Every Field in Minnesota: Building a Geographically Scalable Satellite Imagery Analytics System for Mapping Crop Fields

A Thesis
SUBMITTED TO THE FACULTY OF THE
UNIVERSITY OF MINNESOTA
BY

Jesse Bakker

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
MASTER OF ARTS

Dr. Eric Shook, Adviser

August 2020

Acknowledgements

I could not have done this project alone and I would like to express my deep gratitude to all those who helped make the research possible and support me through its completion.

This research was made possible through generous funding from the Center for Urban and Regional Affairs via the Faculty Interactive Research Program. Thanks to Will Craig and the entire CURA team.

I would like to thank my adviser Dr. Eric Shook for conceiving the project and mentoring me through the entire graduate school process. I am grateful to my committee members, Kathryn Grace and Paul West, who provided guidance and feedback.

Special thanks to the Institute on the Environment, Global Landscapes Initiative Group: Jamie Gerber, Deepak Ray, and Paul West; Jeff Thompson and David Porter at the Minnesota Supercomputing Institute; and the CyberGIS Lab Group: Bryan Runck, Coleman Shepard, Jasper Johnson, Kuan-Jui (Ray) Lin, Mohsen Ahmadkhani, Ali Hossaini, Audrey Gasser.

Finally, I could not have done it without the support of all my friends and family, in particular my parents John and Jeanine, my brother Sam, and my partner Mary.

Table of Contents

List of Figures

List of Tables

List of Equations

Chapter 1 – Introduction

Using Data to Monitor a Changing World

Why do we need to map crop field boundaries?

The planet will face unprecedented challenges in the 21st century. Climate change, population growth, and continued economic development will put immense pressures on both human and environmental systems. By 2050, it is estimated that annual agricultural production will have to increase by 50% to feed the earth's 9 billion people (Fritz et al. 2013). Not only will food production need to increase, but changes in climate systems will make agriculture less predictable around the planet. Climate change will put additional stress on agricultural systems. According to the Intergovernmental Panel on Climate Change (IPCC), as population grows, food production will have to increase amidst the "increasing temperatures, changing precipitation patterns, and greater frequency of...extreme events" due to climate change (IPCC 2019). These environmental pressures will lead to increased food insecurity, particularly in the developing world where there are not already robust agricultural monitoring systems (See et al. 2015).

Understanding where cropland is around the world is critical for the study of food security, agricultural systems, land-use change, water management, and other 21st century challenges (Thenkabail et al. 2010). For areas at greatest risk for food insecurity, particularly in lower income countries that will likely be impacted most severely from climate change and have the most limited resources to target solutions, irregular climate patterns and limited resources for cultivation can lead to drastic differences in agricultural production from year to year. Many countries facing these environmental risks also often have limited infrastructure for collecting critical data to aid in monitoring efforts (Grace et al 2014). Advanced monitoring can improve the response to food shortages at the local scale. High resolution field-level data also enables the integration of local knowledge to shape intervention. Timely, high resolution analysis of cropland area through remote sensing can overcome some of these limitations and aid decision making around food insecurity (Fritz et al. 2013). Utilizing the vast amounts of remote sensing data and computational resources now available, we can realize the Digital Earth envisioned 20 years ago to help solve these emerging global crises (Gore 1998). High resolution crop mapping is needed everywhere to address these growing global challenges (Maciel et al. 2019).

Crop fields are the operational unit of agriculture. Planting, cultivating, and harvesting decisions are all made at the scale of the field. Understanding where and how food production happens in real time, possible only if field locations are known, is a critical tool for informing strategic decision making to intervene in the kinds of agricultural disruptions that are inevitable with a changing climate (Waldner et al. 2015). Remote sensing approaches can fill in data gaps in places where robust agricultural data infrastructure is lacking around the world and form the foundation of higher-level agricultural data products, such as crop classification, yield prediction, and environmental impact modeling (Xiong et al. 2017). Research has demonstrated that incorporating field boundaries improved crop yield estimation, critical for rapid responses to shortages (Lambert et al. 2018). A global map of crop fields would enable local scale analysis of food production for the entire planet (See et al. 2015; Lesiv et al. 2019).

Agricultural fields are different around the world. Farming practices, field size, crop types, soil health, and climate regimes vary drastically across the planet (Lesiv et al. 2019). Capturing this field scale variation globally is not a trivial scientific undertaking. Remote sensing is an effective tool for monitoring global land cover and mapping cropland but has traditionally been constrained by the limitations of scale. Broadly, remote sensing methods either analyze large areas with generalized, coarse resolution outputs or cover small spatial areas to produce highly accurate, finely tuned outputs (Waldner et al. 2015). Recent remote sensing research has started to overcome this limitation, achieving relatively high-resolution field parcel mapping at the regional and continental scales, suggesting global field mapping will soon be possible (Graesser and Ramankutty 2017; Yan and Roy 2016). The development of more high resolution imaging platforms, such as the European Space Agency's (ESA) Sentinel-2 program with 10-meter spatial resolution, and the growth of High Performance Computing (HPC) resources provide the data and processing infrastructure necessary to enable field-scale mapping at global extents. The challenge now is to build an analytical framework to extract fields from the imagery for the entire planet.

Constraints: Computation vs. Accuracy

Karpatne et al (2016) describe the two key constraints for monitoring land cover with remote sensing data at a global scale that apply directly to the field mapping challenge: computation and accuracy. These factors determine the scalability—both computational and spatial—of an analytical approach. Traditionally, these constraints have imposed a tradeoff between the size of a study area and the spatial accuracy of the analysis. However, emerging

sources of satellite imagery and computational infrastructure are enabling new approaches to break through this tradeoff boundary.

Imagery resolution determines the geographic area that can be processed efficiently from a given data set. Computationally, it takes the same amount of processing to analyze a large area of course resolution imagery data as it does a small area of high-resolution imagery. To use three common public imagery sources as an example, a single pixel from the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor at 1-km resolution contains approximately 1,111 pixels of 30-m imagery from the National Aeronautics and Space Administration's (NASA) Landsat program and 10,000 pixels of 10-m imagery from the ESA's Sentinel-2 program. Conducting field-scale analysis with high resolution data at global extent greatly expands the amount of data and computation needed compared to existing coarse resolution global land cover maps.

Accuracy is the other barrier. Generalization is necessary for methods to be geographically transferrable, but high spatial accuracy in the output requires dealing with local variability. By taking generalized characteristics of a relatively large area, coarse resolution imagery helps to alleviate the problem of local heterogeneity, which can show up as natural variation in neighboring high-resolution pixels (Karpatne et al. 2016). Global estimates of cropland typically use coarse resolution imagery and provide a per-pixel value of the probability of the presence of agriculture activity (Ramankutty et al. 2008; Waldner et al. 2016). At the field scale, crop classification and field delineation research that utilizes high resolution imagery is constrained to cover relatively small study areas, relying on other locally tuned methods and data sources (Lebourgeois et al. 2017; Wit and Clevers 2004; Forkuor et al. 2014). Achieving field-scale accuracy over large areas requires flexible methods that are not fine-tuned to a single location or set of training data. Thus, the fine-tuning of methods and reliance on local training data makes many of those techniques unsuitable to scaling up spatially.

This trade-off between computation and spatial accuracy compels much of the current research in the remote sensing land cover research to conduct either coarse, generalizable, global-scale or accurate, fine-tuned, spatially constrained analysis (Waldner, Canto, and Defourny 2015). Global methods are not suited to handle local variability and local methods are not scalable to global extents. To adequately address 21st century challenges, tools like remote sensing data are critical for bridging the gap between global challenges and local solutions. Addressing this trade-off between covering large geographic extents and effectively accounting for the complexities of local variation in agricultural landscapes demands novel approaches to crop mapping.

Connecting global challenges and local solutions with new approaches

Current global scale agricultural data are generally modeled at coarse spatial resolution or collected via census at local administrative units. Many existing global extent cropland and land cover datasets have spatial resolutions ranging from 300-m to 1-km or up to 10-km (Pérez-Hoyos et al. 2017). Limited by this coarse spatial resolution, local variation cannot be captured, especially in places with small-holder farms or across landscapes where cropland is spatially fragmented with other land cover types and vegetation (Lambert et al. 2018; Aguilar et al. 2018). Alternatively, many agricultural data sets are collected via survey and aggregated to census statistics at the national or sub-national level (Xiong et al. 2017). These data are useful for understanding large scale trends, but they are not specific enough to utilize for local spatial analysis and can mask significant sub-national variation in agricultural production. Grace et al (2014) found that "it is a common occurrence to observe notable food production failures in one area of a country while another area experiences average or even above average yields." Additionally, while many developed countries have robust methods for collecting this data such as the National Agricultural Statistics Service (NASS) program of the U.S. Department of Agriculture (USDA), many developing countries facing the highest levels of food insecurity do not have the systems in place to collect such data (Boryan et al. 2011). Spatially generalized cropland mapping can miss important indicators of agricultural production at local scales.

Local level analysis that achieves high levels of accuracy is often finely tuned to the conditions of the study area (White and Roy 2015). In order to achieve the desired granularity of analysis, the geographic extent shrinks so that analysis can be specialized to account for local variability (Waldhoff, Lussem, and Bareth 2017). Often, the limiting factor is training data, which is scarce and time-intensive to create (Belgiu and Csillik 2018; Mahdianpari et al. 2019; Traganos et al. 2018; Debats et al. 2016; Wit and Clevers 2004), or cropland mask information, which is limited by national-level data infrastructure (Yan and Roy 2014; North, Pairman, and Belliss 2019). Imagery resolution can also be a limitation. Landsat data, one of the most widely used imagery sources due to its global coverage, long history, and accessibility, has 30-m spatial resolution, which is too coarse for mapping field boundaries in many places around the world (Fritz et al. 2015). While local scale studies are useful for offering insight into achieving field-scale cropland analysis, these factors can also limit the generalizability of the research methods and findings, anchoring them to site-specific data.

Limitations of a single model approach

A successful approach to global field mapping must include local knowledge as an input. Husak and Grace (2016) found that a single set of variables is not adequate to create a cropland mapping model that spans multiple countries, stating, “there is no such quantifiable universal indicator of cropped area.” Even differences over small distances at the sub-national scale can result in significant changes in physical terrain, climate, growing characteristics, and social practices. These differences suggest that a single set of descriptive data cannot be used to accurately estimate cropland area across large, diverse regions. A single model is not viable for mapping fields everywhere.

A truly global field mapping approach must be flexible to accommodate the local conditions of different spatial and agricultural contexts. Data inputs, like temporal seasonality, spectral signatures of crop species, and the spatial characteristics of fields, must be responsive to the geographic context (Grace et al 2014). Local knowledge about farming practices (irrigation vs. rain-fed), land use traditions (not all regions suitable for agriculture are used for agriculture), season growing patterns, and environmental anomalies in rainfall (like drought or flooding events) are all important indicators for modeling cropland in different geographies. Thus, a global analytical framework needs to both incorporate features from the data itself, through techniques like unsupervised learning, and integrate local knowledge of agricultural practices to be responsive to local context. While a single global model cannot be accurate everywhere, by using local knowledge to inform the methods along with data-driven techniques responsive to conditions within the local spatial context, the limitations of a universal method and reliance on training data can be somewhat overcome.

In a global extent mapping project, local variability and spatial heterogeneity demand that methods be flexible enough to be transferred across geographies. There are three models for this, each with its own set of limitations. The first approach, based on existing generalized global methods, is to build a single universal model that can deal with local and regional variability, then applying it at every location. The critical challenge here is that variability at 10-m is more difficult to generalize than at 1-km and the indicators necessary to map crop fields vary enough across geographies that a single set of global inputs is inadequate (Husak and Grace 2016). The second approach is to scale local analysis up by finding a way to apply local-scale tuning to a global-scale model, which is generally constrained by training data. While there are some universal characteristics about fields that can inform what aspects a field mapping model should take into account—like temporal features and vegetative indices—the local characteristics of a place (e.g., crop, geography, climate, practices) must inform the model. Generating training data

by hand for each locale is not scalable. The third approach—the solution proposed in this work—is to build a flexible analytical framework where unsupervised methods can extract relevant features from the spatio-temporal data and parameters are informed by local knowledge. Instead of relying on a single model or a single set of training data, we construct a workflow that uses fundamental knowledge of spectral and temporal characteristics of cropland to learn from the data and local agricultural practices to set the parameter bounds for deploying the methods. Within this workflow we can take advantage of multiple methods that are data driven and geographically flexible. By proposing a modular type of framework instead of a strict set of methods, our approach enables geographic flexibility, methodological modularity, and computational efficiency. We test this approach by mapping every crop field in Minnesota.

Building and Testing a Scalable Crop Mapping Workflow

Research questions

This thesis research contributes to a larger motivating research question: can we use satellite imagery and geospatial computing to map crop fields with high precision around the world? Since the ultimate goal is to map crop fields in all regions of the planet, the analytical workflow is constrained to using generalizable knowledge about cropland characteristics in multi-temporal satellite imagery and generalizable techniques for extracting those features. Thus, the geographic challenge is not simply to produce field boundaries with high spatial accuracy for the selected study area of Minnesota, but to construct a workflow that is not completely dependent upon hand-derived training data but derived from the data itself and local knowledge of agricultural practices in each locale. Further, the workflow aims to operate at global scale so it must be computationally optimized, .designed to efficiency reduce raw imagery into useful model inputs and process those data to extract field features. The core research objective then becomes constructing a field mapping workflow with enough flexibility to overcome the challenge of spatial heterogeneity inherent in a global scale analysis while being computationally efficient enough to handle global scale data at field scale resolution. To achieve this objective, we must first address three important questions, which are the focus of this thesis.

1. What information can be derived from raw high resolution, multispectral, multitemporal remote sensing imagery that is effective for mapping fields in the state of Minnesota?
2. What generalizable, unsupervised analytical methods can identify cropland and delineate individual fields accurately without relying on ground sampled training data?

3. How can these processes be combined into a geographically scalable, computationally efficient workflow that has the capacity to be deployed in agricultural regions around the world?

Exploring these questions within the variety and geographic scale of Minnesota's landscape will test both the constraints of computation and accuracy.

To answer these three research questions, this thesis will design and develop an analytical workflow to extract crop field boundary polygons from high resolution satellite imagery to map every crop field in the state of Minnesota. Through producing a map of Minnesota crop fields we will develop a scalable prototype of a field mapping workflow, demonstrate the feasibility of such an undertaking, and help identify the core computational and methodological challenges to be overcome in order to scale this project globally. By utilizing globally available 10-m imagery form the Sentinel-2 program with geospatial cyberinfrastructure and high-performance computing methods, we can close the gap between global extent and local analysis. With Minnesota as the study area, we will explore what generalizable characteristics of crop fields in multi-spectral, multi-temporal imagery data are useful for field delineation and gauge the feasibility of designing a robust, flexible field mapping framework that does not rely on large amounts of supervised training data. By understanding these challenges, we can propose a set of methods to create field boundary polygons for the agricultural conditions of Minnesota and describe some of the limitations that must be overcome in order to translate this into a generalizable workflow that can process field-scale data at a global scale with flexibility and efficiency.

Why Minnesota?

The combination of relatively large field sizes, well-established agricultural practices, diversity of land cover types, and abundance of ancillary data makes Minnesota a well-suited study area to identify and overcome some of the core methodological and computational challenges to answer our research questions. Minnesota has a large, mature agricultural sector, with cropland covering much of the state's area. According to the USDA, approximately 42.8% (88,247 km^2) of the state's land area (206,375 km^2) was cropland in 2017 (USDA - National Agricultural Statistics Service. 2019). Minnesota also has a diversity of land cover types, including extensive forested and naturally vegetated areas, freshwater features, and built up urban regions. This variety of land uses is necessary to develop a crop mapping model that is robust to other land cover types.

Minnesota is sufficiently large to test the computational efficiency of the workflow requiring hundreds of billions of pixels to process for a single agricultural season. To achieve full

spatial coverage for the state requires 37 Sentinel-2 tiles, each approximately 100 km by 100 km. At 10-meter spatial resolution, there are 100 million pixels in each of the four bands in each individual image. With 30 usable images from May through October for each tile footprint, accounting for cloudy images that will be omitted from analysis, a statewide analysis would include about 450 billion pixels from more than two terabytes worth of data. The computational demands of analyzing this amount of data requires computational efficiency be designed into the workflow.

What constitutes a field?

The farm field is the operational unit of agriculture and in order to map them we need to define what constitutes a field. Fields can look different under different agricultural practices, environmental regimes, and data resolutions, so researchers have used many various definitions. These can broadly be split into two categories, cropland mapping and field delineation. Cropland mapping tends to focus on the use of the land, specifically that it is cultivated for harvest. Field delineation is interested in identifying discrete objects, so the definition focuses on identifying the boundaries around fields. Both approaches are pertinent in this work, so we draw on the definitions used in each.

In global and continental cropland mapping efforts, the focus is on defining where agriculturally productive land is within the constraints of the data used for analysis. The USDA defines cropland as "all areas used for the production of adapted crops for harvest" (Johnson and Mueller 2010). Xiong et al (2017), in mapping croplands across the African continent, use a similarly broad but data constrained definition, referring to the size of the pixel used in their analysis, defining cropland as "a piece of land of minimum 0.09 ha (30-m 30-m pixel) that is sowed/planted and harvest-able at least once within the 12 months after the sowing/planting date." Ramankutty et al. (2008) combine cropland with pastureland also in part due to data constraints, using United Nations Food and Agriculture Organization's definitions that correspond with how many of the agricultural census data used in the study collect cropland information. This includes both "arable lands," crops planted and harvested each year, and "permanent crops" such as fruit trees and nut trees that are not replanted each year, along with pasture lands. Other global landcover mapping studies use a broader definition that include meadows or pastures in the agricultural class, which limit the utility for use in field mapping.

Fields and field boundaries are not always defined explicitly in research (Watkins and van Niekerk 2019; Belgiu and Csillik 2018) and definitions vary depending on research domain and objectives. Graesser and Ramankutty define fields simply as areas with "individual crops

planted throughout a year" (2017). To further clarify our objectives, we can refer to others who have mapped *field boundaries* specifically, and not just *cropland* (North, Pairman, and Belliss 2019). North et al. (2019) use a slightly updated version of Rydberg and Borgefors' definition, stating that a field boundary is "where a change in crop type takes place or where two similar crops are separated by a natural disruption in the landscape, like a ditch or a road," adding "significant differences in crop management" to clarify certain cases (Rydberg and Borgefors 2001). The USDA, in addition to defining "cropland," uses the term Common Land Unit (CLU) to refer to the field-scale agricultural unit, defined as "the smallest unit of land that has a permanent, contiguous boundary, a common land cover and land management, a common owner and a common producer in agricultural land associated with USDA farm programs. CLU boundaries are delineated from relatively permanent features such as fence lines, roads, and/or waterways" (USDA 2020). While this definition is centered on the producer's relationship to USDA programs, the notions of "common land cover" and a "contiguous boundary" delineated by "relatively permanent features" clarify what agricultural land means at the field scale.

For this paper, we will combine these to define a field as: *the smallest unit of land that has (1) a common land cover and management used for the production of adapted crops for harvest during the study period and (2) a contiguous boundary, where there is a change in crop type, land management, or a landscape feature.* Due to data validation constraints, for the purposes of this thesis we further restrict the minimum size of fields to 10 acres. In USDA agriculture data, 10 acres is the smallest cutoff for reporting farm data. Farm operations under 10 acres only made up about .1% (25,227 acres) of total cropland harvested (20,036,943 acres total) in Minnesota in 2017 (USDA - National Agricultural Statistics Service 2019). Unlike the USDA definition of cropland, ours does not include fruit- and nut-bearing trees.

With this definition, there is still some uncertainty in interpreting satellite imagery to determine what the boundaries of a field are and whether an area is cultivated or not. Even with high resolution imagery and especially with 10-m Sentinel-2 data, the boundaries between neighboring fields or features within individual fields can be difficult for a human eye to interpret. For instance, the border between two fields might be delineated by a wire fence that is not visible in the imagery. Or a field might be bisected by features like grass waterways but still cultivated with a single crop as a unified area. Given the limitations of the imagery resolution, landscape features along boundaries must be at least 100 m^2 (the size of a 10-m by 10-m pixel). Additionally, areas that appear to be crop fields in some images, might be grass or other uncultivated vegetation. This limitation is further discussed in the Data Challenge section below.

Chapter 2 – Background

The current state of remote sensing crop mapping research

Global vs. Local

Remote sensing techniques have been used extensively to map cropland and land cover around the world (Maciel et al. 2019; Schmedtmann and Campagnolo 2015). However, in an analysis of existing global and regional cropland and landcover mapping products, Waldner et al. (2015; 2016) found that accuracy varied widely, validation was lacking or insufficient, and discrepancies in spatial consistency and temporal coverage made utilizing these products difficult at a global scale. While global cropland datasets exist, they are too coarse in spatial resolution and limited in accuracy to be directly applicable to field scale mapping and data gaps persist in many regions (Xiong et al. 2017). Fritz et al. (2015) produced a global map of cropland extent and field size by integrating data from many existing global, regional, and national datasets, with inputs ranked for suitability through crowdsourcing, but the date ranges varied and high resolution datasets were aggregated to 1-km resolution. Recently, global cropland mapping at relatively high resolution has been achieved, but its use at local scales is still limited. Chen et al. created Globeland30, a 30-m global land cover dataset that includes a "cultivated lands" category, demonstrating the viability of high-resolution global mapping (Chen et al. 2015). While a review of global land cover datasets for cropland accuracy found Globeland30 to be the most adequate for cropland monitoring, the accuracy varied regionally, and the analysis is dated to 2010 (Pérez-Hoyos et al. 2017). A precise, up to date global map of agricultural fields does not exist.

Higher resolution agricultural mapping is also an active area of research, but typically these studies only cover relatively small areas (Aguilar et al. 2018; Forkuor et al. 2014). A review of remote sensing papers conducting crop classification found that the majority of the 75 studies used areas smaller than 5,000 sq km, and only one was conducted at a national scale, Myanmar at 676,578 sq km (Orynbaikyzy, Gessner, and Conrad 2019). For reference, Minnesota has an area of 225,180 sq km. The drawback of these analyses is that they tend to be finely tuned to the study areas and therefore the assumptions built into the methods make them implausible to expand to continental and global scales. Along with field mapping, agricultural applications for remote sensing analysis also include yield estimation, drought stress monitoring, water management, crop phenology, land use changes, and emerging work in precision agriculture, but the aims of

these are different from ours and often rely on a prior understanding of cropland location (Atzberger 2013).

New methods are needed to achieve the motivating goals of this project. The growth in the volume, velocity, and resolution of imagery data, the development of novel techniques to extract features and information, and the emergence of geospatial cyberinfrastructure to connect and process these data suggest that remote sensing will be a critical tool in addressing global challenges in the 21st century.

Cropland Mapping and Field Delineation

Existing agricultural mapping research can be split into the categories of Cropland Mapping and Field Delineation. The goal of Cropland Mapping is to identify regions of agricultural production in remotely sensed imagery. This can be computed as a percentage likelihood value, a binary crop/non-crop designation, or as part of a broader land cover classification scheme. Field Delineation, on the other hand, aims to identify discrete field parcels. By delineating boundaries, often from imagery where non-crop areas have been masked out, individual field objects can be extracted. Both approaches are important for a scalable field mapping workflow.

Cropland Mapping consists of computing the likelihood of agricultural activity happening within the geographic area of an imagery pixel. This includes global cropland and regional estimates, which tend to have coarse spatial resolution, cover an extended multi-year window to increase data availability, and use pixel-based analysis methods (Fritz et al. 2015; Pérez-Hoyos et al. 2017; Waldner et al. 2016). Often, this approach is part of land cover mapping, for which cropland is a single class among many and often includes pastures and grazing lands (Gómez, White, and Wulder 2016). Since detailed locations of cropland are important for issues of food security, many national governments have conducted cropland mapping analyses at the national level with varying levels of resolution and accuracy (Waldner et al. 2015). Accurate Cropland Mapping is required for strategic decision making around agriculture production and enables high level analysis like crop classification and yield estimation possible.

Field Delineation is concerned with extracting individual crop field parcels. These approaches tend to focus on much smaller geographic areas, necessarily use higher resolution imagery in which individual field edges are visible, and deploy mixed-method, object-based and edge-finding analytical approaches (North, Pairman, and Belliss 2019; Belgiu and Csillik 2018; Watkins and van Niekerk 2019). Recently, some continental-scale, high resolution field parcel mapping has begun to bridge the gap between global and local field mapping efforts, combining

pixel-based, object-based, and machine learning approaches covering Africa (Xiong et al. 2017), the continental United States (Yan and Roy 2014; 2016), regions of South America (Graesser and Ramankutty 2017), and Europe and the regions surrounding the Mediterranean Sea (Phalke and Özdoğan 2018). We build upon this work to propose a more generalizable and higher resolution approach to this challenge.

Applications of Global Field Boundary Mapping

While remote sensing data is used in many different agricultural research applications such as yield estimation and land cover monitoring (Atzberger 2013), mapping individual field units unlocks the potential for next level analysis. Agricultural remote sensing research largely depends on understanding where cropland is, and High precision, annual maps of cultivated areas are critical for food security analysis (See et al. 2015). A field map can provide a foundational dataset, improving the accuracy of higher-level agricultural data products (Xiong et al 2018). For example, crop masks are routinely used in crop type classification, crop health monitoring, and yield estimation (Yan and Roy 2014, Cai et al. 2018). An accurate field map reduces the uncertainty of existing agricultural research by providing a spatial mask at high precision (He et al. 2018). With the availability of more high-resolution imagery, aggregating a collection of pixels into a unified object can greatly reduce noise and simplify analytical processes.

A field map also discretizes analysis to the operational unit of agriculture, unlocking new potential applications for studying agricultural landscapes. Agricultural statistics are more accurate and meaningful when calculated at the field level (North, Pairman, and Belliss 2019). Currently, the USDA provides crop type classification at the pixel level through the Cropland Data Layer using 30-m Landsat imagery. Often, fields are apparent in the resulting maps but contain mixed classifications, where stray pixels of different crop types introduce noise. With field boundaries, much of that uncertainty can be eliminated. Additionally, field boundaries enable the study of interactions between landscape features. Modeling agricultural run-off into water systems can be improved with a deeper understanding of agricultural practices within precise locations of crop fields. Yield estimation can be done at much more granular scales, revealing differences in agricultural practices and climate stress across geographies. Additionally, field boundaries enable temporal analysis of cropped area, such as near real-time monitoring of food security indicators and comparison of year over year productivity. This kind of baseline spatial data allows for the detection of anomalies, which will be increasingly important as climate change induces more extreme weather events (See et al. 2015). Data-driven food resource planning can be greatly enhanced with high resolution field maps.

What are the challenges to building a global field dataset?

Research is advancing in both global scale remote sensing cropland mapping and in local-scale field delineation. However, there are three critical challenges to merging both goals into a single automated workflow: (1) the need for global satellite imagery with high spatial, spectral, and temporal resolution; (2) the development of methods that are flexible and scalable enough to accommodate a wide variety of climate regimes and agricultural practices without relying on extensive training data; and (3) the cyberinfrastructure capacity to process global-scale data flows in a high performance computing environment. I refer to these three challenges as the: data challenge, methodology challenge, and computation challenge, respectively.

Data Challenge

Satellite Imagery Platforms

Agricultural remote sensing is a vast area of research that utilizes many different satellite imagery sources (Atzberger 2013; Huang et al. 2018). Each has different characteristics and is suited for different research goals based on three core factors:

(1) *Spatial resolution*: the geographic area each pixel covers, ranging from a moderate resolution of 1-km to very high resolution of 1-m or less.

(2) *Spectral range*: the number of bands and their range of wavelengths along the electromagnetic spectrum, typically in the visible (450 - 700 nm), infrared (750 - 2,350 nm), thermal (10,400 - 12,500 nm), or radar (varying wavelengths) ranges.

(3) *Temporal resolution*: the frequency with which the satellite returns to a location, or the time between images.

Optical spectral bands (visible and infrared) are typically used for land cover and agricultural research and are common across all the platforms discussed here (Pérez-Hoyos et al. 2017). Radar data has been used in conjunction with optical imagery for some agricultural related classification work but not typically in cropland mapping or field delineation (Orynbaikyzy, Gessner, and Conrad 2019).

Satellite imaging platforms used in cropland mapping vary widely in spatial and temporal resolution but typically rely on imagery in the visible and near infrared (NIR) spectral ranges. Table 1 details several imaging platforms commonly used in agricultural mapping. Moderate resolution imagery from platforms like NASA's MODIS sensor aboard the TERRA and AQUA satellite constellation and the Vegetation-1 sensor on the French commercial SPOT-4 satellite have been used for large area land cover and agricultural mapping (Pittman et al. 2010; Friedl et

al. 2010; Wardlow and Egbert 2010; Bartholomé and Belward 2005; Ramankutty et al. 2008). The Landsat series of satellites, launched initially in 1972, is a widely used source for multispectral imagery (Huang et al. 2018). The Thematic Mapper, Enhanced Thematic Mapper Plus, and Operational Land Imager sensors on the three most recent Landsat satellites 5, 7, and 8, respectively, provide 30-m multispectral imagery along with a 15-m panchromatic band, with a temporal resolution of 16 days (Jenner 2015). This data has been used in a range of global (Chen et al. 2015) and continental scale agricultural mapping research (Graesser and Ramankutty 2017; Yan and Roy 2016). As part of its Copernicus Program, the ESA's Sentinel-2 platform provides multispectral imagery at 10-m (visible and NIR), 20-m, and 60-m resolution with a five-day temporal resolution (ESA 2020). Many researchers have been using Sentinel-2 imagery for small and large area mapping in place of or in combination with Landsat imagery due to higher resolution and temporal frequency of Sentinel-2 (Belgiu and Csillik 2018; Watkins and van Niekerk 2019; Vuolo et al. 2018; Xiong et al. 2017). Very high-resolution imagery, on the order of 1-m, has become more available in recent years, with the growth in commercial imagery companies. Imagery from DigitalGlobe's WorldView-2 satellite has been utilized for smallholder farm mapping in areas where moderate resolution imagery is not adequate (Aguilar et al. 2018; Debats et al. 2016).

