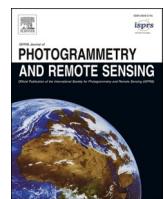


## A machine learning approach for identifying and delineating agricultural fields and their multi-temporal dynamics using three decades of Landsat data

Item Type	Article
Authors	Li, Ting;Johansen, Kasper;McCabe, Matthew
Citation	Li, T., Johansen, K., & McCabe, M. F. (2022). A machine learning approach for identifying and delineating agricultural fields and their multi-temporal dynamics using three decades of Landsat data. <i>ISPRS Journal of Photogrammetry and Remote Sensing</i> , 186, 83–101. <a href="https://doi.org/10.1016/j.isprsjprs.2022.02.002">https://doi.org/10.1016/j.isprsjprs.2022.02.002</a>
Eprint version	Publisher's Version/PDF
DOI	<a href="https://doi.org/10.1016/j.isprsjprs.2022.02.002">10.1016/j.isprsjprs.2022.02.002</a>
Publisher	Elsevier BV
Journal	ISPRS Journal of Photogrammetry and Remote Sensing
Rights	© 2022 The Author(s). Published by Elsevier B.V. on behalf of International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). This is an open access article under the CC BY-NC-ND license ( <a href="http://creativecommons.org/licenses/by-nc-nd/4.0/">http://creativecommons.org/licenses/by-nc-nd/4.0/</a> ).
Download date	2024-08-13 06:43:37
Item License	<a href="http://creativecommons.org/licenses/by-nc-nd/4.0/">http://creativecommons.org/licenses/by-nc-nd/4.0/</a>
Link to Item	<a href="http://hdl.handle.net/10754/675574">http://hdl.handle.net/10754/675574</a>



# A machine learning approach for identifying and delineating agricultural fields and their multi-temporal dynamics using three decades of Landsat data

Ting Li <sup>\*</sup>, Kasper Johansen, Matthew F. McCabe <sup>\*</sup>

*Water Desalination and Reuse Center, Division of Biological and Environmental Sciences and Engineering, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia*

## ARTICLE INFO

### Keywords:

Center-pivot field  
Delineation  
DBSCAN  
Convolution neural networks  
Spectral clustering  
Random forest

## ABSTRACT

High spatial and temporal resolution satellite imagery are essential data for land cover discrimination and mapping of vegetation dynamics, offering insights into the number, extent, and condition of agricultural fields. However, an accurate account of the number, location, and variability of fields can be challenging to obtain in a timely manner, particularly at scale or in regions where ground-supporting data are not available, limiting the capacity for food production and water use assessment and planning. To bridge this capacity gap, a convolutional neural network approach was adopted and combined with two clustering techniques: the density-based spatial clustering of applications with noise (DBSCAN) and spectral clustering, to provide an account of agricultural fields and their extent across a region lacking ground based information. A random forest classification was also employed to discriminate crop types. Using an annual maximum normalized difference vegetation index (NDVI) derived from 2018 Landsat-8 data, the approach was applied to classify the shape and delineate the extent of agricultural fields across an agricultural region in Saudi Arabia that had an area under irrigation exceeding 2,300 km<sup>2</sup>. When assessed against manually identified center-pivot fields (CPFs), the method achieved 97.4% producer's and 98.0% user's accuracies on an object basis, and 81.4% producer's and 85.4% user's accuracies on a pixel basis for identifying non-CPFs (i.e., tree crop plantations and other non-woody crops). The over- and under-segmentation error for CPFs was 1.5% and 1.0%, respectively, with intersection over union errors reported as being 3.5%. The framework showed stability when retrospectively applied to Landsat data from the year 2000, returning 97.5% producer's and 96.6% user's accuracies for CPF identification. In order to characterize the temporal dynamics of agricultural development over the past three decades, an analysis of field behavior between 1988 and 2020 was subsequently undertaken. The analysis indicated that the number of CPFs in the study region increased from 45 (covering 20 km<sup>2</sup>) in 1988 to 5,080 CPFs by 2016 (covering 2,368 km<sup>2</sup>), followed by a recent reduction to 3,700 CPFs in 2020 (covering 1,581 km<sup>2</sup>). Through the multi-temporal analysis, individual fields were able to be characterized in terms of their expansion, contraction and activity throughout the study period. Overall, the proposed method was simple to train, efficient in dealing with large datasets, relied on limited in-situ records to a very small degree, and has the potential to be applied to larger national scales, providing an ongoing assessment of important agroinformatique metrics.

## 1. Introduction

Globally, the agricultural sector accounts for more than two-thirds of global freshwater withdrawals (Pastor et al. 2019). In arid countries such as Saudi Arabia, that figure can exceed 80%. As most arid-land fresh-water abstractions are derived from groundwater, the consequence can be significant declines in groundwater levels and water

availability (Famiglietti 2014). Despite the lack of renewable fresh water, the acreage of agricultural land under irrigation in Saudi Arabia has seen considerable increases over the last four decades. Although published reports indicate that sprinklers account for more than 60% of the water used for irrigation in Saudi Arabia, with center-pivot systems being the dominant irrigation technique (Frenken 2009), that figure likely underestimates the actual amount and volume (López Valencia

<sup>\*</sup> Corresponding author.

E-mail address: [ting.li@kaust.edu.sa](mailto:ting.li@kaust.edu.sa) (T. Li).



et al. 2020). More generally, as in many developing regions of the world, there is a lack of basic information on crop dynamics, acreage, and even the number of center-pivot fields (CPFs). Such information forms the foundation of food- and water-security related studies, including the quantification of crop water use (Aragon et al. 2018; Melton et al. 2012) or crop type identification (Belgiu and Csillik 2018; Li et al. 2015; Peña et al. 2014). While improved use of groundwater resources and irrigation management are required for more sustainable development, this first requires an accurate accounting and assessment of the underlying agricultural activity. Apart from providing a needed historical account of agricultural development, annual or semi-annual mapping of CPFs would provide insight into crop dynamics and field conditions and offer the information required to determine the extent to which current agricultural activity may be complying with policies designed to reduce the ‘pressure on water’ (Frenken 2009).

Satellite images have been widely used to delineate agricultural fields for decades (Graesser and Ramankutty 2017; Rydberg and Borgefors 2001; Song et al. 2009). Indeed, one of the first satellite-based center-pivot monitoring efforts was proposed in the 1970s under NASA’s University Affairs Program (Rundquist et al. 1989), using manual interpretation of Landsat-MSS imagery over Nebraska. Following the development of image software tools in the 1990s, numerous studies have manually delineated CPFs by creating masks covering each individual field from both aerial and satellite imagery (Bauder 2004; Ferreira et al. 2011; Litts et al. 2001; Seth 2015). While certainly still applicable, such approaches are also laborious, time-consuming, and inevitably limited to smaller regions. To improve the scalability of agricultural field mapping, several automated and semi-automated techniques have been proposed to delineate fields, including edge-based, region-based, and shape-based techniques (Mueller et al. 2004; Xia et al. 2018; Zhao et al. 2015). As their name suggests, edge-based techniques focus on identifying the pixels representing field boundaries in an image. In region-based techniques, the homogeneity criterion between pixels forms the basis of grouping pixels into objects, whereas shape-based techniques segment the imagery based on shape features.

Numerous studies have applied various edge-based (Canny 1986) and region-based techniques e.g., multi-resolution segmentation (Baatz 2000) and watershed segmentation (Bleau and Leon 2000), to delineate agricultural fields within satellite imagery (Duro et al. 2012; Turker and Kok 2013; Vieira et al. 2012). In another approach, Yan and Roy (2016) utilized one-year of Landsat-5 and 7 time series data together with the United States Department of Agriculture (USDA) cropland data layer to quantify crop fields over the conterminous United States. When evaluated against reference data, 81.4% of the extracted field objects were correctly matched, with field size being underestimated by 1.2% on average. However, any requirement for crop type information can present as an impediment for wider application in regions where such data are not readily available. Graesser and Ramankutty (2017) proposed a method that incorporated multi-spectral image edge extraction, multi-scale contrast limited adaptive histogram equalization, and adaptive thresholding to detect cropland field parcels in agricultural landscapes across much of South America. Using Landsat Surface Reflectance Climate Data Record products the method achieved a 91% f-score. In another large-scale approach, Johansen et al. (2021) developed an object-based rule set that included edge detection and multiple object shape refinement steps to delineate CPFs across Saudi Arabia. Using Landsat-8 imagery collected over a 3-year period, they achieved a field detection accuracy of 94% for approximately 37,000 fields. Watkins and van Niekerk (2019) compared a framework combining two edge-based and three regional-based techniques to delineate field boundaries in South Africa using Sentinel-2 imagery. They found that Canny edge-detection in conjunction with watershed segmentation, provided an overall accuracy of 92.9% against a manually created reference boundary dataset. While CPFs were successfully delineated into individual fields, it was unclear if the method would be feasible for fields

planted with perennial crops.

In recent times, machine- and deep-learning techniques have become the state-of-the-art in solving agricultural-related segmentation problems (Lebourgeois et al. 2017; Masoud et al. 2020). Most machine learning algorithms can be broadly classified as unsupervised or supervised learning algorithms (Goodfellow et al. 2016). Clustering algorithms, which group data according to the rules within the data itself, are the most frequently employed approaches for unsupervised learning without a training stage. Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is an unsupervised learning algorithm that implements clustering based on the density of data points (Ester et al. 1996). While DBSCAN has been used to segment tree trunks (Tao et al. 2015) and identify woody components of trees (Ferrara et al. 2018) with high accuracies from LiDAR data, no study has yet examined its performance for segmenting agricultural fields within satellite imagery. Although DBSCAN can be expected to group image pixels into potential field units, it might struggle to separate pixels from closely connected fields. In this case, another approach may be required to segment adjacent fields. Spectral clustering is a technique that has been widely used for image segmentation (Shi and Malik 2000; Tung et al. 2010; Zhang et al. 2008), with the technique grouping data based on a similarity graph, i.e. a graph where the data points are linked by edges representing the similarity between each other (Von Luxburg 2007). However, the algorithm requires the number of clusters as model input, which is generally unknown. As such, it is often impractical to run spectral clustering on a satellite image in isolation, especially for data consisting of thousands of CPFs.

In a similar vein, convolutional neural networks (CNN) have become one of the most employed algorithms among the range of deep learning techniques. Many advanced CNNs have been developed based on the initial CNN architecture, referred to as LeNet (LeCun et al. 1989). These evolutions include Alexnet (Krizhevsky et al. 2012), VGGnet (Simonyan and Zisserman 2014), and U-net (Ronneberger et al. 2015), all of which have been successfully applied to the detection of CPFs from satellite imagery (Saraiva et al. 2020; Zhang et al. 2018). Zhang et al. (2018) contributed to one of the first explorations, utilizing Landsat-5 TM to analyze a 20,000 km<sup>2</sup> region in northeast Colorado, US. In that study, the authors compared the performance of three CNN models, including LeNet, Alexnet, and VGGnet, with LeNet outperforming the others, achieving 95.85% user’s and 93.33% producer’s accuracies. However, only the central location of each CPF was identified, with the extent of each field remaining unmapped. The fully convolutional network (FCN), which has its basis in CNN, is another state of the art approach for agricultural field segmentation (Long et al. 2015; Persello et al. 2019). Among various types of FCNs, the U-net model is the most frequently employed method for identifying CPFs. However, U-net has shown significant limitations, including an inability to detect commonly occurring fan-shaped CPFs, where only a part of a CPF is in use (Saraiva et al. 2020) and the merging of all CPFs into one layer instead of delineating individual fields (de Albuquerque et al. 2020; Graf et al. 2020). Instance segmentation using the region-based CNNs (e.g. Mask-RCNN) is another approach that has been applied to delineate CPFs (Carvalho et al. 2021; de Albuquerque et al. 2021a; de Albuquerque et al. 2021b). While CPFs have been delineated successfully with Mask-RCNN, the need for polygons masking individual CPFs in the training stage precludes its application to regions where such data are not available. Instead of employing a single algorithm to delineate CPFs (Mekhalfi et al. 2021; Tang et al. 2021b), some studies have developed hybrid frameworks combining different algorithms for this task. Waldner and Diakogiannis (2020) combined ResUNet-a and watershed segmentation to extract field boundaries using Sentinel-2 image data as input. The method proposed by Tang et al. (2021a) combining PVANET, GoogLeNet, and the Hough transform only required annotation of the location of a CPF using a rectangular box and whether an image object consisted of CPFs or non-CPFs. The delineation was conducted by identifying the center point and radius of a CPF obtained from the Hough transform. Although

incomplete circles, i.e. fan-shaped CPFs, were identified, the boundary of these fields was delineated as a full circle, risking an overestimation of field acreage.

Since the use of any single algorithm has previously demonstrated clear limitations for delineating fields, a hybrid machine learning framework was explored. By employing a hybrid approach, each of the algorithms performs its own specific function, but combine to provide a comprehensive solution. Indeed, while a number of machine learning models (e.g. FCNs and Mask-CNNs) can learn from the field boundary or field mask information in the training dataset (Long et al. 2015; Masoud et al. 2020; Persello et al. 2019), such datasets are not always routinely available (as in Saudi Arabia and most other regions worldwide). Manually creating the field boundaries or field mask is a time-consuming and potentially biased process, albeit one of the most frequent methods used to overcome this training data gap. However, such manual processing is impractical for large regions or dynamic features such as CPFs. Here, we implement an approach that does not require field delineation training data by instead identifying the shape of a field object and determining whether or not it consists of multiple fields: an approach that is both appropriate to the region and considerate of the data that is available. The data required (and available) to drive machine learning approaches is a key consideration in model-choice. In our case, Saudi Arabia is representative of the vast majority of regions outside of the developed world, where even basic agroinformatic data is not routinely available. Ultimately, we seek a compromise between usability, data availability, and computational efficiency. To identify the shape of agricultural fields (circle, fan, or others such as plantations), delineate the extent of the fields, and discriminate broad crop types (i.e., tree crop plantations and other non-woody crops), we propose an innovative framework to delineate individual fields by employing an approach that incorporates four machine learning techniques. The major element of this work combines DBSCAN, CNN classification, and spectral clustering in a stepwise manner to identify and delineate both circular and fan shapes agriculture fields using Landsat-8 data. Following this, a random forest (RF) classification model (Belgiu and Drăguț, 2016; Phalke et al., 2020; Rodriguez-Galiano et al., 2012a,b) was trained and applied to determine if a non-CPF represented a tree crop plantation or other non-woody plant crops. Once the framework was developed and tested, a second objective was to apply the approach to a long time series of Landsat data: in this case, covering the period from 1988 to 2020. This time series analysis allowed an assessment of the change and development in CPFs over the last three decades, including information on acreage for each field, field expansion or field decrease, as well as determining the active status of the fields (i.e. the number of years that

individual fields were in use). The motivating goal for this framework was to provide base-level information of relevance to studies exploring crop type mapping, management of crop water usage, and crop yield prediction, and to propose an approach that could be applied to information-sparse regions typical of the developing world.

