Remote Sensing of Environment 251 (2020) 112048

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

Mapping crops within the growing season across the United States

Venkata Shashank Konduri[a,b,*], Jitendra Kumar[c], William W. Hargrove[d], Forrest M. Hoffman[b,e], Auroop R. Ganguly[a]

[a] Sustainability and Data Sciences Laboratory, Department of Civil and Environmental Engineering, Northeastern University, Boston, MA, USA
[b] Computational Sciences and Engineering Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA
[c] Environmental Sciences Division, Oak Ridge National Laboratory, Oak Ridge, TN, USA
[d] Eastern Forest Environmental Threat Assessment Center (EFETAC), USDA Forest Service, Asheville, NC, USA
[e] Department of Civil and Environmental Engineering, University of Tennessee, Knoxville, TN, USA

ARTICLE INFO

Keywords:
Near real-time crop mapping
Phenoregions
Multivariate spatio-temporal clustering
Cropland data layer
Mapcurves
MODIS
NDVI

ABSTRACT

Timely and accurate knowledge about the geospatial distribution of crops at regional to continental scales is crucial for forecasting crop production and estimating crop water use. The United States (US) is one of the leading food-producing countries, but lacks a nationwide high resolution crop-specific land cover map available publicly during the current growing season. The goal of this study was to map crops across the Continental US (CONUS) before the harvest, and to estimate the earliest date of classification by which crops can be mapped with sufficient accuracy (90% of full-season accuracy). The study employed a scalable *cluster-then-label* model that was trained on multiple years of MODIS NDVI using ground truth data in the form of US Department of Agriculture (USDA) Cropland Data Layer (CDL) products. The first step in the crop classification was to perform Multivariate Spatio-Temporal Clustering (MSTC) of annual MODIS-derived NDVI trajectories to create phenologically similar regions, or *phenoregions*. The second step was to assign crop labels to *phenoregions* based on spatial concordance between *phenoregions* and crop classes from CDL using *Mapcurves*. Assigning crop labels to *phenoregions* was performed within *ecoregions* to reduce classification errors due to spatial variability in phenology caused by variations in climate, agricultural practices, and growing conditions. The crop classifier was trained and validated on the years 2008–2014, then tested independently on 2015–2018. *Ecoregion*-level crop classification performed better than state-level and CONUS-level classification. Pixel-wise accuracy of classification for eight major crops by area was around 70% across the major corn-, soybeans- and winter wheat-producing areas, whereas regions characterized by high crop diversity had slightly lower accuracy. Classification accuracy for dominant crops like corn, soybeans, winter wheat, fallow/idle cropland and other hay/non alfalfa improved with time as they grew, reaching 90% of year-end accuracy by the end of August over each of the four unseen years in the test period. For corn and soybeans, the earliest dates of classification were found to be much earlier in the central regions of the Corn Belt (parts of Iowa, Illinois and Indiana) than in peripheral areas. The ability to map growing crops may permit near real-time monitoring of the health status and vigor of agricultural crops nationally.

1. Introduction

Accurate and timely monitoring of crops over national scales is critical for crop production forecasts, water management, assessment and management of disaster and disturbance impacts and characterizing land use for Earth system modeling (Justice and Becker-Reshef, 2007; Waldner et al., 2015b). Federal agencies and private businesses involved with crop insurance, food and feed processing and financial markets need alerts of impending crop failures and yield shortfalls to avoid human and livestock famine. Extreme events like the 2010 heat wave in Russia and the 2012 drought in the United States (US) result in crop price volatility for food-insecure regions of the world, necessitating an early warning system for agricultural production shortfalls (Welton, 2011; Boyer et al., 2013). The Global Agricultural Monitoring (GLAM) Project (Becker-Reshef et al., 2010) of the US Department of Agriculture (USDA) Foreign Agriculture Service (FAS), the Food and Agricultural Organization (FAO) Global Information and Early Warning System (GIEWS) (FAO, 2019), and the Famine Early Warning System Network (FEWS NET) (USAID, 1985) are continental or global agricultural monitoring systems that provide information on crop

* Corresponding author at: F107-S, Climate Change Science Institute, Oak Ridge National Laboratory, Oak Ridge, TN, USA.
E-mail address: konduri.v@northeastern.edu (V.S. Konduri).

https://doi.org/10.1016/j.rse.2020.112048
Received 21 October 2019; Received in revised form 8 August 2020; Accepted 12 August 2020

conditions and production forecasting for different countries in the world. These systems use a combination of social and remote sensing information, but are generally limited to estimating net production rather than spatially mapping crops.

Mapping the spatial extent and distribution of crops in a timely manner is necessary for near real-time crop health monitoring (Waldner et al., 2015b). The US is a leading food producer in the world, generating about 20% of world grain exports (USDA, 2019c); however, no spatially explicit national crop map is available publicly during the current growing season (Cai et al., 2018). The USDA National Agricultural Statistics Service (NASS) produces the annual Cropland Data Layer (CDL) (Boryan et al., 2011), a crop-specific land cover map for the CONUS at 30 m resolution, but the CDL is not released until the spring of the year following the current growing season, at least four months after the current harvest. Although the USDA issues weekly Crop Progress and Condition Reports (CPCR), tallying growth stages for major crops (USDA, 2019a), these are aspatial, tabular statistics that are often spatially aggregated to administrative units like counties or states.

Past studies have shown the possibility of mapping individual major crops like corn and soybeans with sufficient accuracy as early as July–August (Zhong et al., 2016; Cai et al., 2018) and winter wheat by the end of April (Skakun et al., 2017). Dahal et al. (2018) showed it was possible to map major crops across CONUS by the end of September of the current growing season. However, the scope of these studies was limited either to only a few crops, or to particular counties, states, or groups of states. Near real-time national scale crop mapping is challenging because 1) crop phenology changes quickly over relatively short time scales, thus requiring remote sensing data with a high temporal frequency, 2) crop-specific land cover maps, required for model development, need to be available over large spatial scales, 3) crop phenology varies across space due to differences in environmental growing conditions, 4) interannual variations in crop phenology caused by variations in climate make a classifier trained on a single year perform poorly in another year, and 5) a spatial crop classifier needs to be efficient to be nationally scalable.

Unsupervised methods like *k*-means clustering, the ISODATA algorithm and Gaussian mixture models have been used in the past to cluster features derived from a time series of remotely sensed vegetation indices (Gumma et al., 2016; Skakun et al., 2017; Xiong et al., 2017; Wang et al., 2019). Crop type labels were then assigned to these clusters using spectral matching techniques or using spatially aggregated crop statistics at the administrative level. Supervised methods like decision tree algorithms (Pittman et al., 2010), support vector machines (Waldner et al., 2015a), random forests (Shao and Lunetta, 2012), neural networks (Shao et al., 2010) and, more recently, deep learning approaches (Kussul et al., 2017; Zhong et al., 2019) have also been successfully applied for crop classification at small scales. The choice of classification algorithm requires considering the type and volume of data, target accuracy, ease of use, speed and scalability, usually posing trade-offs and compromises (Gómez et al., 2016). Recent studies have opted for a generalized classifier trained on multiple years, instead of training on just one year (Zhong et al., 2014). Training on multiple years makes the model more robust to phenology shifts due to interannual variations in climate. A model trained on a sufficient number of years would not require re-training for the mapping year, allowing faster near real-time crop mapping. Massey et al. (2017) used a generalized classifier to map major crop types across the CONUS, and found its performance to be almost at par with training and mapping within the same year.

One of the challenges in large area crop mapping is the variation in the timing of crop phenological development across climate zones, since it is influenced by climate, soil, topography, etc., as well as farm technology, management practices, fertilization, irrigation, etc. Growing degree days (GDD) can account for some variations in crop development (Zhong et al., 2014; Skakun et al., 2017). Other studies performed crop classifications at the scale of smaller administrative units like Agriculture Statistics Districts (ASDs) (Sakamoto et al., 2011), states (Zhong et al., 2016), or Agro-ecological zones (AEZs) (Massey et al., 2017), as defined by the United Nations FAO for the year 2000 (Fischer et al., 2000). However, these approaches either do not take into account variations in precipitation and soil properties, or are run within administrative or political boundaries that are not relevant to crop phenology, or are too large to capture phenological variability with climate. Ideally, modeled regions would be described based on environmental variables that reflect crop growing conditions, and would be of small size, created using quantitative analytical methods that are both empirical and reproducible. Multivariate Geographic Clustering algorithms have been successfully used (Hargrove and Hoffman, 2004) to create *ecoregions*: regions on a map within which exist similar combinations of ecologically relevant conditions like temperature, precipitation, soil and topographic properties.

The objectives for this study were as follows:

- To create a national, crop-specific land cover map (with all of the crop types, as included in the CDL) for the CONUS using time series of MODIS-derived Normalized Difference Vegetation Index (NDVI) as inputs to a generalized *cluster-then-label* crop classifier. The model was trained at the scale of individual quantitative *ecoregions* in order to address the spatial variability in phenology. No national-scale cropland maps are available prior to 2008. One of the goals of this study is to generate national-scale crop maps for the years 2000–2007 using MODIS NDVI, before the CDL began. Having national crop maps back to 2000 could help researchers studying land use/land cover change or modeling long-term crop yield.
- To create crop maps in near real-time during the current growing season and to study the rate of increase in mapping accuracy as the season progresses for 8 major crop types grown in the US: corn, soybeans, winter wheat, fallow/idle cropland, other hay/non alfalfa, alfalfa, sorghum and rice. While accuracy may start low early in the growing season, it should improve as the crops grow and mature. The ultimate goal was to estimate the earliest time by which each of the eight major crop types can be mapped with reasonable accuracy across the entire CONUS within the current growing season.

2. Study area and datasets

2.1. Study area and training data

The Cropland Data Layer (CDL), a crop-specific land cover raster map available for the CONUS at 30 m resolution since 2008, was used as the ground truth for classification. USDA NASS creates the CDL using a decision tree-based classifier that uses remote sensing data from Resourcesat-1 Advanced Wide Field Sensor (AWiFS), Deimos-1, UK Disaster Management Constellation-2, Landsat-5/7/8 and MODIS as inputs. Crop type and acreage information collected in surveys from farmers during the current growing season are used to train the CDL classifier. Non-agricultural areas in the CDL are taken from the National Land Cover Database (NLCD) land cover, imperviousness and canopy categories. The CDL has self-reported crop mapping accuracies in the range of 85–95% for major crop categories (Boryan et al., 2011), but the surveys used for training are not made publicly available.

We downloaded the CDL for 2008–2018 from the USDA NASS Data Portal (USDA, 2019b). Our crop classification model was trained over 2008–2014 and applied to the period 2000–2018. The period of 2015–2018 was used as test years for the classifier. Our analysis focused on about 100 major agricultural land cover types out of 122 categories included in the CDL. The study area for each year was obtained by masking out non-agricultural land cover categories like forests, pasture lands, shrub lands, open water, developed spaces, etc. based on the CDL for that year. Provided in Albers Conic Equal Area projection, the CDL was re-projected to Lambert Azimuthal Equal Area for the analysis using a nearest-neighbor resampling technique.