Reference Name	Satellite	Sensor	Spectral Bands	Spatial Resolution	Temporal Resolution
MODIS	NASA Earth Observing System TERRA/AQUA	Moderate Resolution Imaging Spectroradiometer (MODIS)	36 spectral bands	250-m; 500-m; 1000-m	1 day
SPOT 4	SPOT 4	Vegetation-1	4 spectral bands	1000-m	21-26 days
Landsat	NASA/USGS Landsat-5/-7/-8	Multispectral Scanner (MS), Thematic Mapper (TM), Enhanced Thematic Mapper Plus (ETM+), Operational Land Imager (OLI)	4 (MS), 7 (TM), 8 (ETM+), 11 (OLI)	60-m (MS); 30-m (TM, ETM+, OLI); 15-m panchromatic (ETM+, OLI)	16 days
Sentinel-2	ESA Sentinel-2A & 2B	Multi Spectral Instrument	13 spectral bands	10-m; 20-m; 60-m	5 days
WorldView	DigitalGlobe WorldView-2	WorldView-2	8 spectral bands	0.5-m panchromatic, 1.8-m multispectral	1 day

Table 1. Common satellite imaging platforms for cropland mapping

Spatial Resolution

In order to produce high accuracy crop field boundaries, the source imagery needs to have high spatial resolution (Mueller, Segl, and Kaufmann 2004). The relatively coarse resolution of most existing global cropland mapping efforts—from 250-m to 1-km—means they are not directly applicable to field-scale mapping (Waldner et al. 2015). Very high-resolution imagery from platforms like WorldView-2 is useful for delineating cropland in areas with smallholder farming (Debats et al. 2016), but the associated computational challenges and cost to procure commercially limit its viability for use in global analysis (Lebourgeois et al. 2017). Landsat imagery has been used to map global land cover at 30-m resolution (Chen et al. 2015) and demonstrated improved accuracy compared to coarser resolution products (Pérez-Hoyos et al. 2017). It has also been used for field delineation but is not well suited for areas with small field parcel sizes and can introduce inaccuracies even in areas with large fields (Yan and Roy 2016; Graesser and Ramankutty 2017; Phalke and Özdoğan 2018). In the 30-m Landsat derived USDA Cropland Data Layer (CDL), for example, often roads between fields are too narrow or at an angle relative to the pixel grid to be misclassified as cropland. Figure 1 shows a comparison of 10-m Sentinel-2 imagery with 30-m Landsat imagery and the CDL Crop Mask layer where a road

feature has been incorrectly classified as cropland. Yan and Roy (2016) suggest that 10-m Sentinel-2 imagery can be used to build upon the extensive agriculture mapping work that has been done historically with 30-m Landsat data to address some of these limitations. Belgiu and Csillik (2018) were able to successfully delineate field boundaries in three different test areas, including within a fragmented landscape in Romania, citing the 10-m spatial resolution of Sentinel-2 as a key factor. Watkins and van Niekerk (2019) also successfully used Sentinel-2 imagery in South Africa to delineate field boundaries. The increased spatial resolution of Sentinel-2 data enables successful field-level analysis.

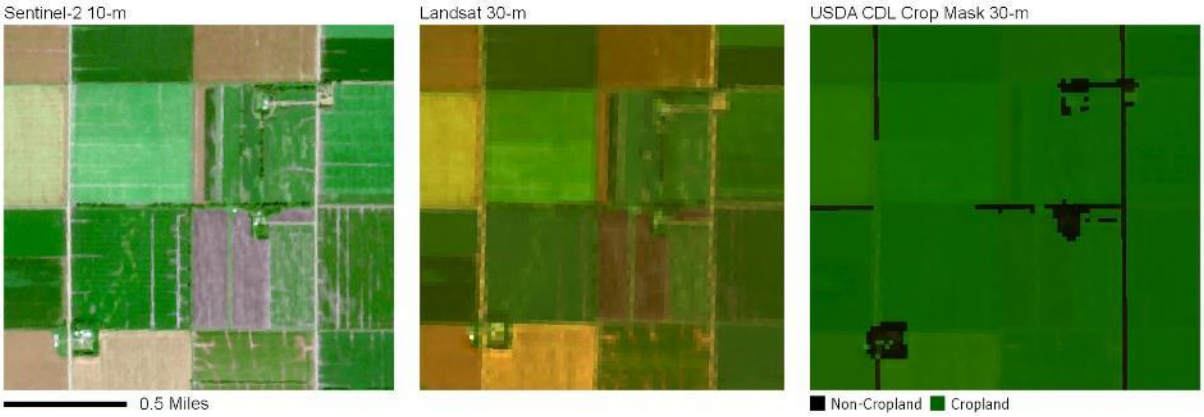

Figure 1. Comparison of 10-m Sentinel-2 imagery (left), 30-m Landsat imagery (middle), and Landsat-derived USDA Cropland Data Layer Crop Mask (right)

Temporal Resolution

The use of time series remote sensing data has been shown to be useful for cropland identification and boundary delineation. Within a single image, the spectral signatures of crops and natural vegetation may be very similar. To differentiate these classes, and to conduct further analysis such as crop type classification, the use of temporal and spectral features improves accuracy (Cai et al. 2018). Phalke and Özdoğan (2018) found that at least 15 valid temporal data points were necessary for a pixel to have robust vegetation statistics for a growing season. In a mixed pixel-based and object-based field delineation study for three agricultural different regions, Belgui and Csillik (2018) found that the more imagery available, the better the results of their time-weighted model. In their study to map smallholder agriculture, Lebourgeois et al. (2017) found that temporal information is essential for differentiating land uses even when spatial resolution is too coarse to produce a pure spectral signal. The short revisit time of the Sentinel-2 program, returning to each location about every five days, enables the necessary temporal resolution to use the seasonal characteristics of different land covers and individual crop types for field mapping.

Cloud cover is a challenge endemic to optical remote sensing analysis but can be mitigated by utilizing high temporal coverage imagery, which increases the likelihood of capturing cloud free data. Many areas around the world with extensive agricultural cultivation have pervasive cloud cover (Whitcraft et al. 2015). Addressing this challenge is necessary for global field mapping. One option is to use a cloud mask to remove cloud-contaminated pixels in an image, which can be done through spectral analysis in each image or through temporal modeling (Candra, Phinn, and Scarth 2020). Once the clouds in an image have been masked out, areas without data remain. To fill in these gaps, researchers often combine images from multiple imagery dates to create a cloud-free temporal composite image. For the case of continental cropland mapping in Africa, Xiong et al. (2017) were limited to only two cloud-free composites for a single year using 30-m Landsat by integrating higher temporal frequency Sentinel-2 data into the processing to fill data gaps. Another approach to minimize the effects of cloud coverage is to reduce the data into temporal features by computing statistics at each pixel from the valid imagery dates, thereby normalizing the data into a consistent structure regardless of the number of snapshots (Phalke and Özdoğan 2018). Both cloud masking and data reduction via robust temporal feature extraction are possible with Sentinel-2 imagery.

Sentinel-2: Suitable for Global Field Mapping

Sentinel-2 data solves the spatial, temporal, and multi-spectral challenges with global coverage. Xiong et al (2017) has shown Sentinel-2 to be effective at mapping cropland over large areas and Watkins and van Niekerk (2019) demonstrated its viability for delineating individual crop fields. Graesser and Ramankutty (2017) mapped field parcels in South America using Landsat imagery and suggested that future work could use Sentinel-2 data to overcome some of the limitations encountered with 30-m imagery for mapping small field parcels accurately. Based on the success of the 30-m Landsat derived Globeland30 land cover product at identifying cropland in areas with small field sizes, Perez-Hoyos et al. (2017) also suggested that Sentinel-2 imagery offers a valuable source for improved cropland mapping in fragmented agricultural landscapes at a global scale. Xiong et al. (2017) demonstrated the utility of Sentinel-2 temporal resolution for mitigating cloud contamination while Belgiu and Csillik (2018) emphasized its importance for improving cropland modeling. While very high-resolution imagery like WorldView-2 holds promise for accuracy in field mapping over small areas, it poses significant challenges for analysis over large areas by drastically increasing computational demands. Sentinel-2 data includes imagery across 13 spectral bands, but those with 10-m resolution (R, G, B, and NIR) are the most critical for agricultural mapping. Based on the imagery available, the

Sentinel-2 data is best suited for our goals because it addresses all three challenges by providing global coverage of high-resolution imagery in the visible and NIR spectrums with high temporal frequency.

Cropland Related Data for Assessing Accuracy

Another component of the crop mapping data challenge is the need for ground truth data to evaluate the accuracy of both cropland mapping and field delineation outputs. Comparison of cropland area with agricultural survey data is a common accuracy assessment approach, particularly for large areas (Xiong et al. 2017). Perez-Hoyos et al. (2017) reviewed existing global land cover datasets to evaluate the accuracy of cropland classes by comparing national level outputs to statistics gathered by the United Nations Food and Agriculture Organization, finding a wide variation of accuracy for different global estimates in different regions. Lesiv et al. (2019) used online crowdsourcing to estimate the distribution of field sizes around the world, an approach that has also been used for identifying the suitability of regional cropland maps (Fritz et al. 2015). In studies that require training data, some samples are held out for model validation (Phalke and Özdoğan 2018) or an independent set of ground truth data is collected for testing and validation (Chen et al. 2015). Another approach is to use existing land cover products as spatial ground truth since authoritative cropland and field parcel data are rare (Graesser and Ramankutty 2017). Crop classification and field delineation studies in the United States often use USDA products like the CDL or the Common Land Unit database, either for training or validation (Yan and Roy 2016; Lark et al. 2017). These external datasets are useful when available, but robust crop mapping products do not exist in most countries and evaluation of global datasets depends on validation samples derived from existing land cover products (Waldner et al. 2015). Field delineation research often relies on hand-digitized field boundaries derived from very high-resolution imagery, which is time consuming and limited in scope when analyzing large areas (Watkins and van Niekerk 2019; North, Pairman, and Belliss 2019). Often, multiple accuracy assessment methods are incorporated based on the data available across multiple geographic scales (Xiong et al. 2017; Phalke and Özdoğan 2018). Our work incorporates USDA survey data, land cover information from the CDL, and hand-digitized field boundaries for validation at the field, county, and statewide levels. In future work, crowdsourcing could provide a useful way to incorporate distributed knowledge into validation of a global-scale fields dataset.

Methodology Challenge

Methods for crop mapping vary widely, often determined by the goals of the research (cropland mapping, field delineation, crop type classification, yield estimation, etc.), the geographic scope of analysis, and the constraints of the data used. There are some commonalities across most studies, though, including the use of spectral vegetation indices, the importance of the temporal dimension of the imagery data, and the use of ground samples or manually digitized training data in the model. While many of these studies provide useful insight for identifying cropland and extracting field features, the supervised nature of most of the models limits their generalizability and geographic scalability. In order to build a truly global workflow, supervised processes and reliance on training data must be largely eliminated from the process.

Pixel-Based and Object-Based Analysis

Pixel-based methods classify each cell in a raster image based on the spectral characteristics of the pixel values. Each pixel becomes a feature vector based on the intensity in each of the spectral bands, which can then be passed to various classification models to assign a prediction to each pixel. Common remotely sensed agriculture data products, such as the USDA Cropland Data Layer, use pixel-based methods to estimate crop types. This approach can be limited by noisy data and local heterogeneity in which natural variability of neighboring pixels emerges when using high resolution imagery (Karpatne et al. 2016), which can lead to a salt and pepper effect in the classification (Watkins and van Niekerk 2019). Most existing global cropland products use pixel-based approaches to classify a pixel as "cropland" or estimate the intensity of agricultural activity (Pérez-Hoyos et al. 2017). Most land cover modeling approaches that use data mining or machine learning approaches typically use pixels as the unit of analysis (Gómez, White, and Wulder 2016).

While pixel-based analyses have traditionally been the most common method for remote sensing land cover classification, object-based image analysis (OBIA) has emerged as a powerful framework, particularly for feature extraction. Blaschke et al. (2014) argue that this approach represents a new paradigm in geospatial analysis. OBIA methods group neighboring pixels into "objects" based on proximity and spectral similarity through a process called segmentation. Once pixels have been grouped into contiguous areas, other characteristics of these objects, such as texture, shape, and average pixel values, can be used for classification and are more effective at handling noisy pixels. These objects often approximate land features more accurately than pixels, given that most features do not conform to a pixel grid. In particular, current crop field research has demonstrated potential for high accuracy outputs by implementing OBIA methods

considering that the outputs represent contiguous field polygons instead of simply groups of pixels (Belgiu and Csillik 2018; Ma et al. 2017; Watkins and van Niekerk 2019; Yan and Roy 2016). However, while OBIA methods solve some shortcomings of pixel-based approaches, they create new challenges of defining what constitutes an object and developing methods for effectively detecting objects (Duro, Franklin, and Dubé 2012). Over-segmentation—where a feature is split into too many objects—and under-segmentation—where neighboring features are joined into a single object—are both common in object-based field mapping. Watkins and Van Niekerk (2019) found that where there is a lot of natural variation within a field it can be over-segmented and small, spectrally similar adjacent fields can be under-segmented. Some segmentation methods used in OBIA approaches can also be dependent on geographically dependent parameters, potentially limiting scalability (Waldner, Canto, and Defourny 2015).

Hybrid methods are becoming more common and have been shown to be effective in large scale agricultural mapping (Xiong et al. 2017). Chen et al (2015) used a hybrid pixel- and object-based approach to global land cover classification at 30-m resolution. They first used spectral and temporal features in a supervised model to classify pixels, then applied labels to object-based polygons through percentage thresholding or expert interpretation, removing the "salt-and-pepper" effects of pixel classification in mixed environments. Incorporating pixel-based data preprocessing with edge-based feature extraction into a multi-step workflow has produced accurate field parcels with both Landsat imagery in South America (Graesser and Ramankutty 2017) and with Sentinel-2 imagery in New Zealand (North, Pairman, and Belliss 2019). These approaches are promising for scalable field mapping efforts because some important spatial features, like edges, can be derived directly from the imagery without relying on a supervised model. Pixel-based analysis is important for classifying areas as cropland and object-based methods can produce distinct field polygons, which is the goal of our work. We combine components of each approach to accomplish both cropland mapping and field delineation.

Time-Series Analysis

Temporal characteristics of remote sensing data can be used effectively in many different methods and are a critically important element of classifying cropland and delineating crop fields (Kussul et al. 2015). For instance, North, Pairman, and Belliss (2019) used images from throughout a growing season to generate multi-channel composites where temporal change in the NIR band is represented through colors to enhance field boundary delineation. Many researchers have used spatio-temporal statistics of imagery data as a means of data reduction for inputs into cropland modeling or field delineation (Phalke and Özdoğan 2018; Graesser and Ramankutty

2017). Phalke and Özdoğan (2018) computed a set of seven temporal statistics of the Enhanced Vegetation Index for each pixel over a growing season that they found to be most effective at differentiating cropland, while simultaneously normalizing the input data across areas with different numbers of cloud-free images. Graesser and Ramankutty (2017) computed ten time-series statistics at each pixel over the growing season for five different vegetation indices, which were then used for both field boundary extraction and object classification. Other approaches use multiple cloud free images throughout a growing season to improve differentiation between neighboring fields through identifying temporal phonological patterns (Belgiu and Csillik 2018). Temporal features were critical for crop classification in the U.S. Midwest (Cai et al. 2018) and in differentiating smallholder agricultural land in Madagascar (Lebourgeois et al. 2017). Like many other researchers, we adopt the temporal analysis of vegetative indices as a core component of our workflow.

Machine Learning and the Limits of Supervised Approaches for Scalability

Machine learning and data mining methods have become popular and effective approaches for land cover and crop type classification in various places with diverse agricultural regimes in both pixel-based and object-based approaches (Mahdianpari et al. 2019; Maxwell, Warner, and Fang 2018). Researchers have demonstrated, for instance, that a random forest algorithm could effectively classify smallholder fields (Debats et al. 2016). Pittman et al (2010) used supervised decision trees (DT) and Random Forest (RF) methods to map global cropland with MODIS data. Duro et al. (2012) demonstrated that DT, RF and Support Vector Machines (SVM) are suitable for agricultural landscape classification with both pixel- and object-based approaches. However, limited amounts of ground-truth training data, heterogeneity in space and time, and varying data quality and resolution all complicate the use of traditional machine learning techniques in spatial projects (Karpatne et al. 2016). Emerging work in deep learning based on model architectures that use layers of hidden neural networks, is particularly promising in the realm of earth systems science as these methods are well-suited to handle complex relationships within spatio-temporal imagery data, but these models also typically require extensive training data (Atluri, Karpatne, and Kumar 2018).

Novel solutions to the processing and analysis of remote sensing data are needed to address the computational challenges of handling the vast amounts of data borne out of recent developments in satellite imaging capabilities. At 5-m spatial resolution, it takes 100 trillion pixels to cover the earth, which is now happening daily. The Trillion Pixel Challenge for GeoAI, described by researchers at the Oak Ridge National Lab, represents an initiative to overcome the

current limits on imagery collection platforms, artificial intelligence methods, and high performance computing infrastructure needed to effectively retrieve and analyze 100 trillion pixels every day (Lunga, Bhaduri, and Stewart 2019). These challenges underscore the need for machine learning methods that are generalizable across space and time, not dependent upon large samples of training data, and efficient to run. Automated feature extraction is one of the key applications for utilizing this emerging field of global big data remote sensing.

Convolutional Neural Networks (CNN) have become some of the most powerful and widespread deep learning models for image analysis (Karpatne et al 2016). Through assessing the spatial context around a pixel at various scales through the model, CNNs can recognize the fine-grain structural elements that make up a feature—like corners, edges, spectral patterns—as well as its relationship to surrounding features. This type of architecture enables the model to identify cohesive, high level features in an image (Maggiori et al. 2017). This is a different paradigm than pixel-level analysis, which looks at each data point in isolation. These models, in addition to labeling what predictive class exists in an image, can classify groups of pixels into individual features, a process called semantic segmentation (Long, Shelhamer, and Darrell 2015).

Applying CNN models to study agriculture using remote sensing imagery is becoming more common, but applications range widely—yield prediction or crop classification, for example—and are not focused on crop field delineation (Kamilaris and Prenafeta-Boldú 2018). Most research utilizing CNNs for semantic segmentation of remote sensing imagery often focus on feature delineation in urban areas, where there are many more classes and objects than in a two-class agricultural area (Marmanis et al. 2017). This is partly due to the availability of several public high resolution labeled training datasets in urban areas (Potsdam and Vaihingen, Germany) published as part of an International Society for Photogrammetry and Remote Sensing challenge to benchmark semantic labeling approaches (Sherrah 2016).

Given the challenge of acquiring precise annotations for machine learning tasks, researchers have experimented with using imperfect labels for training, finding that deep learning models can effectively learn using noisy training data. Mithal et al. (2017) proposed a framework for using imperfect labels to predict rare classes in satellite imagery, with the assumption that if the sum of the error rates of all label classes is less than 1—meaning the noisy labels are mostly true, if not perfect—and that the error rate is conditionally independent—that the errors in the labels are not skewed to one class. They found that these imperfect labels can be almost as good as gold standard training labels in identifying rare classes of forest fires across large geographic areas. Papandreou et al. (2015) also found that training methods using weakly supervised methods and noisy labels such as labeling the bounding box of a feature instead of precise pixel-

level annotations can provide high levels of semantic segmentation accuracy at much lower costs. Deep learning is a promising approach to overcoming some of the limitations of traditional pixel- and object-based approaches but requires more research to make the methods truly generalizable and scalable.

Most existing crop mapping projects rely on supervised classification approaches, limiting the generalizability of the methods to other areas (Gómez, White, and Wulder 2016). Authors across domains have identified training and validation data as a key constraint of any global mapping endeavor and new approaches are needed to overcome this challenge (Traganos et al. 2018; Waldner et al. 2016). Generating training samples manually to develop a model for cropland identification is prohibitively time consuming and complicated by regional agricultural differences for a globally generalizable model. To address this in cropland mapping at the continental scale, areas have been categorized into climatically and agriculturally similar regions each with their own sets of training samples and individually trained models (Phalke and Özdoğan 2018; Xiong et al. 2017). Even in these examples, however, certain areas required the additional creation of training samples throughout the process in areas that proved difficult to model, namely in parts of Sub-Saharan Africa where crop parcels are small and intermixed with natural vegetation or where cloud coverage is pervasive. Due to the complexity, variability, and global extent of agricultural production around the world, field-level labeled data does not exist for supervised approaches to field mapping (S. Wang, Azzari, and Lobell 2019). This approach would require a massive effort to accomplish globally and would limit the analysis to a single time period from which training data was collected.

At the field scale, external data like crop masks or land cover classification are often utilized to spatially subset imagery for field delineation or provide training labels for classification. The Cropland Data Layer has been demonstrated to be very useful in deriving field parcel boundaries (Yan and Roy 2014) and classifying crop types (Cai et al. 2018) in the US but is not applicable to other regions that do not already have a robust national dataset of cropland. This approach is not feasible for building a truly scalable crop mapping workflow. As has been shown, existing cropland maps vary in availability, accuracy, and definition around the world, meaning they cannot be reliably used as crop masks in many areas (Waldner et al. 2015).

The limitations of generalizability for supervised methods demonstrate the necessity and promise of unsupervised approaches, where cropland characteristics are derived from the data themselves. Data mining methods have been used for unsupervised image analysis across multiple domains, including spatio-temporal remote sensing research (Atluri, Karpatne, and Kumar 2018). Clustering has been shown to be effective in identifying similar climate regions in

spatio-temporal data over large areas and shows promise for use in other remote sensing domains (Steinbach et al. 2003). Waldner et al. (2017) implemented an unsupervised approach for cropland mapping at a national scale in South Africa but still depended on external data to generate training samples, using an outdated land cover map to select appropriate pixels for training. Wang et al. (2019) experimented with unsupervised learning for crop type classification in the US Midwest, using k-means and gaussian mixture models to cluster temporal spectral features in conjunction with aggregate crop statistics and the Cropland Data Layer (S. Wang, Azzari, and Lobell 2019). While the Wang et al. study used data inputs that are not available everywhere, it demonstrated the potential for using unsupervised approaches to make more generalizable models that can be deployed as a first step in places that lack more robust agricultural data infrastructure. Further exploration of these unsupervised spatio-temporal clustering of vegetative indices can provide key insight into building a scalable and generalizable crop mapping workflow.

Computation Challenge

In order to handle the volume of data and complex analytics processes for this task, the methods need to be computationally efficient and scalable within a high-performance computing environment. The computation challenge underscores that it is not simply about finding effective methods to delineate crop fields, but also computational infrastructure and efficiency to make it possible at all. Importantly, though, the system of big data infrastructure, computing resources, and connective infrastructure—known as Geospatial Cyberinfrastructure (GCI)—exist to achieve a global field scale analysis (Yang et al. 2010). Computation, which has become a "third pillar" of science alongside theory and experiment (Reed et al. 2005), and the emergence of data as a potential "fourth paradigm" have changed the way that researchers approach their work (Hey 2009). Data and methods alone cannot solve the global field mapping challenge. GCI enables global analysis at high spatial resolution.

Accessing and efficiently making use of high-performance computing resources presents its own set of constraints on the analysis. Utilizing GCI for remote sensing-based research requires access to the computing resources necessary to process the data, connections between the data source and computing resources, and the implementation of HPC methods such as parallel processing to efficiently perform the analysis (Wright and Wang 2011). In order to fully utilize GCI to overcome the computational challenge of global field scale mapping, the analytical workflow must integrate efficiency considerations into the design from the start. Parallel processing in distributed computing environments is critical to efficiently extract useful

information from remote sensing imagery (Lee et al. 2011). Cai et al. (2018) demonstrated that applying state of the art remote sensing methods in a supercomputing environment enables high accuracy crop classification at the field scale. Computational researchers across the earth science domains have developed and adopted HPC software tools for big data, large area, remote sensing analysis through the Pangeo project (Hamman, Rocklin, and Abernathy 2018; "Pangeo" 2020). Some of these open source tools, like the Python packages Xarray and Dask, facilitate flexible raster analysis and parallel processing methods to enable analytical workflows to be deployed on HPC resources.

Public science gateways and commercial cyber infrastructure tools are growing to meet the computational demands of large-scale spatial analysis across domains. The Extreme Science and Engineering Discovery Environment (XSEDE) is a National Science Foundation funded project that provides free resources for education through access to an advanced set of integrated digital resources for intensive computational research (Towns et al. 2014). Open science gateways built upon this GCI, such as the GISandbox project, can provide access to these resources to enable spatial research that would otherwise be limited by computational demands (Shook et al. 2018). Google Earth Engine (GEE) has emerged in recent years as another powerful tool for scalable remote sensing analysis, and a key example of cutting edge GCI, that integrates big data imagery repositories, like Landsat and Sentinel, and automatic parallel computing in a cloud environment (Gorelick et al. 2017). GEE represents the next generation of the Digital Earth, opening access and removing critical barriers for scientists across many domains (Goodchild et al. 2012). The platform has been used for global scale mapping of forest cover (Hansen et al. 2013), regional scale wetland mapping in Canada (Mahdianpari et al. 2019), unsupervised crop type mapping (S. Wang, Azzari, and Lobell 2019) and cropland yield estimation across the United States Midwest (Lobell et al. 2015) and in smallholder farms in Kenya and Tanzania (Jin et al. 2019), and, most similar to our efforts, continental scale cropland mapping of Africa (Xiong et al. 2017), among other applications (Mutanga and Kumar 2019). Going forward, integrating GCI resources like GEE, XSEDE, and its related science gateways will be imperative to empower research in the big data era and explore global challenges.

Remote Sensing Analytical Tools

Common approaches for cropland mapping and field delineation across existing research reveal several remote sensing techniques that can be integrated into a scalable workflow. The methods include spectral indexing, edge detection, clustering, and image analysis through

segmentation and morphology. These core tools can serve as building blocks in a hybrid approach to field mapping, which we deploy in various configurations in our experiments.

Spectral Indices for Vegetation and Water

Almost all crop mapping research utilizes vegetative indices derived from combinations of spectral bands in the remote sensing imagery (Frampton et al. 2013). The Normalized Difference Vegetation Index (NDVI) is a commonly used index for evaluating the health of vegetation in a multispectral image and has been used for decades in remote sensing research (Rouse 1974). NDVI is a ratio between the NIR and red bands that estimates relative vegetative activity in a pixel. High NDVI values suggest healthy vegetation—as healthy plants reflect energy in the NIR wavelengths and absorb energy in the red wavelengths—while bare earth and water have low NDVI values (Rouse 1974; Tucker 1978). The formula for NDVI is:

$$NDVI = (NIR - Red)/(NIR + Red)$$

Equation 1. Normalized Difference Vegetation Index

Over the course of the growing season, manmade features such as roads and buildings, as well as water features like rivers and lakes, will not show very much change in NDVI. Crops, however, will go from bare soil in the spring to high NDVI at peak vegetation and then return to bare soil after harvest, showing a large range in NDVI values. Natural vegetation will also "green up" during the summer, but will not experience as dramatic of a change since it often greens up earlier than crops that are planted each spring and harvested in the fall. Temporal NDVI analysis is an integral part of many crop mapping projects (Waldner et al. 2016).

The Normalized Difference Water Index (NDWI) is a spectral index similar to NDVI that is tuned for detecting open water bodies in multispectral imagery data. It was originally proposed by Gao (1996) as a ration between the NIR and short-wave infrared (SWIR) bands to detect liquid water in vegetation but was modified by McFeeters (1996) for use in identifying water bodies by using the NIR and green spectral bands. Since we are not utilizing the SWIR spectral bands of the Sentinel-2 data and we are interested in identifying water features in the landscape to be masked out of the crop mapping process, we utilize McFeeters' NDWI formula. The formula for NDWI is:

$$NDWI = (Green - NIR)/(Green + NIR)$$

Equation 2. Normalized Difference Water Index

When compared with other indices used for water feature extraction, NDWI performed the best on Landsat imagery (Rokni et al. 2014). NDWI threshold values tend to vary based on temporal

imagery characteristics of sensor angle and atmospheric qualities (Liu 2012). In a mixed landscape urban environment, an NDWI threshold of 0.3 was used to differentiate swimming pool features (Stuart K. McFeeters 2013).

Edges: Sobel Filter

Edge detection is a common analytical technique for image analysis and many different edge-finding algorithms exist, including Sobel, Scharr, Pruitt, Canny, among others (Watkins and van Niekerk 2019). Generally, these algorithms work by passing a weighted convolutional filter over an image and computing the local product of the kernel at each pixel. These edge operators identify gradients in an image according to the kernel inputs, which can vary by size, gradient direction, and weights. Graesser and Ramankutty utilize 2x2 and 3x3 pixel kernels with horizontal, vertical, and diagonal kernel weights (Graesser and Ramankutty 2017). Others have used much larger convolutional operators composed at 16 different angles in order to emphasize edge length over edge magnitude in an image (North, Pairman, and Belliss 2019). The Sobel edge algorithm uses a pair of 3x3 pixel kernels, one a 90-degree rotation of the other, to find a directional gradient magnitude value at each pixel location in the image. These directional gradients are combined to find the absolute magnitude at each pixel.

$$G_x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} G_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

Equation 3. Sobel convolutional filters

$$|G| = \sqrt{G_x^2 + G_y^2}$$

Equation 4. Sobel edge magnitude

We utilize the Sobel edge algorithm in the Python Scikit-Image package.

K-Means Clustering

The k-means algorithm is a popular clustering technique widely used in data mining and remote sensing analysis (Atluri, Karpatne, and Kumar 2018). The algorithm works by setting an initial centroid value for each of the *k* clusters (the number of clusters, *k*, must be defined as an input), then looping through two steps to update the clusters by: *(1)* assigning each observation in the input data to the nearest cluster center based on Euclidean distance and *(2)* recalculating the centroid of each cluster based on the mean of the data points assigned to each cluster. Once the clusters do not change anymore or a set number of iterations are complete, the algorithm stops.

The algorithm is flexible to the size of data features in the input allowing for clustering in n-dimensional feature space. The critical feature of the algorithm is that it is unsupervised, enabling the extraction of pixels with similar spectral and temporal characteristics without having to pre-define what those characteristics are. Once the clusters have been identified, we can identify which have the characteristics of cropland versus non-cropland to aid in analyses.

Segmentation Approaches

Many image segmentation algorithms have been used in remote sensing to delineate landscape features (Jiao et al. 2010). For delineating individual fields with Sentinel-2 imagery, Watkins and Van Niekerk (2019) several edge-based and region-based segmentation approaches. They found that combining Canny edge detection to find cumulative edge gradient values from multi-temporal imagery with a watershed algorithm to "fill" in contiguous regions within edge-bounded areas produced accurate results. Another approach from computer vision is graph-based segmentation, which incorporates spectral values with the spatial relationship of neighboring pixels. We utilize a common graph-based algorithm for segmentation, referred to here as Felzenszwalb segmentation (Felzenszwalb and Huttenlocher 2004). This algorithm works by creating a graph structure with pixels as nodes and edges connecting each pair of neighboring pixels. These edges are weighted based on the similarity of color values in each pixel and segments are created by merging pixels to minimize dissimilarity within a segment and maximize dissimilarity between different segments. One benefit of this segmentation approach is that it is computationally efficient, scaling linearly with the number of pixels in the input image. There is a large body of research on the different approaches to segmentation with remote sensing images and additional exploration of techniques within a combined cropland mapping and field delineation workflow is warranted (Csillik 2017).

Morphological Filters: Post-Processing Noise Removal

Morphological image analysis techniques have become common, useful strategies to complement spectral and temporal analysis of remote sensing imagery. Morphological approaches utilize the spatial relationship and relative intensity of pixels to manipulate spatial structures in an image once it has been reduced to a single channel, where "objects [are] defined as a specific spatial arrangement of image pixels rather than single pixels with a specific spectral signature" (Soille and Pesaresi 2002). This is a common approach to image filtering for noise reduction (the elimination of isolated outlier pixels) and edge enhancement in field delineation workflows. Graesser and Ramankutty (2017) utilize a series of morphological steps to thin edges,

link edge segments, and trim endpoints in order to clean up a binary map of edges derived from temporal vegetation statistics. Yan and Roy (2014) utilize a sequence of dilation and erosion filters on extracted field parcels to clean up edges where pixel-level uncertainty creates jagged boundaries around segmented field objects. North et al (2019) implement thinning and smoothing morphological operations to clean up field boundaries derived from convolutional edge detection. These morphological approaches can be used effectively to clean up binary cropland maps and field boundaries.

Chapter 3 – Proposed Framework

How can we design a scalable, flexible field mapping workflow?

