Remote Sensing of Environment 210 (2018) 35–47

Contents lists available at ScienceDirect

Remote Sensing of Environment

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

A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach

Yaping Cai[a,b], Kaiyu Guan[b,c,*], Jian Peng[d,**], Shaowen Wang[a,b], Christopher Seifert[e], Brian Wardlow[f], Zhan Li[g]

[a] CyberGIS Center for Advanced Digital and Spatial Studies, Department of Geography and Geographical Information Sciences, University of Illinois at Urbana-Champaign, Urbana, IL, United States
[b] National Center for Supercomputing Center, University of Illinois at Urbana Champaign, Urbana, IL, United States
[c] Department of Natural Resources and Environmental Sciences, University of Illinois at Urbana Champaign, Urbana, IL, United States
[d] Department of Computer Science, University of Illinois at Urbana Champaign, Urbana, IL, United States
[e] Department of Earth System Science, Stanford University, Stanford, CA, United States
[f] Center for Advanced Land Management Technologies, School of Natural Resources, University of Nebraska-Lincoln, United States
[g] School for the Environment, University of Massachusetts Boston, Boston, MA, United States

ARTICLE INFO

Keywords:
Crop type classification
Machine learning
Remote sensing
Phenology
Deep Neural Network (DNN)

ABSTRACT

Accurate and timely spatial classification of crop types based on remote sensing data is important for both scientific and practical purposes. Spatially explicit crop-type information can be used to estimate crop areas for a variety of monitoring and decision-making applications such as crop insurance, land rental, supply-chain logistics, and financial market forecasting. However, there is no publically available spatially explicit in-season crop-type classification information for the U.S. Corn Belt (a landscape predominated by corn and soybean). Instead, researchers and decision-makers have to wait until four to six months after harvest to have such information from the previous year. The state-of-the-art research on crop-type classification has been shifted from relying on only spectral features of single static images to combining together spectral and time-series information. While Landsat data have a desirable spatial resolution for field-level crop-type classification, the ability to extract temporal phenology information based on Landsat data remains a challenge due to low temporal revisiting frequency and inevitable cloud contamination. To address this challenge and generate accurate, cost-effective, and in-season crop-type classification, this research uses the USDA's Common Land Units (CLUs) to aggregate spectral information for each field based on a time-series Landsat image data stack to largely overcome the cloud contamination issue while exploiting a machine learning model based on Deep Neural Network (DNN) and high-performance computing for intelligent and scalable computation of classification processes. Experiments were designed to evaluate what information is most useful for training the machine learning model for crop-type classification, and how various spatial and temporal factors affect the crop-type classification performance in order to derive timely crop type information. All experiments were conducted over Champaign County located in central Illinois, and a total of 1322 Landsat multi-temporal scenes including all the six optical spectral bands spanning from 2000 to 2015 were used. Computational experiments show the inclusion of temporal phenology information and evenly distributed spatial training samples in the study domain improves classification performance. The shortwave infrared bands show notably better performance than the widely used visible and near-infrared bands for classifying corn and soybean. In comparison with USDA's Crop Data Layer (CDL), this study found a relatively high Overall Accuracy (i.e. the number of the corrected classified fields divided by the number of the total fields) of 96% for classifying corn and soybean across all CLU fields in the Champaign County from 2000 to 2015. Furthermore, our approach achieved 95% Overall Accuracy by late July of the concurrent year for classifying corn and soybean. The findings suggest the methodology presented in this paper is promising for accurate, cost-effective, and in-season classification of field-level crop types, which may be scaled up to large geographic extents such as the U.S. Corn Belt.

* Correspondence to: K. Guan, National Center for Supercomputing Center, University of Illinois at Urbana Champaign, Champaign, IL, United States.
** Corresponding author.
E-mail addresses: kaiyug@illinois.edu (K. Guan), jianpeng@illinois.edu (J. Peng).

https://doi.org/10.1016/j.rse.2018.02.045
Received 24 April 2017; Received in revised form 3 January 2018; Accepted 22 February 2018
Available online 16 March 2018

1. Introduction

Accurately classifying crop types is important for both scientific and practical purposes. Classifying land cover is a classic question in the remote sensing field, and has been an active research topic for decades (Hansen et al., 2014, 2011, 2000; Hansen and Loveland, 2012; King et al., 2017; Sexton et al., 2013b; Song et al., 2017; Vogelmann et al., 2001; Zhan et al., 2002). However, how to generate accurate and timely maps for crop types with high spatial resolution remains a scientific challenge. Currently, we have no in-season crop type data available for large-scale US croplands. For example, though the USDA publishes the Cropland Data Layer (CDL) data at 30-m spatial resolution, it is usually released in the spring of the subsequent year, with a time lag of at least four to six months after the previous year's harvest time (Boryan et al., 2011). For practical purposes, accurate and timely crop-type classification provides estimations of the planting/harvesting crop areas for a variety of monitoring and decision-making applications of government and private sectors such as crop insurance, land rental, supply-chain logistics, commodity markets, etc. Furthermore, crop-type classification is also the prerequisite for conducting crop yield prediction (Bolton and Friedl, 2013; Lobell et al., 2015). As a result, accurate and in-season information of crop types has considerable importance for management decision-making in public/private sectors and regional economic forecasting.

Extensive research has been done in crop-type classification using two major classification strategies (Chang et al., 2007; Foerster et al., 2012; Lobell and Asner, 2004; Van Niel and McVicar, 2004). One is to solely use the spectral features from a single satellite scene sampled during a certain day within a growing season (Boryan et al., 2011; Van Niel and McVicar, 2004; Yang et al., 2011), and the other is to use both spectral and temporal information during one or multiple growing seasons (Chang et al., 2007; Foerster et al., 2012; Wardlow et al., 2007; Wardlow and Egbert, 2008). The first strategy is based on the rationale that different land covers have distinctive spectral features, and these spectral features in turn can be used for classification. However, some crops have similar spectral information during the peak-growing season when the satellite image is usually acquired, which makes separation of crop types difficult. In addition, spectral differences between crops and natural vegetation (e.g. grass or trees) may be small at certain times of a year. As a result, the similar spectral features between different crops as well as between crops and natural vegetation pose a major challenge for accurate classification. The second strategy utilizes both the spectral and temporal information, which leads to improvements in classification accuracy. Crops usually have different seasonal variations and sowing dates. For example, in the U.S. Corn Belt, corn is usually sown earlier than soybean, and grass usually starts its growing season in spring that is earlier than most crops. These temporal features can be used to improve the accuracy of crop classification. However, the second strategy requires time-series information from multiple satellite images rather than from a single image, and traditionally researchers have implemented this approach using data from sensors with low- or medium- spatial resolution such as the Moderate Resolution Imaging Spectroradiometer (MODIS) (Wardlow et al., 2007; Wardlow and Egbert, 2008).

To achieve field-level classification of crop types, appropriate spatial resolution satellite data inputs to field sizes are required (Lobell, 2013). For the U.S. context, such satellites exist, such as Landsat (Hansen and Loveland, 2012; Roy et al., 2014). Landsat imagery has a higher spatial resolution (30 m) than low- or medium- spatial resolution e.g. MODIS data (gridded at 250 m, 500 m or larger pixel sizes); and unlike SPOT (Duro et al., 2012) and other commercial satellite data, Landsat data is freely available for both concurrent and historical periods. In addition, advanced Landsat products such as the surface reflectance (after atmospheric correction) are readily available from the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) (Schmidt et al., 2013) and the Landsat Surface Reflectance Code (LaSRC) (USGS, 2016) for Landsat 5, 7 and 8. Landsat data has been widely used for land cover classification at local, regional or continental scales (Hansen and Loveland, 2012; Homer et al., 2004; Huang et al., 2007; Liu et al., 2005; Sexton et al., 2013a; Townshend et al., 2012; Yuan et al., 2005). However, Landsat has a low temporal resolution (16-day revisiting cycle compared to the 1–2-day revisiting cycle of MODIS), and clouds frequently contaminate Landsat images. Extracting the continuous time-series information based on Landsat data (especially how to handle missing data because of cloud cover) is a challenge. To utilize both high spatial and temporal information in Landsat, researchers have explored data-fusion approaches to integrate multi-sources of remotely sensed data, for example, fusing MODIS and Landsat data to achieve both high spatial and temporal resolutions (Gao et al., 2015, 2013). However, the existing data-fusion approaches usually fill the gap values from neighboring available pixels by assuming that different periods of satellite images have unchanged land cover types, thus contradicting the purpose of identifying land cover changes over the time. Additionally, fused satellite data is currently not available or operationally provided at a large spatial scale.

As an alternative, we use the Common Land Unit (CLU) to aggregate field level information based on time-series Landsat data. CLUs are generated by the USDA to delineate the field boundary for all registered agricultural fields for the U.S. (Boryan et al., 2011). The average size of a single unit of CLU in Champaign County, IL, is 60.3 ± 52.6 acres (~244,025 ± 212,865 m^2), which is about 16 × 16 ± 15 × 15 30-meter Landsat pixels (Fig. S1). When a CLU field has a sub-field contamination by clouds/shadows in a Landsat scene, we aggregate Landsat information by averaging values from non-cloud only pixels within that field and assign the mean value to that CLU field. Thus the contamination issues can be overcome to a desirable extent, and as a result, the weakness of lower temporal-resolution Landsat data can be largely alleviated. The aggregated and field-level spectral information will then be used for classification. In addition, instead of only using the data for the same year for training/testing for crop-type classification (Boryan et al., 2011; Wardlow and Egbert, 2008), we can also use the data from multiple growing seasons for training our classification model, with the premise that multiple-year data include more scenarios of crop phenology due to various other factors (e.g. sowing date, inter-annual climate variability) and thus can make our classification algorithm more generic and robust when applying to a new year.

Machine learning approaches have been applied to a variety of data-driven predictive applications, such as natural language understanding and image processing (Collobert and Weston, 2008; Hinton et al., 2012; Krizhevsky et al., 2012). Recently, deep learning, including both the Deep Neural Network (DNN) and the Convolutional Neural Network (CNN), shows great potential in various applications compared to other machine learning techniques. Traditionally, classification or regression systems require careful engineering and considerable domain knowledge to extract features from raw data. However, deep learning has the ability to discover informative features with multiple levels of representation, from lower, primitive levels to higher, abstract levels (LeCun et al., 2015; Schmidhuber, 2015). Though neural network methods have been developed several decades ago, recently years see major development in this method through more layers and back-propagation optimization (i.e. deep neural network), which has made significant improvements in classification or other applications (LeCun et al., 2015; Schmidhuber, 2015). Deep learning is still early in its application on remote sensing data for crop-type classification; therefore questions like what information is needed and how to transform the information that can be used in deep learning model need to be answered.