2.2. Remote sensing data

Time series of smoothed and gap-filled NDVI generated from Collection 5 data streams from Terra (MOD13Q1) and Aqua (MYD13Q1) satellite instruments for the CONUS were downloaded from the Oak Ridge National Laboratory (ORNL) Distributed Active Archive Center for Biogeochemical Dynamics (DAAC) for the period 2000-01-01 through 2018-12-31 (Spruce et al., 2016). The MODIS NDVI data set, at a spatial resolution of 231 m and an 8-day temporal frequency, was generated using the NASA Stennis Time Series Product Tool (TSPT) (Spruce et al., 2011) to remove clouds and otherwise clean and filter the time series temporally. The smoothed, gap-filled data set is nearly complete, with few missing values, and is ideal for many phenological analyses and applications. Files are available in netCDF format, one per year, for the period 2000–2018, as a time series of 8-day maximum-value composited MODIS NDVI in Lambert Azimuthal Equal Area projection.

3. Methods

3.1. Development of Phenoregions

Phenoregions are regions having similar annual profiles of NDVI "greenness" phenology through space and time (White et al., 2005). *Phenoregions* capture the gradients of climate and features of known vegetation at large scales. Hargrove and Hoffman (2004) developed Multivariate Spatio-Temporal Clustering (MSTC) based on a non-hierarchical *k*-means algorithm (Hartigan, 1975) for classification of *phenoregions* (White et al., 2005), classification of remote sensing data (Hoffman et al., 2010), analysis of dynamic climate regimes in Global Circulation Models (GCMs) (Hoffman et al., 2008), and detection of disturbance from phenological time series (Mills et al., 2011). A decentralized scalable parallel implementation of the method (Kumar et al., 2011) was employed for the creation of *phenoregions* in this study (Fig. 1(a)). The entire MODIS NDVI time series (2000–2018) was used with MSTC to delineate 5000 *phenoregions* having similar annual phenological profiles. For this study, the exact number of *phenoregions* is not critical as long as they are fine enough to separate phenological diversity sufficiently to distinguish different crop types. MSTC does not explicitly use geographic location during classification and does not impose spatial contiguity. Thus, a *phenoregion* may be comprised of many spatially disjoint agricultural fields, so long as they have similar phenological profiles. After being classified among 5000 *phenoregions*, the data are mapped back to geographical space to create a spatial map. Our method created 19 annual *phenoregions* maps, one per year during 2000–2018. While MODIS NDVI-based *phenoregions* were generated at 231 m resolution, they were regridded to 30 m using the nearest-neighbor resampling technique to match the CDL resolution. Upsampling MODIS resolution *phenoregions* to CDL resolution does not add any information content, but allows all analysis to be performed at native CDL resolution. The reverse option of downsampling CDL to MODIS resolution would have led to loss of information content in CDL. For each year during 2008–2018, the *phenoregions* corresponding to non-crop areas were masked out using the cropland extent from the CDL for that particular year, as shown in Fig. 1(b).

3.2. Spatio-temporal variability in crop phenology

We partitioned variability in crop phenology across time from phenological variability over space. Weather conditions experienced by agricultural regions exhibit immense inter-annual temporal variability that has key implications for planting and harvesting dates, and for the choice of crops planted. Fig. 2(a) shows temporal variability in phenology during 2008–2012 in a highly diverse agricultural region spanning parts of southern Nebraska and northern Kansas. At the continental scale, agricultural regions show spatial variability in phenological timing caused by climate, soils, growing conditions, and crop rotation and management practices. For example, corn growing in northern and southern Kansas during the year 2013 exhibits different phenology, perhaps caused by differences in planting dates, in cultivars and in growing conditions (Fig. 2(b)). At such spatial scales, phenological signatures of a crop type can show large variations, thus causing an overlap with the timing of other crops types, leading to poor classification accuracies. This spatio-temporal variability in crop phenology adds complexity in phenology-based identification of crop types. We address temporal variability by training the classifier on multiple years, and we address spatial variability using *ecoregions* (Section 3.3), thus developing a more robust and accurate general crop classification model.

3.3. Climatic ecoregions

To avoid classification errors due to spatial variability in phenology, we used *ecoregions* to segment the landscape, and we developed a separate crop classification model optimized within each *ecoregion*. *Ecoregions* group together areas with similar climatic, topographic and edaphic conditions. Clustering algorithms have been widely used for classification of *ecoregions* (Hargrove and Hoffman, 2004; Williams et al., 2008; Kumar et al., 2011). We used the same MSTC algorithm (Hargrove and Hoffman, 2004) to divide the CONUS into 500 synoptic *ecoregions*, representing regions with similar crop growing conditions (Fig. 1(c)). The *ecoregions* were developed using 15 environmental variables characterizing bioclimate (Fick and Hijmans, 2017), topography (Saxon et al., 2005) and soil conditions (Global Soil Data Task Group, 2000; Saxon et al., 2005) at 1 km resolution (Table S1).

3.4. Crop classification model

Since they result from a statistical unsupervised classification, *phenoregions* lack any kind of label identifying any particular crop or vegetation type. As part of a *cluster-then-label* classification approach, a supervision step was applied to map each *phenoregion* to a particular crop type, using the CDL. This supervision step was not manual, but was automated, requiring no human interpretation or intervention. Fig. 1(d) summarizes the workflow for this study. Since the CDL is available only since 2008, the model was trained and validated on the years 2008–2014 and tested independently on the years 2015–2018.

3.4.1. Cluster-then-label model training

We labeled each entire *phenoregion* with a single crop type, based on majority spatial overlap, using the CDL as a training data set. To account for spatial variability, crop type assignments were conducted independently within each *ecoregion* across all years in the training period (2008–2014). Crop pixels from the CDL and the spatially concordant pixels from *phenoregions* present within each *ecoregion* were randomly divided into training (70%) and validation (30%) sets for each year (Fig. 1(d i)). *Mapcurves*, a quantitative method that calculates the spatial concordance between two or more categorical maps and provides an assignment of labels between the maps (Hargrove et al., 2006), was used to compare *phenoregions* with the CDL and assign crop type labels to entire *phenoregions*. *Mapcurves* calculates a pairwise Goodness-of-fit statistic (*GOF*) over all categories in the two maps being compared. The *GOF* statistic between a *phenoregion*, *P*, with a crop type, *C*, was defined as follows:

$$GOF_{P,C} = \frac{A_P \cap A_C}{A_P} \times \frac{A_P \cap A_C}{A_C}, \quad (1)$$

where A_P and A_C represent the area under *P* and *C*, respectively, and $A_P \cap A_C$ represents the area that is common to *P* and *C*. The *GOF* statistic increases when areas in the two maps are spatially coincident, but decreases from areas that are not overlapping, so that large crop areas are not selected preferentially. The crop label having the highest *GOF* statistic was assigned to all the cells within that *phenoregion*, as shown in Fig. 1(d ii). Accounting for temporal variability, *Mapcurves* was

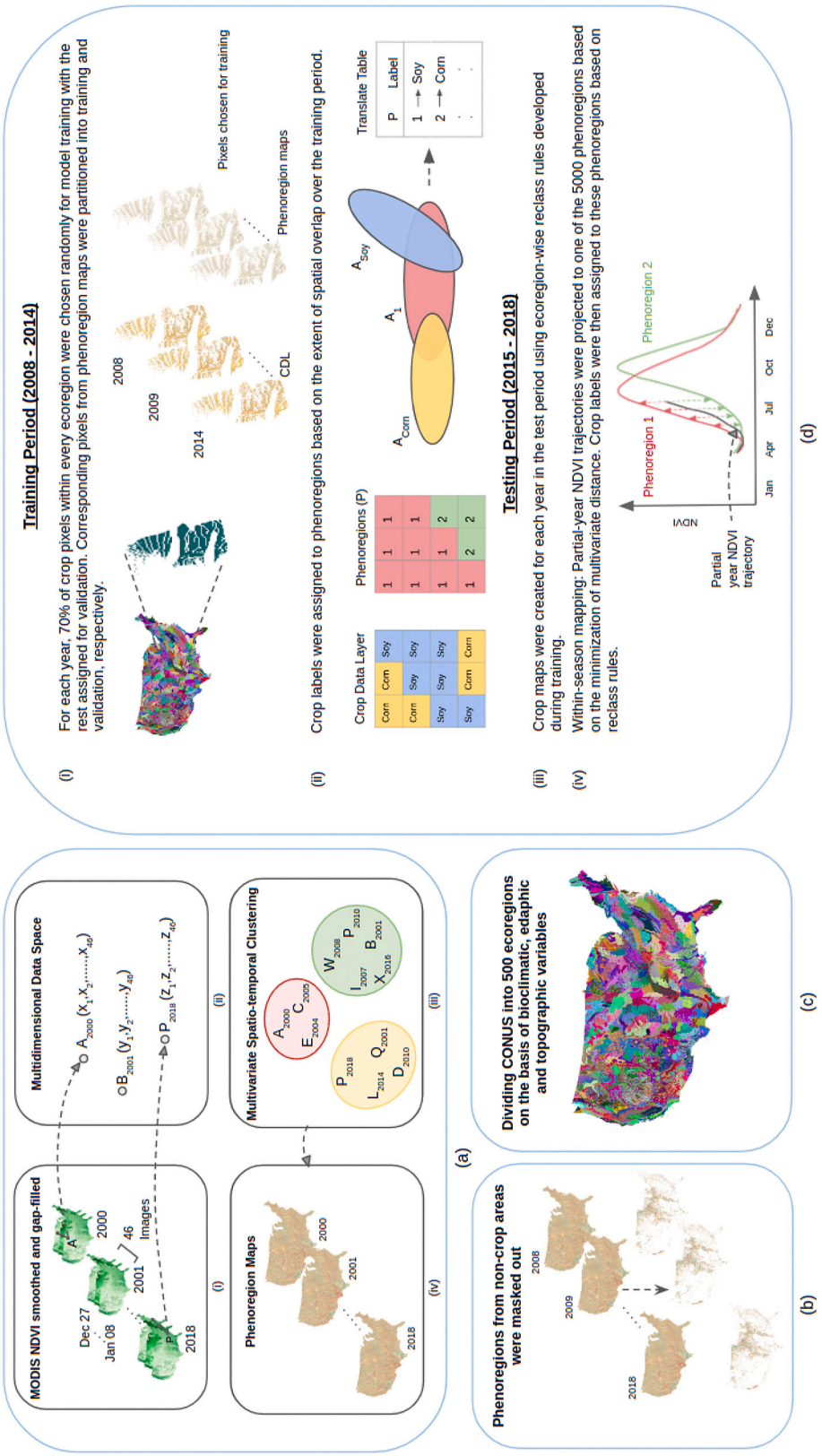

Fig. 1. Steps followed in the *cluster-then-label* -based classification approach: (a) Creation of *phenoregions* using Multivariate Spatio-Temporal Clustering of MODIS NDVI time series. (b) For each year in 2008–2018, *phenoregions* corresponding to non-crop areas were masked out using the cropland extent from CDL for that particular year. (c) Crop type assignment for each *phenoregion* was performed separately within each *ecoregion*, to control for spatial variability in farming methods and growing conditions. (d) Workflow for *Mapcurves*-based training and testing.

