Co-Production of a 10-m Cropland Extent Map for Continental Africa using Sentinel-2, Cloud Computing, and the Open-Data-Cube

Chad Burton[1], Fang Yuan[2], Chong Ee-Faye[1], Meghan Halabisky[2], David Ongo[3], Fatou Mar[4], Victor Addabor[5], Bako Mamane[6], and Sena Adimou[7]

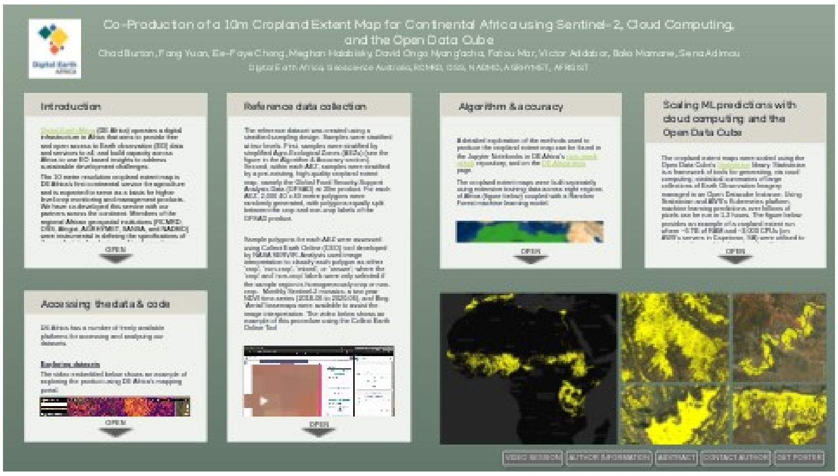
[1]Geoscience Australia
[2]Digital Earth Africa
[3]Regional Centre for Mapping of Resources for Development
[4]Sahara and Sahel Observatory
[5]National Disaster Management Organization
[6]AGRHYMET
[7]AFRIGIST

November 23, 2022

Abstract

A central focus for governing bodies in Africa is the need to secure the necessary food sources to support their populations. It has been estimated that the current production of crops will need to double by 2050 to meet future needs for food production. Higher level crop-based products that can assist with managing food insecurity, such as cropping watering intensities, crop types, or crop productivity, require as a starting point precise and accurate cropland extent maps indicating where cropland occurs. Current continental cropland extent maps of Africa are either inaccurate, have too coarse spatial resolutions, or are not updated regularly. An accurate, high-resolution, and regularly updated cropland extent map for the African continent is therefore recognized as a gap in the current crop monitoring services. Using Digital Earth Africa's Open Data Cube platform, and working in conjunction with multiple regional African geospatial institutions, we co-develop a 10 metre resolution cropland extent map over the African continent using a Random Forest machine learning classifier and an annual time-series of Sentinel-2 satellite images. Members of the regional African geospatial institutions (RCMRD, OSS, Afrigist, AGRHYMET, and NADMO) were instrumental in defining the specifications of the product, in developing and implementing a continental scale reference data collection strategy, and assisted with iterative model building. The cropland extent map comes packaged with three layers: a pixel-based classification, a pixel-based cropland probability layer, and an object-based segmentation filtered classification. All the components of Digital Earth Africa's cropland extent map: models, reference data, code, and results are open source and freely available online through Digital Earth Africa's mapping and analysis platforms. A fuller description of the dataset, including methods, the validation results, and how to access the different datasets can be seen on the DE Africa user guide: https://docs.digitalearthafrica.org/en/latest/data_specs/Cropland_extent_specs.html

Co-Production of a 10m Cropland Extent Map for Continental Africa using Sentinel-2, Cloud Computing, and the Open Data Cube



Chad Burton, Fang Yuan, Ee-Faye Chong, Meghan Halabisky, David Ongo Nyang'acha, Fatou Mar, Victor Addabor, Bako Mamane, Sena Adimou

Digital Earth Africa, Geoscience Australia, RCMRD, OSS, NADMO, AGRHYMET, AFRIGIST

PRESENTED AT:

INTRODUCTION

Digital Earth Africa (https://www.digitalearthafrica.org/) (DE Africa) operates a digital infrastructure in Africa that aims to provide free and o access to Earth observation (EO) data and services to all, and build capacity across Africa to use EO based insights to address sustainable development challenges.

The 10 metre resolution cropland extent map is DE Africa's first continental service for agriculture and is expected to serve as a basis for higl level crop monitoring and management products. We have co-developed this service with our partners across the continent. Members of the regional African geospatial institutions (RCMRD, OSS, Afrigist, AGRHYMET, SANSA, and NADMO) were instrumental in defining the specifications of the product, in developing and implementing a continental scale reference data collection strategy, and assisting with iterativ model building.