In order to overcome the data, methodology, and computation challenges in global field mapping, we propose a framework that builds on existing work to enable efficient, geographically flexible analysis without relying on externally derived training data. Instead of a single approach that is finely tuned to create crop field maps for Minnesota, the framework is modular and divides the process into several distinct operational steps: data preparation, cropland masking, and field delineation. By modularizing each step, this framework can be flexible to accommodate different geographic contexts and agricultural regimes. The result is an analytical workflow for conducting field mapping experiments. It also enables the rapid testing of multiple methods within each step to compare the accuracy and efficiency of combining different approaches into a hybrid model. Through our experiments, we identify some of the key limitations to building a scalable field-level mapping workflow, and through the combined use of traditional remote sensing and emerging machine learning techniques we identify multiple promising solutions.

This modular framework solves the **data challenge** by taking advantage of the high spatial and temporal resolution of Sentinel-2 10-m imagery, **methodology challenge** by building a flexible workflow with hybrid methodologies to achieve semi-supervised cropland masking and field delineation, and **computation challenge** by implementing parallel processing in a HPC environment at the Minnesota Supercomputing Institute and by utilizing optimizations such as temporal data reduction and crop masking to reduce computation load. This framework is the first step toward testing the viability of geographically scalable unsupervised field mapping, so the workflow was designed and developed to allow methods to be tested quickly to identify core data, methodology, and computation challenges.

Methods can be tested independently to achieve the specific sub-goals of each step. By breaking it down into a workflow of processes, methodological approaches and bottlenecks in the computational efficiency can be examined. This enables strategies for Cropland Mapping and Field Delineation to be tested independently as well as integrated into a cohesive process. When put together, this framework will allow us to answer the research questions and explore their geographic and computational scalability.

- *Data Collection and Structuring*
    - Goal: Efficiently read imagery data into an analysis ready data structure that can be deployed in an HPC environment.

    - The data represent all Sentinel-2 images taken over Minnesota for the full growing season, May through October of 2019, comprising billions of pixels. To accommodate this, we need a data structure that efficiently handles multidimensional arrays, supports parallel processing, and enables efficient, flexible implementation of image processing techniques. This step forms the backbone of the data processing pipeline.
- *Data Preprocessing*
    - *Goal*: Reduce raw imagery data into useful inputs for crop mask and field segmentation.
    - This step is critical for designing around the constraints of both computation and accuracy. The efficiency and volume of data going through the workflow is a function of how the preprocessing step is designed. In global scale analysis, this step can greatly diminish the volume of data to be analyzed in subsequent steps. Reducing raw imagery into spectral and temporal features determines what properties of the landscape can be modeled. For delineating crop fields in complex landscapes, the choice of data inputs—like spectral indices, temporal statistics, and edge—is critical to effectively use the analytical tools.
- *Crop Masking*
    - Goal: Differentiate cropland from non-cropland to know where in the landscape to look for field boundaries.
    - Without relying on externally produced crop masks for field delineation, the first analytical step in the workflow is to create a binary layer identifying which pixels are part of any crop field. Within a heterogeneous landscape over large areas, this requires methods that are responsive to the specific spectral and temporal characteristics of cropland. This crop mask layer then address two challenges later in the workflow: (1) it reduces the computational load of the field segmentation step by eliminating non-crop areas; (2) it improves the accuracy of the field outputs by identifying boundaries between crop and non-crop features in the landscape and reducing the amount of local spatial heterogeneity during boundary segmentation.
- *Field Segmentation*
    - *Goal*: Delineate the groups of crop pixels into individual field parcels.
    - Going from pixel analysis to delineating and extracting individual field polygons requires a different set of approaches. While the edge- and region-based methods

for creating field objects complement the crop masking step, they use different inputs and techniques. Segmentation is not a binary task. Methods must be flexible to adequately respond to the visual complexity and ambiguity of automated imagery interpretation.

- *Post-Processing*
    - *Goal:* Use object characteristics such as size and shape to refine field polygons.
    - In this step, field outputs from the analytical processes can be cleaned up, stitched together into a single statewide crop field dataset, and prepared for assessment and implementation.
- *Accuracy Assessment*
    - *Goal*: Measure accuracy in multiple ways to accommodate different types of validation data.
    - Accuracy assessment is conducted on the final output making use of multiple validation sources at various scales. This includes ground truth spatial features from hand-digitized field boundaries at a local scale as well as preexisting crop mask layers and aggregated agricultural survey data at the county and state scales. These assessment methods can be deployed flexibly in other regions depending on availability and type of validation data.

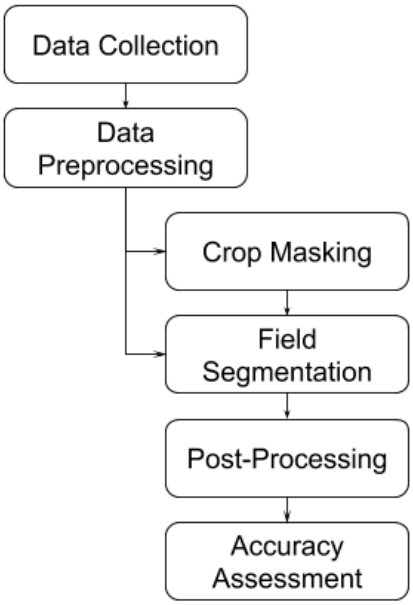

Figure 2. High level conceptual model diagram

Figure 2 shows the conceptual workflow. The methods deployed in our experiments work for Minnesota, but the workflow is designed to be generalizable. Since it does not use hand-curated training data and is a semi-supervised approach, this workflow can be shifted across both space and time. The only inputs are the imagery data and some high level parameters. Changing the Sentinel-2 tiles to a different year or geographic region, or even changing the input source to any set of remote sensing imagery with R, G, B, and NIR spectral bands, will not impede the pipeline from running. The flexibility of the workflow also enables local knowledge of crop types, temporal vegetative patterns, and spectral characteristics of the landscape to be incorporated into methodology decisions. For instance, while Minnesota's growing season typically lasts from planting in the spring to harvest in the fall, other regions might require a different range of imagery dates to capture the full temporal signature of crops.

Chapter 4 – Methodology & Workflow Configurations

How is the workflow deployed in Minnesota?

We tested this framework by conducting experiments in each of the core analytical steps of the workflow: data preprocessing, crop masking and field segmentation. We set up three different Workflow Configurations utilizing different approaches to field mapping. The Thresholding configuration implements a multiple criteria crop mask based on temporal characteristics and a set of user-defined thresholds followed by imagery segmentation to create a set of statewide field maps. The Clustering configuration utilizes unsupervised k-means clustering of temporal pixel features to separate land cover classes and a semi-supervised threshold step to assign crop or non-crop designations to each cluster to create a crop mask. Multi-temporal imagery is used to segment the crop mask into a statewide fields map. The Deep Learning configuration focuses on a smaller study area in order to explore the potential of using a cutting-edge deep learning CNN model for cropland mapping. It uses imperfect training data derived from a threshold-based approach similar to that in the first configuration to assess whether deep learning can help complement traditional methods to make them more generalizable and less dependent on hand-crafted training data. The Deep Learning workflow configuration is aimed at testing a concept so, unlike the first two, it only goes to the crop mask step in the workflow and does not produce a statewide field map. Each configuration aims to answer questions about implementing our framework for scalable crop mapping.

- *Thresholding Configuration*: Combined crop mask creation and field delineation through manual thresholding based on edge values and temporal NDVI change. Can we use a single set of threshold values to effectively mask crops at a statewide level?
- *Clustering Configuration*: Unsupervised cropland masking through k-means clustering of monthly NDVI values. Using clustering and a general understanding of local agricultural conditions in Minnesota, can we identify distinct cropland classes that share similar temporal vegetative signatures without having to manually choose a threshold value along a mask spectrum?
- *Deep Learning Configuration*: Using a deep learning model, can we use noisy crop mask data derived from the semi-supervised mask creation to train a convolutional neural network to improve the crop mask?

Each workflow configuration includes different approaches to data preprocessing to test the utility of various spectral indices, temporal statistics, and data reduction techniques. None of the configurations implement methods that rely on externally derived training data or any data sources other than the imagery. Wherever possible, we implement parallel processing techniques to optimize the computations. The structure of the workflow enabled us to test multiple approaches and string together Put together, these different configurations demonstrate the potential for efficient, scalable crop mapping that can be deployed across different geographies without relying on externally derived training data. Results from each configuration will be discussed in the next chapter to compare the outputs from different experimental approaches.

Imagery Data: Sentinel-2

To test these configurations, we use 10-m Sentinel-2 imagery. Sentinel-2 data is distributed in a standard spatial grid based on the Universal Transverse Mercator (UTM) coordinate system. Each of these tiles is approximately 100-km by 100-km, with overlapping spatial coverage with neighboring tiles, which we remove prior to analysis to ensure non-overlapping outputs. To achieve full spatial coverage for the state of Minnesota over the sixth month growing season of May through October, we incorporate all images over that period taken within the 37 tile footprints that overlap with the area of the state (see Figure 3). We utilize the Level-2A Bottom-of-Atmosphere product, to which corrections have been applied to adjust for distortion of the imagery from atmospheric aerosol particles. We limited the input bands to the four channels that have 10-m resolution and the 20-m cloud mask layer. All data used has been downloaded and stored at the Minnesota Supercomputing Institute and accessed using their HPC resources. Figure 4 shows RGB composite and NIR band images from a single date over an 8-km by 8-km sample area.

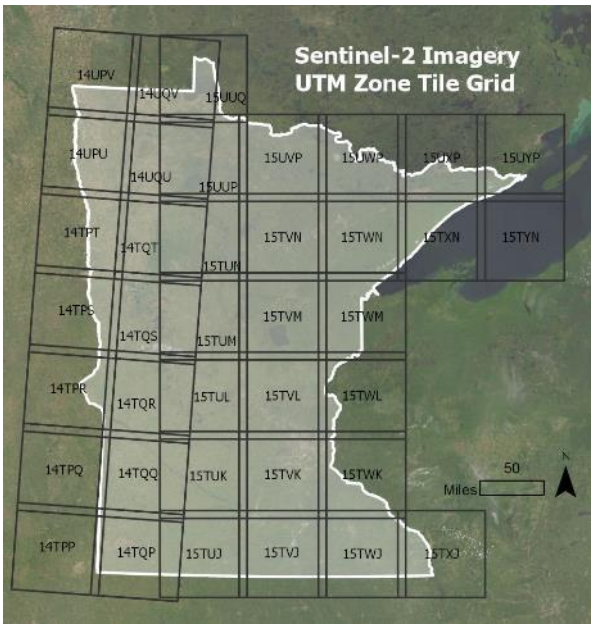

Figure 3. Map of Minnesota with Sentinel-2 imagery tile grid

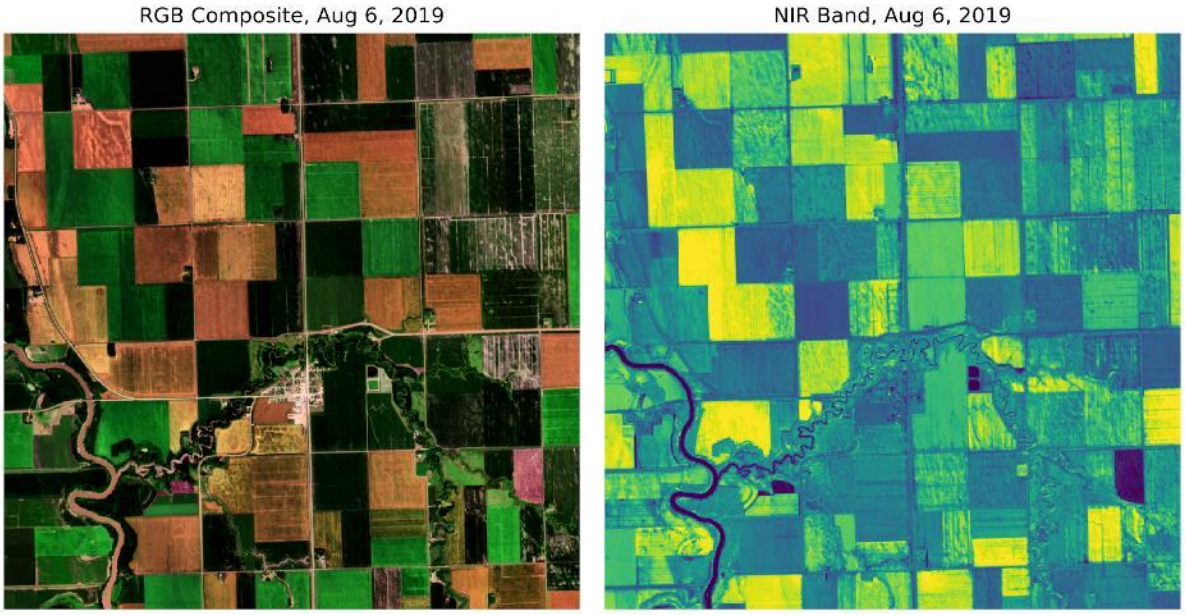

Figure 4. Sample RGB Composite & NIR Band of a tile subset covering an 8-km by 8-km area in MN

Validation Data

One of the challenges of conducting field-scale analysis for large areas is the lack of high-resolution field boundary data for validation over large areas. To overcome this, we utilize three validation data sources at multiple geographic scales. At the pixel-scale, the USDA provides statewide validation data via the Cropland Data Layer (CDL). At the field scale, a public dataset

of field boundary information does not exist, so we manually digitized validation field boundaries for a 15-km by 18-km area in Clay County, Minnesota using 1-m imagery from the USDA National Agriculture Imagery Program (NAIP). At the county-scale the USDA provides the nationwide Census of Agriculture. Using these three datasets, we can measure accuracy at the pixel, field, and county scales to assess at a high level how well the workflow performs at identifying the amount and location of cropland.

The USDA validation sources allow us to measure accuracy at the pixel and county levels. The CDL is a 30-m data product that estimates cropland and crop types at the pixel level for the United States. We utilize the 2018 CDL Crop Mask layer, which was the most recent data available at time of analysis. This Crop Mask layer includes any pixel that has been classified as any type of crop in two out of the five previous years of CDL models. There are two important limitations to this data as a validation source. First, the resolution is coarser than the 10-m Sentinel-2 data, which means that misclassification of small features and edges, particularly roads between fields, is common in the dataset. Second, due to the rolling temporal nature of the product—covering five years instead of just one—areas that are typically planted for cultivation in most years but left fallow in 2019 will show up in the CDL mask but not in our single year model. Therefore, it is not a true one-to-one validation dataset, but it provides a critical way to measure spatial accuracy across the state of Minnesota. The other method for measuring high level accuracy is by using agricultural survey data from the USDA, a common way to validate outputs in crop mapping research, particularly when it comes to yield estimation (Jin, Azzari, and Lobell 2017; Lobell et al. 2015). Data on total cropland area and cropland area harvested were obtained from the most recent USDA NASS Agricultural Census from 2017 (USDA 2019). This data is released nationally every five years, collected via survey to farmers, and aggregated to the county level. The temporal delay in the Agricultural Census means there is a two-year difference in census year and the imagery used in the study. While the census data does not correspond to the same year, it offers a reliable baseline for total cropland area. This provides a benchmark with which to evaluate the high-level accuracy of our crop mask.

In order to assess field-level outputs, crop field boundaries were hand-digitized for a 15-km by 18-km area in Clay County, part of northwest Minnesota's Red River Valley agricultural region (see Figure 5). The validation fields were drawn by hand in ArcGIS as a polygon shapefile from a 1-meter resolution NAIP image from 2019 to match the temporal coverage of the input data. While the input data for the model covers six months of imagery, the data used for digitizing the validation fields is a single snapshot in time. Even with this very high resolution imagery, field boundaries can be difficult to ascertain visually due to the ambiguity of delineating features

like fences between neighboring fields, the similarity between grassy fallow land cultivated cropland, and the occasional fuzzy boundary between fields and natural vegetation. These validation fields may contain errors but represent a close approximation of field-scale ground truth.

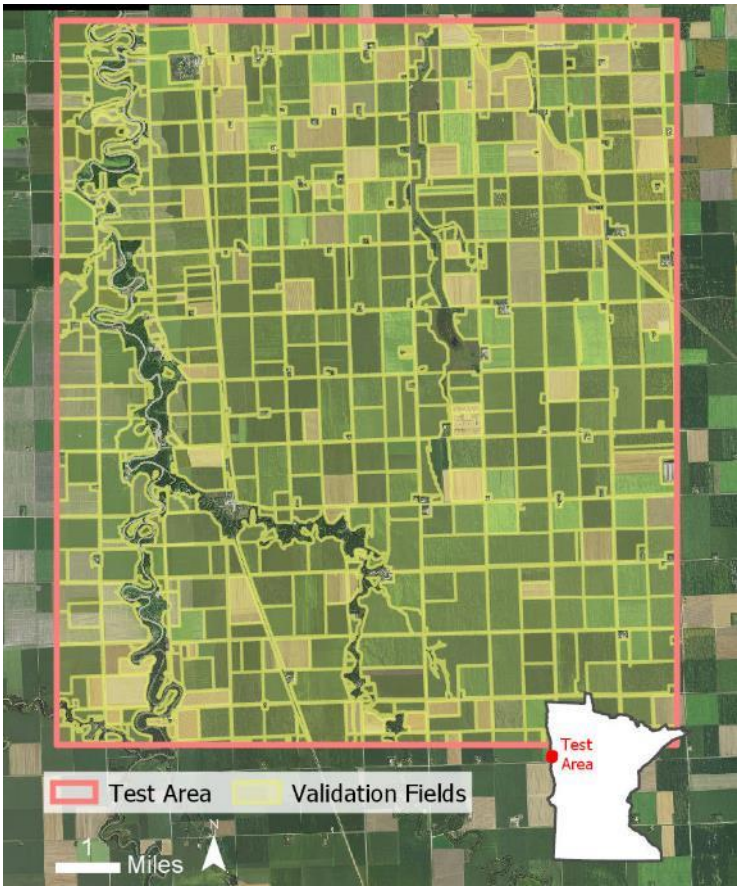

Figure 5. Map of validation fields over imagery with reference location

Using these three validate sources: 1) CDL Crop Mask, 2) USDA Census of Agriculture, and 3) hand-digitized validation fields allows us to assess the accuracy of the statewide field dataset at different spatial scales.

Accuracy Assessment

Using these three validate sources: 1) CDL Crop Mask, 2) USDA Census of Agriculture, and 3) hand-digitized validation fields allows us to assess the accuracy of the statewide field dataset at different scales.

With validation datasets that have spatial information at the pixel scale, we can assess accuracy at the resolution of the imagery. Using the CDL Crop Mask layer and the hand-digitized field boundaries, we compute the intersection of the predicted fields and the validation data to

produce an error matrix, a common approach to remote sensing research (Story and Congalton 1986). The error matrix records the count of True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) values from the intersection. From here, we calculate the Overall Accuracy (OA) value (Equation 5. Overall AccuracyEquation 5), which is the probability that a pixel in the validation layer will be correctly classified.

$$OverallAccuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

Equation 5. Overall Accuracy

While OA is useful at a high level, we also compute the Precision and Recall values. Precision (Equation 6) is the percentage of TP values out of the total number of predicted cropland pixels. It shows the rate at which the workflow is correct when it classifies a pixel as cropland. Recall (Equation 7) is the percentage of TP values out of the total cropland values in the validation data. This shows the percentage of all the cropland in the reference data that the workflow is able to identify. Both metrics are useful for evaluation because the workflow should be able to effectively differentiate cropland from non-cropland (Precision) and capture all the cropland in the study area (Recall). To capture both of these in a single metric, we compute the F1 score (Equation 8), which is the harmonic mean of Precision and Recall, giving a single value estimation of how effectively the workflow achieves these two tasks. With the CDL Crop Mask layer, we can aggregate these metrics to the county-scale for comparison with Agricultural Census data.

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

Equation 6. Precision

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

Equation 7. Recall

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

Equation 8. F1 Score

To compare the outputs to the Agricultural Census data, we compute accuracy at the county scale, which we can aggregate to the total statewide extent. For the statewide extent, we compute the Percent Error (Equation 9) between the total estimated acres of cropland from the workflow output with both the aggregated Cropland Acres and Harvested Acres for all of

Minnesota, which serve as an upper and lower bound for each county, respectively. The Percent Error is a ratio between the absolute value of the difference between the predicted value (V_a) and the expected value (V_e) compared to the expected value. To evaluate the county-level accuracy, we compute the Mean Absolute Error (MAE) (Equation 10) of all the counties compared to the Harvested Acres. This shows how many acres on average a county total differs between the predicted value and the value in the census reference data.

$$Pct.Error = \left| \frac{V_a - V_e}{V_e} \right| * 100$$

Equation 9. Percent Error

$$MAE = \frac{\sum_{i=1}^{n} |y_i - x_i|}{n}$$

Equation 10. Mean Absolute Error

There are multiple approaches for evaluating image segmentation outputs (Watkins and van Niekerk 2019). We implement the Variation of Information (VI) approach proposed by Meila (Equation 11), which estimates the amount of over-segmentation and under-segmentation between two sets of segments by evaluating how much information is shared between segments of the two data sets (Meilă 2003). The VI of two data sets uses ground truth (X) and prediction (Y), and is composed of the amount of over-segmentation $H(X|Y)$ and under-segmentation $H(Y|X)$. In a perfect segmentation these values will be zero. We report these values for each of the experimental outputs in addition to the spatial accuracy metrics.

$$VI(X,Y) = H(X|Y) + H(Y|X)$$

Equation 11. Variation of Information

Data Preparation

In all the workflow configurations, preprocessing of the raw data was conducted to prepare the imagery for analysis. The top and bottom 1% of data values in each of the four spectral bands—red, green, blue, and near infrared (NIR)—were clipped in order to remove outlier pixels that can skew statistical and optical analysis. Each band was then normalized to a range of 0 to 1 to account for differences in pixel value magnitudes between spectral bands.

All spectral analysis in this study is conducted at the unit of the tile footprint. When processing, the 10-m spectral bands are read from the database for each imagery collection date in a tile footprint into a multi-dimensional raster data structure using the open source Python package Xarray. This package is integrated with spatial data packages (like GDAL and rasterio),

multi-dimensional array packages (numpy, pandas), and parallel processing packages (dask). The data structure for each tile is constructed lazily, meaning that the underlying data is not read into memory until computation using that data is necessary. All the imagery for a tile can be combined into a unified dataset with spatial (x, y) and temporal (time) dimensions for each of the four spectral bands with consistent data preprocessing. Once constructed, analytical methods can be applied to this dataset efficiently, only loading data needed for specific operations and with built-in support for parallel processing to take advantage of the resources in an HPC environment. This data structure forms the basis of the pipeline and enables efficient, flexible, and scalable analysis.

Metadata included in the Sentinel-2 data allows for two strategies for handling cloudy images. The Sentinel-2 Level-2A data product includes quality assessment metadata, including the estimated percentage of pixels in the image with clouds. This enables only relatively cloud-free snapshots to be utilized in the processing workflow. There are trade-offs to this approach, however. If a quarter of pixels in an image are cloud-free but the image is not included in order to avoid those cloudy pixels contaminating the temporal analysis, then three-quarters of the usable pixels are discarded. In order to address this limitation, we utilize the cloud mask quality assessment layer included in the imagery data. This layer represents the likelihood of a pixel being cloudy with a value from 0 to 1. It has 20-m resolution and in order to use it to mask out cloudy pixels in the image, we resample it to 10-m resolution to match the spectral bands. This mask has limitations and tends to be conservative in identifying cloudy pixels, only masking out pixels with very high likelihood of being cloud contaminated. It also picks up false positives, where some features like rooftops can display reflectance values that will get picked up by the cloud mask even on clear days.

We utilize these two approaches to address cloud contamination to suit the methods of each configuration (Table 2). For the first, we utilize the cloud coverage percentage metadata value to only include images with less than 20% cloud coverage and we do not use the cloud mask. By excluding data to very clear images, it improves the quality of edges by removing the hard edges created by the cloud mask. In the Clustering configuration, since we depend on monthly median values, we raise the cloud coverage threshold to 50% but utilize the cloud mask to eliminate cloudy areas of image scenes. This enables more monthly values for each pixel but creates artifacts of the masking process. The clustering approach in the Clustering configuration handles these outliers better than the threshold approach used in the Thresholding configuration. Each approach to field mapping requires different attention to pre-processing and image selection for remote sensing data.

	Cloud Mask	Cloud Coverage Threshold
Config. 1 - Thresholding	No	20
Config. 2 - Clustering	Yes	50

Table 2. Cloud masking parameters

Data Post-Processing

The segmented field objects are then exported to a shapefile. The outputs from each tile are passed through a filter to remove any field objects smaller than 40,000 sq. meters, or ~10 acres. For the Thresholding configuration, we also simplified the polygon objects using the Shapely python package to remove polygon vertices while retaining the shape of the original polygon within a certain spatial tolerance. This is primarily to reduce the file size of the final shapefile outputs and reduce irregular edges produced by the vectorization process that converts the raster data to polygons. However, we found that at the field scale, these simplified polygons contained visual artifacts like small slivers and eliminated the alignment of boundaries between neighboring fields, reducing the visual coherence of the field objects. For the Clustering configuration the un-simplified version was chosen for the final output despite being more than twice the file size in order to retain the edges from the segmentation step. A different simplification approach could be beneficial in future work to smooth out remnants of the 10-m pixel grid from field boundaries but maintaining edge contiguity between fields took priority in this work.

The outputs from each tile are merged to form a statewide shapefile of crop fields. Since the state of Minnesota sits in between two UTM zones, the system that forms the Sentinel-2 tile grid, there are overlapping areas in the shapefile outputs. To deal with this, we merge the tiles from each of the UTM zones separately, clip each to the Minnesota state boundary and the UTM zone boundary, then merge these together into a single final statewide shapefile.

Thresholding Configuration: Multiple Threshold Crop Mask

Based on the crop mapping and field delineation literature, we implemented a field mapping workflow that takes advantage of the temporal characteristics of the imagery data to differentiate vegetation patterns and persistent edges over the course of a growing season. Specifically, this configuration compares peak vegetation to early season vegetation and incorporates the average edge magnitude for the entire time series. Generally, cropland goes from bare earth to robust vegetation throughout the growing season, while natural vegetation tends to green up earlier, showing a more consistent vegetative signal over time. Non-vegetation land

covers such as roads, buildings, and water bodies will have relatively low and consistent vegetative signals. Thus, we can use the range and maximum value statistics of NDVI to help differentiate cropland from non-cropland. Additionally, finding persistent edges in the imagery over time tends to reveal boundaries between landscape features and individual fields. Using the NDWI spectral index allows for water features to be identified and masked out of the cropland area.

To separate cropland from non-cropland, we compute a single 'crop mask.' To construct this mask, we compute each of the aforementioned metrics into their own layer, and then we apply a user-defined threshold to each, which enables the separation of cropland from non-cropland based on the information of each layer—vegetation, edges, and water. We then combine all of these layers together into a single, final crop mask layer. The crop mask is applied during the segmentation process to remove non-cropland area, which serves to reduce the amount of processing. We test three different sets of threshold values to explore the viability of using this method over a large area with diverse land cover.

Preprocessing

For this configuration, we did not use a cloud mask, but instead set a relatively restrictive cloud coverage threshold, limiting the images to those with less than 20% cloud coverage based on the Sentinel-2 tile metadata. The R, G, B, and NIR bands for all images that pass this threshold are added to a temporal imagery stack. The data for each tile was broken up into 2000 by 2000-pixel subareas (20-km by 20-km) for processing, resulting in "seams" between neighboring imagery regions. Each of the four bands was preprocessed to remove outliers and normalize pixel values. Wherever possible, analytical steps were parallelized in order to improve computational performance by taking advantage of the high-performance computing environment.

Mask Components

Each of the four mask components—NDWI Mean, NDVI Range, NDVI Maximum, Edges—is computed and thresholded separately. These components are then combined into a single crop mask using an OR function, where if a pixel is masked in any of the component masks, it is classified as a non-crop pixel. The threshold values used in each of the three statewide maps created with this process are shown in Table 3.

Thresholding Configuration	Parameters			
Map Version	NDWI	NDVI Range	NDVI Max	Edges
Thresholding Map 0.3	0.5	0.3	0.3	0.3
Thresholding Map 0.5	0.5	0.5	0.5	0.3
Thresholding Map 0.7	0.5	0.7	0.7	0.3

Table 3. Thresholding Configuration Parameters

1. *NDWI - Mean*

To mask out water bodies from the analysis, we compute the NDWI value at each pixel for all time steps in the input data. We then take the mean NDWI value across time at each location and pass a threshold value to create a binary NDWI mask identifying pixels that are likely water. In the sample image Figure 6 we can clearly see the main branch of the river feature and two man-made retention ponds in the mean NDWI image on the left. These features are masked out in the binary image on the right. Part of the smaller branch of the river can be identified but the NDWI signal is reduced when the feature is too small. We use the same NDWI threshold value—0.5—in each of the three maps in this configuration.

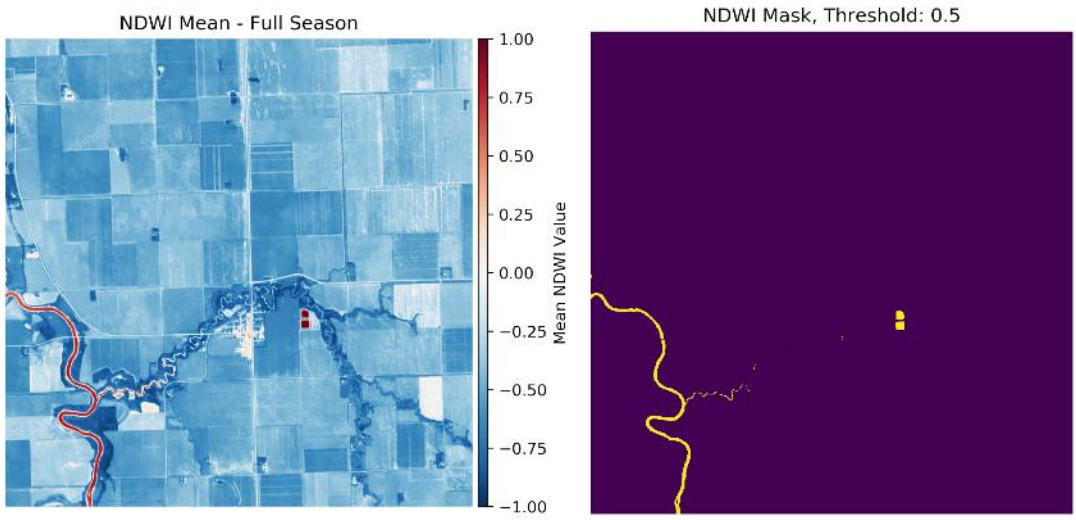

Figure 6. NDWI mask process

2. *NDVI - Change in Monthly Mean*

To assess change in vegetation over the growing season, we compute the monthly mean NDVI value at each pixel from May to October. Figure 7 shows how NDVI changes over time for

different land cover types. We can see that early in the season, naturally vegetated areas around the water features and in the lawns surrounding farmhouses are already displaying high NDVI values while most of the cropland area has negative NDVI values. During the summer months, as the crops in these fields grow, they show high NDVI values and return to low NDVI values at various points depending on when each field is harvested back to bare earth.

Figure 7. NDVI Monthly Mean Values

Using the mean NDVI value for the month of May as a baseline, we can subtract the maximum monthly NDVI value from the baseline to assess how much vegetative change happens throughout the growing season. Figure 8 shows the maximum monthly NDVI value on the left. We can see that cropland and natural vegetation both showed healthy vegetative responses at some point during the season, while roads remain near zero and water features display negative NDVI. By subtracting this from the mean NDVI value in May (middle), we can use the difference (right) to identify areas that had consistent vegetative response throughout the season. This difference map is then thresholded to create a mask of non-crop areas. We test three different values—0.3, 0.5, and 0.7—for this threshold (Figure 9).