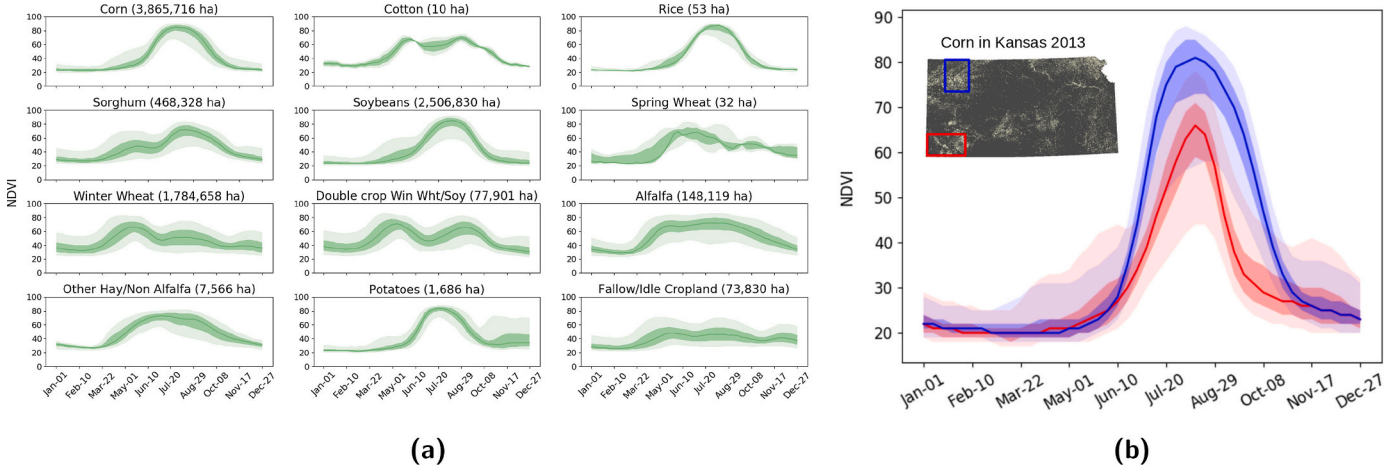

Fig. 2. Crop phenology exhibits a wide range of spatio-temporal variability, thus posing a challenge for national-scale crop mapping. a) Temporal variability in phenology during the period 2008–2012 in a small diverse agricultural region that spans parts of southern Nebraska and northern Kansas. b) Two corn-growing regions within the state of Kansas show large variability in phenology during the same year 2013. (Plot shows the median (solid line), 25th-75th percentile range (dark shade), and 5th-95th percentile range (light shade) of the annual NDVI profile.)

applied to *phenoregion* and CDL maps over all years during the training period to generate a translation table, listing each *phenoregion* and its single corresponding best-fit crop type label. With 5000 *phenoregions*, many *phenoregions* will be assigned to the same crop type label, each representing a variation in climate, edaphic conditions, cultivar, planting date, fertilization, irrigation, and other agricultural factors that may be used or encountered when producing the same crop. In this way, automated supervised labeling of *phenoregions* across all agricultural regions within CONUS was done using *Mapcurves*, while accounting for spatio-temporal variability in crop phenology. The *ecoregion*-specific *Mapcurves* models were applied to validation data sets, the random 30% of data that was set aside each year during 2008–2014, and resulting crop classifications were compared with the CDL to assess accuracy.

3.4.2. Model evaluation within the test period

In addition to evaluating on the validation data collected during the period 2008–2014, we used the *cluster-then-label* approach to create crop maps for the test years 2015–2018. Once the model was trained over the period 2008–2014, unseen years 2015–2018 represented the operational scenario in which trained models were applied and evaluated for their accuracy and applicability at CONUS scale, including within-season classification.

3.4.3. Within-season mapping of crops

We developed a methodology to map croplands at national scale in near real-time as they grow every 8 days (i.e., MODIS composited temporal frequency). The partial phenological MODIS NDVI to date at each cropland pixel within CONUS was assigned to the most-similar *phenoregion* thus far. The assignment was made by identifying the existing *phenoregion* whose profile minimizes the multivariate difference between the existing portions of the two NDVI profiles. Once assigned to a *phenoregion*, the existing trained *cluster-then-label* models (Section 3.4.1) were applied to determine the crop type for each pixel. However, the *cluster-then-label* approach always assigns the best-fitting crop label to a partial NDVI trajectory, even before the crop itself has been planted or has emerged. Early in the season, before crops have substantially emerged, these closest crop type projections are unreliable, yet the best-fitting crop type will still be assigned. To prevent these early misclassifications, partial-year crop classifications were discarded until a minimum spring greenness threshold, defined as 20% of annual amplitude of projected *phenoregion*, was reached (Fig. 3). For each crop in an *ecoregion*, the earliest within-season date by which the crop can be mapped with 90% of the full-season accuracy was identified.

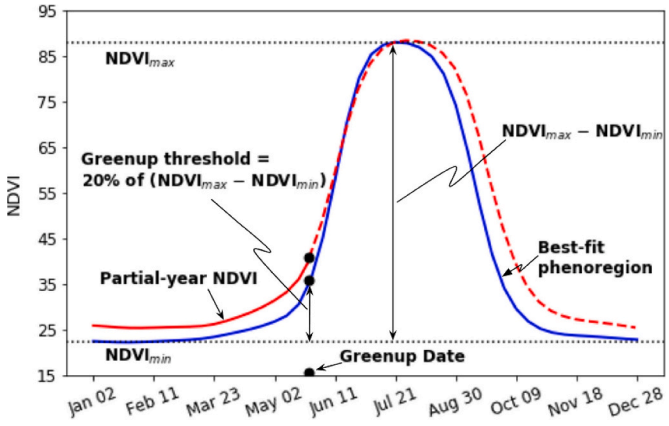

Fig. 3. Assigning crop labels to partial phenological trajectories using the *cluster-then-label* approach was restricted to only those cropland pixels that satisfy a minimum greenness threshold. This greenup threshold was defined as 20% of the annual amplitude (difference between the maximum and minimum) of NDVI for the *phenoregion* (shown in blue), which is the most similar to the partial-year NDVI trajectory (shown in red). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.5. Evaluation metrics

3.5.1. Accuracy assessment

While all crop types contained in the CDL were analyzed and mapped in our study, we focus our accuracy assessment here on the 8 dominant crop types (by area across CONUS): corn, soybeans, winter wheat, fallow/idle cropland, other hay/non alfalfa, alfalfa, sorghum and rice (but see Table S2 for accuracies for all 102 crops, included in the Supplementary Material). Crop types other hay/non alfalfa and fallow/idle cropland are referred to as other hay and fallow, respectively in all tables and figures. Three metrics were used to evaluate the accuracy of classification: *Producer's Accuracy*, *User's Accuracy* and *Overall Accuracy*, as defined in Eqs. (2), (3) and (4). *Producer's Accuracy* is the accuracy of the map from the map producer's point of view, quantifying the probability that a feature class on the ground is correctly classified by the map. *User's Accuracy* is the accuracy from the user's perspective, and quantifies the reliability of the map, i.e., the probability that a feature on the map will actually be present on the

ground. *Overall Accuracy* quantifies the fraction of the reference CDL pixels that are correctly mapped by our crop classification method. *User's Accuracy* is the most relevant for a farmer or resource manager; thus, we focus our discussion on *User's Accuracy*, and include the *Producer's Accuracy* statistics in the Supplementary Material. Errors in classification sometimes happen due to similarity in the phenological signatures of multiple crops, but such confusion can be insightful. We report a confusion matrix with statistics, showing how omission and commission errors are distributed across crop types.

$$\textit{Producer's Accuracy} = \frac{\text{Number of correctly classified pixels of a crop type}}{\text{Total number of pixels of that crop type in the CDL map}} \times 100 \tag{2}$$

$$\textit{User's Accuracy} = \frac{\text{Number of correctly classified pixels of a crop type}}{\text{Total number of pixels of that crop type in the classified map}} \times 100 \tag{3}$$

$$\textit{Overall Accuracy} = \frac{\text{Sum of all correctly classified pixels for all crop types}}{\text{Total number of pixels for all crop types}} \times 100 \tag{4}$$

3.5.2. Shannon diversity of crop types

Omission and commission errors due to confusion between crop types are larger in regions with diverse crop types, and when the cultivated field sizes are smaller than the resolution of MODIS products. We calculate the Shannon Diversity Index (H) to quantify the diversity of crop types within an area:

$$H = -\sum_i p_i \log p_i, \tag{5}$$

where p_i is the proportion of map grid cells belonging to a crop type i in the mapped area. When only one kind of crop exists in an area, H has a value of zero, and crop classification is easy. H increases when there are more crop types present and their probabilities are more uniform within the area. Predictability of crop types decreases with greater crop diversity and greater similarity in the proportional abundances of the crops grown in a region.

4. Results

4.1. Model training to address spatial variability

To address the spatial variability in phenology, independent *cluster-then-label* models were trained within each *ecoregion* (Section 3.3), allowing models that are optimized for each region for higher accuracy. To test the benefits and efficacy of *ecoregion*-specific model training, we also conducted the *cluster-then-label* model training by individual state for each of the 48 states, and as a single model for the entire CONUS. Fig. 4 shows the *User's Accuracy* and *Producer's Accuracy* for the model tested for 2015 when trained at the scales of CONUS, by state, and by *ecoregion*, respectively. Training for smaller regions reduced the spatial variability and thus allowed more specialized models for those regions. *Ecoregions*, derived based on climate, soil and topographic properties, consistently performed the best across all crop types, but models trained at the scale of states provided good accuracy as well. Dominant crops like corn, soybeans and winter wheat are often grown in large fields in concentrated regions of the country, with similar regional crop cultivars and management practices. For such dominant crops, even a single model trained at CONUS-scale produced fairly accurate results. Improvements in models trained at smaller scales of states and *ecoregions* are especially pronounced for crop types that are spatially distributed and/or exhibit a wide range of phenology. For the rest of this paper, we present results based on models trained at the scale of *ecoregions*.

4.2. Mapping crop types across the continental United States

Ecoregion-wise *cluster-then-label* models developed over the training period 2008–2014 were applied to the annual MODIS NDVI-derived *phenoregions* for 2000–2018 to produce crop type maps for each year. The developed crop maps were statistically compared to the CDL to evaluate their accuracy. Evaluation using the 30% validation data set across the period 2008–2014 shows a pixel-wise *Overall Accuracy* of 60–61% for all crop types within CONUS, over 102 crop types (accuracy ranges for individual crops varied widely, and were best for dominant-acreage crops, see Table S2 in the Supplementary Material). *User's Accuracy* for dominant crop types like corn, soybeans and winter wheat varies between 61 and 67%, 58–65% and 60–72% respectively, while for fallow/idle cropland, other hay/non alfalfa, alfalfa, sorghum and rice varies between 50 and 60% (Fig. 5). *Producer's Accuracy* (Fig. S1) for dominant crops like corn and winter wheat varies from 65 to 75% and 69–80%, respectively, and fluctuates between 30 and 40% for less dominant crops like sorghum and rice.