This paper describes a new crop classification system that is targeted at the U.S. Corn Belt, a region dominated by corn and soybeans. We only focused on farmland and pre-filtered other types of land cover (based on CDL), and classified all the patches of farmland into three major categories: corn, soybean and others. We used CLU to aggregate

the information of each field based on time-series Landsat data in order to alleviate the cloud contamination issue, and then built a deep-learning classification model based on DNN using high-performance computing. The research was designed to understand how different spatial and temporal features affect the classification performance. The selected study region of Champaign County, IL is located in the middle of the U.S. Corn Belt with a landscape dominated by corn and soybean production. Two supercomputers, ROGER and Blue Waters at the University of Illinois, are used for generating the time-series Landsat data stack. ROGER has large memory space that allows rapid pre-processing of the large spatial data; and Blue Waters, with richer computing resources, was used to build the classification model through intensive training and testing. Specifically, we address the following two overarching research questions:

(1) What accuracy can the integrated time-series Landsat data and deep learning approach achieve for the crop-type classification?

(2) How early in the growing season can this method achieve the optimal accuracy of crop-type classification using our approach?

2. Data and methods

2.1. Study area

Champaign County, Illinois, has a total area of 638,767 acres and is located in the east-central part of the State (Fig. 1), which falls in the U.S. Corn Belt region. Corn and soybeans are the predominant crops in this study area, which is also the case for the majority of the U.S. Corn Belt. According to the 2012 Census of Agriculture for Champaign County (USDA, 2012), there were a total of 616,493 acres of farmland, within which corn and soybeans account for about 92% of the area. Champaign County has a humid continental climate, typical of the Midwestern United States, with hot summers and cold, moderately snowy winters. As a result, double cropping is not usually practiced here – a single-season cropping system is adopted. The individual farm fields identified by the CLU data do not always grow one type of crop. In fact, many CLUs hold a mixture of both corn and soybeans because a CLU is often subdivided into sub-fields that can be planted with different crops. Thus, we need to segment the mixed CLU farmlands into pure corn and pure soybean fields with information provided by the CDL before field-level information aggregation for crop-type classification can be conducted.

Table 1
Summary of used Landsat data.

	Landsat 5	Landsat 7	Landsat 8	Total
2000	41	45	0	86
2001	40	44	0	84
2002	34	43	0	77
2003	46	40	0	86
2004	45	44	0	89
2005	42	46	0	88
2006	43	45	0	88
2007	34	45	0	79
2008	36	46	0	82
2009	40	46	0	86
2010	44	44	0	88
2011	40	46	0	86
2012	0	45	0	45
2013	0	45	31	76
2014	0	46	46	92
2015	0	45	45	90
Total	485	715	122	1322

2.2. Data

The study area is fully covered by Landsat Path 23 and Row 32, and partially covered by the data of Path 22 and Row 32 in Worldwide Reference System-2 (WRS-2). The Landsat Surface Reflectance Data (LSRD) with 30 m spatial resolution was downloaded from USGS's EarthExplorer web portal (http://earthexplorer.usgs.gov) covering Landsat 5, 7 and 8 for 2000 to 2015 (Table 1). More specifically, the LSRD of Landsat 5 and 7 is generated from the Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) (Schmidt et al., 2013), while the LSRD of Landsat 8 is generated from the Landsat Surface Reflectance Code (LaSRC) (USGS, 2016). There are a total of 1322 LSRD scenes with detailed summary shown in Table 1. We used six spectral bands of each scene for our classification algorithm, so there were $1322 * 6 = 7932$ images in total used. The six spectral bands are Blue, Green, Red, NIR (Near Infrared), SWIR-1 (Shortwave Infrared) and SWIR-2 bands respectively. The specificity of the six spectral bands from Landsat 5, 7 and 8 is shown in Table 2. Though the spectral ranges of the corresponding bands have slight differences between Landsat 5/7 and Landsat 8, we find these differences, which have been well studied in previous work (Flood, 2014; Li et al., 2013), are smaller than one standard deviation of time-series spectral curves shown in Fig. 5, which means that these differences will not have a significant impact on our

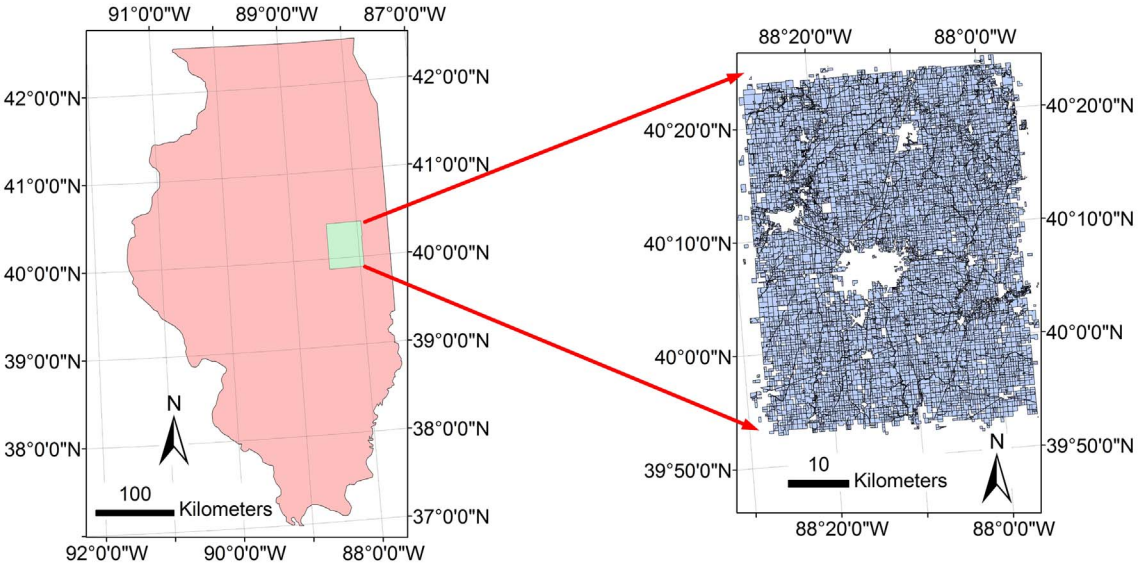

Fig. 1. Geography of the study area (the right panel shows the Common Land Unit (CLU).

Table 2
Detailed information of the six spectral bands from Landsat 5, 7 and 8.

Landsat 5 and 7			Landsat 8		
Band #	Name	Wavelength (μm)	Band #	Name	Wavelength (μm)
Band 1	Blue	0.441–0.514	Band 2	Blue	0.452–0.512
Band 2	Green	0.519–0.601	Band 3	Green	0.533–0.590
Band 3	Red	0.631–0.692	Band 4	Red	0.636–0.673
Band 4	NIR	0.772–0.898	Band 5	NIR	0.851–0.879
Band 5	SWIR-1	1.547–1.749	Band 6	SWIR-1	1.566–1.651
Band 7	SWIR-2	2.064–2.345	Band 7	SWIR-2	2.107–2.294

classification results.

Besides directly using the spectral bands in our classification, we also tested four commonly used vegetation indices (VIs) calculated from the Landsat multispectral data. The four VIs are the Normalized Difference Vegetation Index (NDVI) (Tucker, 1979), Green Chlorophyll Vegetation Index (GCVI) (Gitelson et al., 2003), Enhanced Vegetation Index (EVI) (Huete et al., 2002) and Land Surface Water Index (LSWI) (Xiao et al., 2002), with their formulas shown as follows:

i) $NDVI = (NIR - RED)/(NIR + RED)$
ii) $GCVI = NIR/GREEN - 1$
iii) $EVI = G \times (NIR - RED)/(NIR + C1 \times RED - C2 \times BLUE + L)$
iv) $LSWI = (NIR - SWIR1)/(NIR + SWIR1)$

NDVI is based on the fact that healthy plants usually have a higher reflectance in near infrared (NIR) than visible bands, and NDVI has been widely interpreted as an indicator of photosynthetic capacity (Sellers et al., 1992). However, NDVI can saturate at high leaf area biomass, and the development of GCVI and EVI is largely aimed to reduce this effect (Gitelson et al., 2003; Huete et al., 2002). GCVI has been found to have the most linear relationship with leaf area index (LAI) for corn and soybean compared with other VIs (Gitelson et al., 2003). EVI is designed to reduce the influence of some atmospheric effects by including the blue bands in the VI calculation (Huete et al., 2002). LSWI (Xiao et al., 2002) is developed to approximate canopy water thickness, based on the rationale that the shortwave infrared (SWIR) band is sensitive to leaf water and soil moisture.

The USDA's Cropland Data Layer (CDL) was used as "ground truth" data, which is a raster-formatted, geo-referenced, crop-specific land cover map with the spatial resolution of 30 m/56 m, depending on the sensor used to create the data. The CDL for Champaign County from 2000 to 2015 was obtained from the CropScape website portal (https://nassgeodata.gmu.edu/CropScape/). CDL has the spatial resolution of 30 m derived from Landsat data for most of the years except for 2006–2009, which was derived from the Indian Remote Sensing Satellite (IRS) RESOURCESAT-1 Advanced Wide Field Sensor (AWiFS) with the spatial resolution of 56 m. Since Landsat data is 30 m resolution, we unified all the CDL data (56 m resolution for 2006–2009, and 30 m resolution after 2009 and before 2006) to be 30 m using nearest neighbor interpolation. Although the CDL is not the absolute ground truth, it represents a viable validation data set with a thematic Overall Accuracy greater than 95% (Boryan et al., 2011) for corn and soybean in the U.S. Corn Belt, and it is also the crop-type classification product with the highest accuracy that can be found. As a result, CDL was used as ground truth data for training and testing our crop classification model. Fig. 2(a) presents an example illustrating the corn and soybean distribution from 2015 CDL in Champaign County.

In this study, the USDA's 2008 CLU map was used to identify field boundaries, shown in the right side of Fig. 1. The CLU is an individual contiguous farming parcel, which is the smallest unit of land that has 1) a permanent, contiguous boundary; 2) common land cover and land management; 3) a common owner, and/or a common producer association according to the definition of USDA. There are 19,683 polygons in the Champaign County CLU dataset, however, not all polygons are corn or soybean fields. After applying CDL data to filter CLU data with only corn or soybeans, the polygon number was reduced into 13,959. As a result, these polygons representing corn or soybean fields serve as masks for clipping either LSRD or CDL into field-level segments. The excluded 5724 polygons covered 57,556 acres, which comprises about 9% of Champaign County. These polygons mainly represent non crop types like grass/pasture, forest, impervious surface, open water, undefined, wetland, alfalfa and other crop types such as winter wheat and other small grains. These regions were removed from the classification. Fig. 2(b) shows the CLU aggregated using the information of CDL at the field-level with the crop-type labels generated from the CDL.

