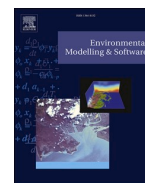




Contents lists available at ScienceDirect

Environmental Modelling and Software

journal homepage: <http://www.elsevier.com/locate/envsoft>

AgKit4EE: A toolkit for agricultural land use modeling of the conterminous United States based on Google Earth Engine

Chen Zhang^{a,b}, Liping Di^{a,b,*}, Zhengwei Yang^c, Li Lin^{a,b}, Pengyu Hao^a^a Center for Spatial Information Science and Systems, George Mason University, Fairfax, VA, 22030, USA^b Department of Geography and Geoinformation Science, George Mason University, Fairfax, VA, 22030, USA^c Research and Development Division, U.S. Department of Agriculture National Agricultural Statistics Service, Washington, DC, 20250, USA

ARTICLE INFO

Keywords:

Google Earth Engine
Cropland Data Layer
Land use modeling
Crop mapping
Geospatial cyberinfrastructure

ABSTRACT

Google Earth Engine (GEE) is an ideal platform for large-scale geospatial agricultural and environmental modeling based on its diverse geospatial datasets, easy-to-use application programming interface (API), rich reusable library, and high-performance computational capacity. However, using GEE to prepare geospatial data requires not only the skills of programming languages like JavaScript and Python, but also the knowledge of GEE APIs and data catalog. This paper presents the AgKit4EE toolkit to facilitate the use of the Cropland Data Layer (CDL) product over the GEE platform. This toolkit contains a variety of frequently used functions for use of CDL including crop sequence modeling, crop frequency modeling, confidence layer modeling, and land use change analysis. The experimental results suggest that the proposed software can significantly reduce the workload for modelers who conduct geospatial agricultural and environmental modeling with CDL data as well as developers who build the GEE-enabled geospatial cyberinfrastructure for agricultural land use modeling of the conterminous United States. AgKit4EE is an open source and it is free to use, modify, and distribute. The latest release of AgKit4EE can be imported to any modeling workflow developed using GEE Code Editor (https://code.earthengine.google.com/?accept_repo=users/czhang11/agkit4ee). The source code, examples, documentation, user community, and wiki pages are available on GitHub (<https://github.com/czhang11/agkit4ee>).

1. Introduction

The conterminous United States (CONUS), which has the largest production areas of corn, soybeans, and sorghum in the world, is an ideal study area for scientists to model natural environment and human activities in agriculture (Feng and Hu, 2004; McCarty et al., 2009; Li et al., 2016; Feng et al., 2019; Flynn, 2019). As the only annual crop-specific land use and land cover (LULC) data product of the CONUS, the Cropland Data Layer (CDL) of the U.S. Department of Agriculture (USDA), National Agricultural Statistics Service (NASS) has been widely used by farmers, educators, students, researchers, and government officers in agricultural business, universities, research institutes, and governments worldwide for agricultural production planning and management, education, government policy formulation and decision making, and various research activities (Boryan et al., 2011; Stern et al., 2012; Mueller and Harris, 2013; Di et al., 2015). The CDL product has been disseminated through CropScape since 2011. According to the report from Google Analytics, more than 208,000 unique users around the

world have visited and interacted with CropScape as of May 2019. As an one-of-a-kind one-stop platform to visualize, retrieve, and analyze CDL, CropScape provides not only various geospatial functionalities such as data customization and downloading, crop acreage statistics, charting and graphing, and multi-temporal change analysis, but also the highly interoperable open standard-based geospatial web services such as Web Map Service (WMS), Web Coverage Service (WCS), and Web Processing Service (WPS) (Han et al., 2012; Zhang et al., 2019d). These functionalities and web services of CropScape have been utilized in many geospatial environmental models, frameworks, and software (Han et al., 2014; Feng et al., 2015; McNider et al., 2015; Groff et al., 2016; Tasdighi et al., 2018). However, due to the limited computing resources, it is difficult to perform advanced modeling and mathematical operations on the multi-year CDL data directly using either the CropScape web portal or its web services. To fill the gap and further facilitate the use of CDL data, a new web-based tool for the CONUS-scale agricultural land use modeling is needed.

Modeling multi-year CDL data for the entire CONUS requires massive

* Corresponding author. Department of Geography and Geoinformation Science, George Mason University, Fairfax, VA, 22030, USA.

E-mail addresses: czhang11@gmu.edu (C. Zhang), ldi@gmu.edu (L. Di), Zhengwei.Yang@usda.gov (Z. Yang), llin2@gmu.edu (L. Lin), phao@gmu.edu (P. Hao).

<https://doi.org/10.1016/j.envsoft.2020.104694>

Received 21 May 2019; Received in revised form 29 January 2020; Accepted 9 March 2020

Available online 13 March 2020

1364-8152/© 2020 Elsevier Ltd. All rights reserved.

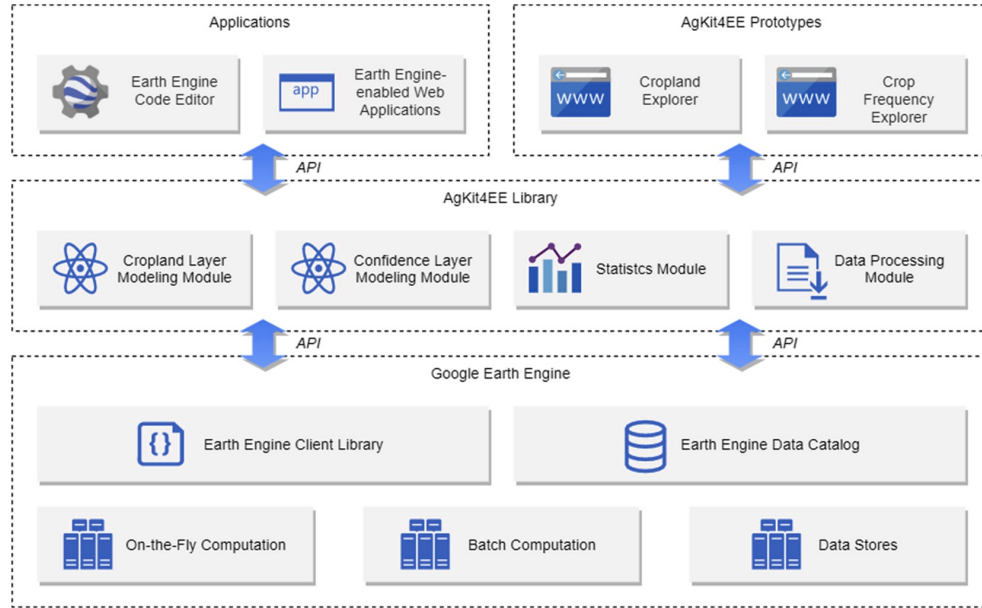


Fig. 1. Architectural context of the proposed GEE-enabled toolkit for agricultural land use modeling.

Table 1

Summary of CDL products in the public data catalog of Google Earth Engine.

	Cropland Layer	Confidence Layer	Cultivated Layer
Availability	1997–2018	2008–2018	2013–2018
Coverage	CONUS (2008–2018) Some states (1997–2008)	CONUS	CONUS
Spatial resolution	30 m	30 m	30 m
Pixel value	0–254	0–100	0–2

Table 2

Data structures defined in the AgKit4EE toolkit.

Data Structure in AgKit4EE	Corresponding GEE Data Structure	Description
Single-band CDL Image	ee.Image()	A single-band image for the specific year
Single-band CDL ImageCollection	ee.ImageCollection()	A collection of single-band CDL images
Stacked CDL Image	ee.Image()	A multi-bands image for the specific CDL product, each band refers to a layer of the specific year
Multi-bands CDL Image	ee.Image()	A multi-bands image for the specific year, each band refers to the specific crop type
Multi-bands CDL ImageCollection	ee.ImageCollection()	A collection of multi-bands CDL images

computational capacity. Therefore, the development of the software should be based on a high-performance geospatial cyberinfrastructure (CI), which supports the collection, management, share, analysis visualization, and dissemination of the geospatial big data over the high-speed network (Yang et al., 2010; Yue et al., 2015). During the past decade, many geospatial applications and tools have been transformed from the traditional geographic information system (GIS) software to the geospatial CI with the rapid advancements in web service technologies (Castronova et al., 2013; Lin et al., 2017; Zhang et al., 2019c), geospatial

information interoperability (Zhao and Di, 2010; Goodall et al., 2013; Nativi et al., 2013; Sun et al., 2017), geospatial cloud computing (Yang and Huang, 2013; Zhang et al., 2017), high-performance computing (Lee et al., 2011), and geospatial big data analytics (Deng and Di, 2014; Vitolo et al., 2015; Di, 2016). Among those geospatial applications, there are many geospatial CIs serving for the environmental modeling community. The Self-adaptive Earth Predictive Systems (SEPS) adopts the service-oriented architecture (SOA) to bridge Earth observation (EO) data and Earth system models using OGC/ISO Sensor Web standards and geospatial interoperability protocols (Di, 2007). CyberGIS (Wang, 2010), a GIS framework based on the advanced CI, has been integrated into many GIS applications and software (Wang et al., 2013; Padmanabhan et al., 2014; Lin et al., 2015). Global Earth Observation System of Systems (GEOSS) platform links a set of EO systems around the world to facilitate the monitoring of the state of the Earth and the sharing of environmental data (Nativi et al., 2013; Santoro et al., 2016). As a collaboration between the Division of Advanced Cyberinfrastructure and the Geosciences Directorate of National Science Foundation (NSF), EarthCube, a community-driven organization for geoscience CI, has funded many projects to improve data access, sharing, visualization, and analysis across geoscience disciplines (Katz, 2015). These projects include GeoLink (Krisnadhi et al., 2015), Cloud Hosted Real-time Data Services for the Geosciences (CHORDS) (Kerkez et al., 2016), Brokering Building Block (BCube) (Khalsa, 2017), CyberConnector (Di et al., 2017; Sun et al., 2017), CyberWay (Di et al., 2019), HydroShare and GeoTrust (Essawy et al., 2018; Xue et al., 2019).