The *cluster-then-label* model was also applied to the test data set with four never-seen-before years 2015–2018. *Overall Accuracy* for the years 2015–2017, over all 102 crop types, is slightly lower (compared to 2008–2014) at ~58% and is 53% for 2018. *User's Accuracy* (Fig. 6) for the eight major crops are fairly consistent over the four test years, with a small reduction compared to the 2008–2014 period, and perform with ~60% accuracy for primary crops like corn, soybeans and winter wheat except for 2018, when the *User's Accuracy* for corn drops to 54%. Pixel-wise *Producer's Accuracy* for the eight major-area crops (Fig. S2) show similar patterns except for sorghum and rice, which have ~40% accuracy for all the years, and soybeans in 2018, the accuracy of which drops to 39%.

Crop type spatial distribution predicted by the *cluster-then-label* model shows broad-level agreement with the CDL (Fig. 7(a)). Three small areas (A, B, and C) from geographically distributed agricultural regions with a wide range of crop diversity were selected for a closer look (Fig. 7(b)). Region A from the Corn Belt, where corn and soybeans are the dominant crops, shows broad agreement between the *cluster-then-label*-based map and CDL. Disagreements between the two maps were prominent along the boundaries of the cultivated fields owing to the coarser resolution of MODIS NDVI products. Region B in winter wheat-producing areas in Kansas demonstrates broad agreement between the two maps. While CDL (at 30 m resolution) is able to resolve the center pivot-irrigated fields very well, our *cluster-then-label*-based map lacks sharpness along fine field boundaries. Region C from Central Valley, California, exhibits immense diversity in crop types grown across small-sized fields and thus represents a difficult-to-classify region; yet the *cluster-then-label* model is able to classify the crop types in this region with reasonable accuracy. Specialty crops like peas, grapes, almonds, walnuts, pistachios, etc. are often challenging to classify accurately (Table S2) as they are grown on small, distributed fields that are smaller than the resolution of MODIS, and may not exhibit distinct, identifiable phenology. Even in the CDL, such specialty crops are based on reported data and are known to have very limited accuracy (Boryan et al., 2011). The *cluster-then-label* model performs well in terms of *Overall Accuracy* (Fig. 8(a)) in major crop growing regions with large field sizes and lower diversity, but has comparatively lower accuracy in regions with high crop diversity and smaller field sizes (Fig. 8(b)).

The *ecoregion*-wise *Overall Accuracy* of pixel-wise classification for 2015 (Fig. 8(a)) is ~70% across much of the Corn Belt, spanning eastern Nebraska, Kansas, Iowa, Illinois, Indiana and western Ohio. Accuracy exceeded 85% in certain regions in major wheat-producing states like Kansas, Oklahoma and Texas. Accuracy in more diverse crop-producing regions like eastern North and South Dakota, western Mississippi and eastern Arkansas and Wisconsin is around 60%. Accuracy is around 50% in the Central Valley, California. Fig. 8(b) shows the Shannon Diversity of crop types (H) across the agriculturally-dominant ecoregions in CONUS. Much of the Corn Belt, growing mostly corn and

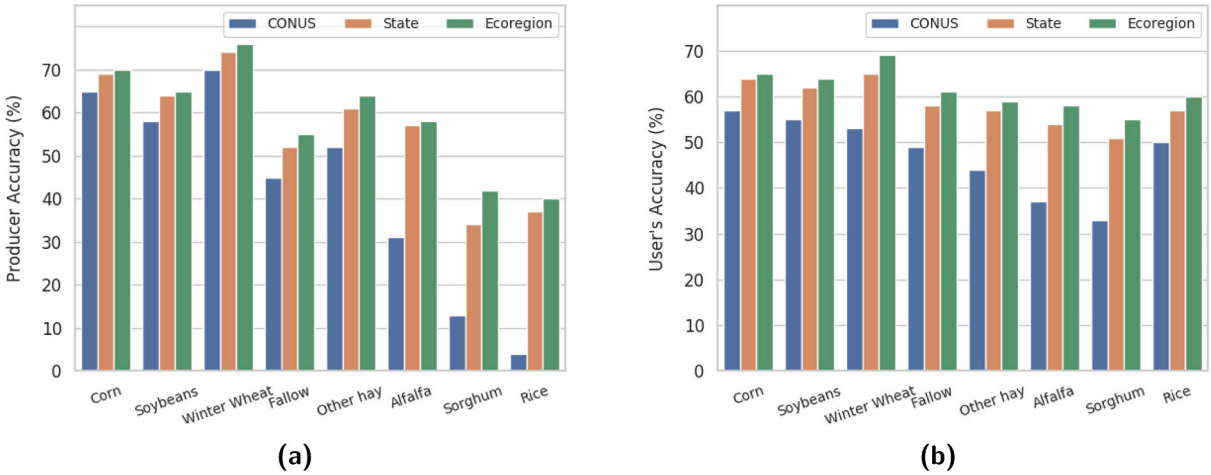

Fig. 4. The *cluster-then-label* model shows an improvement in accuracy when trained and customized for smaller regions, with *ecoregion*-based models showing consistently better performance for all the crop types. (a) Producer's Accuracy when the model is trained at CONUS, state and *ecoregion* scale. (b) User's Accuracy when the model is trained at CONUS, state and *ecoregion* scale.

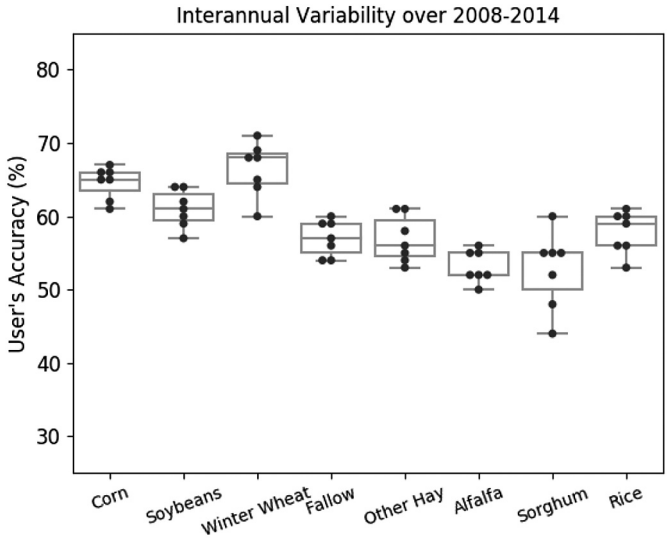

Fig. 5. Evaluation of the *cluster-then-label* model on 30% validation data collected across the period 2008–2014 gives good *User's Accuracy* for eight commonly grown crop types across CONUS.

soybeans, is uniform and has a low value of H. The Central Valley in California, which provides more than half of the fruits, vegetables and nuts grown in the US, has a high H, indicating a high diversity of crops, as do North and South Dakota, which grow multiple crops like corn, soybeans, wheat, hay, sunflower, etc.

Table S2 show the model *User's Accuracy* for all 102 crop types included in the study. Crop types which are grown in limited regions with low acreage, show high variability and modest to low classification accuracy. Such rare crops offer limited samples for model training, and, in addition, training data quality is limited due to the limited accuracy of the CDL. Nevertheless, over time, as the time series of available training data grows, we expect the prognostic power of our models to improve.

4.3. Crop mapping accuracy at the scale of administrative units

Some users of agricultural data need actual, spatially-explicit crop type maps, but many government agencies, private sector organizations, and scientists are interested only in tabular summaries of crop acreage totals over scales of administrative units like counties or states. Classified pixels from the *cluster-then-label* method were aggregated to calculate acreage for each crop type at the county and at the state scale, and totals were compared to corresponding acreages from the CDL (Fig. 9). For the eight area-dominant crop types across the CONUS, good agreement exists between the aggregated *cluster-then-label* model and expected CDL acreage at both county (Fig. 9a) and state (Fig. 9b) scales. There is a slight over-prediction in corn acreages and under-prediction in acreages for soybeans at both scales during the test period (2015–2018), possibly due to confusion between the two crop types. Fallow/idle cropland has a relatively lower value of R^2, attributed to a wide range of conditions (from bare soil to annual cover crop) that fallow/idle cropland may represent; thus, making it more prone to misclassification. Aggregated tabular crop acreage summary products from the *cluster-then-label* model provide a high level of accuracy for applications at the scale of county or state administrative units.

4.4. Within-season mapping of crops

Within-season crop type classification was applied as crops grew in each of the four years from the test period (2015–2018) to test the practicality of mapping crops during the growing season. The model was applied iteratively at every new eight-day interval using the partial-year NDVI observations to-date. Fig. 10(a) shows the performance of within-season crop type classification for the eight area-dominant crop types across CONUS. As the growing season progresses, more crop pixels (shown on Fig. 10(a)) surpass the minimum greenness threshold and are able to be classified. 89% and 94% of all crop pixels of alfalfa and other hay grown across CONUS pass the greenup minimum by early-May and early-June, respectively. For major crops like winter wheat, corn and soybeans, more than 90% of pixels reach the minimum greenness threshold by early-June, mid-July and mid-August, respectively. More than 90% of rice and sorghum pixels exceeded the minimum greenup threshold by mid-August. Pixel-wise *User's Accuracy* improves as crops mature through the growing season and more phenology observations become available. While the accuracy of classification is low during early winter months, a large improvement is observed during July when corn and soybeans reach maturity. The earliest possible date of classification, defined as the date by which the within-season classification accuracy for a crop reaches 90% of the full-season accuracy, varies across crop types, based on differences in their phenology profiles. Fig. 10(a) shows the earliest possible date of classification for eight major crop types for the year 2015. Winter wheat, corn and soybeans are classified at 90% accuracy by early-August, mid-August and late-August, respectively. Fallow/idle cropland and other hay/

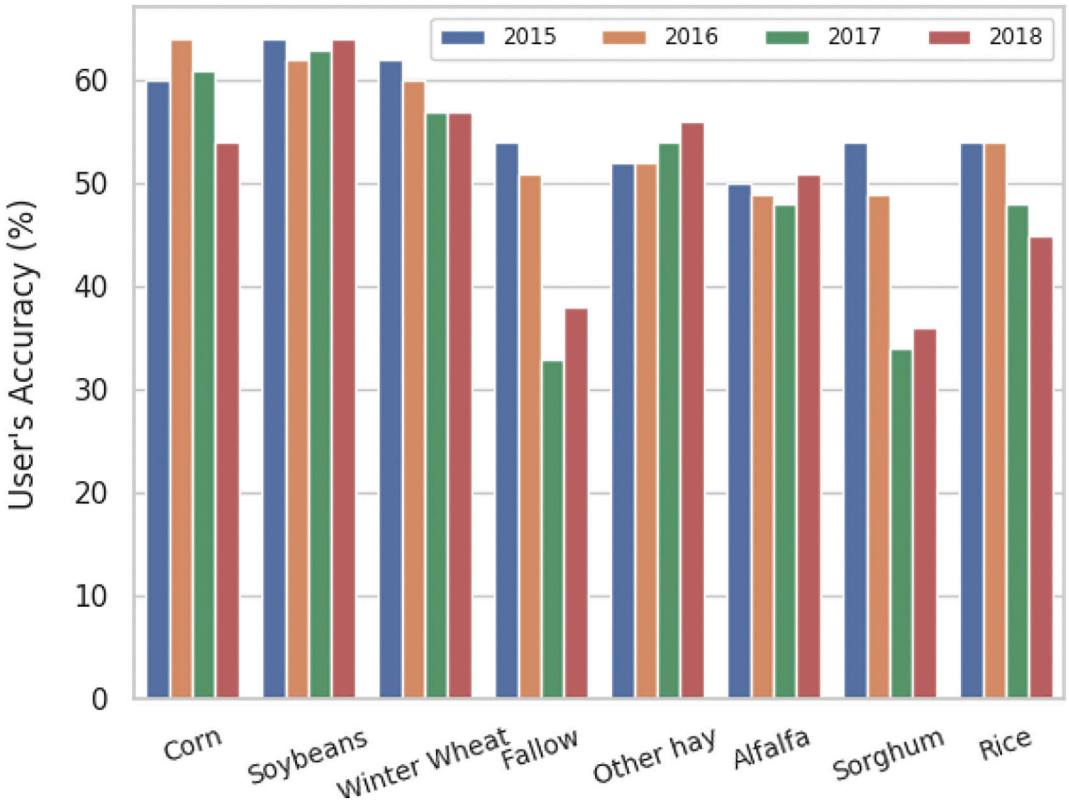

Fig. 6. The *cluster-then-label* model was evaluated on the test data set from never-seen-before years (2015–2018). *User's Accuracy* for the eight major crops is similar across all the four years, except for fallow/idle cropland and sorghum in 2017 and 2018 and corn in 2018.

