Science and Systems, George Mason University, Fairfax, VA 22030 USA (e-mail: hli47@gmu.edu; ldi@gmu.edu; czhang11@gmu.edu; llin2@gmu.edu; lguo2@gmu.edu; gyu@gmu.edu).

Zhengwei Yang is with the National Agricultural Statistics Service, U.S. Department of Agriculture, Washington, DC 20250 USA (e-mail: zhengwei.yang@usda.gov).

Digital Object Identifier 10.1109/TGRS.2024.3361895

source of raw materials for various industries [1]. Consequently, the cultivation and yield of crops in USA significantly influence global food and energy security. Recognizing this, an in-season map detailing crop types across the entire country becomes essential, providing crucial insights into the location and extent of crops planted during the growing season. Simultaneously, an up-to-date national-scale crop-type map proves invaluable for agricultural research and decision-making processes. This includes applications such as comprehensive yield predictions, investigations into biofuel storage, analysis of grain commodity prices, and accurate estimation of crop losses during disasters. The U.S. Department of Agriculture (USDA) National Agricultural Statistics Service (NASS) has consistently generated the annual Cropland Data Layer (CDL) for many years. This comprehensive dataset includes information on over 100 crop types and other significant nonagricultural land use types, covering the entire Conterminous United States (CONUS) with a spatial resolution of 30 m. As reported by Boryan et al. [2], the See5 decision tree classifier model underwent training for each Landsat scene, utilizing ground truths provided by farmers regarding crop types at USDA's Farm Service Agency Common Land Units (CLU). Notably, the overall classification accuracy for major crops, such as corn and soybean, exhibited a range between 85% and 95% in the year 2009. The CDL product is accessible through CropScape [3] and has found extensive utility in diverse agricultural applications and decision-making processes. Its applications include flood monitoring and crop-loss estimation [4], the extraction of crop masks for early season winter wheat identification [5], and the estimation of land cover changes [6]. Nevertheless, the CDL is a post-season product typically accessible in January or February of the subsequent year. This temporal delay renders numerous in-season agricultural applications and decision-making processes impractical.

Previous researchers have delved into in-season crop mapping using remote sensing and machine learning models [7], [8], [9], [10], [11], [12], [13], [14], [15], [16]. The success of some endeavors relies on the availability of high-quality crop-type ground truth data, which is essential for accurately labeling satellite image pixels as training data that play a pivotal role in determining classification performance.

Researchers commonly gather ground truth samples directly from fields to construct the training dataset [17]. In recent years, researchers also extracted crop category data from historical official cropland maps. Wang et al. [18] collected a certain number of random corn and soybean pixels from historical CDL that acted as the crop category data to participate in the supervised classification. Li et al. [19] also used 2011 and 2014 CDL crop pixels without field edge training the random forest classifier to identify the crop in California. In the realm of large-scale mapping, Pazúr et al. [39] employed farmers reported data and Sentinel-2 imagery as the primary data source for classification, relying on random forest classifier to generate distribution maps for cropland and permanent grassland across the entirety of Switzerland. In addition, Blickensdörfer et al. [47] integrated data from Sentinel-2, Landsat 8, Sentinel-1, and other environment and agricultural practices data to map crop types in Germany. Notably, they undertook the resampling of Landsat 8 data to align with the resolution of Sentinel-2. Johnson and Mueller [22] conducted a study testing four data utility strategies (CDL, Landsat 7/8, Sentinel-2, and a combination of all) using a random forest classifier to generate in-season land cover maps for the Corn Belt. Their findings confirmed that the integration of CDL, full-season Landsat 7/8, and Sentinel-2 imagery yielded the best classification performance for corn and soybean. Cai et al. [23] employed USDA CLUs in conjunction with multiple Landsat imagery datasets. They trained a deep neural network model specifically for classifying corn and soybeans in Illinois county, achieving consistently high overall accuracy across the years 2000–2015.

In the context of crop-type mapping for USA, nevertheless, ground survey data, such as CLUs, are usually inaccessible. Moreover, the CDL still possesses misclassifications, introducing uncertainty into the process of directly extracting crop labels. To overcome the challenges, a concept of trusted pixels and an extraction method were introduced, grounded in U.S. agriculture rotation practices. This innovative approach forecasts current-year crop types in specific locations with consistent rotation patterns identified in historical CDL. Several studies have substantiated the viability of this method, showcasing advanced classification performance for crop mapping during the early stages of the growing season [10], [12], [13], [14], [22], [24]. In detail, most of them employed trusted pixels of corn, cotton, rice, and soybean extracted from 2008–2019 CDL in the middle-west of USA based on three kinds of intricate rotation patterns.

Rather than diversifying data source strategies, some peers focus their efforts on innovating machine learning algorithms. Hang et al. [25] developed a cascaded recurrent neural network (RNN) model for hyperspectral image classification. This model has two RNN layers to refine overflow information and enhance correlative features. Subsequently, the features weighted fusion and loss functions weighted combination strategies were built to efficiently connect these two layers. In addition, the convolutional layers were also incorporated to recognize spectral and spatial information. This model achieved higher overall cropland classification accuracy than RNN for two different datasets. Hong et al. [26] introduced

mini graph convolutional networks (miniGCNs) as a solution to address the disadvantages associated with traditional graph convolutional networks (GCNs). These networks were designed to be trained in a minibatch fashion, aiming to discover improved and more robust local optima. The concatenation fusion strategies of miniGCN and convolutional neural network (CNN) reached the highest overall accuracy in crop classification when compared to ten models. Rußwurm et al. [27] developed an end-to-end learned early classification of time-series (ELECTS) RNN. This model utilized multiple remote sensing imagery and official crop datasets to achieve in-season crop-type mapping across four geographically distinct locations, demonstrating a notable level of accuracy.

Nevertheless, there are persistent challenges that require attention. The integration of multiple spectrum datasets with different widths probably introduces uncertainties in the classification process. Extended history periods and complex rotation patterns usually cause a scarcity of trusted pixels and lower accuracy. Machine learning models based on network structures often demand a substantial amount of computation resources when dealing with super large-scale mapping, such as for USA, where careful consideration must be given to computational costs. To efficiently produce U.S. in-season crop-type maps, this article implemented an automated mapping-without-ground-truth workflow on multisource satellite imagery. We utilized the 2017–2021 Crop Data Layer (CDL) to extract trusted pixels serving as land cover labels for the year 2022. Subsequently, we conducted random forest classifications independently for Landsat 8/9 and Sentinel-2A datasets. The results of these classifications were then methodically integrated into the In-Season Crop Data Layer (ICDL) covering the entire CONUS. This approach facilitated the automated generation of ICDLs as early as the end of May. Notably, by the conclusion of July 2022, the accuracy reached exceptional levels, surpassing that of CDL. These ICDL products come out around six months earlier than CDL updates. This research mainly has three contributions.

- 1) Employing the last five-year CDL to extract a substantial quantity of trusted pixels as land cover labels, expanding the trusted pixel methodology from Corn Belt to CONUS for all land cover types, concentrating on two kinds of rotation patterns, as opposed to the previous approach that considered three patterns in approximately ten-year and only for specific crop types in the middle-west of USA.
- 2) Conducting random forest classifications for Landsat 8/9 and Sentinel-2A datasets independently, as opposed to the previous approach of merging datasets before classification, ignoring discrepancies in spectrum widths between corresponding bands of two datasets.
- 3) Pioneering the annual production of CONUS ICDLs during the months of May, June, and July, achieving a satisfactory level of accuracy by July.

The production environment is the public cloud computing platform—Google Earth Engine (GEE) [28] and Python of ArcGIS Pro. The remainder of this article will accomplish the

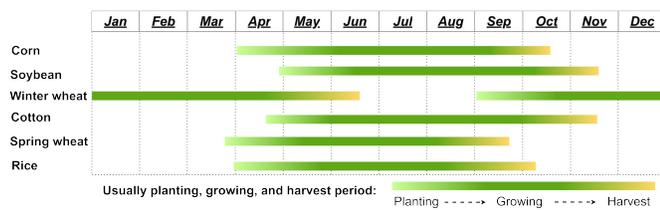


Fig. 1. Major crop's typical planting, growing, and harvest period in USA.

following tasks: elucidating the workflow involved in CONUS ICDL production, evaluating the accuracy of trusted pixel and classification methodologies, assessing the performance of ICDL in regional crop acreage estimation, and analyzing both the advantages and limitations inherent in this workflow.

II. METHODOLOGY

A. Study Area

The study area is the entire CONUS, including 48 states and the District of Columbia with a total size of 7 653 006 KM² [20]. According to the USDA NASS acreage estimates for 2022, the primary crops in USA, ranked from first to sixth in terms of acreage, are corn, soybean, winter wheat, cotton, spring wheat, and rice [21]. Corn and soybean are mainly located in the Corn Belt. Winter wheat and spring wheat are planted in the middle and northern areas such as Kansas, Oklahoma, Texas, South Dakota, North Dakota, Montana, and Washington. A great quantity of cotton is planted in Texas and Georgia. Rice mainly grows in Missouri, Arkansas, Louisiana, and Alabama delta regions. In USA, the annual planting season for major crops, excluding winter wheat, typically commences in April and concludes in June. The harvest, on the other hand, generally takes place between September and November. This established timeframe reflects the customary planting and harvest practices for these crops. This study concentrates on the May–July period, which is the early growing season of corn, soybean, cotton, spring wheat, and rice, and the late growing season of winter wheat. Fig. 1 illustrates the typical planting, growing, and harvest periods for the above major crops in USA [29].