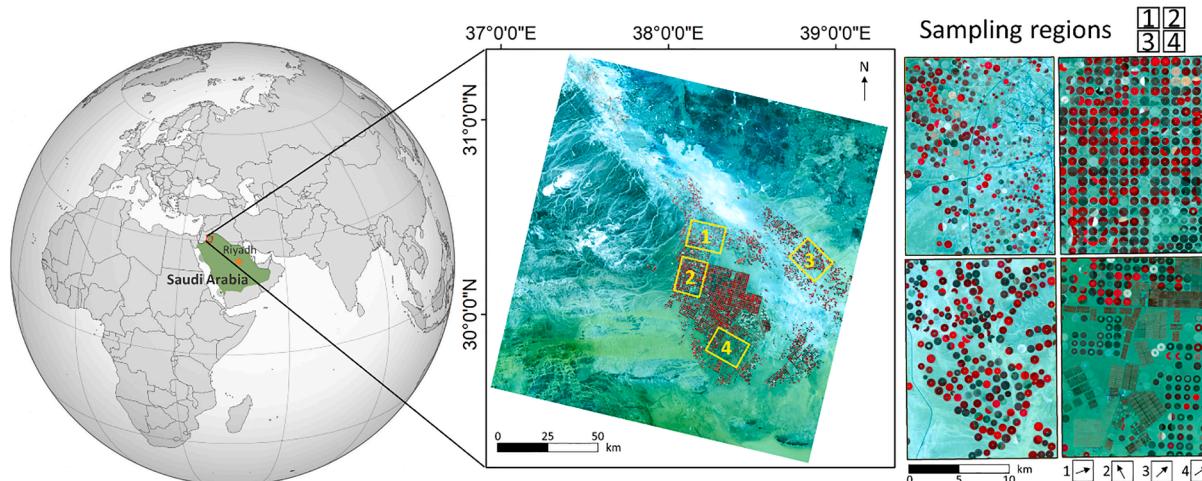
## 2. Study site and data description

### 2.1. Study site: the Al Jawf agricultural region in Saudi Arabia

The Al Jawf region in northwest Saudi Arabia is one of the largest agricultural regions in the Kingdom (see Fig. 1). While agricultural development was initiated in the 1980s, it increased rapidly to cover an irrigated area larger than 1,500 km<sup>2</sup> by 2005 (FAO 2013), with more than 90% of the fields being irrigated by abstracted groundwater (Al-Rumkhani and Din 2004). Most of the center-pivot systems are generally uniform in diameter (approximately 800 m) and adjacent to each other in a well-organized spatial pattern. However, the fields in the northwestern domain are more sporadic, with the field diameter smaller than those in the central portion of the study site. The most recent crop information from the General Authority for Statistics in Saudi Arabia (GASTAT, 2018), indicates that for 2014, the dominant crop types were cereals (including wheat, millet, broom-corn, maize, barley, and sesame), vegetables (including tomato, potato, marrow, eggplant, okra, carrot, dry onion, cucumber, melon, and watermelon), fodder (mainly clover), and fruit (including dates, citrus, and grapes). The average temperature in this region ranges from 10 to 32 °C, with the lowest and highest temperatures occurring in December - January and July - August, respectively (López Valencia et al. 2020).

### 2.2. Landsat data pre-processing

To develop a long-term time series of satellite data, imagery from Landsat-4/5 Thematic Mapper (TM), Landsat-7 Enhanced Thematic Mapper (ETM+), and Landsat-8 Operational Land Imager (OLI) were acquired. Level 2 surface reflectance data of 30-m resolution from path/row 172/39 were downloaded from the United States Geological Survey (USGS) EarthExplorer website (<https://earthexplorer.usgs.gov/>) for the period covering 1988 to 2020 (note that 1988 was the first year when CPFs were observed within the region). The surface reflectance data were created by the Land Surface Reflectance Code algorithm (Vermote et al. 2016) for Landsat-8 imagery and the Landsat Ecosystem Disturbance Adaptive Processing System algorithm (Schmidt et al. 2013) for the Landsat-4/5 and Landsat-7 imagery. Cloud-contaminated images



**Fig. 1.** False-color images of the study site (Landsat-8 path/row 172/39 collected on July 20, 2018). The four yellow rectangles indicate sampling regions to train and validate the CNN model. Each region covered 15 × 20 km. The arrows in the bottom right pointed to the north direction of each sampling region.

**Table 1**

The number of available images and selected images after excluding cloud contaminated images for each year (1988–2020) obtained from January 1 to December 31.

Year	Landsat Mission	Available images	Selected images	Year	Landsat Mission	Available images	Selected images	Year	Landsat Mission	Available images	Selected images
1988	Landsat-4/5 TM	21	18	1999	Landsat-5 TM	21	20	2010	Landsat-5 TM	16	15
1989	Landsat-4/5 TM	15	13	2000	Landsat-5 TM	18	17	2011	Landsat-5 TM	8	8
1990	Landsat-4/5 TM	17	15	2001	Landsat-5 TM	23	18	2012	Landsat-7 ETM+	20	18
1991	Landsat-4/5 TM	20	16	2002	Landsat-7 ETM+	16	16	2013	Landsat-8 OLI	20	17
1992	Landsat-5 TM	16	15	2003	Landsat-7 ETM+	14	13	2014	Landsat-8 OLI	21	17
1993	Landsat-5 TM	16	12	2004	Landsat-5 TM	19	19	2015	Landsat-8 OLI	23	21
1994	Landsat-5 TM	18	15	2005	Landsat-5 TM	13	11	2016	Landsat-8 OLI	23	21
1995	Landsat-5 TM	22	22	2006	Landsat-5 TM	12	9	2017	Landsat-8 OLI	23	23
1996	Landsat-5 TM	20	18	2007	Landsat-5 TM	12	12	2018	Landsat-8 OLI	23	19
1997	Landsat-5 TM	21	21	2008	Landsat-5 TM	13	11	2019	Landsat-8 OLI	23	19
1998	Landsat-5 TM	22	22	2009	Landsat-5 TM	17	15	2020	Landsat-8 OLI	21	18

were manually excluded (Table 1). Landsat-7 data for 2002, 2003, and 2012 were utilized due to the limited availability of Landsat-5 scenes in those years. Co-registration was applied to the satellite scenes using the AROSICS python package (<https://pypi.org/project/arosics/>), following the approach of Scheffler et al. (2017). As the Landsat tile covers slightly different regions, all the scenes were subset to the maximum overlap covering 6084 × 6346 pixels.

NDVI is one of the most frequently employed satellite-based vegetation indices and is used to characterize a variety of vegetation properties. NDVI maps were computed from the red and near-infrared bands for each Landsat scene from 1988 to 2020

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

where *NIR* and *red* are the surface reflectance of the near-infrared and red bands, respectively. Four maps for each year were obtained by extracting 1) the maximum, 2) standard derivation, 3) median, and 4) the difference between the maximum and minimum of the annual NDVI time series. The annual maximum NDVI map served as the base map for the identification and delineation of the fields and all four maps served as the input features of RF for crop type classification.

### 2.3. Ground truth data

Two ground truth datasets were prepared to evaluate the performance of the framework based on data from the years 2000 and 2018. The agricultural fields within the study region in both 2000 and 2018 were manually delineated using ArcMap based on the respective annual maximum NDVI maps. Each field was masked by a polygon and was labeled as being a circular CPF, a fan-shaped CPF, or other (non-CPFs). All fields in 2018, excluding those in the four sampling regions (yellow rectangles in Fig. 1), were used as a ground truth dataset and accounted for a total of 3,690 circular CPF covering 1,815 km<sup>2</sup>, 319 fan-shaped fields covering 88 km<sup>2</sup>, and non-CPFs covering 66.7 km<sup>2</sup>. In a similar manner, all fields obtained using Landsat-5 data in 2000 were used as another ground truth dataset to examine the application of the framework to different epochs. In 2000, a total of 1,640 circular fields covering 884 km<sup>2</sup>, 136 fan-shaped fields covering 28.2 km<sup>2</sup>, and non-CPFs covering 43.3 km<sup>2</sup> were included.

## 3. Description of the machine learning approaches and framework

A combination of four machine learning techniques formed the basis of the delineation framework. These comprised: 1) DBSCAN (Section 3.1) to segment the pixels of the Landsat-based annual maximum NDVI map into potential field objects; 2) CNN classification (Section 3.2) to identify the type of field objects (i.e., one circular CPF, one fan-shaped CPF, one non-CPF, multiple non-CPFs, merged fields, and noise); 3) spectral clustering (Section 3.3) to further segment multiple non-CPFs and merged fields into individual fields; and 4) RF classification

(Section 3.4) to discriminate the vegetation cover of non-CPFs (i.e., tree crop plantations or other non-woody crops). The DBSCAN, spectral clustering, and RF classification were implemented using the “Scikit-learn” Python package (Pedregosa et al. 2011), while the CNN training, validation, and prediction were performed using PyTorch (Mishra 2019). The evaluation process is detailed in Section 3.5, while a summary of the framework is presented in Section 3.6. The methods for multi-temporal analysis of field dynamics using three decades of Landsat data are described in Section 3.7.

### 3.1. Density-based spatial clustering of applications with noise (DBSCAN) clustering

DBSCAN is an algorithm that clusters a dataset based on the density of data points (or pixels in this particular study). The advantage of DBSCAN is that the clustering is performed based on the density of the points such that no prior knowledge about the number of clusters is required. In addition, the algorithm is efficient at dealing with noise within a dataset since points in regions with low density will be classified as noise and omitted. Two global parameters are set by the user, including the radius (*Eps*) for a point to search for its neighbors and the number of neighbors (*MinPts*) within *Eps* (Tan et al. 2006). For a single pixel, the number of border pixels (neighbors) was four. One cluster in DBSCAN is defined as a maximal set of density-connected points or, in other words, the density of points within a cluster is higher than the density of points outside the cluster (Ferrara et al. 2018; Tan et al. 2006). The selection of *Eps* and *MinPts* determines if a point is a core, border, or noise point and further affects the clustering result (Tan et al. 2006). DBSCAN was converted by Campello et al. (2013) into a hierarchical clustering algorithm: HDBSCAN, which applies DBSCAN over various *Eps* and finds the most stable clustering result over different *Eps*. In other words, HDBSCAN can find clusters with varying densities, which DBSCAN is unable to do (McInnes et al. 2017). While HDBSCAN outperforms DBSCAN from several aspects (Campello et al. 2013), it was not deemed suitable for this research since the intensity of CPFs in the study region showed spatial variation (Fig. 1). Therefore, DBSCAN was employed instead of HDBSCAN.

In our study, the data points in DBSCAN represented pixels in a raster image, i.e., the annual maximum NDVI map. Before running the DBSCAN, the annual maximum NDVI map was filtered by a threshold of NDVI ( $\text{NDVI}_{\text{thresh}}$ ) to enhance the difference between the density of pixels within the agricultural fields and the bare soil background. After the filtering, the coordinates of each pixel remaining on the thresholded annual maximum NDVI map were recorded as row and column locations. By doing so, a two-dimensional dataset was created to represent the pixels to be clustered, expressed as  $[(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)]$ , in which *n* was the total number of pixels remaining on the thresholded annual maximum NDVI map. Examination of the annual maximum NDVI image determined that *Eps* and *MinPts*, of 3 and 29, respectively, provided a suitable compromise for the study site for the Landsat imagery. The whole framework was implemented with  $\text{NDVI}_{\text{thresh}} = 0.25$  to

identify the CPFs first. Then, all pixels identified as CPFs were excluded from the annual maximum NDVI map and another  $NDVI_{thresh} = 0.15$  was applied to identify and delineate the non-CPFs.  $NDVI_{thresh} = 0.15$  was used because tree crop plantations in the study region occurred with lower annual maximum NDVI values than crops in CPFs, which could easily be excluded by  $NDVI_{thresh} = 0.25$ . However, initial processing with  $NDVI_{thresh} = 0.15$  for CPFs was not efficient in filtering out the background pixels and hence resulted in CPFs being merged. Therefore, to keep and map tree plantation pixels, while also segmenting the CPFs efficiently, two  $NDVI_{thresh}$  values were applied.

DBSCAN clustering can result in clusters consisting of a different number of pixels ranging from one to over several thousand. Two thresholds were set to identify the “large” clusters (by  $n_{high}$ ) that consisted of more than a single CPF, and the “small” clusters (by  $n_{low}$ ) that consisted of too few pixels to represent a CPF. The predominant center-pivot system had a diameter of 800 m, covering approximately 550 Landsat pixels. Although the size of fields varied in the study region, it was reasonable to assume that a cluster consisting of more than  $n_{high} = 5,000$  pixels was formed by multiple fields, while a cluster consisting of less than  $n_{low} = 20$  pixels was not a valid field. Small clusters were eliminated from the framework, while large clusters were re-clustered by another four iterations of DBSCAN clustering. Large clusters always consisted of pixels from multiple fields with overlapping boundaries. The pixels along the field edges frequently comprised soil background and thus had lower NDVI values than pixels fully covered by vegetation. The elimination of such areas assisted DBSCAN in clustering multiple fields with overlapping boundaries into individual fields. Hence, before re-running DBSCAN to segment the large clusters, pixels with NDVI values in the lowest 5% of the cluster were eliminated.

After running the iterative DBSCAN clustering, pixels within the thresholded annual maximum NDVI map were clustered into various objects: 1) objects that did not form an agricultural field; 2) a single CPF but with different shape (i.e., circular or fan shape); 3) fields not representing a circular or fan shape field (non-CPFs); or 4) merged multiple fields. An  $m \times n$  NDVI matrix was created for pixels within each cluster with their coordinates. Then, the  $m \times n$  matrix was padded to  $m \times m$  (if  $m \geq n$ ) or  $n \times n$  (if  $n > m$ ), with zeros filling the padded cells. The squared NDVI matrices served as input to the subsequent CNN classification. An example of the DBSCAN clustering procedure is shown in Fig. S1.

### 3.2. Convolutional neural networks (CNN)

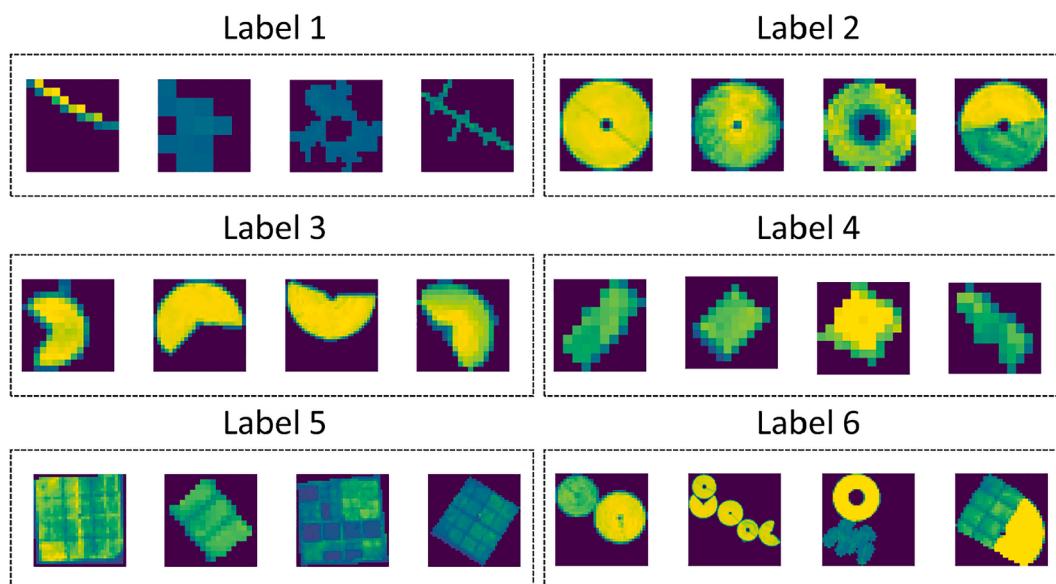
#### 3.2.1. Architecture of CNN

A CNN is an artificial neural network, a computational approach inspired by biological neural networks that process the information by neurons (also referred to as nodes) (Hopfield 1982). The nodes in three layers (i.e., one input layer, several hidden layers, and one output layer) in the artificial neural networks are connected by modeled weights (Schmidhuber 2015). The connections in the networks are updated iteratively to minimize the error between the predicted outputs and the actual outputs. In this application, we selected Alexnet (Fig. S2) due to its previous success in identifying CPFs within satellite imagery (Zhang et al. 2018) and simple architecture. The architecture of the Alexnet model used in our study reflects that published by Krizhevsky et al. (2012) and is shown in Fig. S2.

