HIGH-RESOLUTION MONTHLY CROP WATER DEMAND MAPPING

by

Alec Watkins

A Thesis

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Master of Science in Agricultural and Biological Engineering

School of Agricultural and Biological Engineering

West Lafayette, Indiana

December 2021

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Keith A. Cherkauer, Chair
School of Agricultural and Biological Engineering

Dr. Laura C. Bowling
School of Agronomy

Dr. Melba M. Crawford
School of Agronomy

Approved by:
Dr. Nathan S. Mosier

Dedicated to my wife for her support and patience.

ACKNOWLEDGMENTS

Thanks to Andre de Lima Moraes and Keith Cherkauer for their great help and mentorship with the entire project. Thanks for Dr. Bowling and Dr. Crawford for their great help. Thanks to Jinqiu Chen for her work with CROPWAT. Thanks to Jose Pinto, Alex Mamani, Alex Puma, John Huancahurie, and Mauro Uriarte for helping with ground reference data collection. Thanks to Kim I Luk for his development of the CultiVista Tool. Thanks to Katy Mazer for her help with translation of the tool. Thanks to Larry Biehl for feedback on presentations based on early iterations of this study.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AEA:	Age Estimation Algorithm
AW-SCCA:	Author-Written Spatial Consistency Correction Algorithm, in reference to the author-written sieving algorithm
CDL:	Cropland Data Layer
CI:	Confidence Interval
CMA:	Crop Mapping Algorithm
ET:	Evapotranspiration
FSA:	Farm Service Agency
GDAL:	Geospatial Data Abstraction Library
GDAL-SCCA:	GDAL Spatial Consistency Correction Algorithm, in reference to the GDAL sieving algorithm
GIS:	Geographic Information System
KNN:	K-Nearest-Neighbor classifier
MaxL:	Maximum Likelihood classifier, with multivariate normal distribution assumption
NDVI:	Normalized Difference Vegetation Index
NLCD:	National Land Cover Database
RF:	RandomForest classifier
SCCA:	Spatial Consistency Correction Algorithm
TCCA:	Temporal Consistency Correction Algorithm
UNSA:	Universidad Nacional de San Agustín de Arequipa (English: National University of Saint Augustine of Arequipa)
WDMA:	Water Demand Mapping Algorithm

ABSTRACT

The Department of Arequipa, in Peru, is a region with limited water resources making freshwater management critical and requiring the development of water-demand models, which can be valuable tools for policymakers. This study developed a monthly agricultural water-demand mapping algorithm for the agricultural districts surrounding the city of Arequipa. It was accomplished by: (1) developing a ground-reference data collection method; (2) creating a crop mapping algorithm, which incorporates supervised classification methods, as well as spatial- and temporal-consistency correction methods to create crop maps out of high resolution (~3 m) PlanetScope satellite images; (3) integrating a crop growth-stage prediction algorithm for the crop maps and; (4) applying an algorithm for the estimation of the agricultural-water-demand maps using the results of steps 2 and 3, local climate data, and an irrigation demand estimation tool. The crop mapping algorithm was shown to create maps with acceptable accuracy, with 5 out of 6 months with available data having mean monthly classification accuracies of 69% to 77% for those classes which had available data. Growth stage predictions had mean absolute prediction errors of 0.55 to 0.69 months in 5 out of 6 months. The 6th month (the first with ground reference data collection) had a mean absolute prediction error of 0.90 months because it lacked prior month information to correctly identify planting month. Water demand maps were produced with high spatial (3.0 m) and temporal (monthly) resolution, allowing for a detailed look at local agricultural water demands. This study provides a framework for future large-scale agricultural-water demand mapping for the Department of Arequipa and similar regions around the world.

1. INTRODUCTION

Freshwater management in the Department of Arequipa, Peru is critical, with limited freshwater resources from the Andean Altiplano being highly managed (Stensrud, 2016) and used for the competing needs of over 1.4 million people living in the region (INEI, 2018), a substantial mining industry, and over 700 km^2 of irrigated agriculture (Salmoral et al., 2020), mostly in a desert climate (Moraes et al., 2020). The Department of Arequipa's highly managed and complex water system operates close to its limits and has been challenged by increased climate variability (Moraes et al., 2020) and growing water demands (Salmoral et al., 2020). As in many parts of the world, agriculture represents a large share of the water demand in the region, and, thus, understanding its quantity and variability throughout the year is crucial for local stakeholders, especially water managers and policymakers. However, agriculture in this region is spatially and temporally complicated compared to many other regions of the world. Factors such as a lack of a defined growing season (the region is near the equator), small fields (often less than 0.1 ha), heterogeneity of crop fields (different crops are grown near each other and sometimes in the same fields), and highly variable land management practices all complicate analysis of the region's agriculture and thus agricultural water use.

Crop maps show where crops are being grown and what type of crops are being grown in an area. Crop maps can be used for yield estimation, crop acreage assessments, land use monitoring, and are thus important tools for farmers, land managers, and policymakers (Howard et al., 2012). Moreover, when analyzed with other information, such as weather and climate data and irrigation demand models, crop maps are an important source of information for estimating regional agricultural water demands. Water demand estimates would be valuable for the farmers and policymakers of the region of Arequipa, e.g., to help predict irrigation demand, guide planting patterns, and help manage agricultural land. For example, local stakeholders in the Department of Arequipa have expressed concern about the availability and distribution of irrigation water, as well as a desire for governmental institutions to help more with irrigation water distribution due to the lack of resources and personnel needed for traditional community-based institutions to perform this task (Popovici et al., 2021). Water demand estimates could help by providing near real-time data that would lead to better decision-making at both the local and governmental level.

Crop maps have historically been made through agricultural surveys and traditional cartographic methods but with the advent of remote sensing technology, they are now commonly made by classifying satellite images. The theory of image classification has been around for decades, and many different methods have been developed and applied (Lu and Weng, 2007; Wilkinson, 2005). Many studies on the creation of crop maps have been focused on regions with agriculture management practices that are relatively simple compared to that of Arequipa, such as the United States (Shao et al., 2010; Dahal et al., 2018; Loveland, et al., 2000, Zhong, 2012). For example, the Cropland Data Layer (CDL) project has provided annual crop maps of the contiguous United States since the year 2008 (United States Department of Agriculture, n.d.). As field sizes are typically larger in the U.S. and often planted with a single crop, it is possible for the USDA to rely on fairly coarse spatial resolution satellite images, ranging from about 30 m up to 1 km, that have spectral bands in the visible and near infrared parts of the spectrum. Having a single dominant growing season also means that the USDA can utilize multiple images in a limited time window to capture crop phenological changes.

Recent advances in remote sensing technology, in particular the advent of constellations of satellites with very high spatial resolutions (5 meters or less) that can image every part of the Earth every day, have expanded classification possibilities to include agriculturally complex regions like the Department of Arequipa. However, there is still a lack of case studies and general methods for the classification of such regions. While there exist some recent examples of methods employed to map crops in agriculturally complex regions such as parts of China and Turkey (Zhang et al., 2016; Ozdarici-Ok et al., 2015), these regions do not have all the particularities of the Arequipa region. Most notably, these regions have significant variability in weather that causes there to be defined growing seasons. Zhang et al. (2016) utilized phenological time series of 100-meter resolution PROBA-V data with a 5-day repeat cycle to perform classification of crops in two agricultural regions in China that are similar in size to the agricultural area in and around the City of Arequipa, but that have a dominant growing season and larger, more uniform fields. They found that the higher resolution PROBA-V imagery resulted in more accuracy crop maps compared to MODIS data, especially in the site with more distinct fields. Meanwhile, Ozdarici-Ok et al. (2015) demonstrated that a single very-high-resolution image with only blue, green, red, and NIR spectral bands can be sufficient for mapping up to six different crops using a field site in northwest Turkey. Accuracy was best if the image was taken after the crops had reached an appropriate growth stage,

indicating that the timing of imagery relative to crop growth stage is critical for classification. Given the results of these and other studies, it is hypothesized that it will be possible to map crops in the Department of Arequipa with sufficient accuracy, although it is expected that there will be additional complications in this process due to the specific complexities in agricultural management local to the region.

The goal of this study is to develop a system for creating monthly maps of crop type, crop age and crop water demand for the Department of Arequipa, Peru, with the intent of later turning this system into an operational tool to be used by local stakeholders. This goal can be accomplished in three major steps, which are represented by the following objectives.

1. Develop a crop mapping algorithm (CMA), which uses supervised classification methods, along with high-resolution satellite images and ground reference data, to create monthly crop maps of the Department of Arequipa with acceptable accuracy, despite its agricultural complexities.
2. Develop an age estimation algorithm (AEA), which estimates the ages of crops with acceptable accuracy based on crop progression derived from the crop maps.
3. Create a water demand mapping algorithm (WDMA) for the agriculture in the region of Arequipa by using the data from steps 1 and 2, auxiliary climate data, and an appropriate irrigation demand estimation tool.

Accomplishing these objectives will provide the materials necessary for regular agricultural water demand estimates for the region of Arequipa, Peru. Objective 1 will result in a structure to regularly create and validate crop maps for the region. Time series of the crop maps will be used in objective 2 to estimate crop age, which in turn will be used to estimate crop water demand in objective 3. These water demand maps could then be integrated over crop types or sub-regions to create water demand estimates which provide useful information to policymakers and farmers. While this study focuses on the region of Arequipa, Peru, the methodologies presented should be applicable for regions with a need to map crop water demand but with similarly sparse information on crop types, locations, and ages. Moreover, while the focus of this study is on water demand, the crop maps themselves represent valuable tools that could be used for additional analyses.

2. MATERIALS AND METHODS

2.1 Study Area

The study area is within the Department of Arequipa, in southwestern Peru (Figure 1). The agricultural area to be classified is in and around the city of Arequipa. The region has significant topographic variability, which necessitates the use of small fields, terraces, and less-mechanized agriculture. In addition, much of the agriculture is done in close proximity to urban infrastructure such as buildings and road systems, and many crops are commonly grown in the same immediate area. Small crop field sizes of approximately 0.1 ha or 1000 m^2 are not uncommon in this region, which is much smaller than the 19.3 ha average size of crop fields in the United States (Yan and Roy, 2016). The city of Arequipa is located in an arid region; annual average precipitation in the agricultural areas in and around the city is approximately 123 mm/year, with most of that occurring from December to March (Moraes et al., 2019). Due to its aridity, the agriculture in the city is irrigated and dependent on the highly managed water supply that comes from the higher elevation Andean region (Stensrud, 2016). Due to the low latitude of the region, crops are grown year-round without a defined planting or harvesting season, although there are preferences for growing particular crops during different times of the year.

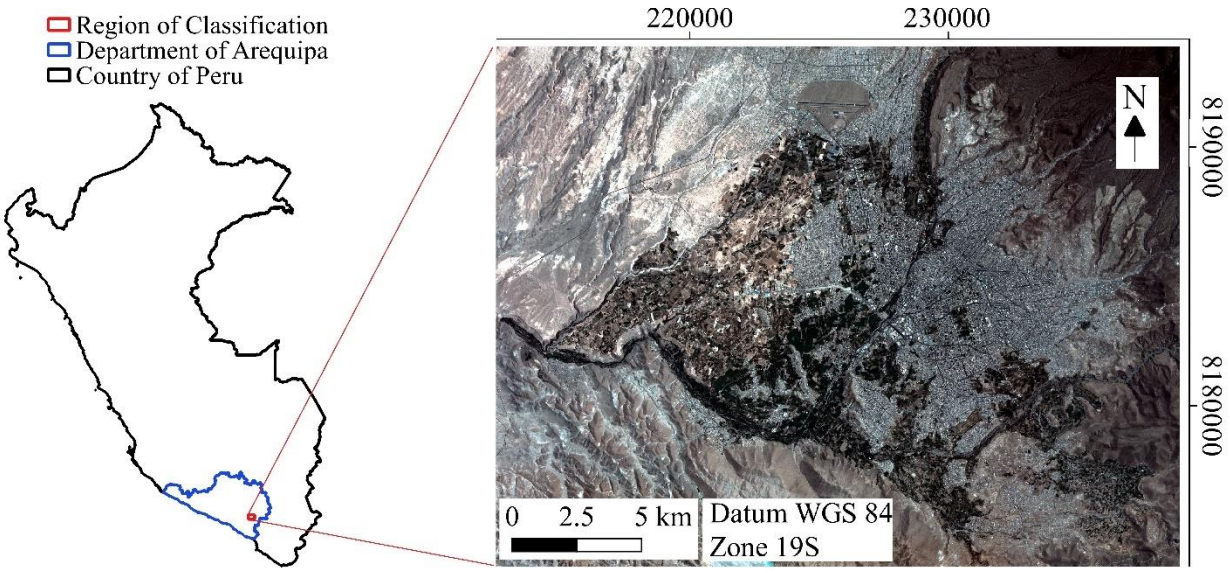

Figure 1. Region of classification for the study (right) and its location relative to the country of Peru and the department of Arequipa (left).

2.2 Satellite Data Acquisition

Satellite images from September 2019 to February 2020 were obtained courtesy of Planet Labs, Inc. The PlanetScope images used in this study have an approximately 3 m spatial resolution, daily repeat cycle, and four spectral bands: blue (455 - 515 nm), green (500 - 590 nm), red (590 - 670 nm), and near infrared (780 - 860 nm) (Planet Labs, Inc., 2018). The images were already converted to surface reflectance values by Planet Labs, Inc. PlanetScope images were chosen due to their high spatial and temporal resolutions, which make the classification process more feasible for the region, given the small field sizes and the lack of a defined crop growing season. To compose a monthly mosaic image that covered the entire region of classification, hereafter referred as a "monthly satellite image", cloud-free images that were nearest to the ground reference data sampling dates were selected (Figure 2). Approximately eight to twelve cloud-free images, taken on one or two days, were used to create each monthly image.

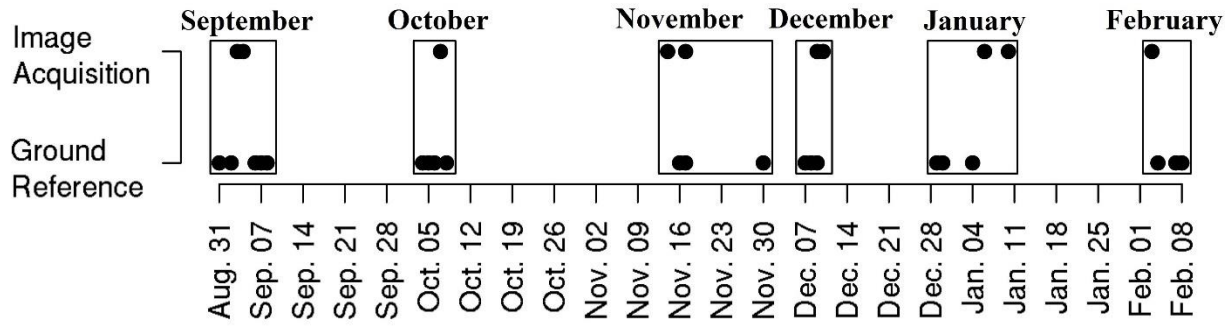

Figure 2. Dates of ground reference sampling and satellite image acquisition for each month of the study period (September 2019 to February 2020). Multiple satellite images may be acquired per day to fully cover the study area.

2.3 Ground Reference Data Acquisition

Supervised classification methods work by using features extracted from reference data of known groups, or classes, to “train” an algorithm, or classifier, and then using the classifier to predict the classes of objects of interest based on the same features of those objects. In this study, the classes are different types of crops, the features of the data are the spectral data in the satellite images, and the objects to be classified are the cells (pixels) of the satellite images. Therefore, to use supervised classification methods to classify different crops in satellite images, ground

reference data are needed. Since it is difficult or impossible to know the class of an object by looking at the satellite image, it is most common to sample reference data in-situ and then locate the areas in geographic information system (GIS) software and extract the desired features from the satellite image once the class is known. When sampling ground reference data for satellite image classification, entire areas (such as agricultural fields or forests) are sampled, and the data of the cells of those areas are extracted. For this study, one sampled field area is considered a single ground reference location regardless of how many cells (image pixels) are contained in it. To extract spectral data from ground reference locations, fields were delineated with polygons in GIS software and R was used to extract the spectral data from the cells contained within or overlapping with the polygons. To avoid edge effects, which is mixing of the spectral signal of nearby cells and which occurs near the edges of fields, generally only the central area of a field was delineated. The distance from the edge of the field to the delineated central region was determined by visual inspection and examination of the spectral data of the pixels to remove those with clear signs of spectral mixing.

To collect the reference data needed in this study, it was necessary to design a reference data collection method. This method needed to allow for accurate identification of ground reference areas (agricultural fields) in GIS software along with the ability to record supplementary crop-related information such as crop age, crop density, and crop height. To serve as an error check and field identification aide, it was also important to be able to associate a picture with the field. Finally, the method needed to be free or low cost and simple to use. To accomplish these goals, a smartphone-based method was developed, based on the smartphone-based application Epicollect5 (Epicollect5, n.d.). Using Epicollect5, it was possible to record a point position with an average accuracy of 3 m (using smartphones built after about 2014) and associate with it a picture of a targeted crop field, the direction in which the picture was taken, and many crop features such as: crop type, age (measured in months, at a resolution of 0.5 months), height (integer inches), and population density (plants/m^2). Epicollect5 is free for research purposes, customizable, has a friendly interface, and perfectly matched the needs of this study. Using a smartphone-based method was beneficial due to the near-ubiquity of smartphones in the region.

To collect data for this study, four undergraduate agronomy students from the Universidad Nacional de San Agustín (UNSA) were selected and trained to use the ground reference data collection method. It was planned to collect ground reference data throughout the region of

classification on the first week of every month for at least a year to develop and evaluate the crop mapping algorithm. Data collection began in September 2019; however, due to the COVID-19 pandemic, field work was interrupted after the February 2020 campaign, yielding only 6 months of ground reference data. Moreover, due to logistical difficulties, the start of the field campaigns in November 2019 and January 2020 were delayed, resulting in longer than optimal gaps in the ground reference data (Figure 2). This is less than ideal, but it did not affect the classification process for those months beyond needing to verify that fields were not harvested between the time of ground reference data and satellite image acquisition.