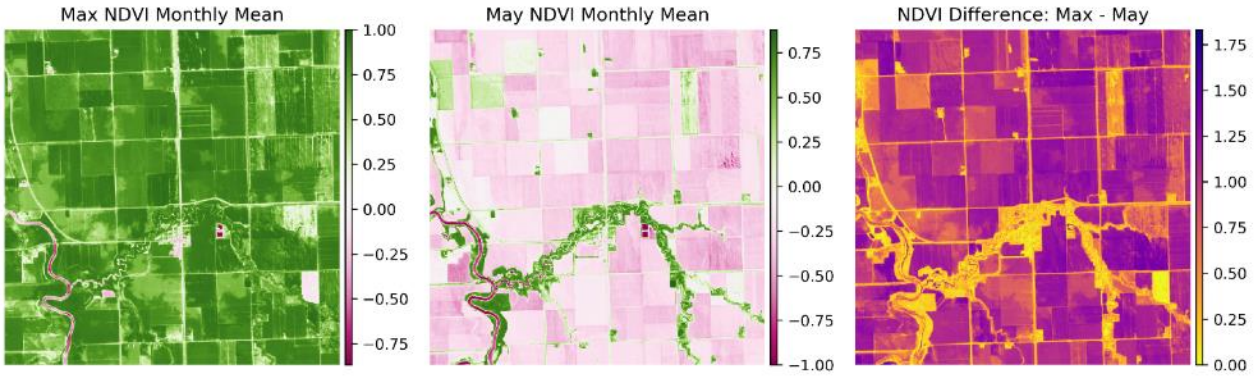

Figure 8. NDVI monthly mean change

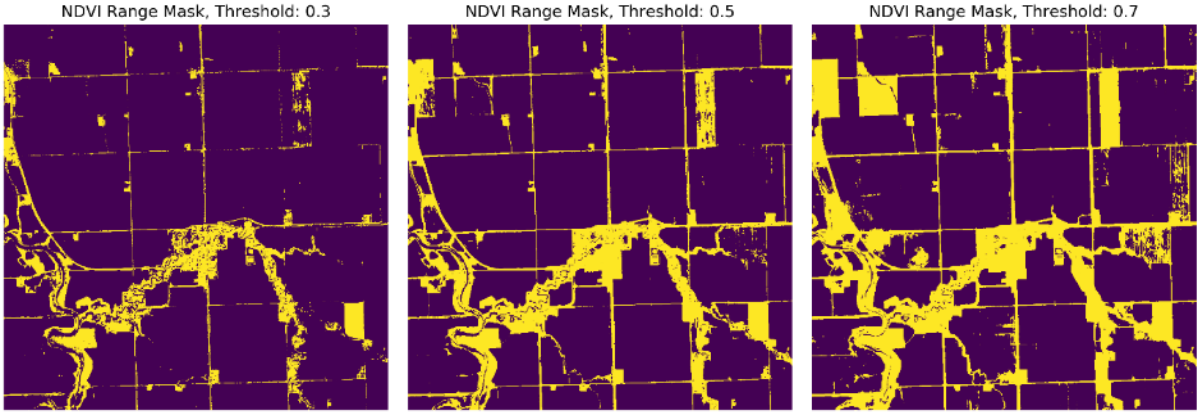

Figure 9. Mask thresholds for NDVI change

3. NDVI - Maximum Monthly Mean

While NDVI change is a useful indicator for masking non-crop areas, we can also use another method to filter out areas that can show a high rate of change but still not be crops by utilizing the maximum NDVI value, as shown in Figure 10. This mask targets vegetation that may change over the season but does not exhibit a strong signal representing a field of healthy, robust crops. Fields that are fallow, grassy areas between and around fields, and non-vegetative areas will be masked by this process. Figure 10 shows the three thresholds used—0.3, 0.5, and 0.7—for a sample area.

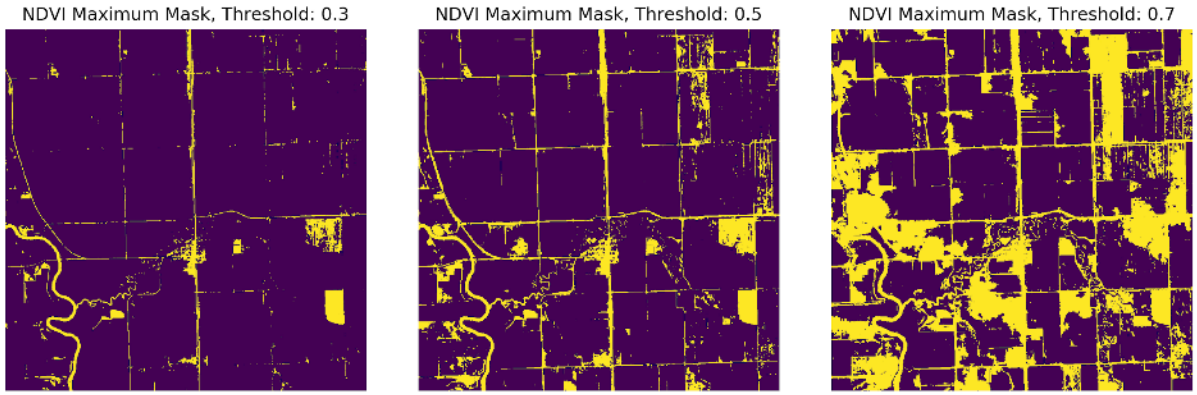

Figure 10. Mask thresholds for NDVI maximum

4. *Edges - Mean Edge Magnitude*

The final component of the mask process is an edge magnitude value. We utilize the common Sobel edge operator in our work, which emphasizes vertical and horizontal edges, which are common in the gridded agricultural landscape of Minnesota (Z. Wang et al. 2018). To compute this, we pass each band of each time step through the Sobel edge algorithm. Figure 11 shows the RGB composite and average edge magnitude for four different imagery dates. We can see that in each image, features like roads and rivers have consistent, high edge magnitude values while edges between fields show different strengths based on the similarity of the field appearance. When neighboring fields are both bare ground like in the June 5 image, the edge value between them is low, but over time the edge emerges more clearly, as we see in the July 2 and September 15 images. In the August 1 image, we see that clouds obscure the land in parts of the image and show up in the edge map. We mitigate this by taking the mean edge value across the full time series, reducing the impact of noise from cloudy images while emphasizing edges from persistent features.

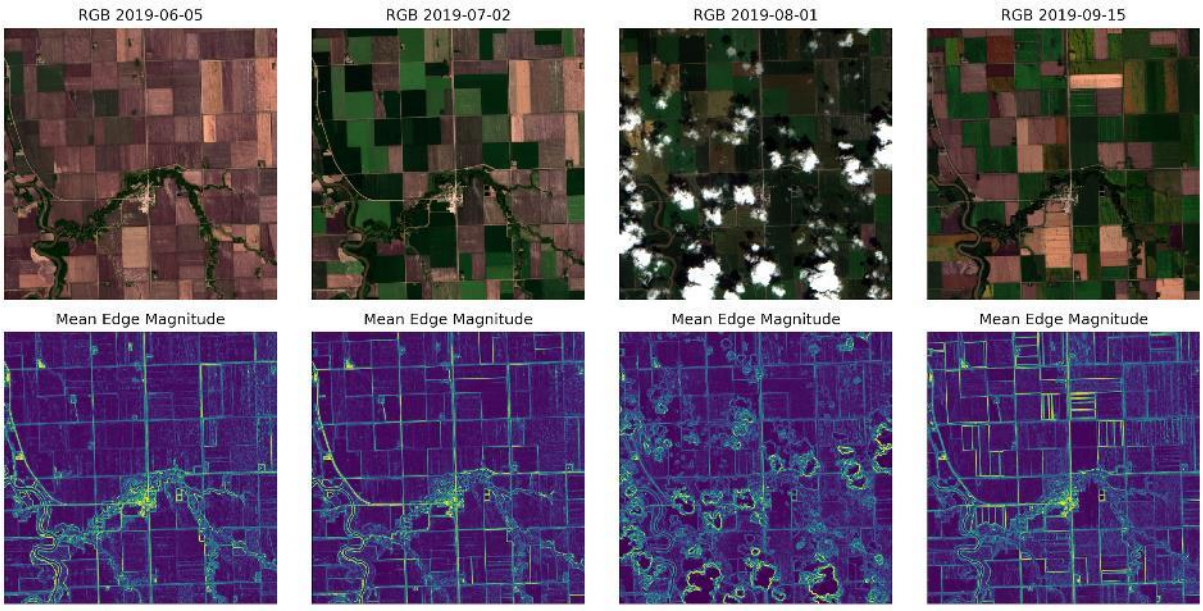

Figure 11. Edge magnitude maps for individual time steps

Figure 12 shows the final edge magnitude map, where noisy edges from the cloudy image have mostly faded to the background. Outlier values are clipped and then the image is normalized to a range of 0 to 1. From there, we can pass a threshold value to create a binary mask of edges. This method identifies many of the same features that the NDVI mask process did, specifically roads and rivers. Importantly, however, it can identify the boundaries between field features where they are otherwise spectrally similar in regard to temporal NDVI. The final image is still noisy and susceptible to masking out areas within fields that may not have consistent crop growth. We use the same edge magnitude threshold value—0.3—in each of the three maps in this configuration.

Figure 12. Mask threshold for full season edge magnitude map

Mask Assembly

Once computed, each of the four masks is combined into a single binary crop mask. If a pixel is classified as non-crop in any of the individual masks, it is classified that way in the final mask. Figure 13 shows each of the four component masks prior to being combined. Once combined, the mask is inverted so that cropland pixels are represented by the True value (shown in yellow). This is then passed to a clean-up process to produce the final crop mask, which will then be used in the segmentation step where the input data is converted to field polygon objects.

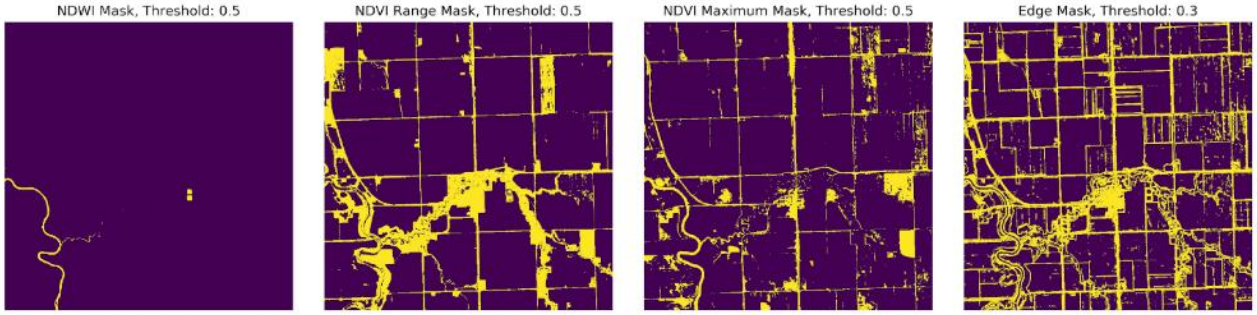

Figure 13. Component mask layers

To clean up the crop mask, which has salt and pepper noise typical of pixel-based image analysis approaches, we pass it through two morphological processes using the SciPy ndimage package. To remove errant pixels from within cropland areas, pixels that are fully enclosed by cropland pixels are converted to cropland. This reduces noise from the edge mapping step. Next,

in order to reduce errant cropland pixels and to smooth edges, we pass a 2 by 2 pixel minimum filter over the image. Figure 14 shows the combined crop mask before and after the cleaning step.

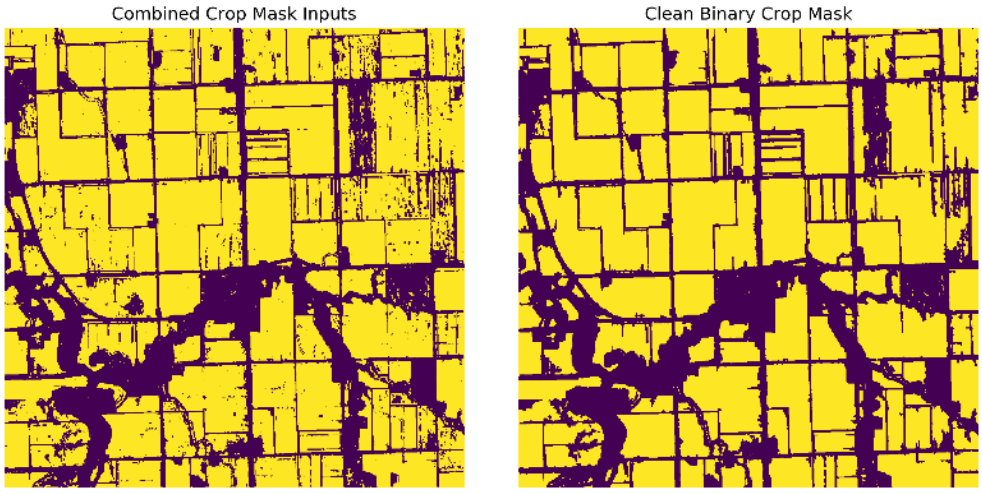

Figure 14. Crop mask cleaning process

The three different sets of threshold values produce different final crop masks, ranging from more optimistic predictions of cropland for the lower threshold mask (0.3) to more restrictive cropland predictions for the higher threshold mask (0.7). Figure 15 shows the final, cleaned up mask for a sample area at each threshold value.

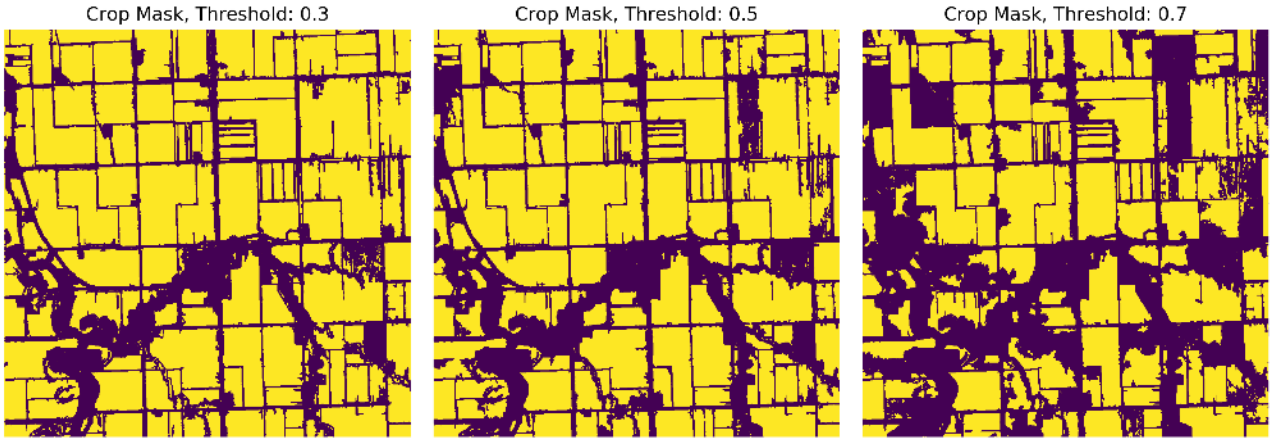

Figure 15. Final crop mask for each set of threshold values

Segmentation

Once the crop mask is computed, we select the RGB bands of a cloud free image to pass to the Felzenszwalb segmentation algorithm. For this configuration, we manually selected the image to be segmented for each tile to ensure that images are cloud free, have full spatial coverage free from data loss due to the path of the satellite over the tile footprint, and come from

part of the season where crops are near full vegetation. An automated process could replace this manual imagery selection step and is implemented in the Clustering configuration. We apply the mask to the image so that only non-masked areas get segmented into polygon objects. Similar to the combined edge-based and region-based approach of Watkins and Van Niekerk (2019) for field delineation, we utilize an edge based method within the mask process to define field boundaries but instead of a region-based segmentation method like watershed, we utilize the Felzenszwalb graph-based algorithm to incorporate spectral qualities of image. Figure 16 shows a masked RGB image used as an input to the segmentation algorithm (left) and the resulting segmentation with field polygons as yellow objects (right). The segmentation outputs are then passed to the post-processing step described above to create a statewide fields map.

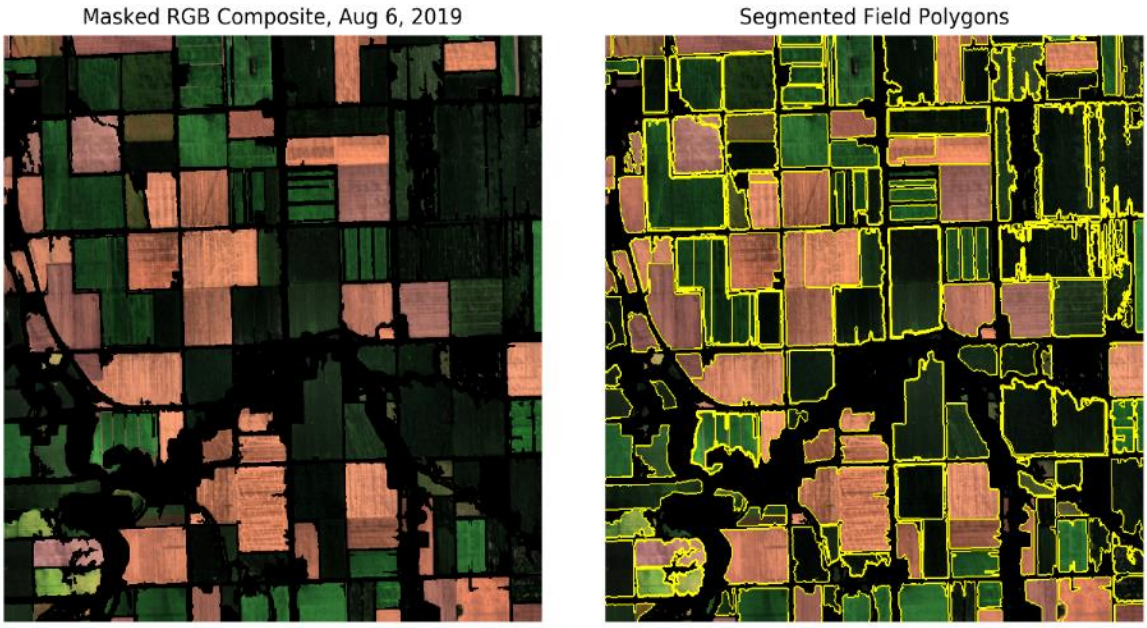

Figure 16. Masked RGB image and segmentation

Clustering Configuration: Semi-Supervised Crop Mask from K-Means Clustering

In order to overcome some of the limitations of using a single set of global threshold values to create the crop mask as in the Thresholding workflow configuration, we used a semi-supervised approach that combines unsupervised clustering with knowledge-based feature selection to differentiate cropland from non-cropland. We utilize the K-Means clustering algorithm identifies pixels that have similar spectral-temporal signatures across the growing season. Using the same observation that cropland has a vegetation curve that is distinct from other

land covers, as was shown to be effective in the prior configuration, the K-Means algorithm can identify those patterns from the data without the need for manual thresholding. We calculate the median monthly NDVI value for May through October, as well as the standard deviation of NDVI for the whole season and the mean NDWI value for the period of August through September, when vegetation is at its peak.

We used these temporal features as inputs to the K-Means algorithm, which clusters them into groups with similar temporal vegetative patterns. We sampled pixels with a stratified random approach across the entire state to ensure that there is a representative sample of each land cover type. We then fit a k-means model for each of the sampling areas with a *k* of 10 clusters. When the entire state had been sampled, we took the cluster centers from each of the sampling area models to fit a statewide K-Means model, reducing the influence of potential outlier samples. We experimented different values for *k* in the statewide model and found that 15 suitably identified variation within cropland areas while differentiating it from other land covers.

Once this unsupervised clustering was complete, we examined the characteristics of the clusters to identify which clusters fit the temporal characteristics of cropland. Based on several temporal vegetative characteristics of these cluster centers—the minimum, maximum, range, and standard deviation of the monthly NDVI median metric used for clustering—we passed a set of thresholds on the clusters to classify each as either Crop or Non-Crop. Unlike in the Thresholding configuration where shifting the threshold values selected pixels along a continuous spectrum, the threshold values here select from a discrete set of clusters to which pixels have already been assigned. Pixels that follow similar temporal vegetative curves will be grouped together without having to find the exact threshold values to characterize individual pixels. Crops demonstrate a distinct phenological vegetative curve over the growing season, which can be used to differentiate cropland areas from other land cover types in different clusters.

Determining the parameters for this binary classification of clusters is the only supervised input for this crop masking process and can be tuned to different locations based on local knowledge of agricultural practices without having to fundamentally change the method or generate training data. Thus, using unsupervised clustering with semi-supervised threshold values enables this approach to be geographically and temporally generalizable. In Minnesota, cropland typically goes from bare soil in the spring during planting to healthy vegetation during the summer with peaks coming at different times depending on the crop type followed by a sharp decline in vegetative value when crops are harvested. This informs the threshold values we use here but can easily be changed if the approach is deployed in a place with different agricultural

characteristics. We find that the accuracy of this approach improves upon the multi-threshold approach in the first configuration.

Preprocessing

For this configuration, we utilize the Sentinel-2 20m cloud probability layer as a cloud mask for each image. We include any image with less than 50% cloud coverage. This allows for more observations over the growing season for each pixel by including more images and removing only the cloudy pixels. This cloud mask is still not perfect and tends to under classify clouds, leaving cloud contaminated pixels in the masked image. However, by computing a monthly median value at each pixel, the influence of the cloudy pixels is diminished and more samples per month at each pixel location improves the robustness of the NDVI statistics.

Mask Components - Cluster Input Features

1. *Monthly median NDVI*

In order to classify cropland pixels from natural vegetation pixels, we computed the median NDVI value for each month from May through October. Like the monthly mean approach in the Thresholding configuration, this enables the differentiation of natural vegetation from cropland. Figure 17 shows monthly median NDVI values for an area in northwestern Minnesota. We can see that natural vegetation greens up earlier in the season and stays green through October after most crops have been harvested. The dark pink rectangular areas are man-made ponds, with water showing low NDVI values. The K-Means algorithm clusters pixels with similar temporal vegetative patterns, capturing the green-up and harvest changes in the time series.

We use the median instead of the mean for this approach because cloud contaminated pixels that are not masked out skew the monthly mean values. In order to limit the influence of cloud noise, we include more images and utilize the cloud mask, so that the median value is a valid cloud-free observation. Sometimes, however, reflective surfaces can be incorrectly classified as clouds in the mask layer and removed from the imagery, as can be seen in the white areas in some of the pond features in the July, September, and October images.

Figure 17. Clustering algorithm inputs: monthly median NDVI

2. *NDVI standard deviation*

Another input for the clustering is the full season standard deviation for NDVI for each pixel. This statistic is helpful for differentiating cropland, which shows large changes in NDVI values over the growing season, from other land cover classes that do not typically show as wide of a spectral change. We did not use mean, minimum, maximum, or range in the NDVI statistics as other researchers have done because errors from the cloud masking process introduced noise in the minimum, maximum and range outputs. Areas where clouds were not masked produced outlier values and visual artifacts. Like the median values in the NDVI time series, standard deviation was much more robust to this noise given enough valid observances throughout the full season. Figure 17 shows that crop fields stand out against road features and natural vegetation. Water features and surrounding vegetation, however, can show large changes in NDVI for areas with aquatic vegetation or where water levels shift over the course of the season.

3. *NDWI for peak growing season (July, August, September)*

To help differentiate water features, we compute the mean NDWI value of each pixel for the time period of July, August, and September, during the time of peak vegetation in Minnesota, when the difference in NDVI and NDWI signals are larger. In Figure 18, we can see the rectangular ponds and natural water features in bright red with high NDWI values, as well as road surfaces and patches of bare earth with positive NDWI values.

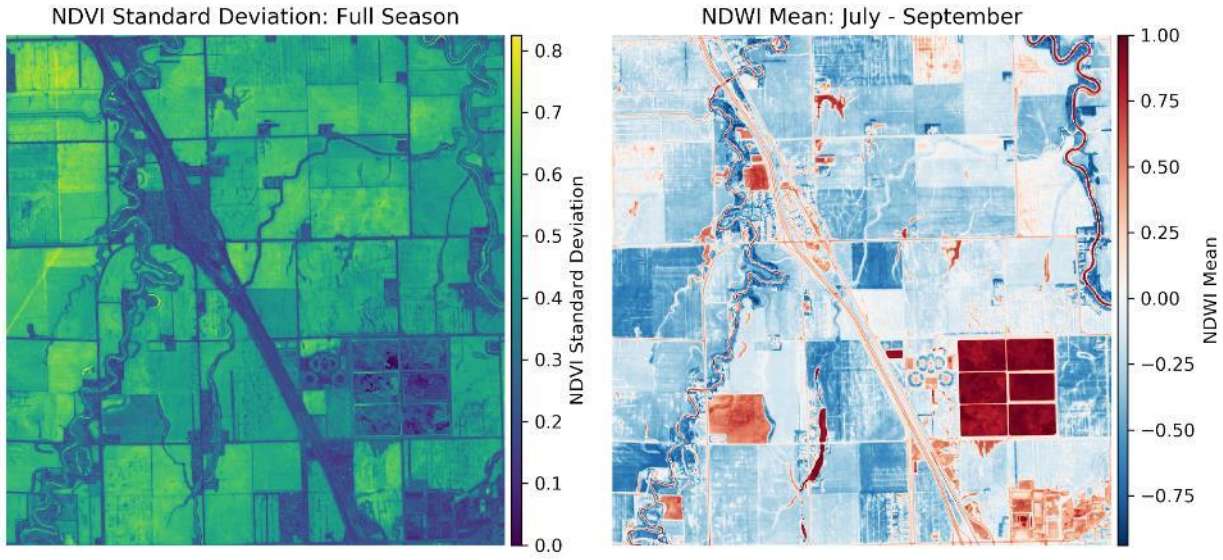

Figure 18. Clustering algorithm inputs: NDVI standard deviation and mean NDWI

Mask Creation

The final set of data features used for the k-means clustering process were monthly median NDVI for May through October, NDVI standard deviation for the whole time series, and mean NDWI for July through September. As each tile was preprocessed to create these input features, 5% of each 5,000 by 5,000 pixel area (four per tile) was extracted to train a k-means classifier with a *k* of 10. Once the entire set of data was sampled, the cluster centers from each of the sample area models were used to train a statewide K-Means model, ensuring representative samples from each land cover type in the state and reducing potential outlier values. We used a *k* value of 15 for the statewide model. These 15 statewide cluster centers were visualized and classified as crop or non-crop based on the temporal NDVI characteristics. We utilize the K-Means implementation found in the Python package Scikit-Learn, with the k-means++ initializer option which speeds up convergence. While this algorithm is computationally intensive, requiring recursive distance calculations for each data point, there are variations of k-means that are more efficient and can be processed in parallel (Kumar et al. 2011; Lv et al. 2010; Pelleg and Moore 1999).

The temporal characteristics of each cluster reveals insight into whether it represents cropland or not. We use the intermediate results of the temporal characteristics of each cluster to identify which clusters represent cropland. Figure 19 shows a heat map of each of the cluster centers (represented as rows) across the eight input features (columns) and Figure 20 plots the same data with NDVI standard deviation and NDWI mean values plotted as points separate from

the monthly median NDVI. Note that the temporal NDVI values go from -1 to 1 while NDVI standard deviation and NDWI mean go from 0 to 1. Several distinct temporal patterns help identify what each cluster represents.

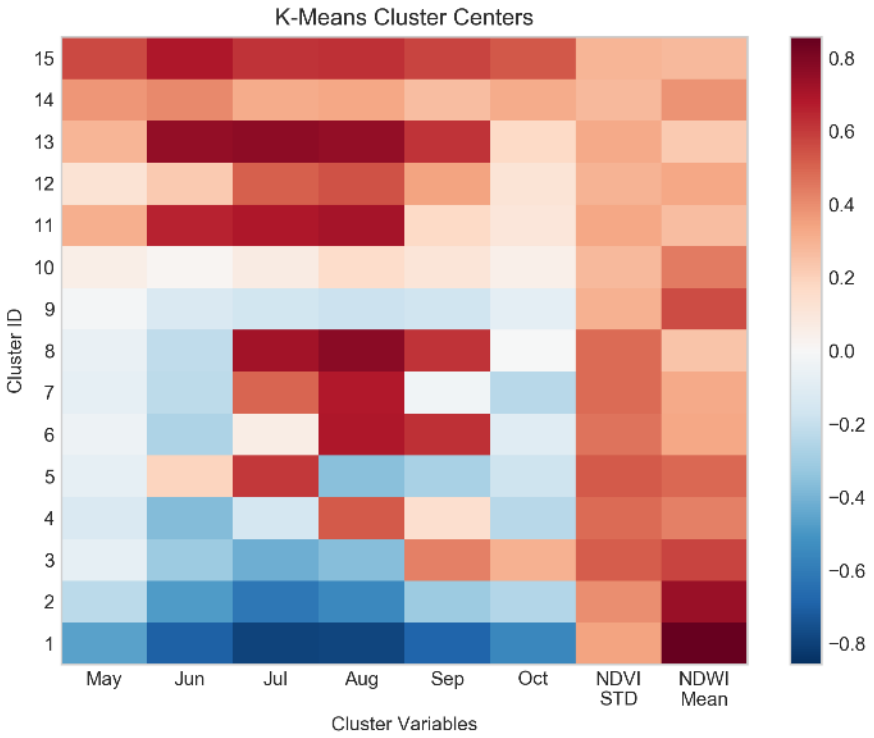

Figure 19. Statewide K-Means Cluster Centers feature heat map

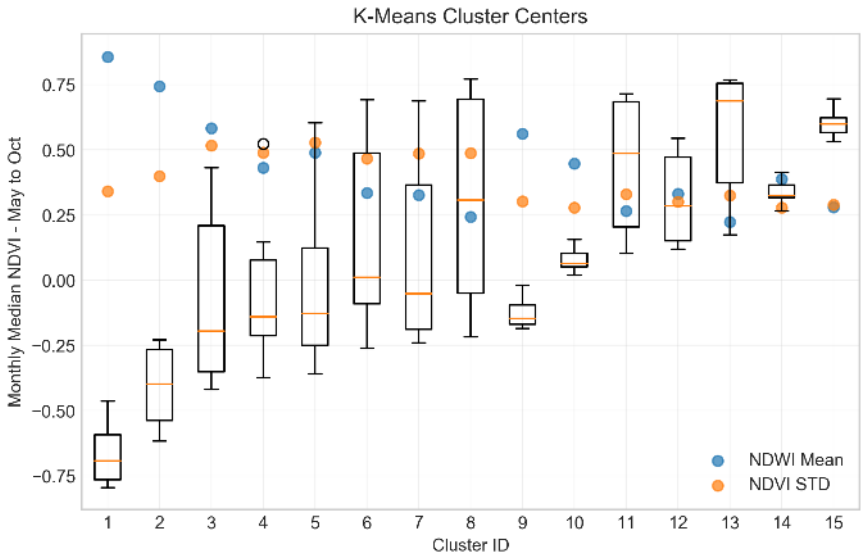

Figure 20. Statewide K-Means Cluster Centers feature plot

Clusters 1 and 2 at the bottom of the heat map in Figure 19 and left side of Figure 20 show negative NDVI values for the full time series, low NDVI standard deviation, and high NDWI values, suggesting that these clusters represent water features.