We have classified cropland using an annual Sentinel-2 time series and a Random Forest machine learning model. The cropland extent produ among the highest resolution and highest accuray products of its type for the continent of Africa. The product comes packaged with three lay pixel-based classification, a pixel-based cropland probability layer, and an object-based segmentation filtered classification. All the compone the service: models, reference data, code, and results are open source and freely available online through Digital Earth Africa's mapping and analysis platforms.

A full description of the dataset, including details on how it was made, the validation results, and how to access the different datasets can be s on the deafrica user guide. (https://docs.digitalearthafrica.org/en/latest/data_specs/Cropland_extent_specs.html)

The cropland extent product is in the final stages of development, with southern and central Africa still to complete. The expected completio is early 2022. The completed regions are available on all of Digital Earth Africa's platforms (see the section below for links and details).

ACCESSING THE DATA & CODE

DE Africa has a number of freely available platforms for accessing and analysing our datasets.

Exploring datasets

The video embedded below shows an example of exploring the product using DE Africa's mapping portal.

[VIDEO] https://res.cloudinary.com/amuze-interactive/video/upload/vc_auto/v1638240962/agu-fm2021/BA-E4-89-84-15-50-25-89-AE-A8 C5-32-7D-19-EE/Video/Digital_Earth_Africa_Map_-_30_November_2021_lzk0h1.mp4

The online mapping portal can be found at this url: maps.digitalearth.africa (https://maps.digitalearth.africa).

A direct link with some of the cropland extent layers pre-loaded can be accessed by following this link. (https://maps.digitalearth.africa/#shar y3JyeK8agnTdwegsSLhlYIhdrqw)

Analysing datasets

The cropland extent data can be analysed alongside a suit of other satellite datasets through on our free JupyterLab Sandbox environment, available at sandbox.digitalearth.africa (http://sandbox.digitalearth.africa/)

The sandbox comes pre-loaded with a comprehensive and well documented set of example Jupyter notebooks and python tools demonstrating Earth Observation workflows using open source libraries. The repository for these notebooks is located on github (https://github.com /digitalearthafrica/deafrica-sandbox-notebooks). The cropland extent maps are utilised in several Jupyter notebooks, including:

- A dataset description notebook (https://github.com/digitalearthafrica/deafrica-sandbox-notebooks/blob/main/Datasets /Cropland_extent.ipynb), including details on how to load the data and a simple example use-case. This notebook is shown in the video below
- Crop phenology (https://github.com/digitalearthafrica/deafrica-sandbox-notebooks/blob/main/Real_world_examples /Phenology_optical.ipynb) using Sentinel-2
- Cropping NDVI Anomalies (https://github.com/digitalearthafrica/deafrica-sandbox-notebooks/blob/main/Real_world_examples /Vegetation_anomalies_monthly.ipynb)

[VIDEO] https://res.cloudinary.com/amuze-interactive/video/upload/vc_auto/v1638329789/agu-fm2021/BA-E4-89-84-15-50-25-89-AE-A8 C5-32-7D-19-EE/Video/Datasets_Cropla_-_JupyterLab_-_1_December_2021_uwk94k.mp4

REFERENCE DATA COLLECTION

The reference dataset was created using a stratified sampling design. Samples were stratified at two levels. First, samples were stratified by simplified Agro-Ecological Zones (AEZs) (see the figure in the Algorithm & Accuracy section). Second, within each AEZ, samples were stra by a pre-existing, high-quality cropland extent map, namely the Global Food Security-Support Analysis Data (GFSAD) at 30m product. For e AEZ, 2,000 40 x 40 metre polygons were randomly generated, with polygons equally split between the crop and non-crop labels of the GFSA product.

Sample polygons for each AEZ were assessed using Collect Earth Online (CEO) tool developed by NASA SERVIR. Analysts used image interpretation to classify each polygon as either 'crop', 'non-crop', 'mixed', or 'unsure'; where the 'crop' and 'non-crop' labels were only sel if the sample region is homogeneously crop or non-crop. Monthly Sentinel-2 mosaics, a two year NDVI time-series (2018-06 to 2020-06), ar Bing 'Aerial' basemaps were available to assist the image interpretation. The video below shows an example of this procedure using the Coll Earth Online Tool

[VIDEO] https://res.cloudinary.com/amuze-interactive/video/upload/vc_auto/v1638239626/agu-fm2021/BA-E4-89-84-15-50-25-89-AE-A8 C5-32-7D-19-EE/Video/Collect_Earth_Online_-_30_November_2021_pbdscc.mp4

Cropland in the reference data is defined as "*a piece of land of minimum 0.01 ha that is sowed/planted and harvestable at least once within 1. months after the sowing/planting date*".