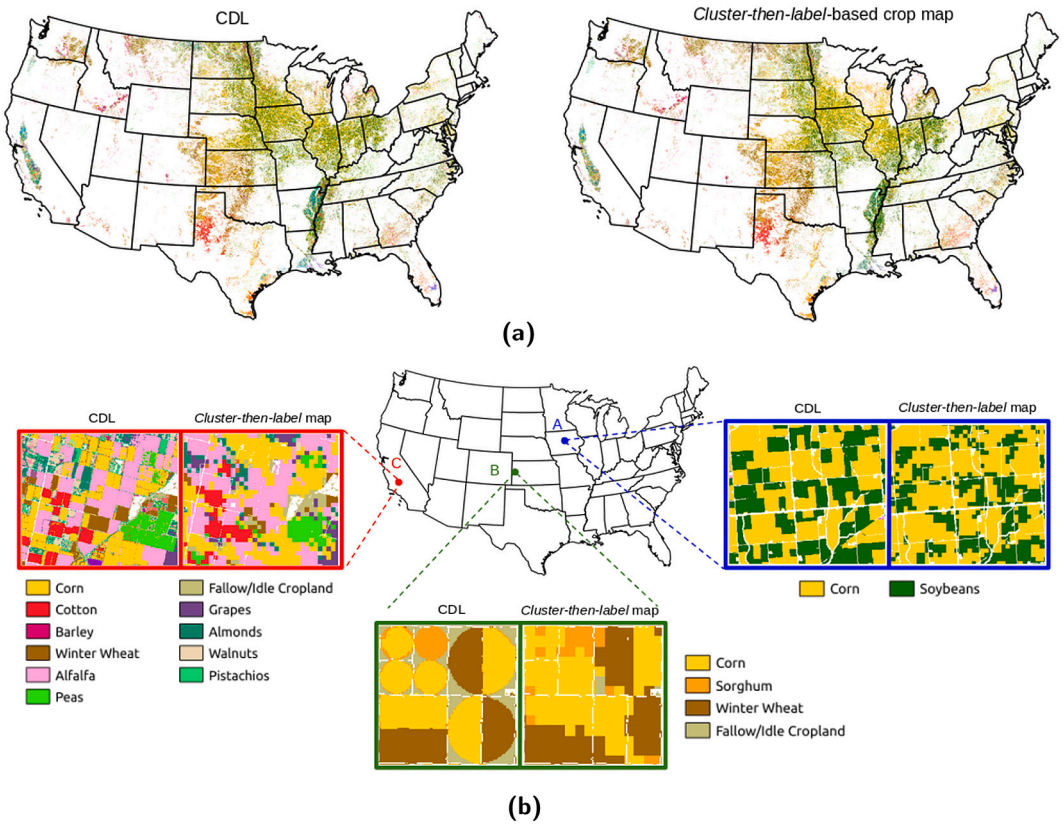

Fig. 7. Comparison of the *cluster-then-label-based* crop map with the USDA Crop Data Layer (CDL) for the year 2015 at different scales. (a) Comparison of the *cluster-then-label-based* crop map with the CDL shows similar patterns at the scale of the CONUS. (b) A closer look at three select regions (A, B, and C) shows a broad-level spatial agreement with CDL, but with some lack of sharpness and accuracy along field boundaries due to the coarser resolution of MODIS products

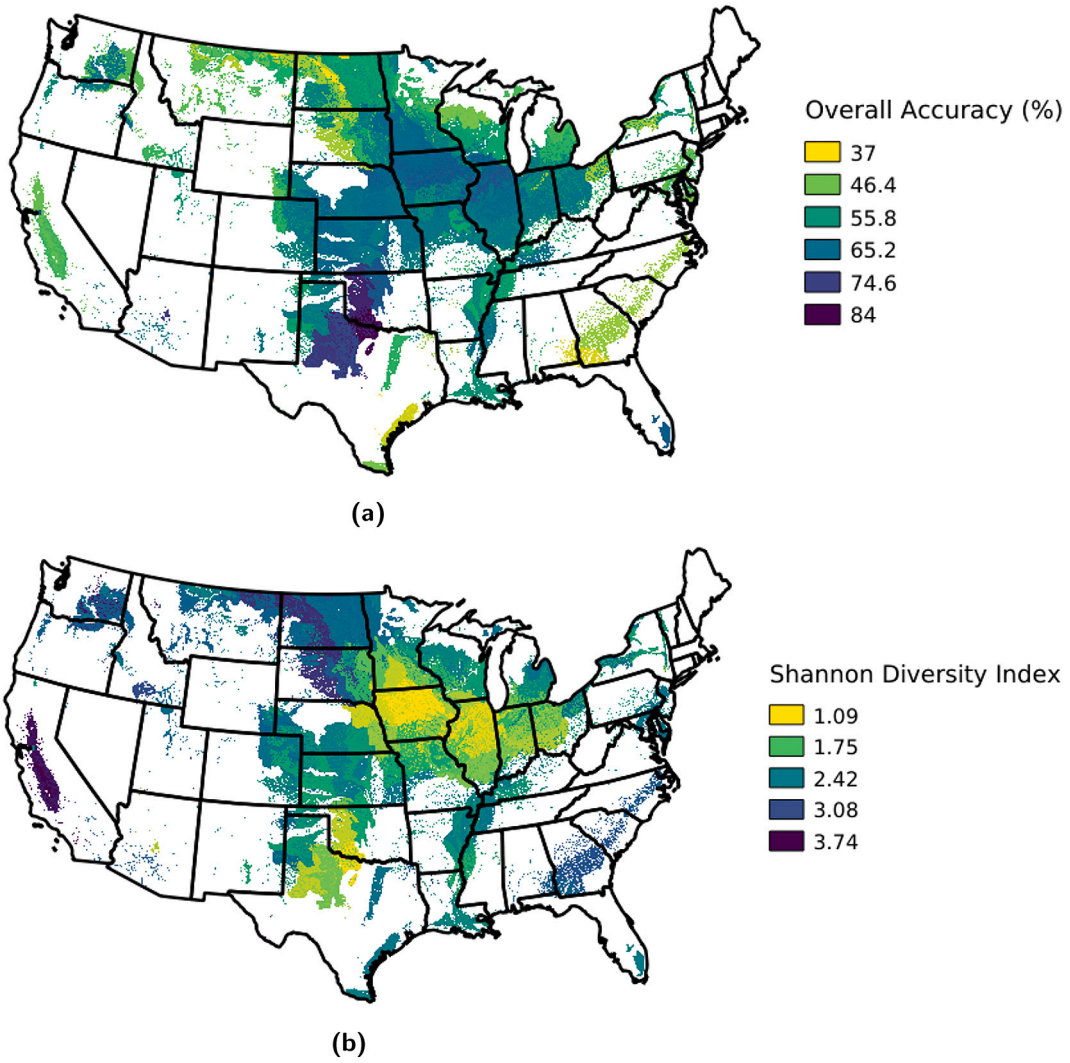

Fig. 8. Pixel-wise *Overall Accuracy* for *cluster-then-label-based* crop classification were found to be lower in regions with higher crop type diversity. (a) *Ecoregion*-wise Overall Accuracy of cluster-then-label-based crop classification for eight area-dominant crop types. (b) Shannon Crop Diversity for agriculturally-dominant ecoregions across the CONUS. The Overall Accuracy/Shannon Crop Diversity values were calculated only for those ecoregions which have at least 20% of their area covered by cropland.

non alfalfa are successfully classified by mid-August, and alfalfa can be mapped by late-September. The earliest possible date of classification for crops like sorghum and rice is around mid-November. These dates show some inter-annual variability during 2015–2018, probably driven by different meteorological conditions during the growing season, among other factors (Fig. 10(b)). Earliest dates of classification occurred around late-July to mid-August for corn, varied between mid-July and late-August for soybeans, and varied between the end of July and mid-August for winter wheat. The earliest date of classification for other hay/non alfalfa varied between July to early August over the four test years. Sorghum and rice had the earliest dates of classification, around mid-September to mid-November, and late-August to mid-November, respectively. Variability in earliest classification date was relatively small for major crop types, and was larger for less dominant crops like other hay, rice and sorghum.

Earliest possible dates of classification are also spatially variable. Fig. 11(a) shows spatial variability in earliest date of classification across the top ten corn-producing states for the year 2015. Corn-producing regions can be classified with 90% of full-season accuracy by mid-May across southern Minnesota, southern Wisconsin and northern Iowa. By early-June, corn can be identified with 90% accuracy across eastern Iowa, much of Illinois, western Indiana and eastern Missouri. By early-August, corn can be identified in western Iowa, eastern South Dakota and eastern Nebraska, and by late-October for eastern Indiana and western Ohio. Earliest date of classification for soybeans is generally later than corn, achievable only by early-July across western Ohio, Indiana, parts of western Iowa, southern Minnesota and eastern Nebraska. By the end of August, soybeans can be identified across much of Illinois and Iowa, but cannot be identified until early-October in eastern parts of North and South Dakota and northern Missouri.

5. Discussion

The goal of this study was to map crops across the CONUS during the active current growing season as they grow, a critical step in near real-time crop health monitoring. The framework outlined here could be extended to make periodically updated projections of yield as currently planted crops develop, thus aiding resource and economic planning and management at regional to national scales.