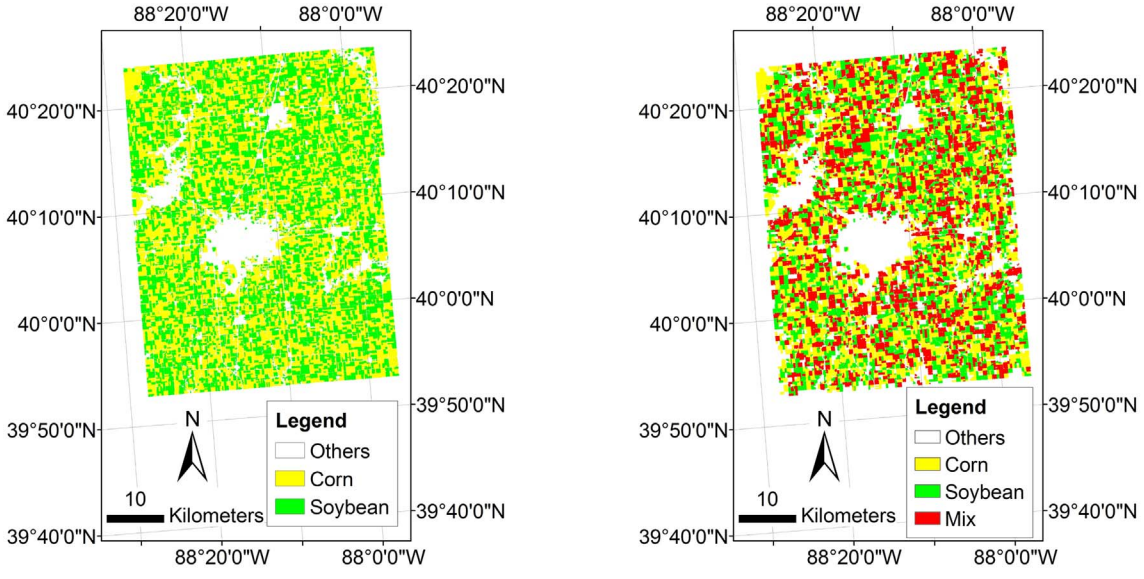

Fig. 2. Example of the 2015 CDL of Champaign County. (a) Raw CDL; (b) the aggregated field-level CDL, where the CLU is used to provide the field-level boundaries.

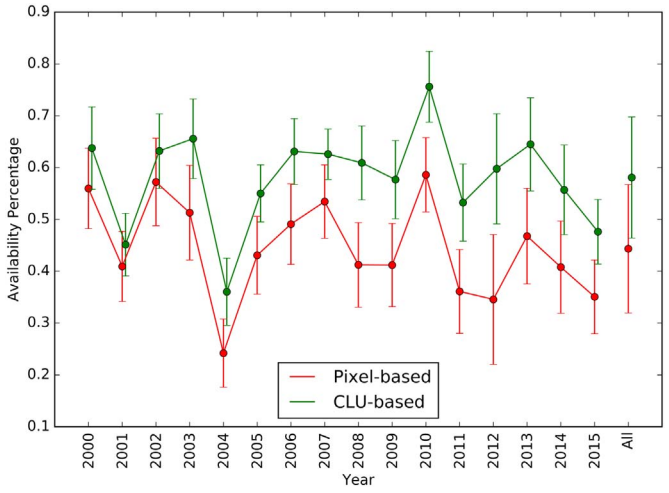

Fig. 3. Data availability improvement: using CLU-based vs. pixel-based information.

2.3. Data preprocessing

We first generated field-level time-series data using the Landsat reflectance data and the CLU for each of the six spectral bands (Fig. S2). For each CLU, the mean value was calculated from all Landsat pixels within the CLU boundary to determine the aggregated field-level information. Even though some pixels in a polygon may be missing, there are still uncontaminated pixels that exist within most CLUs that can be used to calculate field-level information. The availability percentage is calculated through the number of valid values divided by the potential maximum number (the number of total images) of valid values for both pixel and CLU. In general, the availability percentage improved by more than 10% for CLU-based information (CLUs with more than 100 pixels are considered, which comprises 88% total area) instead of pixel-based information, shown in Fig. 3. We then interpolated the field-level, aggregated information into a daily time step for the growing season using the Savitzky–Golay algorithm (Jonsson and Eklundh, 2002) which can further fill certain data gaps. Using different smoothing algorithms had little impact on the classification algorithm, and a simple linear interpolation has achieved almost the same performance as the Savitzky-Golay algorithm. For our study area, an annual time-series span from day of year (DOY) 91 to 270 (early April to late September) was selected to represent the growing season in the U.S. Corn Belt. It is worth noting that before the interpolation of the time series for the different spectral bands, data points that were outside of the 95% percentile of the samples conditioned for each time stamp were filtered. These filtered data were treated as missing data and the interpolation algorithm filled in the missing values. After we derived the field-level surface reflectance data, we then extracted the crop types from the CDL data for each CLU field polygon as labels of training and testing datasets.

2.4. DNN-based classification model

A Deep Neural Network (DNN) model was used to build the crop-type classification model. Fig. 4 shows the schematic diagram of the approach, where $x_{i,\ k}$ ($k \in [1, 2, \ldots, n]$) represents the input data, including different spectral bands or VI at a specific DOY or a section of time series over the growing season as $x_{i,\ k}$ for field i. The dimension n changes according to the experiment's design. The activation function used in our classification model was the Rectified Linear Unit (ReLU) activation function with the form as $f(z) = \max(0, z)$. Our DNN model has four layers, including three hidden layers, and one output layer. Hidden layers in the network transfer raw information into meaningful features for classification. We have tested other configurations of the

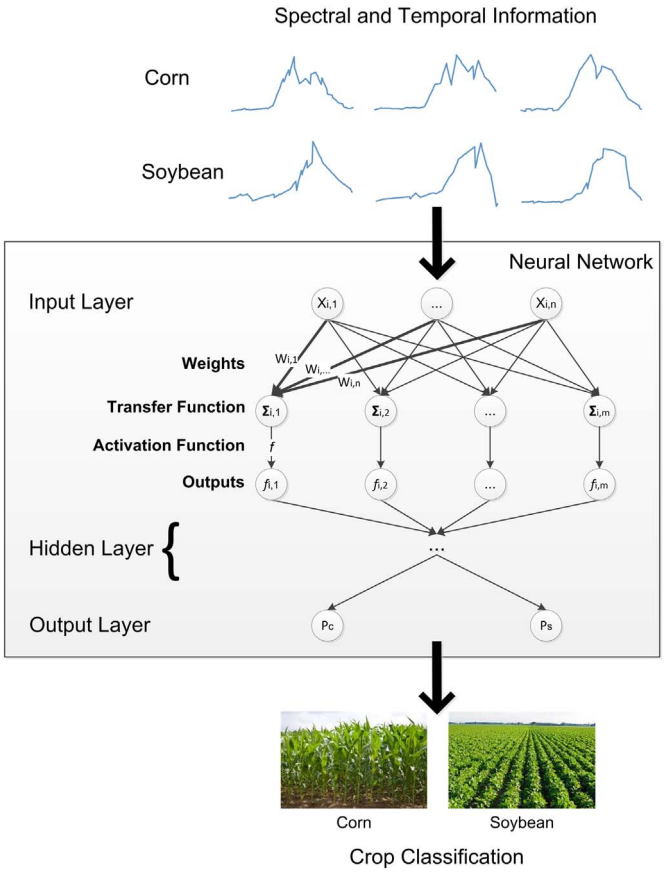

Fig. 4. Framework of the DNN based crop type classification.

DNN model in terms of the hidden layers and different activation functions, and found that the seven hidden layers can achieve the stable and optimal performance in our cases, though the performance difference between three layers and seven layers of the DNN model is less than 1% in terms of Overall Accuracy (see the definition below). Based on the consideration of model parsimony, we use the 3-layer DNN model results in Section 3. We have also tested other machine-learning algorithms, such as the Support Vector Machine, K-means, and Random Forests, and none of them outperforms the DNN approach that is used here (Fig. S3 in the Supplementary material), thus we use the results of the DNN here. In this study, accuracy of the classification is assessed through the Overall Accuracy, which is the number of the corrected classified fields (i.e. the summation from the diagonal in the confusion matrix) divided by the number of the total fields. To achieve Overall Accuracy, we first generated a well-trained model by tuning its hyperparameters through the backpropagation algorithm (Rumelhart et al., 1986) based on the training dataset, and then applied the model on the testing dataset. The detailed strategies to separate dataset into training/testing dataset are addressed in the next section.

2.5. Experiment design

In order to address the overarching questions raised earlier in this paper (i.e. (1) What accuracy can the integrated time-series Landsat data and machine learning approach achieve for the crop-type classification? (2) How early in the growing season can this method achieve the optimal accuracy of crop-type classification using our approach?), we ask sub-questions upon which we designed the experiments. Specifically, we ask the following sub-questions:

Q1. What spectral information is most useful for crop-type classification?

Q1.1. What thematic classification accuracy can be achieved using only the spectral features (at a static time) compared to combining both spectral and temporal phenology information?

Q1.2. What thematic classification accuracy can be achieved using only vegetation indices compared to using the original spectral reflectance?

Q2. How do spatial and temporal sampling strategies affect the accuracy of crop-type classification?

Q2.1. How does the choice of different years of training data affect classification accuracy?

Q2.2. How does the choice of different spatial sampling strategies affect classification accuracy?

To answer the above sub-questions, the following experiments were designed, with the first two groups of experiments addressing Q1, and the third group of experiments for Q2. The **first** group of experiments mainly focuses on the comparison of using information as input for the classification model (Q1). Specifically, for the spectral feature-only approach, the original spectral information from each band at a specific day during the growing season (defined from DOY 91 to 270 with intervals of 5 days) was used in the algorithm. For the spectral and temporal combined approach, data for individual spectral bands were used, but include the time series of observations across the whole growing season (DOY 91 to 270 with intervals of 5 days) as inputs. By comparing these two approaches, the added value of the temporal information in the classification can be quantified, as well as the value of individual spectral bands through the inter-band comparison. We also designed the same experiments as above but substitute spectral bands with different VIs, such that we can answer Q1.2. Additionally, we also investigated whether combining different spectral bands or different VIs could achieve higher classification performance than only using a single band or a single VI. A total of 6 spectral bands and 4 VIs were available for testing and a brute force method was used to explore all possible combinations. For example, for six spectral band combinations, there are total 57 tested combinations ($=C_6^2 + C_6^3 + C_6^4 + C_6^5 + C_6^6$). For experiments in this group, the whole dataset was randomly divided into training data and testing data, which took up 2/3 and 1/3 of the whole dataset respectively. Accuracies were assessed when applying trained models on testing data.

The **second** group of experiments was designed to study the influence of spatial and temporal sampling on crop classification (targeting Q2). The training/testing data were grouped by either different years or by different regions to test the influence of temporal and spatial sampling, respectively. Accuracies were assessed from the independent testing data. To study the influence of different temporal sampling on classification, we either fixed the starting year or fixed the ending year in the experiments (targeting at Q2.1). Specifically, for the experiment that fixes the starting year (2000), data from a different number of continuous years starting at 2000 were combined to predict the crop types in the following year (e.g., all data from 2000 to 2005 was used to predict the crop types in 2006). For the experiment that fixes the ending year (2014), data of a different number of continuous years before 2015 were combined to predict the crop types in 2015 (e.g., all data from 2010 to 2014 is used to predict the crop types in 2015). As a result, the first experiment predicts crop types in different years from 2001 to 2015; however, the second experiment only predicts crop types in 2015. To explore the influence of spatial sampling, Champaign County was evenly divided into 2 regions from north/east to south/west by latitude/longitude (targeting Q2.2). The north/east region and the south/west region were alternately used as the source of training data to train two different models. Then the rest data from the two regions were used as the testing data for comparison purpose to explore spatial factor influence on classification.