B. Data Collection

To produce early season and in-season crop maps, we select Landsat-8, Landsat-9, and Sentinel-2A satellite images in May, May–June, and May–July as the primary sources for the time-series crop-type classification. The primary reason for this phenomenon is that a specific type of satellite imagery often lacks a sufficient number of cloud-free images in certain locations. The presence of cloud coverage can substantially impact the quality of time-series remote sensing classification [30]. Landsat 8 and Landsat 9 Operational Land Imager (OLI) surface reflectance products are generated by the Land Surface Reflectance Code algorithm, which both are atmospherically corrected surface reflectance images. Landsat 8 OLI products are available from 2013 that provide nine spectral bands with 30-m spatial resolution for all bands except the panchromatic band with 15 m, which has a 16-day revisit cycle [31]. The

Landsat 9, launched on September 27, 2021, whose sensor OLI-2 has the same capacity as Landsat 8's OLI, and therefore, they have the same spectral band structure, spatial resolution, and revisit frequency, but different observation times [32]. Landsat 8 and Landsat 9 swath sizes are both 185 × 180 km in a scene, and they have the same scene grids with the same path and row numbers [33].

Thus, we utilized them collectively as Landsat 8/9. There are 459 WRS2 descending scenes covering the CONUS. Sentinel-2A with a five-day revisit frequency on the CONUS published globally in December 2018 and provides bottom-of-atmosphere (BOA) reflectance images, which own 13 spectral bands [34]. Sentinel-2A imagery is presented in the Universal Transverse Mercator (UTM) projection, with a total of 990 scenes covering the CONUS [35], [36]. Each scene encompasses a ground area of 100 × 100 km. Table I shows the Landsat 8, Landsat 9, and Sentinel-2A properties, including spectral resolution, spatial resolution, revisit period, and coverage size. Meanwhile, we collected 2017–2021 CDL data to extract 2022 CONUS trusted pixels as crop-type and land cover labels. To operate the time-series classification scene-by-scene, the scene-tiling grids of Landsat 8/9 and Sentinel-2A were employed as scene-bound features. The WRS-2 Descending (daytime) grid shapefile has Landsat 8/9 scene-tiling grid geometry and path/row number [37]. Meyers [36] developed a Sentinel 2 tiling grid shapefile with every scene's polygon and identification. In addition, the 2021 1:500 000 states boundary shapefile served to restrict the mapping extent [38].

C. Workflow of ICDL Production

The process for generating ICDL is depicted in Fig. 2. Sentinel 2A and Landsat 8/9 images have different spectrum wavelengths and were used to generate the crop-type classification, separately. The shorter revisit period of Sentinel-2A results in a higher quantity of data, which positively impacts the performance of time-series classification. Consequently, Sentinel-2A classification takes precedence, with Landsat 8/9 classification utilized as a supplementary source. The entire production consists of three main steps: 1) scene-by-scene classification of multisource images for a state; 2) every state ICDL; and 3) CONUS ICDL. The top two steps were implemented in GEE, and the third step was conducted using Python of ArcGIS Pro.

Scene-by-scene classification in a specific state is the first step. For the time-series classification of a certain Sentinel-2A scene, a selection of multitemporary high-quality images is made by applying filters based on cloud-cover and no-date percentages. As well as Sentinel-2A is processed via resampling and spectrum calculation. Subsequently, trusted pixels within the extent of this scene are extracted from the 2017–2021 CDL. Following this, the selected trusted pixels, along with their corresponding time-series image pixels, constitute the training data used to train the random forest model, facilitating the execution of the time-series classification. This operation is iteratively applied to each Sentinel-2A scene, with the resulting classifications being mosaicked to construct a comprehensive Sentinel-2A crop-type map for this state.

TABLE I
CHARACTERS OF LANDSAT 8, LANDSAT 9, AND SENTINEL-2A IMAGERY DATA

	Landsat-8/9			Sentinel-2A		
	Band	Resolution	Wavelength (μm)	Band	Resolution	Wavelength (μm)
Spatial and spectral characters	Coastal aerosol	30 meters	0.43 - 0.45	Aerosols	60 meters	0.4423 - 0.4439
	Blue	30 meters	0.45 - 0.51	Blue	10 meters	0.4921 - 0.4966
	Green	30 meters	0.53 - 0.59	Green	10 meters	0.559 - 0.560
	Red	30 meters	0.64 - 0.67	Red	10 meters	0.6645 - 0.665
	NIR	30 meters	0.85 - 0.88	Red Edge 1	20 meters	0.7038 - 0.7039
	SWIR 1	30 meters	1.57 - 1.65	Red Edge 2	20 meters	0.7391 - 0.7402
	SWIR 2	30 meters	2.11 - 2.29	Red Edge 3	20 meters	0.7797 - 0.7825
	Panchromatic	15 meters	0.52 - 0.90	NIR	10 meters	0.833 - 0.8351
	Cirrus	15 meters	1.36 - 1.38	Red Edge 4	20 meters	0.864 - 0.8648
	Thermal infrared 1	30 meters	10.60 - 11.19	Water vapor	60 meters	0.9432 - 0.945
	Thermal infrared 2	30 meters	11.50 - 12.51	SWIR 1	20 meters	1.6104 - 1.6137
				SWIR 2	20 meters	2.1857 - 2.2024
	Revisit interval	16 days			5 days	
Scene coverage size	185×180 KM ²			100×100 KM ²		

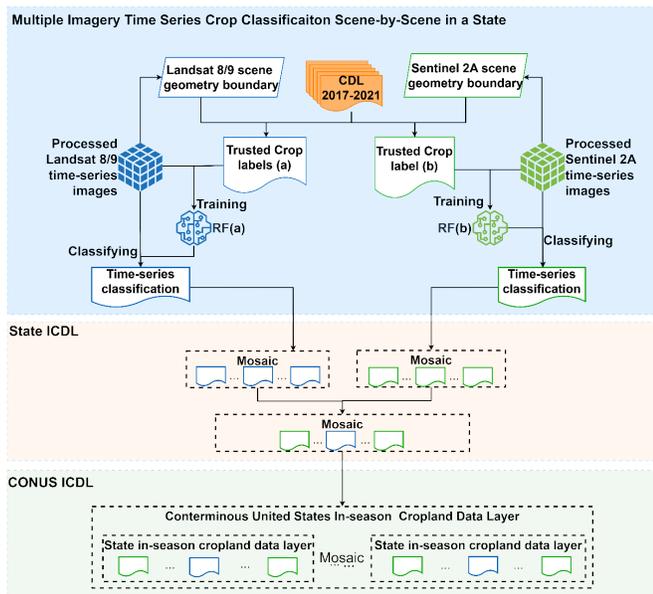


Fig. 2. Workflow of producing ICDL for CONUS.

In cases where the generated map does not cover the entire state area, additional processing is required for the Landsat 8/9 crop-type data to ensure comprehensive coverage. Similarly, the Landsat 8/9 classification for this state can be generated using the same method. Mosaicking these two classification results to generate a new comprehensive in-season crop type map for this state. This step is reiterated for each state, ensuring the repetition of the process to achieve a complete set of ICDL for every individual state. Finally, the ICDLs from all states are mosaicked to create a comprehensive dataset representing the ICDL for the entire CONUS. According to

this workflow, we produced the 2022 CONUS ICDL in May, June, and July.

D. Spectrum Data Processing

This article aims to automate the production of the ICDL using Sentinel-2A and Landsat 8/9 time-series images. Thus, high-quality multisource satellite imagery need to be collected and encapsulated into a time-series image stack in every scene before the classification. Sentinel-2A has cloud-cover and no-data properties, while Landsat 8/9 only has cloud-cover property, facilitating the extraction of high-quality images. Given the necessity for a 30-m resolution in the ICDL and considering that the Sentinel-2A images at a resolution of 10–20 m, a bilinear resampling process was implemented to adjust their resolution to 30-m before conducting the Sentinel-2A classification.

To provide a clear description of the high-quality image selection process, we will use Nebraska in May–July as an illustrative example. During the period from May to July in Nebraska, Sentinel-2A images with cloud cover of less than 5% and no-data properties of less than 8% were chosen to ensure the inclusion of images with minimal cloud interference and missing data. The occurrence of no-data in Sentinel-2A images is a common phenomenon observed across all states. Fig. 3(b) displays Sentinel-2 no-data areas' distribution and empty-image scene in the southeast corner of Nebraska, which need to be filled by Landsat 8/9 classification. The high-quality image selection process for Landsat 8/9 imagery involved setting a threshold of less than 6% for the cloud-cover property. Fig. 3(a) visualizes how Landsat 8/9 images were gathered in areas where Sentinel-2 data had missing values

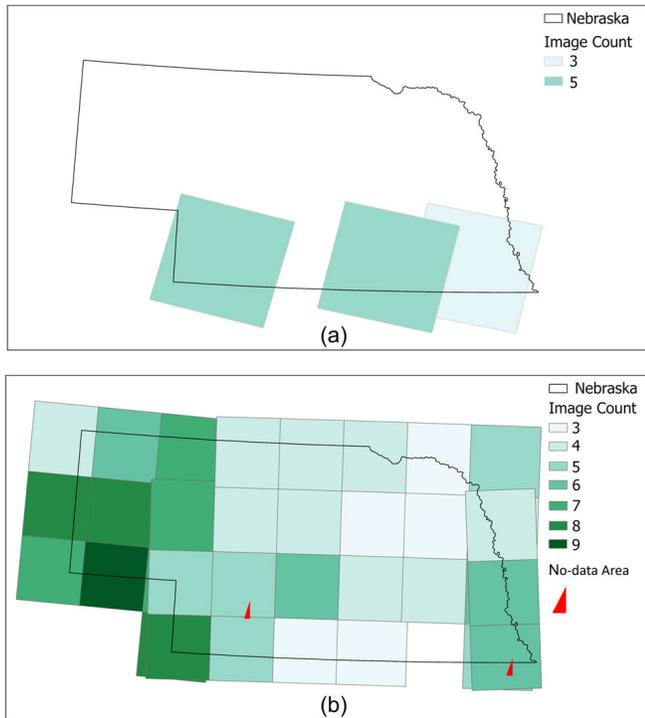


Fig. 3. Example of selected multiple satellite images in May–July in Nebraska. (a) Landsat 8/9, (b) Sentinel 2A (image count—the number of time-series images in the scene).