A *cluster-then-label*-based scheme was developed in this study that 1) classifies the croplands among dynamic *phenoregions* based on MODIS-based phenology and then 2) using CDL as training data, independently trains models within each *ecoregion*, to assign a crop label to an entire *phenoregion*. *Ecoregions* help to address spatial variability in phenology

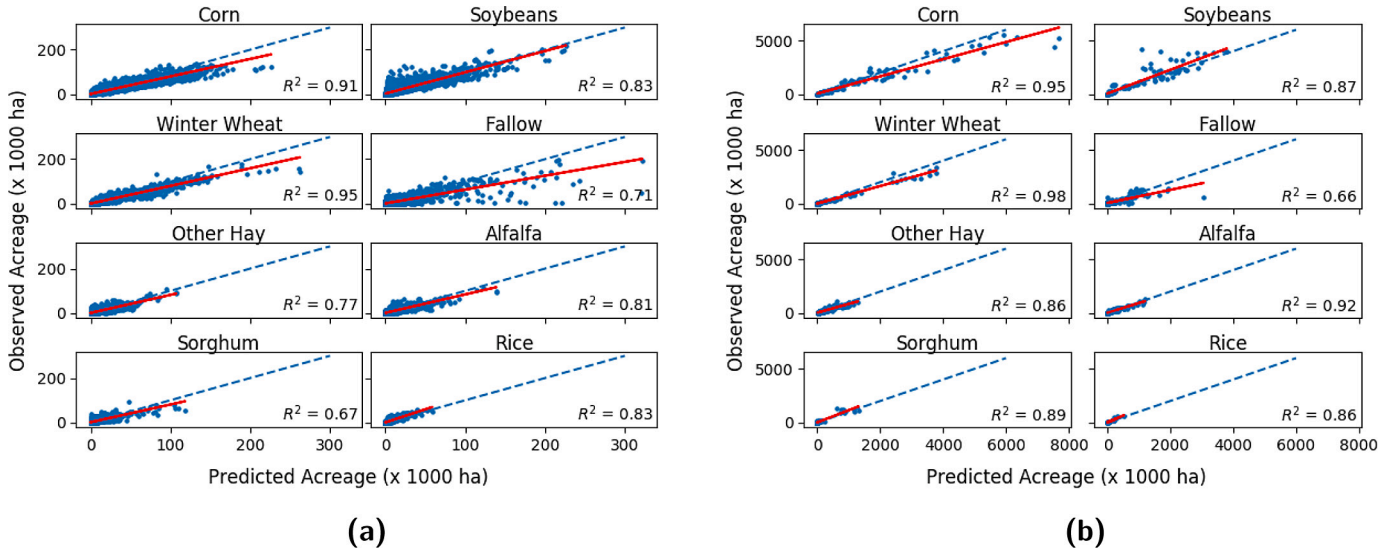

Fig. 9. Comparison of aggregated acreage estimated by *cluster-then-label* model with CDL during the test period 2015–2018 for eight area-dominant crops across CONUS. The dashed line represents the 1:1 line while the red line shows the linear fit for estimated vs expected acreage. Accuracies for aggregated areas show substantial improvement over pixel-wise accuracies in almost all cases, making *cluster-then-label* ideal for tabular crop acreage summaries. (a) Scatter plots comparing county-aggregated crop acreages for *cluster-then-label-based* crop map and CDL (b) Scatter plots comparing state-aggregated crop acreages for *cluster-then-label-based* crop map and CDL. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

due to differences in climate, soil and other growing conditions. Accuracy in crop type classification was improved when *cluster-then-label* models were trained within each *ecoregion*, compared to being trained for individual states or the entire CONUS. However, the increase in accuracy when going from state to *ecoregion* was modest for most crops, showing that even a state-level crop classification was fine enough to capture most local geographic changes in climatic and edaphic conditions, and farming practices like seed varieties, planting, fertilization and irrigation. Increasing spatial specificity to *ecoregions* yields diminishing returns in increased accuracy for major crops, but improvements are still substantial for less-common crops.

Pixel-wise *Producer's Accuracy* and *User's Accuracy* for major crops like corn, soybeans and winter wheat were greater than those of less-commonly grown crops during both training and testing periods. Major crops are generally grown on large, spatially dense fields relative to the size of MODIS pixels (231 m), whereas less dominant crops are grown on smaller, scattered fields, resulting in mixed-pixels at MODIS resolution. We compared our pixel-wise accuracies and R^2 values for county-aggregated acreages with those from two recently published studies on MODIS NDVI-based CONUS-scale crop mapping, Massey et al. (2017) and Dahal et al. (2018) (Tables S3 and S4 in the supplementary section). In spite of considering a greater number of crops, generating and comparing maps at the CDL resolution (30 m) and training and testing on different sets of years, among other differences, our classification accuracies are comparable to these two prior studies. The CDL's published accuracy also tends to be lower for lesser grown crop varieties (Boryan et al., 2011). Accuracy was lower (~10%) for all eight major crop types when the model was run on testing data from unseen-years (2015–2018), as compared to the validation data set representing 30% of the data from 2008 to 2014. The testing data came from unseen-years that were not used in training, thus potentially adding new phenological variation. *Re*-training after exposing the model to additional phenological variability from these additional years would presumably increase classification accuracies even further.

While the *cluster-then-label* method exploits the salient differences in phenological development of the crops, errors in the crop type classification sometimes remain, due to the inherent similarity in NDVI profiles among crop types. Fig. S3 shows box-and-whisker plots of NDVI profiles collected only from pure crop pixels (single crop type growing throughout the entire 231 m MODIS pixel) for major crops grown across Kansas in 2010. Corn and soybeans profiles show close similarity, with soybeans having a slightly later time to peak, due to their later sowing date. Phenology for winter wheat and fallow/idle cropland also are similar. Land left fallow often has grass cover, which grows quickly at the onset of spring, potentially leading to confusion with winter wheat. At the national scale, many of the classification errors occur between corn and soybeans, winter wheat and fallow/idle cropland and other hay/non alfalfa and alfalfa (Table 1).

Pixel-wise *User's Accuracy* for corn and *Producer's Accuracy* for soybeans are unexpectedly lower in 2018 as compared to the other test years (2015–2017). *User's Accuracy* for corn decreases from 69% to 57% for the four states: Iowa, Illinois, Indiana and Nebraska (Table 2) and the *Producer's Accuracy* for soybeans drops from 58% to 27%. The R^2 for corn and soybeans increases after the year 2018 is dropped from the analysis (Fig. S4). Soybeans were planted early in 2018 across these states (Fig. S5), which made its shifted NDVI profile more similar to that of corn, increasing confusion between these two crop types. Other crop types everywhere, as well as corn and soybeans in other geographic regions showed pixel-wise classification accuracies in 2018 that were similar to the rest of the novel testing years 2015–2017.

The *cluster-then-label* method for within-season classification classified major crops like corn and soybeans with 90% of full-season accuracy by the end of August, almost two to three months before harvest (Fig. 10). Other hay/non alfalfa can be mapped earlier in the growing season (early-August). Earliest dates of classification for smaller-area major crops like rice and sorghum is later in the growing season (between mid-September to mid-November), despite having roughly the same planting and harvest schedules as corn and soybeans. Classification of rice and sorghum with reasonable accuracy takes longer in part because of confusion with dominant crops like corn and soybeans. Classification accuracies for rice and sorghum increase in October and November, about when corn and soybeans are harvested.

Winter wheat is phenologically different from the other seven area-dominant crops in terms of its planting, growth and harvest schedule. While we often tend to think of phenological cycle in terms of Gregorian calendar (e.g., Fig. 2), crop phenological cycles are more meaningful in terms of growing season. Winter wheat is planted around September and is harvested in the summer or early Fall of the following calendar year. Hence, the phenological year for winter wheat spans across two calendar years. Fig. 10 shows that winter wheat can be

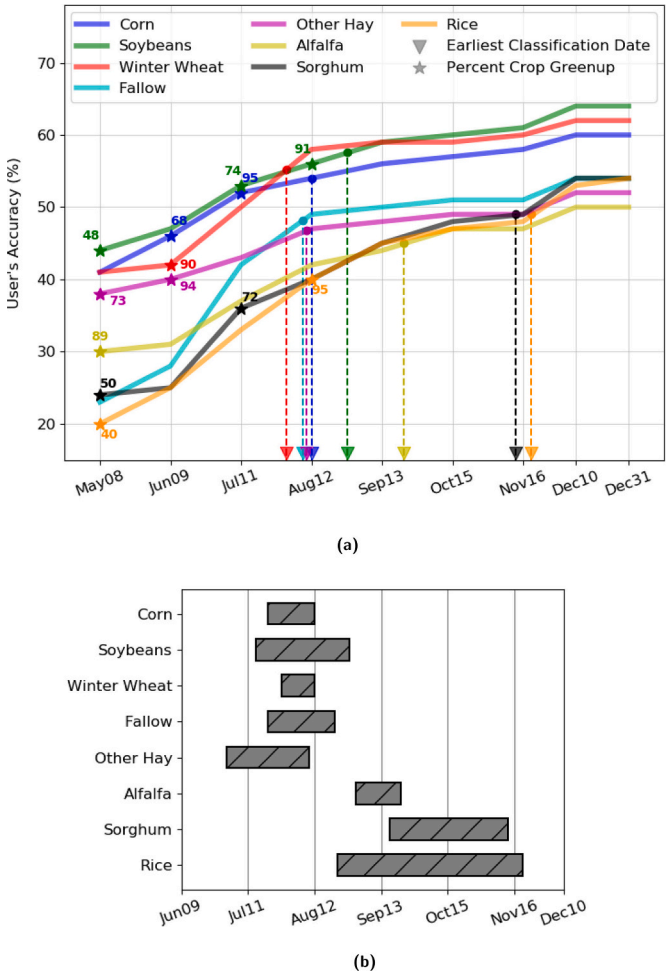

Fig. 10. Pixel-wise classification accuracy for national crop mapping within the growing season, and the earliest possible date of classification across the test period (2015–2018). (a) Improvement in pixel-wise *User's Accuracy* through time for national crop mapping during the growing season for the year 2015 is variable for different crops, but most can be mapped with 90% of full-season accuracy by July–September (stars/numbers represent the percent of corresponding crop pixels which have exceeded the minimum greenup threshold by that time). (b) Inter-annual variability in the earliest possible date of classification is small for major crops like corn and soybeans, compared to crops like sorghum and rice that show larger variability, and are growing in smaller, more spatially scattered fields

mapped with sufficient accuracy by early- to mid-August. There is a sudden increase in the *User's Accuracy* for winter wheat from June to August. Separate special confusion matrices constructed for wheat classification in May, June and July show that this improvement in accuracy is due to a decrease in confusion with corn and soybeans. In order to perform mid-season mapping for winter wheat, the monitoring period should ideally begin from September of the previous year. Waiting to begin the monitoring in January results in a loss of distinct phenological information in the first three to four months of wheat growth. Winter wheat would be the only green crop during this period, presumably improving discrimination of this unique crop type.

Earliest possible dates of classification at 90% of full-season accuracy show spatial variability across *ecoregions*, with dates ranging from early-May to late-October for both corn and soybeans. In general, 90% of full-season mapping accuracy is achieved early in the central regions of Corn Belt (parts of Iowa, Illinois and Indiana), as compared to the peripheral areas. Earliest dates of classification in the central Corn Belt were earlier for corn (June–July) than for soybeans (July–August).