The **third** group of experiments was designed to address the second main question raised in the introduction, i.e. to quantify how early during the concurrent growing season the classification algorithm can accurately predict the crop types in the concurrent year. In order to exclude any correlations between training and testing data, data from 2014 and 2015 are selected as testing data, while data from all years before 2014 were used for model training. Accuracies are assessed for independent testing data in both 2014 and 2015. Starting at the DOY 91, more input Landsat data was gradually included in the algorithm to generate the crop classification until DOY 270. This mimics the real-life situation that more data will be ingested into the crop classification procedure as the growing season progresses and the classification performance would be expected to increase as additional dates of remotely sensed data are added to the classifier until optimal classification performance is achieved.

3. Results

3.1. Time-series profile

Time-series profiles of spectral bands and VIs aggregated from corn/soybean fields in Champaign County from 2000 to 2015 are shown in Fig. 5 to illustrate their potential for contributing to crop-type classification. In Fig. 5, the X-axis is the DOY and the Y-axis is the value of reflectance or VI. The red line stands for corn and the green line represents soybeans. The buffers indicate one standard deviation calculated from all fields and years. For the visible spectral bands shown in Fig. 5(a)–(c), there are large overlaps in their seasonal trajectories between corn and soybean, especially near the late growing season (after ~DOY 200). The NIR band shows more differences between corn and soybeans in the later stages of the growing season. The SWIR bands show a clear difference between corn and soybean during the middle of the growing season (~DOY 190–200). During this period, the SWIR curves of corn and soybeans with one standard deviation have no overlap, which indicates that this feature will be especially useful in differentiating between corn and soybean. Similar features of SWIR are also shown in the LSWI, which incorporates the SWIR band in its calculation. NDVI, GCVI and EVI all show more difference between corn and soybean at the early part and the late part of the growing season, but there are significant overlaps between these two crop types in these VIs throughout the growing season.

3.2. Spectral information-based classification

3.2.1. Classification based on a single spectral band

Here we only used a single spectral band or a single VI at any specific DOY in the growing season to train the classification algorithm. We find that different bands show various performances at different DOYs (Fig. 5a). The visible bands (blue, red and green bands) have a similar performance. The classification accuracy for these visible bands remains relatively low (0.54–0.61) between DOY 90 and 140, followed by a large increase in accuracy (0.56–0.73) between DOY 140 and 180 with a peak accuracy (0.73) around DOY 175. Their performance then drops with the green and red bands having another local optimum around DOY 220–240, but less so for the blue band. NIR shows a similar performance as the visible bands during the early growing season and reaches its peak performance at a relatively later time over a longer period of time (DOY 210–240). The SWIR-1 and SWIR-2 data reaches their highest performance (about 0.85 accuracy) around DOY 195, which is much higher than the performance of any other bands at any DOYs. In addition, SWIR-1 and SWIR-2 attain their peak performance only once, which happen at a different time compared with the visible and NIR bands. Finally, the accuracy performance of SWIR-1 and SWIR-2 are very similar to each other, with SWIR-1 performing slightly better than SWIR-2.

3.2.2. Classification based on a single vegetation index

We conducted the same experiments as above (Section 3.2.1) but for

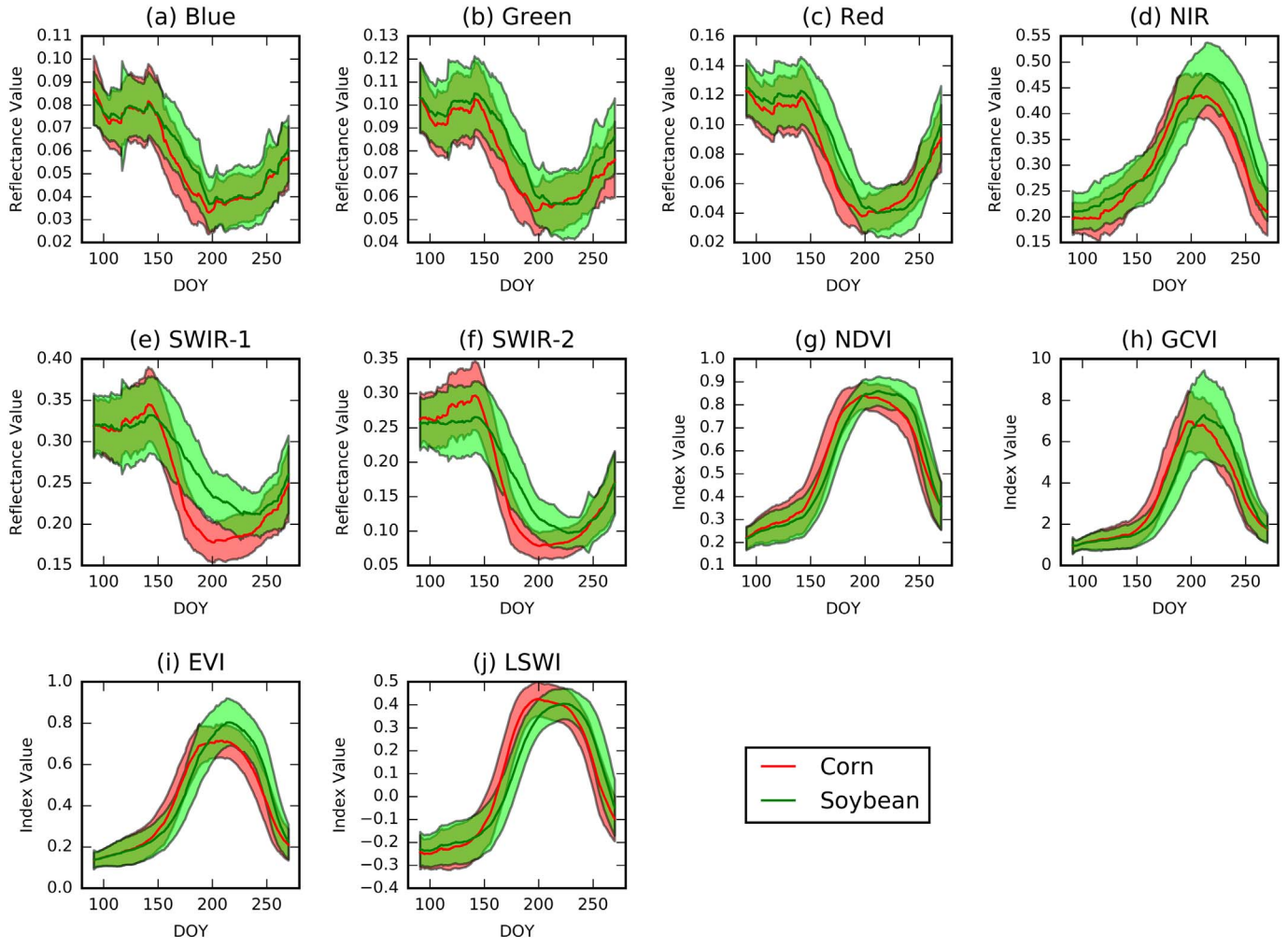

Fig. 5. Time-series spectral band information and vegetation indices are aggregated for all the corn and soybean fields and years for the Champaign County, IL. The red line stands for corn, and the green line represents soybeans. The buffers indicate one standard deviation calculated from all fields and years. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the four VIs (i.e. EVI, GCVI, NDVI, and LSWI) to study their individual performance at any specific DOY, with results shown in Fig. 6(b). EVI, GCVI, and NDVI are found to have a similar pattern in classification performance, which is very different from LSWI. Specifically, the accuracy performances of EVI, GCVI, and NDVI are all low before DOY 140, followed by the first peak in classification performance (~0.70) around DOY 175 and then decline to the local minimum around DOY 195. The three VIs then reach a second peak in classification accuracy with the optimal performance (~0.77) occurring around DOY 235. In contrast, the LSWI only has one peak performance that occurs around DOY 190 with a peak classification performance (~0.87) much higher than the other VIs.

The performances of VIs are consistent with those of individual spectral bands that were used to calculate VIs, which is largely expected. In particular, we find that EVI, GCVI, and NDVI, which all use NIR as a major input, had a similar performance as that of NIR. Since LSWI is calculated based on the two SWIR bands, LSWI's performance is also very similar to those of the two SWIR bands.

3.2.3. Classification using a combination of multiple spectral bands

Intuitively, combining more information (either from more spectral bands or from more VIs) in our model would be expected to improve the overall classification accuracy based on results in Fig. 6 (visual bands have peak performance at the early stage, SWIR bands have peak performance at the middle stage, and NIR has peak performance at the late stage). This hypothesis was explicitly tested here with results presented in Fig. 7. The term *cmbx* (*combination x* = 2, 3, 4, 5, 6) in Fig. 7 stands for x number of bands that are combined for classification experiment. For example, *cmb2* means any two spectral bands are used for classification, thus there exist $C_6^2 = 15$ different combinations and each of them is tested here and only the best one will be recorded. The result in Fig. 7 illustrates that all combinations share a similar classification accuracy pattern across the growing season and the combination of more spectral bands lead to a higher accuracy. The peak accuracy appears around DOY195, which is consistent with the results of single SWIR-1/SWIR-2 band-based classification, indicating that these two bands dominate the classification accuracy around that specific time period. In addition, another peak accuracy appears around DOY 240 due to the contribution of the NIR region. Accuracy increases significantly between DOY 140 and 175, indicating some significant features of corn and soybeans are captured during their early vegetative stages. In summary, combining spectral bands improves the performance at each stage during the growing season.