As one of the major players in the cloud computing business, Google unveiled Google Earth Engine (GEE) in 2010. GEE is a cloud-based platform for planetary-scale geospatial data analysis with diverse geospatial datasets and a variety of ready-to-use application programming interface (API) (Gorelick et al., 2017). It has been used as the major computing platform in many Earth system science studies including LULC change detection (Hansen et al., 2013; Huang et al., 2017; Mid-ekisa et al., 2017; Yu et al., 2018), crop mapping (Shelestov et al., 2017; Teluguntla et al., 2018), digital soil mapping (Padarian et al., 2015), forest mapping (Chen et al., 2017; Koskinen et al., 2019), and wetland mapping (Hird et al., 2017). However, prototyping a complicated modeling algorithm with GEE involves the cost of learning JavaScript and GEE APIs, which is a time-consuming job for modelers, especially

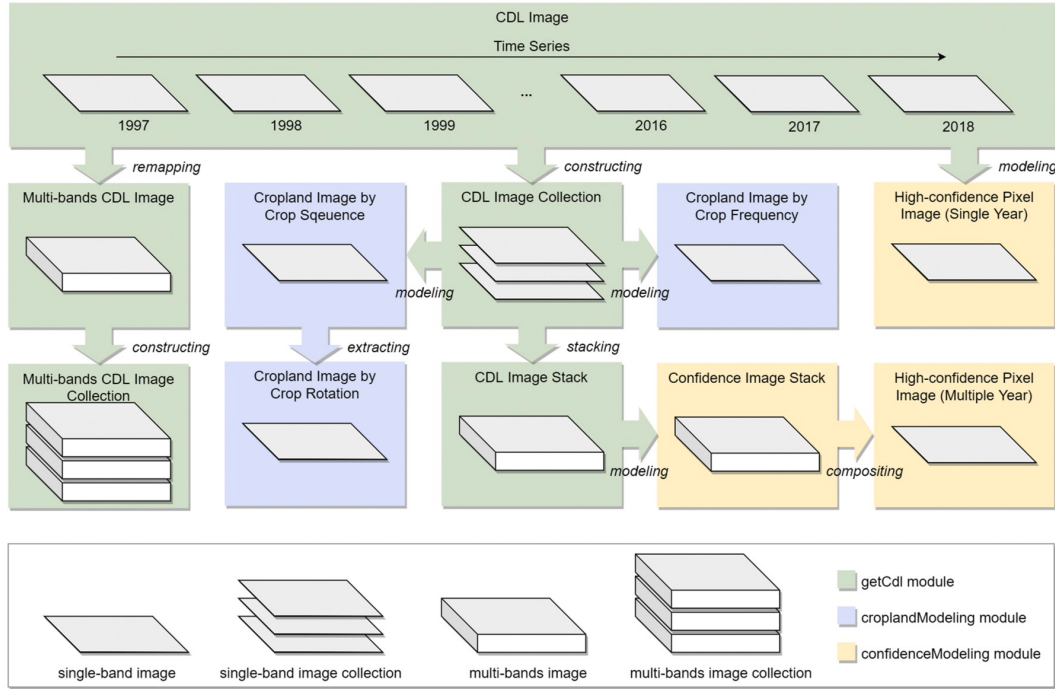


Fig. 2. Development flow of core modules in the AgKit4EE toolkit.

those who do not have a strong technical background. To address this issue, GEE provides the JavaScript/Python client library allowing users to develop GEE-enabled applications and tools as needed. For example, AgriSuit provides a Web-based framework for supporting land-use suitability analysis, which integrating spatial datasets, algorithms, and computing capabilities of GEE platform (Yalew et al., 2016). Collect Earth Online (CEO) offers an open source tool for systematic reference data collection in land cover and use applications (Bey et al., 2016). Flood Prevention and Emergency Response System, a GEE-powered Web-based platform for supporting flood event prevention and emergency response, has been applied in 19 typhoons and torrential rain events from 2013 to 2016 in Taiwan (Liu et al., 2018). The Biomass Estimation platform is a cloud-based application for aboveground biomass mapping and estimation which integrated GEE (Yang et al., 2018). CoastSat is a GEE toolkit to extract shorelines from satellite imagery in large scale (Vos et al., 2019). Li et al., 2019 presents a GEE-enabled toolbox of generating high-quality user-ready Landsat mosaic images.

In this paper, we present the AgKit4EE toolkit to (a) simplify the the CONUS-scale agricultural land use modeling on GEE; (b) derive CDL-based land use data products on-the-fly with GEE; and (c) boost the development of GEE-enabled web applications for agricultural land use modeling. The rest of the paper is organized as follow. Section 2 introduces the architectural context, data, development flow, and core functions. Section 3 presents examples to demonstrate features and capabilities of the toolkit. Section 4 discusses the application scenario, the advantages, and the limitations of the current implementation. The conclusion and future works are given in Section 5.

2. Software design

2.1. Architectural context

Fig. 1 illustrates the architectural context of the proposed software. The development of the AgKit4EE toolkit is fully based on GEE client library and GEE data catalog, which are powered by the high-

performance computation and data stores of Google's cloud infrastructure. The AgKit4EE library contains a suite of modules such as modeling modules, statistics modules, and data processing modules. The library is developed using JavaScript, which is consistent with GEE Code Editor. Users can directly import the library to any project in GEE Code Editor and GEE-enabled web applications. Currently, we have enabled AgKit4EE in two web application prototypes, the Cropland Explorer and the Crop Frequency Explorer, which are published over the Earth Engine Apps platform.

As the core component in the architectural context, the AgKit4EE library includes a variety of frequently used functions for the retrieval, process, modeling, and statistics of CDL data. All functions are developed based on the native GEE APIs and wrapped as the ready-to-use APIs with required and optional arguments. The code example of retrieving a CDL image collection of corn and soybeans for selected years using the `getCdlCollection` function as well as the equivalent code using native GEE APIs are compared in Appendix A, Fig. A1. This example presents a workflow of CDL data retrieval with the readable input arguments in one line of code, where the *product* option refers to the band name of the original CDL data in GEE data catalog, the *remap* option includes a list of the value for the desired crop type codes ("1" and "5" refer to corn and soybeans respectively), the *years* option specifies the years of interest. As a control, the same functionality is equivalent to a series of native GEE APIs including image blending, band selection, value remapping, and mapping over images. It is obvious that such a simplification would significantly reduce the burdens of GEE coding for modelers who are not familiar with GEE APIs and CDL products.

2.2. Data

The objective of the AgKit4EE toolkit is to facilitate the use of CDL data on GEE. The complete collection of historical CDL data has been archived in the GEE data catalog (https://developers.google.com/earth-engine/datasets/catalog/USDA_NASS_CDL). Table 1 summarizes the information of three CDL products, including the cropland layer, the confidence layer, and the cultivated layer. The cropland layer covers the

Table 3

Summary of functions and capabilities offered by the AgKit4EE toolkit.

Module	Function	Output Type	Description
getCdl	getCdlImageByYear (year, options)	Single-band image	Get the CDL image for selected year
	getCdlCollection (options)	Single-band image collection	Get the CDL image collection for selected year range
	getCdlImageStack (options)	Multi-bands image	Get the stacked multi-bands CDL image by stacking multiple single-band CDL images
	getCdlBandsByYear (year, options)	Multi-bands image	Get the multi-bands CDL image of specific crop types for the selected year
	getCdlBandsCollection (options)	Multi-bands image collection	Get the image collection of multi-bands CDL images
confidenceModeling	getCdlPalette (options)	List	Get the list of CDL color scheme
	getConfidenceImageByYear (year, threshold, options)	Single-band image	Get the map of binarized confidence layer for the selected year with assigned threshold value
	getConfidenceImageStack (threshold, options)	Multi-bands image	Get the image stack of binarized confidence layers with assigned threshold value
	getConfidenceImageCollection (threshold, options)	Single-band image collection	Get the image collection of binarized confidence layer with assigned threshold value
croplandModeling	getTrustedConfidence (threshold, options)	Single-band image	Get the map of trusted pixels through multi-year confidence layers with assigned threshold value
	modelingByCropSequence (cropSequence, targetCrop, options)	Single-band image	Get the map of cropland layer that following the common crop rotation pattern
	modelingByPattern (pattern, targetCrop, options)	Single-band image	Get the map of cropland layer that following the common crop rotation pattern
	modelingFrequencyByCrop (targetCrop, options)	Single-band image	Get the frequency map of selected crop type based on the historical cropland layers
getRoi	getFrequencyPalette (options)	List	Get the list of crop frequency color scheme
	getRoiByFips (fips, options)	Feature Collection	Get the region of interest for state, county, or ASD by FIPS code
	getRoiByWrs2 (wrs2Scene, options)	Feature	Get the region of interest for assigned footprint (path/row) of Landsat WRS-2
export	getRoiBySen2Tile (tile, options)	Feature	Get the region of interest for assigned tile of Sentinel-2 image
	exportCdlByRoi (roi, options)	GeoTIFF image	Batch export the historical CDL image of the region of interest to local
	exportCdlByFips (fips, options)	GeoTIFF image	Batch export the historical CDL image of selected state/county/ASD to local
statistics	exportCdlByWrs (footprint, options)	GeoTIFF image	Batch export the historical CDL image of selected WRS-2 footprint to local
	exportCdlBySen2Tile (tile, options)	GeoTIFF image	Batch export the historical CDL image of selected Sentinel-2 tile to local
	statisticsByRoi (roi, options)	Chart	Create a time series chart of crop acreage for the region of interest
	statisticsByCounty (fips, options)	Chart	Create a time series chart of crop acreage for the selected county
	statisticsByPoi (poi, options)	Chart	Create a time series chart of crop type for the selected point

entire CONUS from 2008 to 2018 and some states from 1997 to 2007, which is composed of over 140 land cover classes with 30m spatial resolution. The confidence layer covers the entire CONUS from 2008 to 2018, which reflects the percentage (0–100) of confidence for each cropland pixel (Liu et al., 2004). The cultivated layer covers the entire CONUS from 2013 to present, which is produced based on the most recent five years of cropland layer. The current-year CDL data would be first released through CropScape and the USDA NASS website in the early next year. For example, the 2019 CDL, which is the latest CDL data as of the writing of this article, was published on January 2020. However, the collection of some datasets on GEE might be delayed. The 2019 CDL has not been archived in GEE data catalog as of March 2020.