(no-data). After processing, time-series image stacks for each scene are constructed for both Sentinel-2A and Landsat 8/9. These stacks represent a sequential collection of images over time, capturing the spectral change pattern in two kinds of data. Fig. 3 also illustrates the quantity of time-series images in every scene. The count of time-series images for Sentinel-2A in each scene ranges from 3 to 9, while the additional Landsat 8/9 images supplementing each scene range from 3 to 5.

In certain states, such as Louisiana, where low cloud-cover images are scarce compared to Nebraska, the cloud filter conditions are adjusted accordingly. In that case, the selection criteria for high-quality images involve setting a threshold of less than 12% separately for both Sentinel-2 and Landsat 8/9.

To augment the classification performance, we employed a combination of multiple spectral bands along with two vegetation indexes (VIs) in the classification process. This approach resulted in higher accuracy in crop classification [14], [23]. A similar method was used for mapping pan-European land cover and Switzerland cropland [39], [40]. The well-known vegetable index is the normalized difference vegetation index (NDVI) [41]. Time-series NDVI curve can indicate the vegetable phenology change in the growing period [42]. The normalized difference water index (NDWI) is the indicator to enhance open water body character in multiple spectrum images [43], which was also used in crop identification for MODIS and Sentinel-2A images [14], [44]. In this article, specific bands (blue, green, red, NIR, SWIR_1, and SWIR_2) were extracted from each Landsat 8/9 and Sentinel-2A image. At the same time, the NDVI and NDWI bands also were calculated in these images. As a result, the time-series images

of training and classification spectral data have blue, green, red, NIR, SWIR_1, SWIR_2, NDVI, and NDWI bands. Two VIs formulas are shown as follows:

$$\text{NDVI} = \frac{\text{NIR} - \text{RED}}{\text{NIR} + \text{RED}} \quad (1)$$

$$\text{NDWI} = \frac{\text{GREEN} - \text{NIR}}{\text{GREEN} + \text{NIR}} \quad (2)$$

E. Training Label Extraction

This study enhanced its methodology by incorporating trusted pixel data and time-series satellite images as training samples for the construction of the machine learning model. The pivotal step in this process involves the extraction of trusted pixels, which entails predicting crop types for the upcoming year using historical planting data based on crop rotation patterns. Our prior research has unequivocally validated the efficacy and precision of this method in the middle-west of USA [10], [12], [13], [14], [22], [24]. This study expands the trusted pixel method from Corn Belt to CONUS and concentrates on two kinds of crop rotation patterns in the last five years, as opposed to the previous approach, which considered three in approximately ten years.

The 2022 training pixel data were extracted from 2017–2021 CDL, which serves as an effective predictor for the types of crops anticipated to be planted in 2022. Crop rotation is a farming practice wherein farmers alternate the cultivation of either the same or different crops on a specific field over a defined period. Common rotation patterns in agriculture include both multiple-year and one-year rotations. In the multiple-year pattern, a single type of crop is cultivated continuously on a field for several years. Fields adhering to the multiple-year pattern are expected to plant the same crop in 2022 as they did in the years 2017–2021. Conversely, in the one-year pattern, two different types of crops are alternately grown on the same field during this cycle. Fields following the one-year pattern will cultivate the same crop in 2022 as they did in the years 2020 and 2018.

As an illustration, a field that has been consistently planted with corn from 2017 to 2021 will continue to plant corn in 2022. In contrast, another field with a crop sequence of “corn–soybean–corn–soybean–corn” during this cycle will shift to planting soybeans in 2022. The planting choices for 2022 are determined by the established crop rotation patterns observed in the preceding years. Based on this theory, we successfully extracted a substantial number of trusted pixels for the year 2022 across the CONUS. Each trusted pixel is characterized by its crop category attribution and geospatial location, providing a convenient means to label time-series multisource images for the year 2022.

F. Time-Series Classification

The random forest method [45] was employed for time-series classification in this study. Several studies have observed that this approach demonstrates superior interpretability and stability when dealing with complex input data with a few errors. Moreover, it exhibits higher identification precision compared to both support vector machines (SVM)

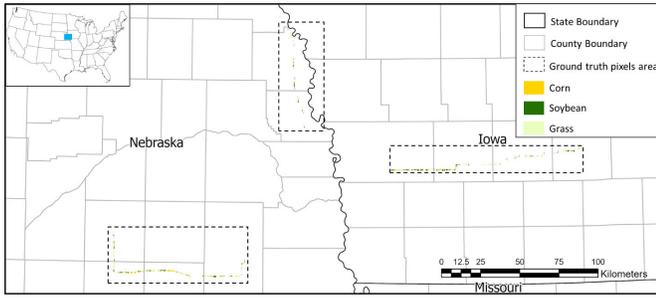


Fig. 4. Distribution of ground truth in Nebraska and Iowa.

and AdaBoost in operational contexts [46]. This model has been used in large-scale optical satellite imagery crop classification [10], [14], [39], [47]. The previous research witnessed random forest classification with a 0.88–0.9 Kappa index when the number of decision trees is 100–1000 and the number of random split variables is 1–8 (number of classification indices) [48]. Nevertheless, the increased number of decision trees necessitates a larger allocation of computational resources and entails more time expenditure, especially in extensive area classification tasks. In this study, we are tasked with processing classifications for hundreds of scene images. For each scene, we have set the number of random split variables to eight, and the number of decision trees is fixed at 500, striking a balance between computational efficiency and classification accuracy. Every processed Landsat 8/9 and Sentinel-2A time-series image stack, along with its corresponding trusted pixel data, is compiled into the training data set. For the training of the random forest model, 5000 random samples are selected to carry out the classification calculation. In this article, the satellite data sources for both training and classification remain consistent across every scene. Specifically, the same Sentinel-2A time-series images are used for both training the model and conducting the classification, and a similar process is followed with Landsat 8/9 data.

G. Validation

This process involves computing pixel-level accuracy and region-level acreage estimates to validate the performance of the ICDL. Pixel-level accuracy evaluates the quality of trusted pixels and classifications by comparing them to independent ground truth data, represented by ground-field raster maps in Nebraska and Iowa. Fig. 4 illustrates the distribution of ground truth data located in the states of Nebraska and Iowa, encompassing areas with corn, soybean, and grass. The assessment process can be segmented into three aspects.

- 1) *Trusted Pixel Accuracy Assessment*: The precision, recall, and F1 score will be calculated between ground truth data and corresponding trusted pixels.
- 2) *Classification Accuracy Assessment*: The precision, recall, and F1 score need to be computed between ground truth pixels and corresponding 2022 ICDL pixels.
- 3) *Crop Acreage Comparison*: The comparison involves assessing the difference between the crop acreage of

ground truth pixels and the corresponding crop acreage predicted by the 2022 ICDL pixels.

Simultaneously, the performance of the CDL 2022 classification is assessed and compared to evaluate the quality of the ICDL. Equations (3)–(5) show precision, recall, and F1 calculation equations, where TP is the number of true positives, FP is the number of false positives, and FN is the number of false negatives. Precision is a value to indicate how many correct pixels are predicted in the trusted pixels or classification pixels. Recall expresses how many correct pixels are predicted in the ground truth pixels. F1 is more like a comprehension value to assess prediction quality

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (3)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4)$$

$$\text{F1} = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}} \quad (5)$$

In the region-level assessment, we compared the regional crop acreage estimates from the 2022 ICDL with the annual crop acreage estimates from the NASS to evaluate the overall regional performance of the ICDL. The region crop acreage estimate is a crucial application of crop maps in agriculture. The NASS Crop Acreage Report (CAR) annually publishes major crop acreage estimates in USA, providing essential support for national agricultural research and decision-making [21]. In addition, the CDL 2022 specific crop-type acreage estimates were also assessed using information from the NASS CAR.

III. EXPERIMENTS AND RESULTS

A. Production of 2022 ICDL

The ICDLs were produced by processing these steps, as shown in Fig. 2. Fig. 5 shows a few detail screenshots of 2022 ICDL products in May, June, and July, illustrating crop field units, grass area, forest area, and river boundary. The 2022 ICDLs are raster maps with 30-m spatial resolution, which includes 86 types of land cover types such as crops, fruits, vegetables, grass, forests, open water, developed areas, and more. The pixel value and color structure are consistent with CDL. The major crop-type land cover distribution areas can be observed easily from ICDL such as corn, soybean, winter wheat, cotton, spring wheat, and rice.

B. Assessment of Trusted Pixel and Classification Quality Using Ground Truth

In this section, we employed ground truth data to assess trusted pixels and classification accuracy at the pixel level. The ground truth data are represented as a ground-field raster map with corn, soybean, and grass. These data were collected in Nebraska and Iowa during July 2022. The 2022 trusted pixel data were extracted from the 2017–2021 CDL. Fig. 6 as an example shows Nebraska trusted pixel data.