Clusters 3 through 8 demonstrate characteristics of cropland. They show low NDVI values at the beginning of the season, when most cropland in Minnesota is bare earth prior to the crops greening up. Throughout the summer months, these clusters show high NDVI values followed by a sharp drop back to negative NDVI, indicating a harvest. The exception is cluster two, which shows late season growth that is not harvested by October. Despite the timing of the vegetative change, all of these clusters show a large NDVI range, with a minimum below zero and a maximum Even though the peak month for NDVI varies between these clusters, they can be identified as cropland without having to identify beforehand when peak vegetation occurs, thus identifying different varieties of crops and planting methods. These clusters also show high NDVI standard deviation, reflecting the sharp changes in vegetation throughout the season, and relatively low NDWI values since most of them exhibited strong vegetation during the months that NDWI was measured.

Clusters 9 and 10 appear to represent non-vegetation land covers that remain consistent throughout the season. Both have NDVI values close to zero and low NDVI standard deviation, suggesting that they are features like buildings or roads.

The remaining clusters, 11 through 15, indicate vegetation throughout the full season, but without as drastic a change as the crop clusters. Each of these five clusters starts and ends the season with a positive NDVI value, indicating the presence of consistent vegetation. Clusters 11, 12, and 13 show a similar green up period as the crop clusters during the summer, but do not show as drastic of a change from the negative NDVI values typical of bare soil. Clusters 14 and 15, on the other hand, do not show very much variation at all, indicating consistent vegetative cover.

Classifying Crop Clusters

Based on the observed characteristics of the clusters we can define criteria to classify each cluster as either crop or non-crop, creating a statewide cropland map that we can use for field delineation. If any of the conditions in Table 4 are true, the cluster is classified as “Not Crop”:

Clustering Configuration. Mask Threshold Values

	NDVI Monthly Median			
NDWI Mask	Standard Dev.	Min.	Max.	Range
>0.2	<0.2	>0	<0.3	<0.7

Table 4. Clustering Configuration Parameters

These parameters were selected based on visual inspection of the characteristics of the cluster centers. Figure 21 shows the classified clusters spatially. We can see based on the reference RGB image that the clusters correspond with landscape features in the reference image. The middle image shows only clusters that fit the crop masking criteria. Crop fields corresponding to different peak vegetative months show up in shades of red, green, and purple. In the central image showing just the crop clusters, man-made features like roads (crossing diagonally from top to bottom) and buildings (in the bottom right corner) as well as water features and the natural vegetation surrounding them have largely been masked out in the clustering process. This approach does not perfectly isolate crop pixels from non-crop pixels. There are mixed classes and areas where pixels appear to be misclassified as cropland, specifically areas classified as cluster 3 in orange. Some fields appear as distinct objects while others have multiple classes within them. Since it is a pixel-based method, clusters are not always discrete units like crop fields. In areas with mixed vegetation, some pixels end up with spectral characteristics similar to crops. When this clustered map is converted into a binary crop mask based on the masking parameters, those mixed areas will be combined into cohesive areas. However, these mixed classes limit the utility of this method to map fields directly into objects, even though in many areas it does a good job at delineating individual fields. A segmentation step is still necessary to delineate field boundaries.

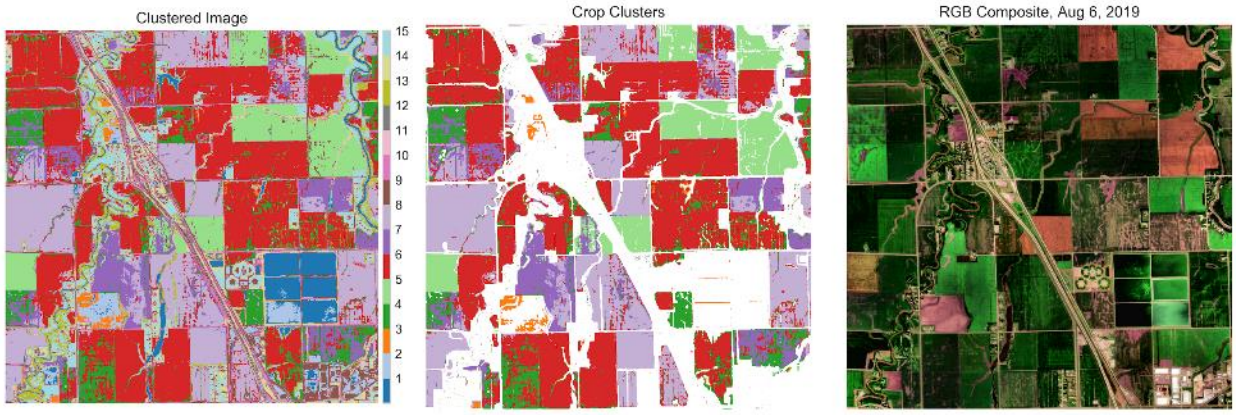

Figure 21. Clustered image with (left) all clusters, (middle) crop-designated clusters, and (right) reference RGB image

In addition to incorporating NDWI in the clustering process, we use a separate input for removing water bodies from the crop mask layer. We pass a single threshold to the average NDWI value during months of peak vegetation. We use a 0.2 threshold value for the NDWI mask threshold. We are less concerned with the confusion between built-up features and water bodies that McFeeters (2013) found when using a 0.3 NDWI threshold in mixed urban landscapes because these human made features would also fall into the non-crop classification through the clustering process.

Mask Assembly

Once the two mask inputs have been generated by thresholding the NDWI and crop cluster layers, we apply morphological methods to clean up the crop mask. This step reduces noise generated by the clustering process, eliminating small holes within fields, smoothing edges, and removing isolated pixels. First, we pass the NDWI mask through a set of closing and dilation filters to remove isolated pixels within the mask. Then we combine the two masks with an OR function, where any pixel classified as crop in either mask is preserved in the combined mask. We then pass this combined mask through a set of filters to fill small holes and smooth edges. Figure 22 shows the two input masks prior to the morphological step and the final mask after it has been cleaned. Notice that some of the misclassified areas from the crop clusters have been removed by incorporating the NDWI mask. While this process does not remove all noise, it eliminates many erroneous pixels. Small unmasked areas will be removed after segmentation once they are converted to objects and can be filtered by size.

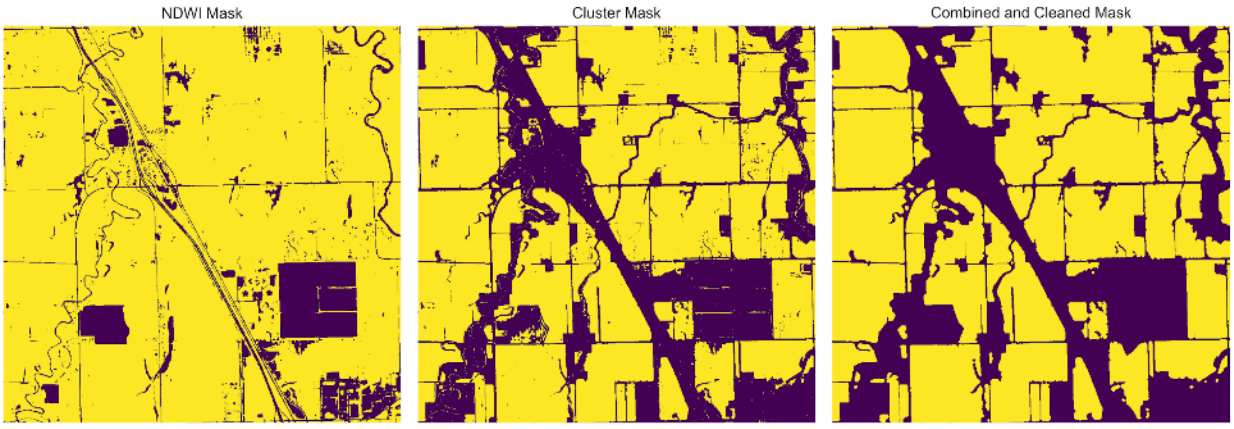

Figure 22. Crop mask components (left) NDWI, (middle) crop clusters, and (right) the final clean mask after morphological processing

Segmentation

Once the crop mask is generated, we apply the crop mask to the raw imagery and pass a set of masked imagery to the segmentation algorithm to delineate boundaries and output a fields shapefile for each tile. Unlike in the first configuration where we used a single RGB image as the input for the segmentation algorithm, for this configuration we utilize a method demonstrated by North et al. (2019) for visualizing fields in distinct ways. They combined the NIR band from three different cloud free imagery dates into a single three-channel RGB composite. This way, the color space in the image includes temporal differences across the growing season, highlighting boundaries between fields that might otherwise be difficult to see in a single image where neighboring fields both have a healthy green vegetative appearance. In order to select cloud free images, we selected the image closest to July 15, August 15, and September 15 with cloud coverage below 5% for each tile, ensuring that the same image was not selected for multiple dates. Figure 23 shows a comparison between the NIR temporal composite image and an RGB composite from a single date with segmentation outputs from each. We can see that the different input images produce different segmentation results. Once the segmentation step is complete, the shapefiles are passed to the post-processing step described above where they are merged into a final statewide fields map.

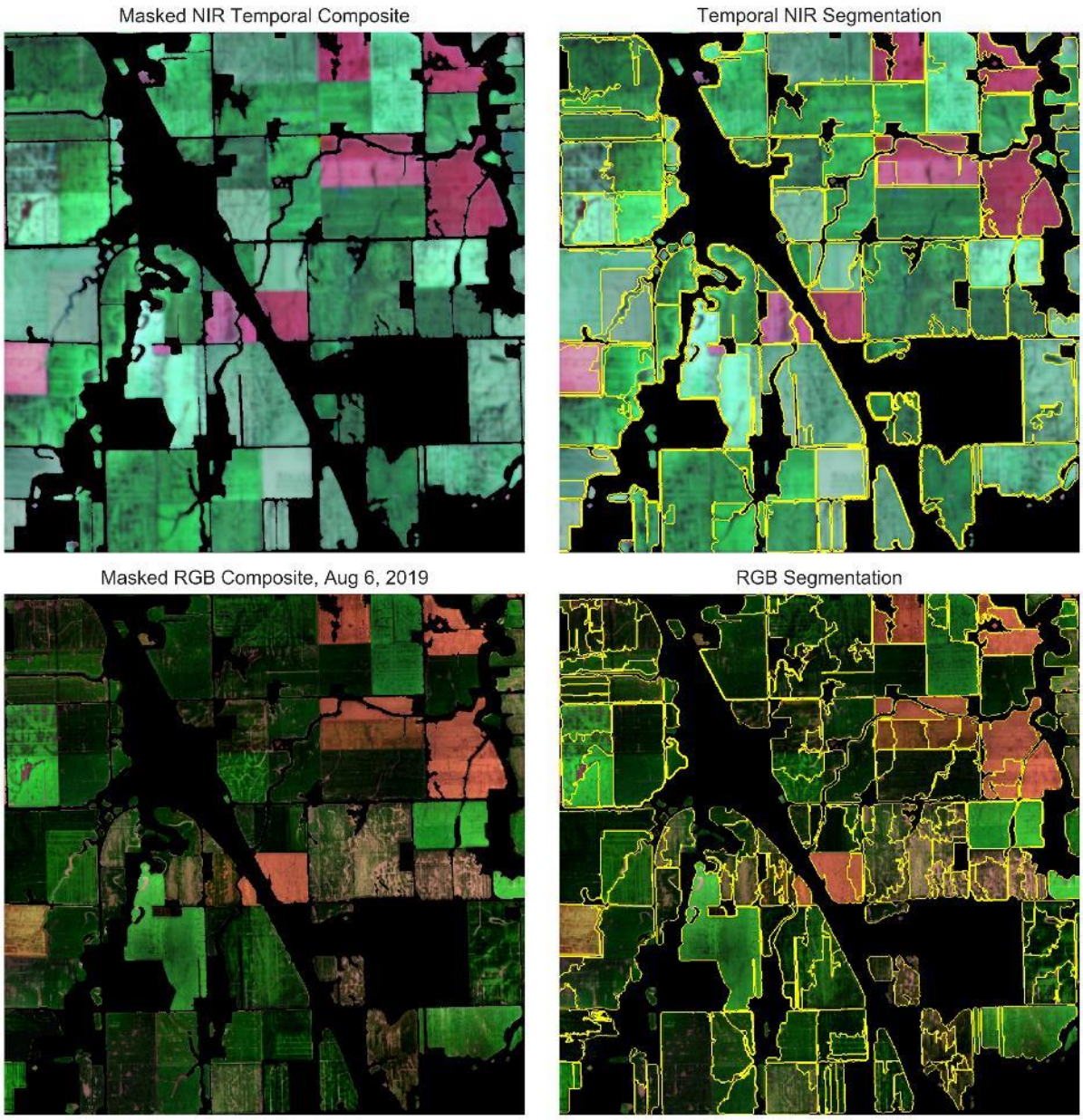

Figure 23. Comparison of segmentation inputs. NIR temporal composite image (top left) with resulting segmentation (top right) and single date RGB image (bottom left) with resulting segmentation (bottom right)

Deep Learning Configuration: Using Imperfect Training Data to Improve Crop Mask Accuracy

One of the key limiting challenges for an unsupervised pixel-based approach like k-means clustering is that it looks at each pixel individually. For problems like field mapping, and many other spatial data applications, the context of a pixel is integral to understanding what it represents. A pixel in a field could have similar spectral characteristics as a pixel in another

vegetated area, which is why the techniques used in the first two configurations misclassify areas. However, looking at the neighborhood of that pixel reveals that it is surrounded by similar pixels that are in turn bounded by distinct edges, which is what allows humans to interpret imagery. The emergence of deep learning models like CNNs, which combine layers of convolutions that look at pixel neighborhoods across multiple scales in an image, enables spatial context to be incorporated into spectral analysis for advanced image classification and segmentation.

Like many other machine learning methods, however, CNNs typically require large amounts of training data to learn the characteristics of spatial features. In our case, using hand-digitized validation fields to train a model to map crop fields from imagery would be impossible for large extents like the state of Minnesota. Pixel-level labels are labor intensive to create by hand, but with the methods demonstrated in the first two configurations, we can generate pixel-level crop masks with high precision across large areas. If traditional remote sensing methods can generate training data for deep learning models, these two approaches can complement one another. The automation of training data could potentially remove the core barrier to deploying the strengths of deep learning for cropland classification and feature extraction across large, diverse areas. We test a common CNN architecture, known as a U-Net, using noisy crop masks as training data to explore the potential of unsupervised deep learning for crop mapping.

The approach for the Deep Learning configuration differs from the first two configurations in that it was only conducted for a single Sentinel-2 tile, not the entire state, and it uses hand-picked imagery from throughout the 2017 growing season, not all available imagery for the 2019 season. This configuration is designed to be exploratory, and in particular to assess whether integrating deep learning into the workflow could overcome some of the limitations of traditional remote sensing and machine learning methods. This approach focuses on the crop masking step of the framework and does not include a segmentation step. Instead, the crop mask is compared to the outputs to a rasterized version of the hand-digitized validation fields.

We conduct four experiments with the CNN model testing different model inputs and ensemble predictions with multiple models. The first experiment tests whether the approach is feasible. We assess whether a model trained with imperfect labels can improve upon the test accuracy of those same labels. In the second experiment, we vary the input data to test whether including an edge magnitude layer in the inputs improves the accuracy. Third, we train a model with multiple crop mask inputs of varying precision to assess whether the addition of more imperfect training data improves accuracy. And finally, we use multiple models trained with different imperfect labels to make ensemble predictions.

Study Area

The study area for this configuration is an approximately 100 km by 100 km region in northwestern Minnesota and eastern North Dakota. This region, part of the Red River Valley, is a predominantly agricultural area, but it also includes river features, natural vegetation, and some urban areas, namely parts of Fargo, ND. The area coincides with the coverage of a single Sentinel-2 tile (ID: T14TPT) as shown in Figure 24.

Figure 24. Study area and imagery tile extent for Experiment 3

Data

The imagery source for this experiment is Sentinel-2 10-m imagery comprised of the R, G, B, and NIR bands. The study period covers five dates covering the 2017 growing season: May 13, June 22, July 20, August 29, and October 20. This is a different time period than the previous two workflow configurations. The selection of dates allows for a phenological curve of vegetative growth over time to be present in the data. The full imagery tile was not used due to regions of no data in several images from the path of the satellite swath. Identical to the first two configurations, the imagery bands were normalized, and outliers clipped. The validation area remained the same, but the fields were manually revised to correspond with the study year using NAIP 1-m imagery to 2017 delineate field boundaries. In order to capture the field edges, which are often closer than 10-m in the polygon validation layer, the edges were buffered by -10 meters prior to being converted to the raster layer that was used for accuracy assessment. The dates were hand-picked to be cloud-free so a cloud mask was not applied. Figure 25 shows the area used for training and testing. The imagery within the test area was withheld from the imagery used for training.

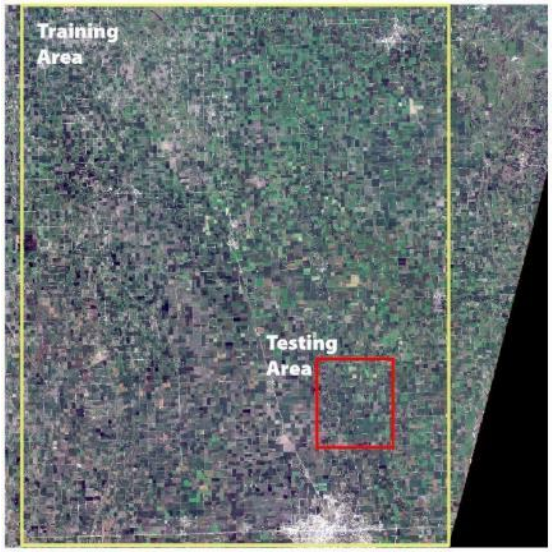

Figure 25. Training area and testing area for the Deep Learning configuration

Model Architecture: U-Net

This configuration implements a U-Net model, which is an encoder/decoder style CNN originally proposed for medical image segmentation that has become popular for semantic segmentation of remote sensing imagery (McGlinchy et al. 2019). This architecture uses a series of convolutional layers analyzing the 3x3 pixel neighborhood at each location of the image and max pooling to down-sample the input image to coarser resolutions by taking the maximum value of each 2x2 pixel area. Through this down-sampling process, the model is able to learn characteristics of the imagery at both a high resolution, like a single pixel, and zoomed out to see that pixel within the surrounding field area. This process encodes the image into a feature space that is no longer explicitly spatial (Ronneberger et al. 205). From there, the right side of the model decodes the information from the feature space back into the original imagery resolution using up-convolutional filters that reintroduce information using neighboring cells. Skip connections link the left side to the right side so that higher resolution information from closer to the original image can be integrated into the up-sampling process. During learning, weights in each layer are updated by minimizing a loss function that compares the output to the training labels. Figure 26 visualizes an example U-Net, with the encoder on the left and decoder on the right, connected by skip connections. The model outputs a probability value for each pixel, with 1 being cropland and 0 being non-cropland. To convert this to a binary classification, we rounded each pixel probability to the nearest whole number. We trained each model for 30 epochs in batches of 20 using the Adam optimizer and binary cross entropy as the loss function. The model was implemented in Python using the Keras package with a TensorFlow backend. This type of

architecture is suitable for cropland mapping because it is flexible in the amount of input layers it can take, allowing for multi-spectral, multi-temporal imagery inputs.

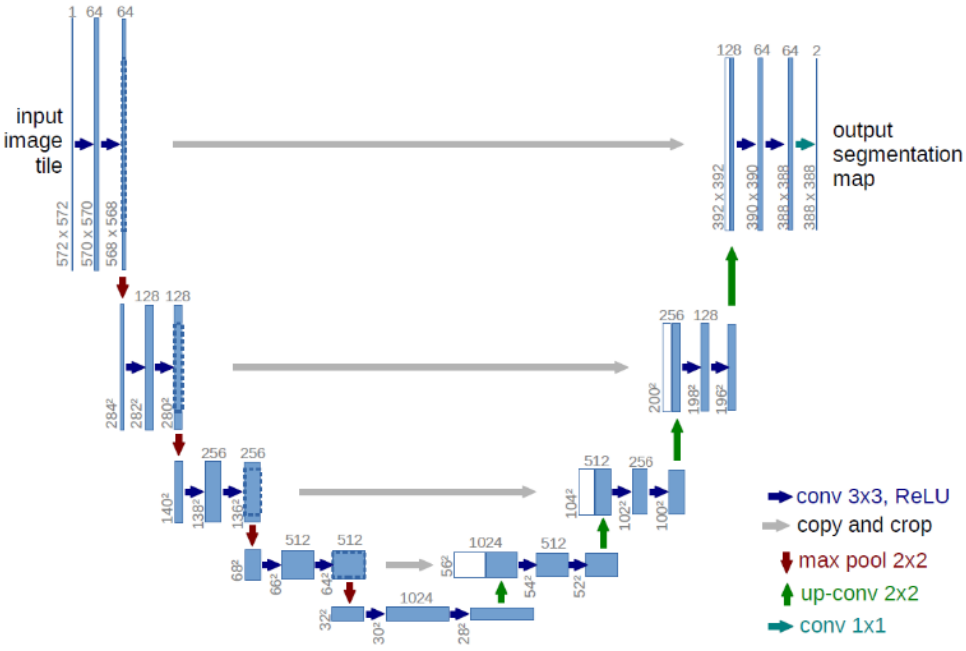

Figure 26. U-Net architecture example (Ronneberger et al 2015)

Preprocessing

The first preprocessing step was to create a series of crop masks to use as training data for the CNN model. We will refer to these layers as the Crop Mask Labels, as they provide the labeled data to train the models. We the same preprocessing steps as in the Thresholding configuration, but instead of thresholding each component individually to create the mask, we combined the mask components into a single composite crop probability layer prior to thresholding. For the mask inputs we computed mean edge magnitude and the NDVI range over the five imagery dates. We normalized each of these layers from 0 to 1, weighted the NDVI range to be one half of the edge magnitude based on visual inspection, and added them together. We then pass a single threshold to this crop mask layer. Once created, we passed a minimum filter to the mask to reduce noise within the fields. Figure 27 shows the simplified masking process.

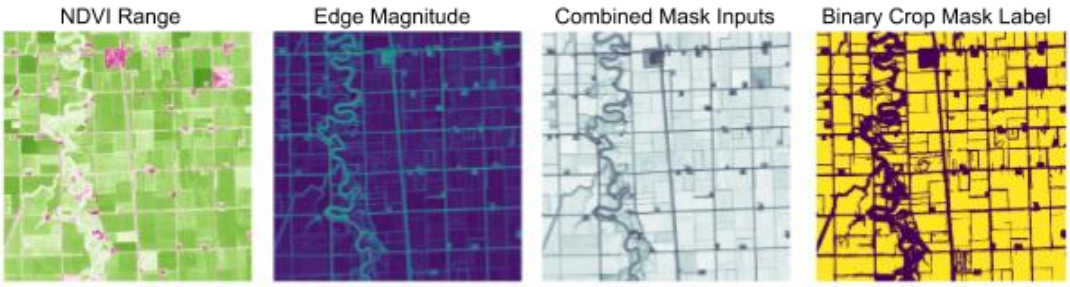

Figure 27. The NDVI Range (left) and Edge Map (left-center) are combined to create a crop mask probability map (right-center) and then converted to a binary crop mask (right) by a simple threshold.

The binary Crop Mask Labels were calculated with six different thresholds (16, 18, 20, 22, 24, and 26) based on visual inspection to identify the range of values that represent over- and under-estimation of crop area. Figure 28 shows what the range of thresholds look like on a sample patch. These are split into corresponding patches and used to train the CNN model. To prepare the input data for the CNN model, we computed NDVI for each time step and then the R, G, B, NIR, and NDVI bands were stacked along the temporal dimension to effectively create a raster with 25 bands. This raster was split into 256 by 256-pixel patches to be passed as input to the model. The edge map layer was also split into patches and used as an input to the model in the second experiment.

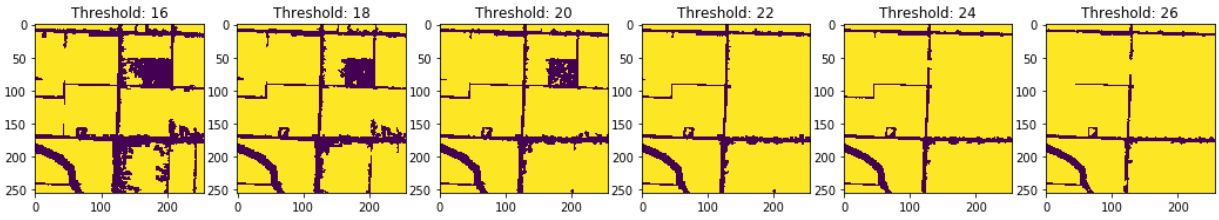

Figure 28. Sample patch at each crop mask threshold

We compared each of the Crop Mask Labels to the hand-validated test labels to determine the baseline accuracy for each Crop Mask Label threshold value. The model accuracy will be compared to these baseline values to assess whether they improved the cropland prediction. One of the core assumptions of Mithal et al. (2017) for using imperfect training labels is that the majority of the labels are correct, even if imperfect. We can see in Table 5 that despite having a range of class ratios between the different mask thresholds, the majority of the Crop Mask Labels are correct in all of the training masks. The precision in all the Crop Mask Labels is above 94%, but the recall varies widely, indicating that at lower thresholds many crop pixels are missed. Additionally, in the validation set, 80.4% of the pixels were classified as crop. The higher threshold Crop Mask Labels had a similar distribution, but the ones with lower threshold were

much noisier, with the threshold 16 mask being split almost evenly between the two classes. These Crop Mask Labels provide a range of "noise" to test with the CNN model.

Label Set	Class = 1 (Crop)	Class = 0 (Non-crop)	Class Ratio	Precision	Recall	F1
Crop Mask Label Threshold: 16	1,387,488	1,365,024	0.504	98.2%	61.6%	75.7%
Crop Mask Label Threshold: 18	1,771,703	980,809	0.644	97.8%	78.3%	87.0%
Crop Mask Label Threshold: 20	1,991,000	761,512	0.723	97.2%	87.5%	92.1%
Crop Mask Label Threshold: 22	2,115,048	637,464	0.768	96.2%	92.0%	94.0%
Crop Mask Label Threshold: 24	2,191,547	560,965	0.796	95.2%	94.3%	94.8%
Crop Mask Label Threshold: 26	2,243,303	509,209	0.815	94.3%	95.6%	94.9%
Test Labels	**2,211,884**	**540,628**	**0.804**			

Table 5. Evaluation of noisy crop mask label sets compared to hand-validation test labels

Model Experiments

Experiment 1: Training a crop mapping CNN with imperfect Crop Mask Labels

The first experiment tested whether a model trained on a single Crop Mask Label threshold can outperform the training data. For this, we trained three models on each of the six Crop Mask Labels training sets to ensure that at least one of the models at each threshold was viable.

Experiment 2: Using an edge magnitude input layer to improve accuracy

Marmanis et al. (2016) demonstrated that adding an edge detection model prior to image segmentation with a CNN is a way to include class boundaries as part of the input, resulting in improved accuracy and more precise object boundaries in the output. For the second experiment, we repeated the structure of the first experiment, but added the edge magnitude layer used in the crop mask preprocessing as an additional input layer.

Experiment 3: Training on multiple Crop Mask Label thresholds

For the third experiment, we trained a single model iteratively on multiple Crop Mask Label thresholds. By increasing the number of training samples with varying levels of noise, this experiment tests whether the model can learn the crop features better with more samples of varying accuracy.

Experiment 4: Ensemble prediction from the best models

For the final experiment, we tested whether using an ensemble of models trained on different Crop Mask Label thresholds would improve the accuracy by combining their predictions. We took the mean prediction value from the three best models (based on F1 score) from experiments 1 and 2 and computed the accuracy of the combined prediction. Ensemble models have been found to outperform single models in many cases and we tested that hypothesis here (Marmanis et al. 2016).

Chapter 5 – Results

The outputs from the Thresholding and Clustering configurations are compared using validation data at the pixel, county, and field scales. The Deep Learning configuration experiments are evaluated separately because they did not produce statewide outputs and only utilize the hand-digitized validation data for assessment.

Thresholding & Clustering Configuration Evaluation

Overall, field outputs from the Clustering Configuration were more accurate than all three map versions from the Thresholding approach at the pixel, county, and field scales. Despite testing three sets of parameters for the Thresholding configuration, all the outputs were less effective at differentiating cropland from non-cropland, more susceptible to variation in the imagery, and less consistent over the geographic extent of Minnesota. The outputs from three sets of threshold values, referred to as Low, Medium, and High (see Table 6), reveal a trade off between precision and recall in the Thresholding configuration. The output from the Clustering configuration, the Clustering Map, captured more cropland statewide with fewer misclassification errors than the Thresholding configuration and approached the accuracy of the CDL Crop Mask at the field scale.

Map Version	Experimental Parameters
Low Threshold Map	NDVI Range 0.3
Medium Threshold Map	NDVI Range 0.5
High Threshold Map	NDVI Range 0.7
Clustering Map	N/A

Table 6. Statewide field map versions

The statewide field maps in Figure 29 show the outputs from the Thresholding and Clustering workflow configurations, with the CDL Crop Mask as a reference. Visually, the Low and Medium Threshold Maps display cropland predictions across the northeastern part of the state, where there is little cropland according to the CDL Crop Mask. The Medium and High Threshold Maps also show indications of artifacts from the path of the satellite swath over the state, resulting in diagonal marks across the western edge and southeastern corner of the state. From the statewide view, the High Threshold Map and the Clustering Map appear to predict cropland in areas that align with the CDL Crop Mask, which is verified by the pixel-scale accuracy analysis in Table 7.