#### 3.2.2. Alexnet training and validation

Alexnet is a supervised learning technique requiring ground truth labels to train and validate the model. As described in Section 3.1, DBSCAN clustering was performed on the four sampling regions (see yellow rectangles in Fig. 1) to obtain representative ground truth data. The objects obtained after running DBSCAN were labeled as 1–6, representing: (1) noisy objects that did not form a valid field, (2) a single circular CPF, (3) a single fan CPF (part of the CPF), (4) a single non-CPF, (5) multiple non-CPFs, and (6) merged fields consisting of multiple fields (Fig. 2). The objects were mirrored and rotated to different angles (i.e., 90, 180, and 270 degrees) to balance the number of samples for each label. To increase the number of samples, small random perturbations were added to or subtracted from the NDVI matrices, while ensuring that the values within the matrix only ranged between 0 and 1. In total, 6,125 samples were created for each category, with 70% of these samples being randomly selected (from each category) to train the Alexnet model, while the remainder was used for validation.

The cross-entropy loss in the Alexnet model was updated by back-propagation using the mini-batch gradient descent method with the momentum, which is common in CNN training for classification to reduce the fluctuation of loss degradation and speed up the model training (Garcia-Garcia et al. 2018). As suggested by Ruder (2016), the learning rate and momentum coefficient were set to 0.001 and 0.9, respectively. In the first iteration of training, the weights for the



**Fig. 2.** Examples of field objects in six categories. Label 1 presented noisy objects that did not form a valid field. Labels 2, 3, and 4 represented a single circular center-pivot field (CPF), a single fan-shaped CPF, and a single non-CPF, respectively. Label 5 presented objects consisting of multiple non-CPFs, while Label 6 characterized objects consisting of merged fields from Labels 1 to 5. These objects were manually labeled for training and validation of the CNN classification model.

convolutional and fully connected layers in the Alexnet were initialized by the He initialization (He et al. 2015): a popular weight initialization method for ReLu activation functions. Fifty percent of the nodes in the first two fully-connected layers were randomly dropped from the network to preclude model overfitting, as suggested by Hinton et al. (2012).

Among all the hyperparameters needed to train the Alexnet model, the mini-batch size was the only one requiring tuning based on our parameter selection. The general mini-batch size for training a CNN model ranges from 2 to 2014, increasing by powers of 2 (Radiuk 2017). Experiments were conducted with different batch sizes (including 2, 4, 8, 16, 32, 64, and 128 images) to find the optimal mini-batch size for our dataset. The training stage stopped when the average loss for 400 consecutive iterations was smaller than 0.005, or the epoch (i.e., the number of passes of the entire training dataset by the algorithm) reached 100. The Alexnet performance was evaluated by the training time and four indices that are usually employed to assess a supervised classification algorithm: producer's accuracy, user's accuracy, F1, and overall accuracy (Olson and Delen 2008). The four indices are calculated as: *Producer accuracy* =  $TP/(TP + FN)$ , *User accuracy* =  $TP/(TP + FP)$ ,  $F1 = 2/(1/Producer\ accuracy + 1/User\ accuracy)$ , and *overall accuracy* =  $(TP + TN)/(TP + TN + FP + FN)$ , where  $TP$ ,  $FN$ ,  $FP$ , and  $TN$  are the numbers of true positives, false negatives, false positives, and true negatives, respectively. Balancing the model training time and accuracy, the experiment suggested that a batch size of 16 was optimal, with the producer's accuracy, user's accuracy, F1, and overall accuracy all exceeding 0.99. Thus, the Alexnet model was trained with a batch size of 16 to predict the shape of the fields.

### 3.3. Spectral clustering

#### 3.3.1. Algorithm of spectral clustering

Spectral clustering was employed to re-cluster the objects identified as multiple fields by Alexnet (Labels 5 and 6 in Fig. 2) into individual fields. Spectral clustering is a classic approach for clustering non-linear dataset based on a similarity graph, which is expressed as  $G = (V, E, W)$ , in which  $V$  is the set of vertices  $v_i$  (with each  $v_i$  representing a data point  $x_i$ );  $E$  is the set of edges between  $v_i$ ; and  $W$  represents the weight for the edges (Dhillon et al. 2004). In this study, the normalized spectral clustering proposed and described by Ng et al. (2002) was employed. The weight matrix was formed by the Gaussian kernel function of the gradient of NDVI values of adjacent pixels, i.e., the difference in NDVI values of two adjacent pixels, with the weights being zero for nonadjacent pixels, and is expressed as:

$$W_{ij} = \exp(-\frac{g_{ij}^2}{2\sigma^2}) \text{ if } i \neq j, \text{ and } W_{ii} = 0 \quad (2)$$

where  $W_{ij}$  was the weight of pixel  $i$  and  $j$ ;  $g_{ij}$  was the difference in NDVI values of pixels  $i$  and  $j$ ;  $g_{ij}$  was the gradient of NDVI values of adjacent pixels; and  $\sigma$  was the standard deviation of the values in the distance matrix formed by  $g_{ij}$ . The degree of point  $x_i$  is the sum of all edge weights between  $x_i$  and its adjacent data points, and is defined as (Von Luxburg 2007):

$$d_i = \sum_{j=1}^n w_{ij} \quad (3)$$

The degree matrix  $D$  for a dataset with  $n$  data points is a matrix with  $d_1, d_2, d_3, \dots, d_n$  on the diagonal. Then the Laplacian matrix is then defined as (Von Luxburg 2007):

$$L = D - W \quad (4)$$

The target of spectral clustering was to find the eigenvectors with the first  $k$  smallest eigenvalues by solving  $Lx = \lambda Dx$ , where  $k$  was the target number of clusters input by the user.

Spectral clustering requires the number of clusters to be produced

( $N_{clu}$ ) as input to run the algorithm. In our study, thousands of objects required re-clustering by the spectral clustering step. A procedure to automatically search for the optimal number of clusters of field objects consisting of multiple fields (Label 6 in Fig. 2) was explored by assessing the performance of two indices: the Silhouette coefficient (Rousseeuw 1987) and the Calinski-Harabasz index (Calinski and Harabasz 1974), which combine both cohesion and separation of clustering results for indicating the optimal number of clusters. Our experiment suggested that the Calinski-Harabasz index performed well in predicting the optimal number of clusters (result not shown here). However, a list of  $N_{clu}$  candidates for each object was required to reduce the computational time when searching for the optimal  $N_{clu}$  value. The number of pixels for each re-clustered object was divided by a threshold value ( $p_{thresh}$ ), which indicated the approximate number of pixels forming a CPF in the study region. The value of  $p_{thresh}$  was selected based on a CPF with an 800 m diameter, resulting in approximately 550 Landsat-8 pixels. Relaxed upper and lower boundaries of the  $N_{clu}$  candidates were required to adapt the framework to varying sizes of center-pivot systems. After testing on the sampling regions, the range of  $N_{clu}$  candidates was set based on the following rules:

```
if ratio ≤ 1, Nclu ∈ [2,3];
if 1 < ratio ≤ 10, Nclu ∈ [2, ⌈(ratio+5)⌉], with ⌈ ⌉ indicating round up to an integer; and
if ratio > 10, Nclu ∈ [⌈(ratio/3)⌉, ⌈(ratio+5)⌉ ], with ⌈ ⌉ indicating round down to an integer;
```

where  $ratio = \text{the number of pixels of an object}/p_{thresh}$ . On the other hand,  $N_{clu}$  was constrained to two for re-clustering objects consisting of multiple non-CPFs (Label 5 in Fig. 2) because of the uncertain pixel number that formed a non-CPF. This re-clustering for non-CPFs was necessary since the CNN might classify some objects consisting of CPFs as non-CPFs. By re-clustering objects that consisted of non-CPFs, CPFs could be segmented from the non-CPFs and classified to the correct category in the subsequent analysis.

#### 3.3.2. Enhancement of spectral clustering

The optimal  $N_{clu}$  obtained in Section 3.3.1 for label 6 might result in over-segmentation or under-segmentation errors when  $N_{clu}$  was higher or lower than the actual number of fields. Procedures were designed to reduce such errors. Over-segmentation error caused a single circular field to be segmented into several evenly divided fans, which resulted in linear boundaries of geographically adjacent clusters. To address this, the boundary pixels of two geographically adjacent clusters were counted. If the number of boundary pixels of two geographically adjacent clusters exceeded a value of 8, the Pearson correlation coefficient of the pixel coordinates (with  $x$  and  $y$  indicating the rows and columns of pixels in the object) was computed. If that correlation coefficient exceeded a value of 0.8, the two clusters were merged into one cluster. These two constraints were selected based on a priori experiment. For under-segmentation error, the algorithm failed to segment pixels from multiple fields into individual fields. Spectral clustering was run multiple times to address this problem. Before being re-clustered, pixels with NDVI values in the lowest 5% of the cluster, primarily occurring along the field edge, were eliminated to assist spectral clustering. The re-clustering procedure stopped when the iteration reached five, the object with label 5 consisted of less than 500 pixels, or no object was predicted as Label 5 or 6 by Alexnet. Three examples of spectral clustering with different Alexnet predicted labels are shown in Fig. S3.

### 3.4. Random forest classification

After running the procedures described in Sections 3.1 to 3.3, two labels were assigned to the pixels of the annual maximal NDVI map, i.e. the clustering label for each field and the CNN classification label

indicating the field shape. The non-CPFs were classified by an RF model to discriminate if they presented tree crops or other non-woody crops. Four features were employed as the input for the RF model, including the maximum, the standard derivation, the median, and the difference of maximum and minimum of the annual NDVI time series values.

#### 3.4.1. Random forest algorithm

RF classification is based upon building a number of individual decision trees, with each decision tree constructed by bootstrapping samples from the training dataset with replacement. A random subset or combination of variables is tested at each tree node to partition the training data into increasingly homogeneous subsets. The variable subset, which results in the greatest increase in data purity (e.g., highest Gini), is chosen for this node (Breiman 2001). After the forest of decision trees is constructed, the final classification result is accomplished by voting based on all the trees, i.e., each tree will give a label to the input data. The label voted by most of the trees would be assigned as the final classification result.

#### 3.4.2. Random forest classification model training and validation

Pixels occupied by tree crop plantations and other crops in sampling region 4 (shown in Fig. 1) were manually labeled using Landsat-8 imagery in 2018 based on information obtained from a previous in-person field survey. A total of 58,000 pixels were selected for each label, and five-fold cross-validation (Schaffer 1993) was applied to select the optimal input hyperparameters of the model. In five-fold cross-validation, the dataset is split into five subsets. The RF model is trained using four of the folds and is evaluated on the rest of the data. A grid search was applied to identify the optimal combination of the hyperparameters, including the number of trees ([10, 50, 100, 150, 200, 250, 300]), the number of variables to consider when looking for the best split ( $\lceil \sqrt{N} \log_2 N \rceil$ , where  $N$  is the number of input features) (Breiman 2001), and the function to measure the quality of a split ('gini', 'entropy') (Raileanu and Stoffel 2004). The combination of the hyperparameters that gave the highest mean value of overall accuracy in the five-fold cross-validation was employed to retrain the model using all the sampling data. The five-fold cross-validation experiment suggested that the optimal parameter set was [300,  $\log_2 N$ , 'entropy'] for the number of trees, the number of features when looking for the best split, and the function to measure the quality of a split, with a mean overall accuracy of 0.98. The RF model was trained on a pixel basis. However, the prediction was applied on an object basis since the standard deviation, median, and maximum-minimum difference maps for Landsat-7 data were affected by the SLC-off from May 31, 2003. Pixels affected by the SLC-off error showed extremely low NDVI values and resulted in wrong measures of the input features, which would mislead the RF classification when identifying the crop types. To overcome this influence, the RF prediction from 1988 to 2020 was constructed at the field level using the field delineation result obtained from Sections 3.1 to 3.3. The median value of the maximum, the standard deviation, median, and maximum-minimum difference value of each delineated field was computed based on the pixels that were unaffected by the SLC-off error and presented the input feature of the RF classification.

#### 3.5. Evaluation of the framework

As there was no ground-truth data available to actually evaluate the output from the RF model, we were unable to identify the crop type (tree crop plantation or other crops) of non-CPFs, except for within sampling region 4. There were fewer than 70 km<sup>2</sup> of fields that were not representing circular or fan-shaped fields over the approximately 1,970 km<sup>2</sup> of the agricultural area in the evaluation region in 2018. Our main focus was to assess the field identification and delineation results from Sections 3.1 to 3.3. A shapefile consisting of polygons representing each of the individual fields was produced. A polygon created by boundary

pixels of a cluster was better than the pixel-based mask since some of the pixels within the fields might be eliminated during the processing. These eliminated pixels can be recovered when we use polygons to extract all pixels forming each field from the base maps (e.g., annual maximum NDVI map).

The framework was evaluated based on the field detection accuracy and the delineation accuracy. The field detection accuracy was assessed at an object level for circular and fan-shaped CPFs and a pixel level for non-CPFs. The pixel-level evaluation was used for non-CPFs because the 30 m spatial resolution imagery precluded the detection of field edges for some of the non-CPFs (but especially for tree crop plantations), resulting in objects consisting of indivisible multiple non-CPFs. The producer's accuracy and user's accuracy were presented to quantify the detection accuracy of CPFs. However, the overall accuracy was excluded, since the framework-delineated fields did not completely match the fields in the ground truth map due to over- and under-segmentation errors.

The delineation accuracy of CPFs was evaluated by the over-segmentation ( $S_{\text{over}}$ ) and under-segmentation ( $S_{\text{under}}$ ) indices described by Clinton et al. (2010) and utilized by Yan and Roy (2014) for assessment of the field delineation accuracy, which are defined as:

$$S_{\text{under}}(E_j; T_i) = \left( 1 - \frac{|T_i \cap E_j|}{|E_j|} \right) \times 100\% \quad (5)$$

$$S_{\text{over}}(E_j; T_i) = \left( 1 - \frac{|T_i \cap E_j|}{|T_i|} \right) \times 100\% \quad (6)$$

where  $E_j$  represented the pixels from a framework-delineated field and  $T_i$  included the manually delineated pixels of a field from the ground truth map. If  $E_j$  intersected with multiple fields in the ground truth map or  $T_i$  intersected with multiple fields on the framework-delineated map, Equations (5) and (6) were computed only for the field that had the most intersecting pixels.  $S_{\text{union}}$ , an index computed from the root mean square of  $S_{\text{under}}$  and  $S_{\text{over}}$ , was also employed:

$$S_{\text{union}}(i,j) = \sqrt{\frac{S_{\text{over}(i,j)}^2 + S_{\text{under}(i,j)}^2}{2}} \quad (7)$$

$S_{\text{under}}$ ,  $S_{\text{over}}$ , and  $S_{\text{union}}$  provide values ranging from 0 to 100%, with a value of 0 defining a perfect delineation result and a value of 100% indicating either a high degree of under-segmentation error, over-segmentation error, or combined under- and over-segmentation error, respectively. The Intersection over Union (IoU) error (Huang et al. 2020) was also employed, which is defined as:

$$\text{IoU}(A, B) \text{ error} = (1 - \text{area}(A \cap B) / \text{area}(A \cup B)) \times 100\% \quad (8)$$

For a CPF,  $A$  was a mapped field based on the machine learning approaches, and  $B$  represented a ground truth field. The IoU error ranges from 0 to 100%, with a value of 100% indicating an incorrect result and a value of 0 indicating a correct result. In addition, four standard statistical metrics frequently employed to evaluate model performance were computed based on the area of the fields in the framework-delineated map and in the ground truth map. These included the coefficient of determination ( $R^2$ ), the mean absolute bias deviation (MAD), the relative MAD (rMAD), and relative mean bias deviation (rMBD), with the latter two being the absolute or mean bias divided by the mean of the number of pixels of extracted fields.

#### 3.6. Framework overview

A summary of the processing workflow is presented in Fig. 3. Using annual Landsat surface reflectance imagery as initial input (Step 1), the annual maximum NDVI map was first developed (Step 2) and then used to eliminate pixels with NDVI values below  $NDVI_{\text{thresh}} (=0.25)$  (Step 3).