Besides the CDL data, the toolkit integrated a collection of boundary data that frequently used in agricultural and environmental modeling, including U.S. county boundary, U.S. state boundary, USDA Agricultural Statistics District (ASD) boundary, Landsat World Reference System-2 (WRS-2) scene boundary, and Sentinel-2 tile boundary. These boundary data will improve efficiency while preparing data for region of interest. For example, users can export the CDL data of specific U.S. state, county, or ASD by using the *exportCdlByFips* function with the specific Federal Information Processing Standards (FIPS) code.

2.3. Implementation

GEE provides two basic geospatial data structures, *Image* and *ImageCollection*, to manipulate raster data. The *Image* is a single raster image data composed by one or multiple bands. The *ImageCollection* is a stack of the *Image*. As of December 2019, the GEE data catalog has archived all historical CDL data from 1997 to 2018. The CDL product of

each year is saved as an *Image*. Depending on the data availability (Table 1), each CDL *Image* contains either one, two, or three bands. Based on *Image* and *ImageCollection*, we defined and implemented five extended data structure options to manipulate CDL images. They are *Single-band CDL Image*, *Single-band CDL ImageCollection*, *Stacked CDL Image*, *Multi-bands CDL Image*, and *Multi-bands CDL ImageCollection*. Table 2 summarizes all data structures defined in AgKit4EE.

Fig. 2 illustrates the architecture and the development flow of the core modules in AgKit4EE. The development is composed of a suite of modules, and each module contains a group of functions. The *getCdl* module provides the capability of getting original CDL data according to user requirements, such as the product type, year, or crop type. Meanwhile, all modeling functions are implemented based on the *getCdl* module. According to the data product type, the modeling functions are implemented as *croplandModeling* module and *confidenceModeling* module respectively. The *croplandModeling* module consists of functions related to the cropland layer, such as crop sequence modeling, crop rotation modeling, and crop frequency modeling. The *confidenceModeling* module handles the functions of pixel-level confidence percentage modeling based on the confidence layer. Additionally, there are several miscellaneous modules offering common geospatial functions. The *getRoi* module manages the U.S. boundary files. The *export* module allows user to batch export of the on-demand CDL data. The *statistics* module provides the statistical functions for agricultural land use change analysis.

2.4. Functions and capabilities

All functions in the AgKit4EE toolkit contain one or more arguments.

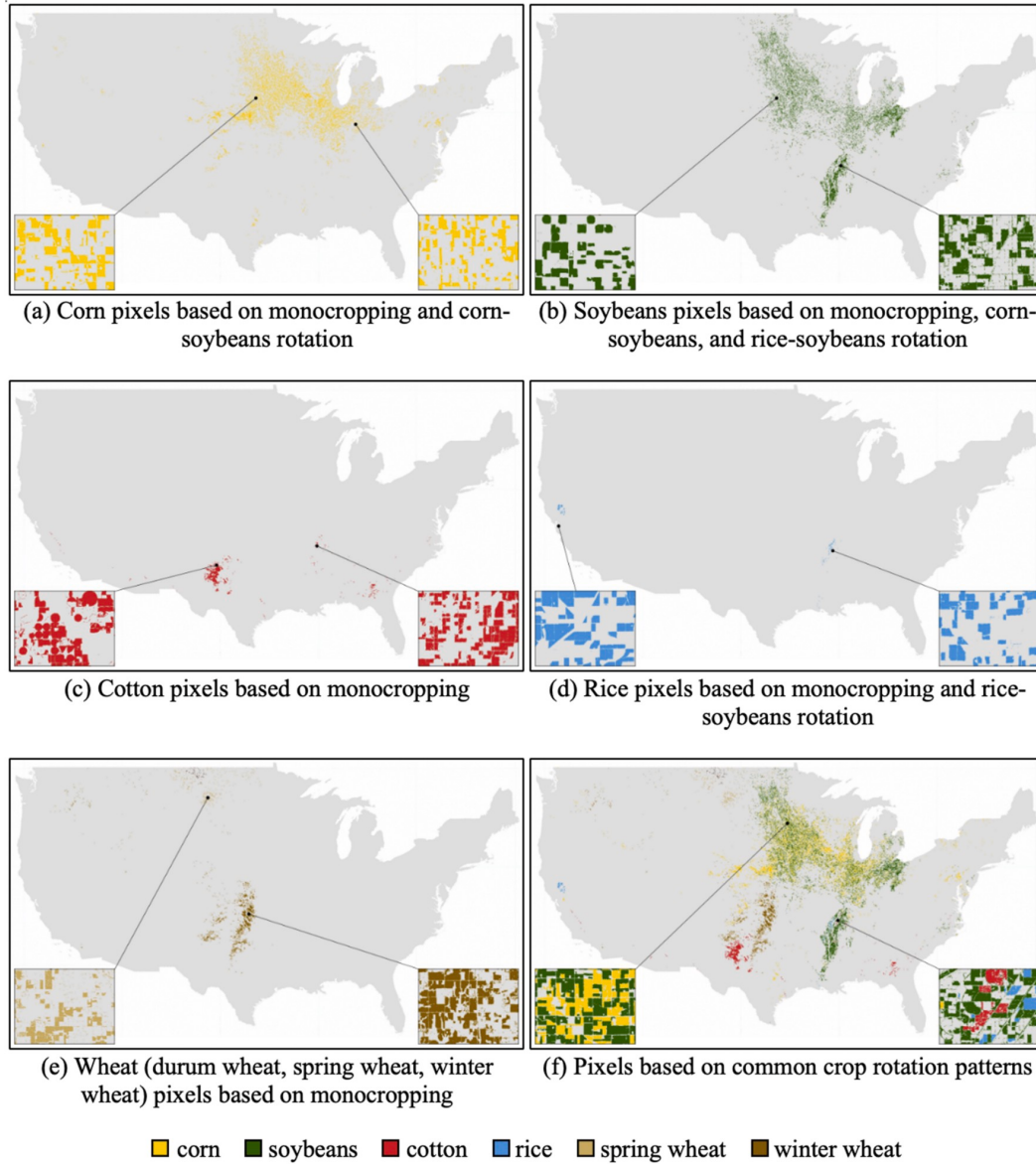


Fig. 3. Mapping major crops for the CONUS based on the common crop rotation patterns.

Besides the required arguments that users must assign before use, most functions also have several optional arguments, which are passed through the *options* object. For example, the optional arguments of the *getCdlCollection* function in the *getCdl* module consist of *product*, *years*, *remap*, and *defaultValue*. The *product* option is the product name of the desired CDL layer. The *years* option is a list of years of interest. The *remap* option is a list of the codes of target crops. The *defaultValue* option is the code value of no-data pixels. Table 3 summarizes all functions offered by the current release of the toolkit.

3. Examples

This section gives several examples of agricultural land use modeling and analysis using the AgKit4EE toolkit. First, we demonstrate the capability of crop mapping for the entire CONUS using the crop sequence modeling function (Section 3.1). Then we present an example of crop frequency modeling and compare the result with the official crop

frequency map produced by USDA NASS (Section 3.2). In addition, we also illustrate how to extract the high-confidence CDL pixels by stratifying the historical confidence layers (Section 3.3). Finally, two AgKit4EE-enabled web applications are prototypically implemented and published through the Earth Engine Apps platform (Section 3.4).

3.1. Mapping cropland by crop sequence

The cropping sequence can affect crop yield (Edwards et al., 1988; Crookston et al., 1991; Berzsenyi et al., 2000) as well as soil quality, soil fertility, and soil physical/chemical properties (Janzen et al., 1992; Karlen et al., 2006; Van Eerd et al., 2014; Triberti et al., 2016). For instance, the corn-soybean rotation, a widely adopted common cropping practice, helps preserve the croplands productivity in the U.S. Corn Belt. Based on the reliable crop sequence information, the types of crops to be planted can be predicted before the growing season starts (Zhang et al., 2019a; Zhang et al., 2019b). These prediction and pre-season crop

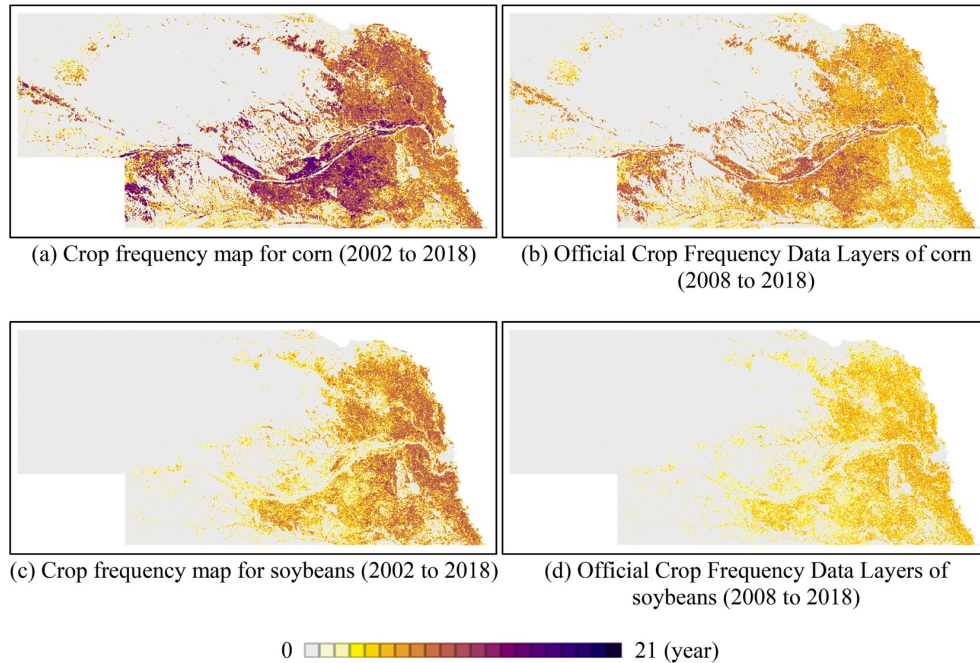


Fig. 4. Comparison of crop planting frequency maps by AgKit4EE with the official Crop Frequency Data Layers by USDA NASS.