The region of classification was roughly split into three agricultural areas (Cerro Colorado, Characato-Paucarpata, and the Chili River Valley) based on the advice of collaborator and UNSA Agronomist Jose Pinto (personal communication, 2019) who indicated that farmers in each of these regions preferred different crops. Cerro Colorado and Characato-Paucarpata are higher in altitude and separated by the Chili River Valley (Figure 3). Ground reference data were collected from each of these areas in each month. Due to logistical difficulties and the highly-variable topography in the region, it was difficult to get a spatially-randomized sample of agriculture within each of the three areas. It was, therefore, assumed that the mixture of crops being grown in each of these areas for each month was fairly similar, with the most variability in regional agriculture occurring between these areas. This allowed the ground reference sampling teams to focus on capturing a similar, maximal number of samples within a constrained spatial subset of each area instead of on obtaining a spatially-representative sample across each area. Ground reference sampling teams were trained to sample different fields along the route in each month that would capture a spatially-representative sample from the area in which they were sampling. With this sampling plan, it is assumed that the monthly sample is spatially representative of the entire region.

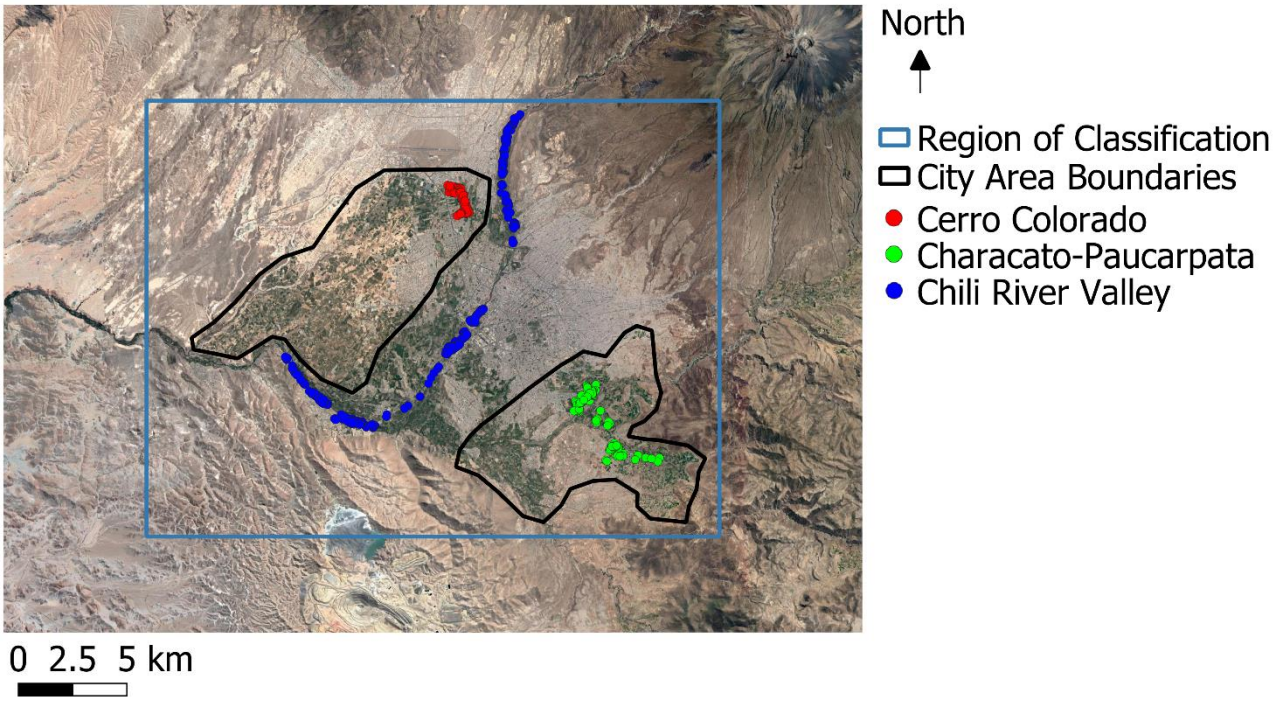

Figure 3. A map of the region of classification with all ground reference data points, color coded by agricultural area within the city. There are three agricultural areas defined: Cerro Colorado, Characato-Paucarpata, and the Chili River Valley. The first two are roughly outlined by black polygons and are identified by the color of the points within them. The Chili River area is identified by the blue points that run along it and includes the agricultural areas (identified by their greenness) surrounding the points. The background of the map is a Google Satellite map (Google, 2016).

2.4 Ground Reference Data Pre-Processing

2.4.1 Assessment of Ground Reference Data

Six months of ground reference data were collected by trained undergraduate volunteers in Arequipa. During this period some crop types were found over a substantial portion of the study domain in all six months (Table 1); these include alfalfa, peas, broccoli, celery, and lettuce. For other crops, there appeared to be a "season" in which they were dominant. For example, garlic, quinoa, and onion were dominant in September and October, then became less abundant in November, and were largely gone from the region by December. Meanwhile, corn was minimal in September, but started to appear in November, and became a dominant crop in the region in December. Fava beans were sampled from a significant number of fields only in September, which suggests that they may have been more dominant in the months prior to the start of ground-reference sampling.

A total of 32 cover types were sampled (Table 1). Monthly sample sizes started at 148 total fields in September and generally declined over time, ending at 103 total fields sampled in February. The ten most commonly sampled cover types had total sample sizes (across all months) ranging from 43 (onion) to 94 (alfalfa). Many cover types were less common in the ground reference database and did not have a sufficient number of samples to classify them well. Indeed, 20 of the cover types had 10 or fewer total samples, usually spread out over several months. Assuming the sample frequencies represent the actual distribution of the crops, then these crops represent a small proportion of the regional crops (often <1%). Based on this analysis, the subsequent generation of crop maps was restricted to the cover types that had more than 10 total samples, which includes the following cover types: alfalfa, peas, garlic, broccoli, green beans, corn, bare soil, celery, onion, lettuce, quinoa, and oats. Representation of the other crops is limited in the ground reference database and assumed to be a small overall fraction of crops planted based on the ground reference sampling strategy. Therefore, these crops were not included in the development of regional crop maps as there is an insufficient number of samples to classsify their spectral signature and including them could cause spectral overlap with more representative crops that would detract from the overall classification accuracy.

Table 1. Sampled cover types and corresponding number of ground reference samples for each month of the study. Cover types are sorted by the total number of times they were sampled in the 6-month preliminary study period. Bolded cover types have more than 10 total samples and are included in the crop mapping process.

Cover	September	October	November	December	January	February	Total
Alfalfa	23	18	14	13	15	11	94
Peas	13	16	13	13	18	5	78
Broccoli	16	15	15	12	9	7	74
Garlic	27	26	18	2	1	0	74
Green beans	1	7	11	12	14	17	62
Corn	2	2	7	16	17	15	59
Bare soil	13	9	7	8	7	13	57
Celery	5	6	10	10	13	10	54
Lettuce	6	6	8	12	8	4	44
Onion	11	16	9	6	1	0	43
Quinoa	5	5	3	4	0	0	17
Oats	4	7	4	0	0	1	16
Carrot	1	1	1	4	1	2	10
Fava beans	9	1	0	0	0	0	10
Potatoes	2	4	3	1	0	0	10
Pumpkin	0	1	1	1	2	3	8
Cauliflower	1	1	0	0	2	3	7
Leek	0	1	1	1	2	2	7
Potato	0	0	0	0	2	4	6
Illusion flower	3	1	0	0	1	0	5
Cabbage	1	0	0	1	1	1	4
Pasture	0	0	1	0	1	2	4
Parsley	0	0	1	1	1	0	3
Carnations	0	0	0	0	1	1	2
Chinese onion	2	0	0	0	0	0	2
Good herb	1	0	0	0	0	1	2
Rue	1	1	0	0	0	0	2
Turnip	0	0	2	0	0	0	2
Chard	0	0	0	0	0	1	1
Cilantro	0	1	0	0	0	0	1
Italian zucchini	0	0	0	0	1	0	1
Wheat	1	0	0	0	0	0	1
Total	148	145	129	117	118	103	760

2.4.2 Quantifying the Visibility Threshold of Crops

Crops in early development stages can easily be classified as soil due to the amount of soil exposed between the limited canopy of the small plants. A data filtering strategy was created to remove young crop samples from the ground reference dataset so that those fields were not used in the development of monthly crop classifications. For this process, the reported age or height (for alfalfa) of the crop in the field (from the ground reference dataset) was evaluated against the mean Normalized Difference Vegetation Index (NDVI) (Rouse et al., 1974) values for the ground reference field, as calculated from the corresponding remote sensing imagery.

The extent of canopy closure for each crop can be inferred from the change in average field NDVI over time (Figure 4). The NDVI is highly sensitive to the presence of chlorophyll, resulting in higher values for healthy plants than for soils. Soil NDVI values for the study area range from 0.10 to 0.27. Young crops with little canopy cover will have NDVI values indistinguishable from soil because the PlanetScope pixels will reflect more soil than canopy. As the crops age and their canopies become fuller, the field average NDVI increases, and it becomes possible for the image classification algorithm to differentiate between fields with crops and fields without crops.

The relationship between NDVI and crop height for alfalfa (Figure 4a) and crop age for corn (Figure 4b) are typical for the majority of crops analyzed. They illustrate that for most crops, NDVI has a logarithmic relationship with crop height or age, with a sharp increase in NDVI soon after planting (or cutting for alfalfa), followed by a stable phase once full canopy cover is reached. If crops were sampled after senescence or harvest, then a declining NDVI phase is also observed, such as the slight declining phase observed for corn (Figure 4b). The relationships found for oats (Figure 4c) and quinoa (Figure 4d) are different from the rest of the crops, likely due to the small number of samples or the fact that both crops change color as they near harvest.

The proportion of green canopy cover has been found to have a very strong linear relationship with the NDVI of crops,, often with a slope of nearly 1 (Tenreiro et al., 2021), which signifies that NDVI is a good representation of canopy cover for many crops. Several studies have found a double-logistic relationship between NDVI and crop age (represented as day of the year), i.e., a constant relation followed by a sharp increase, a stable relation, a sharp decrease, and then another constant relation (Doraiswamy et al., 2007; Li et al., 2007; Huang et al., 2019). However, in these studies, NDVI was measured for the days of the year before the crops were planted and after the crops were harvested. In this study, NDVI was only measured when crops were planted

and usually before they were harvested, and so, in most cases, only the sharp increase and stable relation phases were captured, leading to the apparently logarithmic relationship observed in Figure 4.

Even when full canopy cover is reached, there is a variation in NDVI with age or height (Figure 4a,b). There are multiple potential sources for this variance, including sampling error (i.e., misinterpreting the age or height of the plant), different planting conditions (e.g., variable plant density, different breeds or varieties), and different growing conditions (variable water content, microclimates or pest pressure) (Shanmugapriya et al., 2019). Large variance in spectral data could cause low classification accuracies due to class overlap and reduce the ability to fully represent classes with a limited number of samples. For this reason, it was decided to use NDVI thresholds to "filter" samples. The NDVI thresholds were chosen to remove samples that had not reached a state of canopy closure, thereby ensuring that crops were distinguishable from soil and reducing the variance of the data. By using a fixed NDVI threshold for each crop, crops that are too young to be classified can easily be removed, as can substantial NDVI outliers that could be due to diseased crops, partially-harvested fields, and other factors. Ground reference fields where the average NDVI does not exceed the threshold are excluded from the classification process

Lower bounds for ages (heights) and NDVI thresholds used in the filtering process were determined using NDVI vs. age (height) graphs, a selection of which are presented in Figure 4. NDVI thresholds were determined by visually approximating the NDVI value above which the within-group (age or height) NDVI means began to vary minimally with age (height); lower bounds for ages (heights) were found by visually approximating the age (height) value that corresponded to the chosen NDVI value. In general, NDVI values were chosen conservatively to retain as many samples as possible while ensuring the remaining samples had the desired qualities. NDVI thresholds used for filtering ground reference data for each crop type are provided in Appendix A (Table A-1). This process was difficult for oats and quinoa due to their small sample sizes and decreasing NDVI with height and age. The lowest NDVI threshold was 0.35 for lettuce and onion, two small plants that even when densely planted will not completely close their canopy between rows. Alfalfa, a dense perennial grass, had the highest NDVI threshold of 0.60. Based on the observed NDVI ranges, cells with an NDVI value of 0.30 or less were assumed to be soil and pre-classified as such to decrease algorithm runtimes. The number of ground reference fields available after filtering for minimum NDVI are listed in Table 2.

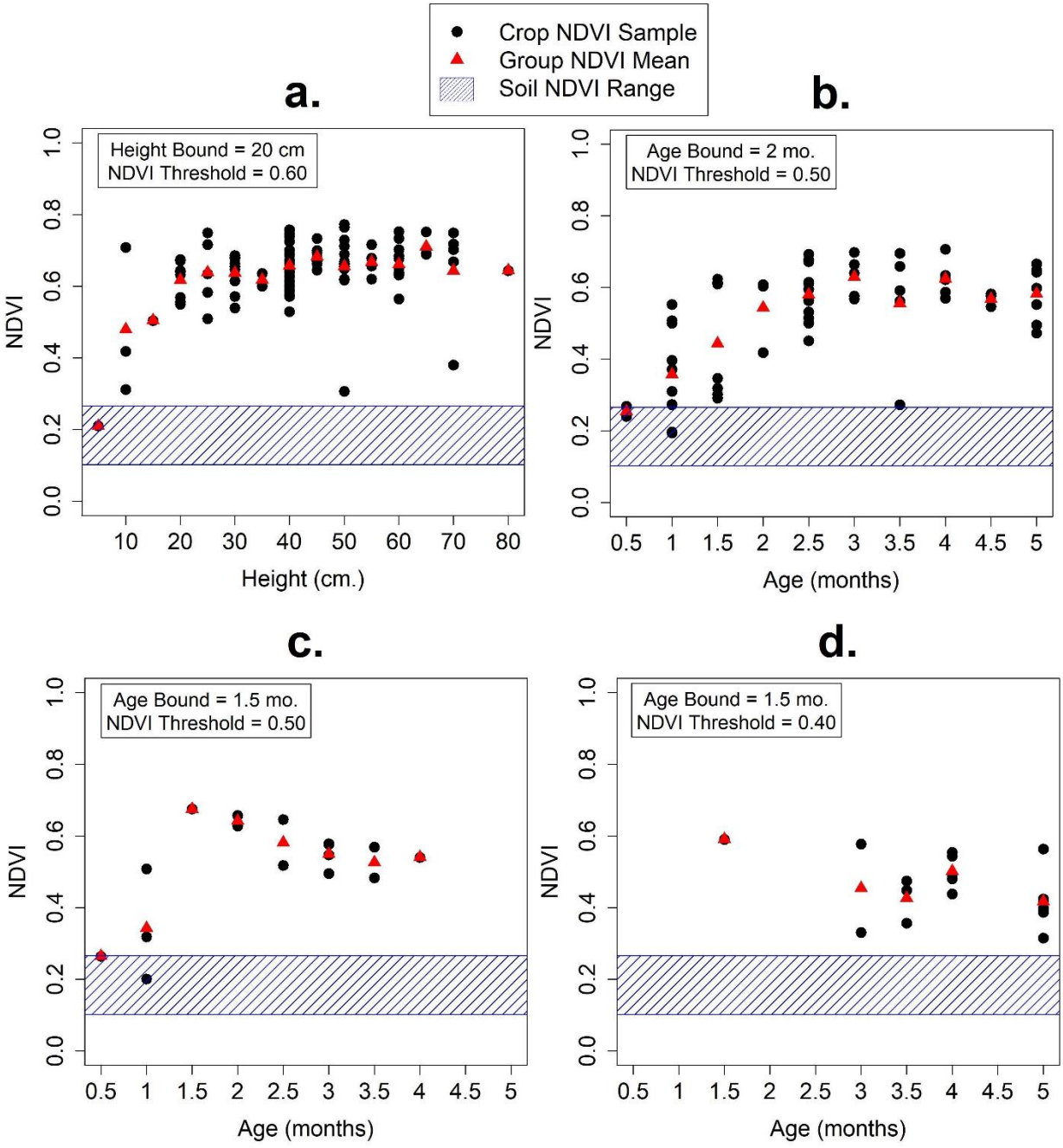

Figure 4. NDVI vs. age (height) plots for: (a) alfalfa, (b) corn, (c) oats, and (d) quinoa. Black points represent means of cell NDVI values for sample fields and red points represent means of black points at different age (height) levels. The blue, cross-hatched interval represents the range of means of cell NDVI values for sample fields which had the bare soil cover type.

2.4.3 Identification of Additional Ground Reference Data

The number of fields that can be sampled each month as ground reference data is low relative to the number of different crops being grown and the number of small fields being cultivated. Fortunately, crops are persistent developing from seeds to mature plants over several months, so a field observed in one month can be used in the classification process for every month from planting to harvest. For example, recently planted crops that failed the NDVI threshold filtering process in the month they were observed as part of the ground reference survey due to their small size, might be large enough a month later to be used for image classification in that month. For a field from one month to be used for extracting spectral data from a different sampling month, the field had to meet the following conditions: (1) it could not overlap spatially with any other fields in the given sampling month (i.e. we cannot reuse the same field); (2) the reference age of the crop associated with the field must be between zero (just planted) and the maximum observed reference age for that crop, after the reference age is adjusted to account for the change in month (e.g., if a field in November had a crop of age 2 months, then its adjusted age in December would be 3 months); and (3) the field passes the NDVI threshold filter in the given sampling month. This process resulted in additional ground reference fields being available for the classification process (Table 2). Note that older crops in fields included in the ground reference database can be used to increase the number of ground reference locations in earlier months if they meet the same criteria.

Table 2. Number of ground reference fields by crop type that passed the NDVI threshold filtering (TF) and after the identification of additional ground reference fields that meet the selection criteria (IA). Values indicate the total number of ground reference fields by crop type after each pre-processing step. The number of fields listed under IA is the number actually used in the classification process.