3.2.4. Classification using a combination of multiple VIs

The same analysis as Section 3.2.3 was also conducted for different combinations of VIs at specific DOYs with the results shown in Fig. 8. For the four VIs, a total 11 combinations were tested, and the results show a similar pattern to those from the combination of the spectral bands. From DOY 170 to 240, a relatively high classification accuracy is

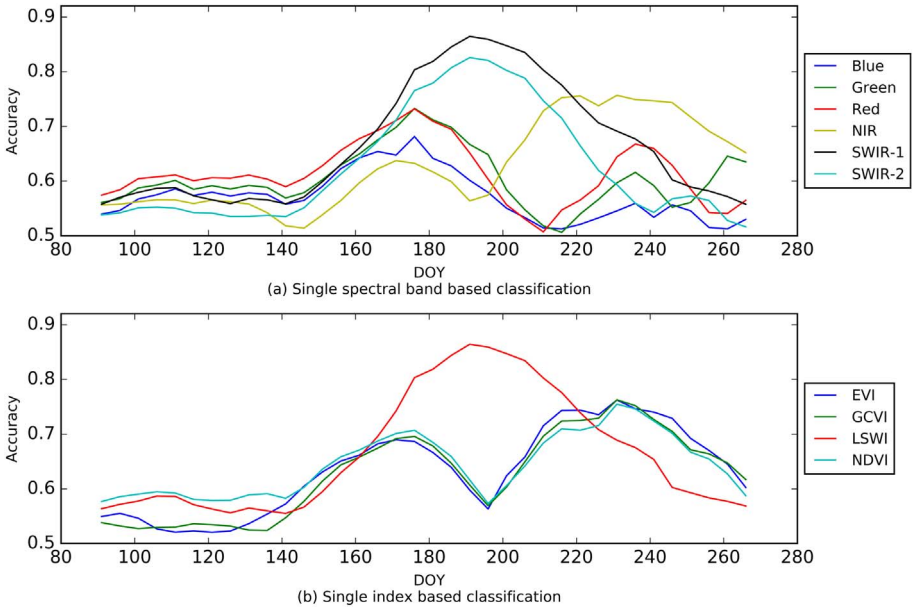

Fig. 6. Classification performance of using single spectral band (a) or a single VI (b) at any specific DOY.

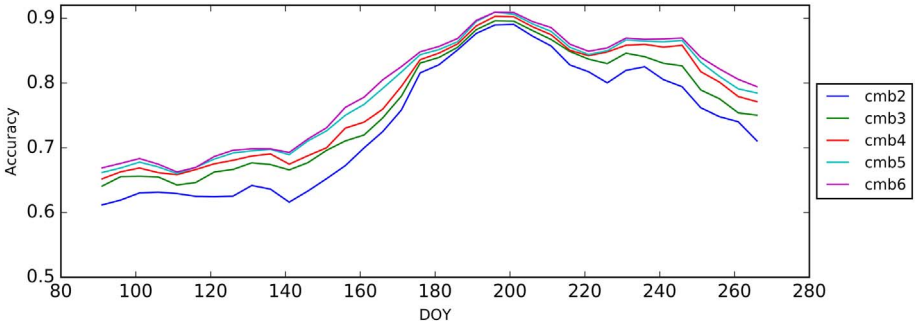

Fig. 7. Classification performance of using a combination of spectral bands at any specific DOY. The term cmbx stands for x number of spectral bands that are combined for classification experiment.

achieved, while there is a steep increase and decrease in accuracy before and after that time period, respectively. In sum, combining VIs improves performance at each stage during the growing season with little difference in classification accuracy observed among the different combinations of VIs.

3.3. Classification using both spectral and temporal information

While in Section 3.2 spectral bands or VIs either in individually or in a combination were evaluated, but only at a specific DOY (i.e., single date). In this section, the spectral information is combined with the temporal information for the entire growing season, i.e. time series of the spectral bands or VIs are used here for crop classification.

3.3.1. Classification based on a single spectral band/VI

Fig. 9 shows the classification results of a single band or VI using all growing season data in the classification model. The accuracies of individual spectral bands have greater accuracy variations (from 0.87 to 0.93) than the results obtained using different VIs, which shows a more consistent classification accuracy pattern (the accuracy difference is within 0.01). The accuracy of VIs does not significantly exceed the accuracy of individual spectral bands. In general, the VIs have higher accuracy than all the visible bands, but VIs have a similar accuracy as infrared band (i.e. NIR and SWIR-1) results.

3.3.2. Classification using a combination of multiple spectral bands/VIs

Like the previous experiments using various combinations of data,

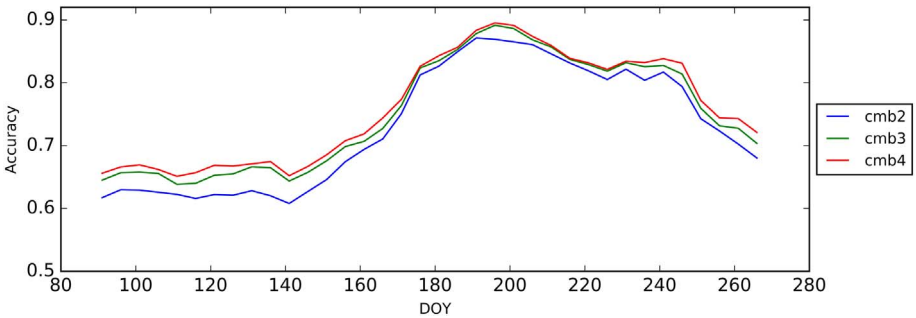

Fig. 8. Classification performance of using a combination of VIs at any specific DOY. The term cmbx stands for x number of VIs that are combined for classification experiment.

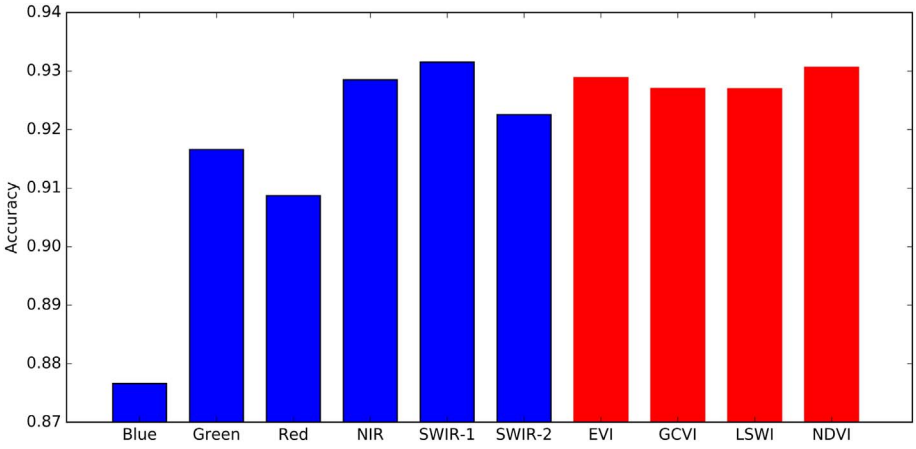

Fig. 9. Classification performance of a single band or a single VI, when incorporating the temporal phenology information in the classification.

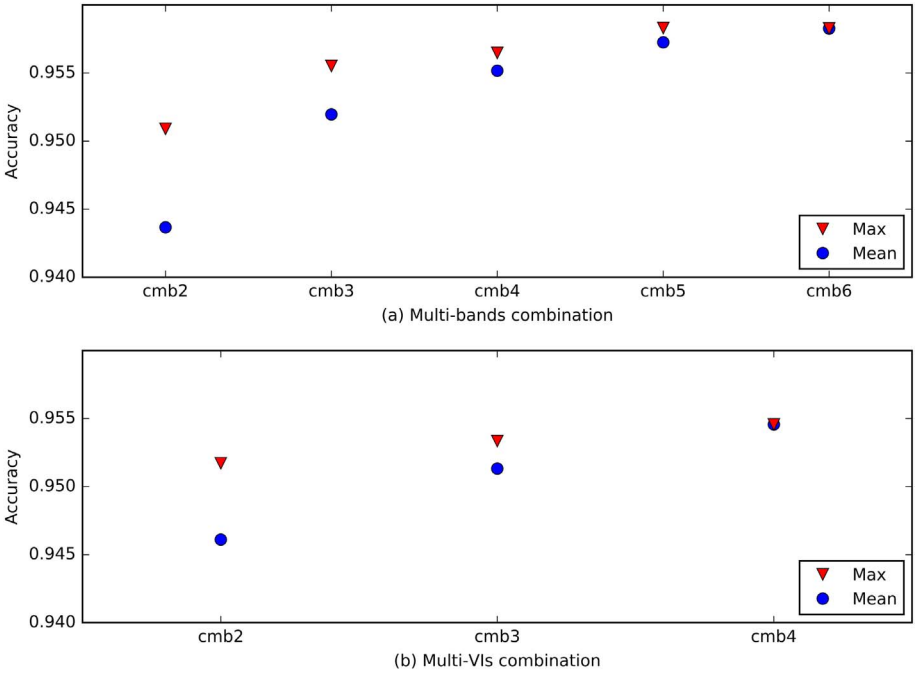

Fig. 10. Classification performance of using a combination of spectral bands (a) or VIs (b) with the phenology information, and both mean and maximum performance of different combination are shown. The term cmbx stands for x number of spectral bands (or VIs) that are combined for classification experiment.

the same combinations were tested using the entire time series of individual spectral bands or VIs data across the growing season, as shown in Fig. 10. The incorporation of more bands or VIs leads to improved classification performance, but this improvement is marginal (only within 0.02 between the lowest and highest accuracy). Table 3 shows the specific spectral bands used in combination with the highest classification performance. Green and SWIR-1 appear in all the best performing combinations, which is identical to the combinations in spectral band-based classifications.

Comparing results in Section 3.2 (only using the data from a specific DOY) and in Section 3.3 (using the whole growing season time series), we can increase accuracy by 5% when using the temporal information; thereby demonstrating the added value of using temporal information for the classification results.

Table 3
The best combinations for different number combinations.

	cmb2	cmb3	cmb4	cmb5	cmb6
Combination	Green, SWIR-1	Green, SWIR-1, SWIR-2	Green, Red, NIR, SWIR-1	All except Blue	All

3.4. Impact of temporal and spatial sampling on classification

3.4.1. Impact of temporal sampling for classification

Considering real world applications, a more useful way to conduct classification is to use only previous continuous years' data to train the model, and to use the current year's data for testing, such that we have a predictive and forecasting capability for the current year's crop types. To explore this scenario, two experiments are designed using combined spectral bands with temporal information. One has a fixed starting year (2000) and different numbers of the continuous years' data are used to train and predict crop types in the following year. For example, all data from 2000 to 2005 is used to predict crop types in 2006. Fig. 11(a) shows a generally increasing trend in performance (except for some outliers), indicating that including more years of data for classification training can achieve a higher performance for classifying crop types for