planting information are critical to many early-season environmental modeling and applications. To facilitate the pre-season crop mapping, AgKit4EE offers an innovative function to extract pixels that following the specific cropping sequence in the recent years. The code example of mapping major crops based on the common crop rotation patterns using the *modelingByCropSequence* function can be found in Appendix A, Fig. A2. In this example, the variable *corn_mono* and *corn_rotation* refer to the corn pixels following the monocropping and corn-soybeans rotated cropping pattern. The variable *soybean_mono*, *soybean_rotation*, and *soybean_rotation_alt* refer to the soybeans pixels following the monocropping, corn-soybeans, and rice-soybean rotated cropping pattern. The variable *cotton_mono* refers to the cotton pixels following the monocropping pattern. The variable *rice_mono* and *rice_rotation* refer to the rice pixels following the monocropping pattern and rice-soybeans rotated cropping pattern. The variable *durumWheat_mono*, *springWheat_mono*, and *winterWheat_mono* refer to the wheat pixels following the monocropping pattern. Fig. 3 shows the spatial distribution of major crops extracted from the historical CDL data with the above-mentioned crop sequences.

3.2. Mapping cropland by crop frequency

USDA NASS releases the Crop Frequency Data Layers accompanying with the release of annual cropland layer. This product identifies the planting frequency for the specific crop type based on the CDL from 2008 to present (Boryan et al., 2014). The Crop Frequency Data Layers for four major crops in the U.S., corn, cotton, soybeans, and wheat, are available to the public. The latest Crop Frequency Data Layer products are available on CropScape and the USDA NASS website. The AgKit4EE toolkit provides the capability of modeling crop frequency based on the historical cropland layers. The code example of mapping the frequency of major crops (i.e., corn and soybeans) for the state of Nebraska can be found in Appendix A, Fig. A3. The variable *freq_corn* refers to the planting frequency map of corn and *freq_soybean* refers to the planting frequency map of soybeans. Fig. 4 compares the crop planting frequency maps with the NASS official Crop Frequency Data Layers.

As shown in Fig. 4a and c, the AgKit4EE derived crop planting

frequency maps contain more information than the NASS Crop Frequency Data Layers as shown in Fig. 4b and d. More importantly, the toolkit provides the capability of modeling the frequency of any crop type from CDL within any available year range. For example, users can model the frequency of major crops for the state of Arkansas in the recent five years (2014–2018) by assigning the *years* option of the *modelingFrequencyByCrop* function (see Appendix A, Figure, A4). Fig. 5 shows the mapping results of corn (Fig. 5a), cotton (Fig. 5b), rice (Fig. 5c), soybeans (Fig. 5d), winter wheat (Fig. 5e), and double cropping of winter wheat and soybeans (Fig. 5f).

3.3. Mapping confidence layer by threshold

The CDL data are produced using C5.0/See5 decision tree algorithm. For every classified CDL pixel, there is an associated classification confidence measure, which measures the confidence percentage of the corresponding CDL pixel. A map of the CDL pixels that are higher or lower than a specific confidence level threshold can be produced by binarizing the confidence layer with the specific threshold. Furthermore, by modeling the multi-years of confidence layers, we can observe the spatial distribution of the trusted cropland pixels maintaining high confidence for identifying the crop type of the corresponding pixel. The code example of modeling confidence layers with the *modelingConfidenceByThreshold* function can be found in Appendix A, Fig. A5. The variable *conf_70*, *conf_80*, *conf_90*, and *conf_100* represent the maps of cropland pixels which the confidence value is consistently greater than 70%, 80%, 90%, and 100%, respectively, for the given time period. Fig. 6 shows the binarized maps of the modeling result, in which the bright pixels represent the pixels higher than the threshold value.

3.4. Enabling AgKit4EE in web applications

In this study, we enabled AgKit4EE in two web application prototypes, the Cropland Explorer (<https://czhang11.users.earthengine.app/view/agkit4ee-cdl-explorer>) and the Crop Frequency Explorer (<https://czhang11.users.earthengine.app/view/agkit4ee-crop-frequency-explorer>). These prototypes are published as the GEE app

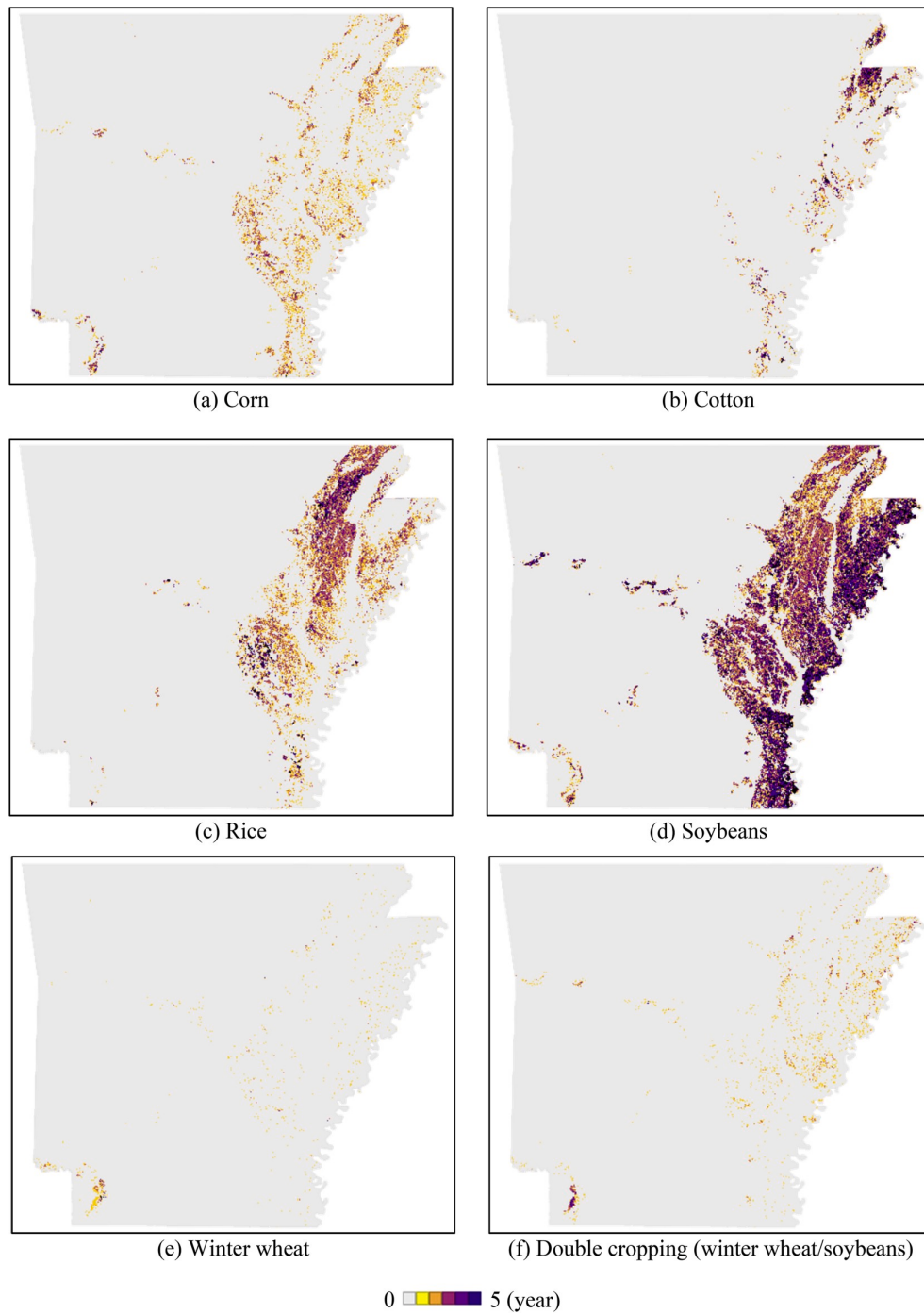


Fig. 5. Mapping crop frequency for the state of Arkansas from 2014 to 2018.