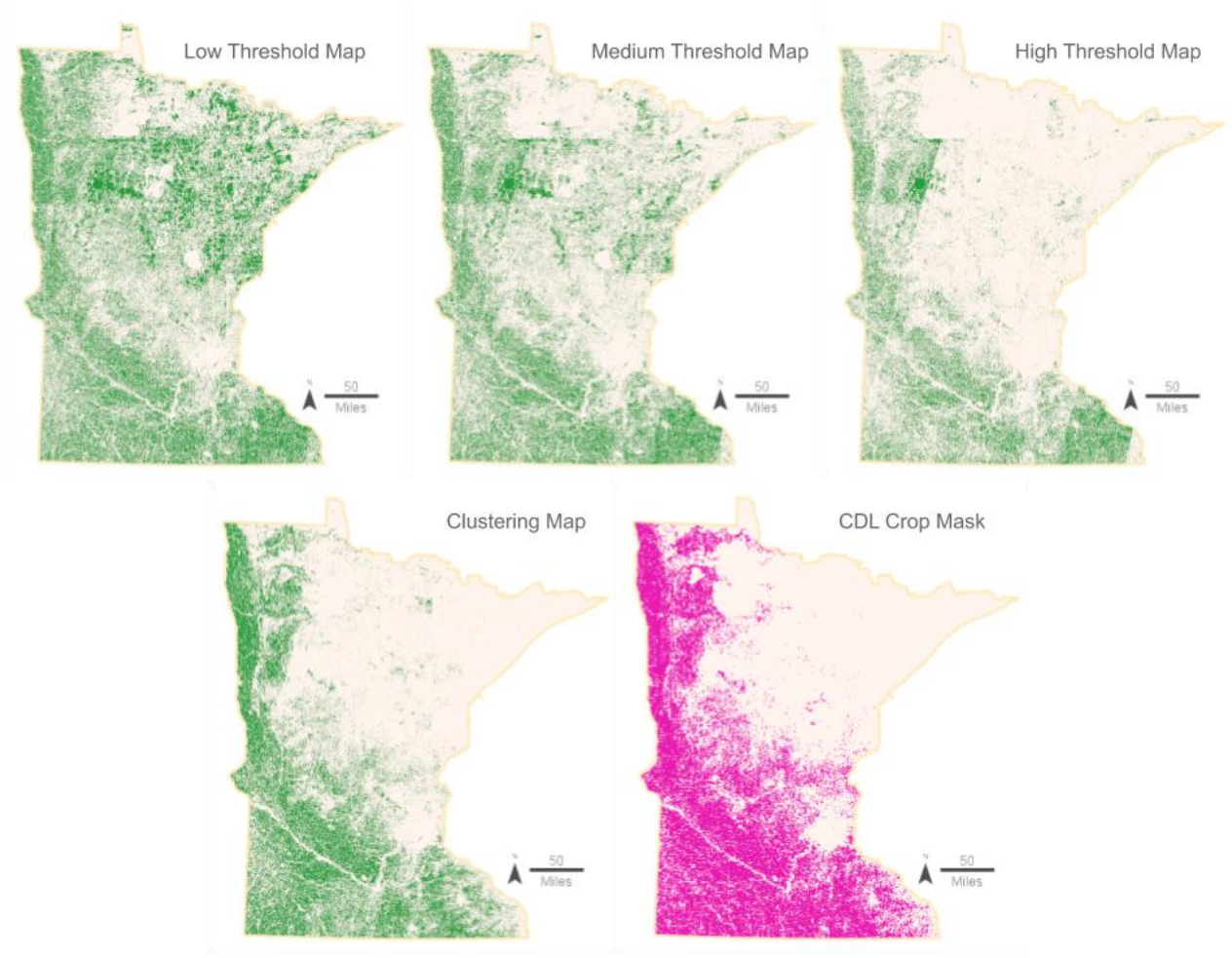

Figure 29. Statewide field map outputs from the Thresholding and Clustering configurations, with CDL Crop Mask layer reference

Pixel-Scale Accuracy: CDL Crop Mask Layer

Pixel-Scale Accuracy: CDL Crop Mask Layer

Statewide Map Version	Statewide Precision	Statewide Recall	Statewide F1 Score	Statewide Overall Acc.
Low Threshold Map	53.0%	65.6%	58.6%	65.6%
Medium Threshold Map	65.0%	60.0%	62.4%	73.1%
High Threshold Map	80.5%	47.0%	59.3%	76.1%
Clustering Map	91.6%	74.7%	82.3%	88.1%

Table 7. Pixel-Scale Accuracy

We computed pixel-scale accuracy at the statewide extent by intersecting the CDL Crop Mask layer with the statewide outputs. The Clustering configuration produced a more spatially accurate statewide crop map than the Thresholding configuration (Table 7). Compared to the CDL Crop Mask, the clustering map had the highest OA of any of the maps at 88.1%. It also had the highest precision (91.6%), recall (74.7%), and F1 score (82.3%), indicating that the Clustering configuration identified more cropland area statewide and did so with fewer misclassification errors.

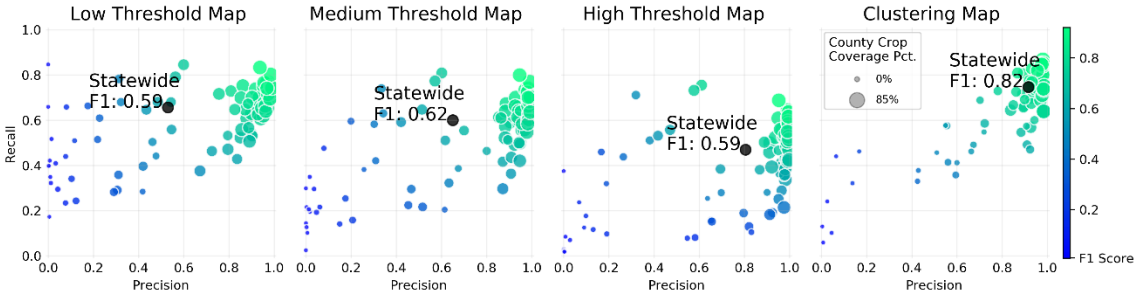

Figure 30. Pixel-scale accuracy with CDL Crop Mask layer by county

Computing pixel-scale accuracy for each county reveals that the Clustering configuration is more consistent in cropland prediction across counties with varying amounts of agricultural activity. Figure 30 shows a pixel-scale accuracy plot of precision vs. recall for each map by county and overall. The size of the points represents the percentage of land in each county that is cropland and the color represents the F1 score. In the Low Threshold Map, the overall precision was the lowest of any of the maps but for ag-dominated counties, the larger dots, the precision was relatively high. For the High Threshold Map, with much more restrictive threshold values, the precision reached 80.5% with many agricultural counties at or near 100%, but the recall dropped to 47%, indicating that while the model was very discriminative it missed many cropland areas. Despite these differences, the F1 scores for the Low Threshold and High Threshold Maps were nearly identical (58.6% and 59.3%, respectively). The Medium Threshold Map, with a threshold value in the middle, improves upon the precision of lower threshold map and the recall of higher threshold map, increasing the F1 score to 62.4%.

There is a trade-off between precision—accurately classifying cropland—and recall—capturing more cropland area—in the Thresholding approach. Comparing the Thresholding maps, we see that at the low threshold value the recall is higher, capturing more cropland statewide, but it comes at the expense of precision. As the threshold value increases in the Medium Threshold and even further in High Threshold Map the points shift down and to the right, where precision for many counties is nearing 100% but recall is lower, missing more than half of the cropland

area in the case of the High Threshold Map. We can see this trade-off between precision and recall in the Thresholdind configuration clearly in Figure 31. All three outputs had similar F1 scores, ranging from 58.6% to 62.4%, while precision and recall varied inversely to each other. With the Low Threshold Map, more area was classified as cropland, improving recall (65.6%), but at the expense of precision (53.0%). The more conservative threshold of the High Thresholdd Map reduced the amount of these false positives, resulting in a precision score of 80.5%. However, this was at the expense of recall (47.0%), which indicates that less than half of the total cropland in the state was classified with this set of parameters. The Medium Threshold Map resulted in precision (65.0%) and recall (60.0%) in between the other two sets of parameters, indicating that this approach has a trade-off between precision and recall, limiting the potential overall accuracy. The three statewide points the Thresholding maps indicate the trade-off boundary, which is surpassed in both precision and recall by the Clustering Map.

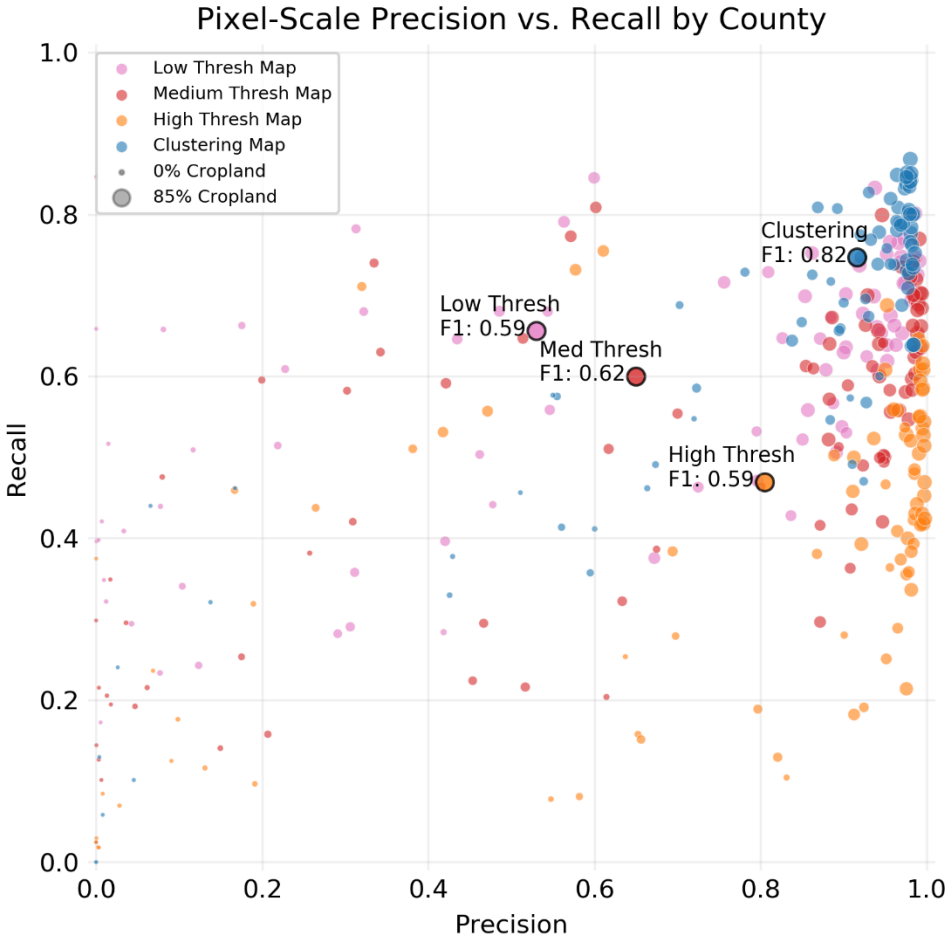

Figure 31. Pixel-Scale precision vs. recall by county, comparison of Thresholding and Clustering Map outputs

The Clustering Map, on the other hand, has very high precision (91.6%), indicating that this approach effectively differentiates between crop and non-crop pixels, and has high recall (74.7%), capturing most of the cropland in the state. Figure 31 highlights the differences between the two approaches, showing the pixel-scale accuracy of each of the maps by county and overall. In all the maps, accuracy was low for the counties where there is not very much cropland, represented by the small points on the plots. In these counties, a small error in total predicted cropland results in large relative error and low accuracy. Analyzing these areas at the county-scale helps illuminate the performance of each configuration across the state.

County-Scale Accuracy: USDA Census of Agriculture

County-Scale Accuracy: USDA Agricultural Census 2017

Statewide Map Version	Total Predicted Acres	State Cropland Area Pct. Error	State Harvested Area Pct. Error	County Mean Absolute Error (Harvested Area)
Low Threshold Map	24,843,930	14.0%	23.9%	154,331
Medium Threshold Map	18,540,623	14.9%	7.55%	124,712
High Threshold Map	11,693,357	46.3%	41.7%	122,678
Clustering Map	16,354,601	24.9%	18.4%	49,471
CDL Crop Mask	20,037,084	8.0%	0.1%	19,114
Ag. Census Harvested Acres	20,054,132	8.0%	0.0%	-
Ag. Census Cropland Acres	21,786,756	0.0%	8.6%	-

Table 8. County-Scale Accuracy

Using the Census of Agriculture as validation data to evaluate county-scale accuracy, the Clustering configuration also outperformed the Thresholding configuration at the county extent, despite having a greater statewide percent error than the Medium Threshold Map. Table 8 shows the predicted acres of cropland from each map and the percent error compared with both the Cropland Acres and Harvested Acres reference points. The CDL Crop Mask layer is included to gauge the accuracy between spatial pixel-scale ground truth and census county-scale validation statistics. We see that the DCL Crop Mask has only a 0.1% error compared to the census Harvested Acres at the statewide level. Among the map outputs, Medium Threshold Map is the closest in aggregate to both the Cropland Area and the Harvested Area. Notably, the predicted area varies from above the reference to well below the reference data across the three Thresholding map versions, indicating that accuracy is highly dependent upon input the

parameters. While the Medium Threshold Map has a lower percent error than the Clustering Map compared to the statewide totals, this only measures accuracy in the aggregate scale.

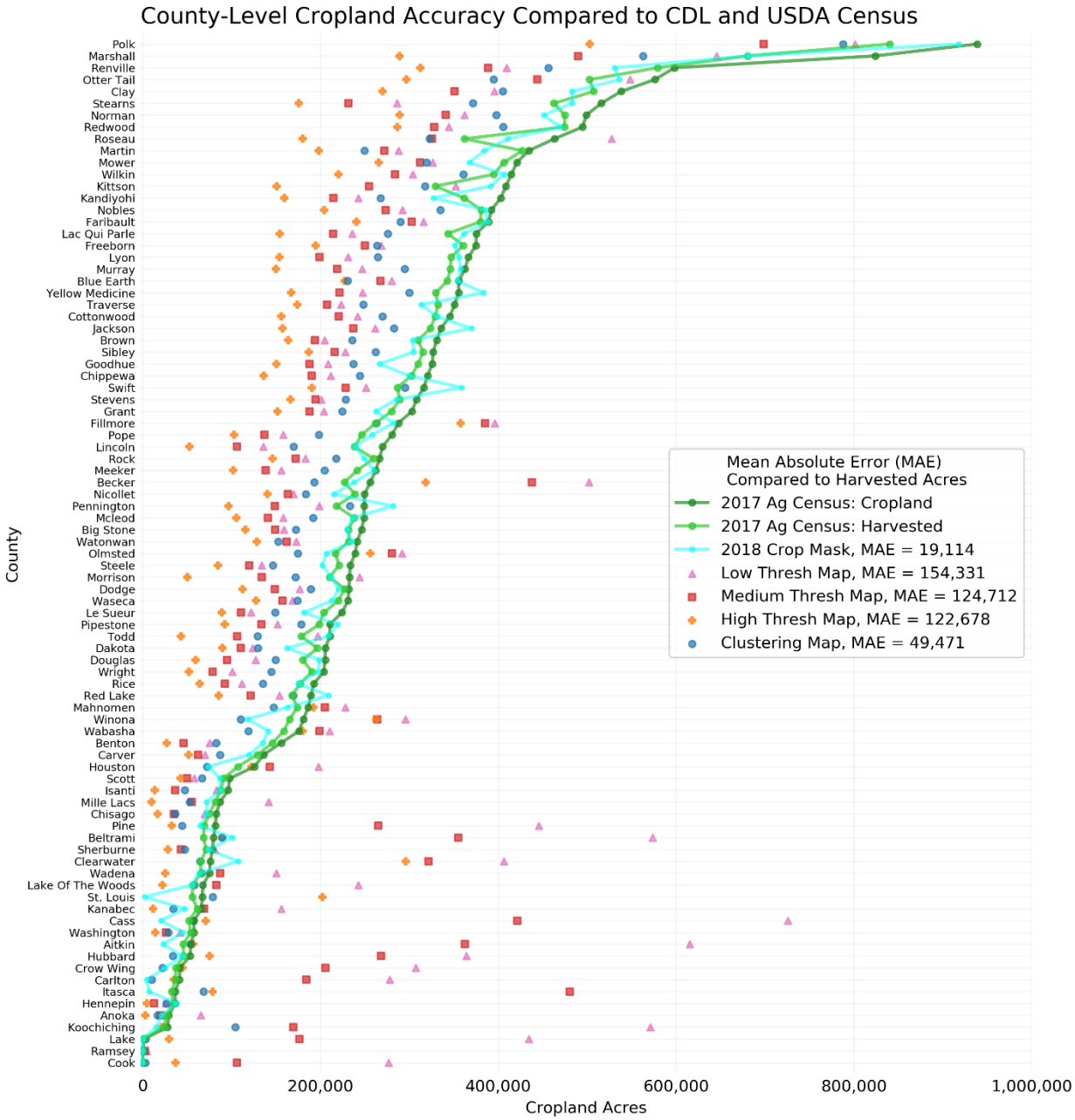

Figure 32. County-Scale Cropland Accuracy

A county-scale breakdown, however, indicates that the Clustering Map is much more consistently accurate across the state. Figure 32 shows the cropland area in acres for each of the three reference data sets (the lines) and the estimated values for each map version (the points) sorted by total Cropland Area from the census. Despite having a lower percent error at the statewide extent, Medium Threshold Map is inconsistent county to county. It greatly

overestimates many counties with little cropland while underestimating counties with large amounts of cropland, resulting in a total predicted cropland area similar to the statewide census data but not spatially accurate. All of the three Thresholding maps tend to vary more widely from county to county compared to the Clustering Map and they over-estimate cropland in many of the counties where there is little cropland, resulting in MAE values of over 120,000 acres for all three maps. Counties like Cook and Lake in far northern Minnesota are dominated by forested areas and the Thresholding maps overestimate cropland. Urban counties like Ramsay and Hennepin, on the other hand, have close agreement between all the outputs. This demonstrates that much of the misclassification comes from natural vegetation. The Clustering Map is much more consistent across county geographies, underestimating cropland area in nearly every county, but it is more accurate overall, with a MAE of less than 50,000 acres. The largest errors come primarily from underestimating cropland in ag-dominated counties, suggesting that this approach is more effective at differentiating cropland from natural vegetation. We also evaluated the CDL Crop Mask against the census data at the county level, where it shows some variation county to county but tracks closely with the Harvested Area value, resulting in a MAE of just under 20,000 acres.

Field-Scale Accuracy: Hand-Digitized Validation Fields

Field-Scale Accuracy - Validation Fields

Statewide Map Version	Field Precision	Field Recall	Field F1	Overall Acc.
Low Threshold Map	98.9%	83.6%	90.6%	85.1%
Medium Threshold Map	99.1%	80.4%	88.8%	82.5%
High Threshold Map	99.8%	71.2%	83.1%	75.1%
Clustering Map	98.4%	93.0%	95.6%	92.7%
CDL Crop Mask Layer	94.6%	97.7%	96.1%	93.2%

Table 9. Field-Scale Spatial Accuracy

The field-scale accuracy assessment indicates both how well the different approaches classify cropland at the local scale and how effective the segmentation techniques are at delineating individual fields compared to the validation field polygons. The spatial accuracy at the field scale is higher than at the statewide scale for all map versions, notably the precision scores, indicating that both the Thresholding and Clustering configurations perform better in areas that are predominantly cropland, see Table 9. The Clustering Map, however, still outperforms the Thresholding maps in OA, at 92.7%, and F1 score, at 95.6%, when compared to

the hand digitized validation fields. Field-scale accuracy for the Thresholding configuration showed a similar trade-off between precision and recall as the pixel-scale accuracy, but in this case the Low Threshold Map had higher OA instead of The High Threshold Map. This suggests that a lower threshold better identifies in agriculturally dominated areas while a higher threshold is better suited for large area differentiation of cropland and non-cropland. We also evaluated the spatial accuracy of the CDL Crop Mask at the field scale, where it had slightly lower precision but higher recall, F1, and OA than the other maps, demonstrating that it effectively captures cropland at the local level but also includes non-crop areas when compared to fields derived from very high resolution imagery. Figure 33 shows the that while all the maps have very high precision at the field-scale, the Clustering Map captures more of the field area and, notably, has nearly the same F1 score as the CDL Crop Mask Layer.

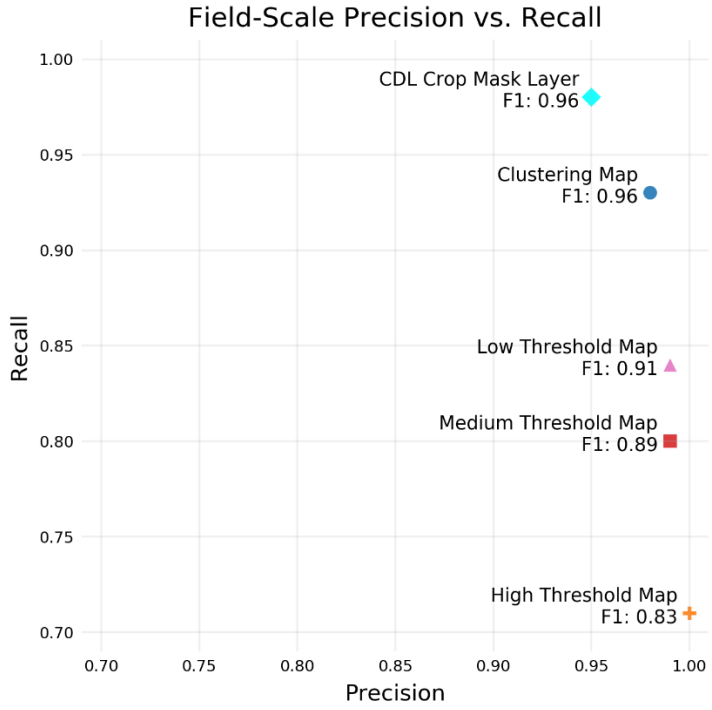

Figure 33. Field-Scale Accuracy Precision vs. Recall

To evaluate the accuracy of the segmentation—individual field delineation—we compared the field polygons from each output to the validation fields (Table 10). There were 440 total validation fields covering a total of 50,996 acres, resulting in an average field size of 116 acres. The Thresholding maps were closer in the total number of fields than the Clustering Map but were not as accurate in total field area. Using the Variation of Information (VI) metric from Meilă (2003) for evaluating the segmentation, we found that the Thresholding maps tend to be under-segmented in many areas, merging multiple fields together into a single object. The

Clustering Map shows slightly more over-segmentation, where individual fields are split into multiple field objects. Visual inspection shows that much of the over-segmentation in the Clustering Map is at the boundaries of fields, where instead of a single field edge there are multiple polygons. This is likely due to the gaussian smoothing done in the preprocessing of the image used for segmentation, where the noise within fields is reduced but strong edges are not maintained. Incorporating edges into the crop mask in the Thresholding configuration improves the integrity of field boundaries by masking them out instead of relying on the segmentation function to delineate them. Overall, the VI is lower for the Clustering Map, but visual inspection of the field boundaries in Figure 34 shows that segmentation errors in both configurations result in field boundary inaccuracies compared to the validation fields.

	Field Segmentation Statistics			***Variation of Information (VI)***		
Statewide Map Version	**Total Fields**	**Total Field Area**	**Average Field Area**	**False Splits (Over-segment.)**	**False Merges (Under-segment.)**	**Total VI**
Low Threshold Map	495	43,137	87.1	0.66	2.36	3.02
Medium Threshold Map	492	41,384	84.1	0.67	2.58	3.25
High Threshold Map	475	36,394	76.6	0.67	3.17	3.84
Clustering Map	698	48,185	69.0	0.91	0.98	1.89
Validation Fields	*440*	*50,996*	*115.9*	*0.00*	*0.00*	*0.00*

Table 10. Field-Scale Segmentation Accuracy

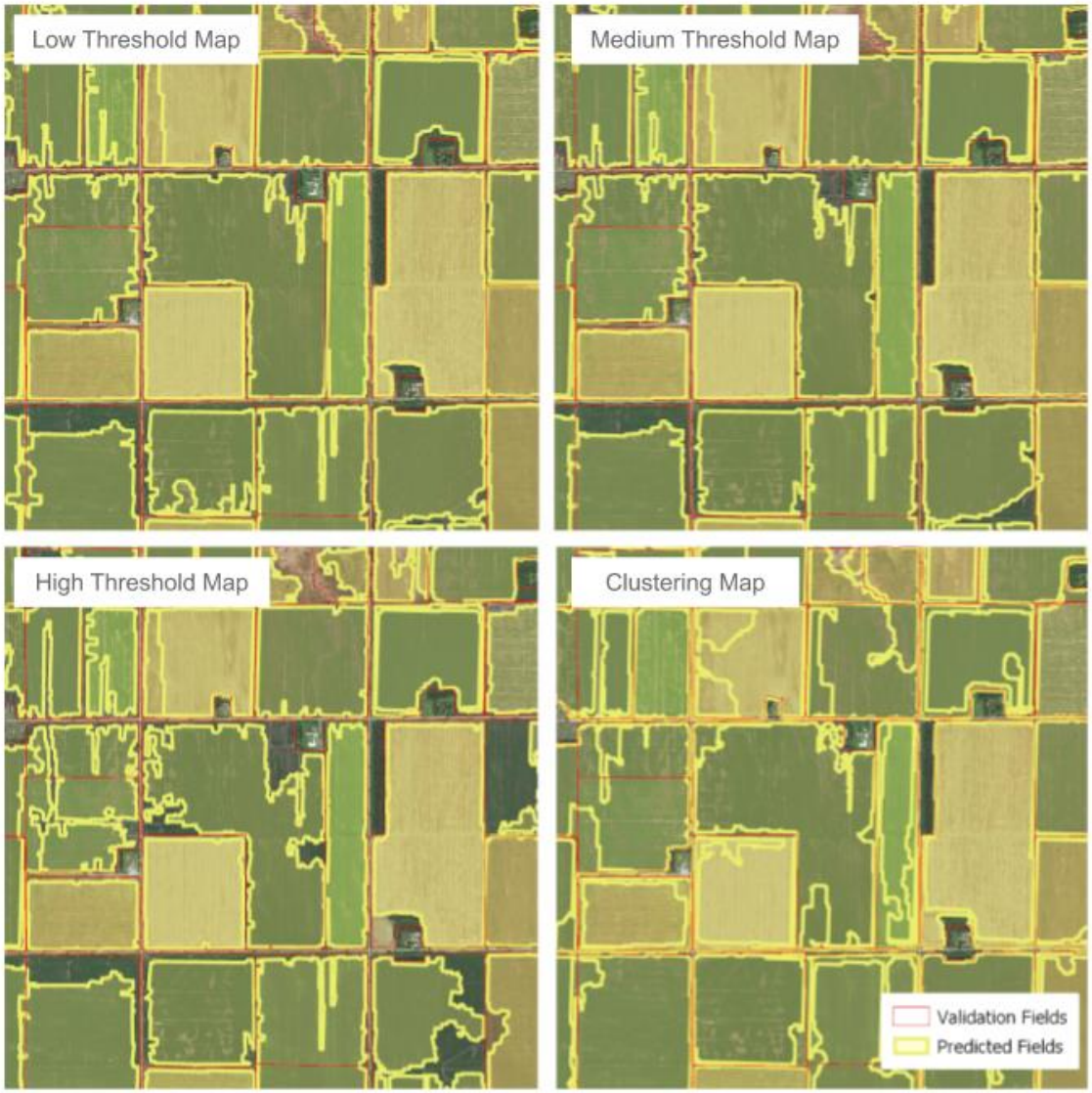

Figure 34. Field-Scale segmentation maps for detail of validation area

Overall, the Clustering approach is more effective at mapping cropland over the large, diverse landscape of Minnesota and delineating field boundaries than the Thresholding approach.

Deep Learning Configuration Evaluation

The Deep Learning configuration was only evaluated at the pixel-scale, assessing the cropland area predicted by the model. Since this configuration was exploratory, the outputs were not segmented into field polygons and therefore we did not evaluate field-scale metrics.

The cropland predictions produced by the CNN deep learning models matched and, in some cases, exceeded the accuracy of the Crop Mask Labels used as training data, indicating that a CNN can learn from imperfect labels to improve predictive accuracy. The most accurate model

in each of the trials exceeded the most accurate baseline Crop Mask Label, based on F1 scores. Models trained on Crop Mask Labels with high precision but relatively low recall—in other words, masks that miss a lot of field pixels but are almost always correct when they positively classify cropland—showed the largest increase in accuracy. These models retained similar precision while improving recall and identifying a larger share of the cropland pixels. Using edges as an input layer produced the highest accuracy of any of the models but the difference between the best models in each trial was not very large.

For Experiment 1, each CNN model was trained on a single Crop Mask Label with varying threshold values ranging from 16, the most restrictive, to 24, the most optimistic. We computed the precision, recall, and F1 scores for the best performing model trained at each Crop Mask Label threshold to compare with the baseline accuracy of the corresponding labels used for training (Table 11). The F1 score improved for the model at each Crop Mask Label threshold compared to the baseline except for Threshold 26 where it remained even. Models trained on the two noisiest, least accurate Crop Mask Labels thresholds improved the most compared to the baseline, with each model reaching 92.9% F1 or better. The three highest overall accuracy scores were from the models trained on the thresholds 22, 24, and 20. For the threshold 22 model, both precision and recall improved from the baseline.

Threshold	Model	Precision	Recall	F1
16	Crop Mask Label	**98.2%**	61.6%	75.7%
	U-Net Model	98.1%	**88.2%**	**92.9%**
18	Crop Mask Label	**97.8%**	78.3%	87.0%
	U-Net Model	95.7%	**93.8%**	**94.7%**
20	Crop Mask Label	**97.2%**	87.5%	92.1%
	U-Net Model	97.1%	**93.2%**	**95.1%**
22	Crop Mask Label	96.2%	92.0%	94.0%
	U-Net Model	**96.4%**	**94.5%**	**95.4%**
24	Crop Mask Label	**95.2%**	94.3%	94.8%
	U-Net Model	94.3%	**96.1%**	**95.2%**
26	Crop Mask Label	**94.3%**	95.6%	**94.9%**
	U-Net Model	93.6%	**96.1%**	94.8%

Table 11. Deep Learning model accuracy compared to baseline training label accuracy

Plotting the precision vs. recall (Figure 35) for the baseline Crop Mask Labels and the model outputs reveals a trade-off boundary for accuracy with the Crop Mask Labels thresholds.

As the threshold gets higher, more pixels are estimated to be in the crop class and so precision is lowered (more non-crop pixels are misclassified as crops), but recall improves (as more of the crop pixels are accurately identified). Notably, several of the models break beyond this trade-off boundary to improve both precision and recall compared to the baseline Crop Mask Labels. For example, each of the models trained on the Threshold 16 mask retained about the same level of precision but jumped dramatically in recall, meaning the models were able to learn enough characteristics of the crop features from the most conservative Crop Mask Labels to identify a lot more crop pixels without sacrificing precision. Several other models trained on Thresholds 20 and 22 also beat the baseline in both precision and recall. Models trained on the higher Thresholds (24 and 26) tended to have lower precision than their baseline training sets and follow the pareto boundary. The dots clustered in the top left are models whose training loss plateaued and ended up classifying every pixel as a crop pixel.