	September		October		November		December		January		February	
Cover	TF	IA	TF	IA	TF	IA	TF	IA	TF	IA	TF	IA
Alfalfa	11	15	17	21	14	14	12	12	10	10	10	10
Bare soil	13	19	9	15	7	11	8	12	7	8	13	15
Broccoli	8	8	11	14	15	17	9	10	8	8	5	5
Celery	2	2	1	2	5	6	6	6	10	10	5	5
Corn	0	0	1	1	4	5	13	13	11	11	12	12
Garlic	24	30	24	34	14	15	0	0	0	0	0	0
Green beans	0	0	0	0	2	2	6	6	11	11	10	10
Lettuce	2	2	4	5	5	5	8	8	8	8	2	2
Oats	2	5	7	8	4	6	0	0	0	0	0	0
Onion	4	4	13	16	9	14	5	7	1	1	0	0
Peas	4	4	10	12	4	5	4	4	9	9	2	2
Quinoa	4	6	2	5	3	4	3	3	0	0	0	0
Total	74	95	99	133	86	104	74	81	75	76	59	61

2.5 Creating A Map Of Agricultural Areas

Agricultural production in and around the city of Arequipa is often done in close proximity to infrastructure, including roads, building, bridges and terrace walls. Near the river, the agriculture is intermingled with native vegetation including riparian woodlands and pastures. The city also has sports fields and parks, although lawns are uncommon. To avoid the classification of non-agricultural vegetation as agriculture and to reduce computational time when classifying images for crop production, a map of agricultural areas (or mask of non-agricultural areas) was developed.

Supervised classification (maximum likelihood) was used to map five different cover classes: barren, urban/built-up (infrastructure), turf grass, trees, and agriculture. These classes were chosen because they broadly represent land cover in the region. Note that the barren class includes non-agricultural soil but not agricultural soil. Twenty-five ground reference samples of each non-agriculture class were obtained by visual identification in Google Satellite images, while field-sampled ground reference data were used to represent the agriculture class. Agricultural samples were used only if the field was sampled in only a single month (to avoid duplicate samples in the training and validation data pool) and they were not bare soil at the time of sampling (to ensure the field had vegetation in at least one month and therefore represented typical agriculture). Application of these constraints resulted in a total of 202 reference samples for agricultural fields. Thirty percent of the reference samples of each class were used for validation and the remaining data were used for training of the classifier. Each sample (agricultural field or non-agricultural area) defines a region of interest (ROI) containing several image pixels corresponding to the monthly raster image. For each sample cell, the mean, standard deviation, and range (maximum minus minimum) of the time series of monthly NDVI values were calculated and used as features to construct a new multi-layer raster file. The features of mean, standard deviation, and range of monthly NDVI values were used because they capture both the central tendencies of all classes and the greater variability in NDVI present in the agriculture classes as fields are harvested and replanted. The non-agricultural classes should have fairly-stable NDVI values over time, even the non-agricultural vegetative classes because of the low latitude and constant climate of the region. Probability histograms were then computed for each feature of each training data class. Then, to classify each cell, the features of the cell were computed, and the probabilities of belonging to each class's feature histogram were computed; the cell was then assigned to the class with the largest

sum of probabilities across the features. After classification, the non-agricultural classes were aggregated into a single non-agricultural class.

The accuracy assessment (Table 3) shows that the classification was generally successful in differentiating agricultural and non-agricultural classes, but there was substantial confusion between the non-agricultural classes. As the focus of this classification process is to separate agricultural from non-agricultural areas, the confusion between classes within the two major classifications can be neglected. At the end of the classification process, the raster was reduced to a binary mask of agricultural and non-agricultural pixels.

Table 3. Validation results for the non-agriculture mask raster classification. Rows represent the actual class from the validation samples, whereas columns represent the classified class.

Class	Agriculture	Barren	City	Grass	Trees	Class Accuracy (%)	Ag/Non-Ag Accuracy (%)
Agriculture	3750	1	1	92	162	94	94
Barren	0	62270	26553	0	4661	67	100
City	12	5932	14981	120	1251	67	100
Grass	24	19	46	320	1894	14	99
Trees	5	1	8	102	232	67	99

On inspection, the masking raster was found to have spatial consistency issues. Visually, these appear as scattered, isolated classified cells (a “salt-and-pepper" effect) and poorly-mapped class boundary regions due to the per-pixel approach used for classification. Additionally, several areas were found with substantial classification errors. To fix the issues with the masking raster, the raster was first "sieved" using the sieving algorithm from the Geospatial Data Abstraction Library (GDAL) (Warmerdam, n.d.). The sieving process removes small groups of cells that are isolated within larger groups of cells of a different class. The end result is to produce larger, contiguous groupings of similar classes. The raster was then converted to polygons for comparison with high-resolution satellite images of the region. Polygons of the sieved mask were overlain on raw PlanetScope images, and a Google Satellite map (Google, 2016) and final corrections were made to the classification using manual digitization techniques. During this manual-correction process, field boundaries were fixed, and the Chili River was mapped as non-agricultural. The process was deemed completed when visual inspection found no additional errors of significance. The polygon file was converted back to a raster image of the same domain and resolution as the

original mask, and the validation accuracy of the new map was found to be perfect for each of the five classes (note that the manual-correction process was performed without knowledge of the exact location of the reference data). While this process was time-consuming, it is expected that this masking raster will be useful for future research in the region. Most of the terraced fields have been used for agricultural production for centuries, while expansion of the city has generally occurred along the margins of the agricultural areas or in the hills far from irrigation water. Identifying these changes should be much easier than creating an entirely new masking raster and can be done infrequently or in response to known development. The resulting raster mask (Figure 6) reduces runtimes for the crop mapping algorithm by allowing it to skip non-agricultural cells. This also results in the production of higher quality maps.

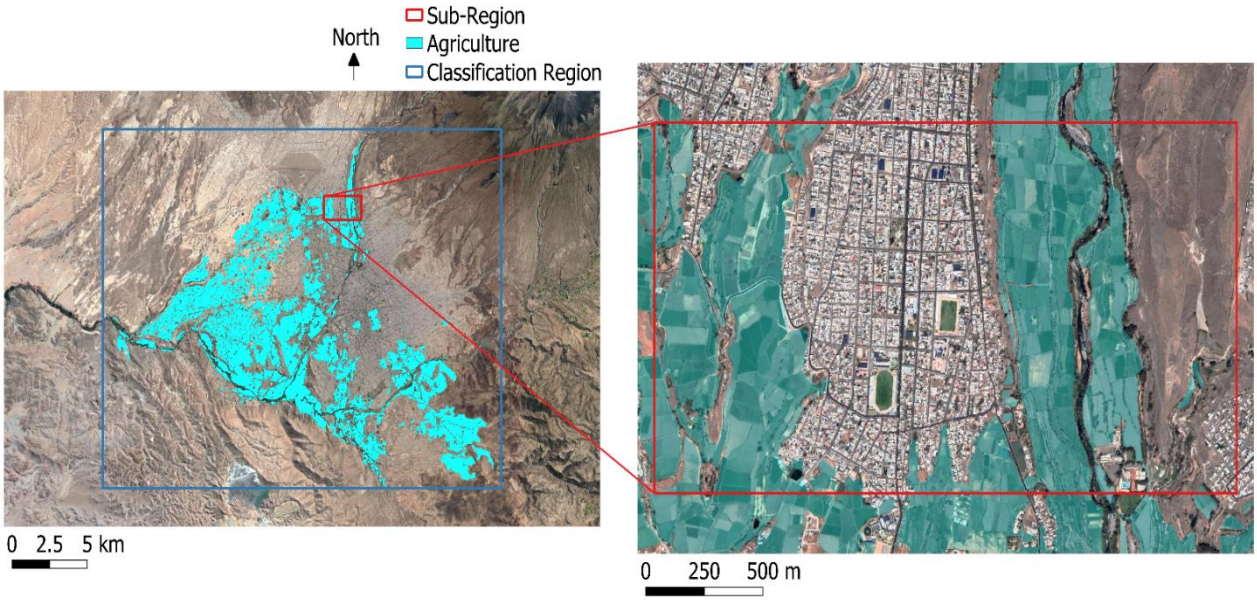

Figure 5. Left: Map of the agricultural area in and around the city of Arequipa, Peru. Right: Sub-region of the map of agriculture with opacity turned down to see underlying land cover. Classified map was adjusted manually to improve accuracy and is used to mask non-agricultural areas out of imagery as part of the crop mapping process. The background of the map is a Google Satellite map (Google, 2016).

2.6 Classification Methods And Validation

Generally speaking, classification methods can be considered as either supervised or unsupervised classifiers; they may also be grouped as per-pixel, sub-pixel, object-oriented, and field-oriented classifiers (Lu and Weng, 2007). The classifiers in this study were applied as supervised per-pixel classifiers, meaning that they were trained using one set of data (training data) and validated using another, independent set of data. Both the training and the evaluation datasets must identify areas in the imagery of known cover classes. Supervised classification was used because there were many classes (crop types) to be identified that likely have significant overlap between their spectral signatures, making unsupervised methods unsuitable (Song et al., 2005) and increasing the need for an accuracy assessment of the final product. The per-pixel nature of the classification scheme means that each pixel in the remote sensing image is assigned to a cover class. The per-pixel method was selected due to the high variability in shape and size of the agricultural fields in the region and lack of field parcel data.

Three supervised classification methods were used in this study: K-nearest-neighbor (KNN), maximum likelihood with multivariate normal distribution assumption (MaxL), and RandomForest (RF) (Breiman, 2001). The KNN and MaxL classifiers are discussed in detail in most books on pattern classification (e.g. Stork et al., 2000), whereas the RandomForest method is described in detail by Breiman (2001), so only a brief explanation of each classifier is included here. The KNN and MaxL methods were coded in the R language, whereas the RF method was used through the randomForest package of RStudio (R Core Team, 2020).

1. The K-nearest-neighbor (KNN) method is a generalization of the nearest neighbor search technique (Fix and Hodges, 1951). It works by, for each data point to be classified, computing the distance in feature space from the data point to each training point ("neighbor" in the feature space), and assigning the data point to the class that occurred the most often among the K training points with least distance from the data point ("K nearest neighbors"). The KNN method is easy to implement and when K is small has short runtimes compared to many other classification methods due to its computational simplicity. Is it also worth noting that, under certain conditions and with the limit of infinite training data, the K-nearest-neighbor method has the optimal

behavior of a classifier (Gyorfi and Gyorfi, 1978), i.e. its error rate converges to the irreducible error rate.

2. The maximum likelihood (MaxL) method used in feature classification is a particular case of maximum likelihood estimation and a description of it can be found in many textbooks on remote sensing and pattern classification (e.g., Richards, 1999, p. 240). It works by assuming a family of distributions for the spectral data of the classes (e.g. multivariate normal), obtaining estimates for the parameters of the class distributions using the training data, and then assigning data points to the class that maximizes a likelihood function. The history of maximum likelihood estimation itself goes back hundreds of years to famous historical figures such as Bernoulli, Lambert, Lagrange, and Laplace, but its use was not widespread until Fisher popularized it in the early 20th century (Edwards, 1974). The MaxL method has several attractive features, such as ease of implementation, ease of interpretation, and good convergence properties with increasing sample sizes (Stork et al., 2000, pp. 85 and 101). The MaxL method has been shown to be competitive with other, more complicated methods in several studies (e.g. Ozdarici-Ok et al., 2015).
3. RandomForest (RF) is a relatively recent addition to image classification, being first described by Breiman in 2001. It works by using a large number of decision trees. The decision trees are grown using randomly chosen samples from the training data; nodes are split using the best split (most distinguishable features) on a subset of the input variables (in this case, spectral bands of the data); the trees are grown as large as possible and are not pruned (Breiman, 2001). In this study, the RandomForest method was used through the randomForest package of RStudio (The R Foundation, 2020). The order of the subset used for node splitting was determined by minimizing the out of bag error resulting from each possible subset order. The number of trees used was 2000, as no improvement in the classification accuracies was observed beyond this number. The Random Forest (RF) method has outperformed many other classification methods in studies using satellite image classification (Lawrence et al., 2006; Heung et al., 2016; Meier et al., 2018), making it an appealing choice for this study.

Classifier skill was judged using per-class validation accuracies (or simply accuracies), calculated as the number of correctly classified cells of a given class divided by the total number of reference cells of that class; this accuracy is sometimes called the "producer's accuracy". Overall accuracy and the kappa coefficient were not used because of the highly variable sizes of agricultural fields and variable number of samples for each class, which could cause these error metrics to be unrepresentative of all classes. Accuracies are always expressed as percentages. Ground reference data for each month are split into training and validation datasets, which further limits the availability of data for both processes.

In general, there is no inherently superior classification method, with the most suitable classifier depending on the characteristics of the available data (Stork et al., 2000, p. 456). The KNN, MaxL, and RF classifiers were chosen for this study because each one offers a unique approach to classification: KNN is a non-parametric geometry-based classifier; MaxL is a parametric likelihood-based classifier; and RF is a non-parametric decision tree classifier. Using a wide range of classifier types is useful as an exploratory tool and may help determine which type of classifier is best suited for the data and would thus be appropriate to use in the future when the tool is deployed. In this study, for each month, all three classification methods were applied to the reference data to examine the performance of the different methods. The classifier with the highest accuracy within a given month was chosen as the classifier for that month, to be used throughout the entire crop mapping process (Section 2.7). However, this is not recommended as a permanent part of the process when viewed as an operational tool. Using the same classifier in every month would be simpler for users, so that exploratory data analysis is not necessary to choose a classifier in every month. Choosing one classifier that is robust enough to perform well on most data encountered by the tool is something that ought to be done before the tool is deployed.

In addition to being useful as an exploratory tool, using multiple classifiers for this study was deemed acceptable due to the following reasons. Firstly, because the dataset characteristics in each month are different: crop types and distributions generally differ in each month; certain crops undergo general trends in age (Table 1); and spectral data differ for each crop type and for its growth stage. These differences in spectral data could cause different classifiers to be best suited to the monthly data. Secondly, although the use of a different classifier in each month may reduce inter-month consistency, the use of the temporal-correction algorithm should remove most inconsistencies between the different classification methods (Section 2.7.3). Nevertheless, when

establishing this process as a long-term operational tool, it is recommended that a single algorithm is identified as most suited for the data typically encountered by the tool. This will make the tool simpler for users. However, this may require the collection of a larger ground reference dataset and significant exploratory analysis.

2.7 The Crop Mapping Algorithm (Cma)

The Crop Mapping Algorithm (CMA) was designed to have filtered ground reference data and monthly satellite images as inputs and to output crop maps, map accuracies, and crop statistics (areas and percentages of agricultural cover for each cover type). Prior to image classification, the CMA determines the classification method for each month by running each classification model with a large number of sets of training and validation ground reference data subset at random from the full ground reference dataset, and comparing the mean validation accuracies from the runs. Each time the classification process is performed, 60% of the ground reference data are selected at random for training, and the remaining 40% are used for validation. Small sample sizes cause significant differences in classification outcomes depending on how the training and validation datasets are sampled, and so multiple runs are performed to capture the variability in outcome and obtain an estimate of the average model performance. The monthly classification method is determined prior to image classification by classifying reference data because classifying high-resolution images has a large computational cost. Since the chosen method will later be used to classify images, which are then subject to post-processing, it is assumed here that the method with the larger mean accuracies prior to post-processing would have the larger mean accuracies after post-processing.

Due to spatial and temporal inconsistencies in the classified maps, the CMA was designed to incorporate spatial- and temporal-consistency correction algorithms during post-processing. The CMA performs image classification, post-processing, and validation assessment using randomly-sampled training and validation datasets for a user-defined number of runs. Each training set consists of 60% of the full ground reference dataset, selected at random, with the remaining 40% used as the validation set. The validation accuracies and crop statistics from each run are used to form mean confidence intervals for these results, which gives an estimate of the average performance of the model. The number of runs used is user-defined so that users can select a number that fits the scope of their data and allows for flexibility in the widths of the confidence

intervals. For this study, the number of runs used was set to 108, as it was deemed sufficient to develop sufficiently narrow confidence intervals and is divisible by the number of classification processes that could be run in parallel, which is 36 (3 machines were available, with each machine having enough memory to run 12 classification processes in parallel). Since there were 6 months in the study, this resulted in a total of 648 crop maps, 108 for each month. The output crop maps are based on the ensemble of crop maps created in the multiple runs, with each cell value representing the most common class for that cell among all crop maps (for a given month).

There are some specific criteria and limitations related to the use of the CMA: (1) The CMA does not classify crops with 2 or fewer ground reference samples (including samples that are gained from the identification of additional ground reference fields detailed in Section 2.3.3) as these cases were determined to have an inadequate number of samples for accurately representing a crop class. (2) Celery and oats were not classified, despite having a sufficient number of samples in many months, because these crops almost always had low validation accuracies (less than 50%) and including these crops in the overall classification process led to significant decreases in validation accuracies (greater than 10%) for crops that represented a large fraction of planted area (e.g., alfalfa, corn, and onion). This means that those fields will be misclassified in the resulting crop maps, but as they are fairly minor crops in the region, this misclassification is assumed to result in minimal additional error in estimation of water demand. (3) Since accuracies and crop statistics are considered on a monthly basis, the accuracy and crop statistic confidence intervals output from the CMA were designed to be joint (by month) confidence intervals; the method used for creating joint confidence intervals was the Bonferroni method (NIST/SEMATECH e-Handbook of Statistical Methods, 2003).

The remainder of this section will explain details of the CMA related to the choice of the monthly classification methods, the development of the Spatial Consistency Correction Algorithm (SCCA) and the Temporal Consistency Correction Algorithm (TCCA), and the ensemble crop maps.

2.7.1 Determination of Monthly Classification Methods

To quickly determine the classification method for each month, 400 randomized training and validation datasets were generated, and each algorithm was trained and validated on the randomized datasets to obtain 400 sets of accuracies for each method and month combination (e.g.

MaxL had 400 sets of accuracies in September, as well in each of the other months). Only validation data were classified at this stage due to the much faster runtime of classifying validation data compared to classifying entire images and performing post-processing (a few seconds versus a few hours), allowing for better estimation of mean accuracy. Four values of K (1, 2, 3 and 4) were used in the KNN algorithm (denoted as KNN (K=x), where x is the value of k). The Bonferroni method was used to establish joint 90% (or greater) confidence intervals (CIs) for the monthly means of the accuracies. The large number of classifications resulted in narrow confidence intervals, and so boxplots of mean accuracy distributions were used to visually determine the classification methods in each month.