This definition excludes non-planted grazing lands and perennial crops which can be difficult for satellite imagery to differentiate natural vegetation.

In addition to the samples collected through CEO, additional samples were digitized through a GIS application in locations where the classifi poorly. This process was done in an iterative fashion, incrementally improving classifications through targeted training data collection.

In total, >24,000 training data samples were collected, with an additional ~1,800 samples isolated as an independent validation dataset (see th figure in the slideshow in the bottom right of the poster).

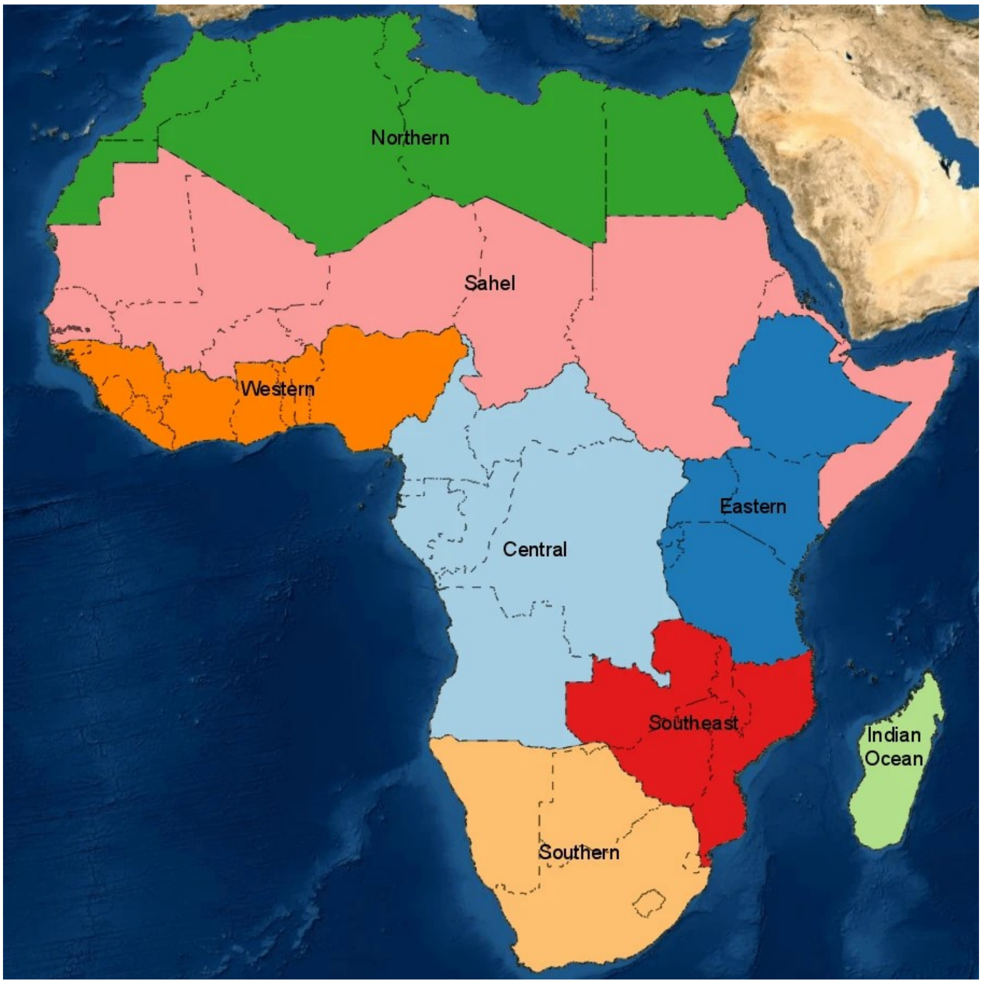
The accuracy of the method for collecting reference samples described above was independently evaluated by Radiant Earth (https://www.radiant.earth/). Radiant Earth's team developed a visualisation app to validate a random subset of the labels collected using the tool. The app retrieves Airbus SPOT imagery over the area of interest in four 6-month windows (covering the two-year of the training data specification). For each AEZ, the app would sequentially retrieve the images for each individual polygon, visualize all the available imagery overlay the polygon on top of it. A member of Radiant's team would then interpret the class (Crop, No-Crop, Mixed) based on the guideline examples provided by Digital Earth Africa, and record that in the app. After validating all polygons, a new GeoJSON file would be generated the additional property for validation labels. The results of this independent validation of DE Africa's reference samples are shown in the tabl below. The overall accuracy is 96.3 %, indicating that the reference samples are of a high quality and fit-for-purpose.