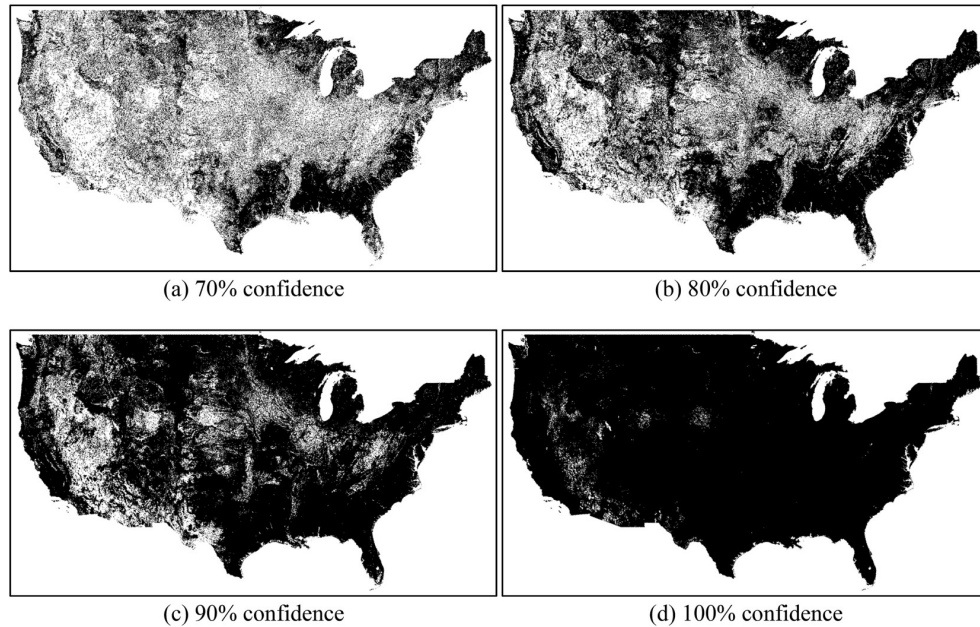


Fig. 6. Mapping high confidence pixels from 2014 to 2018.

over the Earth Engine Apps platform (<https://www.earthengine.app/>). As shown in Fig. 7, the graphic user interface includes a configuration panel and a map explorer. The app calls the specific modeling functions of the toolkit based on the user's choice and dynamically reload the on-demand results on the map explorer. With the Cropland Explorer (Fig. 7a), users can select the product layer, year, crop types, and boundary layer on the configuration panel. The prototype currently provides the crop area statistics and crop sequence statistics, which are created using *statisticsByROI* and *statisticsByPoi* functions of the *statistics* module. For a selected region (county, ASD, state) of interest and point of interest on the map explorer, the time series chart of changes of crop area for the selected region of interest and crop sequence for the point of interest will be plotted on the panel. Similarly, the Crop Frequency Explorer (Fig. 7b) will produce the crop frequency map according to the crop type and years of interest.

4. Discussion

4.1. Contribution of the study

The AgKit4EE toolkit makes the CONUS-scale agricultural land use modeling more effectively and efficiently. First of all, it remedies the limitation of CropScape and provides various ready-to-use modeling functions, such as crop sequence modeling, crop frequency modeling, and confidence layer modeling. These functions can be coupled with many modeling workflows. For example, the pixels derived from the common crop sequence can be potentially used as training samples to train the classification model, which would be a low-cost but reliable way to produce reference data for the early-season crop mapping over a large geographic area. The crop frequency map can provide information regarding the potential geospatial distribution of crop planting in the future. By modeling the confidence layer, the current CDL users can easily find out the trusted CDL pixels over time thus better assessing the modeling result.

Computing resource is critical to the performance of geospatial environmental modeling. Because of the limited computing capacity, the CONUS-scale modeling might take a few hours or days to run on a single server or personal computer. By taking advantage of GEE's powerful cloud computing environment, it takes only a few seconds to

process the CDL data for the entire CONUS. This will save a considerable amount of time for modelers. We believe that the environmental modeling community, LULC community, and agricultural sectors will be benefited from using GEE along with AgKit4EE.

4.2. Application scenarios

The AgKit4EE toolkit is designed for users who deal with CDL data on the GEE platform and GEE-enabled web applications. Here are some common application scenarios. First, all functions can be directly requested to visualize, explore, and export the on-demand CDL products through GEE Code Editor. Another application scenario is integrating the toolkit with other modeling workflow while CDL is one of the data sources. Moreover, the AgKit4EE library can be imported as JavaScript module into the web frameworks. In this way, the implementation of GEE-based web applications and geospatial CI can be significantly accelerated. Developers can just focus on the implementation of the high-level architecture without figuring out the details of GEE APIs and CDL data.

4.3. Data derived from CDL

The AgKit4EE toolkit is more than a GEE extension for CDL data visualization. It is characterized by many modeling functions and operations. The original CDL data can be fully utilized and various agricultural land use products can be derived using the toolkit. Table 4 compares the data offered by CropScape, GEE data catalog, and AgKit4EE. The original CDL products (cropland layer, confidence layer, and cultivated layer) can be directly exported using the *export* module. For example, user can start a job to batch download the cropland layer maps from 2010 to 2018 of corn and soybeans for Iowa by calling the *exportCDLByFips* function in the *export* module. All other products that derived from the CDL data (e.g., crop frequency map and crop sequence map) can be retrieved through the modeling functions summarized in Table 3 then exported to the local path with the GEE's built-in export functions. AgKit4EE is an open source software with a fully extensible structure. It is easy and free to extend the modules and develop more CDL-based agricultural land use products as needed.

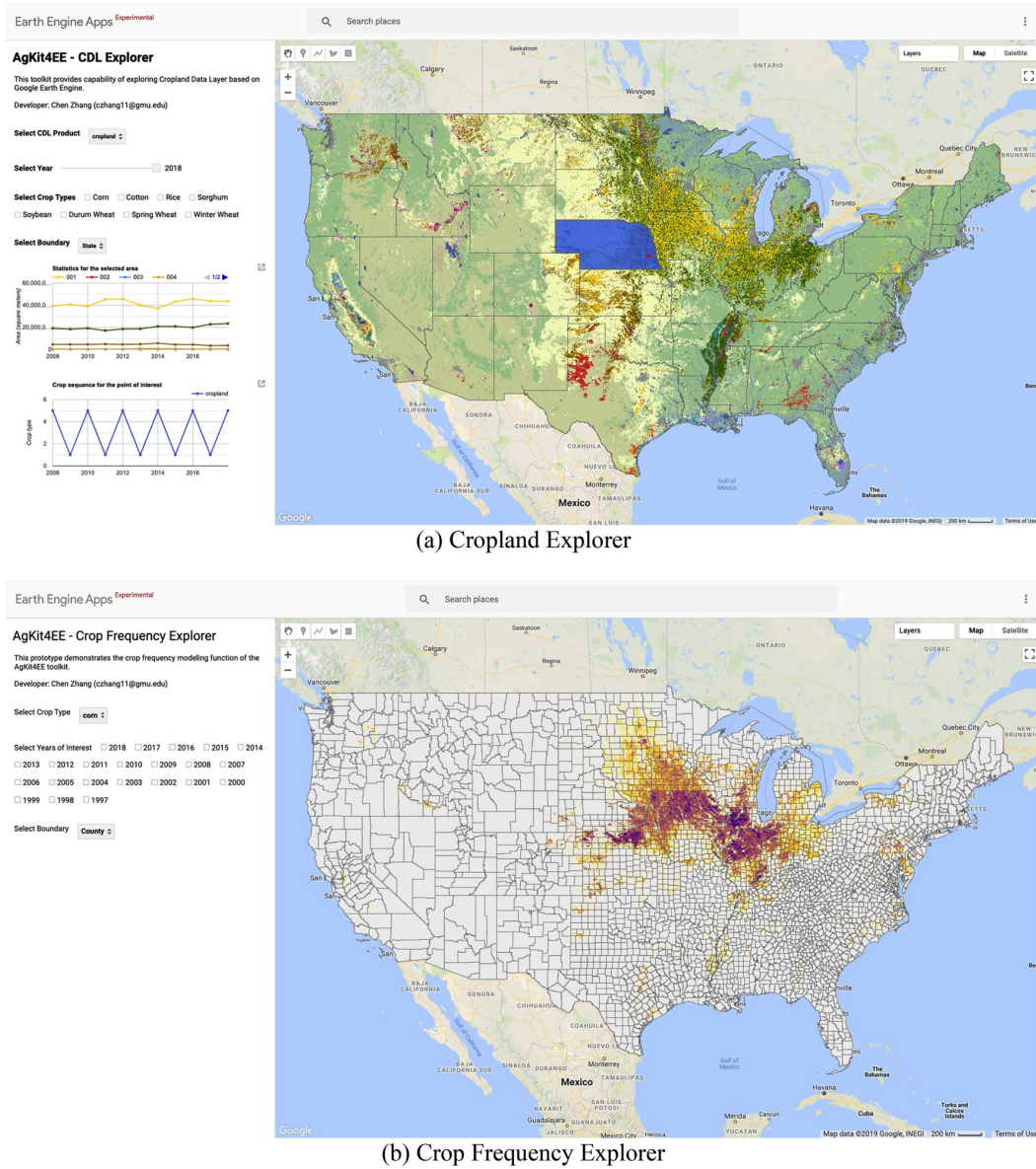


Fig. 7. The graphic user interface of the AgKit4EE-enabled web application prototypes.

Table 4
Comparison of CDL-based agricultural land use data from three sources.

Data	CropScape	GEE Data Catalog	AgKit4EE
Cropland layer	Available	Available	Available
Confidence layer	N/A	Available	Available
Cultivated layer	Available	Available	Available
Cropland map by crop sequence	N/A	N/A	Available
Crop frequency map	Available for major crop types (corn, soybean, wheat, and cotton)	N/A	Available for all crop types
Confidence pixel map by threshold	N/A	N/A	Available

4.4. Limitations

The current release of AgKit4EE still has some limitations. On the one hand, computing capacity of some geospatial functions are restricted by GEE. For example, we currently only support the county-level statistics due to the limited number of pixels per each process (10 million pixels) allowed by GEE. When scaling up to the state level or larger geographic area, the process will be stopped with the “too many pixels in the region” error. A solution to bypass the limitation is to increase the scale for the reducer operation. However, this tweak would affect the accuracy of the statistics result. A better solution is to sum up the county-level results to ASD or state level, but it would take more processing time. Moreover, the Earth Engine Apps platform does not support the data export function in the current phase. To export the modeling result, the user must run the toolkit through the GEE Code Editor. On the other hand, there were many misclassified pixels in the early-year CDL data because of

cloud cover and lack of satellite images. Also, the coverage of the early-year CDL was incomplete. Only a few states were fully covered before 2008. These quality and coverage issue of the early-year CDL data can potentially affect the follow-on studies (Zhang et al., 2020).