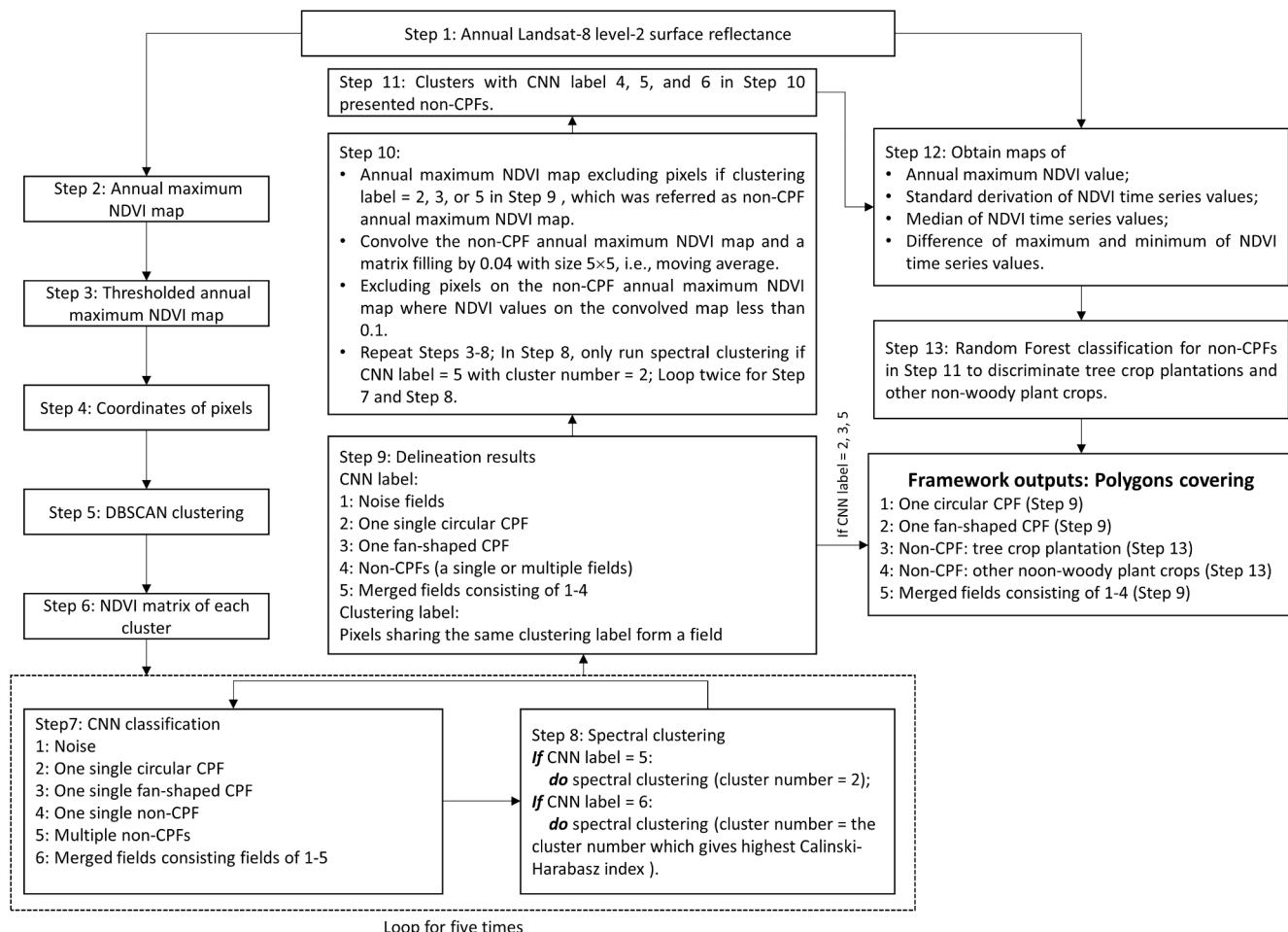


Fig. 3. Schematic of the methodological framework.

Then, the coordinates of pixels (row and column locations) of the thresholded NDVI map were obtained and used as input to the DBSCAN clustering (Steps 4–6). After that, a pre-trained CNN (Section 3.2.2) model was applied to identify the shape of field objects (clusters) obtained from DBSCAN (Step 7). Spectral clustering was applied to re-cluster the objects identified as non-CPFs or objects consisting of multiple fields (Step 8). After running five iterations of CNN predictions and spectral clustering, each pixel will have two types of label, i.e., the CNN label (indicating the shape of the field that a pixel belonged to) and the cluster label (pixels sharing the same cluster label forming a field). However, tree crop plantations present lower annual maximum NDVI values, which could easily be excluded by the multiple iterations described above. To address this, a non-CPF annual maximum NDVI map was created by excluding the pixels identified as CPFs and objects consisting of multiple fields (Step 10). Then the non-CPF annual maximum NDVI map was smoothed using a 5 × 5 pixels moving window to remove noise. After that, Steps 3 to 8 were repeated on the denoised non-CPF annual maximal NDVI map (starting from thresholding  $NDVI_{thresh} = 0.15$ ). The pre-trained RF model (Section 3.4.2) was then applied to further classify non-CPF crops as either tree crop plantations or non-woody plants (Steps 12 – 13). The final outputs of the framework were polygons covering individual fields (delineation result) and field shape maps showing if a field represented a single circular CPF; a single fan-shaped CPF; non-CPFs, i.e. tree crop plantations or other non-woody crops; or merged fields. The parameter settings for the framework are summarized in Table S1.

### 3.7. Field dynamics analysis from 1988 to 2020

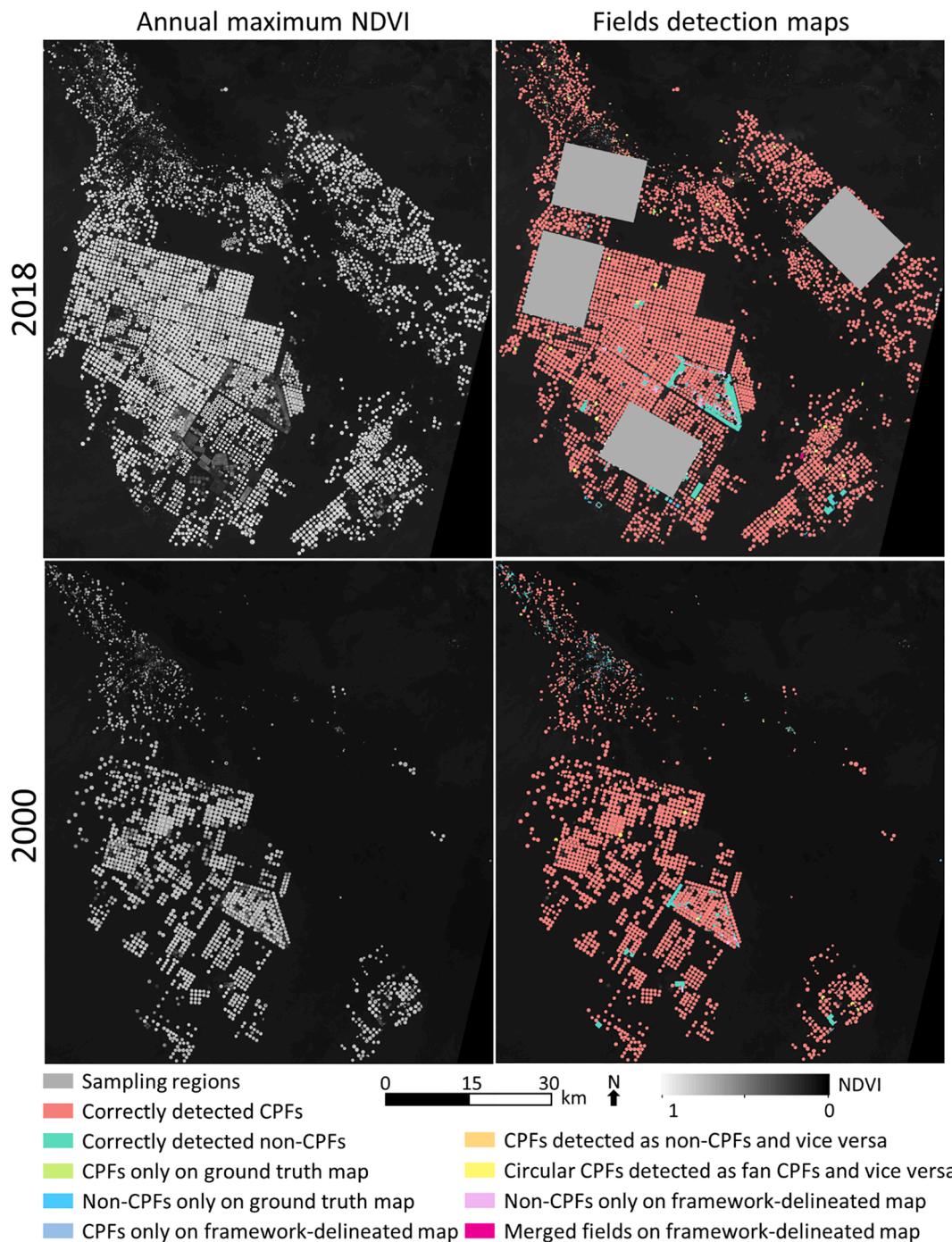
The 2018-trained framework was used to identify and delineate the fields in the study region for each year from 1988 to 2020. Firstly, the number of circular and fan-shaped fields for each year was counted. We did not count the number of non-CPFs as the 30 m resolution imagery failed to present the field edge for non-CPFs, especially for tree plantations, resulting in merged plantation fields. Then, the acreage of each field was obtained by the sum of areas corresponding to the pixels. Pixels forming a CPF with an NDVI value of less than 0.25 were excluded from acreage counting. This filtering was conducted as some of the CPFs represented donut shapes with no crops grown in the center of the field. The polygon created in the framework covered all the pixels of a CPF including those with no crops grown. These pixels were excluded when calculating the acreage of a CPF. Thirdly, the delineated fields for every two consecutive years were compared. Herein, if more than 60% of pixels forming a CPF observed in one year were duplicated with pixels forming a CPF observed in the year after, the two CPFs were defined as a duplicated field. Then, the relative bias of field acreage of the duplicated fields was calculated, which was defined as  $bias = (A_{af} - A_{bf})/A_{af}$ , where  $a$  takes its value from [1989, 1990, 1991, ..., 2020],  $bf = af - 1$ , and  $A_{af}$  and  $A_{bf}$  indicated the acreage of a duplicated CPF observed in years  $bf$  and  $af$ . A field was defined as a “duplicated field” if the intersection of the field acreage in two consecutive years ( $af$  and  $bf$ ) was  $\geq 0.6 A_{af}$  or  $0.6 A_{bf}$ . Based on the duplicated fields, we also identified fields that were observed in year  $af$  but were not observed in year  $bf$ , and vice versa. By doing so, the field dynamics were quantified on a yearly basis.

Moreover, the field expansion and contraction maps were obtained by the earliest and latest (before 2020) detection year of a pixel that formed a field. Finally, a CPF was defined as active if the median of annual maximum NDVI pixel values forming the field exceeded 0.35. For a non-CPF pixel, it was identified as active as long as it was detected. The number of years that a CPF was detected as active was counted as relative activity of the fields representing the number of years out of the time series that the field was in production.

## 4. Results

### 4.1. Evaluation of the framework

As can be seen from Fig. 4, the field delineation maps indicate that the framework performed well at detecting the fields using Landsat-8 (2018) and Landsat-5 (2000) imagery. In 2018, 97.4% of the 4,009 CPFs from the ground truth map were delineated correctly (i.e., producer's accuracy), while 98.0% of the 3,983 framework-delineated CPFs were correctly present on the ground truth map (i.e., user's accuracy)



**Fig. 4.** Maximum annual NDVI maps and associated validation results of field detection for both 2018 and 2000, highlighting correctly detected fields and different types of errors. The framework was trained using data from the four sampling regions in 2018 (grey box), with the machine learning model then applied to delineate fields across the rest of the region in 2018 and for the entire region of 2000. Center-pivot fields (CPFs) represented circular and fan-shaped fields, while non-CPFs represented the union of tree crop plantations and other non-woody crops.

**Table 2**

The accuracy for field detection using an object-basis for center-pivot fields (CPFs) in 2018 and 2000. CPFs represented circular (Circle) and fan-shaped (Fan) fields.

Framework-delineated map		Ground Truth			
		2018		2000	
		Circle	Fan	Circle	Fan
Total CPFs on ground truth map		4,009		1,776	
Total CPFs on the framework-delineated map		3,983		1,793	
Merged fields on the framework-delineated map		1		0	
CPFs not detected		16		3	
CPFs detected as non-CPFs		33		16	
Non-CPFs detected as CPFs		23		36	
Producer's accuracy		97.4%		97.5%	
User's accuracy		98.0%		96.6%	

(Table 2). The producer's and user's accuracy for CPF detection in 2000 was 97.5% and 96.6% (Table 3), respectively, which were similar accuracies to those in 2018. Training the machine learning framework on 2018 imagery and applying this to the imagery from 2000 suggested good multi-temporal stability of the framework for the same region (Fig. 4).

The main error for incorrect detection of CPFs in 2018 was due to circular CPFs being identified as fan-shaped CPFs and vice versa, accounting for 52.4% and 69.6% of the omission and commission errors of CPFs in 2018 (Table 2), respectively. CPFs being identified as non-CPFs accounted for 31.4% of omission errors, while non-CPFs being identified as CPFs accounted for 29.1% of the commission errors in 2018. Four main reasons were causing these two types of errors. The first was field size, with fields of less than 50 Landsat pixels precluding shape identification at the 30 m spatial resolution (Fig. 5a and b). Heterogeneity in the NDVI values of pixels forming some of the CPFs was another cause. Pixels with relatively low NDVI values (e.g.,  $<= 0.25$ ) were excluded by the framework, causing misrepresentation of circular/fan-shaped fields by the remaining pixels  $>0.25$ , and were thus incorrectly identified (Fig. 5e). The third cause was the over-segmentation of fields. In the example of Fig. 5d, a circular field was divided into, and subsequently identified as, two fan-shaped fields. Such over-segmentation error occurred for some fields due to the irrigation sprinkler system forming a line through the field with low NDVI values and high NDVI gradients, which resulted in fields being partitioned into multiple clusters. The fourth cause occurred due to under-segmentation, where multiple merged fields were incorrectly identified as fan-shaped fields or non-CPFs (Fig. 5f). This type of error was mainly due to overlapping edge pixels between neighboring fields, which reduced the gradient of NDVI values in the adjoining area (as opposed to CPF edge pixels adjoining bare ground) and hence prevented field separation. The over-segmentation and under-segmentation errors could also lead to CPFs being identified as non-CPFs, e.g. if parts of a field or merged fields appeared non-circular or non-fan-shaped. In addition, 16 of CPFs on the

ground truth map in 2018 were not detected by the framework, accounting for 15.2% of the identification errors of CPFs. This type of error mainly occurred to small CPFs, CPFs with low NDVI values (Fig. 5b), and when small CPFs were closely surrounded by large CPFs (Fig. 5g, some pixels belonging to smaller CPFs were clustered to the surrounding large CPFs, resulting in the small CPFs being misidentified as non-CPFs).

Omission errors were mainly caused by 14.5% of the non-CPF pixels from the ground truth map not being mapped by the framework, which accounted for 77.9% of the omission error. This was due to two main reasons. First, non-CPFs consisting of more than 20 pixels on the ground truth map but containing less than 20 pixels on the framework-delineated map were eliminated by the framework (Fig. 5a and b). The second reason was the exclusion of low pixels values of NDVI on the framework-delineated map, which might result in (1) a non-CPF on the ground truth map being over-segmented into several subparts each of which consisting of less than 20 pixels and thus being excluded by the framework, and (2) some non-CPFs, especially for young tree crop plantations (Fig. 5g) blended into the sandy background and were hence not identified by the framework. The primary error in the framework-delineated map was due to the incorrect identification of non-CPF pixels that did not exist on the ground truth map, which accounted for 72.4% of the commission errors. This was mainly due to the non-CPFs' detection procedure utilizing a value of 0.15 of NDVI to eliminate bare soil background pixels, resulting in some background pixels surrounding the non-CPFs (with NDVI exceeding 0.15) being identified as part of non-CPFs (pixels in pink color in Fig. 5h).