		Collect Earth Label			
		non-crop	**crop**	**total**	**producer's**
RE Label	**non-crop**	289	18	307	94.1
	crop	3	258	261	98.9
	total	292	276	568	
	user's	99.0	93.5		96.3

ALGORITHM & ACCURACY

A detailed exploration of the methods used to produce the cropland extent map can be found in the Jupyter Notebooks in DE Afr crop-mask github (https://github.com/digitalearthafrica/crop-mask) repository, and on the DE Africa docs (https://docs.digitalearthafrica.org/en/latest/data_specs/Cropland_extent_specs.html) page.

The cropland extent maps were built seperately using extensive training data across eight regions of Africa (figure below) couple with a Random Forest machine learning model.



The algorithm is summarised in the main figures of the poster (bottom right) as a flow chart. The annual time-series of Sentinel-2 images is f divided into two six-month seasons: Jan-Jun and Jul-Dec. The two seasons are composited through the calculation of a high-dimensional geometric median ('geomedian') which preserves the relationship between bands, allowing the accurate calculation of band indices like NDVI LAI and MNDWI. In addition to the calculation of geomedians, three measures of variability around the geomedian are calculated: bray-curt euclidean, and spectral median absolute deviations. Ancillary datasets such as slope and the CHIRPS rainfall climatologies are also appended These features are fed into a Random Forest classifier to produce the pixel-based estimates of crop/non-crop. The pixel based estimates are th improved through an image segmentation and majority vote filter to produce an object-based cropland extent.

The accuracy of this approach is summarised in the confusion matrices below (for the regions that have been completed). The headline statist are bulleted below.

- The **Eastern Africa** cropland extent map has an **overall accuracy of 90.3 %, and an f-score of 0.85**
- The **Western Africa** cropland extent map has an **overall accuracy of 83.6 %, and an f-score of 0.75**

- The **Northern Africa** cropland extent map has an **overall accuracy of 94.0 %, and an f-score of 0.91**
- The **Sahel Africa** cropland extent map has an **overall accuracy of 87.9 %, and an f-score of 0.78**

Western		*Map Label*			
		non-crop	**crop**	**total**	**producer's**
Validation Label	**non-crop**	177	24	201	88.1
	crop	25	72	97	74.2
	total	202	96	298	--
	user's	87.6	75.0	--	83.6

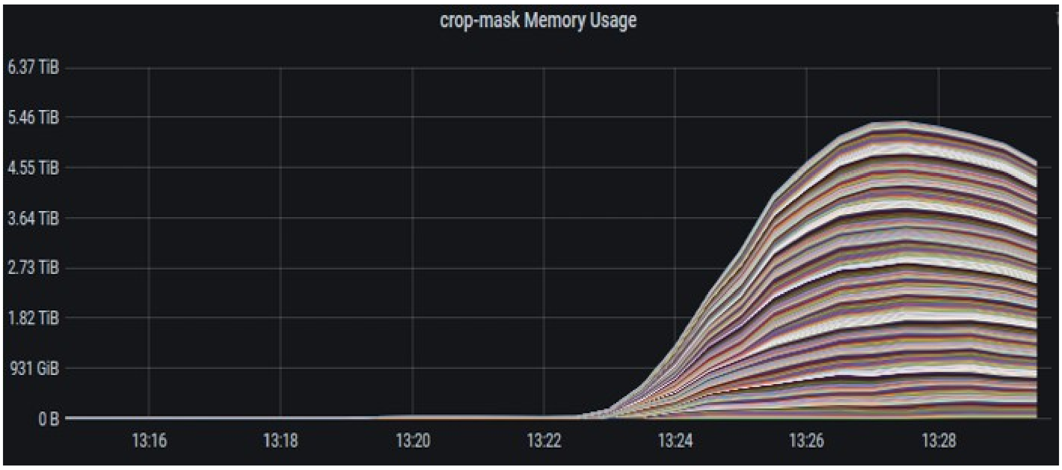
Northern		*Map Label*			
		non-crop	**crop**	**total**	**produc**
Validation Label	**non-crop**	192	8	200	[illegible]
	crop	10	90	100	[illegible]
	total	202	98	300	
	user's	95.1	91.8		[illegible]