In the assessment process, we extracted trusted pixels, ICDL pixels, and CDL 2022 pixels corresponding to ground truth

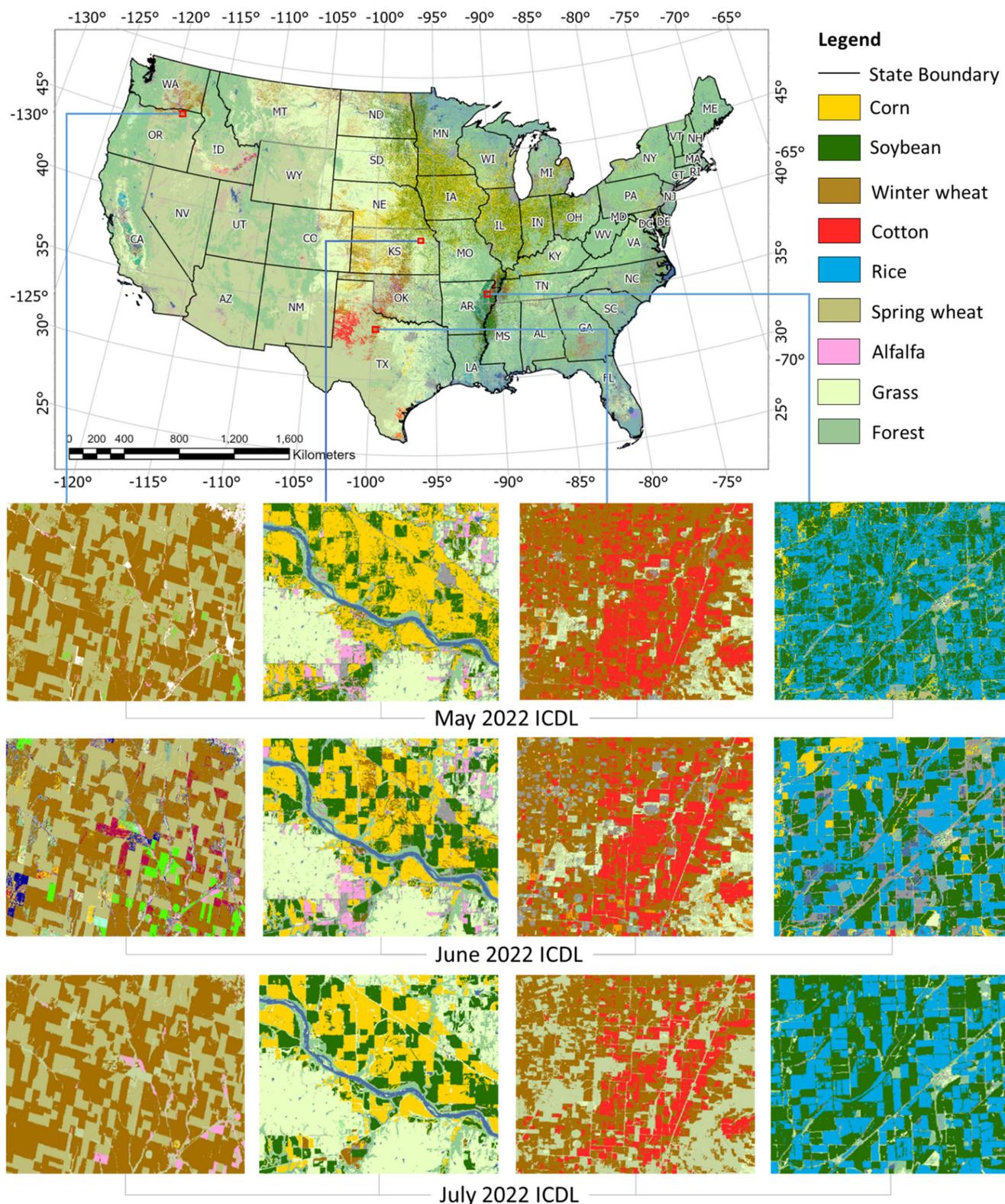


Fig. 5. 2022 CONUS ICDL produced by May, June, and July.

pixels. Subsequently, precision, recall, and F1 scores were calculated for each crop category separately. These measurement values of corn and soybean’s trusted pixels are presented in Table II. In Nebraska, the precision of corn trusted pixels is

higher than that of soybean, although the recall is lower. The F1 score for corn trusted pixels is 0.930, while for soybean, it is 0.888. Similarly, in Iowa, corn trusted pixel precision remains higher than soybean, with both having recalls greater

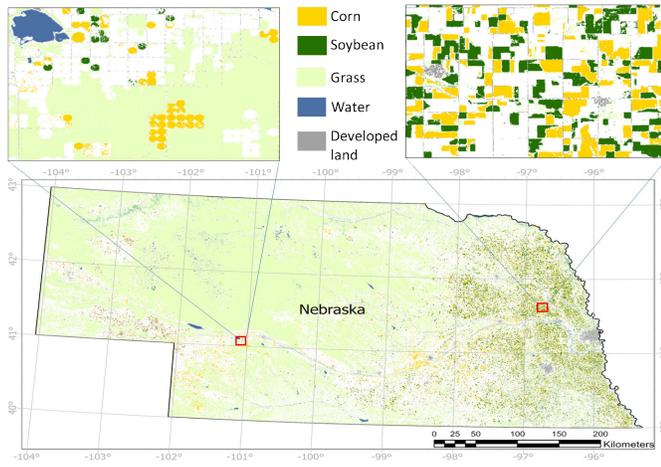


Fig. 6. Nebraska trusted pixel data.

TABLE II
TRUSTED PIXEL ACCURACY ASSESSMENT

State	Crop	Precision	Recall	F1
Nebraska	corn	0.953	0.909	0.930
	soybean	0.849	0.930	0.888
Iowa	corn	0.997	0.905	0.949
	Soybean	0.918	0.993	0.954

than 0.9. Consequently, the F1 score for corn trusted pixels is slightly smaller than soybean, with values of 0.949 and 0.954.

In addition, the precision, recall, and F1 scores for the time-series classification results of corn and soybean from May to July, considering both ICDL and CDL 2022 pixels corresponding to ground truth pixels, were computed and are presented in Table III. The results indicate a time-series increasing accuracy tendency in all items of ICDL. In Nebraska, the classification performances of corn and soybean in ICDL exhibit an increasing trend from May to July, reaching peak accuracy in July. The F1 scores for corn and soybean in July are 0.911 and 0.845, respectively. These scores are slightly higher than CDL for corn but slightly lower than CDL for soybean. During the period from May to July, the classification quality of corn in ICDL is superior to that of soybean, aligning with the accuracy distribution observed in trusted pixels in Nebraska. In Iowa, within the growing season, ICDL from May to July demonstrates a similar increasing tendency in corn and soybean classification performance, climbing to the top accuracy in July with corn F1 = 0.959 and soybean F1 = 0.969, both higher than CDL 2022. In the July ICDL F1 scores for Iowa, corn shows a slightly lower score than soybean, which corresponds to their trusted pixel accuracy distribution, as indicated in Table II. Both corn and soybean of Iowa exhibit better classification performance compared to Nebraska. These assessments still indicate that July 2022 ICDL obtains stratified classification accuracy.

The corn and soybean acreage estimates from ground truth data, as well as the corresponding estimates from ICDL

and CDL 2022, were calculated separately. In addition, the percentage differences in acreage between them were computed. Table IV shows the acreage estimates and percentage differences. In ICDL, Iowa corn and soybean acreage estimates are much closer to ground truth crop acreage than Nebraska. In Nebraska, the soybean ICDL acreage has a bigger gap compared to the ground truth acreage than corn. However, in Iowa, the soybean ICDL acreage is closer to the ground truth acreage than corn. The observed crop acreage comparisons align with the accuracy distribution of trusted pixels and the classification results discussed earlier. For CDL, in Nebraska, the biases in corn and soybean acreage for CDL compared to the ground truth (11.26%, 11.77%) are significantly larger than those for ICDL. In Iowa, the percentage difference in corn acreage compared to ground truth data (−1.43%) is higher than ICDL (−1.15%), but for soybean acreage, the gap (0.04%) is smaller than ICDL (−0.39%).

C. Assessment of Region-Level Crop Acreage Using Official Crop Acreage Data

Crop acreage estimates in extensive areas are widely applied in agricultural research, such as total yield prediction. In this section, major crops, such as corn, soybean, winter wheat, cotton, spring wheat, and rice, are presented as acreage estimates in the 2022 July CONUS ICDL. These estimates are then compared with the 2022 official figures from the NASS CAR [21] and the CDL. In this context, the NASS CAR is considered as reference data, while ICDL and CDL serve as test data. We calculated specific crop planted acreage in each main plantation region, typically encompassing specific states. Fig. 7 shows the comparisons.

1) *Corn and Soybean Acreage Assessment:* According to the NASS acreage data, corn and soybean are recognized as the most important crops in USA [21]. Their primary cultivation region is the Corn Belt, a traditional agricultural plantation area consisting of 13 states in the Midwest of the CONUS. Acreages for corn and soybean in the Corn Belt were estimated using ICDL and CDL, and these estimates were compared with NASS CAR data. Fig. 7(a) and (b) illustrates their comparisons in this region. In corn acreage estimate, most states' estimates in 2022 ICDL are quite near to the NASS acreage data. For corn acreage estimates, the 2022 ICDL estimates in most states closely align with NASS acreage data. While there are instances of underestimation and overestimation in certain states, such as North Dakota and Missouri, the overall corn acreage estimation in the Corn Belt region for 2022 ICDL is very close to the 2022 NASS corn acreage data, with only a 0.01% gap. Regarding the soybean acreage estimate in the Corn Belt region, the 2022 ICDL provides estimates that are generally close to the NASS acreage data, with some variations observed in states such as Kansas and Minnesota. The overall soybean acreage estimate exhibits only a −0.68% gap compared to NASS soybean acreage. In the case of CDL, underestimates and overestimates are observed in certain states, leading to an overall corn acreage difference of −1.25% compared to NASS CAR. Similarly, for soybean, the overall acreage difference is −2.59%. Both differences are larger than those observed with ICDL.