2.7.2 The Spatial Consistency Correction Algorithm (SCCA)

The process of spatial-consistency correction, sometimes called "sieving", is a process whereby groups of cells of a certain class are assigned to another class (e.g., by using the class of the largest neighboring group) if the number of cells in the group is less than a threshold. Sieving is often performed when using per-pixel classifiers due to the presence of scattered, isolated classified cells (a “salt-and-pepper" effect) that occurs due to per-pixel differences in spectral data. These per-pixel differences could be due to noise, spectral blending of nearby pixels, or natural within-class variability. There are many types of spatial consistency correction algorithms (SCCAs) available, but most of these require the setting of a minimum threshold area, below which small patches are deemed too small to remain independent. In the case of the CMA, the SCCA was developed to reduce the misclassification of pixels within or along the edges of the small, terraced fields found in and around the city of Arequipa.

In order to determine an appropriate sieving threshold, the 148 ground reference fields from September (Table 1), were delineated with polygons, excluding cells with edge effects, using the unclassified PlanetScope image from September and a Google Satellite map (Google, 2016) as visual aides. The ground reference fields in September were used due to the fairly large sample size and the broad spatial distribution of the samples in this month. The polygons were used to calculate field sizes using an approximate PlanetScope image cell size of 3 m by 3 m (giving a cell area of 9 m^2). The field median size was 1755 m^2, the mean was 2254 m^2, the standard deviation was 1474 m^2, and the range was 423 to 8127 m^2. The distribution of field area had a strong right skew (Figure 6).

A classified image was then sieved using GDAL at various sieving thresholds between the lower end of the range of field sizes (423 m^2) and the median of field sizes (1755 m^2). The sieved images were visually examined to determine an appropriate sieving threshold. Positive factors that were looked for were the ability to remove noise (random clusters of erroneously classified cells, including those within larger fields) and the ability to retain small fields. Based on visual examination, a sieving threshold of 900 m^2 was chosen, meaning, when subjected to an SCCA, groups of cells with less than 100 contiguous cells of the same class would be assigned to another class based on the decision rule of the SCCA, for example by using the class of the largest neighboring group.

This threshold was chosen as a compromise; smaller sieving thresholds left much noise, whereas larger sieving thresholds sieved many small fields. While 11% of the fields examined in September had area less than this threshold, this would not be the case for many of these fields if cells with edge effects were included in the area calculation. Since cells with edge effects may be classified correctly, and, since the SCCA in this study was designed to correct edge effects around fields, the actual percentage of fields sieved at this threshold should be less than 11%, which is supported by the sieving results in Section 3.1.2. Given these considerations, the importance of removing noise from the classified image, and the large variation in field size in this region which makes any sieving threshold a compromise between removing noise and sieving small fields, the chosen threshold was deemed appropriate.

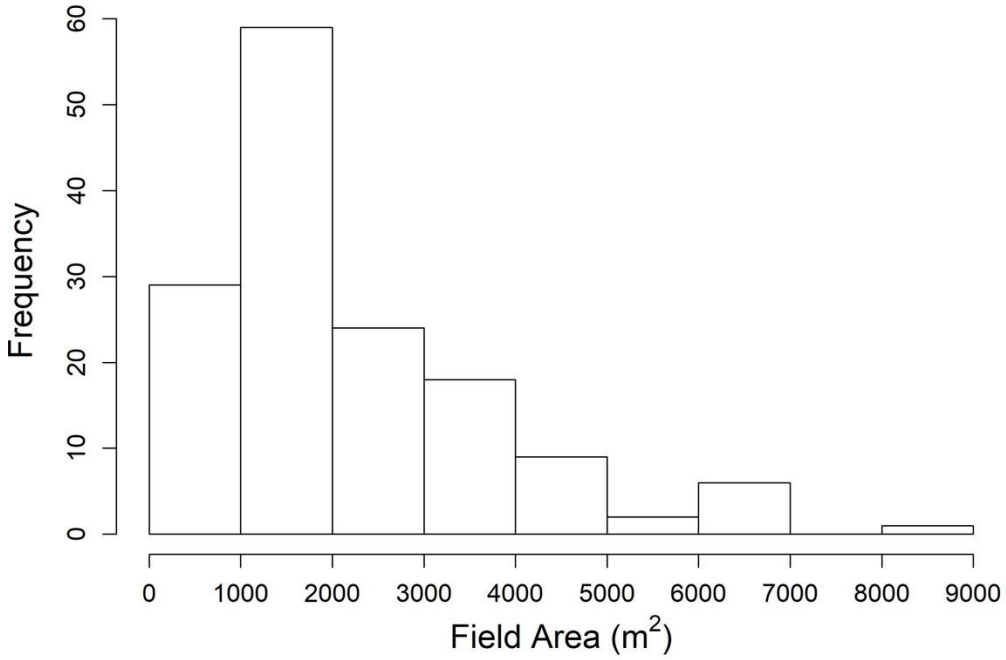

Figure 6. Distribution of field area for 148 ground reference fields sampled in the region of the study in September 2019.

While the GDAL sieving algorithm (henceforth called the GDAL-SCCA) was available for use in the QGIS software, the authors made the decision to write their own sieving algorithm in the R language. This decision was made for several reasons: (1) the GDAL-SCCA does not consider shapes of fields and fails to account for edge effects of fields, and so fields often end up with long, winding, and spatially inconsistent field boundaries that are physically unreasonable; (2) the GDAL-SCCA sieves agricultural cells into non-agricultural cells, which should not be done; (3) having all calculations done in R would be preferable, to make the system less complex for future users. Edge effects, in particular, were a significant issue in this region due to the small fields and proximity of different crops to each other, to urban areas and to terrace walls, which often resulted in spectral blending around fields.

The spatial-consistency correction algorithm designed in this study, henceforth to be called the Author-Written SCCA (AW-SCCA), was designed to account for these issues while still achieving the goal of an SCCA (removal of scattered groups of isolated, mis-classified cells). The AW-SCCA uses a pixel-based approach that, given a pixel, grows a region around that pixel and

groups the classes of the same type as the given pixel and considers that group for sieving based on its size and geometry. This makes the AW-SCCA slower than the GDAL-SCCA. However, it aims to minimize the occurrence of edge effects by using the modal value of the agricultural cells in a region constructed around the centroid of the group of cells being considered for sieving, instead of using the largest contiguous neighbor to determine the replacement cell values. The constructed region is rectangular and has an area approximately twice as large as the sieving size to capture the nearby distribution of classes. To further minimize the occurrence of edge effects and physically unreasonable field shapes, the region-growing algorithm in the SCCA avoids multiple 90-degree branches, so that long, winding fields are more often sieved.

In order to quantify the performance of the AW-SCCA, sample means of the classification accuracies for each sieving case (GDAL-SCCA sieved, AW-SCCA sieved, and unsieved) were visually compared. Additionally, the statistical significances of the pair-wise differences of the accuracies were evaluated using large sample t-tests. Visual inspection of maps produced by each sieving case was also done to compare the methods.

2.7.3 The Temporal Consistency Correction Algorithm (TCCA)

The goal of the temporal consistency correction algorithm (TCCA) was to make the sequence of crop maps physically reasonable with respect to time (temporally consistent). This was necessary due to the differences in crop distributions and ages, the filtering of young crops (Section 2.4.2), and the potential use of different classification methods (Section 2.6). For example, it is temporally inconsistent to have a different type of crop at the same spatial location in each month because it is highly unlikely for crops to change that quickly or consistently.

The TCCA considers the time series of monthly cells (each cell being at the same spatial location) and establishes a new crop growth sequence based on when cells are classified as soil, based on the assumption that soil values represent the transitions between crops. The crop class within each crop growth sequence is forced to be identical in value. To do so, the TCCA first checks each crop growth sequence to identify all crop classes that occur two or more times (i.e., have multiplicity two or greater), assuming that this means the value correctly represents the crop planted in the field during the crop growth sequence. If more than one crop has a multiplicity greater than two in the same crop sequence, ties are broken by choosing the crop with the highest average validation accuracy amongst the crops with multiplicity two or greater. If a crop class is

determined in this way, all crop values in the crop growth sequence are made identical to this value. This method was used to avoid favoring classes with higher-accuracy or that are more common as could occur in a probabilistic approach that makes replacement decisions based on accuracy. If there are no such crops, then the discriminant function, d, defined below, is used to choose the replacement cell value.

Given a set of assigned cell values (classes), say $S = \{\omega_1, \omega_2, \ldots, \omega_n\}$, where $i = 1, 2, \ldots, n$ are months and there are at least two distinct classes, then a decision rule needs to be defined to choose a single class to be mapped to all elements in S. An obvious choice for this decision rule is to choose the class that maximizes the probability of all replacements being correct. To approximate this probability, define the function $d(\omega_i) = \sum_{j=1}^{n} P(\varphi_{i,j})$, where $P(\varphi_{i,j})$ is the probability that the cell in month *j* should have been assigned the value ω_i and was mistakenly assigned the value ω_j, where ω_i and ω_j are the assigned class values in months i and j in S so j indexes both S (assigned classes) and months. In other words, for a fixed month i, the probability in each month *j* that the assigned value in month *j* was wrong and should instead be the value assigned in month i is calculated, and the probabilities are summed to give an approximation for the total likelihood that the assigned value in month i should be chosen to replace the assigned values in all other months in order to make S homogeneous. $P(\varphi_{i,j})$ was approximated by the proportion of reference data in month *j* of class ω_i times the proportion of cells of class ω_i that were classified as cells of class ω_j in month *j* (obtained from error matrices). Note that this approximation follows from the definition of conditional probability. With d so defined, $d(\omega_i)$ is calculated for all $i = 1, 2, \ldots, n$, and the potential for error is minimized by the following assignment of cell values: Choose ω_i and replace S by $S' = \omega_i, \omega_i, \ldots, \omega_i$, where $|S'| = \mathrm{n}$, if $d(\omega_i) > d(\omega_j)$ for all $j \neq i$. In the case of ties, the choice was made arbitrarily.

Finally, the TCCA regressively corrects the classification of crops onto earlier-in-time soil values for a location based on the crop age, so that fields that might have been misclassified as soil are updated to reflect the mature crops classified in later imagery. The number of months for which this regressive reclassification takes place is crop-dependent and based on the age at which the crop becomes distinguishable, determined from visual analysis of NDVI vs. age graphs, examples of which are shown in Figure 4. In the case of alfalfa, there was no age value to use for this part of the algorithm, since height, not age, was used as a predictor for NDVI. Alfalfa can reach the

harvesting stage in as little as 40 days (Anderson, 2015) and is often cut in 28–40-day intervals, especially during warmer weather (Undersander, et al., 2000). Since the temporal resolution of the images is monthly (with some discrepancies), this makes it difficult to regressively assign values accurately, as alfalfa could possibly be harvested again between two images. Moreover, while alfalfa is often grown continuously, it is sometimes grown as a cover crop, so one cannot simply assume it was present in multiple months because it was present in a single month. As a result of these considerations, it was decided that no regressive correction would be applied to alfalfa.

To quantify the performance of the algorithm, sample means of the temporally and non-temporally corrected accuracies were visually compared. Additionally, the statistical significances of the pair-wise differences of the accuracies were evaluated using large sample t-tests. It is worth noting that, if the crop mapping process is continued in the future, the sequence of maps could become very large, which could make applying the TCCA to the entire sequence infeasible, although this was not an issue in this study because the sequence was only six months long. Since most crops in this region grow to a maximum age of four to five months (Table A-1), application of the TCCA should be limited to the previous six to seven months to allow it to capture the before and after soil values and the crop growth sequence for all crop classes. As the multiplicity of two or more months for a crop type should be established in three to four months' time, the expectation is that updates to crop maps from the TCCA will be limited to three months total, so that the current month's map will be a first draft without future knowledge of what is being planted, the previous month's map will have improved accuracy as more recently planted fields can be identified, and the map from the month before that will be run through a final round of TCCA adjustments. Thus maps from four or more months prior to the current month will no longer need to be adjusted.

2.7.4 Ensemble Crop Maps

For visual interpretation of the results, monthly ensemble crop maps were created. The cells of the ensemble maps have values representing the most common class among the corresponding cells of all (108 for this study) crop maps generated using 60-40 splits of the ground reference data for that month. The ensemble process is done on a per cell basis, meaning, for a given month and spatial location, the ensemble cell at that location is assigned the most common class among all cells at that location in the crop maps of the given month. This method was chosen because it was felt to be most indicative of all classified maps, made the most of the limited ground

reference data, and allowed for the calculation of confidence intervals based on the distribution of crop classes for each pixel from the ensemble of crop maps.

2.8 The Age Estimation Algorithm (Aea)

The Age Estimation Algorithm (AEA) is applied to each of the crop maps produced by the CMA and predicts the age, in months, of all relevant cells in each crop map to produce crop age maps. Following the application of the TCCA, the AEA assumes that the length of a sequence of crop values that are equal in value and contiguous in time represents the maximum age of this crop, and then determines the age in a given month by finding the position of the month in this sequence. The AEA counts crop age in terms of integer months, since imagery and ground reference data are collected on a monthly basis. Because of this, a crop planted since the previous map is identified as being 1 month old, and crop ages are reported based on the age they will reach between the current map and the next map. The only complication is when the crop sequence within a field does not include a clear starting point, which can happen when previous images are not available from which to identify a soil value marking the change between one crop and the next. In this case, a correction factor, equal to the maximum age of the crop under consideration, as found in the reference data, minus the length of the sequence is added to each predicted age in the sequence.

Since crops are classified at a monthly resolution, average monthly water demand at all potential growth stage intervals were used to estimate monthly crop water demand (see Section 2.9.). However, since alfalfa goes through the same water demand cycle approximately every month using the average monthly water demand value for alfalfa should give a good approximation for its monthly water demand, regardless of its growth stage. Based on these considerations, it was considered unnecessary to predict the growth stage of alfalfa using the AEA. This also conveniently bypasses the need to build a more complicated growth stage prediction model specifically for alfalfa, since the prediction method used for other crops would not work for a crop that can be harvested multiple times.

The AEA was validated on all reference data cells for the crops of interest (including the unfiltered data cells but excluding alfalfa, the growth stage of which was not estimated) by comparing the reference ("observed") ages to the predicted ages using various error metrics. Only cells that were correctly classified by the CMA were used for validation. The error metrics used were mean prediction error or MPE, calculated as $mean(predicted - observed)$, mean

absolute prediction error or MAPE, and root mean square error or RMSE. The average number of cells used for validation or Sample Size was also calculated. The z-statistic was used to construct 90% confidence intervals for the means of these error metrics.

2.9 The Water Demand Mapping Algorithm (Wdma)

For each month, crop type, and crop age group (where ages were broken into monthly intervals), average monthly water demand per unit area were estimated from CROPWAT, a decision support tool developed by the Land and Water Development Division of the Food and Agriculture Organization of the United Nations (FAO; FAO, 1992). To use as inputs to CROPWAT, contemporary daily weather data from Moraes et al. (2021) were used, as were long-term monthly average reference evapotranspiration values derived from climate data from 1988-2014, cleaned and filled by Moraes et al., (2021). Additionally, four representative soil profiles for the agricultural area were obtained from Barriga (2016) and used in CROPWAT simulations. Little variation was observed in the simulations with respect to soil profiles and so results were averaged over the four soil profiles. Planting dates in the simulations were varied to obtain monthly average water demand data for each month, crop type, and crop age group combination. Monthly water demand by crop growth stage was used to match the monthly crop map product and to account for the small changes in regional weather through the year.

Each cell of each crop map was matched with the corresponding cell of the crop age map to obtain crop age data, and then matched with CROPWAT average monthly crop water demand, depending on month, crop type, and crop age group. The associated average monthly crop water demand value was then multiplied by cell area to create a cell of a water demand map. In this way, a corresponding water demand map was created for each crop map. Since the growth stage of alfalfa was not predicted (see Section 2.8.), the average monthly water demand per unit area over the entire growth cycle was used. Regional water demand for each classified crop are obtained monthly via integration of the water demand maps over area planted in each crop. The Bonferroni method and the z-statistic were then used to construct 90% joint (by month) confidence intervals for the regional crop water demands.

3. RESULTS

3.1 Crop Mapping Algorithm (Cma) Evaluation

3.1.1 Choice of Monthly Classification Methods

The RF and MaxL methods generally produced higher accuracy results than the KNN method, likely because of the small sample sizes in many of the classes, which result in only small values of K being viable (Figure 7). When considering all six months of data, the MaxL method generally produced the highest accuracies, with the largest medians and (generally) 1st and 3rd quartiles in October, November, December, and January. However RF had a larger median in February, and a slightly larger median and 1st quartile in September. Based on these results, it was decided to use the RF method in September and February, and the MaxL method for the other four months of the study period. When viewed in the context of choosing a single classification method to use for the tool the future, the results in Figure 7 imply that the MaxL method might be the best choice of those tested, given that it had higher median accuracy in four of six months. However, the MaxL method is parametric and therefore less robust than the other methods; although it resulted in higher median accuracy on the study data, it may not generalize as well to other months, wherein spectral data may be different due to different crop types and crop ages. Therefore, it may be better to investigate additional non-parametric methods that may be better suited to the data encountered by the tool.