5. Conclusion and future works

This paper presents the design, implementation, and use examples of AgKit4EE. As a GEE-enabled toolkit for the CONUS-scale agricultural land use modeling, AgKit4EE contains a variety of frequently used functions for retrieving, visualizing, modeling, and analyzing CDL data. The major functions and capabilities of the proposed software, including crop sequence modeling, crop frequency modeling, confidence layer modeling, and geospatial statistics were demonstrated. Additionally, two AgKit4EE-enabled web applications, the Cropland Explorer and the Crop Frequency Explorer, were prototyped and published over the Earth Engine Apps platform. The result suggests this toolkit would greatly reduce the workload of modelers and developers who deal with CDL data.

In the next phase of development, we will enhance and extend the AgKit4EE toolkit, integrate more modeling functions, develop the machine learning module, and support more LULC data products such as the National Land Cover Database (NLCD) of U.S. Geological Survey (USGS). A geospatial CI with all features of the AgKit4EE toolkit is under development. Currently, the toolkit is implemented in JavaScript only. We will implement the core modules in Python to support more third-party applications developed with GEE Python API.

Software availability

Software Name: AgKit4EE.

Developer: Center for Spatial Information Science and Systems, George Mason University.

Technical support: Chen Zhang (czhang11@gmu.edu).

Programming language: JavaScript.

License: MIT.

Software required: Google Earth Engine.

Project: https://code.earthengine.google.com/?accept_repo=user/s/czhang11/agkit4ee.

Earth Engine repository: <https://earthengine.googleusercontent.com/users/czhang11/agkit4ee>.

Github repository: <https://github.com/czhang11/agkit4ee>.

Cropland Explorer: <https://czhang11.users.earthengine.app/view/agkit4ee-cdl-explorer>.

Crop Frequency Explorer: <https://czhang11.users.earthengine.app/view/agkit4ee-crop-frequency-explorer>.

Declaration of competing interest

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

Acknowledgments

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

Appendix A. Code Examples

```
// getCdlCollection function of AgKit4EE
getCdlCollection({product:'cropland', remap:[1,5], years:['1997','2007','2017']});

// Equivalent code using native GEE APIs
var cdl1997 = ee.Image('USDA/NASS/CDL/1997');
var cdl2007 = ee.Image('USDA/NASS/CDL/2007a').blend(ee.Image('USDA/NASS/CDL/2007b'));
var cdl2017 = ee.Image('USDA/NASS/CDL/2017');
var cdlCollection = ee.ImageCollection([cdl1997, cdl2007, cdl2017]);
var conditional = function(item) {
  var cdlImage = ee.Image(item);
  var cdlImage_cropland = cdlImage.select('cropland');
  return cdlImage_cropland.remap([1,5],[1,5]);
};
cdlCollection.map(conditional);
```

Fig. A1. Example of retrieving CDL image collection using AgKit4EE.

```
var corn_mono = croplandModeling.modelingByCropSequence([1,1,1,1,1]);
var corn_rotation = croplandModeling.modelingByCropSequence([1,5,1,5,1]);
var soybean_mono = croplandModeling.modelingByCropSequence([5,5,5,5,5]);
var soybean_rotation = croplandModeling.modelingByCropSequence([5,1,5,1,5]);
var soybean_rotation_alt = croplandModeling.modelingByCropSequence([5,3,5,3,5]);
var cotton_mono = croplandModeling.modelingByCropSequence([2,2,2,2,2]);
var rice_mono = croplandModeling.modelingByCropSequence([3,3,3,3,3]);
var rice_rotation = croplandModeling.modelingByCropSequence([3,5,3,5,3]);
var durumWheat_mono = croplandModeling.modelingByCropSequence([22,22,22,22,22]);
var springWheat_mono = croplandModeling.modelingByCropSequence([23,23,23,23,23]);
var winterWheat_mono = croplandModeling.modelingByCropSequence([24,24,24,24,24]);
```

Fig. A2. Mapping major crops based on common crop rotation patterns for the entire CONUS.

```

1 var freq_corn = croplandModeling.modelingFrequencyByCrop(1).clip(getRoi.getRoiByFips('38'))
2 var freq_soybean = croplandModeling.modelingFrequencyByCrop(5).clip(getRoi.getRoiByFips('38'))

```

Fig. A3. Mapping frequency of corn and soybeans for the state of Nebraska.

```

1 var _years = ['2014','2015','2016','2017','2018'];
2 var ar_corn = croplandModeling.modelingFrequencyByCrop(1, {years:_years}).clip(getRoi.getRoiByFips('05'));
3 var ar_cotton = croplandModeling.modelingFrequencyByCrop(2, {years:_years}).clip(getRoi.getRoiByFips('05'));
4 var ar_rice = croplandModeling.modelingFrequencyByCrop(3, {years:_years}).clip(getRoi.getRoiByFips('05'));
5 var ar_soy = croplandModeling.modelingFrequencyByCrop(5, {years:_years}).clip(getRoi.getRoiByFips('05'));
6 var ar_wht = croplandModeling.modelingFrequencyByCrop(24, {years:_years}).clip(getRoi.getRoiByFips('05'));
7 var ar_db1WhtSoy = croplandModeling.modelingFrequencyByCrop(26, {years:_years}).clip(getRoi.getRoiByFips('05'));

```

Fig. A4. Mapping crop frequency for the state of Arkansas in the recent five years.

```

1 var _years = ['2014','2015','2016','2017','2018'];
2 var conf_70 = confidenceModeling.modelingConfidenceByThreshold(70, {years:_years});
3 var conf_80 = confidenceModeling.modelingConfidenceByThreshold(80, {years:_years});
4 var conf_90 = confidenceModeling.modelingConfidenceByThreshold(90, {years:_years});
5 var conf_100 = confidenceModeling.modelingConfidenceByThreshold(100, {years:_years});

```

Fig. A4. Modeling confidence layer for the entire CONUS.