TABLE III
CORN AND SOYBEAN CLASSIFICATION ACCURACY ASSESSMENT FOR ICDL AND CDL 2022 IN NEBRASKA AND IOWA

State	Data	Corn			Soybean		
		Precision	Recall	F1	Precision	Recall	F1
Nebraska	May ICDL	0.764	0.654	0.705	0.707	0.601	0.649
	June ICDL	0.800	0.734	0.766	0.683	0.676	0.680
	July ICDL	0.883	0.940	0.911	0.878	0.814	0.845
	CDL	0.903	0.914	0.908	0.840	0.860	0.850
Iowa	May ICDL	0.719	0.683	0.701	0.741	0.772	0.756
	June ICDL	0.937	0.867	0.901	0.930	0.914	0.922
	July ICDL	0.970	0.949	0.959	0.970	0.968	0.969
	CDL	0.950	0.937	0.943	0.962	0.950	0.956

TABLE IV
CORN AND SOYBEAN ACREAGE ESTIMATES IN GROUND TRUTH DATA AND CLASSIFICATION DATA

	Crop	Ground truth acreage (acres)	ICDL acreage (acres)	Difference percentage (ICDL & Ground)	CDL acreage (acres)	Difference percentage (CDL & Ground)
Nebraska	corn	3410.87	3611.9	5.89%	3794.95	11.26%
	soybean	2133.88	2280	6.85%	2384.96	11.77%
Iowa	corn	2155.23	2130.5	-1.15%	2124.32	-1.43%
	soybean	2477.26	2467.7	-0.39%	2478.15	0.04%

2) *Spring Wheat Acreage Assessment*: According to the 2022 NASS crop acreage estimate, a total of 11 110 000 acres of spring wheat were planted in USA. This cultivation predominantly occurred in states such as North Dakota, Montana, Minnesota, Washington, Idaho, and South Dakota. Spring wheat acreage was reported to be less than that of winter wheat but considerably more than rice [21]. We estimated spring wheat planted acreage in 2022 ICDL and CDL, as well as compared them with NASS acreage data in these six states. The calculation and comparison results are shown in Fig. 7(c). North Dakota, Montana, and Minnesota emerged as the leading states in terms of spring wheat cultivation, with a substantial portion of their acreage being identified in the 2022 ICDL. The states of Washington, Idaho, and South Dakota planted relatively few acres of spring wheat, and extremely low quantities of them were estimated in the 2022 ICDL. Possible reasons for this could include the presence of numerous incorrectly trusted pixels and biased distribution of spring wheat in these three states. These factors may contribute to the challenges in accurately estimating spring wheat acreage in the ICDL for these states. The overall spring wheat acreage in this region from ICDL is 11 131 000 acres, which has a 0.19% gap with NASS acreage data. The total acreage of spring wheat from CDL in this region is 12 200 000, drifting around 9.81% from NASS CAR. In general, the overall spring wheat acreage estimates from CDL exhibit a larger difference compared to NASS when compared to the estimates from ICDL.

3) *Cotton Acreage Assessment*: USA exports one-third of the cotton in the world and stands as the third cotton producer in the global [49]. In the 2022 ICDL, the predominant cotton cultivation in USA was observed in Texas, Georgia, Oklahoma, and Arkansas, collectively contributing to over 70% of the total cotton production. In addition, limited cotton cultivation was identified in the Delta region and California. This section estimated cotton acreage in four specific states from ICDL and CDL, comparing them with NASS acreage data. Fig. 7(d) illustrates the comparison outcomes.

Notably, approximately 6 915 000 acres of cotton were planted in Texas, closely aligning with the NASS acreage data of 7 100 000 acres. Similarly, the cotton acreage estimate in Georgia closely matches the NASS data, with both indicating the same acreage of 1 200 000. However, in Arkansas and Oklahoma, their cotton acreages exhibit big differences from the NASS acreage data. In this region, 2022 ICDL overall cotton acreage shows a -4.39% difference from the NASS data. The 2022 CDL cotton acreage appears to significantly overestimate in the four states mentioned, leading to an overall acreage that exhibits a substantial gap with NASS CAR (26.04%).

4) *Winter Wheat Acreage Assessment*: Winter wheat is predominantly cultivated in the central and western regions of the CONUS. In this analysis, the 2022 ICDL was employed to estimate winter wheat acreage in eight specific states, namely, Kansas, Oklahoma, Texas, Colorado, Washington, Montana, Idaho, and Oregon. These estimates were then compared

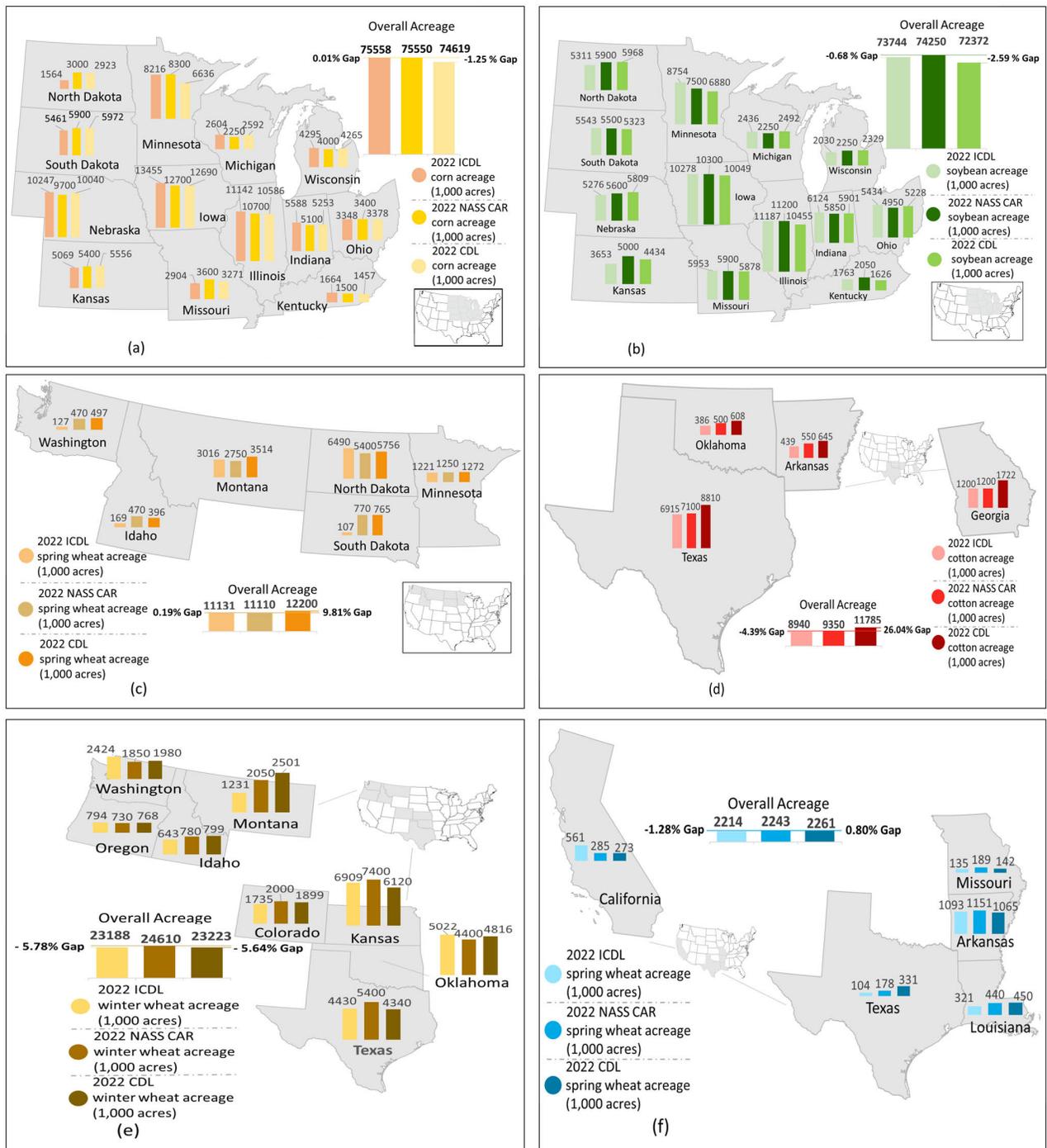


Fig. 7. Comparison of six major crop acreage between 2022 July ICDL, NASS Crop Acreage Report (NASS CAR), and CDL. (a) Corn acreage comparison in Corn Belt. (b) Soybean acreage comparison in Corn Belt. (c) Spring Wheat acreage comparison. (d) Cotton acreage comparison. (e) Winter Wheat acreage comparison. (f) Rice acreage comparison.

with corresponding NASS acreage data. Fig. 7(e) shows the calculation results. Kansas stands out as the leading state in the nation for winter wheat cultivation, with approximately 6 909 000 acres estimated in the 2022 ICDL. This estimate closely aligns with the corresponding NASS winter wheat acreage data. Oklahoma, Texas, Colorado, and Washington also demonstrate substantial winter wheat cultivation, with sizable acreages identifiable in the 2022 ICDL. The total winter wheat acreage in this region is approximately 23 188 000 acres, indicating a -5.84% difference compared

to the corresponding NASS acreage data. Notably, the overall acreage estimation for winter wheat in CDL (23 223 000 acres) exhibits a slightly smaller difference (-5.64%) with NASS CAR compared to ICDL. Both estimates are remarkably close to each other. Winter wheat is typically harvested in early June. The ICDL in 2022 July, employing only three months (May–July) of spectrum information for classification, likely contributes to the significant underestimation of winter wheat in certain states, such as Montana and Texas.

5) *Rice Acreage Assessment*: Rice holds the rank of the sixth most-grown grain in USA, with cultivation primarily concentrated in southern states such as Arkansas, Louisiana, Texas, Missouri, and the state of California. This section processed similar estimates and comparisons with the above. Fig. 7(d) demonstrates the outcomes. According to the 2022 ICDL, Arkansas boasted the largest acreage of rice fields in the entire nation, having cultivated 1 093 000 acres of rice in 2022. Notably, this estimate aligns closely with the 2022 NASS acreage data. In the other four states (Louisiana, Texas, Missouri, and California), there are notable differences between the two datasets. In this region, the overall rice acreage in the 2022 ICDL is quite close to NASS acreage data, and their gap is -1.28% . The total acreage estimation in this region of rice from 2022 CDL exhibits a smaller gap (0.8%) with NASS CAR than ICDL.