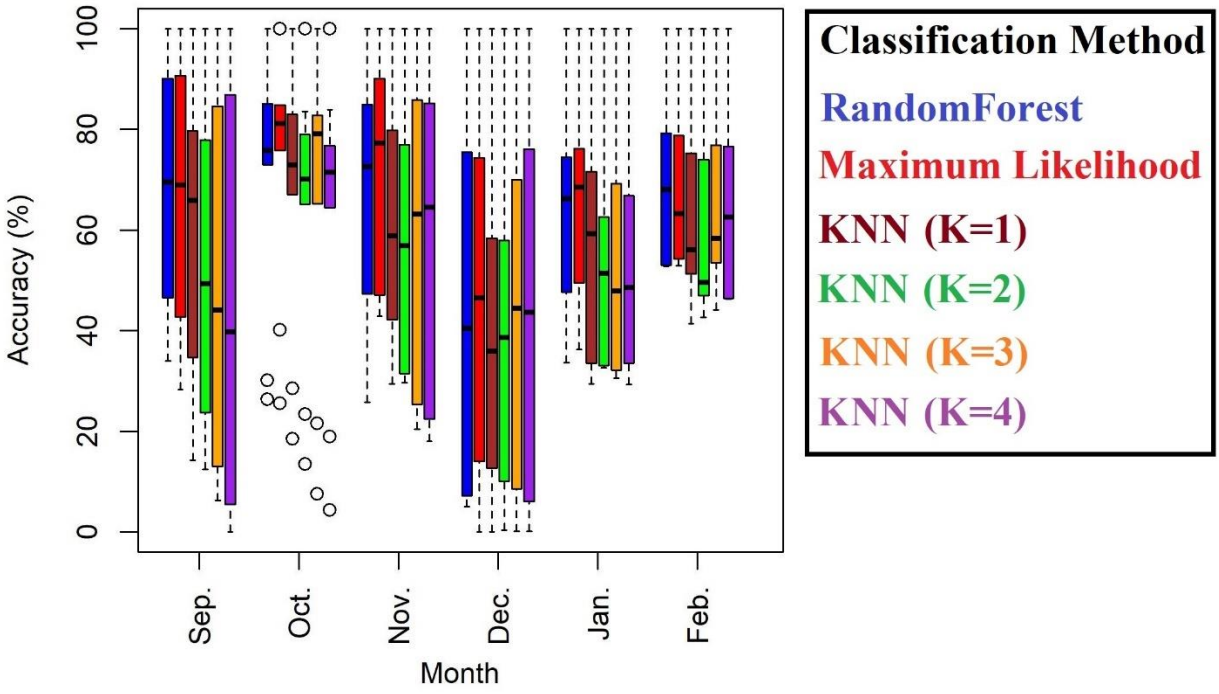

Figure 7. Boxplots of sample means of monthly accuracies for each classification method. N=400 for each month and classification method combination. Box is defined by the 25th percentile (top) and 75 percentile (bottom) of the data, with the median (50th percentile) marked as the line in the box. Bottom whisker is 1st quartile minus the inter-quartile range, and top whisker is 3rd quartile plus the inter-quartile range. Values outside of the whiskers are outliers and marked with circles. Accuracy cannot be higher than 100% or less than 0%. All 90% joint CIs were within ±5.5% of the sample mean, with 98% of the CIs being within ±4% of the sample mean.

After performing the image classification process with the chosen algorithms, the mean accuracies were generally high (greater than 70%) for alfalfa, broccoli, and garlic, whereas they were generally low (60% or less) for corn, green beans, lettuce, peas, and quinoa. Onion was classified well in October and November but poorly in September and December. The joint 90% (or greater) CIs for the monthly mean accuracies of the chosen monthly classification methods are presented in Figure 8. Note that bare soil was perfectly classified; this is because bare soil was overclassified due to the data-filtration process; however, the over-classification will be corrected using the TCCA, which will result in increased error for the classification of bare soil.

Comparing Figure 8 with Table 2 indicates that small sample sizes might lead to low accuracies, e.g., onion was classified poorly in September and December (with sample sizes of 4 and 7) but classified well in October and November (with sample sizes of 16 and 14). Lettuce was

generally classified poorly, possibly due to its small number of ground reference samples (5 to 8). There were, however, cases where crops were classified poorly even when sample sizes were not very small. For example, corn was classified poorly in December, January, and February (with sample sizes of 13, 11, and 12), yet these sample sizes are generally comparable to those of alfalfa, which was classified well by comparison. This variation in classification accuracy with moderate ground reference sample sizes could be due to difficulty in establishing certain classes due to their high spectral variability or spectral overlap with other classes. For example, corn changes color in its later growth stages and without a defined growing season corn can be present simultaneously in the study area at every growth stage. Future work should focus on quantifying spectral overlap and establishing a lower threshold for the number of ground reference samples that must be collected to maintain sufficient classification accuracy for the mapping tool. It is also worth noting that spectral overlap is more likely when there are fewer spectral bands, since less information about the complete spectral signature is available. Classification errors due to spectral overlap could possibly be alleviated by using satellite images with more bands in the future. Planet Labs, Inc. is currently testing a replacement for the existing PlanetScope 4-band sensor that has 8-bands and might improve the results from this analysis.

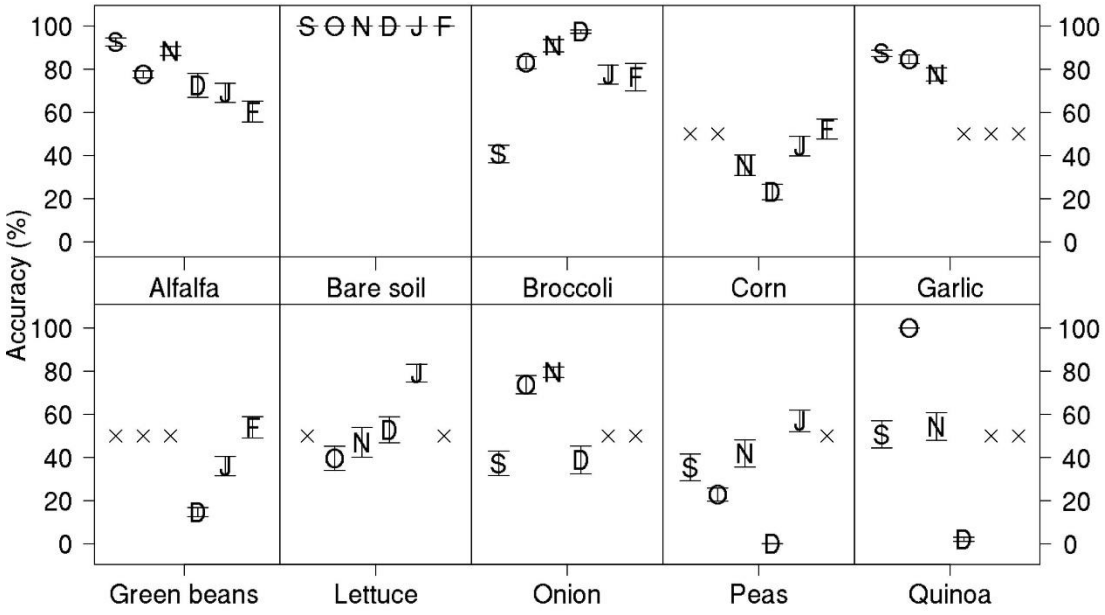

Figure 8. Sample means (letter centroid) and 90% joint (by month) CIs (whiskers) for classification accuracies of crop maps with no post-processing performed on them. Letters represent each month; months with no data are represented by the X letter. Sample means and confidence intervals were calculated with 108 samples in each case.

3.1.2 Spatial Consistency Correction Algorithm (SCCA)

Spatially, there are three goals that the AW-SCCA was designed to achieve: to remove small clusters of incorrectly-classified cells, to prevent the sieving of agricultural cells into non-agricultural cells, and to create more physically reasonable fields than the GDAL-SCCA. The first point should be a goal of any SCCA, and its success can be measured by comparing SCCA results with unsieved results, whereas the latter two points are more specific to this study, and their success can be measured by comparing the AW-SCCA to the GDAL-SCCA. Figure 9 provides a comparison of unsieved, GDAL-SCCA-sieved, and AW-SCCA-sieved cases and highlights cases where the GDAL-SCCA sieved agricultural cells into non-agricultural cells and cases where the GDAL-SCCA resulted in physically unreasonable field shapes (shown as black ellipses and rectangles overlain on the GDAL-SCCA image subset).

Both the GDAL-SCCA and the AW-SCCA did a good job of removing most small clusters of incorrectly-classified cells (Figure 9). Cases where the GDAL-SCCA sieved agricultural cells into non-agricultural cells are easily seen in Figure 9b and occurred when small fields were

contiguous to non-agricultural areas. Most cases where the GDAL-SCCA resulted in physically unreasonable field shapes were due to edge effects at field boundaries, i.e., the blending of cell spectral values at the edges of fields, where crops or bare soil transition into other classes. When crop cells blended with bare soil cells, the blended cells were often classified as garlic or onion, likely due to the sparse canopy of these crops, resulting in a border of erroneously classified garlic or onion around fields. Moreover, when two types of crops were in neighboring fields, edge effects sometimes resulted in a sort of salt-and-pepper mixture of the two classes at the boundary. These edge effect issues were mostly corrected by application of the AW-SCCA (Figure 9c).

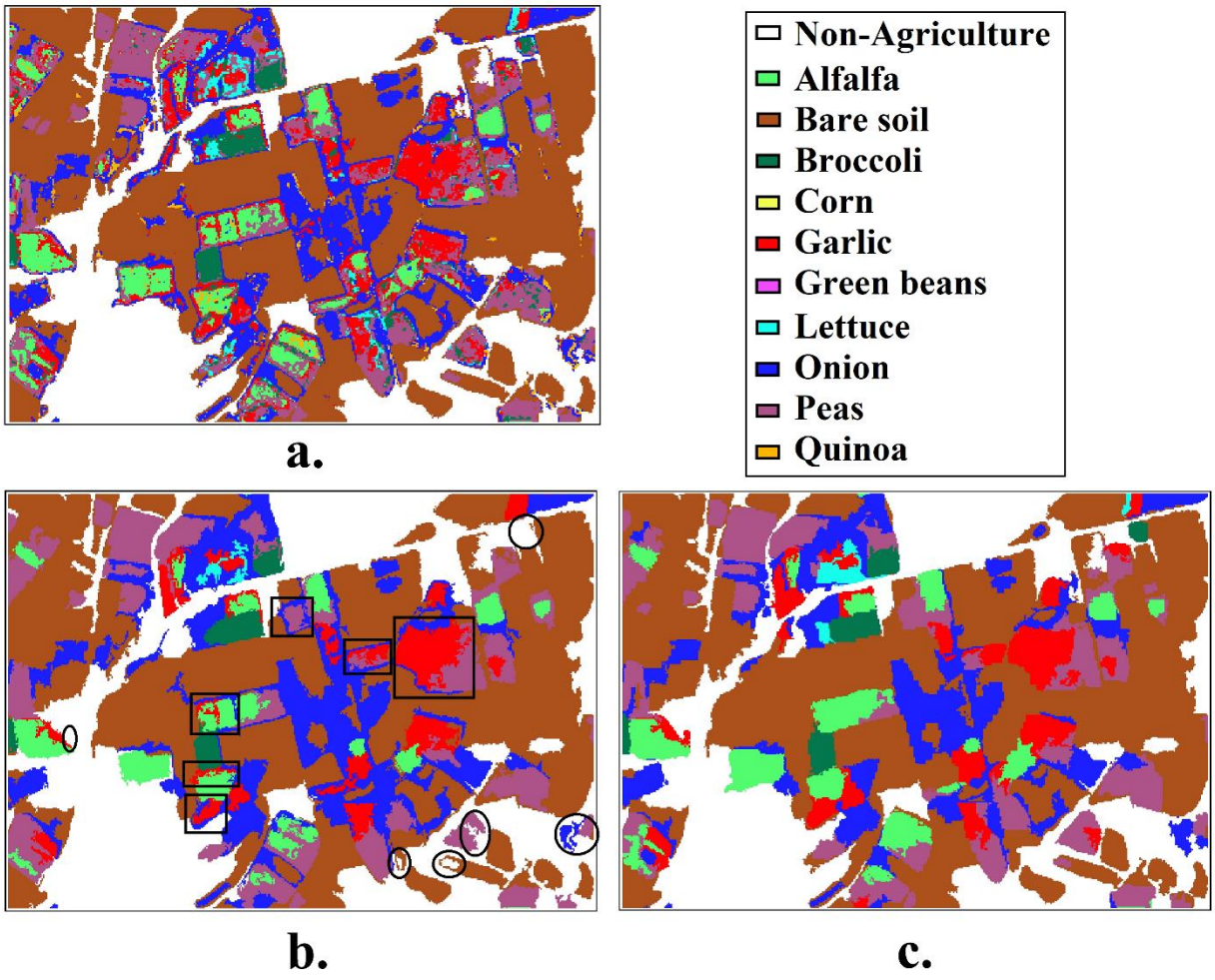

Figure 9. Image subset comparison for the three sieving cases: (a) raw unsieved classified image, (b) sieved using the GDAL-SCCA, and (c) sieved using the AW-SCCA . Black ellipses on the GDAL-SCCA image subset represent cases where the GDAL-SCCA sieved agricultural cells into non-agricultural cells and black rectangles represent cases where the GDAL-SCCA had physically unreasonable field shapes.

It should be noted that the AW-SCCA does not always reclassify groups of cells that are slightly smaller than the sieving threshold because the decision rule for choosing a replacement cell value uses the modal value of the cells contained in a region around the centroid of the group of cells being considered for sieving. This makes it possible for the algorithm to choose the same class as the replacement value if the same class is present nearby or if there is no dominant class surrounding the group of cells. While this may result in some errors, this feature was left in the AW-SCCA since it is may be beneficial in the presence of the very small fields in this region (Figure 6).

The AW-SCCA generally produces higher validation accuracies than does the GDAL-SCCA (Figure 10), though the mean differences are generally small in magnitude, usually $\leq 5\%$. Using paired t-tests to check the statistical significance of accuracy differences shows that, of the 45 crop and month accuracy combinations shown in Figure 10, there was no statistical difference (at a type I error rate of 10%) between the AW-SCCA and GDAL-SCCA accuracies in 40% of cases. However, the AW-SCCA produced significantly larger accuracies than the GDAL-SCCA in 47% of cases and significantly smaller accuracies in only 13% of cases, and so the AW-SCCA was determined to perform with higher accuracy than the GDAL-SCCA. Performing the same process but comparing the AW-SCCA and unsieved accuracies shows that the AW-SCCA accuracies were not significantly different than the unsieved accuracies in 36% of cases, significantly larger in 36% of cases, and significantly less in 29% of cases. Therefore, the AW-SSCCA was determined to improve the accuracies of the crop maps in the majority of cases. Note that t-tests were appropriate because the distributions of accuracy differences were highly normal due to the underlying random sampling procedure leading to these accuracies. .

Although there were many statistically significant changes in accuracy, there are five cases that stand out wherein the accuracy largely decreases with sieving using both the AW-SCCA and the GDAL-SCCA: alfalfa and broccoli in September, corn in November, green beans in December and lettuce in October (Figure 10). These large decreases in accuracies likely mean that entire fields were sieved into the wrong class. However, one can see that the AW-SCCA had a much smaller decrease than did the GDAL-SCCA for alfalfa in September, corn in November, and lettuce in October. Additionally, with the exception of alfalfa in September, these cases had low classification accuracies prior to sieving (less than 50%) and small sample sizes (5 to 8; see Table 2), which indicates that the raw unsieved fields were poorly classified and were therefore difficult

to correct via sieving alone. Based on visual inspection and the validation accuracy results, the AW-SCCA was considered a better choice for use in the CMA than the GDAL-SCCA.

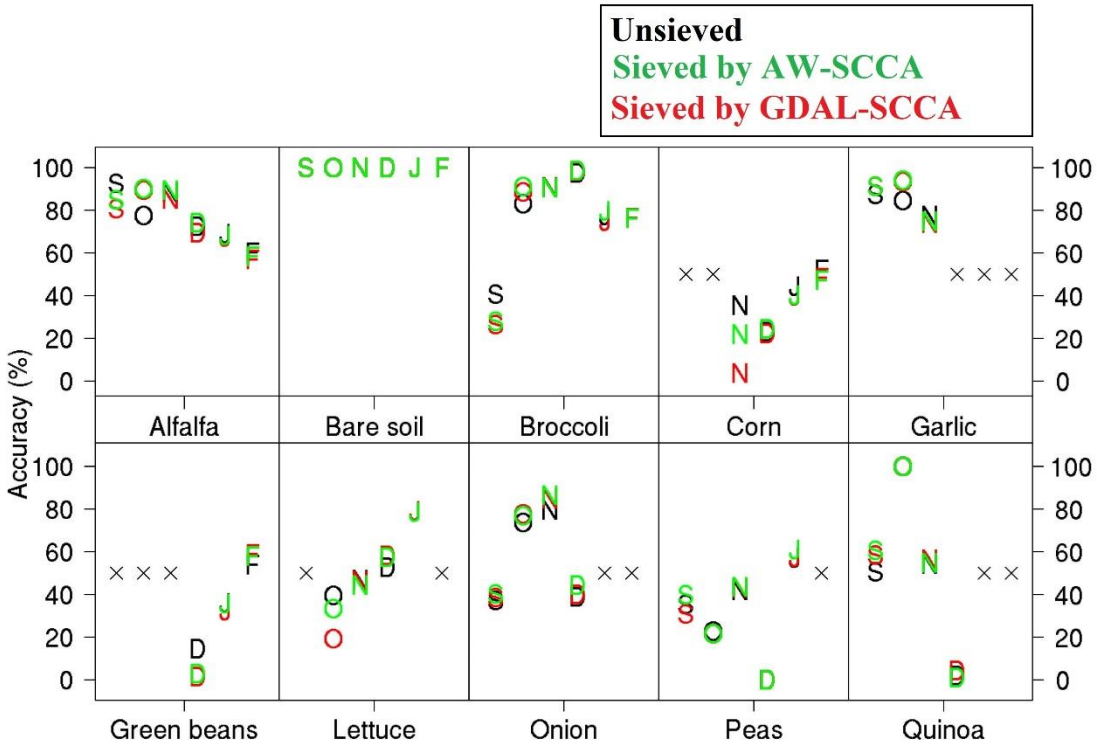

Figure 10. Samples means (letter centroid) of classification accuracies for the three sieving cases (unsieved, GDAL-SCCA sieved, and AW-SCCA-sieved). Letters represent each month; months with no data are represented by the x letter. Each sample mean was computed with 108 samples.