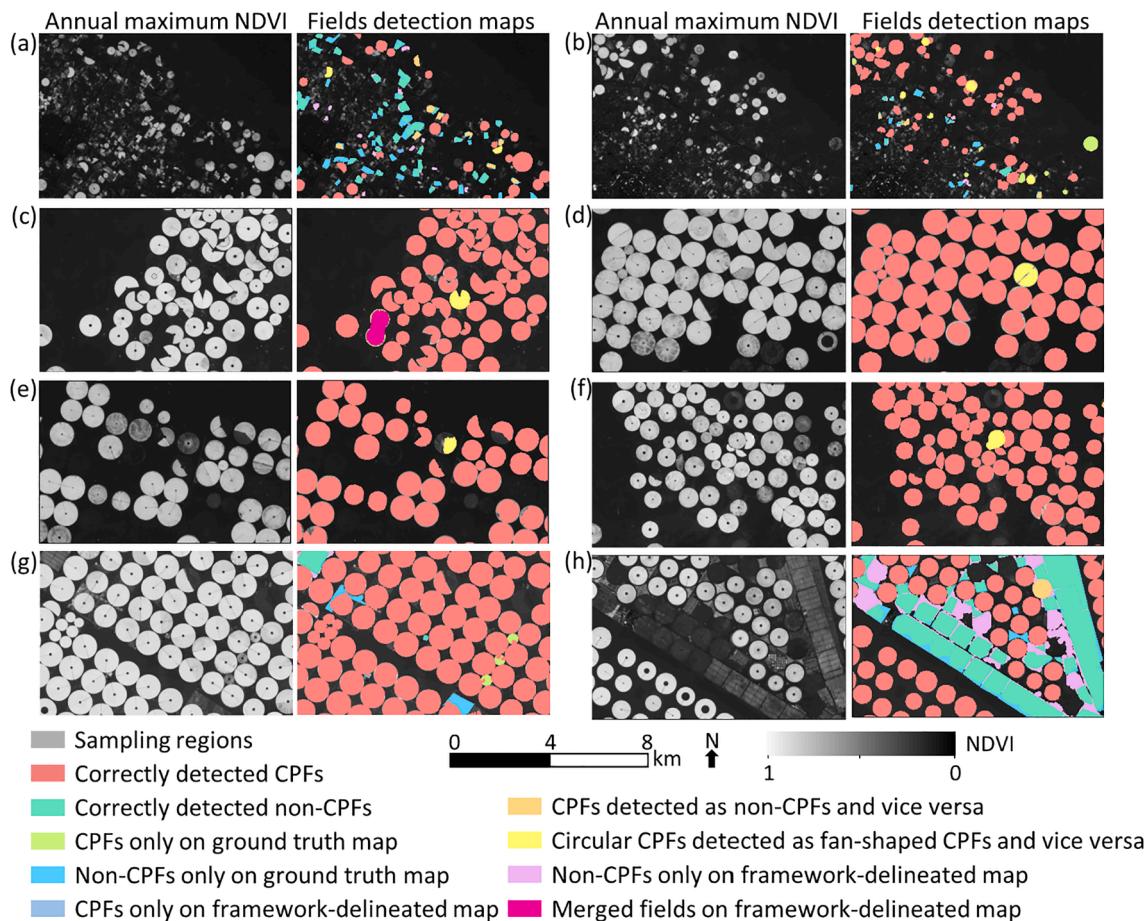
The non-CPF detection accuracies in 2000 were lower than those in 2018 (Tables 2 and 3), suggesting a limitation in applying the model trained in 2018 to other periods in terms of identifying non-CPFs. This was because the non-CPFs in 2000 appeared more complex with more within-field variation than those in 2018, as many crop fields in the northwestern region were not irrigated by center-pivot systems in 2000. The 2018-trained framework was not familiar with such conditions, since the main samples representing the non-CPFs in the training dataset were from plantations in 2018.

A second validation step assessed the delineation accuracy of CPFs at the pixel level. The median values of  $S_{over}$ ,  $S_{under}$ ,  $S_{union}$ , and  $IoU$  errors were 1.5% (1.7%), 1.0% (0.5%), 1.4% (1.6%), and 3.5% (3.7%) for 2018 and 2000 (in brackets), respectively (Fig. 6a), which indicated high overlap between the estimated fields and ground truth fields. The median and the upper quartile of  $S_{over}$  was slightly higher than  $S_{under}$ , suggesting that the framework produced slightly more over-segmentation errors than under-segmentation errors for both years. The  $R^2$  of the acreage of individual fields between the CPFs on the framework-delineated map and the ground truth map was 0.99 and 0.98 for 2018 and 2000, respectively (Fig. 6b and c). The negative  $rMBD$  values for both years indicated that the acreage of CPFs was generally under-estimated. This was due to the elimination of pixels with NDVI values higher than 0.25 (5% lowest NDVI value in each cluster) in the repeated DBSCAN and spectral clustering procedures, which were mainly pixels along the field boundary. The pixel-level assessment also suggested that the model trained using 2018 data performed well in 2000 in the same region for delineating the extent of individual CPFs.

**Table 3**

The accuracy for field detection using a pixel-basis for non-center-pivot fields (non-CPFs) in 2018 and 2000.

	2018	2000
Correctly detected non-CPF pixels	60,352	34,213
Total non-CPF pixels on ground truth map	74,115	48,081
Total non-CPF pixels on the framework-delineated map	70,641	41,721
Non-CPF pixels not detected	10,718	10,709
Non-CPF pixels detected as CPFs	3,045	3,519
Non-CPF pixels not present on ground truth map but present on framework delineated map	7,451	5,669
Producer's accuracy	81.4%	71.2%
User's accuracy	85.4%	82.0%



**Fig. 5.** Examples of field identification errors showing (a) small circular center-pivot fields (CPFs) being identified as fan-shaped CPFs and vice versa, and non-CPFs present in the ground truth map not being detected by the framework; (b) CPFs and non-CPFs present in the ground truth map but not detected by the framework; (c) merged fields and a fan-shaped CPF detected as a circular CPF; (d) a circular CPF being identified as fan-shaped CPF due to over-segmentation; (e) a circular CPF being identified as a fan-shaped CPF due to elimination of pixels with low NDVI values; (f) two circular CPFs being identified as a fan-shaped CPF; (g) three small circular CPFs and some non-CPFs with low NDVI values not detected by the framework; and (h) non-CPF pixels only on the framework-delineated map or only on the ground truth map. CPFs represented circular and fan-shaped center-pivot fields, while non-CPFs represented the union of tree crop plantations and other non-woody crops.

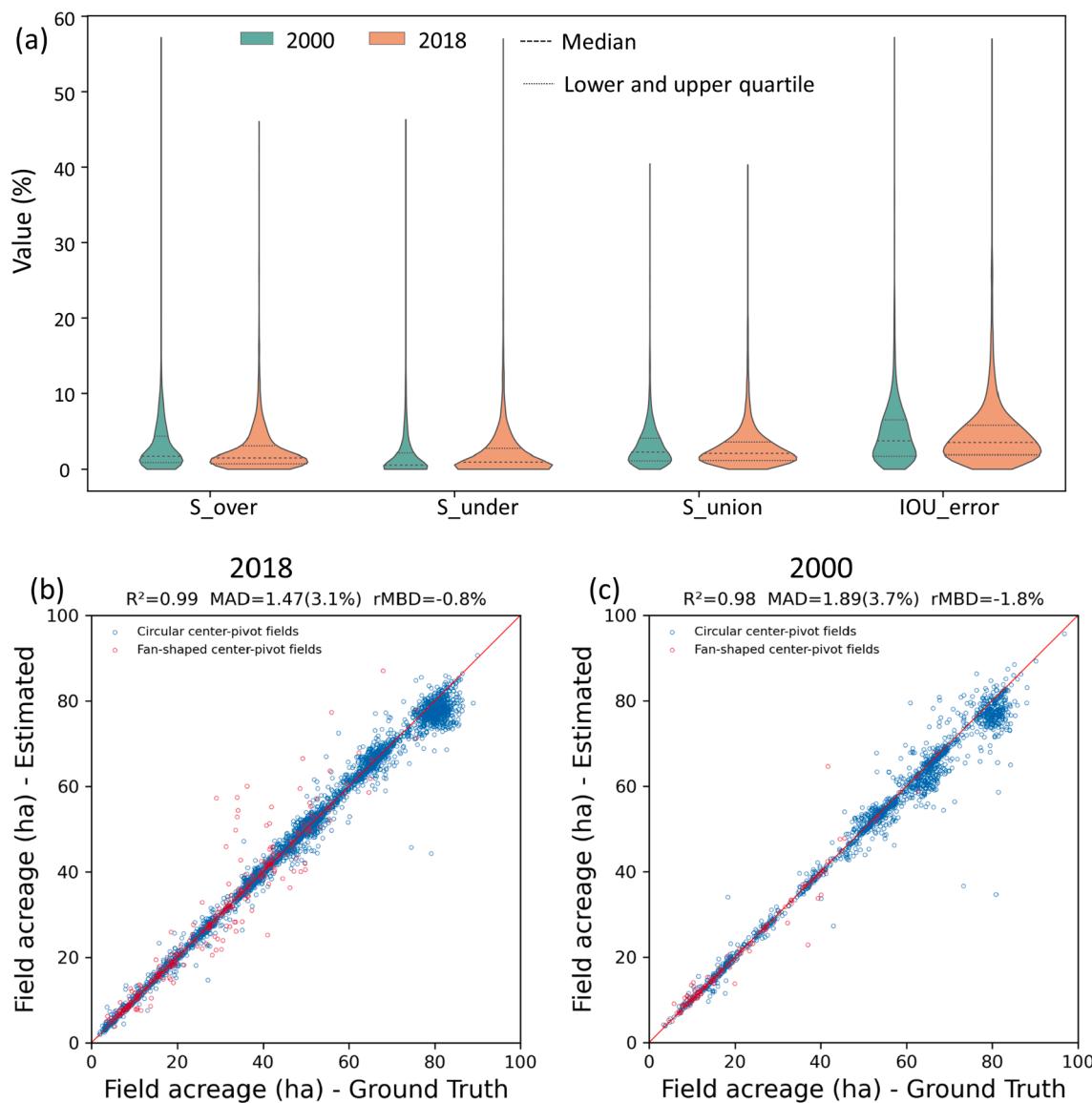
#### 4.2. Mapping agricultural field dynamics from 1988 to 2020

The framework trained using data from 2018 was applied to delineate the fields from 1988 to 2020 using Landsat data (Table 1). There were 45 CPFs detected in 1988, accounting for approximately 20 km<sup>2</sup> of irrigated acreage. A sharp increase was observed in 1993 when 410 (230 km<sup>2</sup>) additional CPFs were detected compared to 1992 (Fig. 7a and d). The sharpest increase of CPFs occurred in 2002 with approximately 660 new CPFs detected compared to 2001, resulting in an additional 370 km<sup>2</sup> of irrigated center-pivot systems (Fig. 7d). A significant decrease of CPFs was observed in 2009 (Fig. 7a and d) with the removal of 517 (374 km<sup>2</sup>) CPFs compared to 2008. Although around 310 CPFs were newly created in 2009 compared to 2008, this was the first time since 1988 that the overall acreage of CPFs decreased by any significant amount (a reduction of approximately 195 km<sup>2</sup>). However, from 2010 to 2016 the area of CPFs increased every year, reaching its largest extent in 2016, with around 5,080 CPFs covering 2,368 km<sup>2</sup>. Since 2016, the number of CPFs started to decrease, with the largest observed reduction occurring in 2020. By 2020, the number of CPFs had dropped to around 3,700, covering an area of 1,581 km<sup>2</sup> (Fig. 7a).

The median values of CPF acreage ranged from 50 to 60 ha (Fig. 7c), which was consistent with the length of the rotating center-pivot systems covering 800 m in diameter. The width of the violins in Fig. 7c indicated the count of the fields in each acreage range. The result

suggested (1) the dominant CPFs had an acreage ranging from 40 to 60 ha before 2000 and after 2009; (2) from 2000, small CPFs of less than 30 ha started to appear; and (3) from 2001 to 2008, the dominant CPF acreage ranged from 60 to 80 ha. Herein, if more than 60% of pixels forming a CPF observed in one year were also mapped as active in the year after, the two CPFs were defined as a duplicated field. The acreage bias of duplicated fields that were detected in two consecutive years (Fig. 7e) suggested that until 2010, most of the duplicated CPFs increased in size compared to the previous year. Since 2011, the duplicated fields were generally shrinking, especially for duplicated CPFs detected in 2019 and 2020, when their field diameter was reduced to cover an acreage of about 10% less.

Non-CPFs, representing a tree crop plantation or other non-woody plant crops, were mainly present in the northwestern region of the study site, with the area increasing from 1988 to 1992 from around 68 km<sup>2</sup> to 145 km<sup>2</sup> (Fig. 7b). In 1993, the non-woody non-CPFs started to decrease. On the other hand, a significant increase in tree crop plantations was observed from 1994 (Fig. 7b). The tree crop plantation acreage displayed an increasing trend until 2020. The minor fluctuations of the tree crop plantation acreage before 2010 are likely caused by errors introduced when identifying the non-CPF pixels, especially when the trees were young with tree crop pixels appearing with low NDVI values, which may be the discrimination from background pixels challenging. However, such errors will decrease as the trees grow. The largest acreage



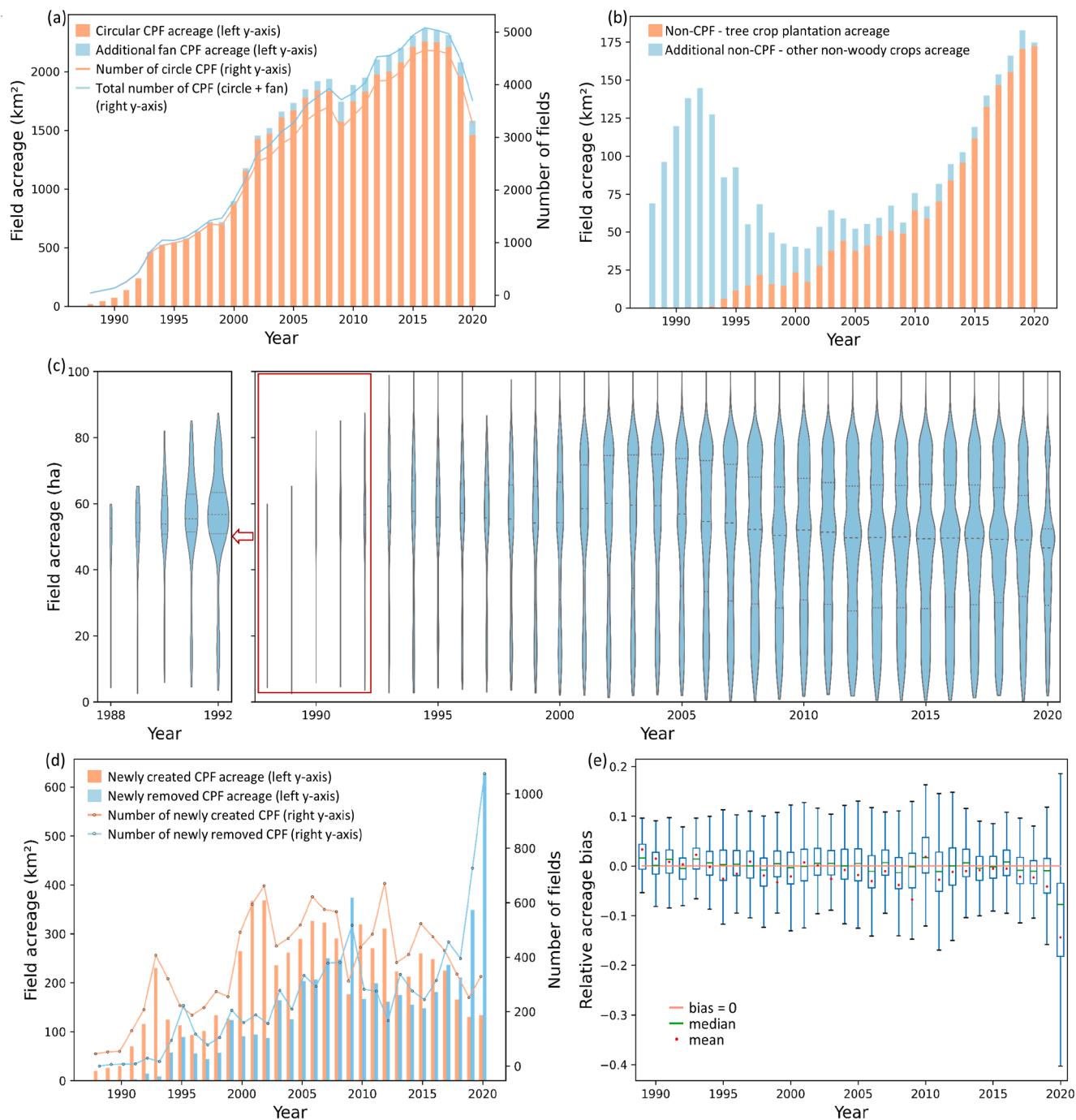
**Fig. 6.** Delineation accuracy measures for 2018 and 2000 showing: (a) a violin plot of  $S_{\text{over}}$ ,  $S_{\text{under}}$ ,  $S_{\text{union}}$  and  $\text{IoU}_{\text{error}}$  values; and (b) the scatterplots of the field acreage obtained from the ground truth and framework-estimated maps ( $n = 3,994$  and  $1,773$ ) for the years 2018 and 2000, respectively. The scatterplots were based on the center-pivot fields (CPFs) from the ground truth maps, but excluding the CPFs that were not detected by the framework. In (a),  $S_{\text{over}}$  and  $S_{\text{under}}$  indicate the over-segmentation and under-segmentation errors.  $S_{\text{union}}$  indicates an index computed from the root mean square of  $S_{\text{under}}$  and  $S_{\text{over}}$ .  $\text{IoU}_{\text{error}}$  indicates the intersection over union error. In (b) and (c),  $R^2$  is the coefficient of determination.  $\text{MAD}$  indicates the mean absolute bias deviation, with relative  $\text{MAD}$  subsequently shown in the parentheses.  $rMBD$  indicates the relative mean bias deviation. The red lines indicate  $x = y$ .

of tree crop plantations was observed in 2020 and covered an area of approximately  $172 \text{ km}^2$ .