IV. DISCUSSION

A. Significance of ICDL

The 2022 ICDL data product, as derived from this study, stands as the first in-season crop-type land cover map for the CONUS. Notably, the ICDL successfully identified various land cover categories encompassing all major crops in USA, such as corn, soybean, winter wheat, cotton, spring wheat, and rice. An outstanding feature of the ICDL is its early production and publication, occurring in early August. This timing is approximately six months ahead of the CDL release schedule. This endeavor represents an exploration into utilizing the capabilities of a public cloud platform for the automated production of an extensive area crop map. Through multiple assessments, the classification quality of ICDL demonstrates an increasing trend throughout the growing season, achieving satisfactory accuracy in the July data product when compared to ground truth data. In addition, the acreage estimates for major crops, such as corn, soybean, spring wheat, and cotton, exhibit higher agreement with NASS CAR in certain regions compared to CDL 2022. This suggests that ICDL performs well in capturing the dynamics of crop cultivation and provides more accurate estimates in specific areas.

B. Application Scenario of ICDL

The ICDL product serves as a nationwide raster cropland map, offering valuable information for agriculture applications and decision-making processes. This comprehensive map provides a detailed and accessible overview of cropland distribution across the entire nation, supporting a range of applications in the agricultural sector. The potential applications might involve the following.

- 1) *Yield Prediction*: By monitoring crop conditions throughout the growing season, ICDL supports the estimation of potential yields. This information is valuable for farmers, agribusinesses, and policymakers for planning and decision-making.
- 2) *Harvest Planning*: ICDL aids in harvest planning by providing basic field data to monitor the spatial and temporal variability of crop maturity. This information helps optimize harvest timing and logistics.

- 3) *Land Use Planning*: ICDL contributes to land use planning by providing information on the distribution and types of crops grown in specific regions. These data are useful for policymakers, land managers, and researchers.
- 4) *Insurance and Risk Assessment*: Crop insurance agencies can use ICDL to assess and mitigate risks associated with yield fluctuations. Timely information on crop distribution enhances the accuracy of insurance assessments.
- 5) *Environmental Monitoring*: Monitoring changes in crop cover contributes to environmental monitoring efforts. This includes assessing the impact of agriculture on ecosystems, tracking land use changes, and studying the relationship between crop patterns and environmental conditions.
- 6) *Precision Agriculture*: Farmers can utilize ICDL to facilitate precision agriculture practices. This involves tailoring inputs (such as water, fertilizers, and pesticides) based on the specific needs of different areas within a field, optimizing resource use and enhancing overall crop productivity.
- 7) *Loss Estimates*: ICDL can be fundamental data to be used in crop loss estimates during the growing season. For some extremely natural disasters such as flooding, rescue departments or insurance companies can use ICDL to estimate the affected crop types and areas, assess the economic loss, and assist disaster relief decision-making.

C. Advantages of the Mapping-Without-Ground-Truth Approach

The ICDL offers the capability to generate early in-season crop maps for major crop types across large-scale counties, utilizing an automated and efficient approach that does not necessarily rely on ground truth data. The central focus of this study lies in mapping without the reliance on ground truth samples. We employed trusted pixels, derived as land cover labels from historical CDL data. This approach not only circumvented the need for extensive time and financial resources but also facilitated a freely automated extraction process. Notably, our study expanded the scope of trusted pixels to encompass not only major crop types but also various other land cover categories such as developed areas, grass, forest, and open water. This inclusive approach significantly broadened the identification capabilities of remote sensing cropland supervised classification. Consequently, the 2022 CONUS ICDL encompasses more than 80 distinct land cover types.

D. Limitations and Potential Solutions

According to the acreage assessment, their accuracies are not equal everywhere in the CONUS, and there are significantly low accuracies in some states such as corn acreage estimates in North Dakota and spring wheat acreage estimates in South Dakota. Meanwhile, ICDL and CDL both have big percentage differences in overall winter wheat acreage compared to NASS CAR. The reasons may include unequal crop rotation patterns and misclassification of CDL. In some

areas of CONUS, on one hand, the crop rotation pattern could be variable dramatically in different places, which will significantly impact the trusted pixel quality, but our trusted pixel is based on the constant rotation patterns. On the other hand, the historical CDL is the basic data of trusted pixels, which still own a certain number of misclassification pixels. To improve crop-type classification accuracy, such as winter wheat, merging the confidence data layer with CDL probable is a method to improve the CDL accuracy [50]. Another approach is CDL purification and refinement to improve the quality of the trusted pixels [51]. Meanwhile, some areas are covered by clouds for a long period: thus, there are insufficient high-quality time-series satellite imageries that miss crop phenology information and reduce the classification performance. Synthetic aperture radar (SAR) data such as Sentinel-1 can be a potential supplementary data in crop classification that is insensitive to the cloud.

E. Computational Complexity

In the production of ICDLs, various tasks, such as collecting multiple satellite imagery, image processing, CDL collection, trusted pixel extraction, training the random forest model, scene-by-scene classification, and state ICDL, primarily rely on the GEE platform to the automated process. During the process, computational instability was only encountered when training with more than 5000 samples. Consequently, in this study, we opted to randomly select 5000 samples from each scene's trusted pixels to address this computational challenge. The computation time for each state varies, ranging from 0.5 to 2 h, depending on the size of the state's geographical area. Moreover, it is worth noting that multiple states can be processed simultaneously. For the CONUS ICDL, the "Set Null" and "Mosaic To New Raster" Python APIs in ArcGIS Pro were employed to achieve efficient outcomes in the processing of the third step of the workflow. This program completed its execution in no more than 2 h, even when running on a standard laptop without additional GPU. In summary, the entire process took approximately 40 h to produce an ICDL map.

V. CONCLUSION

This study explored an automated workflow for the in-season crop-type data layer mapping without ground truth in the CONUS using Landsat 8, Landsat 9, and Sentinel-2A datasets. The trusted pixels from historical CDL acted as land cover labels based on the crop rotation patterns. Thousands of high-quality satellite images in May–July, integrating land cover labels to train random forest models, conducting the time-series classification. The approach successfully produced the CONUS ICDLs product from May to July 2022, including major crop types and other land use types. The pixel- and region-level data assessments were executed by utilization of the ground truth data and NASS acreage estimates data. The ICDL product at the end of July reaches higher accuracy than CDL in most regions. This study offered a set of in-season CDL-like data products, as well as provided an option for automated large-scale in-season crop mapping.

In the future, some relevant work should be procedure to further increase the accuracy and applications of ICDL. First, the refinement of trusted pixels should be a pivotal focus for enhancing mapping quality. On the one hand, future research endeavors could delve into the intricate study of diverse crop rotation patterns across different regions and different crop types. The trusted pixel pre-refinement using early season spectrum information, on the other hand, could be explored to reduce noise pixels in each crop type. Concurrently, there are plans to explore the integration of microwave satellite data into the crop-type identification process, aiming to mitigate the impact of atmospheric conditions. This innovative approach seeks to enhance the robustness and accuracy of crop mapping. In addition, the introduction of transfer learning in crop-type mapping is on the horizon. This involves extracting high-quality trusted pixels to train classifiers in specific locations and utilizing the trained model to process crop mapping in other areas, extremely reducing the uncertainty of large-scale trusted pixels. The versatility of this algorithm is envisioned to transcend national and temporal boundaries. Simultaneously, those applications of ICDL will be explored, aiming to unlock and maximize its potential uses in various contexts.

DATA AVAILABILITY

The ICDL data product is GeoTIFF 8-bit integer file format with 30-m resolution and EPSG:5070 coordinate system. The full dataset and associated web services are accessible on the iCrop service system (<https://cloud.csiss.gmu.edu/icrop/>) [52], served and managed by the Centre for Spatial Information Science and System, George Mason University.

REFERENCES

- [1] H. Huang and M. Khanna, Eds., "An econometric analysis of U.S. crop yield and cropland acreage: implications for the impact of climate change," in *Proc. Agricult. Appl. Econ. Assoc. Conf.*, Jul. 2010, pp. 1–34, doi: [10.22004/ag.econ.61527](https://doi.org/10.22004/ag.econ.61527).
- [2] C. Boryan, Z. Yang, R. Mueller, and M. Craig, "Monitoring US agriculture: The US department of agriculture, national agricultural statistics service, cropland data layer program," *Geocarto Int.*, vol. 26, no. 5, pp. 341–358, Aug. 2011, doi: [10.1080/10106049.2011.562309](https://doi.org/10.1080/10106049.2011.562309).
- [3] W. Han, Z. Yang, L. Di, and R. Mueller, "CropScape: A web service based application for exploring and disseminating US conterminous geospatial cropland data products for decision support," *Comput. Electron. Agricult.*, vol. 84, pp. 111–123, Jun. 2012, doi: [10.1016/j.compag.2012.03.005](https://doi.org/10.1016/j.compag.2012.03.005).
- [4] C. G. Boryan, Z. Yang, A. Sandborn, P. Willis, and B. Haack, "Operational agricultural flood monitoring with Sentinel-1 synthetic aperture radar," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2018, pp. 5831–5834, doi: [10.1109/IGARSS.2018.8519458](https://doi.org/10.1109/IGARSS.2018.8519458).
- [5] C. Boryan, Z. Yang, P. Willis, and A. Sandborn, "Early season winter wheat identification using Sentinel-1 synthetic aperture radar (SAR) and optical data," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2019, pp. 6239–6242, doi: [10.1109/IGARSS.2019.8898511](https://doi.org/10.1109/IGARSS.2019.8898511).
- [6] K. Copenhaver, Y. Hamada, S. Mueller, and J. B. Dunn, "Examining the characteristics of the cropland data layer in the context of estimating land cover change," *ISPRS Int. J. Geo-Inf.*, vol. 10, no. 5, p. 281, Apr. 2021, doi: [10.3390/ijgi10050281](https://doi.org/10.3390/ijgi10050281).
- [7] V. Demarez, F. Helen, C. Marais-Sicre, and F. Baup, "In-season mapping of irrigated crops using Landsat 8 and Sentinel-1 time series," *Remote Sens.*, vol. 11, no. 2, p. 118, Jan. 2019, doi: [10.3390/rs11020118](https://doi.org/10.3390/rs11020118).
- [8] J. Dong et al., "Early-season mapping of winter wheat in China based on Landsat and sentinel images," *Earth Syst. Sci. Data*, vol. 12, no. 4, pp. 3081–3095, Nov. 2020, doi: [10.5194/essd-12-3081-2020](https://doi.org/10.5194/essd-12-3081-2020).