3.1.3 Temporal Consistency Correction Algorithm (TCCA)

Application of the TCCA improves classification accuracies for most cover types and months (Figure 11). In many cases, the improvement in accuracy is as large as 10-20% or more. Using paired t-tests to check the statistical significance of the temporally-corrected and non-temporally-corrected accuracy differences shows that, of the 45 crop and month accuracy combinations shown in Figure 11, 13% of cases had no significant differences, whereas the temporally corrected accuracies were significantly larger in 58% of cases and significantly less in

29% of cases. Crops that benefited greatly from application of the TCCA include alfalfa, corn, and onion, all of which had 10-20% or greater increases in accuracy in almost every month they were classified. Classes which generally did not benefit from application of the TCCA are (1) bare soil (the accuracy decreased in 4 of 6 cases) and (2) peas (the accuracy decreased in 3 of 5 cases). In case the case of bare soil, the loss of classification accuracy for bare soil is expected because the TCCA regressively reclassifies bare soil cells in some past months based on what was growing there in later months, and the number of months for which this reclassification takes place for each crop is only approximate (Section 2.7.3.). In the case of peas, the loss of classification accuracy is likely because peas had generally low validation accuracies in the original classification (see Figure 8) and so were unlikely to be classified correctly in two or more months and were not favored by the discriminant function. Perhaps a greater number of samples and higher temporal resolution of images could address the issue in case (1) by allowing for the determination and application of more accurate regression values. Despite these issues, the TCCA was considered a success based on its beneficial effect for most crops.

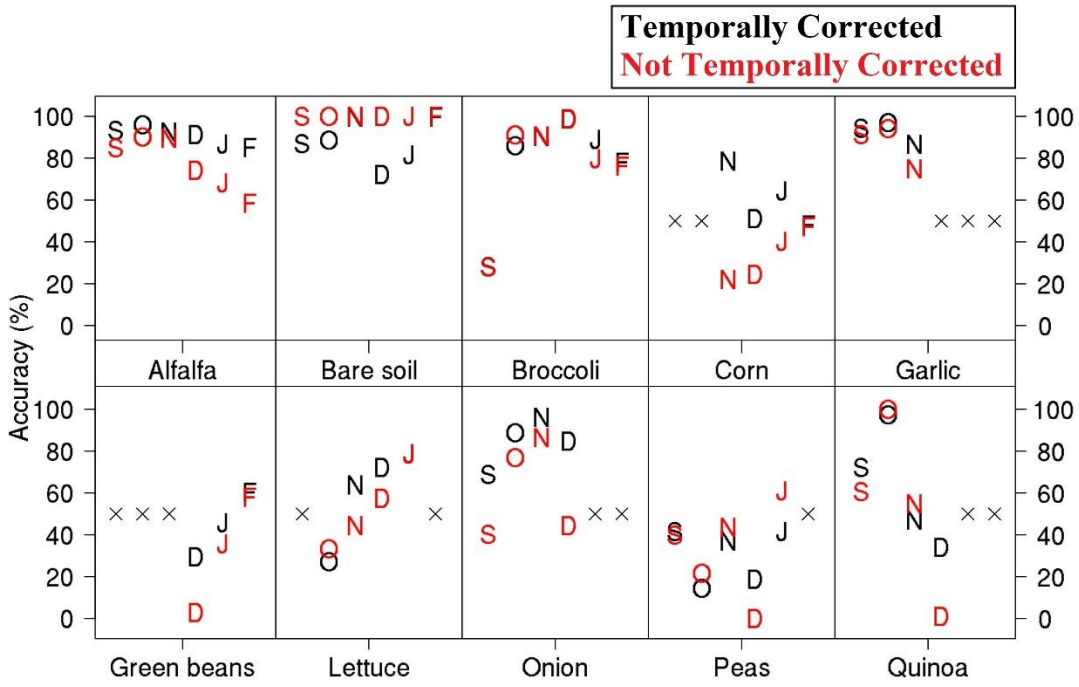

Figure 11. Sample means (letter centroids) of the accuracies of non-temporally-corrected images and temporally-corrected images. Letters represent each month; months with no data are represented by the x letter. Each sample mean was computed with 108 samples.

3.1.4 Crop Mapping Algorithm (CMA) Outputs

Of the 45 classification accuracies available after all post-processing, 22 have "high" means of 80% or greater, 10 have “acceptable” means between 60% and 79%, and 13 have “low” means below 60%, meaning that 71% of cases had acceptable or high accuracies (Table 4). The classes with the best accuracies were alfalfa (6/6 high accuracies), bare soil (5/6), broccoli (4/6), garlic (3/3), and onion (3/4). Of the 13 low accuracies, 7 of these came from peas and green beans, which were classified with low accuracy in every case except for green beans in February, which had 60% accuracy. The remainder of the classes, including corn, lettuce, and quinoa, had a mixture of low, acceptable, and high accuracies depending on the month. The monthly average of all class accuracy means was always acceptable and ranged from 61% in December to 77% in November (Table 4). However, these monthly averages are often heavily influenced by very low classification accuracies; for example, if not for the accuracies of lettuce and peas in October (27% and 14%), the monthly average of all accuracies would have been 92% rather than 74%.

Peas and green beans were the most difficult classes to classify well. This could be because they are usually planted sparsely to allow room for branching and vine growth, which limits the density of the canopy early in the growing season and because their growth patterns are similar. Difficulty in classifying quinoa and lettuce was likely due in part to the small sample sizes for these classes (5-8 for lettuce and 3-6 for quinoa; Table 2). Corn was difficult to classify well even in cases where its sample size was comparable to well-classified crops like alfalfa and broccoli; this could be because corn changes its spectral signature over time, which could make it hard to represent as a single class with such a limited sample.

Table 4. Joint (by month) 90% CIs for means of map accuracies (%) after all post-processing. The first and second numbers in the parentheses are the lower and upper bounds of the CIs. Sample means are shown above the confidence intervals. The monthly average row denotes the monthly average of all sample means in that month. Green entries have mean accuracies 80% or greater; yellow entries have mean accuracies between 60% and 79%; and red entries have mean accuracies less than 60%.

Cover	September	October	November	December	January	February
Alfalfa	93 (91,95)	96 (95,97)	92 (90,95)	91 (89,94)	87 (83,90)	85 (82,88)
Bare soil	87 (84,89)	89 (86,92)	100 (99,100)	72 (67,77)	82 (78,85)	100 (100,100)
Broccoli	28 (23,33)	86 (82,90)	90 (87,94)	99 (98,99)	89 (85,93)	78 (71,85)
Corn	NA	NA	79 (73,84)	51 (45,57)	64 (58,71)	48 (43,53)
Garlic	94 (93,96)	97 (96,98)	87 (83,90)	NA	NA	NA
Green beans	NA	NA	NA	30 (22,37)	46 (39,52)	60 (55,66)
Lettuce	NA	27 (19,35)	64 (57,70)	72 (65,79)	79 (75,83)	NA
Onion	69 (60,77)	89 (86,91)	96 (94,98)	85 (79,91)	NA	NA
Peas	42 (34,50)	14 (11,18)	37 (28,46)	19 (13,25)	42 (35,49)	NA
Quinoa	72 (65,79)	97 (96,99)	47 (37,57)	34 (24,45)	NA	NA
Monthly Average	69	74	77	61	70	74

The temporal (Figure 12) and spatial (Figure 13) trends in planted area for mapped crops can be extracted from the final ensemble crop maps. Some crops have fairly consistent coverage for five to six months, including alfalfa (averaging 9% of the agricultural area), broccoli (4%), lettuce (9%), and peas (4%) (Figure 12). Other crops experience substantial changes in planted area (Figure 12). Some, such as garlic (23% to 0%) and onion (21% to 0%) were clearly dominant in the region at the start of the 6-month study period, while other crops replaced those once harvested, including green beans (0% to 26%), and corn (0% to 11%). Some of these crops, such as alfalfa, may be grown year-round as they are economic staples, whereas other crops may be grown at particular times of the year due to regional changes in weather, water or demand that

restrict or promote their planting. It is difficult to make any general claims about the cycles of crops in the region due to having only six months of reference data and a minimal seasonal cycle.

Transitions from one crop to another can be seen clearly from the spatial maps (Figure 13). For example, garlic was a dominant crop in the center of the city in September and October, but is largely harvested by December and left fallow or replaced with corn and green beans by January and February. Onion dominates to the northwest of the city center in October, November and December before being harvested. Farther to the west, the area is dominated by lettuce and alfalfa, while dominance in the far east switches from onions to green beans and corn. The presence of spatial trends suggests the existence of hyper-local preferences in crops, perhaps due to microclimates, variable sources of irrigation water, or social networks.

Some potential issues that appear in these results are that (1) the bare soil percentages are much higher in September and February than in the other months and (2) there is no lettuce classified in February despite the crop existing in that month (Table 1). Issue (1) is easy to explain, in the case of February, because the TCCA could not regressively assign crop values to the soil values in this month, as it was the last month in the set of images. In September, it could perhaps be explained by September being a transitional month between growing periods of certain crops, although this cannot be verified due to the lack of reference data prior to September. As for issue (2), this is because most of the sampled lettuce in this month was too young to classify so those ground reference samples were filtered before classification and therefore do not appear in the resulting maps. If the mapping process had not been interrupted by COVID-19 this issue would have been fixed by temporal correction using data from the next 1 to 2 months. Issues (1) and (2) highlight the importance of the TCCA and its regressive assignment of crop values. Crops less than about 2 months of age were not classified but were regressively corrected by the TCCA using classification results from later months, which implies that 1 to 2 follow-up months are needed.

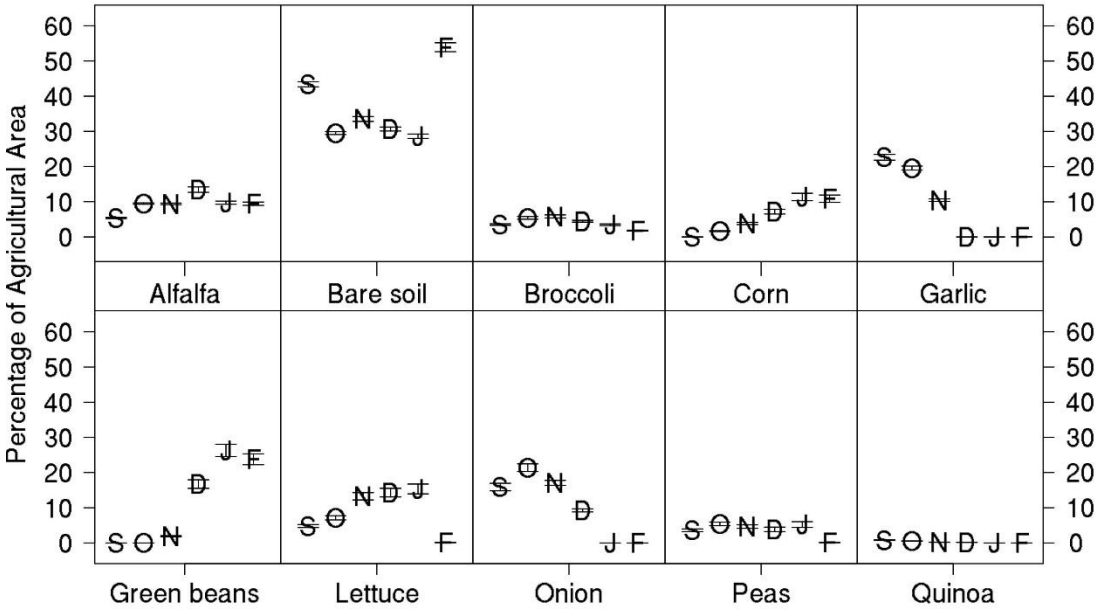

Figure 12. Sample means (letter centroid) and 90% joint (by month) CIs (error bars) for percentage of agricultural area covered by each crop based on the final ensemble crop map. Letters indicate the first letter of each month. Sample means and confidence intervals are based on 108 samples in each case.

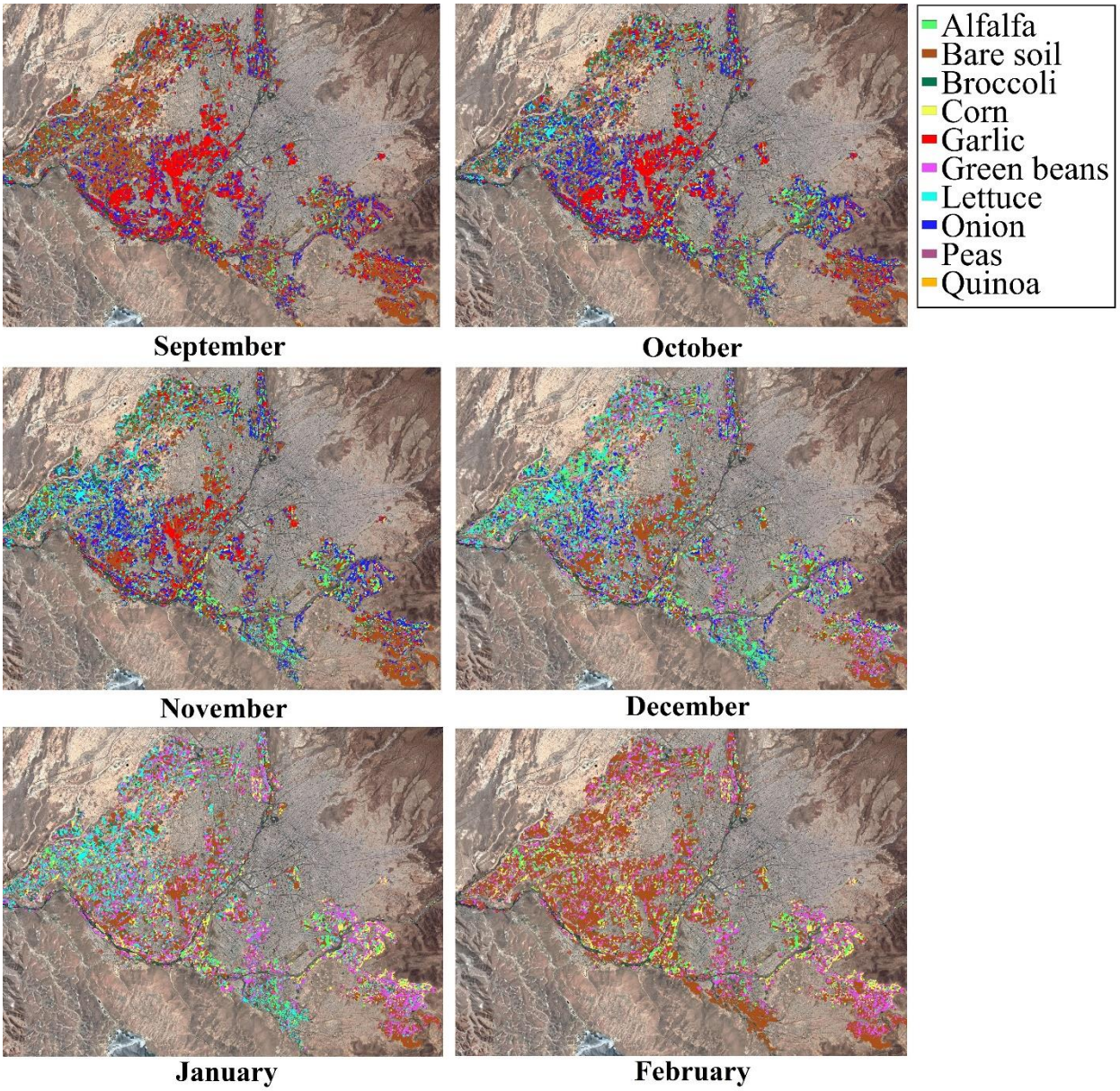

Figure 13. Ensemble crop maps with cell classes representing the modal values of 108 crop maps. The background of the crop maps is a Google Satellite map (Google, 2016).

3.2 Age Estimation Algorithm (AEA)

The MPE, MAPE, and RMSE validation metrics (Table 5) show that the AEA is generally a better predictor for the months nearer to the middle of the sequence (such as November and December) than those at the edges of the sequence (September and February). In particular, September has an MAPE of 0.90 months, which is about 1.5 times the average of the other months. Both September February have an RMSE greater than 1 month, whereas the other months have an average RMSE of 0.87 months and all have an RMSE of below 1 month. Moreover, September has an MPE of 0.33 months, and February has an MPE of –0.36 months, whereas the other months (excluding October) have MPEs closer to 0 (ranging from –0.06 to 0.17 months). Therefore, ages in September tend to be overpredicted, and ages in February tend to be underpredicted, whereas the prediction error in other months tends to be nearly random.

These results are expected, given the way the TCCA works. In September, there is no knowledge of previous months and a correction factor has to be applied when estimating ages in certain cases, leading to errors and a tendency to overpredict ages since the correction factor assumes crops will always be harvested at their maximal age as seen in the reference data (Section 2.8.). In February, there is no knowledge of future months and the regressive reassignment of crop values to bare soil cannot be performed, leading to much smaller sample sizes (Table 5) and a tendency for the observed ages used for validation to be higher and thus underpredicted. While these issues can also affect other months, they are most likely to affect the months at the ends of the sequence, and the results are therefore most pronounced in these months. These considerations imply that adding more months to the process can be expected to lead to higher-quality results, like those seen in October, November, December, and January.

Since the resolution of the predicted values is 1 month, and the resolution of the observed values is 0.5 months, there are both "resolution error" and prediction error in the AEA. The ideal MPE value is still 0, which would indicate a tendency for random error. However, the ideal MAPE and RMSE are not 0. Assuming each observed age is at least 0.5 months and has an equal chance of being a multiple of 0.5 months or being an integer, then a predictor with no prediction error would have an expected MAPE of 0.25 months and an expected RMSE of $\sqrt{1/8}$ or about 0.35 months (this is easily verified from the formulas or numerical simulation). Similarly, in this situation a predictor that has an equal chance of –1, 0, and 1 months of prediction error would have an expected MAPE of 0.75 months and an expected RMSE of $\sqrt{19/24}$ or about 0.90 months.

Comparison of these theoretical results with the results in Table 5 shows that the MAPE and RMSE are less than or equal to 0.75 and 0.90 months in October, November, and January (with MAPE values of 0.59, 0.61, and 0.55 months and RMSE values of 0.83, 0.90, and 0.79 months), suggesting reasonable prediction error given the resolution of the data and that most crops grow to be about 4.5 to 5 months of age (Table A-1). The MAPE and RMSE values in December are 0.69 and 0.96 months, while its MPE is only 0.17 months, suggesting that this month has about 1 month of nearly random prediction error. The predictive power of the model suffers in September and February, shown by the RMSE values of 1.10 and 1.02 months in September and February and the MAPE value of 0.90 months in September. This is due to a lack of knowledge of past and future months, as discussed previously. However, collecting more sequential monthly data in the future could greatly lower the percentage of maps with poorer predictions.