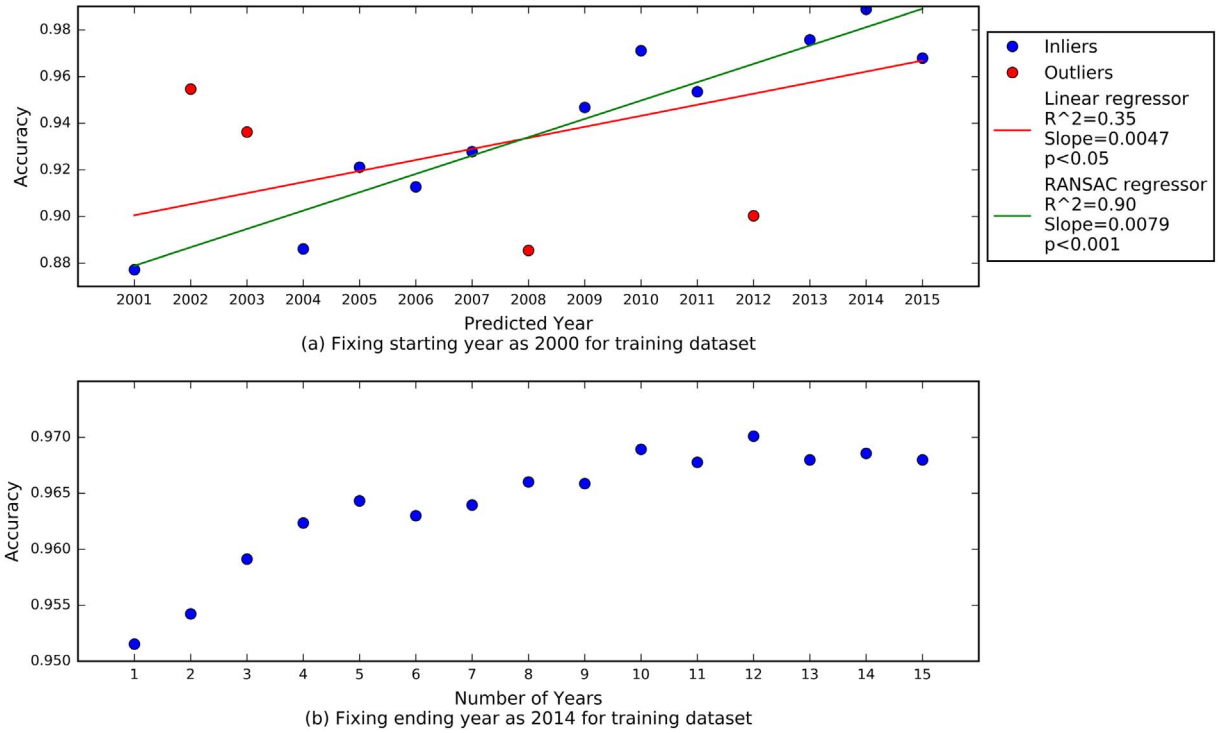

Fig. 11. Impacts of using different temporal sampling for the final classification results. (a) Using the data of different numbers of continuous years starting in 2000 to predict crop types in the following year (for example, we use all the data from 2000 to 2005 to predict the crop types in 2006). To reveal the trend, we conducted linear regression and also applied RANdom SAmple Consensus (RANSAC) algorithm to linear regression, shown in red and blue lines respectively. The outliers are detected by RANSAC algorithm. (b) Using data of different numbers of continuous years ending in 2014 to predict crops type in 2015 (for example, we use all the data from 2010 to 2014 to predict the crop types in 2015). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the next year. The outliers occur in 2002, 2003, 2008 and 2012. Since SLC-off occurred after May 31, 2003, that may explain why performance in 2002 is high but slightly decreases in 2003. Additionally, if we compare results in Fig. 11(a) with Fig. 3, we can find low data availability will lead to poor performance, which can further explain low accuracies in 2004 and 2008. 2012 is a drought year, and the temporal phenology information of crop is different from other years. As a result, the model trained with data only in normal years performs poorly for this exceptional drought year. The other experiment uses a fixed ending year (2014), and uses different numbers of previous continuous years' combination data to train and predict crop types in 2015 (e.g., all data from 2010 to 2014 is used to predict crop types in 2015), with the results shown in Fig. 11(b). For the second experiment, if we use more years of data to only predict crop types in 2015, a clear increase in performance occurs when more years of data are included (Fig. 11(b)) with the performance increase plateauing after around 10 years.

3.4.2. Impact of spatial sampling for classification

This experiment aims to test the impact of different spatial sampling schemes on crop-type classification accuracy. In Fig. 12, format "*X-Y*" refers to the scenario that uses data within X region as the training data and tests the model output using data within Y region. Specifically, Champaign County is evenly divided into 2 regions from north to east and from south to west evenly by latitude/longitude, respectively, with the north/east area marked as "N"/"E" and the south/west area marked as "S"/"W". For each region, we divide the data into training and testing data with no overlaps. Fig. 12 shows that N-N and S-S (i.e. training and testing samples are from the same regions) have higher performance than N-S and S-N (i.e. training and testing samples are from different regions) respectively, similar to E-E, W-W, E-W, and W-E. In addition, models with combined regions (NS or EW) as training data perform slightly better than that training and testing samples from the same regions. These results indicate that spatial heterogeneities exist, but benefits from building different specific models for sub-regions are not obvious since Champaign County is a small region and spatial heterogeneities are not significantly large within the county. As a result, one model for the whole region is appropriate. We suggest that future applications of our method for a larger area (e.g. the whole state of Illinois, or many states) should consider the spatial heterogeneity and possibly developing region-specific training algorithms for different regions.

3.5. In-season forecasting: pushing to early stage

For any given year, the earlier a relatively high accuracy in classification can be achieved, the more valuable these results are for decision-making activities. Here, an experiment is developed to test how early during the growing season and acceptable level of classification accuracy can be achieved. To achieve this goal, a fixed starting date is set at DOY 91 and the ending date is varied from DOY 96 to 270. For each ending date, the time series of historical data used to train the classification model spans from DOY 91 to that specific end date. The classification model uses the temporal phenology information of six spectral bands combined, which represents the best performance across different combinations (Fig. 10(a)). The trained classification model is then applied to data in the next year for the test (e.g. The top panel of Fig. 13 shows the results of using the data from 2000 to 2013, and testing using the 2014 data; the bottom panel of Fig. 13 shows the results of using the data from 2000 to 2014, and testing using the 2015 data). The blue line in Fig. 13 shows the classification accuracy at different ending dates, and the red line shows the rate of accuracy change. We smoothed the accuracy curve in Fig. 13 by applying a moving average with the window size of 3.

For both cases in Fig. 13, we find that the classification accuracy increases with the progression of time (i.e. more inputs from satellite data), and the accuracy reaches a plateau around DOY 200 (i.e. middle

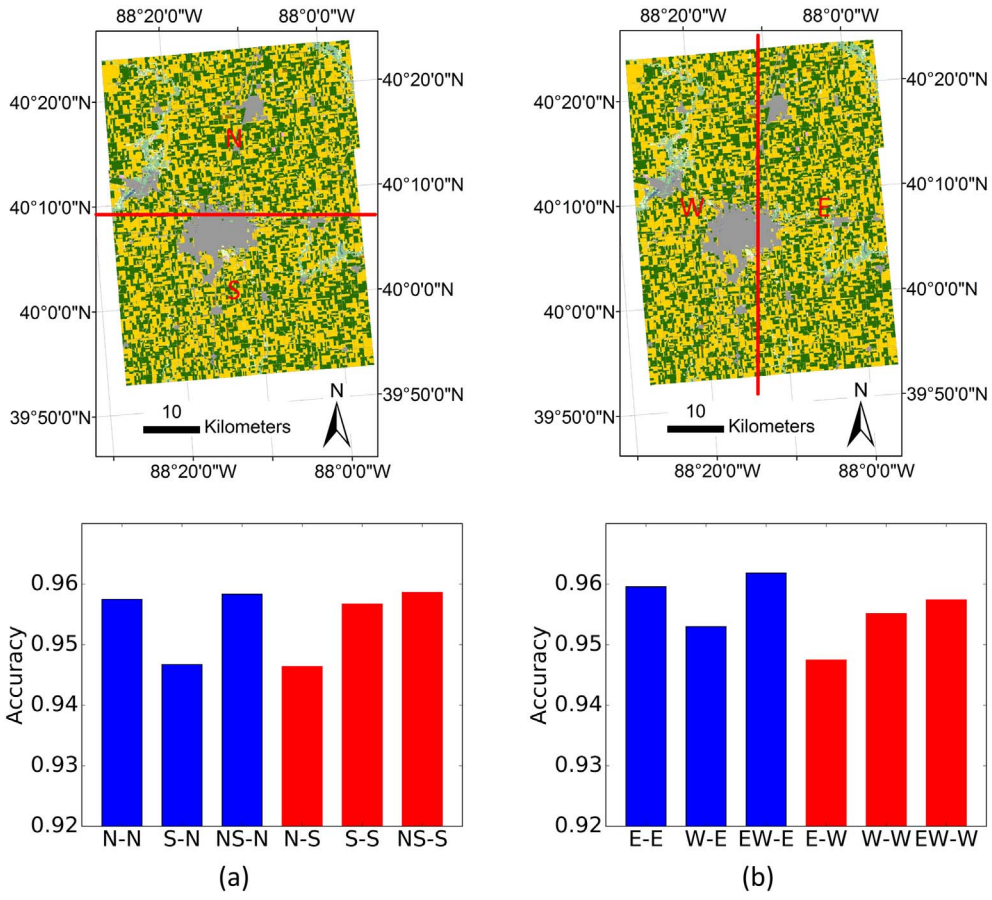

Fig. 12. Impacts of using different spatial sampling for the final classification results. (a) shows north-to-south division strategy and its performance, while (b) shows east-to-west division strategy and its performance. Labels in histogram represent where the training and testing samples come from, e.g. N-N means using the North-region samples to train and North-region samples to test; EW-W means using both the East-region and West-region samples to train and West-region samples to test.

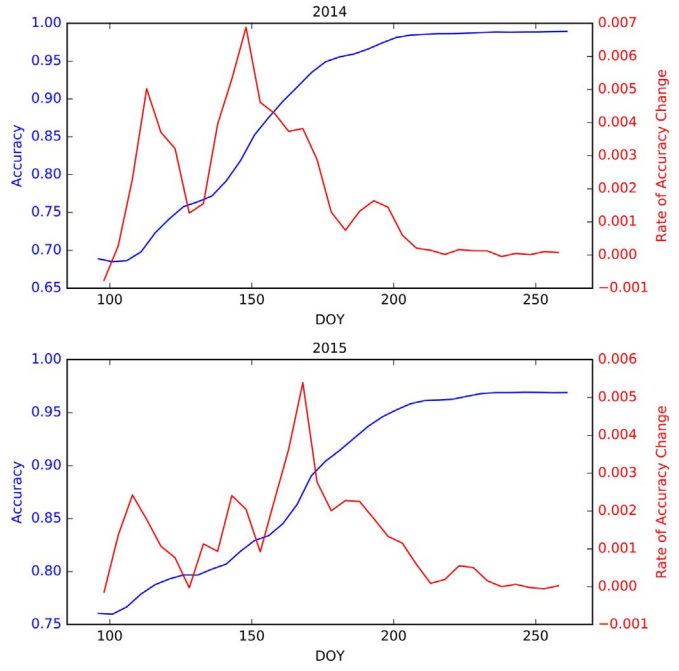

Fig. 13. Classification performance as a function of time (blue line). For example, the DOY 125 means that all the available data from the starting day (DOY 91) to DOY 125 from the historical years are used to train the model and the current year's satellite data from DOY 91–125 are used to predict the current year's crop types. The red line refers to the rate of change in the blue line, which indicates the rate of performance change with more data ingested for the classification. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

of July), which is confirmed by a commensurate decrease in the change in accuracy rate (i.e. the red line approaches zero at and after ~DOY 200). We also find that the largest positive changing accuracy rate change occurs between DOY 140 and 160 (late May and early June) for both years. This is expected, as late May and early June is the early vegetative growth period for both corn and soybean, and their different sowing dates and different canopy development patterns should lead to the most obvious spectral-temporal differences between them. For both test years, the classification model reached 95% classification accuracy by around DOY 180 (end of June).