- [9] G. Forkuor, C. Conrad, M. Thiel, T. Ullmann, and E. Zougrana, "Integration of optical and synthetic aperture radar imagery for improving crop mapping in northwestern Benin, West Africa," *Remote Sens.*, vol. 6, no. 7, pp. 6472–6499, Jul. 2014, doi: [10.3390/rs6076472](https://doi.org/10.3390/rs6076472).
- [10] H. Li, L. Di, C. Zhang, L. Lin, and L. Guo, "Improvement of in-season crop mapping for Illinois cropland using multiple machine learning classifiers," in *Proc. 10th Int. Conf. Agro-Geoinformatics*, Jul. 2022, pp. 1–6, doi: [10.1109/Agro-Geoinformatics55649.2022.9859153](https://doi.org/10.1109/Agro-Geoinformatics55649.2022.9859153).
- [11] S. Meng, X. Wang, X. Hu, C. Luo, and Y. Zhong, "Deep learning-based crop mapping in the cloudy season using one-shot hyperspectral satellite imagery," *Comput. Electron. Agricult.*, vol. 186, Jul. 2021, Art. no. 106188, doi: [10.1016/j.compag.2021.106188](https://doi.org/10.1016/j.compag.2021.106188).
- [12] M. S. Rahman, L. Di, E. Yu, C. Zhang, and H. Mohiuddin, "In-season major crop-type identification for US cropland from Landsat images using crop-rotation pattern and progressive data classification," *Agriculture*, vol. 9, no. 1, p. 17, Jan. 2019, doi: [10.3390/agriculture9010017](https://doi.org/10.3390/agriculture9010017).
- [13] C. Zhang et al., "Rapid in-season mapping of corn and soybeans using machine-learned trusted pixels from cropland data layer," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 102, Oct. 2021, Art. no. 102374, doi: [10.1016/j.jag.2021.102374](https://doi.org/10.1016/j.jag.2021.102374).
- [14] C. Zhang et al., "Towards automation of in-season crop type mapping using spatiotemporal crop information and remote sensing data," *Agricult. Syst.*, vol. 201, Aug. 2022, Art. no. 103462, doi: [10.1016/j.agsy.2022.103462](https://doi.org/10.1016/j.agsy.2022.103462).
- [15] B. Zhao et al., "Automatic wheat lodging detection and mapping in aerial imagery to support high-throughput phenotyping and in-season crop management," *Agronomy*, vol. 10, no. 11, p. 1762, Nov. 2020, doi: [10.3390/agronomy10111762](https://doi.org/10.3390/agronomy10111762).
- [16] L. Zhong, L. Hu, and H. Zhou, "Deep learning based multi-temporal crop classification," *Remote Sens. Environ.*, vol. 221, pp. 430–443, Feb. 2019, doi: [10.1016/j.rse.2018.11.032](https://doi.org/10.1016/j.rse.2018.11.032).
- [17] X.-P. Song et al., "National-scale soybean mapping and area estimation in the United States using medium resolution satellite imagery and field survey," *Remote Sens. Environ.*, vol. 190, pp. 383–395, Mar. 2017, doi: [10.1016/j.rse.2017.01.008](https://doi.org/10.1016/j.rse.2017.01.008).
- [18] S. Wang, S. Di Tommaso, J. M. Deines, and D. B. Lobell, "Mapping twenty years of corn and soybean across the US Midwest using the Landsat archive," *Sci. Data*, vol. 7, no. 1, p. 307, Sep. 2020, doi: [10.1038/s41597-020-00646-4](https://doi.org/10.1038/s41597-020-00646-4).
- [19] Z. Li, G. Chen, and T. Zhang, "A CNN-transformer hybrid approach for crop classification using multitemporal multisensor images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 847–858, 2020, doi: [10.1109/JSTARS.2020.2971763](https://doi.org/10.1109/JSTARS.2020.2971763).
- [20] US Census Bureau. *State Area Measurements and Internal Point Coordinates*. Accessed: Aug. 19, 2022. [Online]. Available: <https://www.census.gov/geographies/reference-files/2010/geo/state-area.html>
- [21] NASS. (Jun. 2022). *United States Total Planted Acreage*. [Online]. Available: <http://downloads.usda.library.cornell.edu/usda-esmis/files/j098zb09z/0z70b374s/w9506686w/acrg0622.pdf>
- [22] D. M. Johnson and R. Mueller, "Pre- and within-season crop type classification trained with archival land cover information," *Remote Sens. Environ.*, vol. 264, Oct. 2021, Art. no. 112576, doi: [10.1016/j.rse.2021.112576](https://doi.org/10.1016/j.rse.2021.112576).
- [23] Y. Cai et al., "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.*, vol. 210, pp. 35–47, Jun. 2018, doi: [10.1016/j.rse.2018.02.045](https://doi.org/10.1016/j.rse.2018.02.045).
- [24] C. Zhang, L. Di, L. Lin, and L. Guo, "Extracting trusted pixels from historical cropland data layer using crop rotation patterns: A case study in Nebraska, USA," in *Proc. Int. Conf. Agro-Geoinformatics*, Jul. 2019, pp. 1–6, doi: [10.1109/Agro-Geoinformatics.2019.8820236](https://doi.org/10.1109/Agro-Geoinformatics.2019.8820236).
- [25] R. Hang, Q. Liu, D. Hong, and P. Ghamisi, "Cascaded recurrent neural networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5384–5394, Aug. 2019, doi: [10.1109/TGRS.2019.2899129](https://doi.org/10.1109/TGRS.2019.2899129).
- [26] D. Hong, L. Gao, J. Yao, B. Zhang, A. Plaza, and J. Chanussot, "Graph convolutional networks for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 7, pp. 5966–5978, Jul. 2021, doi: [10.1109/TGRS.2020.3015157](https://doi.org/10.1109/TGRS.2020.3015157).
- [27] M. Rußwurm, N. Courty, R. Emonet, S. Lefèvre, D. Tuia, and R. Tavenard, "End-to-end learned early classification of time series for in-season crop type mapping," *ISPRS J. Photogramm. Remote Sens.*, vol. 196, pp. 445–456, Feb. 2023, doi: [10.1016/j.isprsjprs.2022.12.016](https://doi.org/10.1016/j.isprsjprs.2022.12.016).
- [28] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, "Google Earth Engine: Planetary-scale geospatial analysis for everyone," *Remote Sens. Environ.*, vol. 202, pp. 18–27, Dec. 2017, doi: [10.1016/j.rse.2017.06.031](https://doi.org/10.1016/j.rse.2017.06.031).
- [29] NASS. (Oct. 2010). *Field Crops Usual Planting and Harvesting Dates*. [Online]. Available: <https://usda.library.cornell.edu/concern/publications/vm40xr56k>
- [30] V. H. R. Prudente, V. S. Martins, D. C. Vieira, N. R. D. F. E. Silva, M. Adami, and I. D. Sanches, "Limitations of cloud cover for optical remote sensing of agricultural areas across south America," *Remote Sens. Appl., Soc. Environ.*, vol. 20, Nov. 2020, Art. no. 100414, doi: [10.1016/j.rsase.2020.100414](https://doi.org/10.1016/j.rsase.2020.100414).
- [31] USGS-Landsat 8 Mission. *Landsat 8 | US Geological Survey*. Accessed: Aug. 22, 2022. [Online]. Available: <https://www.usgs.gov/landsat-missions/landsat-8>
- [32] USGS-Landsat 9 Mission. *Landsat 9 | US Geological Survey*. Accessed: Aug. 22, 2022. [Online]. Available: <https://www.usgs.gov/landsat-missions/landsat-9>
- [33] NASA-Landsat Science. *Landsat 8 | Landsat Science, Landsat 8 | Landsat Science*. Accessed: Aug. 23, 2022. [Online]. Available: <https://landsat.gsfc.nasa.gov/satellites/landsat-8/>
- [34] ESA. *User Guides—Sentinel-2 MSI—Level-2A Product—Sentinel Online*. Accessed: Aug. 22, 2022. [Online]. Available: <https://sentinel.esa.int/web/sentinel/user-guides/sentinel-2-msi/product-types/level-2a>
- [35] eAtlas. *Sentinel-2 UTM Tiling Grid (ESA) | eAtlas*. Accessed: Aug. 23, 2022. [Online]. Available: <https://eatlas.org.au/data/uuid/f7468d15-12be-4e3f-a246-b2882a324f59>
- [36] Justin Meyers. (Feb. 2, 2016). *Sentinel-2-Scene-Tile-Grid-Shapefile*. Accessed: Oct. 6, 2022. [Online]. Available: <https://github.com/justinelliotmeyers/Sentinel-2-Shapefile-Index>
- [37] USGS. (Feb. 27, 2019). *Landsat Shapefiles and KML Files*. Accessed: Aug. 30, 2022. [Online]. Available: <https://www.usgs.gov/landsat-missions/landsat-shapefiles-and-kml-files>
- [38] US Census Bureau. (2021). *United States Census Bureau Cartographic Boundary Files*. Accessed: Aug. 30, 2022. [Online]. Available: <https://www.census.gov/geographies/mapping-files/time-series/geo/cartographic-boundary.html>
- [39] R. Pazir, N. Huber, D. Weber, C. Ginzler, and B. Price, "A national extent map of cropland and grassland for Switzerland based on Sentinel-2 data," *Earth Syst. Sci. Data*, vol. 14, no. 1, pp. 295–305, Jan. 2022, doi: [10.5194/essd-14-295-2022](https://doi.org/10.5194/essd-14-295-2022).
- [40] D. Pflugmacher, A. Rabe, M. Peters, and P. Hostert, "Mapping pan-European land cover using Landsat spectral-temporal metrics and the European Lucas survey," *Remote Sens. Environ.*, vol. 221, pp. 583–595, Feb. 2019, doi: [10.1016/j.rse.2018.12.001](https://doi.org/10.1016/j.rse.2018.12.001).
- [41] C. J. Tucker, "Red and photographic infrared linear combinations for monitoring vegetation," *Remote Sens. Environ.*, vol. 8, no. 2, pp. 127–150, May 1979, doi: [10.1016/0034-4257\(79\)90013-0](https://doi.org/10.1016/0034-4257(79)90013-0).
- [42] B. A. Bradley, R. W. Jacob, J. F. Hermance, and J. F. Mustard, "A curve fitting procedure to derive inter-annual phenologies from time series of noisy satellite NDVI data," *Remote Sens. Environ.*, vol. 106, no. 2, pp. 137–145, Jan. 2007, doi: [10.1016/j.rse.2006.08.002](https://doi.org/10.1016/j.rse.2006.08.002).
- [43] S. K. McFeeters, "The use of the normalized difference water index (NDWI) in the delineation of open water features," *Int. J. Remote Sens.*, vol. 17, no. 7, pp. 1425–1432, May 1996, doi: [10.1080/01431169608948714](https://doi.org/10.1080/01431169608948714).
- [44] P. Hao, Y. Zhan, L. Wang, Z. Niu, and M. Shakir, "Feature selection of time series MODIS data for early crop classification using random forest: A case study in Kansas, USA," *Remote Sens.*, vol. 7, no. 5, pp. 5347–5369, Apr. 2015, doi: [10.3390/rs70505347](https://doi.org/10.3390/rs70505347).
- [45] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, pp. 5–32, Oct. 2001, doi: [10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324).
- [46] M. Pal, "Random forest classifier for remote sensing classification," *Int. J. Remote Sens.*, vol. 26, no. 1, pp. 217–222, Jan. 2005, doi: [10.1080/01431160412331269698](https://doi.org/10.1080/01431160412331269698).
- [47] L. Blickensdörfer, M. Schwieder, D. Pflugmacher, C. Nendel, S. Erasmi, and P. Hostert, "Mapping of crop types and crop sequences with combined time series of Sentinel-1, Sentinel-2 and Landsat 8 data for Germany," *Remote Sens. Environ.*, vol. 269, Feb. 2022, Art. no. 112831, doi: [10.1016/j.rse.2021.112831](https://doi.org/10.1016/j.rse.2021.112831).
- [48] V. F. Rodriguez-Galiano, B. Ghimire, J. Rogan, M. Chica-Olmo, and J. P. Rigol-Sanchez, "An assessment of the effectiveness of a random forest classifier for land-cover classification," *ISPRS J. Photogramm. Remote Sens.*, vol. 67, pp. 93–104, Jan. 2012, doi: [10.1016/j.isprsjprs.2011.11.002](https://doi.org/10.1016/j.isprsjprs.2011.11.002).