Table 5. 90% confidence intervals for Age Estimation Algorithm (AEA) validation metrics. Metrics are mean prediction error (MPE), mean absolute prediction error (MAPE), root mean square error (RMSE), and sample size. MPE, MAPE, and RMSE have units of months, and sample size has units of pixels. The first and second numbers in the parentheses are the lower and upper bounds of the confidence interval, and the number above the confidence interval is the sample mean.

Validation Metric	September	October	November	December	January	February
MPE	0.33 (0.33,0.34)	0.38 (0.38,0.39)	-0.06 (-0.07,-0.05)	0.17 (0.16,0.18)	0.10 (0.09,0.12)	-0.36 (-0.39,-0.33)
MAPE	0.9 (0.89,0.90)	0.59 (0.59,0.60)	0.61 (0.60,0.61)	0.69 (0.68,0.69)	0.55 (0.54,0.56)	0.68 (0.65,0.7)
RMSE	1.10 (1.09,1.10)	0.83 (0.83,0.84)	0.90 (0.89,0.91)	0.96 (0.95,0.97)	0.79 (0.77,0.80)	1.02 (0.98,1.06)
Sample Size	5423 (5399,5448)	6875 (6839,6911)	5934 (5896,5973)	3128 (3092,3163)	2666 (2630,2702)	1460 (1438,1482)

3.3 Water Demand Mapping Algorithm (WDMA)

As water demand is a function of crop and crop stage, the monthly water demand by volume for each crop type (Figure 14) roughly follows the planted area for each crop type (Figure 12). Differences in the temporal trends of water demand volume and crop area are due to the variation in water demand based on crop type and age. For example, corn area increases fairly

linearly from September to February (Figure 12), but water demand increases almost exponentially from September to December, is relatively constant in January, and decreases substantially in February. Corn water demand does increase exponentially in the first two months after planting, and stays high from about 2 to 3.5 months after planting. After that, the crop senesces, and water demand drops before harvest. Peak water demand for green beans is on the order of 1.5 to 2 months after planting, thus green bean water demand peaks in December, even though planted area does not peak until January. Crop water demand falls off quickly as beans are typically harvested at about 2 months. The biggest water demand in September and October is from garlic, followed by onion. In November, garlic, onion, and lettuce water demand is similar as garlic and onion are being harvested, and lettuce is reaching maximum extent. In December, green beans, lettuce, and corn dominate. Green beans and corn continue to be the big users of water in January. February water demand estimates are low as many fields were transitioning and no follow-up data were collected to confirm what was planted in many of the fields that were left classified as bare soil with no water demand (Figure 12).

Spatial correlation between crop type and water demand is observed as well (Figures 13 and 15), with the highest demand occurring in September and October in the central part of the city where garlic is the dominant crop. The impact of onion on water demand in those months is more difficult to see as it is more distributed across the entire area, but it does contribute to higher water demand in the north-central and eastern parts of the agricultural area. Water demand in the central part of the city decreases in November and December as crops, especially garlic, are harvested. Higher water demand appears outside the city center, to the northwest and southeast, as crops planted in September and October in those regions mature and require additional water. Primary water demand in January comes from green beans, lettuce, and corn, and most of these crops reached peak water demand in the previous month resulting in lower water demand across the map. Water demand maps for January and February also suffered from lack of information on what crops were planted in later months. For January, this is most clear in the central part of the city, but the problem affects much of the region for February as many of the fields classified in the January image were harvested resulting in classification as bare soil. Many crops planted in January and February would have reached the NDVI threshold for classification in March, so the TCCA would have allowed them to be classified in the earlier imagery, and their water demand could have been included in the analysis, providing more accurate estimates for regional water

needs. Since spatial trends are present, it could be useful for future studies to partition the region of classification and to calculate water demand for each subregion. When combined with data about local sources of irrigation water, such an analysis would allow for water management at a very fine level.

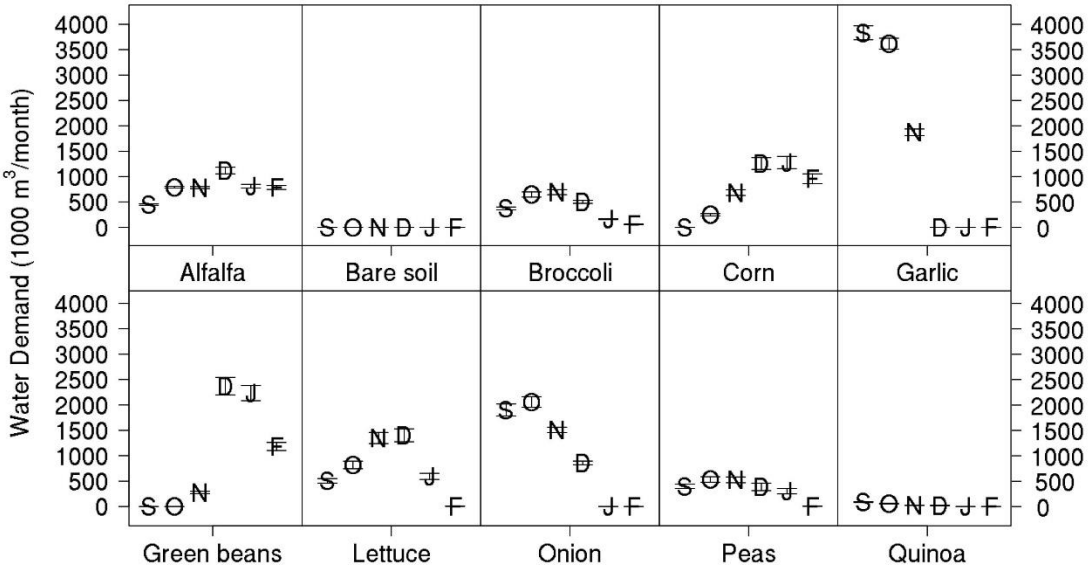

Figure 14. Sample means and 90% joint (by month) CIs for water demands of crop maps after all post-processing. Letters represent each month. Sample means and confidence intervals are based on 108 samples in each case.

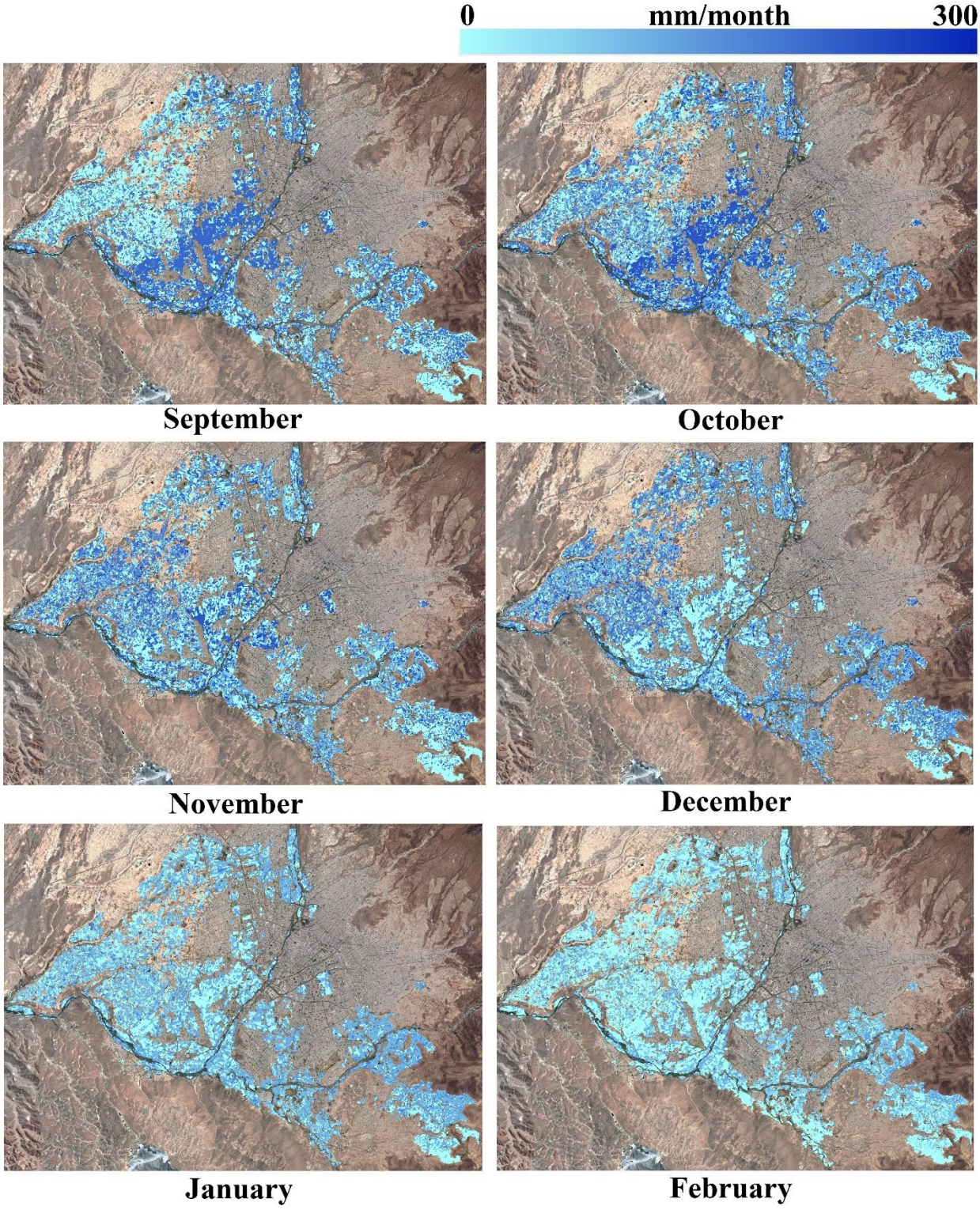

Figure 15. Water demand maps corresponding to ensemble crop maps in Figure 13. The background of the water demand maps is a Google Satellite map (Google, 2016).

4. DISCUSSION

4.1 Ground Reference Data Collection

Perhaps the most well-known crop mapping service is the Cropland Data Layer (CDL) (United States Department of Agriculture, n.d.), which has been providing annual crop maps of the continental United States since 2008. As the CDL has a long history in the field of remote sensing, it is worthwhile to compare the methodologies and results in this project with those of the CDL.

Although the developed data collection method served the purpose of this paper, as evidenced by the CMA and AEA results, there were issues with the method. The biggest issue was that, even with a team of local agronomists, it was difficult to collect sufficient data for crop mapping on a monthly basis. This difficulty is highlighted in the temporal spread of the ground reference samples (Figure 2) and the fairly small monthly sample sizes given the number of crops and classification features (Table 1). Difficulties in collecting reference data in the region include the distances between primary agricultural areas, negotiating urban traffic, the reliance on taxis and public transportation to reach sample sites, and the dominance of thousands of small, terraced fields cultivated independently by families and small landowners. This makes it especially challenging to implement a randomized, spatially-representative sampling scheme. Plans to conduct a preliminary sampling blitz, organizing a larger number of volunteer observers to canvas a larger number of fields to produce a pool of data large enough to test the effect of ground reference sample size on classification accuracy, were put on hold due to the COVID-19 pandemic. Ground reference sampling was restarted in September 2021 with a small team, but the pandemic continues to be a major issue.

Another challenge was estimating crop age by visual inspection. This was done with good accuracy in this study by using a volunteer team of agronomy students led by an experienced agronomist. However, the need for accurate ground reference data ages limits the pool of people who can volunteer or be hired to perform the data collection, and this ought to be considered when planning future data collection. This issue could perhaps be circumvented in the future by developing a guide to help users estimate crop age by relating age to other variables such as general appearance, height, or number of leaves.

There are, of course, alternatives to manually collecting reference data. For example, the CDL creators directly obtain agricultural ground reference data polygons from the Farm Service Agency (FSA) and obtain non-agricultural ground reference data from the National Land Cover Database (NLCD) (Johnson and Mueller, 2010). The FSA and NLCD provide a large and fairly up-to-date (in the case of the NLCD) or contemporary (in the case of the FSA) source of ground reference data for the CDL. The NLCD is itself a classified map and is therefore subject to its own inaccuracies, which may be propagated by the CDL. Unfortunately, there are no known similar programs in place in Peru. Moreover, working with farmers to produce a Peruvian analogue of the FSA data may prove challenging due to the different socioeconomic, educational, and technical status of farmers in Peru as compared with those in the United States. Nevertheless, developing infrastructure and governmental programs in Peru may lead to better sources of ground reference data in the future, although the method used in this project may be the best option at the present.

4.2 The Crop Mapping Algorithm (Cma)

The CDL generally has area-weighted average crop accuracies around 90% (Lark et al., 2021). However, these accuracies are heavily weighted towards the most representative crops in the United States (wheat, soybeans, and corn), which tend to have among the highest classification accuracies in the CDL (Lark et al., 2021; Boryan et al., 2011). In fact, in the 2009 CDL, many less-representative crops had producer's accuracies of 40% or less, despite the nearly-ideal reference data collection method used in the CDL and the large size and homogeneity of the fields in the United States. These results clearly show the theoretical limitations of classification: there are only so many classes that one can classify before the spectral overlap reduces class accuracies to unacceptable levels. Given the scope of this study and the fact that this is the first iteration of the CMA, whereas the CDL has undergone many improvements since its creation in 1997 (Boryan et al., 2011), the CMA results seem acceptable. Indeed, the accuracies for several major crops (alfalfa, broccoli, garlic, and onion) are generally comparable to the accuracies of the major crops in the CDL (Table 4), whereas the other crops suffered from lower accuracies, as did the minor crops in the CDL.

While one of the goals of this study was to create water demand estimates, the crop maps themselves represent valuable tools. The many uses of the CDL have been enumerated by Mueller and Harris (2013) and include applications to agricultural economics, food security, land

conversion and transition, and crop rotations. It is possible that, in the future, the CMA will play a similar role in Peru as the CDL has historically played in the United States. Certainly, if the CMA is continued and more months of crops are mapped, other researchers could use the crop maps in ways similar to how the results of the CDL have been used historically. For example, a long-term set of CMA crop maps would enable one to look for common crop rotations in the region by examining the agricultural cover of different crops over time. If the crop rotations in the region are less than ideal, it could be beneficial for the region to help educate farmers on the importance of proper crop rotations, for example, the benefits that they bring to soil and environmental health. Another example is that a long-term set of CMA crop maps could be used to determine what, if any, dominant crop cycles occur in the region. For example, one might find that garlic and onion are often dominant from July to November and are generally followed by corn and green beans from December to March. This knowledge could help plan for market fluctuations in crop availability.

4.3 The Age Estimation Algorithm (Aea)

A review of current literature suggests that it is rather uncommon to estimate crop age using remote sensing methods. There are likely several reasons for this, including: (1) in most regions there is a fairly consistent and widespread planting pattern for crops; for example, in the Midwestern United States, crops are generally planted in the early summer and nearly all farmers subscribe to the chosen planting date for a given year. Therefore, many authors interested in predicting water demand already have a good idea of the crop ages in their region of interest. Those interested in mapping water demand can therefore obtain crop evapotranspiration (ET) coefficients from tables after a review of planting patterns for the season (this was done, for example, by Casa et al., 2007). However, this is not the case in this region. (2) for those interested in mapping water demand, it is possible to use remote sensing data to predict crop ET coefficients by regressing the crop ET coefficients on variables such as vegetation indices (Kamble et al., 2013; El-Shiberny et al., 2014; Reyes-Gonzalez, 2018). However, creating such a regression model for this region would require ground reference data for crop ET coefficients, which the authors of this study did not have. Additionally, the small field size limits the availability of remote sensing products useful for the area.

The most similar methodology found in literature is that of Di et al. (2015) who developed a remote sensing-based crop growth stage model for nine different crops in the U.S. They used a double-sigmoid model, the CDL, and a crop model to predict phenological state with RMSE errors ranging from 8% to 14% with an average RMSE of 11%. The study did not directly predict crop age, but instead predicted phenological stage in an environment where there is a dominant growing season. Additionally, as RMSE is scale-dependent, comparisons to the AEA in this study cannot be made. Although the results in this study could not be compared to those in other studies, the results of the AEA were deemed acceptable based on the external validation metrics.

4.4 The Water Demand Mapping Algorithm (Wdma)

Remote sensing is commonly used to find crop water demand estimates, often in regions with limited freshwater resources where management is crucial (Casa et al., 2007; El-Shirbeny et al., 2014; Reyes-Gonzalez et al., 2018; El-Magd et al., 2005). The most common method of doing so is by using crop maps along with the equation: $ET_C = K_C \cdot ET_R$, where K_C is the crop ET coefficient, ET_R is a reference crop evapotranspiration, and ET_C is the evapotranspiration for the crop. K_C values are typically found by using tabulated values derived from field experiments along with knowledge of crop ages but can also be derived by regressing locally-relevant K_C on some remote-sensing-based values such as NDVI (Kamble et al., 2013; El-Shiberny et al., 2014; Reyes-Gonzalez, 2018). The reference crop ET values are generally found by using the FAO Penman-Monteith equation (Allen et al., 1998), with remote sensing or in-situ meteorological data as inputs. This was fundamentally the same process used in this study. The key difference was that CROPWAT was used to obtain monthly average ET_C values for crops at monthly age intervals, and these values were combined with crop maps and crop age maps to create the water demand maps.

Crop water demand estimates can help manage freshwater resources, for example by predicting water demand and its spatial distribution. Estimates can help users, such as farmers and policymakers, plan planting cycles, plan and design infrastructure, and allocate water resources. However, caution should be used when interpreting the water demand values given in this study, due to the propagation and introduction of errors at the various steps in the water demand mapping process. It is recommended that the water demand values in this study be used as supplementary or exploratory only, and not as a sole basis for decision making. Indeed, prior to their use in

decision-making, it would be beneficial to perform validation on the final water demand results. This could be done by obtaining in-situ ET measurements at various fields, integrating them to obtain water demands for fields, and comparing them to the predicted values; however, this was beyond the scope of the study. For an overview of in-situ ET measurement techniques, see Allen et al. (2011). Another option is to monitor the irrigation distribution system in the region and make comparisons to the predicted water demand.