4. Discussion

The discussion is organized to first summarize the answers to the questions that are posed at the end of Section 1 and in Section 2.5, and then addresses some limitations of this study. The first research question is related to the best classification accuracy that our approach can achieve, and explored what information is most useful to achieve the highest accuracy performance. We answer this question from the following five aspects.

(1) The results demonstrate that combining the CLU and Landsat image data provides an effective solution to generate time-series data that can be used for the subsequent crop-type classification. The data availability is improved by using CLU-based information instead of pixel-based information. In Fig. 3, the improvement increases after 2003 since the Landsat 7 began to have SLC-off, which indicates our method is efficient to handle the SLC-off effect.

(2) The SWIR bands, which have not been commonly used in many previous crop-type classification efforts, are found to have notably better ability to distinguish between corn and soybeans than visible and NIR bands (Figs. 6 & 9). SWIR bands are related to the crop water content and have been used to detect crop water stresses (Ghulam et al.,

2008; Xiao et al., 2006). A recent study also found the potential use of SWIR in crop yield estimation (You et al., 2017). SWIR also contributes to the higher performance of LSWI than other VIs evaluated in this study. One notable feature of the SWIR bands is that they achieve the highest classification performance in classification near the middle of the growing season, which is a period during the growing season that the NIR has been largely saturated (Fig. 6). We thus infer that the largest difference between corn and soybeans in terms of canopy water contents occurs during the peak-growing season.

(3) The results confirm that the inclusion of temporal information through time-series data inputs improves classification accuracy compared to using single-date data (i.e. snapshots). Using temporal and spectral information together achieves 10–15% higher classification accuracy than only using spectral information at one specific date.

(4) The results also confirm that combining all spectral bands leads to the highest classification performance. However, if only a subset of the Landsat spectral bands are to be used, the green band in combination with one of the two SWIR bands can provide the highest classification performance among other combinations. In addition, VIs do not perform significantly better than original spectral data (especially the SWIR bands), and the performance among different VIs are similar.

(5) We assess the impact of different spatial and temporal sampling strategies on crop classification performance. We find that increasing the number of years in training data and collecting samples more evenly across the spatial domain of the study area usually lead to higher classification accuracy. Though we also find that the classification performance stabilized after about 10 years of training data is used, and has little gain in thematic accuracy when additional years of data are added.

The second research question focuses on the ability of our approach to make the in-season classification. The model with the highest classification performance (i.e., using all spectral bands and temporal information) trained with all the previous year's data has attained high classification accuracy when applied to the next year (Fig. 13), with the final accuracy reaching 96%. For within-season classification, our best model to classify corn and soybeans reached 95% classification accuracy by DOY ~180 (end of June). This result means that corn and soybeans should be able to be classified at a relatively high thematic accuracy by early to mid-growing season using satellite data collected for all prior dates acquired earlier in the growing season. The latter condition can be fulfilled now, as the newly collected Landsat 7 and 8 data become available for processing within one week of data acquisition (http://landsat.usgs.gov/CDR_LSR.php). As Sentinel-2 data becomes available, an additional source of near-real-time data will further improve the satellite data availability and thus the classification accuracy. Finally, the USDA CDL layer is used here as the benchmark data set for locating historical corn and soybean fields across the study area. It is worth noting that the major motivation of this work is not to substitute the CDL, but rather to provide an in-season classification approach/product that has the similar performance of CDL for the U.S. Corn Belt, but can provide near-real-time, in-season crop maps for decision-making activities. The results here demonstrate that our proposed methodology has substantial potential to meet this goal.

It is worth noting the following limitations and uncertainties of this study, some of which represent future directions to further improve this classification method. One major limitation is that all land covers that have no corn or soybean are pre-filtered; in other words, we only focus on classifying a binary system (corn vs. soybean). However, as previous studies (Wardlow et al., 2007; Wardlow and Egbert, 2008) pointed out for the broader U.S. Corn Belt, differentiating corn and soybean remains the biggest challenges as other types of crops (e.g. winter wheat, sorghum) or natural vegetation (e.g. grass and trees) usually have a very different temporal phenology than corn and soybean. Furthermore, only focusing on corn and soybean is justified for the targeted study area, as they are the dominant crop types in the main states of the U.S. Corn Belt. However, future research efforts should be devoted to extending this approach and including other types of crop types and natural vegetation types. There are also expected uncertainties in the classification approach, and reducing these uncertainties is another future research direction. First, when a time series stack from Landsat data is generated for each field, interpolation is used to fill gaps in the Landsat data. We have tested a series of other interpolation methods and found little influence on the final classification performance. With more satellite data becoming available, such as Sentinel-2 data, the gap filling performance can be further improved. Another possible approach is to use multi-sensor fusion data, such as the STARFM algorithm developed by (Gao et al., 2006) to integrate MODIS (low spatial, but high temporal resolution) and Landsat (low temporal, but high spatial resolution), for the classification purpose (Gao et al., 2017). Second, we use CLU to aggregate the field or sub-field level information (see details in Section 2.3). The major motivation of using CLU is to aggregate the spectral information from many pixels in order to largely avoid the cloud cover issue. However, it should be recognized that within a CLU there can be a mixture of crops and this mixture may also change over time, and that our current approach assumes that the division of different fields is known.

5. Conclusion

This study has demonstrated that combining time-series Landsat Surface Reflectance Data and CLUs with a machine learning approach provides a cost-effective and high-performance option for field-level and in-season crop-type classification for corn/soybean dominated Corn Belt landscape, with a detailed case study in Champaign County, Illinois. This study is important for both scientific and practical uses. The classification model based on DNN was applied to distinguish corn and soybean patches for each CLU field, which was trained and tested using the CDL as the ground truth. Systematic experiments were conducted to determine which information is most useful for classifying corn and soybeans. Overall, a high Overall Accuracy of classification (~97%) is achieved using this method. In addition, the ability to perform within-season crop-type classification can be achieved at a relatively early stage of the growing season at about DOY 210 (late July or early August), with equivalent accuracy (~95%) to classification results at the end of year based on an entire growing season of data inputs. All data sources (LSRD, CLU, and CDL) used in this study are publicly available. The current approach has a great potential to be scaled up to other counties, states, and possibly the whole U.S. Corn Belt. Further improvements can be developed, such as using texture features extracted from high temporal-spatial resolution fusion data (e.g. fusion data MODIS, Landsat).

Acknowledgements

Kaiyu Guan acknowledges the support by the Blue Waters Professorship awarded by National Center for Supercomputing Applications of University of Illinois at Urbana Champaign, NASA New Investigator Award (NNX16AI56G), and NASA Carbon Monitoring System Award (80NSSC18K0170). Yaping Cai and Shaowen Wang acknowledge the support of an Illinois Distinguished Graduate Fellowship. We acknowledge the use of the Blue Waters and ROGER supercomputers at the University of Illinois at Urbana Champaign. Blue Waters is supported by the National Science Foundation under award numbers 0725070 and 1238993, and the state of Illinois while ROGER is supported by the National Science Foundation under grant number 1429699.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.rse.2018.02.045.

References

Bolton, D.K., Friedl, M.A., 2013. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. Agric. For. Meteorol. 173, 74–84. http://dx.doi.org/10.1016/j.agrformet.2013.01.007.

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 (5), 341–358. http://dx.doi.org/10.1080/10106049.2011.562309.

Chang, J., Hansen, M.C., Pittman, K., Carroll, M., DiMiceli, C., 2007. Corn and soybean mapping in the United States using MODIS time-series data sets. Agron. J. 99 (6), 1654. http://dx.doi.org/10.2134/agronj2007.0170.

Collobert, R., Weston, J., 2008. A unified architecture for natural language processing. In: Proceedings of the 25th International Conference on Machine Learning - ICML '08. ACM Press, New York, New York, USA, pp. 160–167. http://dx.doi.org/10.1145/1390156.1390177.

Duro, D.C., Franklin, S.E., Dubé, M.G., 2012. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. Remote Sens. Environ. 118, 259–272. http://dx.doi.org/10.1016/j.rse.2011.11.020.

Flood, N., 2014. Continuity of reflectance data between Landsat-7 ETM+ and Landsat-8 OLI, for both top-of-atmosphere and surface reflectance: a study in the Australian landscape. Remote Sens. 6, 7952–7970. http://dx.doi.org/10.3390/rs6097952.

Foerster, S., Kaden, K., Foerster, M., Itzerott, S., 2012. Crop type mapping using spectral–temporal profiles and phenological information. Comput. Electron. Agric. 89, 30–40. http://dx.doi.org/10.1016/j.compag.2012.07.015.

Gao, F., Masek, J., Schwaller, M., Hall, F., 2006. On the blending of the Landsat and MODIS surface reflectance: predicting daily Landsat surface reflectance. IEEE Trans. Geosci. Remote Sens. 44 (8), 2207–2218. http://dx.doi.org/10.1109/TGRS.2006.872081.

Gao, F., Wang, P., Masek, J., 2013. Integrating remote sensing data from multiple optical sensors for ecological and crop condition monitoring. In: Gao, W., Jackson, T.J., Wang, J., Chang, N.-B. (Eds.), International Society for Optics and Photonics, pp. 886903. http://dx.doi.org/10.1117/12.2023417.

Gao, F., Hilker, T., Zhu, X., Anderson, M., Masek, J., Wang, P., Yang, Y., 2015. Fusing Landsat and MODIS data for vegetation monitoring. IEEE Geosci. Remote Sens. Mag. 3 (3), 47–60. http://dx.doi.org/10.1109/MGRS.2015.2434351.

Gao, F., Anderson, M.C., Zhang, X., Yang, Z., Alfieri, J.G., Kustas, W.P., Mueller, R., Johnson, D.M., Prueger, J.H., 2017. Toward mapping crop progress at field scales through fusion of Landsat and MODIS imagery. Remote Sens. Environ. 188, 9–25. http://dx.doi.org/10.1016/j.rse.2016.11.004.

Ghulam, A., Li, Z.-L., Qin, Q., Yimit, H., Wang, J., 2008. Estimating crop water stress with ETM+ NIR and SWIR data. Agric. For. Meteorol. 148 (11), 1679–1695. http://dx.doi.org/10.1016/j.agrformet.2008.05.020.

Gitelson, A.A., Viña, A., Arkebauer, T.J., Rundquist, D.C., Keydan, G., Leavitt, B., 2003. Remote estimation of leaf area index and green leaf biomass in maize canopies. Geophys. Res. Lett. 30 (5) n/a-n/a. https://doi.org/10.1029/2002GL016450.