6. Limitations, challenges and future steps

The ability to create a gap-filled remote sensing product that spans the whole CONUS is critical for near real-time crop health monitoring and commodity yield predictions. It requires remote sensing products that are corrected for missing values due to clouds/snow cover. NDVI values were used as an integrative proxy to capture crop land surface phenology. Past studies have included additional spectral bands spanning optical, Near Infrared, Short Wave Infrared (SWIR) and Synthetic-Aperture Radar (SAR), as well as indices that are derived from them, like Enhanced Vegetation Index (EVI), Green Chlorophyll Vegetation Index (GCVI), Land Surface Water index (LSWI), Normalized Difference Tillage Index (NDTI), among others. The addition of these bands and indices has been shown to improve classification accuracy, and future studies could include such additional metrics. The 231 m spatial resolution of MODIS is coarse for capturing the phenology of less dominant crops, which are typically grown on smaller fields, which may contribute to lower classification accuracies for these rarer crop types. Cross-sensor fusion could create a data product with high spatial and temporal frequency, thereby addressing the problem of mixed-pixel effect. The *Mapcurves* algorithm and *cluster-then-label* model assigned a single crop label (having the best Goodness-of-fit) to the entire *phenoregion* based on a single majority "winner takes all" strategy; however, other overlapping crop types might also be significant. A fuzzy labeling approach could be applied, or even more phenoregions could be used to distinguish even more-similar crop type phenology profiles from each other. Given the amount of remote sensing and CDL data available, more sophisticated machine learning, deep learning or Bayesian algorithms could also be tested.

Prior work has shown impressive results, but often with more specialized models classifying fewer crops, on smaller geographic regions, and/or tested on the same years as they were developed. We tested our general crop classification and mapping rigorously, on novel future years with which the model had no prior experience. Practical application of our crop classifier will likely be on the next unfolding growing season in the upcoming year, with unknown phenological deviations, and with which the model has no prior experience. Despite the enhanced difficulty of realistic testing on unseen years, the R^2 values of our general model were reasonable across all eight area-dominant crops during normal phenological years within the CONUS.

Summing crop acreages by type up to ever-larger accounting units also generally increases the accuracies. As we spatially aggregated, it is possible that some of the classification errors at the pixel level might be canceling out, leading to a dramatic improvement in accuracy results. Some prior efforts reported only these greater accuracies from such spatially aggregated results, instead of cell-by-cell accuracy/confusion results. The scale of results that are needed depends on the intended use, but, if actual spatially explicit national maps of crops are required, then the relevant accuracies are those reported at that finest spatial scale. These results are already nationally-scaled and predict all crop types, which are the desired features of a fully functional production system. We also have produced annual CDL-style maps from 2000 to 2007, during which no CDL maps were produced, and have made them available to download for general use.

Shifts in crop phenology from year to year were the major source of variability in crop mapping accuracy. Extreme weather conditions, such as floods or droughts, that may significantly affect timing of crop planting, resulting in unexpected shifts of deviations in phenology, can affect the accuracy of our phenology based classifications. For the year 2018, crop classification accuracies were unexpectedly low for the two crops, corn and soybeans, in one particular four-state geographic area, the US "breadbasket". Accuracies for other crops in this same location, and for all crops outside this region, were comparable with the other novel years during the testing period. Unusually early planting of soybeans, coupled with unusually fast phenological development of soybeans (Fig. S5), led to increased confusion between corn and soybeans

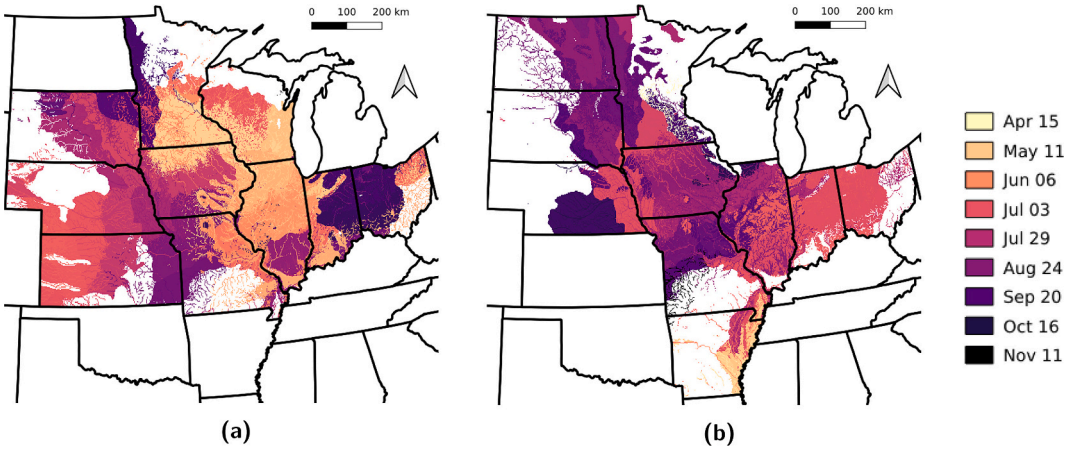

Fig. 11. Earliest possible date of crop classification for corn and soybeans during the year 2015 is highly variable across space. In general, corn is identified with 90% of full-season accuracy several weeks before soybeans. Earliest dates of classification are calculated for only those *ecoregions* which have at least 5% area of the respective crop. (a) Earliest possible date of crop classification for the top ten corn-producing states by area. (b) Earliest possible date of classification for the top ten soybeans-producing states by area.

Table 1
Pixel-wise confusion matrix for 2015 national crop type mapping (area in thousands of hectares). Diagonal values (shaded) represent crop classifications that agreed with the USDA Crop Data Layer (CDL). Winter Wheat abbreviated as Win Wht. Common classification confusions were corn for soybeans, winter wheat for fallow/idle cropland, and other hay/non alfalfa for alfalfa.

		Crop Data Layer							
		Corn	Soybeans	Win Wht	Fallow	Other Hay	Alfalfa	Sorghum	Rice
Reclassed Map	Corn	25,519	10,748	742	976	864	956	689	72
	Soybeans	5,735	18,161	329	950	420	234	423	429
	Win Wht	521	254	10,535	1,772	428	422	704	1
	Fallow	328	464	1,008	5,709	733	254	222	84
	Other Hay	1,088	916	477	706	6,272	1,271	76	2
	Alfalfa	823	311	496	407	966	4,098	46	1
	Sorghum	155	62	137	175	59	6	861	3
	Rice	53	106	1	63	1	1	26	365

within this region during this year.

These results underscore the overarching importance of inter-year phenological variability, and the resultant effects on the timing of planting and development. The relatively fine distinctions between "normal" phenologies of soybeans and corn can be overwhelmed by the magnitude of between-year phenological variability in some locations during some years. Similarly, crop failures and late re-planting will lower pixel-wise accuracies, as will any agricultural practices that alter the expected timing of phenological development upon which the separation of crop types are distinguished.

The CDL, which served as our training data, itself is a classification product and likely contains errors that will be propagated forward. Underlying ground-based observations used to train CDL themselves are not publicly available for independent use or testing, due to privacy and proprietary agri-business concerns. The CDL, therefore, represents the only data source for training and testing crop classification approaches such as ours. As crop mapping models increase in prognostic power, this limited availability of public, error-free training data may become the greatest limitation to future progress in remote sensing-based national crop classification.

Table 2
Pixel-wise mapping accuracy for the unusual year 2018 was lower in the four major crop-producing states: Iowa, Illinois, Indiana and Nebraska due to earlier planting of soybeans, which resulted in lower accuracy at CONUS-scale.

Accuracy Metric	2018	2015–2017
User's Accuracy for Corn (%)	57	69
Producer's Accuracy for Soybeans (%)	27	58

The *cluster-then-label* method developed here could be fully automated and integrated into an online mapping system, like the United States Forest Service's *ForWarn* (https://forwarn.forestthreats.org/). *ForWarn* is a vegetation change recognition and tracking system that provides near real-time change maps for the continental United States that are updated every eight days, using MODIS NDVI. *ForWarn* tracks disturbance in all vegetation, not just forests, including potential disturbances in rangeland vegetation and agricultural crops. Unlike forests that (usually) remain growing in the same places from year to year, farmers often plant different crops in the same field, using an unpredictable rotation system. If the crop planted this year has been changed, the normal NDVI value that is used for baseline comparison with the current observed NDVI will be inappropriate, and the relative crop health status shown by *ForWarn* will be incorrect. However, if *ForWarn* could be provided with temporally improving maps of crop types planted in this current growing season, then crop health could be monitored nationally every 8 days, along with the health of forests and rangelands. Such spatially explicit crop predictions are possible, starting as early as August of the current growing season.

7. Conclusions

Developing crop maps over large areas during the growing season is important for forecasting crop yield and food production at national scales. The goal of this study was to produce national-scale crop maps at 8-day intervals during the growing season. We first developed a *cluster-then-label* approach to create end-of-growing season crop maps for CONUS. This was done using a generalized classification approach consisting of two steps: 1) creating *phenoregions* based on Multivariate Spatio-Temporal Clustering of annual time series of 8-day NDVI collected for every 231 m pixel on the ground across CONUS for the years 2000–2018, and 2) assigning crop labels to *phenoregions* based on the degree of spatial concordance between crop growing areas and entire, individual *phenoregions*. Spatial and temporal variability in phenology increases the challenges of national crop mapping and were addressed by training the *cluster-then-label* models within each *ecoregion* and on multiple years (2008–2014), respectively. The resulting maps compare well with the CDL. Overall accuracy of classification was around 70% across major corn, soybeans and winter wheat-producing regions, while accuracy was lower in areas with greater crop diversity.

We then used this approach to generate crop maps for CONUS well-before harvest, and to estimate the earliest time during the growing season by which crops could be mapped with sufficient accuracy. Major crops like corn, soybeans, winter wheat, fallow/idle cropland and other hay/non alfalfa could be mapped as early as August across CONUS with 90% of the full-season accuracy. We also produced CDL-like maps for the years 2000–2007, before any such maps existed, and we have made them available to download. More than a demonstration of feasibility on a limited geographic area or for only a few crop types, our *cluster-then-label* method provides a fully scaled production capability for practical near real-time mapping of all crops as they grow and mature across CONUS. Running updated projections of final crop yields for each planted crop during the growing season, estimated from historical productivity data per hectare within each ecoregion, may be a feasible next step.

Data availability

The data products from this study are publicly available at doi:https://doi.org/10.5281/zenodo.3478335. The data collection includes annual crop type maps for the period 2000–2018. It also includes the earliest dates of classification for eight dominant crop types (corn, soybeans, winter wheat, fallow/idle cropland, alfalfa, other hay/non-alfalfa, sorghum, and rice) during 2015.

Declaration of Competing Interest

None.

Acknowledgments

This research was partially sponsored by the US Department of Agriculture, US Forest Service, Eastern Forest Environmental Threat Assessment Center. Additional support was provided by the Reducing Uncertainties in Biogeochemical Interactions through Synthesis and Computation (RUBISCO) Science Focus Area (SFA), which is sponsored by the Regional and Global Model Analysis (RGMA) activity in the Earth and Environmental Systems Sciences Division (EESSD) of the Biological and Environmental Research (BER) office in the US Department of Energy Office of Science. This research used resources of the Oak Ridge Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC05-00OR22725. This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the US Department of Energy. The United States Government retains and the publisher, by accepting the article for publication, acknowledges that the United States Government retains a non-exclusive, paid-up, irrevocable, world-wide license to publish or reproduce the published form of this manuscript, or allow others to do so, for United States Government purposes. The Department of Energy will provide public access to these results of federally sponsored research in accordance with the DOE Public Access Plan (http://energy.gov/downloads/doe-public-access-plan). Funding for Auroop R. Ganguly was provided by National Science Foundation (NSF) through its BIGDATA award 1447587, CyberSEES award 1442728 and CRISP Type 2 award 1735505.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rse.2020.112048.