5. CONCLUSIONS

The goal of this study was to develop a monthly agricultural crop and water-demand mapping algorithm for the agriculture around the city of Arequipa, Peru. This goal was accomplished in three steps. For the first step, a crop mapping algorithm was created, which incorporates supervised classification methods, as well as spatial- and temporal-consistency correction algorithms. Secondly, a crop growth-stage prediction algorithm for the crop maps was created. Finally, an algorithm for creating agricultural-water-demand maps using the results of the steps 1 and 2 and auxiliary data was made.

The developed algorithms were tested on 6 months of reference data, which were collected monthly from September 2019 to February 2020. The crop mapping algorithm was shown to create maps with acceptable accuracy, with 5 maps having mean monthly classification accuracies of 69% or greater, with the 6th month having a mean monthly accuracy of 61%. The growth-stage prediction algorithm was found to be an acceptable predictor of age, mean absolute prediction errors being between 0.55 and 0.69 months in 5 out of 6 months, with the 6th month having an MAPE of 0.90 months while being inherently difficult to predict due to the lack of maps prior to it. Water demand maps were produced with high spatial (3.0 m) and temporal (monthly) resolution, allowing for a detailed look at the agricultural water demands of the region, as well as easy integration of areas of interest.

This study produced methods to create useful information for policymakers and land managers in the Arequipa region. The water demand estimates of crops at different growth stages can be used to help guide planting patterns, manage water resources, and manage agricultural land, which activities are crucial in this region. This study also laid the foundation for future crop water demand mapping in the region by creating a functional methodology that can be used to create more water demand maps on a monthly basis and is easily improved upon. The method is expected to become more robust as more months of ground reference data are collected, due to an improved temporal consistency correction algorithm. Moreover, this study is proof that it is possible to successfully map crops in this region and it serves as a unique case study for a novel area of remote-sensing research.

6. FUTURE WORK

This research was conducted as part of the Arequipa Nexus Institute (Arequipa Nexus Institute - Purdue University/UNSA), which is a partnership between the Universidad Nacional de San Agustín (UNSA) in the city of Arequipa, Peru and Purdue University in the city of West Lafayette, Indiana. One goal of the Institute is to increase the capacity of UNSA faculty to conduct independent research. Because of this, all GIS work was done in QGIS (QGIS Development Team, 2018) and all algorithm development was done in RStudio (R Core Team, 2019). These are free software packages and were chosen to reduce the costs as much as possible for the future users of the system and to simplify the map creation process. Additionally, a prototype web tool was developed to make the information of this study accessible to stakeholders. The "Cultivista" (link embedded) web tool freely provides crop maps, mapping accuracies, crop areas, and crop water demand results for the city of Arequipa and its surrounding agriculture. Planning for continued monitoring of agriculture and water demand for the region is in place, as well as expansion of the tool to other agricultural areas of the department.

If the results produced in this study are to be created on a monthly basis, it would be beneficial to reduce repetitive, time-consuming processes. Moreover, the results ought to be reproducible by users with less technical experience and with less powerful computing hardware. There are many possibilities to streamline the process used in this paper. One major way of reducing labor input to the process would be to try to get farmers involved in the ground reference data collection process. Perhaps farmers would be willing to report what they are growing in their field in return for agronomic advice or help accessing and utilizing the results of the study, to which their data would contribute. A second major way of reducing labor input would be to use code to automate many of the steps. For example, in this study, ground reference data had to be plotted in GIS software and the fields corresponding to the points delineated manually, which was time consuming and would only become more so if more data are collected. During this process, the spectral data were also examined to check for outlying measurements (e.g. dead or diseased fields). This process could be automated by manually creating a vector layer of all the fields in the region (which would be a significant upfront cost but not infeasible due to the size of the region), then using the GPS coordinates and direction of the field data to automatically select the field. Spectral data could then be subject to various quality control checks to remove outlying

measurements automatically. Additionally, graphical analysis checks could likely be replaced by code to reduce user input and required technical knowledge. If the end users of the product are typically less technically experienced, it may not be worthwhile to present confidence intervals and instead simply present sample means. A lower number of runs may also be acceptable to largely reduce computational demand, although it would be useful to first establish how much confidence in the results is needed.

There are also ways in which the results of this study could possibly be improved. One possibility would be to use satellite images with comparable spatial resolution yet more spectral bands, which would provide more details on the spectral signatures of crops and could result in higher accuracies. The classification method generally most suited for the data encountered by the tool should also be determined. It is recommended that the same method is used for every month to make the tool simpler to use and more robust. Determining the most suitable method may require a larger ground reference dataset and significant exploratory analysis. The results in this study suggested that the MaxL classifier was well-suited for the data. However, this is a parametric method and may not be suitable for this tool because significant changes in monthly data may occur. Additional non-parametric methods that may be more suitable for the tool and can produce accuracies comparable to the MaxL method should be investigated. Preliminary investigations suggest that the support vector machine method may be a good choice.

Finally, it is worth considering when the newest month of results should be released. End users may want the results within the same month, but these results are highly inaccurate due to the inability to map young crops and the need for data from later months for use in the temporal consistency correction algorithm. These inaccuracies were seen in the highly-underestimated water demand in the month of February, which was the last month of the study.

APPENDIX A

Table A-1. NDVI thresholds, approximate age (height) thresholds corresponding to the NDVI thresholds, and maximum observed ages (heights) for select crops. Sample fields with means of cell NDVI values less than the NDVI threshold associated with the crop growing in the field were not used in the classification process. NDVI thresholds were defined by, for each crop, plotting NDVI vs. growth stage and identifying the NDVI value above which NDVI varied minimally with respect to growth stage, which approximately represents when full canopy structure has been developed and the crop is fully distinguishable from soil.

Crop	NDVI Threshold	Approx. Age/Height Threshold	Maximum Observed Age/Height
Alfalfa	0.60	20 cm	81 cm
Broccoli	0.50	2 months	4.5 months
Celery	0.50	2 months	4 months
Corn	0.50	2 months	5 months
Garlic	0.45	2 months	5 months
Green beans	0.45	2 months	4 months
Lettuce	0.35	2 months	4 months
Oats	0.50	1.5 months	4 months
Onion	0.35	2.5 months	5 months
Peas	0.45	2.5 months	4.5 months
Quinoa	0.40	1.5 months	5 months

REFERENCES

Allen, R.G.; Pereira, L.S.; Howell, T.A.; and Jensen, M.E. Evapotranspiration information reporting: I. Factors governing measurement accuracy. *Agricultural Water Management 98* ***2011****,* pp. 899-920.

Allen, R.G.; Pereira, L.S.; Raes, D.; and Smith, M. Crop evapotranspiration guidelines for computing crop water requirements. *FAO irrigation and drainage paper No. 56* ***1998***, p. 300.

Anderson, B. (2015). Managing 1st Year Alfalfa. Available at: https://cropwatch.unl.edu/managing-1st-year-alfalfa (accessed 1 January 2020)

Barriga, R. A. A., 2016. Preparation of the Soil Study and the Classification of Lands Because of Its Capacity of Greater Use in the Region. Report of the Capacity Development Project in Ecological Zoning Economic for Territorial Management in the Arequipa Region.

Boryan, C.; Yang; Z.; Mueller, R.; and Craig, M. Monitoring US agriculture: The US Department of Agriculture, National Agricultural Statistics Service, Cropland Data Layer Program. *Geocarto International* ***2011***, pp. 1-18.

Breiman, L. Random Forests. *Machine Learning* ***2001****, 45*, pp. 5-32.

Casa, R.; Rossi, M.; Sappa, G.; and Trotta, A. Assessing Crop Water Demand by Remote Sensing and GIS for the Pontina Plain, Central Italy. *Springer, Water Resour Manag* ***2007***.

Dahal, D.; Wylie, B.; and Howard, D. Rapid Crop Cover Mapping for the Conterminous United States. *Scientific Reports* ***2018****, 8:8631*.

Di, L.; Yu, E.G.; Yang, Z.; Shrestha, R.; Kang, L.; Zhang, B.; and Han, W. Remote sensing based crop growth stage estimation model. *IEEE IGARSS* ***2015***.

Doraiswamy, P.C.; Akhmedov, B.; Beard, L.; Stern, A.; and Mueller, R. Operational Prediction of Crop Yields Using MODIS Data and Products. *2007 International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences Special Publications (In Press)* ***2007***.

Edwards, A.W.F. The History of Likelihood. *Int. Stat. Rev* ***1974****, Vol. 42, No. 1*, pp. 9-15.

El-Magd, I.A. and Tanton, T. Remote sensing and GIS for estimation of irrigation crop water demand. *International Journal of Remote Sensing* ***2005****, Vol. 26., No. 11.*, pp. 2359-2370.

El-Shirbeny, M.A.; Ali, A-E.M.; and Saleh, N.H. Crop Water Requirements in Egypt Using Remote Sensing Techniques. *Journal of Agricultural Chemistry and Environment* ***2014****, 3,* pp. 57-65.

Epicollect5. Available online: https://five.epicollect.net/ (accessed 1 July 2019)

FAO. (1992). *CROPWAT, a computer program for irrigation planning and management by Martin Smith*. FAO Irrigation and Drainage Paper No. 26. Rome.

Fix, E. and Hodges, J. Discriminatory analysis. Nonparametric discrimination: Consistency properties. In *Technical Report 4;* USAF School of Aviation Medicine: Randolph Field, Texas, 1951.

Google. (2016). 地图数据 ©2016 Google

Gyorfi, L. and Gyorfi, Z. An Upper Bound on the Asymptotic Error Probability of the k-Nearest Neighbor Rule for Multiple Classes. *IEEE Transactions on Information Theory **1978**, vol. IT-24*, pp. 512-514.

Heung, B.; Ho, H.C.; Zhang, J.; Knudby, A.; Bulmer, C.E.; and Schmidt, M.G.. An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. *Geoderma **2016**, 265*, pp. 62-77.

Howard, D.M.; Wylie, B.K.; and Tieszen, L.L. Crop classification modelling using remote sensing and environmental data in the Greater Platte River Basin, USA. *International Journal of Remote Sensing **2012**, 33:19*, pp. 6094-6108.

Huang, X.; Liu, J.; Zhu, W.; Atzberger, C.; and Liu, Q. The Optimal Threshold and Vegetation Index Time Series for Retrieving Crop Phenology Based on a Modified Dynamic Threshold Method. *Remote Sens. **2019**, 11, 2725*.

INEI. *Compendio Estadístico 2018;* Instituto Nacional de Estadística e Informática: Lima, Peru, 2018.

Johnson, D.M. and Mueller, R. The 2009 Cropland Data Layer. *Photogrammetric Engineering & Remote Sensing **2010***, pp. 1201-1205.

Kamble, B.; Kilic, A.; and Hubbard, K. Estimating Crop Coefficients Using Remote Sensing-Based Vegetation Index. *Remote Sens. **2013**, 5*, pp. 1588-1602.

Lark, T.J., Schelly, I.H.; and Gibbs, H.K. Accuracy, Bias, and Improvements in Mapping Crops and Cropland across the United States Using the USDA Cropland Data Layer. *Remote Sens., **2021**, 13, 968*

Lawrence, R.L; Wood, S.D; and Sheley, R.L. Mapping invasive plants using hyperspectral imagery and Breiman Cutler classifications. *Remote Sensing of Environment **2006**, 100* (3), pp. 356-362.

Li, A.; Liang, S.; Wang, A.; and Qin, J. Estimating Crop Yield from Multi-temporal Satellite Data Using Multivariate Regression and Neural Network Techniques. *Photogrammetric Engineering & Remote Sensing **2007***.

Loveland, T.R., B.C. Reed, J.F. Brown, D.O. Ohlen, Z. Zhu, L.Yang, and J.W. Merchant, 2000. Development of a global land cover characteristics database and IGBP DISCover from 1 km AVHRR data. *Int. J. Remote Sensing, 21(6–7),* 1303– 1365

Lu, D.; Weng, Q. A survey of image classification methods and techniques for improving classification performance. *Int. J. Remote Sens.*, ***2007***, *28*, 823–870.

Meier, M., de Souza, E.; Francelino, M.R.; Filho, E.I.F; and Schaefer, C.E.G.R. Digital Soil Mapping Using Machine Learning Algorithms in a Tropical Mountainous Area. *Revista Brasileira de Ciência de Solo* ***2018***, *42*, pp. 1-22.

Moraes, A.G.; Bocardo, E.; Bowling, L.C.; Daneshvar, F.; Pinto, J.; Watkins, A.H.; and Cherkauer, K.A. Assessment of Arequipa's hydrometeorological monitoring infrastructure to support water management decisions. *Purdue University Research Repository* ***2020***.

Moraes, A.G.; Bowling, L.C.; Zeballos Velarde, C.R.; and Cherkauer, K.A. Arequipa Climate Maps – Normals. *Purdue University Research Repository* ***2019***. doi:10.4231/490D-HC66

Mueller, R. and Harris, M. Reported Uses of CropScape and the National Cropland Data Layer Program. *Sixth International Conference on Agricultural Statistics* ***2013***.

NIST/SEMATECH e-Handbook of Statistical Methods. (2003). Bonferroni's Method. Retrieved from https://www.itl.nist.gov/div898/handbook/prc/section4/prc473.htm (accessed 1 June 2020)

Ozdarici-Ok, A.; Ok, A.O.; Schindler, K. Mapping of Agricultural Crops from Single High-Resolution Multispectral Images--Data-Driven Smoothing vs. Parcel-Based Smoothing. *Remote Sens.* ***2015***, *7*, 5611-5638.

Planet Labs, Inc. (2018). Planet Surface Reflectance Product. Retrieved from https://assets.planet.com/marketing/PDF/Planet_Surface_Reflectance_Technical_White_Paper.pdf

Popovici, R.; Erwin, A.; Ma, Z.; Prokopy, L.S.; Zanotti, L.; Bocardo Delgado E.F.; Pinto Cáceres, J.P.; Zeballos, E.Z.; Salas O'Brien, E.P.; Bowling, L.C.; Arce Larrea, G.R. Outsourcing governance in Peru's integrated water resources management. *Land Use Policy*, ***2021***, *Volume 101.*

QGIS Development Team. (2018). QGIS Geographic Information System. Open Source Geospatial Foundation Project. http://qgis.osgeo.org

R Core Team. (2019). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

R Core Team. (2020) Package 'randomForest'. Available at: https://cran.r-project.org/web/packages/randomForest/randomForest.pdf (accessed 1 November 2019)

Reyes-Gonzalez, A.; Kjaersgaard, J.; Trooien, T.; Hay, C.; and Ahiablame, L.. Estimation of Crop Evapotranspiration Using Satellite Remote Sensing-Based Vegetation Index. *Hindawi, Advances in Meterology* ***2018***.

Richards, J. *Remote Sensing Digital Image Analysis;* Springer-Verlag: Berlin, 1999; p. 240.

Rouse, J.W.; Haas, R.H.; Schell, J.A.; Deering, D.W.; and Harlan, J.C. *Monitoring the vernal advancement of retrogradation of natural vegetation;*NASA/GSFC (Type III, Final Report), **1974**; p.371.

Salmoral, G.; Zegarra, E.; Vázquez-Rowe, I.; González, F.; Del Castillo, L.; Saravia, G.R.; Graves, A.; Rey, D.; and Knox, J.W. Water-related challenges in nexus governance for sustainable development: Insights from the city of Arequipa, Peru. *Science of The Total Environment* ***2020****, 747:141114.*

Shao, Y.; Lunetta, R.S.; Ediriwickrema, J.; and Iiames, J. Mapping Cropland and Major Crop Types across the Great Lakes Basin using MODIS-NDVI Data. *Photogrammatric Engineering & Remote Sensing* ***2010****, Vol. 75, No. 1*, pp. 73-84.

Shanmugapriya, P.; Rathika, P.; Ramesh, T.; and Janaki, P. Applications of Remote Sensing in Agriculture – A Review. *International Journal of Current Microbiology and Applied Sciences* ***2019****, Volume 8 Number 01.*

Song, M.; Civco, D.L.; Hurd, J.D. A competitive pixel-object approach for land cover classification. *Int. J. Remote Sens.* ***2005****, 26*, pp. 4981–4997.

Stensrud, A.B. Dreams of growth and fear of water crisis: The ambivalence of "progress" in the Majes-Siguas irrigation project, Peru. *History and Anthropology* ***2016,*** *27(5)*, 569-584.

Stork, D.G.; Hart, P.E.; and Duda, R.O. *Pattern Classification, 2nd* ed; Wiley-Interscience, 2000.

Tenreiro, T.R.; Garcia-Vila, M.; Gomez, J.A.; Jimenez-Berni, J. A.; and Elias, F. Using NDVI for the assessment of canopy cover in agricultural crops within modelling research. *Computers and Electronics in Agriculture* ***2021***, 182:106038.

Undersander, D.; Cosgrove, D.; Cullen, E.; Grau, C.; Rice, M.E.; Renz, M.; Sheaffer, C.; Shewmaker, G.; Sulc, M. *Alfalfa Management Guide*; American Society of Agronomy, Inc., Crop Science Society of America, Inc., and Soil Science Society of America, Inc., 2000.

United States Department of Agriculture. (n.d.) CropScape and Cropland Data Layer. Available online: https://www.nass.usda.gov/Research_and_Science/Cropland/Release/ (accessed 1 May 2021)

Warmerdam, F. (n.d.). gdal_alg.h: GDAL Algorithms C API. Retrieved from https://gdal.org/api/gdal_alg.html#_CPPv415GDALSieveFilter15GDALRasterBandH15GDALRasterBandH15GDALRasterBandHiiPPc16GDALProgressFuncPv (accessed 1 December 2019)

Wilkinson, G.G. Results and implications of a study of fifteen years of satellite image classification experiments. *IEEE Trans. Geosci. Remote Sens.* ***2005****, 43,* 433–440.

Yan, L. and Roy, D.P. Conterminous United States crop field size quantification from multi-temporal Landsat data. *Remote Sensing of the Environment* ***2016****, 172,* pp. 67-86.

Zhang, X.; Zhang, M.; Zheng, Y.; and Wu, B. Crop Mapping Using PROBA-V Time Series Data at the Yucheng and Hongxing Farm in China. *Remote Sens.,* ***2016****, 8,* 915.

Zhong, L. Efficient crop type mapping based on remote sensing in the Central Valley, California. *UC Berkeley Electronic Theses and Dissertations* ***2012***.