References

- Berzsenyi, Z., Györfi, B., Lap, D., 2000. Effect of crop rotation and fertilisation on maize and wheat yields and yield stability in a long-term experiment. *Eur. J. Agron.* 13, 225–244. [https://doi.org/10.1016/S1161-0301\(00\)00076-9](https://doi.org/10.1016/S1161-0301(00)00076-9).
- Bey, A., Díaz, A.S.P., Maniatis, D., Marchi, G., Mollicone, D., Ricci, S., Bastin, J.F., Moore, R., Federici, S., Rezende, M., Patriarca, C., Turia, R., Gamoga, G., Abe, H., Kaidong, E., Miceli, G., 2016. Collect earth: land use and land cover assessment through augmented visual interpretation. *Rem. Sens.* 8, 1–24. <https://doi.org/10.3390/rs8100807>.
- 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>.
- Boryan, C.G., Yang, Z., Willis, P., 2014. US geospatial crop frequency data layers. In: *Agro-Geoinformatics, 2014 3rd International Conference*. Presented at the International Conference on Agro-Geoinformatics. <https://doi.org/10.1109/Agro-Geoinformatics.2014.6910657>.
- Castranova, A.M., Goodall, J.L., Elag, M.M., 2013. Models as web services using the open geospatial consortium (OGC) web processing service (WPS) standard. *Environ. Model. Software* 41, 72–83. <https://doi.org/10.1016/j.envsoft.2012.11.010>.
- Chen, B., Xiao, X., Li, X., Pan, L., Doughty, R., Ma, J., Dong, J., Qin, Y., Zhao, B., Wu, Z., Sun, R., Lan, G., Xie, G., Clinton, N., Giri, C., 2017. A mangrove forest map of China in 2015: analysis of time series Landsat 7/8 and Sentinel-1A imagery in Google Earth Engine cloud computing platform. *ISPRS J. Photogrammetry Remote Sens.* 131, 104–120. <https://doi.org/10.1016/j.isprsjprs.2017.07.011>.
- Crookston, R.K., Kurl, J.E., Copeland, P.J., Ford, J.H., Lueschen, W.E., 1991. Rotational cropping sequence affects yield of corn and soybean. *Agron. J.* 83, 108–113. <https://doi.org/10.2134/agronj1991.00021962008300010026x>.
- Deng, M., Di, L., 2014. Building open environments to meet big data challenges in earth sciences. In: *Big Data*. <https://doi.org/10.1201/b16524-7>.
- Di, L., 2016. Big data and its applications in agro-geoinformatics. In: *2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. Presented at the 2016 IEEE International Geoscience and Remote Sensing Symposium. IGARSS, pp. 189–191. <https://doi.org/10.1109/IGARSS.2016.7729040>.
- Di, L., 2007. Geospatial sensor web and self-adaptive earth predictive systems (SEPS). In: *Proceedings of the Earth Science Technology Office (ESTO)/Advanced Information System Technology (AIST) Sensor Web Principal Investigator (PI) Meeting*, pp. 1–4. San Diego, USA.
- Di, L., Sun, Z., Yu, E., Zhang, C., Guo, L., 2019. CyberWay—An integrated geospatial cyberinfrastructure to facilitate innovative Way of Inter-and Multi-disciplinary Geoscience Studies. *Geophysical Research Abstracts* 21.
- Di, L., Sun, Z., Zhang, C., 2017. Facilitating the easy use of earth observation data in earth system models through CyberConnector. In: *AGU Fall Meeting Abstracts*.
- Di, L., Yu, E.G., Yang, Z., Shrestha, R., Kang, L., Zhang, B., Han, W., 2015. Remote sensing based crop growth stage estimation model. In: *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. Presented at the 2015 IEEE International Geoscience and Remote Sensing Symposium. IGARSS, pp. 2739–2742. <https://doi.org/10.1109/IGARSS.2015.7326380>.
- Edwards, J.H., Thurlow, D.L., Eason, J.T., 1988. Influence of tillage and crop rotation on yields of corn, soybean, and wheat. *Agron. J.* 80, 76–80. <https://doi.org/10.2134/agronj1988.00021962008000010018x>.
- Essawy, B.T., Goodall, J.L., Zell, W., Voce, D., Morsy, M.M., Sadler, J., Yuan, Z., Malik, T., 2018. Integrating scientific cyberinfrastructures to improve reproducibility in computational hydrology: example for HydroShare and GeoTrust. *Environ. Model. Software* 105, 217–229. <https://doi.org/10.1016/j.envsoft.2018.03.025>.
- Feng, Q., Chaubey, I., Her, Y.G., Cibir, R., Engel, B., Volenc, J., Wang, X., 2015. Hydrologic and water quality impacts and biomass production potential on marginal land. *Environ. Model. Software* 72, 230–238. <https://doi.org/10.1016/j.envsoft.2015.07.004>.
- Feng, Q., Engel, B.A., Flanagan, D.C., Huang, C., Yen, H., Yang, L., 2019. Design and development of a web-based interface for the Agricultural Policy Environmental eXtender (APEX) model. *Environ. Model. Software* 111, 368–374. <https://doi.org/10.1016/j.envsoft.2018.09.011>.
- Feng, S., Hu, Q., 2004. Changes in agro-meteorological indicators in the contiguous United States: 1951–2000. *Theor. Appl. Climatol.* 78, 247–264. <https://doi.org/10.1007/s00704-004-0061-8>.
- Flynn, K.C., 2019. Site suitability analysis for tef (*Eragrostis tef*) within the contiguous United States. *Comput. Electron. Agric.* 159, 119–128. <https://doi.org/10.1016/j.compag.2019.02.016>.
- Goodall, J.L., Saint, K.D., Ercan, M.B., Briley, L.J., Murphy, S., You, H., DeLuca, C., Rood, R.B., 2013. Coupling climate and hydrological models: interoperability through web services. *Environ. Model. Software* 46, 250–259. <https://doi.org/10.1016/j.envsoft.2013.03.019>.
- Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., 2017. Google earth engine: planetary-scale geospatial analysis for everyone. *Remote Sens. Environ.* 202, 18–27. <https://doi.org/10.1016/j.rse.2017.06.031>.
- Groff, S.C., Loftin, C.S., Drummond, F., Bushmann, S., McGill, B., 2016. Parameterization of the InVEST Crop Pollination Model to spatially predict abundance of wild blueberry (*Vaccinium angustifolium* Aiton) native bee pollinators in Maine, USA. *Environ. Model. Software* 79, 1–9. <https://doi.org/10.1016/j.envsoft.2016.01.003>.
- Han, W., Yang, Z., Di, L., Mueller, R., 2012. CropScape: a Web service based application for exploring and disseminating US continuous geospatial cropland data products for decision support. *Comput. Electron. Agric.* 84, 111–123. <https://doi.org/10.1016/j.compag.2012.03.005>.
- Han, W., Yang, Z., Di, L., Zhang, B., Peng, C., 2014. Enhancing agricultural geospatial data dissemination and applications using geospatial web services. *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.* 7, 4539–4547. <https://doi.org/10.1109/JSTARS.2014.2315593>.
- Hansen, M.C.C., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A. a, Tyukavina, A., Thau, D., Stehman, S.V.V., Goetz, S.J.J., Loveland, T.R.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O.O., Townshend, J.R.G.R.G., Potapov, P.V., Moore, R., Hancher, M., Turubanova, S.A. a, Tyukavina, A., Thau, D., Stehman, S.V.V., Goetz, S.J.J., Loveland, T.R.R., Kommareddy, A., Egorov, A., Chini, L., Justice, C.O.O., Townshend, J.R.G.R.G., 2013. High-resolution global maps of 21st-Century forest cover change. *Science* 342, 850–854. <https://doi.org/10.1126/science.1244693>.
- Hird, J.N., DeLancey, E.R., McDermid, G.J., Kariyeva, J., 2017. Google earth engine, open-access satellite data, and machine learning in support of large-area probabilistic wetland mapping. *Rem. Sens.* 9, 1315. <https://doi.org/10.3390/rs9121315>.
- Huang, H., Chen, Y., Clinton, N., Wang, J., Wang, X., Liu, C., Gong, P., Yang, J., Bai, Y., Zheng, Y., Zhu, Z., 2017. Mapping major land cover dynamics in Beijing using all Landsat images in Google Earth Engine. *Remote Sens. Environ.* 202, 166–176. <https://doi.org/10.1016/j.rse.2017.02.021>.
- Janzen, H.H., Campbell, C.A., Brandt, S.A., Lafond, G.P., Townley-Smith, L., 1992. Light-fraction organic matter in soils from long-term crop rotations. *Soil Sci. Soc. Am. J.* 56, 1799–1806. <https://doi.org/10.2136/sssaj1992.03615995005600060025x>.
- Karlen, D.L., Hurlley, E.G., Andrews, S.S., Cambardella, C.A., Meek, D.W., Duffy, M.D., Mallarino, A.P., 2006. Crop rotation effects on soil quality at three Northern corn/