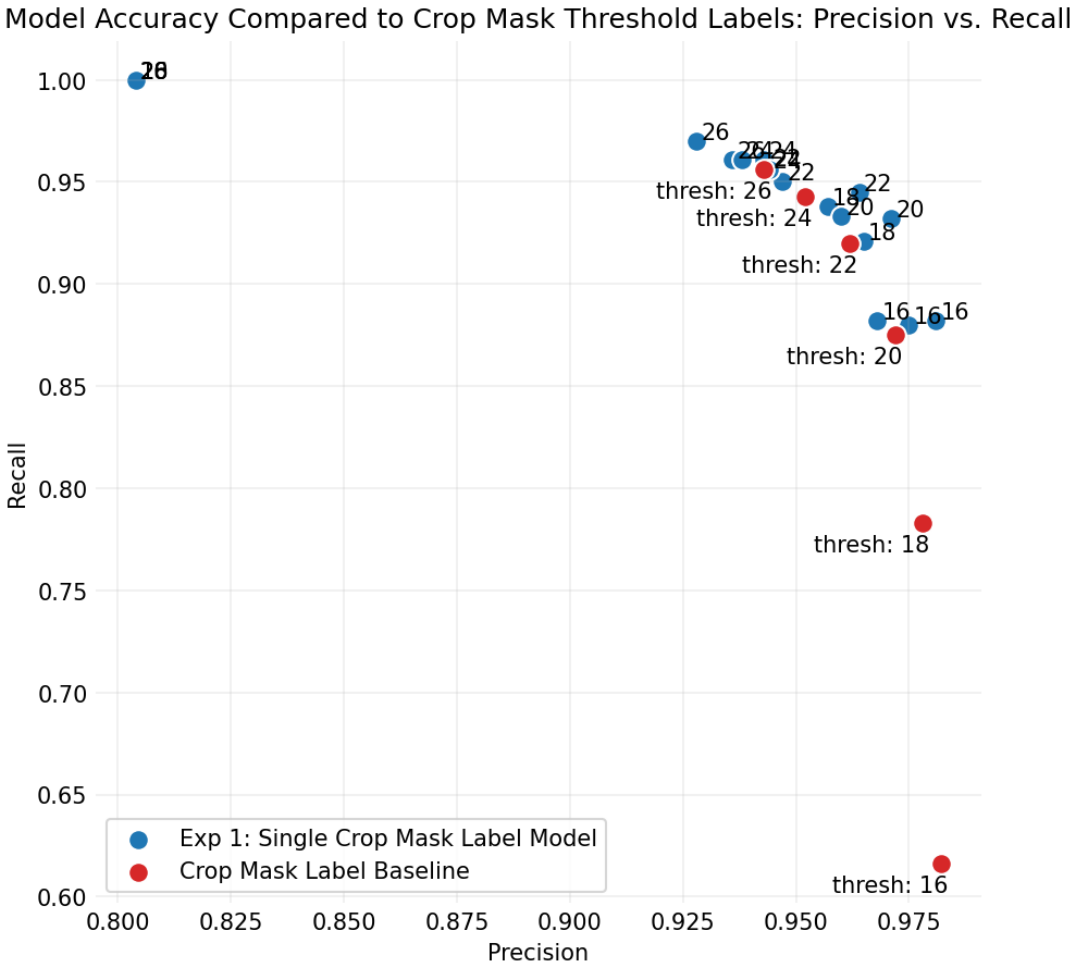

Figure 35. Deep Learning Experiment 1. precision vs. recall

Visualizing the outputs from the individual Crop Mask Label threshold models reveals some insight into what each model is learning. In Figure 7, the leftmost images show an RGB composite from the July imagery date and the ground truth hand-digitized label. The top row shows the Crop Mask Labels at each threshold and the bottom row shows the output from the corresponding model. This example imagery patch has crop fields in most of the image and a waterway with natural vegetation along the right side. We can see that the Threshold 16 mask misses much of the crop area and has very noisy edges for the other parts, but the model has still learned enough to identify most of the fields and field boundaries. The green color of the predicted field areas reflects the relative uncertainty of the model. The two models on the far right, Threshold 24 and 26, overestimate the crop area and classify much of the natural vegetation along the stream as crops. The two in the middle, Threshold 20 and 22, appear to perform the best, identifying fields and boundaries without misclassifying the natural vegetation.

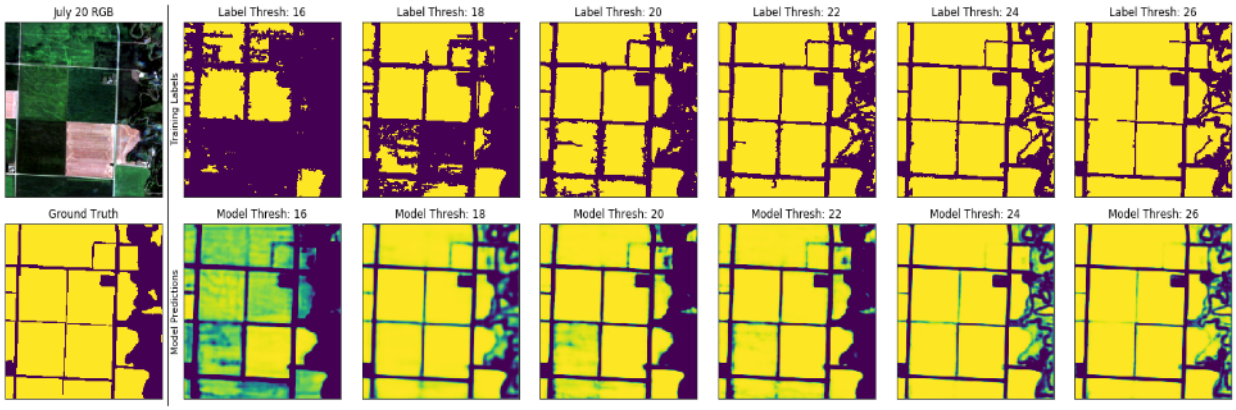

Figure 36. Experiment 1. Crop Mask Label thresholds (top row) and the corresponding CNN model results (bottom row)

The outputs from each of the other experiments tended to fall along the precision/recall trade-off boundary but the model from Experiment 2, which included an edge magnitude layer in the inputs, produced the highest overall accuracy. Experiment 3, which tested models trained on multiple Crop Mask Label thresholds, showed the highest recall, but the lowest precision, suggesting that it misidentified natural vegetation and was not as effective at predicting crop pixels. Experiment 4 combined the predictions from three individual models as an average value, with less accurate results than the individual models of the ensemble. The best model from each of the experiments exceeded the F1 score of the most accurate baseline Crop Mask Label threshold, suggesting that all the approaches are effective enough to at least match the input training layer. Table 12 shows the accuracy of the best model from each experiment along with a

confusion matrix. The models trained with and without edges had the highest overall accuracy and outperformed the baseline in both precision and recall.

Model	Experiment	Precision	Recall	F1	True Positive	False Positive	True Negative	False Negative
U-Net Threshold: 22	Exp 1: Single Crop Mask Label	**96.4%**	94.5%	95.4%	2,090,488	77,995	462,633	121,396
U-Net w Edges Threshold: 24	Exp 2: Single Crop Mask Label with Edges	95.7%	96.5%	**96.1%**	2,133,602	95,650	444,978	78,282
U-Net Multiple Threshold: All	Exp 3: Multiple Crop Mask Label Thresholds	93.0%	**97.0%**	95.0%	2,145,435	161,637	378,991	66,449
Ensemble Edges Prediction Threshold: 20, 22, 24	Exp 4: Ensemble Mean Prediction	94.8%	95.8%	95.3%	2,119,603	117,279	423,349	92,281
Crop Mask Label Threshold: 26	Crop Mask Label Baseline	95.7%	96.5%	**96.1%**	2,114,899	128,404	412,224	96,985

Table 12. Accuracy for the best model from each experiment and the most accurate Crop Mask Label threshold

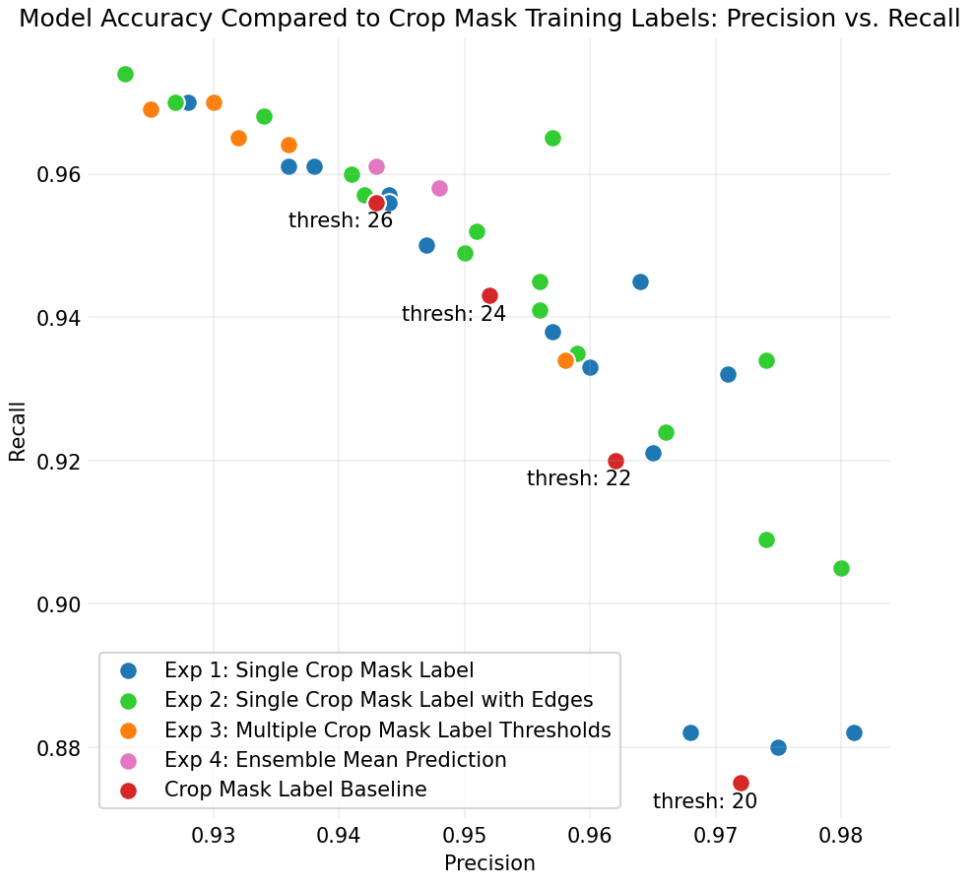

Figure 37. Precision vs. Recall for all model experiments and baseline Crop Mask Label Thresholds

Plotting the precision vs recall of the models from each of the experiments in Figure 37 reveals that most of them follow along the same accuracy boundary of the baseline crop mask label thresholds. However, many of the models from Experiments 1 and 2, trained with single Crop Mask Labels without and with edges, respectively, outperformed the baseline on both axes. These models indicate the possibility of using deep learning approaches to overcome some of the trade-offs between precision and recall from a more traditional thresholding approach, but additional work is needed to test this over larger areas.

Chapter 6 – Discussion

Answering the Research Questions

Based on the results of the three workflow configuration experiments we conducted, we were able to answer the core research questions that motivated this work:

1. What information can be derived from raw high resolution, multispectral, multitemporal remote sensing imagery that is effective for mapping fields in the state of Minnesota?
2. What generalizable, unsupervised analytical methods can identify cropland and delineate individual fields accurately without relying on ground sampled training data?
3. How can these processes be combined into a geographically scalable, computationally efficient workflow that has the capacity to be deployed in agricultural regions around the world?

Through exploring these questions, we identified potential solutions for each of the data, methodology, and computational challenges facing global-scale field mapping. We answered the first research question by identifying characteristics of high-resolution remote sensing imagery that are useful for cropland mapping and field delineation, specifically monthly NDVI statistics, seasonal NDVI standard deviation, edge magnitude, and mean NDWI values. To answer the second research question, we explored the potential and encountered the limitations of several approaches to conducting field mapping without reliance on training data. The Clustering configuration demonstrated that accurate field mapping can be achieved over large areas using unsupervised k-means clustering and a general knowledge of local agricultural practices, while the Deep Learning configuration yielded promising results for extending traditional methods with automated generation of training data. The three different workflow configurations illustrated the potential for building a flexible, efficient workflow for deploying field mapping approaches, providing multiple examples of how to achieve the third research question. This framework can be extended to build upon the findings of our work.

Sentinel-2 Temporal Vegetative Features to Address the Data Challenge

In addressing the first research question, we found that temporal vegetative characteristics, as demonstrated in the cropland mapping and field delineation research, were critical to the success of the different workflow configurations. The temporal variation seen in the monthly median NDVI values used as k-means inputs in the Clustering configuration were able to extract distinct temporal patterns associated with land cover classes. The 5-day temporal resolution of the Sentinel-2 imagery enabled robust monthly statistics to be generated for most

parts of the state. The spatial resolution of the imagery also proved to be adequate for delineating Minnesota crop fields, suggesting that it is suitable for high resolution field mapping in many places globally. Sentinel-2 imagery is sufficient for solving the data challenge by enabling the use of robust temporal features at high spatial resolution necessary for field mapping.

While Sentinel-2 imagery was suitable for crop mapping, the difference in spatial and temporal resolution with the CDL Crop Mask Layer is indicative of the challenges of large area field mapping. Since high resolution field maps do not exist for most places and are usually tied to a specific time period, imperfect data will be necessary to evaluate the accuracy of field mapping, which will introduce an element of uncertainty in the accuracy measurements (Pérez-Hoyos et al. 2017). Unlike the 10-m Sentinel-2 imagery used for our analysis, the CDL Crop Mask Layer is derived from coarser 30-m Landsat imagery. Additionally, the CDL Crop Mask Layer is a composite of five years of CDL data, meaning that is a pixel was classified as a crop in two of the previous five years, then it is classified as crop in the mask layer. Thus, if areas are typically used as crop fields but were left fallow during our study period, then they are likely to be flagged as false negatives in the accuracy assessment. This likely contributed to the relatively low recall value of the Clustering Map (74.7% across the state), but we are unable to evaluate by how much. Expanding temporal analysis of the Sentinel-2 imagery to align more closely with the time period of the CDL data could be effective in improving the identification of edges in the imagery. Given that the CDL Crop Mask was within 0.1% error of the statewide Harvested Area from the Census of Agriculture, this temporal discrepancy likely does not have a large impact on the accuracy and is an effective validation dataset for evaluating statewide pixel-level accuracy.

Data-Driven Techniques to Address the Methodology Challenge

Testing three different configurations of the analytical workflow revealed that manual thresholding is not suitable for scalable analysis over large areas but data-driven techniques like k-means clustering and deep learning with imperfect training data have potential to be applied as generalizable methods for large area crop mapping. This addresses the second research question. Simple thresholding techniques can be very accurate when fine-tuned to local conditions but are less suitable for generalizable mapping tasks, while the data-driven and local knowledge-informed Clustering approach demonstrated high accuracy over large areas. Taking advantage of the strengths of emerging deep learning models by creating a hybrid approach can further extend crop mapping beyond the limitations of more traditional methods.

Thresholding Workflow Configuration

In the Thresholding workflow configuration, we found that using temporal NDVI statistics and edge magnitude can effectively identify cropland in certain regions but is not accurate over large areas. Tuning the parameters to dense agricultural regions produced accurate cropland classification in many counties that are primarily agricultural but leads to misclassification of natural vegetation in other areas. With a low set of thresholds on temporal NDVI change and monthly maximum, many forested areas in northeastern Minnesota were classified as cropland. With a higher set of NDVI thresholds, the model had higher precision, but only identified 47% of the state's cropland. At the field scale, however, the results showed higher accuracy. The edge magnitude analysis used in this configuration could be a beneficial step to improve the delineation of inter-field boundaries in the Clustering configuration.

The Medium Threshold map, for instance, achieved 82.5% OA and a 88.8% F1 score at the field-scale covering a small area (Table 9), but only 73.1% OA and 62.4% F1 at the statewide extent (Table 7). A single set of parameters likely will not be effective over large, diverse areas. Additional experimentation of Thresholding parameters could be undertaken to find the optimal values for each parameter to reduce the confusion between crop and non-crop areas and potentially improve the outputs, but the results suggest that there is a limit to overall accuracy for this approach due to the trade-off between precision and recall shown in Figure 31. The Crop Mask Labels used as training data in the Deep Learning configuration showed the same trade-off between precision and recall. These masks were generated with the same overall approach as the Thresholding configuration—user defined threshold values for NDVI change and edge magnitude. The accuracy for these Crop Mask Labels (F1 score topping out at 94.9%) were higher than for the outputs of the Thresholding configuration (F1 score of 90.6% at the field-scale), but they were finely tuned to the smaller, more ag-dominated study area of the Deep Learning configuration. This suggests that over small areas a Thresholding approach could produce highly accurate field maps, but it is not suitable over large areas where fine-tuning is not feasible. While the Thresholding workflow configuration did not rely on external training data, it failed to produce accurate cropland predictions over the state of Minnesota.

All the statewide outputs from this workflow had artifacts from the difference in imagery availability, most notably in areas where the satellite path was evident in the outputs. This could be due to the low cloud threshold that was used, leaving some parts of the state with too few observations to produce robust temporal statistics. Additional preprocessing steps are needed to improve the performance of this configuration to overcome the limited data problem. Implementing a cloud masking approach could produce more valid imagery pixels, but it would

also introduce noise into the edge mapping step. Based on the accuracy results, universal threshold is not feasible for large area crop mapping and this approach is not geographically scalable.

Clustering Workflow Configuration

The Clustering workflow configuration approached validation data in terms of precision, topping 91% statewide based on the pixel-scale accuracy with the CDL Crop Mask layer. Despite this high precision, it still tends to miss cropland areas, resulting in a statewide recall score of 74.7%. This approach is promising because it is primarily unsupervised and learns from the data itself. The selection of the number of clusters and the parameters to identify crop and non-crop clusters depend on testing and local knowledge, so the approach still requires some manual input to tune parameters. However, it does not require external training data, meaning that it can be deployed in different geographic areas with minimal user input.

Importantly, the spectral information and temporal resolution of Sentinel-2 data seem to be adequate for deriving the temporal vegetative information that is necessary to identify most of the cropland in the mixed landscape environment of Minnesota. This might not be the case in all environments, particularly where cropland and natural vegetation have similar temporal vegetative characteristics. Even in Minnesota, about one quarter of the cropland area was missed by the Clustering configuration. To address this, additional inputs can be used for clustering. We tested one set of preprocessing inputs to use in the k-means clustering, but different features can be derived from the raw imagery to use in the clustering for more discriminative results. Additional experiments should be undertaken in other agricultural landscapes where differentiating cropland from non-cropland requires different vegetative or temporal features to test the robustness of the Clustering configuration.

Additionally, this approach has potential to be extended to the field delineation step, too. In many cases, prior to merging into a binary mask, the cropland clusters formed discrete objects that visually corresponded with field boundaries, as shown in Figure 21. These clustered objects could be integrated as field polygons in an object-based approach, but noisy, over-segmented areas would need to be addressed. This approach could form the foundation of a field mapping workflow configuration. Refining the clustering method and building out additional steps within the workflow to address the underestimation of cropland and improve field delineation could produce a scalable, generalizable, and accurate analytics system for mapping crop fields.

Comparing Limitations of the Statewide Configurations

The Clustering workflow configuration was much more robust to the diversity of land cover across the state than the Thresholding configuration. It did not rely on externally derived training data but utilized the characteristics of the imagery data itself to differentiate cropland from non-cropland. Incorporating local knowledge about agricultural characteristics, such as the seasonal vegetative patterns used to define the crop cluster classification, will be necessary to make the Clustering configuration truly scalable. This type of human input is much less time consuming than creating training data around the world and allows for interdisciplinary collaboration to incorporate local agricultural expertise from different regions. We only tested a single version of the Clustering configuration and additional experiments are warranted to explore whether a similar trade-off between precision and recall exists when parameters like the value of *k* in the k-means clustering is changed. The Clustering configuration also misses about one quarter of the cropland in the state of Minnesota, achieving a statewide recall value of 74.7%, indicating that additional work is needed to achieve accuracy on par with the CDL Crop Mask. The difference in spatial and temporal resolution between the Sentinel-2 imagery used in the Clustering configuration and the CDL data can account for some of the differences, but the Clustering Map consistently underestimates cropland area in nearly every county. This indicates that some cropland is being clustered in with non-crop clusters during the k-means process or that the parameters used to classify the clusters as cropland are too restrictive. The consistency of underestimating predicted cropland area suggests that the method is robust but could be further refined. Notably, the Clustering Map had accuracy on par with the CDL Crop Mask at the field-scale, within 0.5% for both OA and F1 (Table 9). Additional experimentation on the input features, the number of clusters, and the crop cluster thresholds is warranted to explore ways to improve the recall of the Clustering configuration, but its overall accuracy indicates that it is a promising approach for unsupervised field mapping.

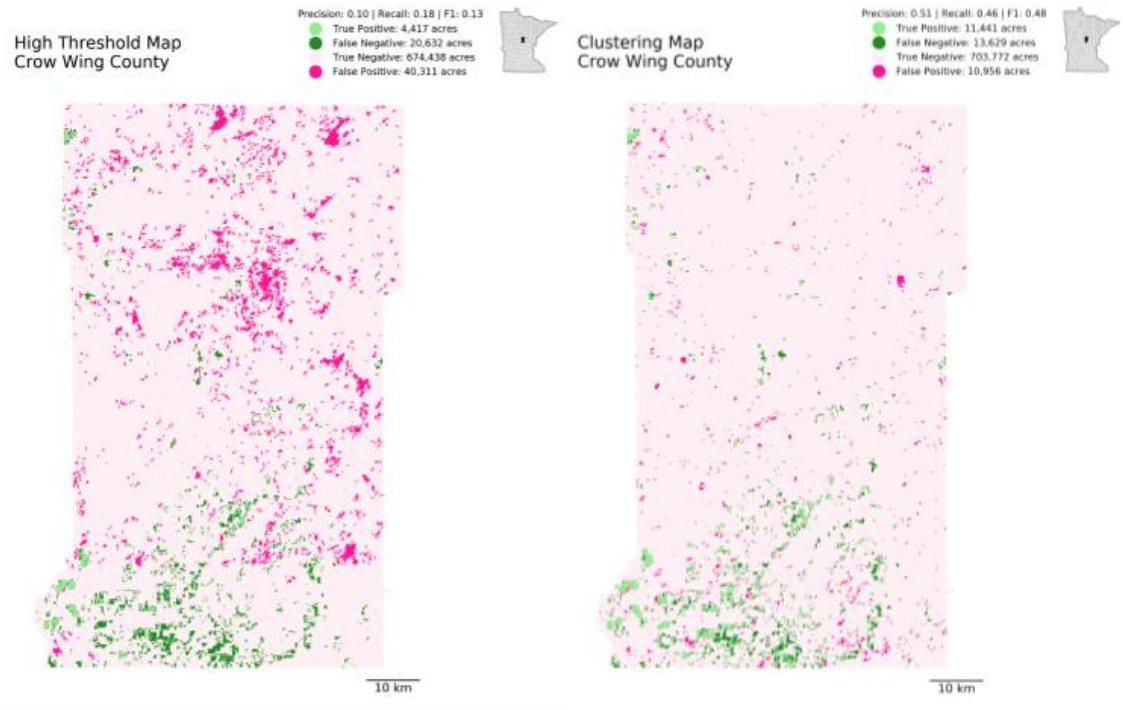

Figure 38. Crow Wing County pixel-scale accuracy comparison of natural vegetation

The Thresholding approach has difficulty differentiating natural vegetation from cropland, particularly evident in the northeast portion of the state where there is little cropland. The Clustering approach also misclassifies some natural vegetation areas, but at a much lower rate, resulting in a precision value greater than even the most conservative parameters of the High Threshold Map. Figure 38 shows the pixel-level accuracy of High Threshold Map and the Clustering Map in Crow Wing County, in the northeast part of the state. The county is primarily natural vegetation, with small amounts of cropland in the southern portion. Even though the accuracy in the Clustering Map is only near 50% (51% precision, 46% recall, 48% F1), this approach is much more accurate than the most conservative High Threshold Map (10% precision, 18% recall, 13% F1). The use of all the monthly median NDVI values in the Clustering configuration allows for more of the temporal vegetative information to be incorporated into the cropland mapping than in the Thresholding configuration, where only the monthly maximum and range values were used. As indicated in the literature, incorporating more temporal features into the analysis can improve the differentiation of cropland from natural vegetation (Kussul et al. 2015). Clustering, as a data-driven unsupervised technique, is a more effective than user-defined Thresholding as a crop mapping approach for large, diverse landscapes.

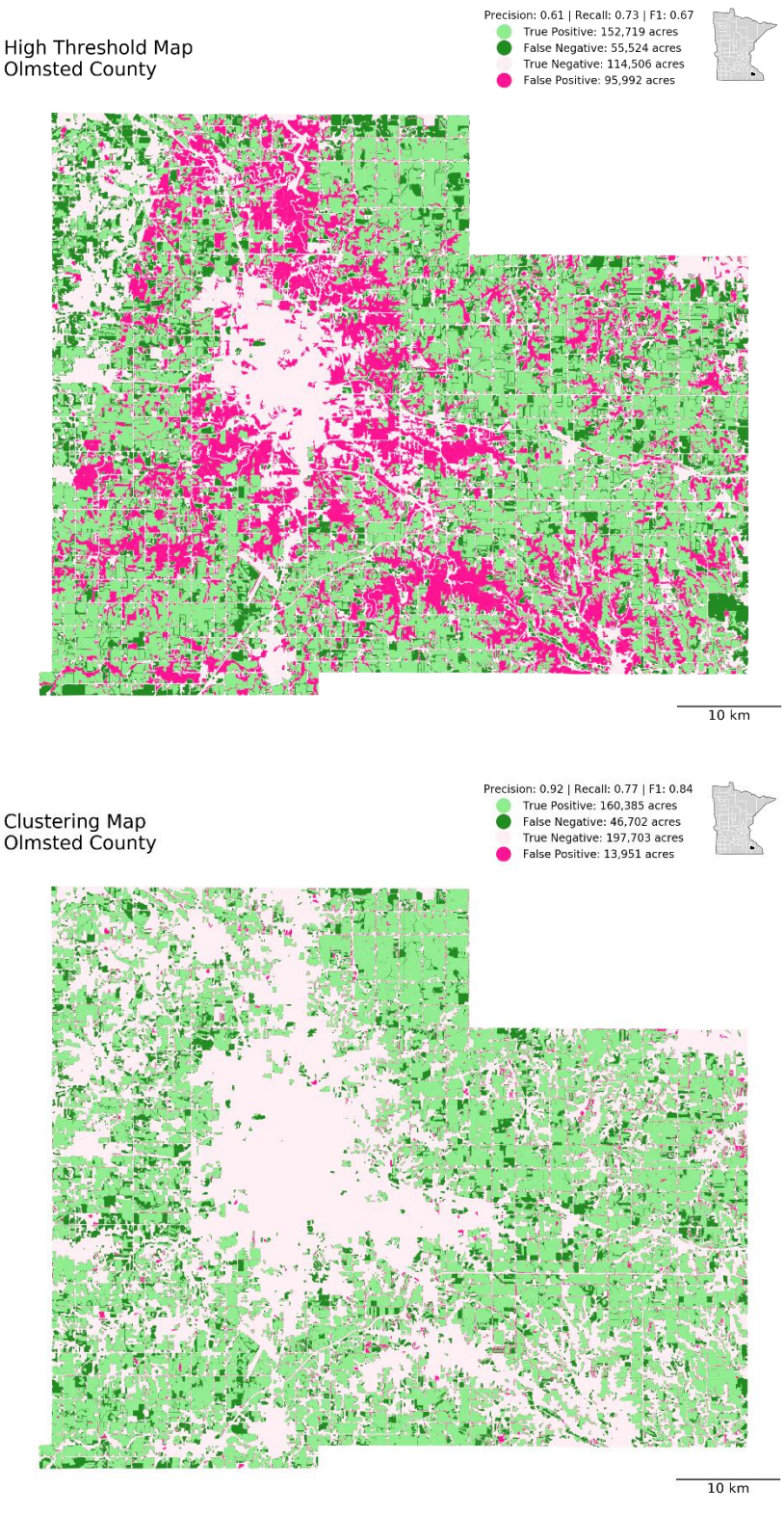

Figure 39. Olmsted County pixel-scale accuracy comparison of imagery artifacts

The Thresholding workflow configuration was also more sensitive to noise and artifacts in the imagery. With the lower threshold value for cloud coverage in the Thresholding configuration, tiles with uneven coverage due to the path of the satellite can produce artifacts in the cropland mapping. If there is an imagery date that only partially covers a tile, it is more likely to influence the temporal NDVI statistics if there are fewer total temporal observations. Thus, by utilizing the cloud mask with a higher cloud coverage threshold in the Clustering configuration, we can incorporate more cloud-free observations at each pixel and limit the influence of these types of differences in data availability across space. This does not completely eliminate these types of errors, as we can see in some counties in the Clustering Map, but the distinct swath of the satellite path is much less evident than in some of the Thresholding maps. Figure 39 shows High Threshold Map and the Clustering Map pixel-scale accuracy for Olmsted County in southeast Minnesota, with the city of Rochester at its center. The diagonal line evident in the High Threshold Map image on the left indicates where a difference in imagery due to the path of the satellite resulted in an arbitrary boundary in the areas being predicted as cropland. The Clustering Map, however, shows no indication of this artifact, indicating a more robust approach to normalizing the temporal data for areas with uneven imagery coverage.

Deep Learning Workflow Configuration

A deep learning convolutional model trained on noisy crop mask data shows potential for improving crop mask accuracy. It is potentially scalable because the training data is derived from the workflow itself, which we demonstrated with the Clustering workflow configuration could be accurate over large areas. One of the limitations of the clustering approach was the ambiguity in some classes between natural vegetation and cropland. The result was somewhat conservative overall, missing nearly one quarter of the cropland pixels statewide. Importantly, this method demonstrated that a CNN model trained on a mask with high precision and low recall, like the mask derived from the clustering approach, could significantly improve recall without sacrificing the precision of the prediction. Coupling these two configurations—Clustering and Deep Learning—has the potential to overcome the limitations of each individual approach.

There were limitations in the deep learning experiment we conducted. First, the experiment used clean data, hand-picked for cloud-free imagery, and focused on a region dominated by agriculture with consistent field feature shapes. Over the full geographic extent of the state, the model might not be able to learn the cropland features as effectively. Second, the crop mask labels that were tested had relatively high accuracy, suggesting that they were a relatively good proxy for fields, which might not be the case in other regions for the automated

Crop Mask Labels used for training. One other drawback to deep learning approaches is that training is often computationally intensive, requiring significantly more time and computational resources. Based on current results, the geographically scalable for a model trained for a single imagery tile or a single state remains unclear. Additional research would be required to test the generalizability of a model trained on imperfect crop masks over large areas. These results suggest that using imperfect data for training can effectively train a deep learning model to predict cropland. However, this was a very limited experiment and the generalizability and scalability of this approach is still to be tested.

Felzenszwalb Segmentation

The Felzenszwalb segmentation algorithm implemented in the Thresholding and Clustering workflows could delineate fields well in areas where the crop mask was accurate, but it suffered from over-segmentation in some areas and under-segmentation in other areas. The method takes several input parameters and the field size outputs are sensitive to those user-defined inputs. A single set of parameters for scale, size, and image blurring will not work for every location. This reiterates a similar limitation of the Thresholding configuration. User-defined parameters require local fine tuning to be accurate and this is not feasible for Minnesota-wide analysis. This will be crucial for designing configurations for global-scale field mapping. Different field delineation approaches would likely be needed to make the workflow more generalizable. Changing the inputs for segmentation to incorporate temporal characteristics can improve field delineation accuracy, but more robust segmentation methods will likely be needed to achieve generalizable high accuracy in field delineation. The incorporation of edge-finding techniques once the crop mask has been applied seems like a promising approach. The field boundary delineation method proposed by North et al (2019) effectively created field polygon by building edge maps based on edge length instead of simply edge magnitude. In places with large, usually rectangular fields like Minnesota, this approach could be effective when paired with a crop masking step.

A Flexible Workflow to Address the Computational Challenge

By testing three different configurations, we demonstrated that viability of our proposed workflow for being flexible to implementing different methods and quickly testing multiple approaches, providing a prototype for addressing the third research question. This is important both for continued research on field mapping to further refine our work but also for enabling geographic flexibility. For addressing the differences in fields around the world, a single

workflow configuration might not be effective everywhere but having the flexibility within the workflow to change the preprocessing data features, specific parameters, or even entire methods enables the configuration to be tailored based on local knowledge without having to rely on local external data. Being set up to run in a high-performance computing environment also reduces the limitations on processing. This workflow could be tested for different geographic regions or time periods simply by replacing the input imagery. It can serve as a foundational prototype for scaling up this field mapping research.