To better understand the field dynamics temporally and spatially, the field expansion and contraction maps were produced by the earliest detection year and latest detection year (before 2020) of pixels forming a field. The highest expansion rate of fields was observed for the periods between 1990 and 1995 and 2000–2005, resulting in approximately  $708 \text{ km}^2$  and  $1,070 \text{ km}^2$  of added acreage (Fig. 8). CPFs contributed to the majority of the expanding acreage, occupying 82.8% and 95.4% for the two periods, respectively. The increase of CPFs resulted in a higher density of CPFs within the central and southern parts of the region (green and pink color in Fig. 8). The expansion of non-CPFs was primarily observed in the northwestern region before 1995. After 2005, additional CPFs mainly occurred in the northeastern area of Al Jawf (Fig. 8). Contraction of the field acreage occurred mainly after 2015, resulting in a reduction of approximately  $1,181 \text{ km}^2$  of acreage by the end of 2020 (Fig. 9), with the removal of CPFs in 2020 contributing

51.6%. The field contraction was mainly observed in the central and northern parts of the study site, where the fields were primarily created before 2010 (Figs. 8 and 9). A few fields created after 2010 in the northeastern region were removed by 2020 (Figs. 8 and 9). The hollow circles in Fig. 9 represent the footprint of shrunk fields being observed after 2010, where the length of the center-pivot sprinkling system was shortened, causing the field diameter to be reduced. These pixels were the main reason behind the negative bias of duplicated field acreage after 2010 (Fig. 7e).

A map was produced to understand the number of years within the time series that a field was active (Fig. 10). A CPF was defined as active within a year if the median of the annual maximum NDVI pixel values forming the field exceeded 0.35. A non-CPF was defined as active as long as it was detected. Then the sum of years that a pixel belonged to an active field was calculated as the active index map. As can be seen, most fields were active for a period of 25–30 years (approximate  $1,385 \text{ km}^2$ ), followed by 15–20 years (approximate  $1,180 \text{ km}^2$ ). Many of the CPFs in

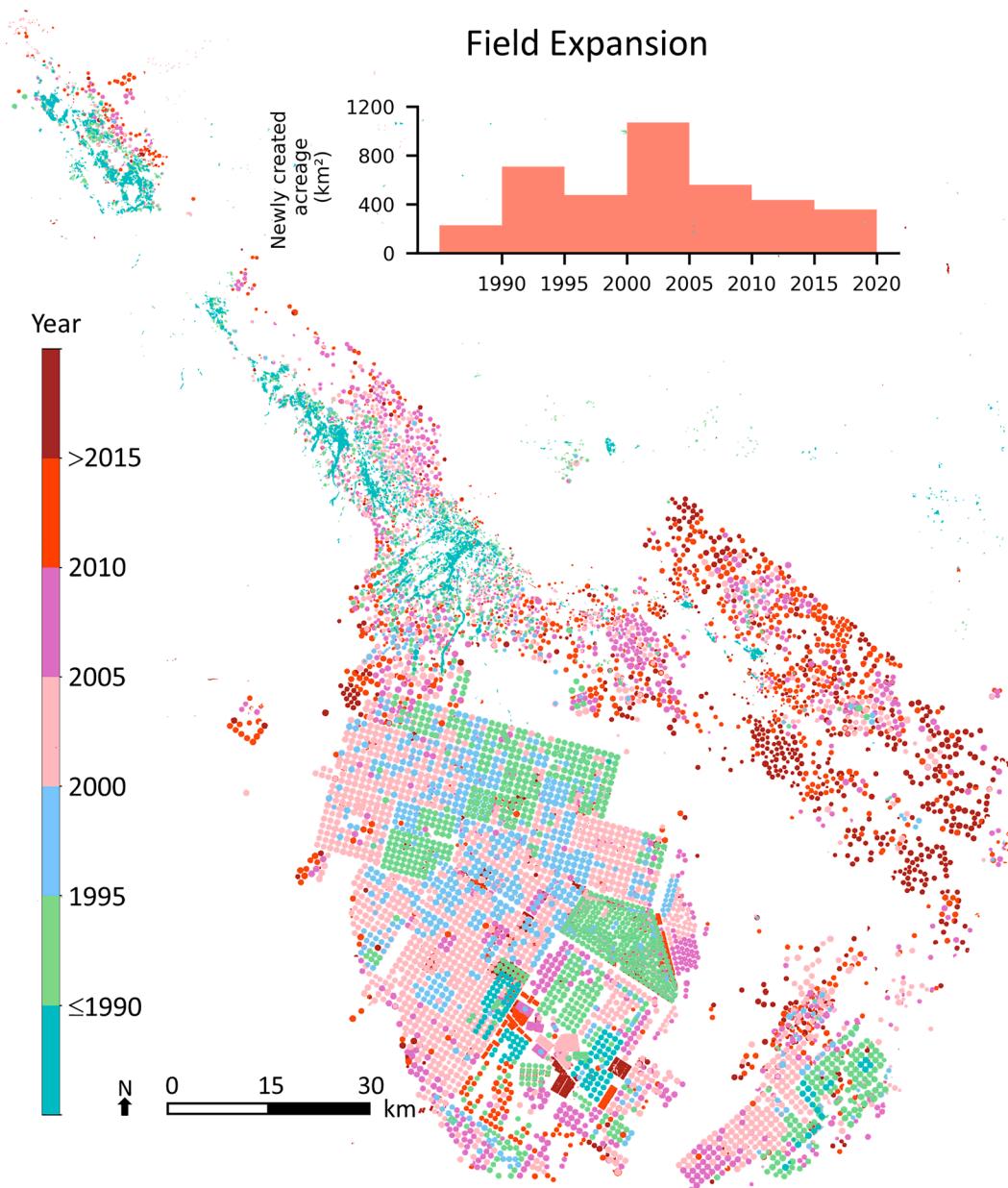


**Fig. 7.** Agricultural field dynamics of the Al Jawf region from 1988 to 2020 showing: (a) the acreage (left y-axis) and number (right y-axis) of center-pivot fields (CPFs), with blue bars indicating additional fan-shaped CPF; (b) the acreage of non-CPFs, i.e. tree crop plantations and other non-woody crops classified by the random forest model; (c) violin plots (with quartiles as dashed lines) showing the distribution of CPF acreage for each year; (d) the newly added/removed acreage (left y-axis) and number (right y-axis) of CPFs compared to the previous year; and (e) the relative bias of field acreage of the duplicated fields being detected for two consecutive years, which was defined as  $bias = (A_{af} - A_{bf})/A_{af}$ , where  $af$  takes its value from [1989, 1990, 1991, ..., 2020],  $bf = af - 1$ , and  $A_{af}$  and  $A_{bf}$  indicated the acreage of a duplicated CPF observed in year  $bf$  and  $af$ . A field was defined as a “duplicated field” if the intersection of the field acreage in consecutive two years ( $af$  and  $bf$ ) was  $\geq 0.6 A_{af}$  or  $0.6 A_{bf}$ .

the central to the southern part of Al Jawf were active for more than 15 consecutive years. The CPFs in the northern region were generally mapped as being active for less than 15 years, as they were created after 2005 (Fig. 8). The non-CPFs occupied by non-woody crops were active for less than 15 years, which was caused by the elimination of non-woody crop fields that first appeared from 1993 to the early 2000s (Fig. 7b).

## 5. Discussion

The delineation of individual field maps is fundamental to a variety of studies, including dynamic field modeling (e.g., crop type classification and crop yield forecasting) (Cai et al. 2018), evapotranspiration modeling (Aragon et al. 2018), water resource management (López Valencia et al. 2020), as well as in driving and delivering policy and management efforts. The results presented herein advance upon a

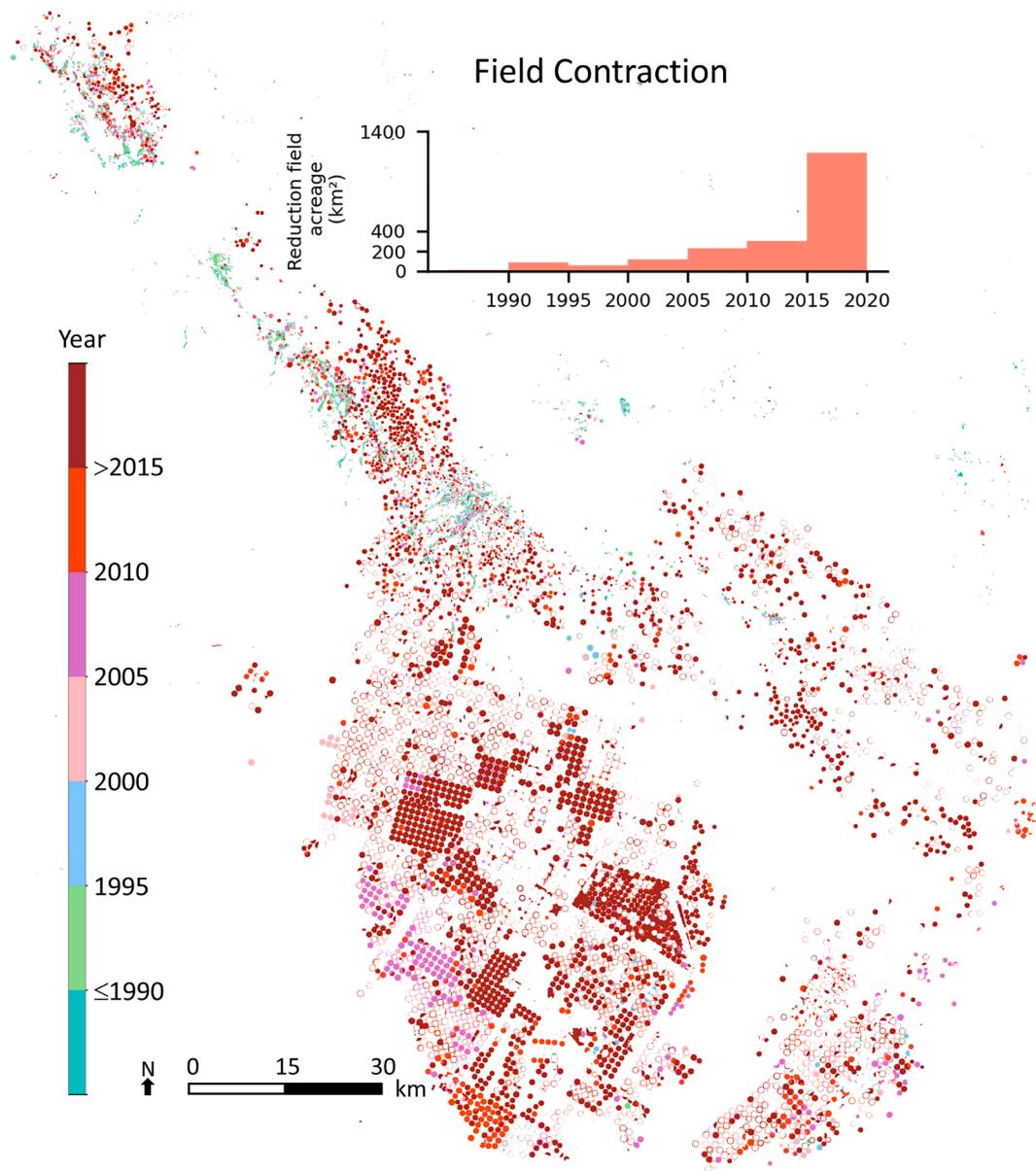


**Fig. 8.** Agricultural field expansion map. The map was generated by identifying the year that each pixel within an agricultural field was first detected as being active by the machine learning framework. The bar plot at the top represents the field acreage, calculated from the point of the first detection of each field in successive 5-year periods.

number of earlier machine-learning applications in terms of delineating individual CPFs and identifying the field shape, i.e., circular, fan shape or other, providing additional insights into field dynamics and characterization. While it is not always meaningful to intercompare results from studies that have used different data or are based in different locations, a broad comparison can provide some insight into relative performance. In this case, our method showed an advantage in delineating individual fields relative to studies by Zhang et al. (2018) and Tang et al. (2021b), which identified the location of CPFs without segmenting the extent of individual CPFs, or studies by Graf et al. (2020), Saraiwa et al. (2020), and de Albuquerque et al. (2020), which merged all CPFs as a single layer map. Carvalho et al. (2021) proposed an innovative method using Mask-RCNN to delineate CPFs for three regions located in the Cerrado biome, Central Brazil, using seven spectral bands of Landsat-8 images. The study regions consisted of around 3,730 CPFs in 2016. Ground truth data were created based on surveyed vector polygons masking individual CPFs. Their method was trained using

4,762 objects and evaluated against 650 objects. In our study, the framework was trained using 4,287 objects and evaluated against 5,785 CPFs collected across different years. We derived our collected training data by labeling the shape of the field candidates and identifying if it consisted of multiple fields, which was an easy and fast process. The studies by de Albuquerque et al. (2021a) and de Albuquerque et al. (2021b), which mapped individual CPFs using Mask-RCNN across the same region in the study conducted by Carvalho et al. (2021), showed similar limitations in terms of requiring CPF polygons in the training stage, which was time-consuming to collect for regions where such data were not available.