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

Hansen, M.C., Defries, R.S., Townshend, J.R.G., Sohlberg, R., 2000. Global land cover classification at 1 km spatial resolution using a classification tree approach. Int. J. Remote Sens. 21 (6–7), 1331–1364. http://dx.doi.org/10.1080/014311600210209.

Hansen, M.C., Egorov, A., Roy, D.P., Potapov, P., Ju, J., Turubanova, S., Kommareddy, I., Loveland, T.R., 2011. Continuous fields of land cover for the conterminous United States using Landsat data: first results from the Web-Enabled Landsat Data (WELD) project. Remote Sens. Lett. 2 (4), 279–288. http://dx.doi.org/10.1080/01431161.2010.519002.

Hansen, M.C., Egorov, A., Potapov, P.V., Stehman, S.V., Tyukavina, A., Turubanova, S.A., Roy, D.P., Goetz, S.J., Loveland, T.R., Ju, J., Kommareddy, A., Kovalskyy, V., Forsyth, C., Bents, T., 2014. Monitoring conterminous United States (CONUS) land cover change with Web-Enabled Landsat Data (WELD). Remote Sens. Environ. 140, 466–484. http://dx.doi.org/10.1016/j.rse.2013.08.014.

Hinton, G., Deng, L., Yu, D., Dahl, G., Mohamed, A., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., Sainath, T., Kingsbury, B., 2012. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. IEEE Signal Process. Mag. 29 (6), 82–97. http://dx.doi.org/10.1109/MSP.2012.2205597.

Homer, C., Huang, C., Yang, L., Wylie, B., Coan, M., 2004. Development of a 2001 national land-cover database for the United States. Photogramm. Eng. Remote Sens. 70 (7), 829–840. http://dx.doi.org/10.14358/PERS.70.7.829.

Huang, H., Legarsky, J., Othman, M., 2007. Land-cover classification using radarsat and Landsat imagery for St. Louis, Missouri. Photogramm. Eng. Remote Sens. 73 (1), 37–43. http://dx.doi.org/10.14358/PERS.73.1.37.

Huete, A., Didan, K., Miura, T., Rodriguez, E., Gao, X., Ferreira, L., 2002. Overview of the radiometric and biophysical performance of the MODIS vegetation indices. Remote Sens. Environ. 83 (1), 195–213. http://dx.doi.org/10.1016/S0034-4257(02)00096-2.

Jonsson, P., Eklundh, L., 2002. Seasonality extraction by function fitting to time-series of satellite sensor data. IEEE Trans. Geosci. Remote Sens. 40 (8), 1824–1832. http://dx.doi.org/10.1109/TGRS.2002.802519.

King, L., Adusei, B., Stehman, S.V., Potapov, P.V., Song, X.-P., Krylov, A., Di Bella, C., Loveland, T.R., Johnson, D.M., Hansen, M.C., 2017. A multi-resolution approach to national-scale cultivated area estimation of soybean. Remote Sens. Environ. 195, 13–29. http://dx.doi.org/10.1016/j.rse.2017.03.047.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet Classification With Deep Convolutional Neural Networks. Advances in Neural Information Processing Systems.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444. http://dx.doi.org/10.1038/nature14539.

Li, P., Jiang, L., Feng, Z., 2013. Cross-comparison of vegetation indices derived from Landsat-7 Enhanced Thematic Mapper Plus (ETM+) and Landsat-8 Operational Land Imager (OLI) sensors. Remote Sens. 6, 310–329. http://dx.doi.org/10.3390/rs6010310.

Liu, J., Liu, M., Tian, H., Zhuang, D., Zhang, Z., Zhang, W., Tang, X., Deng, X., 2005. Spatial and temporal patterns of China's cropland during 1990–2000: an analysis based on Landsat TM data. Remote Sens. Environ. 98 (4), 442–456. http://dx.doi.org/10.1016/j.rse.2005.08.012.

Lobell, D.B., 2013. The use of satellite data for crop yield gap analysis. F. Crop. Res. 143, 56–64. http://dx.doi.org/10.1016/j.fcr.2012.08.008.

Lobell, D.B., Asner, G.P., 2004. Cropland distributions from temporal unmixing of MODIS data. Remote Sens. Environ. 93 (3), 412–422. http://dx.doi.org/10.1016/j.rse.2004.08.002.

Lobell, D.B., Thau, D., Seifert, C., Engle, E., Little, B., 2015. A scalable satellite-based crop yield mapper. Remote Sens. Environ. 164, 324–333. http://dx.doi.org/10.1016/j.rse.2015.04.021.

Roy, D.P., Wulder, M.A., Loveland, T.R., W., C.E., Allen, R.G., Anderson, M.C., Helder, D., Irons, J.R., Johnson, D.M., Kennedy, R., Scambos, T.A., Schaaf, C.B., Schott, J.R., Sheng, Y., Vermote, E.F., Belward, A.S., Bindschadler, R., Cohen, W.B., Gao, F., Hipple, J.D., Hostert, P., Huntington, J., Justice, C.O., Kilic, A., Kovalskyy, V., Lee, Z.P., Lymburner, L., Masek, J.G., McCorkel, J., Shuai, Y., Trezza, R., Vogelmann, J., Wynne, R.H., Zhu, Z., 2014. Landsat-8: science and product vision for terrestrial global change research. Remote Sens. Environ. 145, 154–172. http://dx.doi.org/10.1016/j.rse.2014.02.001.

Rumelhart, D.E., Hinton, G.E., Williams, R.J., 1986. Learning representations by back-propagating errors. Nature 323, 533–536. http://dx.doi.org/10.1038/323533a0.

Schmidhuber, J., 2015. Deep learning in neural networks: an overview. Neural Netw. 61, 85–117. http://dx.doi.org/10.1016/j.neunet.2014.09.003.

Schmidt, G., Jenkerson, C., Masek, J., Vermote, E., Gao, F., 2013. Landsat Ecosystem Disturbance Adaptive Processing System (LEDAPS) Algorithm Description. US Geological Survey.

Sellers, P.J., Berry, J.A., Collatz, G.J., Field, C.B., Hall, F.G., 1992. Canopy reflectance, photosynthesis, and transpiration. III. A reanalysis using improved leaf models and a new canopy integration scheme. Remote Sens. Environ. 42 (3), 187–216. http://dx.doi.org/10.1016/0034-4257(92)90102-P.

Sexton, J.O., Song, X.-P., Feng, M., Noojipady, P., Anand, A., Huang, C., Kim, D.-H., Collins, K.M., Channan, S., DiMiceli, C., Townshend, J.R., 2013a. Global, 30-m resolution continuous fields of tree cover: Landsat-based rescaling of MODIS vegetation continuous fields with lidar-based estimates of error. Int. J. Digit. Earth 6 (5), 427–448. http://dx.doi.org/10.1080/17538947.2013.786146.

Sexton, J.O., Urban, D.L., Donohue, M.J., Song, C., 2013b. Long-term land cover dynamics by multi-temporal classification across the Landsat-5 record. Remote Sens. Environ. 128, 246–258. http://dx.doi.org/10.1016/j.rse.2012.10.010.

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

Townshend, J.R., Masek, J.G., Huang, C., Vermote, E.F., Gao, F., Channan, S., Sexton, J.O., Feng, M., Narasimhan, R., Kim, D., Song, K., Song, D., Song, X.-P., Noojipady, P., Tan, B., Hansen, M.C., Li, M., Wolfe, R.E., 2012. Global characterization and monitoring of forest cover using Landsat data: opportunities and challenges. Int. J. Digit. Earth 5 (5), 373–397. http://dx.doi.org/10.1080/17538947.2012.713190.

Tucker, C.J., 1979. Red and photographic infrared linear combinations for monitoring vegetation. Remote Sens. Environ. 8 (2), 127–150. http://dx.doi.org/10.1016/0034-4257(79)90013-0.

USDA, 2012. 2012 Census of Agriculture – Champaign County, Illinois.

USGS, 2016. Landsat Surface Reflectance Higher-level Data Products [WWW Document]. http://landsat.usgs.gov/CDR_LSR.php, Accessed date: 22 April 2016.

Van Niel, T.G., McVicar, T.R., 2004. Determining temporal windows for crop discrimination with remote sensing: a case study in south-eastern Australia. Comput. Electron. Agric. 45 (1), 91–108. http://dx.doi.org/10.1016/j.compag.2004.06.003.

Vogelmann, J.E., Howard, S.M., Yang, L., Larson, C.R., Wylie, B.K., Van Driel, N., 2001. Completion of the 1990s National Land Cover Data Set for the conterminous United States from Landsat Thematic Mapper data and ancillary data sources. Photogramm. Eng. Remote Sens. 67 (6).

Wardlow, B.D., Egbert, S.L., 2008. Large-area crop mapping using time-series MODIS 250 m NDVI data: an assessment for the U.S. Central Great Plains. Remote Sens. Environ. 112 (3), 1096–1116. http://dx.doi.org/10.1016/j.rse.2007.07.019.

Wardlow, B.D., Egbert, S.L., Kastens, J.H., 2007. Analysis of time-series MODIS 250 m vegetation index data for crop classification in the U.S. Central Great Plains. Remote Sens. Environ. 108 (3), 290–310. http://dx.doi.org/10.1016/j.rse.2006.11.021.

Xiao, X., Boles, S., Liu, J., Zhuang, D., Liu, M., 2002. Characterization of forest types in Northeastern China, using multi-temporal SPOT-4 VEGETATION sensor data. Remote Sens. Environ. 82 (2), 335–348. http://dx.doi.org/10.1016/S0034-4257(02)00051-2.

Xiao, X., Boles, S., Frolking, S., Li, C., Babu, J.Y., Salas, W., Moore, B., 2006. Mapping paddy rice agriculture in South and Southeast Asia using multi-temporal MODIS images. Remote Sens. Environ. 100 (1), 95–113. http://dx.doi.org/10.1016/j.rse.2005.10.004.

Yang, C., Everitt, J.H., Murden, D., 2011. Evaluating high resolution SPOT 5 satellite imagery for crop identification. Comput. Electron. Agric. 75 (2), 347–354. http://dx.doi.org/10.1016/j.compag.2010.12.012.

You, J., Li, X., Low, M., Lobell, D., Ermon, S., 2017. Deep Gaussian Process for Crop Yield Prediction Based on Remote Sensing Data. AAAI.

Yuan, F., Sawaya, K.E., Loeffelholz, B.C., Bauer, M.E., 2005. Land cover classification and change analysis of the Twin Cities (Minnesota) Metropolitan Area by multitemporal Landsat remote sensing. Remote Sens. Environ. 98 (2), 317–328. http://dx.doi.org/10.1016/j.rse.2005.08.006.

Zhan, X., Sohlberg, R., Townshend, J.R., DiMiceli, C., Carroll, M., Eastman, J., Hansen, M., DeFries, R., 2002. Detection of land cover changes using MODIS 250 m data. Remote Sens. Environ. 83 (1), 336–350. http://dx.doi.org/10.1016/S0034-4257(02)00081-0.