- soybean belt locations. *Agron. J.* 98, 484–495. <https://doi.org/10.2134/agronj2005.0098>.
- Katz, A., 2015. *EarthCube: A Community-Driven Organization for Geoscience Cyberinfrastructure*.
- Kerkez, B., Daniels, M., Graves, S., Chandrasekar, V., Keiser, K., Martin, C., Dye, M., Maskey, M., Vernon, F., 2016. Cloud hosted real-time data services for the geosciences (CHORDS). *Geosci. Data J.* 3, 4–8. <https://doi.org/10.1002/gdj3.36>.
- Khalsa, S.J., 2017. Data and metadata brokering – theory and practice from the BCube project. *Data Sci. J.* 16, 1. <https://doi.org/10.5334/dsj-2017-001>.
- Koskinen, J., Leinonen, U., Vollrath, A., Ortmann, A., Lindquist, E., d'Annunzio, R., Pekkarinen, A., Käyhkö, N., 2019. Participatory mapping of forest plantations with open foris and Google earth engine. *ISPRS J. Photogrammetry Remote Sens.* 148, 63–74. <https://doi.org/10.1016/j.isprsjprs.2018.12.011>.
- Krisnadhi, A., Hu, Y., Janowicz, K., Hitzler, P., Arko, R., Carbotte, S., Chandler, C., Cheatham, M., Fils, D., Finin, T., Ji, P., Jones, M., Karima, N., Lehnert, K., Mickle, A., Narock, T., O'Brien, M., Raymond, L., Shepherd, A., Schildhauer, M., Wiebe, P., 2015. The GeoLink modular oceanography ontology. In: Arenas, M., Corcho, O., Simperl, E., Strohmaier, M., d'Aquin, M., Srinivas, K., Groth, P., Dumontier, M., Hefflin, J., Thirunarayan, K., Staab, S. (Eds.), *The Semantic Web - ISWC 2015, Lecture Notes in Computer Science*. Springer International Publishing, pp. 301–309.
- Lee, C.A., Gasser, S.D., Plaza, A., Chang, C.-I., Huang, B., 2011. Recent developments in high performance computing for remote sensing: a review. *IEEE J. Sel. Topics Appl. Earth Obs. Remote Sens.* 4, 508–527. <https://doi.org/10.1109/JSTARS.2011.2162643>.
- Li, H., Wan, W., Fang, Y., Zhu, S., Chen, X., Liu, B., Hong, Y., 2019. A Google Earth Engine-enabled software for efficiently generating high-quality user-ready Landsat mosaic images. *Environ. Model. Software* 112, 16–22. <https://doi.org/10.1016/j.envsoft.2018.11.004>.
- Li, J., Mahalov, A., Hyde, P., 2016. Impacts of agricultural irrigation on ozone concentrations in the Central Valley of California and in the contiguous United States based on WRF-Chem simulations. *Agric. For. Meteorol.* 221, 34–49. <https://doi.org/10.1016/j.agrformet.2016.02.004>.
- Lin, L., Di, L., Zhang, C., Hu, L., Tang, J., Yu, E.G., 2017. Developing a Web Service Based Application for Demographic Information Modeling and Analyzing. 2017 6th International Conference on Agro-Geoinformatics. <https://doi.org/10.1109/Agro-GeoInformatics.2017.8047069>.
- Lin, T., Wang, S., Rodríguez, L.F., Hu, H., Liu, Y., 2015. CyberGIS-enabled decision support platform for biomass supply chain optimization. *Environ. Model. Software* 70, 138–148. <https://doi.org/10.1016/j.envsoft.2015.03.018>.
- Liu, C.-C., Shieh, M.-C., Ke, M.-S., Wang, K.-H., Liu, C.-C., Shieh, M.-C., Ke, M.-S., Wang, K.-H., 2018. Flood prevention and emergency response system powered by Google earth engine. *Rem. Sens.* 10, 1283. <https://doi.org/10.3390/rs10081283>.
- Liu, W., Gopal, S., Woodcock, C.E., 2004. Uncertainty and confidence in land cover classification using a hybrid classifier approach. *Photogramm. Eng. Rem. Sens.* 70, 963–971. <https://doi.org/10.14358/PERS.70.8.963>.
- McCarthy, J.L., Korontzi, S., Justice, C.O., Loboda, T., 2009. The spatial and temporal distribution of crop residue burning in the contiguous United States. *Sci. Total Environ.* 407, 5701–5712. <https://doi.org/10.1016/j.scitotenv.2009.07.009>.
- McNider, R.T., Handyside, C., Doty, K., Ellenburg, W.L., Cruise, J.F., Christy, J.R., Moss, D., Sharda, V., Hoogenboom, G., Caldwell, P., 2015. An integrated crop and hydrologic modeling system to estimate hydrologic impacts of crop irrigation demands. *Environ. Model. Software* 72, 341–355. <https://doi.org/10.1016/j.envsoft.2014.10.009>.
- Midekisa, A., Holl, F., Savory, D.J., Andrade-Pacheco, R., Gething, P.W., Bennett, A., Sturrock, H.J.W., 2017. Mapping land cover change over continental Africa using Landsat and Google Earth Engine cloud computing. *PLoS One* 12, e0184926. <https://doi.org/10.1371/journal.pone.0184926>.
- Mueller, R., Harris, M., 2013. Reported uses of CropScape and the National cropland data layer program. In: *Proceedings of the International Conference on Agricultural Statistics VI*. Presented at the Sixth International Conference on Agricultural Statistics, p. 9.
- Nativi, S., Mazzetti, P., Geller, G.N., 2013. Environmental model access and interoperability: the GEO Model Web initiative. *Environ. Model. Software* 39, 214–228. <https://doi.org/10.1016/j.envsoft.2012.03.007>.
- Padarian, J., Minasny, B., McBratney, A.B., 2015. Using Google's cloud-based platform for digital soil mapping. *Comput. Geosci.* 83, 80–88. <https://doi.org/10.1016/j.cageo.2015.06.023>.
- Padmanabhan, A., Wang, S., Cao, G., Hwang, M., Zhang, Z., Gao, Y., Soltani, K., Liu, Y., 2014. FluMapper: a cyberGIS application for interactive analysis of massive location-based social media. *Concurrency Comput. Pract. Ex.* 26, 2253–2265. <https://doi.org/10.1002/cpe.3287>.
- Santoro, M., Nativi, S., Mazzetti, P., 2016. Contributing to the GEO Model Web implementation: a brokering service for business processes. *Environ. Model. Software* 84, 18–34. <https://doi.org/10.1016/j.envsoft.2016.06.010>.
- Shelestov, A., Lavreniuk, M., Kussul, N., Novikov, A., Skakun, S., 2017. Exploring Google earth engine platform for big data processing: classification of multi-temporal satellite imagery for crop mapping. *Front. Earth Sci.* 5, 1–10. <https://doi.org/10.3389/feart.2017.00017>.
- Stern, A.J., Doraiswamy, P.C., Raymond Hunt, E., 2012. Changes of crop rotation in Iowa determined from the United States Department of Agriculture, National Agricultural Statistics Service cropland data layer product. *J. Appl. Remote Sens.* 6, 063590. <https://doi.org/10.1117/1.JRS.6.063590>.
- Sun, Z., Di, L., Hao, H., Wu, X., Tong, D.Q., Zhang, C., Virgei, C., Fang, H., Yu, E.G., Tan, X., Yue, P., Lin, L., 2017. CyberConnector: a service-oriented system for automatically tailoring multisource Earth observation data to feed Earth science models. *Earth Sci. Inf.* 1–17. <https://doi.org/10.1007/s12145-017-0308-4>.
- Sun, Z., Di, L., Zhang, C., Fang, H., Yu, E., Lin, L., et al., 2017. Building robust geospatial web services for agricultural information extraction and sharing. 2017 6th International Conference on Agro-Geoinformatics. <https://doi.org/10.1109/Agro-GeoInformatics.2017.8047055>.
- Tasdighi, A., Arabi, M., Harmel, D., Line, D., 2018. A Bayesian total uncertainty analysis framework for assessment of management practices using watershed models. *Environ. Model. Software* 108, 240–252. <https://doi.org/10.1016/j.envsoft.2018.08.006>.
- Teluguntla, P., Thenkabail, P.S., Oliphant, A., Xiong, J., Gumma, M.K., Congalton, R.G., Yadav, K., Huete, A., 2018. A 30-m landsat-derived cropland extent product of Australia and China using random forest machine learning algorithm on Google Earth Engine cloud computing platform. *ISPRS J. Photogrammetry Remote Sens.* 144, 325–340. <https://doi.org/10.1016/j.isprsjprs.2018.07.017>.
- Triberti, L., Nastri, A., Baldoni, G., 2016. Long-term effects of crop rotation, manure and mineral fertilisation on carbon sequestration and soil fertility. *Eur. J. Agron.* 74, 47–55. <https://doi.org/10.1016/j.eja.2015.11.024>.
- Van Eerd, L.L., Congreves, K.A., Hayes, A., Verhallen, A., Hooker, D.C., 2014. Long-term tillage and crop rotation effects on soil quality, organic carbon, and total nitrogen. *Can. J. Soil Sci.* 94, 303–315. <https://doi.org/10.4141/cjss2013-093>.
- Vitolo, C., Elkhatib, Y., Reusser, D., Macleod, C.J.A., Buytaert, W., 2015. Web technologies for environmental big data. *Environ. Model. Software* 63, 185–198. <https://doi.org/10.1016/j.envsoft.2014.10.007>.
- Vos, K., Splinter, K.D., Harley, M.D., Simmons, J.A., Turner, I.L., 2019. CoastSat: a Google Earth Engine-enabled Python toolkit to extract shorelines from publicly available satellite imagery. *Environ. Model. Software* 122, 104528. <https://doi.org/10.1016/j.envsoft.2019.104528>.
- Wang, S., 2010. A CyberGIS framework for the synthesis of cyberinfrastructure, GIS, and spatial analysis. *Ann. Assoc. Am. Geogr.* 100, 535–557. <https://doi.org/10.1080/00045601003791243>.
- Wang, S., Anselin, L., Bhaduri, B., Crosby, C., Goodchild, M.F., Liu, Y., Nyerges, T.L., 2013. CyberGIS software: a synthetic review and integration roadmap. *Int. J. Geogr. Inf. Sci.* 27, 2122–2145. <https://doi.org/10.1080/13658816.2013.776049>.
- Xue, Z., Couch, A., Tarboton, D., 2019. Map based discovery of hydrologic data in the HydroShare collaboration environment. *Environ. Model. Software* 111, 24–33. <https://doi.org/10.1016/j.envsoft.2018.09.014>.
- Yalew, S.G., van Griensven, A., van der Zaag, P., 2016. AgriSuit: a web-based GIS-MCDA framework for agricultural land suitability assessment. *Comput. Electron. Agric.* 128, 1–8. <https://doi.org/10.1016/j.compag.2016.08.008>.
- Yang, C., Huang, Q., 2013. *Spatial Cloud Computing: A Practical Approach*. CRC Press.
- Yang, C., Raskin, R., Goodchild, M., Gahegan, M., 2010. Geospatial cyberinfrastructure: past, present and future. *Comput. Environ. Urban Syst.* 34, 264–277. <https://doi.org/10.1016/j.compenvurbysys.2010.04.001>.
- Yang, Z., Li, W., Chen, Q., Wu, S., Liu, S., Gong, J., 2018. A scalable cyberinfrastructure and cloud computing platform for forest aboveground biomass estimation based on the Google Earth Engine. *Int. J. Digit. Earth* 1–18. <https://doi.org/10.1080/17538947.2018.1494761>.
- Yu, Z., Di, L., Tang, J., Yu, E.G., Rahman, MdS., Zhang, C., 2018. *Land Use/Land Cover Classification and Change Analysis for Ganges River Basin from 2000 to 2010*. In: *AGU Fall Meeting Abstracts*, pp. B23J-B2661.
- Yue, P., Baumann, P., Bugbee, K., Jiang, L., 2015. Towards intelligent GIServices. *Earth Sci. Inf.* 1373–1376. <https://doi.org/10.1109/IGARSS.2015.7326032>, 2015-Novem.
- Zhang, C., Di, L., Lin, L., Guo, L., 2019a. Machine-learned prediction of annual crop planting in the U.S. Corn Belt based on historical crop planting maps. *Comput. Electron. Agric.* 166, 104989. <https://doi.org/10.1016/j.compag.2019.104989>.
- Zhang, C., Di, L., Lin, L., Guo, L., 2019b. Extracting trusted pixels from historical cropland data layer using crop rotation patterns: a case study in Nebraska. In: 2019 8th International Conference on Agro-Geoinformatics (Agro-Geoinformatics). Presented at the 2019 8th International Conference on Agro-Geoinformatics (Agro-Geoinformatics), pp. 1–6. <https://doi.org/10.1109/Agro-GeoInformatics.2019.8820236>. USA.
- Zhang, C., Di, L., Sun, Z., Lin, L., Yu, E.G., Gaigalas, J., 2019c. Exploring cloud-based web processing service: a case study on the implementation of CMAQ as a service. *Environ. Model. Software* 113, 29–41. <https://doi.org/10.1016/j.envsoft.2018.11.019>.
- Zhang, C., Di, L., Sun, Z., Yu, E.G., Hu, L., Lin, L., Tang, J., Rahman, MdS., 2017. Integrating OGC web processing service with cloud computing environment for earth observation data. In: 2017 6th International Conference on Agro-Geoinformatics. Presented at the International Conference on Agro-Geoinformatics. <https://doi.org/10.1109/Agro-GeoInformatics.2017.8047065>.
- Zhang, C., Di, L., Yang, Z., Lin, L., Yu, E.G., Yu, Z., Rahman, M.S., Zhao, H., 2019d. Cloud environment for disseminating NASS cropland data layer. In: 2019 8th International Conference on Agro-Geoinformatics (Agro-Geoinformatics). Presented at the 2019 8th International Conference on Agro-Geoinformatics (Agro-Geoinformatics), pp. 1–5. <https://doi.org/10.1109/Agro-GeoInformatics.2019.8820465>.
- Zhang, C., Yang, Z., Di, L., Lin, L., Hao, P., 2020. Refinement of Cropland Data Layer Using Machine Learning. *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences* 42, 161–164. <https://doi.org/10.5194/isprs-archives-XLII-3-W11-161-2020>.
- Zhao, P., Di, L., 2010. Geospatial web services: advances in information interoperability. *IGI Global*. <https://doi.org/10.4018/978-1-60960-192-8>.