Scaling to Global-Extent with Geospatial Cyberinfrastructure

Overall, the conceptual framework was successful in our experiments, but additional work is required to transfer the workflow to an environment that can process the analysis globally. We were able to implement three different configurations to test multiple methods efficiently. Given the modular nature of the workflow, the strengths of each method can be integrated together in future research. While we were able to utilize storage and computational resources at the Minnesota Supercomputing Institute (MSI), we were still geographically limited to Minnesota during the 2019 growing season due to storage limitations at MSI.

To extend this kind of work globally and solve some of these computational limitations, incorporating additional HPC resources will be necessary. Extensive work is underway to develop the systems to analyze remote sensing data in real time at a global scale to handle the vast imagery data flows coming from remote sensing systems (Lunga, Bhaduri, and Stewart 2019). Transferring the workflow to resources on the XSEDE would provide an open science solution to global-scale analysis (Towns et al. 2014). This would both solve the computational processing challenges and enable collaboration on a global field mapping project with researchers across domains and around the world. Google Earth Engine has also been demonstrated as an effective geospatial cyber infrastructure resource for global-scale analysis (Gorelick et al. 2017). It is particularly well-suited for remote sensing as it connects the global database of Sentinel-2 imagery with cloud computing resources and automatic parallel processing for many tasks. Advances in HPC infrastructure and emerging deep learning techniques will soon solve the computational challenges of high-resolution global analysis. Field mapping is a challenge well suited to taking advantage of these advances.

Chapter 7 – Conclusion

We explored the potential for building a scalable field mapping workflow utilizing high-resolution Sentinel-2 imagery with flexible methods that do not require training data. The framework proved to be successful for rapid testing of different analytical approaches and integration of different methods for cropland mapping and field delineation through three different workflow configurations. We showed that fields can be mapped with high accuracy across the state of Minnesota with semi-supervised methods such as k-means clustering when paired with robust temporal vegetative features derived from Sentinel-2 data. We identified the limitations of using universal thresholds for large areas and identified a promising approach for implementing deep learning to extend traditional remote sensing techniques.

Future work in field mapping can utilize the flexible workflow framework introduced here to test new methods at each step in the workflow. The modular nature enables other to take the useful approaches demonstrated here and extend the analysis with new methods. With this scalable satellite imagery analytics system, we demonstrated that field mapping over large areas is possible, deployed unsupervised methods to make it geographically scalable, and explored promising cutting-edge deep learning methods that could hold potential for more advances in scalable analysis.

Advances in global field mapping will be critical for addressing the food security challenges facing the world in the coming century (Fritz et al. 2013). To provide solutions at the same scale as these challenges, we will need accurate, timely, and high-resolution data on where food is produced around the world (Maciel et al. 2019). A global field map would serve as a critical foundational dataset for ensuring there is enough food to feed the planet in the face of climate change and a growing population (Thenkabail et al. 2010). The data and computational resources necessary to bridge the gap between global-scale analysis and local-scale accuracy are now available. Building upon the framework presented here can help to meet the need for a high-resolution global map of crop fields.

Bibliography

Aguilar, Rosa, Raul Zurita-Milla, Emma Izquierdo-Verdiguier, and Rolf A. de By. 2018. “A Cloud-Based Multi-Temporal Ensemble Classifier to Map Smallholder Farming Systems.” *Remote Sensing* 10 (5): 729. https://doi.org/10.3390/rs10050729.

Atluri, Gowtham, Anuj Karpatne, and Vipin Kumar. 2018. “Spatio-Temporal Data Mining: A Survey of Problems and Methods.” *ACM Comput. Surv.* 51 (4): 83:1–83:41. https://doi.org/10.1145/3161602.

Atzberger, Clement. 2013. “Advances in Remote Sensing of Agriculture: Context Description, Existing Operational Monitoring Systems and Major Information Needs.” *Remote Sensing* 5 (2): 949–81. https://doi.org/10.3390/rs5020949.

Bartholomé, E., and A. S. Belward. 2005. “GLC2000: A New Approach to Global Land Cover Mapping from Earth Observation Data.” *International Journal of Remote Sensing* 26 (9): 1959–77. https://doi.org/10.1080/01431160412331291297.

Belgiu, Mariana, and Ovidiu Csillik. 2018. “Sentinel-2 Cropland Mapping Using Pixel-Based and Object-Based Time-Weighted Dynamic Time Warping Analysis.” *Remote Sensing of Environment* 204 (January): 509–23. https://doi.org/10.1016/j.rse.2017.10.005.

Boryan, Claire, Zhengwei Yang, Rick Mueller, and Mike Craig. 2011. “Monitoring US Agriculture: The US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program.” *Geocarto International* 26 (5): 341–58. https://doi.org/10.1080/10106049.2011.562309.

Cai, Yaping, Kaiyu Guan, Jian Peng, Shaowen Wang, Christopher Seifert, Brian Wardlow, and Zhan Li. 2018. “A High-Performance and in-Season Classification System of Field-Level Crop Types Using Time-Series Landsat Data and a Machine Learning Approach.” *Remote Sensing of Environment* 210 (June): 35–47. https://doi.org/10.1016/j.rse.2018.02.045.

Candra, Danang Surya, Stuart Phinn, and Peter Scarth. 2020. “Cloud and Cloud Shadow Masking for Sentinel-2 Using Multitemporal Images in Global Area.” *International Journal of Remote Sensing* 41 (8): 2877–2904. https://doi.org/10.1080/01431161.2019.1697006.

Chen, Jun, Jin Chen, Anping Liao, Xin Cao, Lijun Chen, Xuehong Chen, Chaoying He, et al. 2015. “Global Land Cover Mapping at 30m Resolution: A POK-Based Operational Approach.” *ISPRS Journal of Photogrammetry and Remote Sensing*, Global Land Cover Mapping and Monitoring, 103 (May): 7–27. https://doi.org/10.1016/j.isprsjprs.2014.09.002.

Debats, Stephanie R., Dee Luo, Lyndon D. Estes, Thomas J. Fuchs, and Kelly K. Caylor. 2016. "A Generalized Computer Vision Approach to Mapping Crop Fields in Heterogeneous Agricultural Landscapes." *Remote Sensing of Environment* 179 (June): 210–21. https://doi.org/10.1016/j.rse.2016.03.010.

Duro, Dennis C., Steven E. Franklin, and Monique G. Dubé. 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 Sensing of Environment* 118 (March): 259–72. https://doi.org/10.1016/j.rse.2011.11.020.

ESA. 2020. "Sentinel-2 - ESA Operational EO Missions - Earth Online - ESA." 2020. https://earth.esa.int/web/guest/missions/esa-operational-eo-missions/sentinel-2.

Felzenszwalb, Pedro F., and Daniel P. Huttenlocher. 2004. "Efficient Graph-Based Image Segmentation." *International Journal of Computer Vision* 59 (2): 167–81. https://doi.org/10.1023/B:VISI.0000022288.19776.77.

Forkuor, Gerald, Christopher Conrad, Michael Thiel, Tobias Ullmann, and Evence Zoungrana. 2014. "Integration of Optical and Synthetic Aperture Radar Imagery for Improving Crop Mapping in Northwestern Benin, West Africa." *Remote Sensing* 6 (7): 6472–99. https://doi.org/10.3390/rs6076472.

Friedl, Mark A., Damien Sulla-Menashe, Bin Tan, Annemarie Schneider, Navin Ramankutty, Adam Sibley, and Xiaoman Huang. 2010. "MODIS Collection 5 Global Land Cover: Algorithm Refinements and Characterization of New Datasets." *Remote Sensing of Environment* 114 (1): 168–82. https://doi.org/10.1016/j.rse.2009.08.016.

Fritz, Steffen, Linda See, Ian McCallum, Liangzhi You, Andriy Bun, Elena Moltchanova, Martina Duerauer, et al. 2015. "Mapping Global Cropland and Field Size." *Global Change Biology* 21 (5): 1980–92. https://doi.org/10.1111/gcb.12838.

Fritz, Steffen, Linda See, Liangzhi You, Chris Justice, Inbal Becker-Reshef, Lieven Bydekerke, Renato Cumani, et al. 2013. "The Need for Improved Maps of Global Cropland." *Eos, Transactions American Geophysical Union* 94 (3): 31–32. https://doi.org/10.1002/2013EO030006.

Gómez, Cristina, Joanne C. White, and Michael A. Wulder. 2016. "Optical Remotely Sensed Time Series Data for Land Cover Classification: A Review." *ISPRS Journal of Photogrammetry and Remote Sensing* 116 (June): 55–72. https://doi.org/10.1016/j.isprsjprs.2016.03.008.

Goodchild, Michael F., Huadong Guo, Alessandro Annoni, Ling Bian, Kees de Bie, Frederick Campbell, Max Craglia, et al. 2012. "Next-Generation Digital Earth." *Proceedings of the*

National Academy of Sciences of the United States of America 109 (28): 11088–94. https://doi.org/10.1073/pnas.1202383109.

Gore, Al. 1998. "The Digital Earth." *Australian Surveyor* 43 (2): 89–91. https://doi.org/10.1080/00050348.1998.10558728.

Gorelick, Noel, Matt Hancher, Mike Dixon, Simon Ilyushchenko, David Thau, and Rebecca Moore. 2017. "Google Earth Engine: Planetary-Scale Geospatial Analysis for Everyone." *Remote Sensing of Environment*, Big Remotely Sensed Data: tools, applications and experiences, 202 (December): 18–27. https://doi.org/10.1016/j.rse.2017.06.031.

Grace, Kathryn, Greg Husak, and Seth Bogle. 2014. "Estimating Agricultural Production in Marginal and Food Insecure Areas in Kenya Using Very High Resolution Remotely Sensed Imagery." *Applied Geography* 55 (December): 257–65. https://doi.org/10.1016/j.apgeog.2014.08.014.

Graesser, Jordan, and Navin Ramankutty. 2017. "Detection of Cropland Field Parcels from Landsat Imagery." *Remote Sensing of Environment* 201 (November): 165–80. https://doi.org/10.1016/j.rse.2017.08.027.

Hamman, Joseph, Matthew Rocklin, and Ryan Abernathy. 2018. "Pangeo: A Big-Data Ecosystem for Scalable Earth System Science" 20 (April): 12146.

Hansen, M. C., P. V. Potapov, R. Moore, M. Hancher, S. A. Turubanova, A. Tyukavina, D. Thau, et al. 2013. "High-Resolution Global Maps of 21st-Century Forest Cover Change." *Science* 342 (6160): 850–53. https://doi.org/10.1126/science.1244693.

Hey, Anthony J. G., ed. 2009. *The Fourth Paradigm: Data-Intensive Scientific Discovery*. Redmond, Washington: Microsoft Research.

Huang, Yanbo, Zhong-xin Chen, Tao Yu, Xiang-zhi Huang, and Xing-fa Gu. 2018. "Agricultural Remote Sensing Big Data: Management and Applications." *Journal of Integrative Agriculture* 17 (9): 1915–31. https://doi.org/10.1016/S2095-3119(17)61859-8.

Husak, G. J., M. T. Marshall, J. Michaelsen, D. Pedreros, C. Funk, and G. Galu. 2008. "Crop Area Estimation Using High and Medium Resolution Satellite Imagery in Areas with Complex Topography." *Journal of Geophysical Research: Atmospheres* 113 (D14). https://doi.org/10.1029/2007JD009175.

Intergovernmental Panel on Climate Change. "Special Report on Climate Change and Land — IPCC Site." 2019. Accessed May 16, 2020. https://www.ipcc.ch/srccl/.

Jenner, Lynn. 2015. "Landsat Overview." Text. NASA. April 1, 2015. http://www.nasa.gov/mission_pages/landsat/overview/index.html.

Jiao, Licheng, Maoguo Gong, Shuang Wang, Biao Hou, Zhi Zheng, and Qiaodi Wu. 2010. "Natural and Remote Sensing Image Segmentation Using Memetic Computing." *IEEE Computational Intelligence Magazine* 5 (2): 78–91. https://doi.org/10.1109/MCI.2010.936307.

Jin, Zhenong, George Azzari, and David B. Lobell. 2017. "Improving the Accuracy of Satellite-Based High-Resolution Yield Estimation: A Test of Multiple Scalable Approaches." *Agricultural and Forest Meteorology* 247 (December): 207–20. https://doi.org/10.1016/j.agrformet.2017.08.001.

Jin, Zhenong, George Azzari, Calum You, Stefania Di Tommaso, Stephen Aston, Marshall Burke, and David B. Lobell. 2019. "Smallholder Maize Area and Yield Mapping at National Scales with Google Earth Engine." *Remote Sensing of Environment* 228 (July): 115–28. https://doi.org/10.1016/j.rse.2019.04.016.

Johnson, David M., and Richard Mueller. 2010. "The 2009 Cropland Data Layer." *Photogramm. Eng. Remote Sens.* 76: 1201–1205.

Kamilaris, A., and F. X. Prenafeta-Boldú. 2018. "A Review of the Use of Convolutional Neural Networks in Agriculture." *The Journal of Agricultural Science* 156 (3): 312–22. https://doi.org/10.1017/S0021859618000436.

Karpatne, A., Z. Jiang, R. R. Vatsavai, S. Shekhar, and V. Kumar. 2016. "Monitoring Land-Cover Changes: A Machine-Learning Perspective." *IEEE Geoscience and Remote Sensing Magazine* 4 (2): 8–21. https://doi.org/10.1109/MGRS.2016.2528038.

Kussul, N., S. Skakun, A. Shelestov, M. Lavreniuk, B. Yailymov, and O. Kussul. 2015. "Regional Scale Crop Mapping Using Multi-Temporal Satellite Imagery." *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* XL-7/W3 (April): 45–52. https://doi.org/10.5194/isprsarchives-XL-7-W3-45-2015.

Lambert, Marie-Julie, Pierre C. Sibiry Traoré, Xavier Blaes, Philippe Baret, and Pierre Defourny. 2018. "Estimating Smallholder Crops Production at Village Level from Sentinel-2 Time Series in Mali's Cotton Belt." *Remote Sensing of Environment* 216 (October): 647–57. https://doi.org/10.1016/j.rse.2018.06.036.

Lark, Tyler J., Richard M. Mueller, David M. Johnson, and Holly K. Gibbs. 2017. "Measuring Land-Use and Land-Cover Change Using the U.S. Department of Agriculture's Cropland Data Layer: Cautions and Recommendations." *International Journal of Applied Earth Observation and Geoinformation* 62 (October): 224–35. https://doi.org/10.1016/j.jag.2017.06.007.

Lebourgeois, Valentine, Stéphane Dupuy, Élodie Vintrou, Maël Ameline, Suzanne Butler, and Agnès Bégué. 2017. "A Combined Random Forest and OBIA Classification Scheme for Mapping Smallholder Agriculture at Different Nomenclature Levels Using Multisource Data (Simulated Sentinel-2 Time Series, VHRS and DEM)." *Remote Sensing* 9 (3): 259. https://doi.org/10.3390/rs9030259.

Lee, Craig A., Samuel D. Gasster, Antonio Plaza, Chein-I Chang, and Bormin Huang. 2011. "Recent Developments in High Performance Computing for Remote Sensing: A Review." *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 4 (3): 508–27. https://doi.org/10.1109/JSTARS.2011.2162643.

Lesiv, Myroslava, Juan Carlos Laso Bayas, Linda See, Martina Duerauer, Domian Dahlia, Neal Durando, Rubul Hazarika, et al. 2019. "Estimating the Global Distribution of Field Size Using Crowdsourcing." *Global Change Biology* 25 (1): 174–86. https://doi.org/10.1111/gcb.14492.

Lobell, David B., David Thau, Christopher Seifert, Eric Engle, and Bertis Little. 2015. "A Scalable Satellite-Based Crop Yield Mapper." *Remote Sensing of Environment* 164 (July): 324–33. https://doi.org/10.1016/j.rse.2015.04.021.

Long, Jonathan, Evan Shelhamer, and Trevor Darrell. 2015. "Fully Convolutional Networks for Semantic Segmentation." *ArXiv:1411.4038 [Cs]*, March. http://arxiv.org/abs/1411.4038.

Lunga, Dalton D., Budhu Bhaduri, and Robert Stewart. 2019. "The Trillion Pixel GeoAI Challenge Workshop." ORNL/TM-2019/1442. Oak Ridge National Lab. (ORNL), Oak Ridge, TN (United States). https://doi.org/10.2172/1606744.

Ma, Lei, Manchun Li, Xiaoxue Ma, Liang Cheng, Peijun Du, and Yongxue Liu. 2017. "A Review of Supervised Object-Based Land-Cover Image Classification." *ISPRS Journal of Photogrammetry and Remote Sensing* 130 (August): 277–93. https://doi.org/10.1016/j.isprsjprs.2017.06.001.

Maciel, Adeline Marinho, Gilberto Camara, Lubia Vinhas, Michelle Cristina Araújo Picoli, Rodrigo Anzolin Begotti, and Luiz Fernando Ferreira Gomes de Assis. 2019. "A Spatiotemporal Calculus for Reasoning about Land-Use Trajectories." *International Journal of Geographical Information Science* 33 (1): 176–92. https://doi.org/10.1080/13658816.2018.1520235.

Maggiori, Emmanuel, Yuliya Tarabalka, Guillaume Charpiat, and Pierre Alliez. 2017. "High-Resolution Semantic Labeling with Convolutional Neural Networks." IEEE Transactions on Geoscience and Remote Sensing 55 (12): 7092–7103. https://doi.org/10.1109/TGRS.2017.2740362.

Mahdianpari, Masoud, Bahram Salehi, Fariba Mohammadimanesh, Saeid Homayouni, and Eric Gill. 2019. “The First Wetland Inventory Map of Newfoundland at a Spatial Resolution of 10 m Using Sentinel-1 and Sentinel-2 Data on the Google Earth Engine Cloud Computing Platform.” *Remote Sensing* 11 (1): 43. https://doi.org/10.3390/rs11010043.

Marmanis, Dimitrios, Konrad Schindler, Jan Dirk Wegner, Silvano Galliani, Mihai Datcu, and Uwe Stilla. 2017. “Classification With an Edge: Improving Semantic Image Segmentation with Boundary Detection.” ArXiv:1612.01337 [Cs], December. http://arxiv.org/abs/1612.01337.

Maxwell, Aaron E., Timothy A. Warner, and Fang Fang. 2018. “Implementation of Machine-Learning Classification in Remote Sensing: An Applied Review.” *International Journal of Remote Sensing* 39 (9): 2784–2817. https://doi.org/10.1080/01431161.2018.1433343.

McGlinchy, Joe, Brian Johnson, Brian Muller, Maxwell Joseph, and Jeremy Diaz. 2019. “Application of UNet Fully Convolutional Neural Network to Impervious Surface Segmentation in Urban Environment from High Resolution Satellite Imagery.” In *IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium*, 3915–18. https://doi.org/10.1109/IGARSS.2019.8900453.

Meilă, Marina. 2003. “Comparing Clusterings by the Variation of Information.” In Learning Theory and Kernel Machines, edited by Bernhard Schölkopf and Manfred K. Warmuth, 173–87. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-45167-9_14.

Mithal, Varun, Guruprasad Nayak, Ankush Khandelwal, Vipin Kumar, Nikunj C. Oza, and Ramakrishna Nemani. 2017. “RAPT: Rare Class Prediction in Absence of True Labels.” IEEE Transactions on Knowledge and Data Engineering 29 (11): 2484–97. https://doi.org/10.1109/TKDE.2017.2739739.

Mueller, Marina, Karl Segl, and Hermann Kaufmann. 2004. “Edge- and Region-Based Segmentation Technique for the Extraction of Large, Man-Made Objects in High-Resolution Satellite Imagery.” *Pattern Recognition* 37 (8): 1619–28. https://doi.org/10.1016/j.patcog.2004.03.001.

Mutanga, Onisimo, and Lalit Kumar. 2019. “Google Earth Engine Applications.” *Remote Sensing* 11 (5): 591. https://doi.org/10.3390/rs11050591.

North, Heather C., David Pairman, and Stella E. Belliss. 2019. “Boundary Delineation of Agricultural Fields in Multitemporal Satellite Imagery.” *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 12 (1): 237–51. https://doi.org/10.1109/JSTARS.2018.2884513.

Orynbaikyzy, Aiym, Ursula Gessner, and Christopher Conrad. 2019. “Crop Type Classification Using a Combination of Optical and Radar Remote Sensing Data: A Review.” *International Journal of Remote Sensing* 40 (17): 6553–95. https://doi.org/10.1080/01431161.2019.1569791.

“Pangeo.” 2020. 2020. https://pangeo.io/.

Papandreou, George, Liang-Chieh Chen, Kevin Murphy, and Alan L. Yuille. 2015. “Weakly- and Semi-Supervised Learning of a DCNN for Semantic Image Segmentation.” ArXiv:1502.02734 [Cs], October. http://arxiv.org/abs/1502.02734.

Pérez-Hoyos, Ana, Felix Rembold, Hervé Kerdiles, and Javier Gallego. 2017. “Comparison of Global Land Cover Datasets for Cropland Monitoring.” *Remote Sensing* 9 (11): 1118. https://doi.org/10.3390/rs9111118.

Phalke, Aparna R., and Mutlu Özdoğan. 2018. “Large Area Cropland Extent Mapping with Landsat Data and a Generalized Classifier.” *Remote Sensing of Environment* 219 (December): 180–95. https://doi.org/10.1016/j.rse.2018.09.025.

Pittman, Kyle, Matthew C. Hansen, Inbal Becker-Reshef, Peter V. Potapov, and Christopher O. Justice. 2010. “Estimating Global Cropland Extent with Multi-Year MODIS Data.” *Remote Sensing* 2 (7): 1844–63. https://doi.org/10.3390/rs2071844.

Ramankutty, Navin, Amato T. Evan, Chad Monfreda, and Jonathan A. Foley. 2008. “Farming the Planet: 1. Geographic Distribution of Global Agricultural Lands in the Year 2000.” *Global Biogeochemical Cycles* 22 (1). https://doi.org/10.1029/2007GB002952.

Reed, Daniel, Ruzena Bajcsy, Manuel Fernandez, Jose-Marie Griffiths, Randall Mott, Jack Dongarra, Chris Johnson, et al. 2005. “Computational Science: Ensuring America’s Competitiveness,” June, 117.

Ronneberger, Olaf, Philipp Fischer, and Thomas Brox. “U-Net: Convolutional Networks for Biomedical Image Segmentation.” *ArXiv:1505.04597 [Cs]*, May 18, 2015. http://arxiv.org/abs/1505.04597.

Rydberg, A., and G. Borgefors. 2001. “Integrated Method for Boundary Delineation of Agricultural Fields in Multispectral Satellite Images.” *IEEE Transactions on Geoscience and Remote Sensing* 39 (11): 2514–20. https://doi.org/10.1109/36.964989.

Schmedtmann, Jonas, and Manuel Campagnolo. 2015. “Reliable Crop Identification with Satellite Imagery in the Context of Common Agriculture Policy Subsidy Control.” *Remote Sensing* 7 (7): 9325–46. https://doi.org/10.3390/rs70709325.

See, Linda, Steffen Fritz, Liangzhi You, Navin Ramankutty, Mario Herrero, Chris Justice, Inbal Becker-Reshef, et al. 2015. “Improved Global Cropland Data as an Essential Ingredient

for Food Security." *Global Food Security* 4 (March): 37–45. https://doi.org/10.1016/j.gfs.2014.10.004.

Sherrah, Jamie. 2016. "Fully Convolutional Networks for Dense Semantic Labelling of High-Resolution Aerial Imagery." ArXiv:1606.02585 [Cs], June. http://arxiv.org/abs/1606.02585.

Shook, Eric, Davide Del Vento, Andrea Zonca, and Jun Wang. 2018. "GISandbox: A Science Gateway for Geospatial Computing." In Proceedings of the Practice and Experience on Advanced Research Computing - PEARC '18, 1–7. Pittsburgh, PA, USA: ACM Press. https://doi.org/10.1145/3219104.3219150.

Soille, P., and M. Pesaresi. 2002. "Advances in Mathematical Morphology Applied to Geoscience and Remote Sensing." *IEEE Transactions on Geoscience and Remote Sensing* 40 (9): 2042–55. https://doi.org/10.1109/TGRS.2002.804618.

Steinbach, Michael, Pang-Ning Tan, Vipin Kumar, Steven Klooster, and Christopher Potter. 2003. "Discovery of Climate Indices Using Clustering." In *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 446–455. KDD '03. Washington, D.C.: Association for Computing Machinery. https://doi.org/10.1145/956750.956801.

Story, Michael, and Russell G. Congalton. 1986. "Accuracy Assessment: A User's Perspective." Photogrammetric Engineering and Remote Sensing, 3.

Thenkabail, Prasad S., Munir A. Hanjra, Venkateswarlu Dheeravath, and Muralikrishna Gumma. 2010. "A Holistic View of Global Croplands and Their Water Use for Ensuring Global Food Security in the 21st Century through Advanced Remote Sensing and Non-Remote Sensing Approaches." *Remote Sensing* 2 (1): 211–61. https://doi.org/10.3390/rs2010211.

Towns, John, Timothy Cockerill, Maytal Dahan, Ian Foster, Kelly Gaither, Andrew Grimshaw, Victor Hazlewood, et al. 2014. "XSEDE: Accelerating Scientific Discovery." *Computing in Science Engineering* 16 (5): 62–74. https://doi.org/10.1109/MCSE.2014.80.

Traganos, Dimosthenis, Bharat Aggarwal, Dimitris Poursanidis, Konstantinos Topouzelis, Nektarios Chrysoulakis, and Peter Reinartz. 2018. "Towards Global-Scale Seagrass Mapping and Monitoring Using Sentinel-2 on Google Earth Engine: The Case Study of the Aegean and Ionian Seas." *Remote Sensing* 10 (8): 1227. https://doi.org/10.3390/rs10081227.

USDA. 2020. "Common Land Unit (CLU)." Page. Common Land Unit (CLU). 2020. https://www.fsa.usda.gov/programs-and-services/aerial-photography/imagery-products/common-land-unit-clu/index.

USDA - National Agricultural Statistics Service. 2019. “USDA - National Agricultural Statistics Service - Census of Agriculture.” 2019. https://www.nass.usda.gov/AgCensus/.

Vuolo, Francesco, Martin Neuwirth, Markus Immitzer, Clement Atzberger, and Wai-Tim Ng. 2018. “How Much Does Multi-Temporal Sentinel-2 Data Improve Crop Type Classification?” *International Journal of Applied Earth Observation and Geoinformation* 72 (October): 122–30. https://doi.org/10.1016/j.jag.2018.06.007.

Waldhoff, Guido, Ulrike Lussem, and Georg Bareth. 2017. “Multi-Data Approach for Remote Sensing-Based Regional Crop Rotation Mapping: A Case Study for the Rur Catchment, Germany.” *International Journal of Applied Earth Observation and Geoinformation* 61 (September): 55–69. https://doi.org/10.1016/j.jag.2017.04.009.

Waldner, François, Diego De Abelleyra, Santiago R. Verón, Miao Zhang, Bingfang Wu, Dmitry Plotnikov, Sergey Bartalev, et al. 2016. “Towards a Set of Agrosystem-Specific Cropland Mapping Methods to Address the Global Cropland Diversity.” *International Journal of Remote Sensing* 37 (14): 3196–3231. https://doi.org/10.1080/01431161.2016.1194545.

Waldner, François, Guadalupe Sepulcre Canto, and Pierre Defourny. 2015. “Automated Annual Cropland Mapping Using Knowledge-Based Temporal Features.” *ISPRS Journal of Photogrammetry and Remote Sensing* 110 (December): 1–13. https://doi.org/10.1016/j.isprsjprs.2015.09.013.

Waldner, François, Steffen Fritz, Antonio Di Gregorio, and Pierre Defourny. 2015. “Mapping Priorities to Focus Cropland Mapping Activities: Fitness Assessment of Existing Global, Regional and National Cropland Maps.” *Remote Sensing* 7 (6): 7959–86. https://doi.org/10.3390/rs70607959.

Wang, Sherrie, George Azzari, and David B. Lobell. 2019. “Crop Type Mapping without Field-Level Labels: Random Forest Transfer and Unsupervised Clustering Techniques.” *Remote Sensing of Environment* 222 (March): 303–17. https://doi.org/10.1016/j.rse.2018.12.026.

Wang, Zhibin, Kaiyi Wang, Feng Yang, Shouhui Pan, and Yanyun Han. 2018. “Image Segmentation of Overlapping Leaves Based on Chan–Vese Model and Sobel Operator.” *Information Processing in Agriculture* 5 (1): 1–10. https://doi.org/10.1016/j.inpa.2017.09.005.

Wardlow, Brian D., and Stephen L. Egbert. 2010. “A Comparison of MODIS 250-m EVI and NDVI Data for Crop Mapping: A Case Study for Southwest Kansas.” *International Journal of Remote Sensing* 31 (3): 805–30. https://doi.org/10.1080/01431160902897858.

Watkins, Barry, and Adriaan van Niekerk. 2019. “A Comparison of Object-Based Image Analysis Approaches for Field Boundary Delineation Using Multi-Temporal Sentinel-2 Imagery.” *Computers and Electronics in Agriculture* 158 (March): 294–302. https://doi.org/10.1016/j.compag.2019.02.009.

Whitcraft, Alyssa K., Eric F. Vermote, Inbal Becker-Reshef, and Christopher O. Justice. 2015. “Cloud Cover throughout the Agricultural Growing Season: Impacts on Passive Optical Earth Observations.” *Remote Sensing of Environment* 156 (January): 438–47. https://doi.org/10.1016/j.rse.2014.10.009.

White, Emma, and David Roy. 2015. “A Contemporary Decennial Examination of Changing Agricultural Field Sizes Using Landsat Time Series Data: Landsat Field Size Change.” *Geo: Geography and Environment* 2 (April). https://doi.org/10.1002/geo2.4.

Wit, A. J. W. De, and J. G. P. W. Clevers. 2004. “Efficiency and Accuracy of Per-Field Classification for Operational Crop Mapping.” *International Journal of Remote Sensing* 25 (20): 4091–4112. https://doi.org/10.1080/01431160310001619580.

Wright, D. J., and S. Wang. 2011. “The Emergence of Spatial Cyberinfrastructure.” *Proceedings of the National Academy of Sciences* 108 (14): 5488–91. https://doi.org/10.1073/pnas.1103051108.

Xiong, Jun, Prasad S. Thenkabail, James C. Tilton, Murali K. Gumma, Pardhasaradhi Teluguntla, Adam Oliphant, Russell G. Congalton, Kamini Yadav, and Noel Gorelick. 2017. “Nominal 30-m Cropland Extent Map of Continental Africa by Integrating Pixel-Based and Object-Based Algorithms Using Sentinel-2 and Landsat-8 Data on Google Earth Engine.” *Remote Sensing* 9 (10): 1065. https://doi.org/10.3390/rs9101065.

Yan, L., and D. P. Roy. 2014. “Automated Crop Field Extraction from Multi-Temporal Web Enabled Landsat Data.” *Remote Sensing of Environment* 144 (March): 42–64. https://doi.org/10.1016/j.rse.2014.01.006.

———. 2016. “Conterminous United States Crop Field Size Quantification from Multi-Temporal Landsat Data.” *Remote Sensing of Environment* 172 (January): 67–86. https://doi.org/10.1016/j.rse.2015.10.034.

Yang, Chaowei, Robert Raskin, Michael Goodchild, and Mark Gahegan. 2010. “Geospatial Cyberinfrastructure: Past, Present and Future.” *Computers, Environment and Urban Systems*, Geospatial Cyberinfrastructure, 34 (4): 264–77. https://doi.org/10.1016/j.compenvurbsys.2010.04.001.