- [49] USDA.ERS. *USDA ERS—Cotton & Wool*. Accessed: Oct. 31, 2022. [Online]. Available: <https://www.ers.usda.gov/topics/crops/cotton-wool/>
- [50] T. J. Lark, I. H. Schelly, and H. K. Gibbs, "Accuracy, bias, and improvements in mapping crops and cropland across the United States using the USDA cropland data layer," *Remote Sens.*, vol. 13, no. 5, p. 968, Mar. 2021, doi: [10.3390/rs13050968](https://doi.org/10.3390/rs13050968).
- [51] L. Lin et al., "Validation and refinement of cropland data layer using a spatial-temporal decision tree algorithm," *Sci. Data*, vol. 9, no. 1, p. 63, Mar. 2022, doi: [10.1038/s41597-022-01169-w](https://doi.org/10.1038/s41597-022-01169-w).
- [52] C. Zhang et al., "Cyberinformatics tool for in-season crop-specific land cover monitoring: Design, implementation, and applications of iCrop," *Comput. Electron. Agricult.*, vol. 213, Oct. 2023, Art. no. 108199, doi: [10.1016/j.compag.2023.108199](https://doi.org/10.1016/j.compag.2023.108199).



Hui Li received the M.S. degree in cartography and geography information system from the Xi'an University of Science and Technology, Xi'an, China, in 2012. He is currently pursuing the Ph.D. degree in Earth systems and geoinformation sciences with George Mason University, Fairfax, VA, USA.

He is a Research Assistant with the Center for Spatial Information Science and Systems, George Mason University. His research interests include machine learning technology, agro-geoinformatics, remote sensing, and geographic information systems.

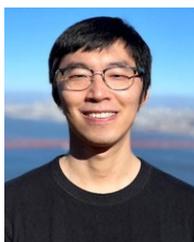


Liping Di (Senior Member, IEEE) received the Ph.D. degree in remote sensing/GIS (geography) from the University of Nebraska—Lincoln, Lincoln, NE, USA, in 1991.

He is a Professor and the Founding Director of the Center for Spatial Information Science and Systems (CSISS) and a Professor at the Department of Geography and Geoinformation Science, George Mason University, Fairfax, VA, USA. He has engaged in geoinformatics and remote sensing research for more than 30 years, published over 550 publications, and

received research grants of more than \$70 million.

Dr. Di has actively participated in the activities of a number of professional societies and international organizations, such as IEEE GRSS, ISPRS, CEOS, ISO TC 211, OGC, INCITS, and GEO. He chaired INCITS/L1, a U.S. national committee responsible for setting U.S. national standards on geographic information and representing USA at ISO Technical Committee 211 (ISO TC 211) from 2010 to 2016, and served as the Convenor of ISO TC 211 Working Group 9 from 2019 to 2022. He was elected as the Inaugural President of the International Society of Agromatics (ISAM) in 2022. He served as the Co-Chair for the Data Archiving and Distribution Technical Committee (DAD TC) of IEEE GRSS from 2002 to 2005 and the Chair for DAD TC from 2005 to 2009. He was one of the founders of the Earth Science Informatics Technical Committee (ESI TC) of IEEE GRSS. In 2012, he founded the annual International Conference on Agro-geoinformatics, an IEEE GRSS-sponsored specialty conference, and has served as the General Conference Chair since 2012. He is also the General Chair of IGARSS 2026. He has received many awards for his work, including the R&D100 Award from R&D Magazine in 2008, the Secretary's Honor Award from USDA in 2011, the Merit Award from the InterNational Committee for Information Technology Standards (INCITS) in 2016, and the Lifetime Achievement Award from INCITS in 2022.



Chen Zhang (Member, IEEE) received the Ph.D. degree in Earth systems and geoinformation sciences from George Mason University, Fairfax, VA, USA, in 2021.

He is a Research Assistant Professor with the Center for Spatial Information Science and Systems, George Mason University, Fairfax, VA. His research interests include geographic information science and systems, remote sensing, agro-geoinformatics, geospatial information interoperability and standards, and land use and land cover change. He has

published over 80 peer-reviewed journal articles, book chapters, conference papers, and engineering reports. He leads the development of many geospatial data service systems to serve the geosciences and remote sensing community.



Li Lin (Member, IEEE) received the Ph.D. degree in Earth systems and geoinformation sciences from George Mason University, Fairfax, VA, USA, in 2022.

He is a Research Assistant Professor with the Center for Spatial Information Science and Systems, George Mason University. His research interests include agro-geoinformatics, remote sensing, urban development, geospatial metadata, and catalog federation. He has authored or coauthored more than 50 peer-reviewed journal articles, book chapters,

conference papers, and engineering reports. He leads the development of many geospatial data service systems to serve the geosciences and remote sensing community.



Liying Guo (Member, IEEE) is a Research Professor at the Center for Spatial Information Science and Systems, George Mason University, Fairfax, VA, USA. She has engaged in geospatial information research for more than 25 years. Her research interests include geospatial science applications, food security, ecosystem sustainability, public health, climate change, and land use and land cover change. She has actively participated in the activities of a number of IEEE GRSS, ISO TC211, INCITS, and OGC, and served as the PI and the Co-PI on

multiple federal and national research grants. She is the Secretary of the International Society of Agromatics (ISAM), the Co-Chair of the Organizing Committee of the annual International Conference on Agro-Geoinformatics, and the Manager of the NSF-funded multi-institutional and multidisciplinary CropSmart project.



Eugene G. Yu (Member, IEEE) received the B.Sc. degree in physical geography from Peking University, Beijing, China, the M.Sc. degree in environmental remote sensing from the University of Aberdeen, Aberdeen, U.K., the M.S. degree in information systems and the M.S. degree in computer science from George Mason University, Fairfax, VA, USA, and the Ph.D. degree in geography with focus on remote sensing and geographic information systems from Indiana State University, Terre Haute, IN, USA.

His research interests include geographic information systems, remote sensing, intelligent image understanding, sensor web, semantic web, computational vision, and robotics. He is a Research Professor in geographic information systems and remote sensing and an Associate Director of the Center for Spatial Information Science and Systems, George Mason University.

Dr. Yu is a member of the American Geophysical Union, IEEE Geoscience and Remote Sensing Society, and Computer Society.



Zhengwei Yang (Member, IEEE) received the B.S. degree in electrical engineering from the Shanghai University of Science and Technology, Shanghai, China, in 1982, the M.S. degree in systems engineering from Shanghai Jiaotong University, Shanghai, in 1985, and the Ph.D. degree in electrical engineering from Drexel University, Philadelphia, PA, USA, in 1997.

He is a Geographer with the Research and Development Division, National Agricultural Statistics Service, since 2005. He has published over

90 papers. His research interests include AI/machine learning, remote sensing methods, and GIS technology and their application in agriculture. His work spans conceptual, theoretical, and application research as well as system development.