The method proposed by Tang et al. (2021a) is perhaps the most similar to our approach. Three modules were implemented in a stepwise manner, including PVANET, GoogleLeNet, and the Hough transform. PVANET was employed first to detect CPF candidates. In this stage, the CPFs were annotated by rectangular boxes in the training dataset. GoogleLeNet was then applied to identify the false detections by



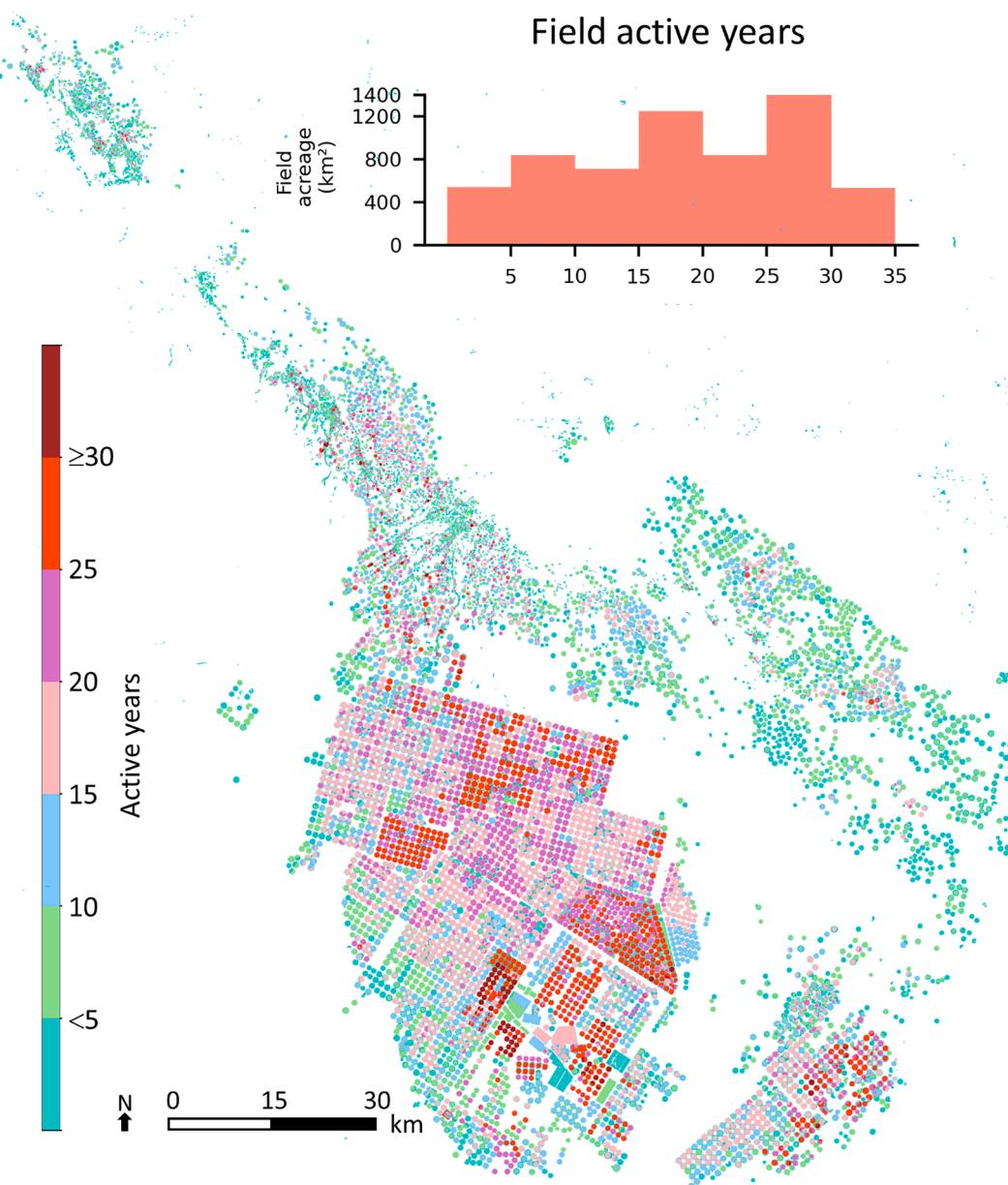
**Fig. 9.** Agricultural field contraction map. The map was generated by identifying the year that each pixel within an agricultural field was last detected as being active before 2020 by the machine learning framework. The bar plot at the top represents the field acreage of removed fields for different 5-year periods.

PVANET, with the images labeled to present whether they consisted of CPFs or not. The Hough transform was subsequently applied to delineate the shape of CPFs by identifying the center point and radius of a CPF. The framework was applied to map CPFs in Mato Grosso in Brazil using a true color image of Sentinel-2 and achieved 95% and 95.5% user's and producer's accuracies, respectively. Our method performed slightly better than this framework, i.e., 97.4% and 98.8% for user's and producer's accuracies, respectively, and managed to delineate small CPFs with radii shorter than 200 m and fan-shaped fields. Overall, compared to previous related studies, the proposed method herein showed advantages of delineating individual CPFs. For instance, our method produced high mapping accuracies when evaluated against large datasets across different years: even when being trained with limited data. In addition, observations in the form of field shape and whether an object consisted of multiple fields enabled easy and quick generation of training data. Importantly, our study presents the first attempt to discriminate between complete CPFs (circular CPFs), incomplete CPFs (fan-shaped CPFs), and non-CPFs representing rectangular shapes (tree plantation). However, as already noted, it is challenging to intercompare

distinct approaches performed in different locations and under different local conditions. Future efforts may focus on performing a thorough intercomparison of machine learning approaches for mapping agricultural fields in a range of landscapes. Here, we also present the first analysis of center-pivot dynamics extending over a period of three decades across a key agricultural region of Saudi Arabia. Future work will also explore the implementation of the framework to the entirety of Saudi Arabia to provide a deeper understanding of multi-temporal agricultural dynamics in relation to policy initiatives.

### 5.1. The benefit of a hybrid framework

In our hybrid framework, each of the four machine learning algorithms performed its own function. DBSCAN clustering was primarily used for segmenting those fields that were geographically separated and showed good performance when dealing with this large geospatial dataset. However, it was less successful at segmenting neighboring fields with adjoining edges. In these cases, spectral clustering proved to be useful but also presented limitations in terms of the requirement for



**Fig. 10.** Map showing the number of years that a field was active between 1988 and 2020. A field was defined as active within a year if the median of the annual maximum NDVI (normalized differential vegetation index) pixel values forming the field exceeded 0.35. The bar plot at the top shows the acreage of active fields in relation to the number of years being active.

setting the number of clusters and a reduced ability to deal with this large dataset requiring significantly longer run time. It took less than a second on a desktop computer for spectral clustering to segment an input field object consisting of three or four CPFs. However, it could take several minutes if the input field object consisted of more than ten CPFs. Using field objects that were preliminarily segmented by DBSCAN as input to the spectral clustering considerably reduced the runtime. Indeed, the increased computational complexity of spectral clustering can be a key limitation for employing the technique (Miyahara et al. 2014).

The minimal value of the number of clusters is two when running spectral clustering. As such, a procedure was required to identify if a field object (obtained from DBSCAN) consisted of a single field or not. Spectral clustering should only be run to segment field objects consisting of multiple fields. In addition, the irregular shape of fields has proven to be a challenge for identifying and delineating CPFs (Saraiva et al. 2020; Zhang et al. 2018). For example, Mekhalfi et al. (2021) excluded fan-

shaped CPFs in their center-pivot detection study. CNN classification is capable of identifying the shape of the field and if a field object consists of multiple fields.

CPFs were primarily covered by non-woody crops, while the non-CPFs might consist of tree crops or non-woody crops, which could successfully be discriminated by an RF classifier. However, the accuracy of RF classification was only evaluated using data from sampling region 4 (Fig. 1). We were unable to evaluate the performance of the RF model using the same dataset as evaluating the delineation results (as shown in Fig. 4) since no in situ crop type information was provided. The dependence on ground truth data was one of the most significant limitations of applying any supervised learning approaches in this research.

## 5.2. Socio-political drivers of field dynamics

The observed increases in center-pivot fields since the 1980s (Fig. 7a and b) were largely driven by policies initiated in the early 1970s to

boost agricultural production and achieve a measure of food self-sufficiency (Frenken 2009; Ouda 2014). Likewise, the significant field acreage reduction observed in 2009 (Fig. 7a and d) was a consequence of policy initiatives in 2008, which banned the planting of crops such as wheat and provided incentives to switch to the cultivation of high-value crops such as fruit (Frenken 2009; Ouda 2014). The most recent decrease in field acreage that was observed after 2018 was a manifestation of new regulations in Saudi Arabia to reduce water use in the agricultural sector due to the current levels of unsustainable use of non-renewable water resources (Belgiu and Csillik 2018). Based on our work, it is predictable that the crop water consumption in Al Jawf since 2018 would have significantly decreased due to the removal of CPFs. Although the acreage of tree plantations keeps increasing, there is a significant net reduction in agricultural acreage. Our results may support decision-makers in determining if implemented policy initiatives have the intended outcome.

### 5.3. Potential improvements to the framework

Although our results achieved high mapping accuracies, there may be potential to further improve the processing framework to be adapted for other regions. Generally, for machine learning approaches to make a sustained impact on delivering new and insightful agricultural informatics, the transferability of developed approaches needs to be thoroughly explored. For instance, the soil background can make it difficult to mask tree crop plantation pixels (Fig. 5h), especially when the tree crops are young. Alternative indices considering the soil spectra such as the Soil Adjusted Vegetation Index may offer a potential solution (Belgiu and Csillik 2018; Napoli et al. 2018) and should be explored in future work. The framework parameters were initialized based on the dominant size of the center-pivot fields (800 m in diameter), but this might introduce problems when identifying small fields, such as those in the northwestern region of the study site (Fig. 5a and b). A self-learning procedure, such as gradient descent (Waldner and Diakogiannis 2020), may be useful to update the parameters automatically during the learning process. It is also feasible to divide the large satellite image scene into several subsets and adjust the framework parameters based on the field condition that best fit each of the subsets, i.e. dominant shape and size of fields. This may shorten the runtime of the framework, especially when applying the framework to finer scale satellite data such as Sentinel-2 or PlanetScope, as the processing can be employed on independent subsets and in parallel by a machine. A reduction of ‘over-segmentation errors caused by spectral clustering is an area where further research is also justified (Nadler and Galun 2007), e.g. by training a CNN model to count the number of fields in a given image.

## 6. Conclusions

A three-decades long spatial and temporal analysis of center-pivot fields and tree plantations was performed across a major agricultural region of Saudi Arabia. Based on a unique processing workflow that integrated four machine learning algorithms and exploited a 33-year time-series of Landsat imagery, the approach was used to characterize agricultural field delineation and their multi-temporal dynamics. Results established the high accuracy of the framework for identifying a variety of agricultural field shapes, i.e., circular and fan shape (or neither) and their size, as well as for delineating individual center-pivot fields. Using imagery from 2018, individual center-pivot fields were mapped and their shapes classified, resulting in a 97.4% producer's and 98.0% user's accuracy on an object basis. The over-segmentation and under-segmentation errors were less than 1.5% and the coefficient of determination for the field acreage between the delineated center-pivot fields and ground truth data was 0.99. In terms of identifying non-center-pivot fields, the method achieved 81.4% producer's and 85.4% user's accuracy on a pixel basis. A subsequent model trained on 2018 data showed high stability when applied to the detection and

delineation of fields for the year 2000, allowing subsequent application of this approach across the time series of satellite data. Through this analysis, new insights into the number and acreage of fields, along with their increase or decrease on a yearly basis, the expansion and contraction of individual fields through time, and the number of years an individual field was active, were obtained. These results provide additional insight into important agricultural informatics that may assist related studies such as crop type discrimination, yield estimation, water consumption, and other food and water security-related concerns. The machine learning algorithms used in this framework provided a largely automated, computationally efficient, and accurate solution to field identification and delineation. Field dynamics obtained from the 33-year Landsat time-series provided a temporal exploration of agricultural development in the Al Jawf region, and clearly reflected how various policy initiatives affected field management and production practices. Future work should focus on exploring potential improvements to the accuracy of the framework. These may be achieved by: (1) exploring additional spectral bands and vegetation indices as input; (2) developing a self-learning procedure to find the optimal model parameters suited for CPFs with various diameters; and (3) adapting the framework to identify field types in environments that differ from those studied herein. By developing and applying this novel temporal analysis on high-resolution satellite data, an improved understanding of the ability of machine learning methods in detecting fields with different shapes and delineating the extent of the fields has been gained.

### CRediT authorship contribution statement

**Ting Li:** Conceptualization, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. **Kasper Johansen:** Methodology, Validation, Writing – review & editing. **Matthew F. McCabe:** Conceptualization, Supervision, Writing – review & editing.

### Declaration of Competing Interest

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

### Acknowledgments

Research reported in this publication was supported by the King Abdullah University of Science and Technology (KAUST).

### Appendix A. Supplementary material

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.isprsjprs.2022.02.002>.

### References

- Al-Rumkhani, Y.A., Din, S.U., 2004. Use of remote sensing for irrigation scheduling in arid lands of Saudi Arabia. *J. Indian Soc. Remote Sens.* 32 (2), 225–233.
- Aragon, B., Houborg, R., Tu, K., Fisher, J.B., McCabe, M., 2018. CubeSats enable high spatiotemporal retrievals of crop-water use for precision agriculture. *Remote Sensing* 10 (12), 1867. <https://doi.org/10.3390/rs10121867>.
- Baatz, M., 2000. Multi resolution Segmentation: an optimum approach for high quality multi scale image segmentation. In: Beitrage zum AGIT-Symposium. Salzburg, Heidelberg, 2000, pp. 12–23.
- Bauder, T.A., 2004. Center pivot irrigation in Colorado as mapped by landsat imagery. Technical bul.(Colorado Agricultural Experiment Station); TB-04-4.
- Belgiu, M., Csillik, O., 2018. Sentinel-2 cropland mapping using pixel-based and object-based time-weighted dynamic time warping analysis. *Remote Sens. Environ.* 204, 509–523.
- Belgiu, M., Drăgut, L., 2016. Random forest in remote sensing: a review of applications and future directions. *ISPRS J. Photogramm. Remote Sens.* 114, 24–31.
- Bleau, A., Leon, L.J., 2000. Watershed-based segmentation and region merging. *Comput. Vis. Image Underst.* 77 (3), 317–370.
- Breiman, L., 2001. Random forests. *Mach. Learn.* 45, 5–32.