References

Becker-Reshef, I., Justice, C., Sullivan, M., Vermote, E., Tucker, C., Anyamba, A., Small, J., Pak, E., Masuoka, E., Schmaltz, J., et al., 2010. Monitoring global croplands with coarse resolution earth observations: the global agriculture monitoring (GLAM) project. Remote Sens. 2, 1589–1609. https://doi.org/10.3390/rs2061589.

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.

Boyer, J.S., Byrne, P., Cassman, K.G., Cooper, M., Delmer, D., Greene, T., Gruis, F., Habben, J., Hausmann, N., Kenny, N., et al., 2013. The US drought of 2012 in perspective: a call to action. Global Food Security 2, 139–143. https://doi.org/10.1016/j.gfs.2013.03.002.

Cai, Y., Guan, K., Peng, J., Wang, S., Seifert, C., Wardlow, B., Li, Z., 2018. A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach. Remote Sens. Environ. 210, 35–47. https://doi.org/10.1016/j.rse.2018.02.045.

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.

FAO, 2019. Global Information and Early Warning System. http://www.fao.org/giews/en/ (last date accessed: 06/18/2019).

Fick, S.E., Hijmans, R.J., 2017. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. Int. J. Climatol. 37, 4302–4315. https://doi.org/10.1002/joc.5086.

Fischer, G., van Velthuizen, H.T., Nachtergaele, F.O., 2000. Global agro-ecological zones assessment: methodology and results. In: Technical Report. IIASA, Laxenburg, Austria URL. http://pure.iiasa.ac.at/id/eprint/6182/1/IR-00-064.pdf (Last date accessed: 10/14/2019).

Global Soil Data Task Group, 2000. Global Gridded Surfaces of Selected Soil Characteristics (IGBP-DIS). [Global Gridded Surfaces of Selected Soil Characteristics (International Geosphere-Biosphere Programme - Data and Information System)]. https://doi.org/10.3334/ORNLDAAC/569.

Gómez, C., White, J.C., Wulder, M.A., 2016. Optical remotely sensed time series data for land cover classification: a review. ISPRS J. Photogramm. Remote Sens. 116, 55–72. https://doi.org/10.1016/j.isprsjprs.2016.03.008.

Gumma, M.K., Thenkabail, P.S., Teluguntla, P., Rao, M.N., Mohammed, I.A., Whitbread, A.M., 2016. Mapping rice-fallow cropland areas for short-season grain legumes intensification in south asia using MODIS 250 m time-series data. Int. J. Digital Earth 9, 981–1003. https://doi.org/10.1080/17538947.2016.1168489.

Hargrove, W.W., Hoffman, F.M., 2004. Potential of multivariate quantitative methods for delineation and visualization of ecoregions. Environ. Manag. 34, S39–S60. https://doi.org/10.1007/s00267-003-1084-0.

Hargrove, W.W., Hoffman, F.M., Hessburg, P.F., 2006. Mapcurves: a quantitative method for comparing categorical maps. J. Geograph. Syst. 8, 187. https://doi.org/10.1007/s10109-006-0025-x.

Hartigan, J., 1975. Clustering Algorithms. John Wiley & Sons.

Hoffman, F.M., Hargrove, W.W., Mills, R.T., Mahajan, S., Erickson, D.J., Oglesby, R.J., 2008. Multivariate spatio-temporal clustering (MSTC) as a data mining tool for environmental applications. In: Sànchez-Marrè, M., Béjar, J., Comas, J., Rizzoli, A.E., Guariso, G. (Eds.), Proceedings of the iEMSs Fourth Biennial Meeting: International Congress on Environmental Modelling and Software Society (iEMSs 2008), pp. 1774–1781. https://scholarsarchive.byu.edu/iemssconference/2008/all/27/?utm_source=scholarsarchive.byu.edu%2Fiemssconference%2F2008%2Fall%2F27&utm_medium=PDF&utm_campaign=PDFCoverPages.

Hoffman, F.M., Mills, R.T., Kumar, J., Vulli, S.S., Hargrove, W.W., 2010. Geospatiotemporal data mining in an early warning system for forest threats in the United States. In: Proceedings of the 2010 IEEE International Geoscience and Remote Sensing Symposium (IGARSS 2010), pp. 170–173. https://doi.org/10.1109/IGARSS.2010.5653935.

Justice, C.O., Becker-Reshef, I., 2007. Report from the Workshop on Developing a Strategy for Global Agricultural Monitoring in the Framework of Group on Earth Observations (GEO). Avaliable online: http://www.fao.org/gtos/igol/docs/meeting-reports/07-geo-ag0703-workshop-report-nov07.pdf (accessed on 11 June 2015).

Kumar, J., Mills, R.T., Hoffman, F.M., Hargrove, W.W., 2011. Parallel *k*-means clustering for quantitative ecoregion delineation using large data sets. Proc. Comput. Sci. 4, 1602–1611. https://doi.org/10.1016/j.procs.2011.04.173.

Kussul, N., Lavreniuk, M., Skakun, S., Shelestov, A., 2017. Deep learning classification of land cover and crop types using remote sensing data. IEEE Geosci. Remote Sens. Lett. 14, 778–782. https://doi.org/10.1109/LGRS.2017.2681128.

Massey, R., Sankey, T.T., Congalton, R.G., Yadav, K., Thenkabail, P.S., Ozdogan, M., Meador, A.J.S., 2017. MODIS phenology-derived, multi-year distribution of conterminous US crop types. Remote Sens. Environ. 198, 490–503. https://doi.org/10.1016/j.rse.2017.06.033.

Mills, R.T., Hoffman, F.M., Kumar, J., Hargrove, W.W., 2011. Cluster analysis-based approaches for geospatiotemporal data mining of massive data sets for identification of forest threats. Proc. Comput. Sci. 4, 1612–1621. https://doi.org/10.1016/j.procs.2011.04.174.

Pittman, K., Hansen, M.C., Becker-Reshef, I., Potapov, P.V., Justice, C.O., 2010. Estimating global cropland extent with multi-year MODIS data. Remote Sens. 2, 1844–1863. https://doi.org/10.3390/rs2071844.

Sakamoto, T., Wardlow, B.D., Gitelson, A.A., 2011. Detecting spatiotemporal changes of corn developmental stages in the US Corn Belt using MODIS WDRVI data. IEEE Trans. Geosci. Remote Sens. 49, 1926–1936. https://doi.org/10.1109/TGRS.2010.2095462.

Saxon, E., Baker, B., Hargrove, W., Hoffman, F., Zganjar, C., 2005. Mapping environments at risk under different global climate change scenarios. Ecol. Lett. 8, 53–60. https://doi.org/10.1111/j.1461-0248.2004.00694.x.

Shao, Y., Lunetta, R.S., 2012. Comparison of support vector machine, neural network, and CART algorithms for the land-cover classification using limited training data points. ISPRS J. Photogramm. Remote Sens. 70, 78–87. https://doi.org/10.1016/j.isprsjprs.2012.04.001.

Shao, Y., Lunetta, R.S., Ediriwickrema, J., Iiames, J., 2010. Mapping cropland and major crop types across the Great Lakes Basin using MODIS-NDVI data. Photogramm. Eng. Rem. Sens. 76, 73–84. https://doi.org/10.14358/PERS.76.1.73.

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.

Spruce, J., Gasser, G., Hargrove, W., 2016. MODIS NDVI Data, Smoothed and Gap-Filled, for the Conterminous US: 2000–2015. https://doi.org/10.3334/ORNLDAAC/1299.

Spruce, J.P., Sader, S., Ryan, R.E., Smoot, J., Kuper, P., Ross, K., Prados, D., Russell, J., Gasser, G., McKellip, R., et al., 2011. Assessment of MODIS NDVI time series data products for detecting forest defoliation by gypsy moth outbreaks. Remote Sens. Environ. 115, 427–437. https://doi.org/10.1016/j.rse.2010.09.013.

USAID, 1985. Famine Early Warning System Network. http://fews.net/ (last date accessed: 07/03/2019).

USDA, 2019a. Crop Progress Reports. Website. URL. https://usda.library.cornell.edu/concern/publications/8336h188j?locale=en (last date accessed: 06/14/2019).

USDA, 2019b. Cropland Data Layer. Website. URL. https://www.nass.usda.gov/Research_and_Science/Cropland/Release/index.php (last date accessed: 06/18/2019).

USDA, 2019c. World Agricultural Supply and Demand Estimates. Website. URL. https://www.usda.gov/oce/commodity/wasde/index.htm (last date accessed: 06/14/2019).

Waldner, F., Canto, G.S., Defourny, P., 2015a. Automated annual cropland mapping using knowledge-based temporal features. ISPRS J. Photogramm. Remote Sens. 110, 1–13. https://doi.org/10.1016/j.isprsjprs.2015.09.013.

Waldner, F., Fritz, S., Di Gregorio, A., Defourny, P., 2015b. Mapping priorities to focus cropland mapping activities: fitness assessment of existing global, regional and national cropland maps. Remote Sens. 7, 7959–7986. https://doi.org/10.3390/rs70607959.

Wang, S., Azzari, G., Lobell, D.B., 2019. Crop type mapping without field-level labels: random forest transfer and unsupervised clustering techniques. Remote Sens. Environ. 222, 303–317. https://doi.org/10.1016/j.rse.2018.12.026.

Welton, G., 2011. The Impact of Russia's 2010 Grain Export Ban, Oxfam in Association With GSE Research. pp. 76–107 URL. https://www-cdn.oxfam.org/s3fs-public/file_attachments/rr-impact-russias-grain-export-ban-280611-en_3.pdf.

White, M.A., Hoffman, F., Hargrove, W.W., Nemani, R.R., 2005. A global framework for monitoring phenological responses to climate change. Geophys. Res. Lett. 32. https://doi.org/10.1029/2004GL021961.

Williams, C.L., Hargrove, W.W., Liebman, M., James, D.E., 2008. Agro-ecoregionalization of Iowa using multivariate geographical clustering. Agric. Ecosyst. Environ. 123, 161–174. https://doi.org/10.1016/j.agee.2007.06.006.

Xiong, J., Thenkabail, P.S., Gumma, M.K., Teluguntla, P., Poehnelt, J., Congalton, R.G., Yadav, K., Thau, D., 2017. Automated cropland mapping of continental Africa using Google earth engine cloud computing. ISPRS J. Photogramm. Remote Sens. 126, 225–244. https://doi.org/10.1016/j.isprsjprs.2017.01.019.

Zhong, L., Gong, P., Biging, G.S., 2014. Efficient corn and soybean mapping with temporal extendability: a multi-year experiment using Landsat imagery. Remote Sens. Environ. 140, 1–13. https://doi.org/10.1016/j.rse.2013.08.023.

Zhong, L., Yu, L., Li, X., Hu, L., Gong, P., 2016. Rapid corn and soybean mapping in US Corn Belt and neighboring areas. Sci. Rep. 6, 36240. https://doi.org/10.1038/srep36240.

Zhong, L., Hu, L., Zhou, H., 2019. Deep learning based multi-temporal crop classification. Remote Sens. Environ. 221, 430–443. https://doi.org/10.1016/j.rse.2018.11.032.