- Cai, Y., Guan, K., Peng, J., Wang, S., Seifert, C., Wardlow, B., Li, Z., 2018. A high-performance and in-season classification system of field-level crop types using time-series Landsat data and a machine learning approach. *Remote Sens. Environ.* 210, 35–47.
- Calinski, T., Harabasz, J., 1974. A dendrite method for cluster analysis. *Commun. Stat.-Theory Methods* 3 (1), 1–27.
- Campello, R.J., Moulavi, D., Sander, J., 2013. Density-based clustering based on hierarchical density estimates. In: *Pacific-Asia Conference On Knowledge Discovery And Data Mining*. Springer, pp. 160–172.
- Canny, J., 1986. A computational approach to edge detection. *IEEE Trans. Pattern Anal. Machine Intell.* 679–698.
- Carvalho, O.L.F.d., de Carvalho Júnior, O.A., Albuquerque, A.O.d., Bem, P.P.d., Silva, C.R., Ferreira, P.H.G., Moura, R.D.S.d., Gomes, R.A.T., Guimarães, R.F., Borges, D.L., 2021. Instance segmentation for large, multi-channel remote sensing imagery using Mask-RCNN and a Mosaicking approach. *Remote Sensing* 13 (1), 39. <https://doi.org/10.3390/rs13010039>.
- Clinton, N., Holt, A., Scarborough, J., Yan, L.i., Gong, P., 2010. Accuracy assessment measures for object-based image segmentation goodness. *Photogramm. Eng. Remote Sens* 76 (3), 289–299.
- de Albuquerque, A.O., de Carvalho Júnior, O.A., Carvalho, O.L.F.d., de Bem, P.P., Ferreira, P.H.G., de Moura, R.D.S., Silva, C.R., Trancoso Gomes, R.A., Fontes Guimarães, R., 2020. Deep semantic segmentation of center pivot irrigation systems from remotely sensed data. *Remote Sensing* 12 (13), 2159. <https://doi.org/10.3390/rs12132159>.
- de Albuquerque, A.O., de Carvalho, O.L.F., e Silva, C.R., de Bem, P.P., Trancoso Gomes, R.A., Borges, D.L., Guimarães, R.F., Pimentel, C.M.M., de Carvalho Júnior, O.A., 2021a. Instance segmentation of center pivot irrigation systems using multi-temporal SENTINEL-1 SAR images. *Remote Sens. Appl.: Soc. Environ.* 23, 100537. <https://doi.org/10.1016/j.rse.2021.100537>.
- de Albuquerque, A.O., de Carvalho, O.L.F., Silva, C.R.E., Luiz, A.S., de Bem, P.P., Gomes, R.A.T., Guimaraes, R.F., Junior, O.A.d.C., 2021b. Dealing with clouds and seasonal changes for center pivot irrigation systems detection using instance segmentation in sentinel-2 time series. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 14, 8447–8457.
- Dhillon, I.S., Guan, Y., Kulis, B., 2004. Kernel k-means, spectral clustering and normalized cuts. In: *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 551–556.
- Duro, D.C., Franklin, S.E., Dubé, M.G., 2012. A comparison of pixel-based and object-based image analysis with selected machine learning algorithms for the classification of agricultural landscapes using SPOT-5 HRG imagery. *Remote Sens. Environ.* 118, 259–272.
- Ester, M., Kriegel, H.-P., Sander, J., Xu, X., 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In: *Kdd*, pp. 226–231.
- Famiglietti, J.S., 2014. The global groundwater crisis. *Nat. Clim. Change* 4 (11), 945–948.
- Fao, 2013. Global map of irrigation areas - Saudi Arabia. Food and Agriculture Organization of the United Nations, Rome.
- Ferrara, R., Virdis, S.G.P., Ventura, A., Ghisu, T., Duce, P., Pellizzaro, G., 2018. An automated approach for wood-leaf separation from terrestrial LIDAR point clouds using the density based clustering algorithm DBSCAN. *Agric. For. Meteorol.* 262, 434–444.
- Ferreira, E., Toledo, J.H.d., Dantas, A.A.A., Pereira, R.M., 2011. Cadastral maps of irrigated areas by center pivots in the State of Minas Gerais, using CBERS-2B/CCD satellite imaging. *Engenharia Agrícola* 31 (4), 771–780.
- Frenken, K., 2009. Irrigation in the Middle East region in figures AQUASTAT Survey-2008. Water Reports.
- Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., Martinez-Gonzalez, P., Garcia-Rodriguez, J., 2018. A survey on deep learning techniques for image and video semantic segmentation. *Appl. Soft Comput.* 70, 41–65.
- Goodfellow, I., Bengio, Y., Courville, A., 2016. Deep Learning. MIT Press.
- Graesser, J., Ramankutty, N., 2017. Detection of cropland field parcels from Landsat imagery. *Remote Sens. Environ.* 201, 165–180.
- Graf, L., Bach, H., Tiede, D., 2020. Semantic segmentation of sentinel-2 imagery for mapping irrigation center pivots. *Remote Sensing* 12 (23), 3937. <https://doi.org/10.3390/rs12233937>.
- GASTAT, 2018. Agricultural Area Cultivated and Production by Type of Crop, Statistical Yearbook of 2018.
- He, K., Zhang, X., Ren, S., Sun, J., 2015. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In: *Proceedings of the IEEE international conference on computer vision*, pp. 1026–1034.
- Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.R., 2012. Improving neural networks by preventing co-adaptation of feature detectors. arXiv preprint arXiv:1207.0580.
- Hopfield, J.J., 1982. Neural networks and physical systems with emergent collective computational abilities. *Proc. Natl. Acad. Sci.* 79 (8), 2554–2558.
- Huang, L., Luo, J., Lin, Z., Niu, F., Liu, L., 2020. Using deep learning to map retrogressive thaw slumps in the Beiluhe region (Tibetan Plateau) from CubeSat images. *Remote Sens. Environ.* 237, 111534. <https://doi.org/10.1016/j.rse.2019.111534>.
- Johansen, K., Lopez, O., Tu, Y.-H., Li, T., McCabe, M.F., 2021. Center pivot field delineation and mapping: A satellite-driven object-based image analysis approach for national scale accounting. *ISPRS J. Photogramm. Remote Sens.* 175, 1–19.
- Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*, pp. 1097–1105.
- Lebourgeois, V., Dupuy, S., Vintrou, É., Ameline, M., Butler, S., Bégué, A., 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>.
- LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard, R.E., Hubbard, W., Jackel, L.D., 1989. Backpropagation applied to handwritten zip code recognition. *Neural Comput.* 1 (4), 541–551.
- Li, Q., Wang, C., Zhang, B., Lu, L., 2015. Object-based crop classification with Landsat-MODIS enhanced time-series data. *Remote Sensing* 7 (12), 16091–16107.
- Litts, T., Russell, H., Thomas, A., Welch, R., 2001. Mapping irrigated lands in the ACF river basin. In: Georgia Institute of Technology.
- Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–3440.
- López Valencia, O.M., Johansen, K., Aragón Solorio, B.J.L., Li, T., Houborg, R., Malbeteau, Y., AlMashharawi, S., Altaf, M.U., Fallatah, E.M., Dasari, H.P., Hoteit, I., McCabe, M.F., 2020. Mapping groundwater abstractions from irrigated agriculture: big data, inverse modeling, and a satellite-model fusion approach. *Hydro. Earth Syst. Sci.* 24 (11), 5251–5277.
- Masoud, K.M., Persello, C., Tolpekin, V.A., 2020. Delineation of agricultural field boundaries from Sentinel-2 images using a novel super-resolution contour detector based on fully convolutional networks. *Remote Sensing* 12 (1), 59. <https://doi.org/10.3390/rs12100059>.
- McInnes, L., Healy, J., Astels, S., 2017. hdbscan: Hierarchical density based clustering. *J. Open Source Softw.* 2 (11), 205. <https://doi.org/10.21105/joss10.21105/joss.00205>.
- Mekhalfi, M.L., Nicolò, C., Bazi, Y., Al Rahhal, M.M., Al Maghayreh, E., 2021. Detecting crop circles in google earth images with mask R-CNN and YOLOv3. *Appl. Sci.* 11 (5), 2238. <https://doi.org/10.3390/app11052238>.
- Melton, F.S., Johnson, L.F., Lund, C.P., Pierce, L.L., Michaelis, A.R., Hiatt, S.H., Guzman, A., Adhikari, D.D., Purdy, A.J., Roosevelt, C., Votava, P., Trout, T.J., Temesgen, B., Frame, K., Sheffner, E.J., Nemani, R.R., 2012. Satellite irrigation management support with the terrestrial observation and prediction system: A framework for integration of satellite and surface observations to support improvements in agricultural water resource management. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* 5 (6), 1709–1721.
- Mishra, P., 2019. Introduction to neural networks using PyTorch. In: *PyTorch Recipes*. Springer, pp. 111–126.
- Miyahara, S., Komazaki, Y., Miyamoto, S., 2014. An algorithm combining spectral clustering and DBSCAN for core points. *Knowledge and Systems Engineering*. Springer, pp. 21–28.
- Mueller, M., Segl, K., Kaufmann, H., 2004. Edge-and region-based segmentation technique for the extraction of large, man-made objects in high-resolution satellite imagery. *Pattern Recogn.* 37 (8), 1619–1628.
- Nadler, B., Galun, M., 2007. Fundamental limitations of spectral clustering. In: *Advances in neural information processing systems*, pp. 1017–1024.
- Napoli, C., Wise, B., Wogan, D., Yaseen, L., 2018. Policy options for reducing water for agriculture in Saudi Arabia. In: *Assessing Global Water Megatrends*. Springer, pp. 211–230.
- Ng, A.Y., Jordan, M.I., Weiss, Y., 2002. On spectral clustering: Analysis and an algorithm. In: *Advances in neural information processing systems*, pp. 849–856.
- Olson, D.L., Delen, D., 2008. Performance evaluation for predictive modeling. In: *Advanced Data Mining Techniques*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 137–147. [https://doi.org/10.1007/978-3-540-76917-0\\_9](https://doi.org/10.1007/978-3-540-76917-0_9).
- Ouda, O.K.M., 2014. Impacts of agricultural policy on irrigation water demand: a case study of Saudi Arabia. *Int. J. Water Resour. Dev.* 30 (2), 282–292.
- Pastor, A.V., Palazzo, A., Havlik, P., Biemans, H., Wada, Y., Obersteiner, M., Kabat, P., Ludwig, F., 2019. The global nexus of food–trade–water sustaining environmental flows by 2050. *Nat. Sustainability* 2 (6), 499–507.
- Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., 2011. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* 12, 2825–2830.
- Peña, J., Gutiérrez, P., Hernández-Martínez, C., Six, J., Plant, R., López-Granados, F., 2014. Object-based image classification of summer crops with machine learning methods. *Remote Sensing* 6 (6), 5019–5041.
- Persello, C., Tolpekin, V.A., Bergado, J.R., de By, R.A., 2019. Delineation of agricultural fields in smallholder farms from satellite images using fully convolutional networks and combinatorial grouping. *Remote Sens. Environ.* 231, 111253. <https://doi.org/10.1016/j.rse.2019.111253>.
- Phalke, A.R., Özdogan, M., Thenkabail, P.S., Erickson, T., Gorelick, N., Yadav, K., Congalton, R.G., 2020. Mapping croplands of Europe, middle east, russia, and central asia using landsat, random forest, and google earth engine. *ISPRS J. Photogramm. Remote Sens.* 167, 104–122.
- Radiuk, P.M., 2017. Impact of training set batch size on the performance of convolutional neural networks for diverse datasets. *Inf. Technol. Manage. Sci.* 20, 20–24.
- Raileanu, L.E., Stoffel, K., 2004. Theoretical comparison between the gini index and information gain criteria. *Ann. Math. Artif. Intell.* 41 (1), 77–93.
- Rodríguez-Galiano, V.F., Chica-Olmo, M., Abarca-Hernandez, F., Atkinson, P.M., Jeganathan, C., 2012a. Random Forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture. *Remote Sens. Environ.* 121, 93–107.
- Rodríguez-Galiano, V.F., Ghimire, B., Rogan, J., Chica-Olmo, M., Rigol-Sánchez, J.P., 2012b. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS J. Photogramm. Remote Sens.* 67, 93–104.

- Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. Springer, pp. 234–241.
- Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *J. Comput. Appl. Math.* 20, 53–65.
- Ruder, S., 2016. An overview of gradient descent optimization algorithms. arXiv preprint arXiv:1609.04747.
- Rundquist, D.C., Hoffman, R.O., Carlson, M.P., Cook, A.E., 1989. Nebraska center-pivot inventory: an example of operational satellite remote sensing on a long-term basis. *Photogramm. Eng. Remote Sens.* 55, 587–590.
- Rydberg, A., Borgefors, G., 2001. Integrated method for boundary delineation of agricultural fields in multispectral satellite images. *IEEE Trans. Geosci. Remote Sens.* 39 (11), 2514–2520.
- Saraiva, M., Protas, É., Salgado, M., Souza Jr, C., 2020. Automatic mapping of center pivot irrigation systems from satellite images using deep learning. *Remote Sensing* 12, 558.
- Schaffer, C., 1993. Selecting a classification method by cross-validation. *Mach. Learn.* 13 (1), 135–143.
- Scheffler, D., Hollstein, A., Diedrich, H., Segl, K., Hostert, P., 2017. AROSICS: an automated and robust open-source image co-registration software for multi-sensor satellite data. *Remote Sensing* 9 (7), 676. <https://doi.org/10.3390/rs9070676>.
- Schmidhuber, J., 2015. Deep learning in neural networks: An overview. *Neural Netw.* 61, 85–117.
- Schmidt, G., Jenkerson, C.B., Masek, J., Vermote, E., Gao, F., 2013. Landsat ecosystem disturbance adaptive processing system (LEDAPS) algorithm description. In: US Geological Survey.
- Seth, N.M., 2015. Analyzing the increase in center pivot irrigation systems in Custer County, Nebraska USA from 2003 to 2010. *Pap. Resour. Anal.* 17, 15.
- Shi, J., Malik, J., 2000. Normalized cuts and image segmentation. *Departmental Papers (CIS)* 107.
- Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.
- Song, X., Wang, J., Huang, W., Liu, L., Yan, G., Pu, R., 2009. The delineation of agricultural management zones with high resolution remotely sensed data. *Precis. Agric.* 10 (6), 471–487.
- Tan, P.-N., Steinbach, M., Kumar, V., 2006. Cluster analysis: basic concepts and algorithms. *Introd. Data Min.* 8, 526–533.
- Tang, J., Arvor, D., Corpetti, T., Tang, P., 2021a. Mapping center pivot irrigation systems in the Southern Amazon from sentinel-2 images. *Water* 13 (3), 298. <https://doi.org/10.3390/w13030298>.
- Tang, J., Zhang, Z., Zhao, L., Tang, P., 2021b. Increasing shape bias to improve the precision of center pivot irrigation system detection. *Remote Sensing* 13 (4), 612. <https://doi.org/10.3390/rs13040612>.
- Tao, S., Wu, F., Guo, Q., Wang, Y., Li, W., Xue, B., Hu, X., Li, P., Tian, D.i., Li, G., Yao, H., Li, Y., Xu, G., Fang, J., 2015. Segmenting tree crowns from terrestrial and mobile LiDAR data by exploring ecological theories. *ISPRS J. Photogramm. Remote Sens.* 110, 66–76.
- Tung, F., Wong, A., Clausi, D.A., 2010. Enabling scalable spectral clustering for image segmentation. *Pattern Recogn.* 43 (12), 4069–4076.
- Turker, M., Kok, E.H., 2013. Field-based sub-boundary extraction from remote sensing imagery using perceptual grouping. *ISPRS J. Photogramm. Remote Sens.* 79, 106–121.
- Vermote, E., Justice, C., Claverie, M., Franch, B., 2016. Preliminary analysis of the performance of the Landsat 8/OLI land surface reflectance product. *Remote Sens. Environ.* 185, 46–56.
- Vieira, M.A., Formaggio, A.R., Rennó, C.D., Atzberger, C., Aguiar, D.A., Mello, M.P., 2012. Object Based Image Analysis and Data Mining applied to a remotely sensed Landsat time-series to map sugarcane over large areas. *Remote Sens. Environ.* 123, 553–562.
- von Luxburg, U., 2007. A tutorial on spectral clustering. *Stat. Comput.* 17 (4), 395–416.
- Waldner, F., Diakogiannis, F.I., 2020. Deep learning on edge: extracting field boundaries from satellite images with a convolutional neural network. *Remote Sens. Environ.* 245, 111741. <https://doi.org/10.1016/j.rse.2020.111741>.
- Watkins, B., van Niekerk, A., 2019. A comparison of object-based image analysis approaches for field boundary delineation using multi-temporal Sentinel-2 imagery. *Comput. Electron. Agric.* 158, 294–302.
- Xia, X., Liang, H., RongFeng, Y., Kun, Y., 2018. Oil tank extraction in high-resolution remote sensing images based on deep learning. In: 2018 26th International Conference on Geoinformatics. IEEE, pp. 1–6.
- Yan, L., Roy, D.P., 2016. Conterminous United States crop field size quantification from multi-temporal Landsat data. *Remote Sens. Environ.* 172, 67–86.
- Yan, L., Roy, D.P., 2014. Automated crop field extraction from multi-temporal Web Enabled Landsat Data. *Remote Sens. Environ.* 144, 42–64.
- Zhang, C., Yue, P., Di, L., Wu, Z., 2018. Automatic identification of center pivot irrigation systems from landsat images using convolutional neural networks. *Agriculture* 8 (10), 147. <https://doi.org/10.3390/agriculture8100147>.
- Zhang, X., Jiao, L., Liu, F., Bo, L., Gong, M., 2008. Spectral clustering ensemble applied to SAR image segmentation. *IEEE Trans. Geosci. Remote Sens.* 46 (7), 2126–2136.
- Zhao, W., Yang, H., Shen, Z., Luo, J., 2015. Oil tanks extraction from high resolution imagery using a directional and weighted hough voting method. *J. Indian Soc. Remote Sens.* 43 (3), 539–549.