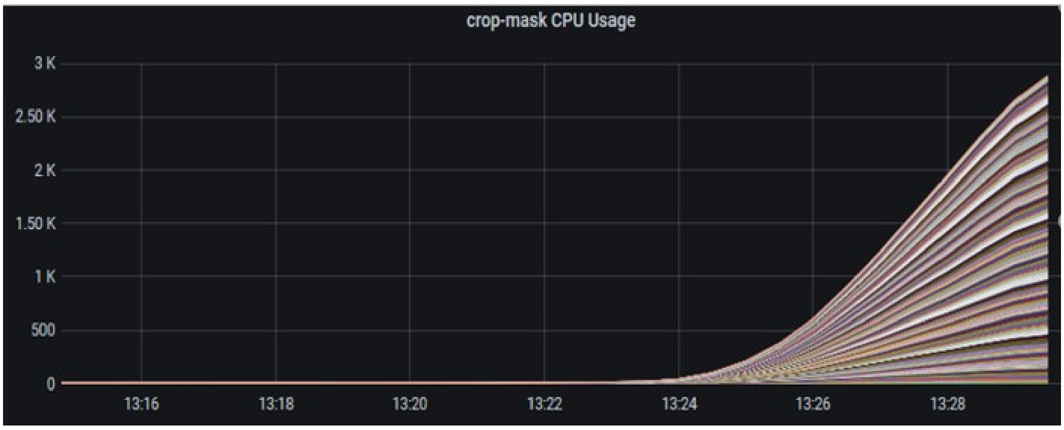
Eastern		*Map Label*			
		non-crop	**crop**	**total**	**producer's**
Validation Label	**non-crop**	190	10	200	95.0
	crop	19	81	100	81.0
	total	209	91	300	
	user's	90.9	89.0		90.3

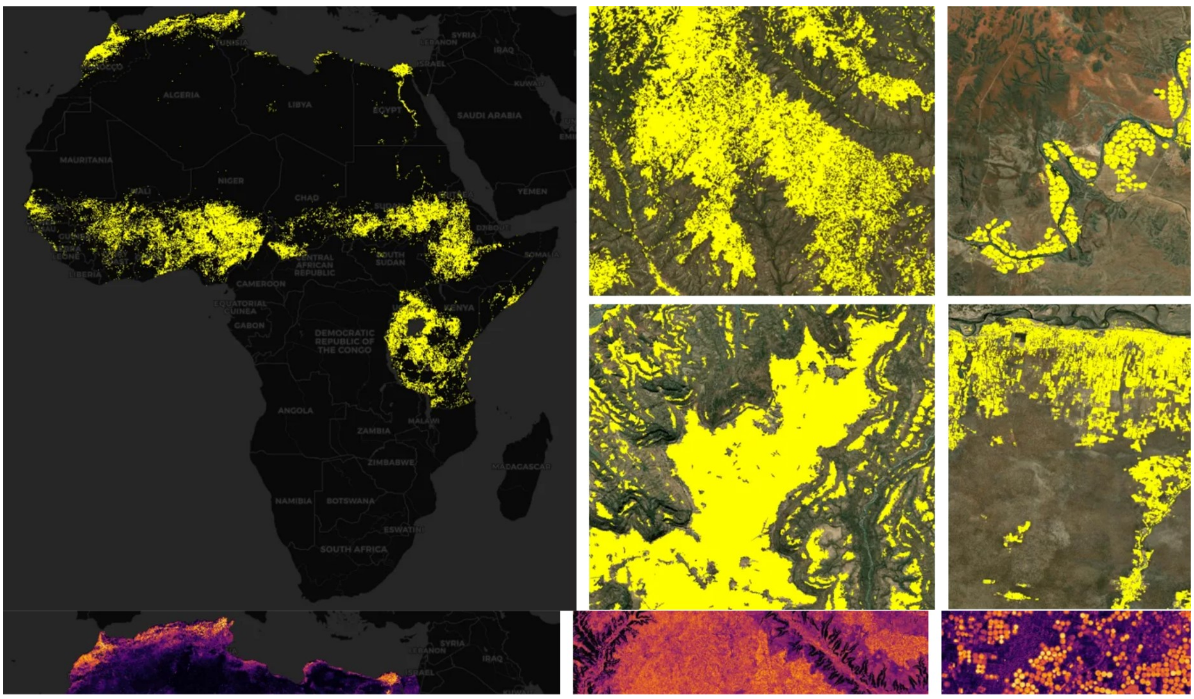
Sahel		*Map Label*			
		non-crop	**crop**	**total**	**produc**
Validation Label	**non-crop**	191	9	200	[illegible]
	crop	26	62	88	[illegible]
	total	217	71	288	[illegible]
	user's	88.0	87.3		[illegible]

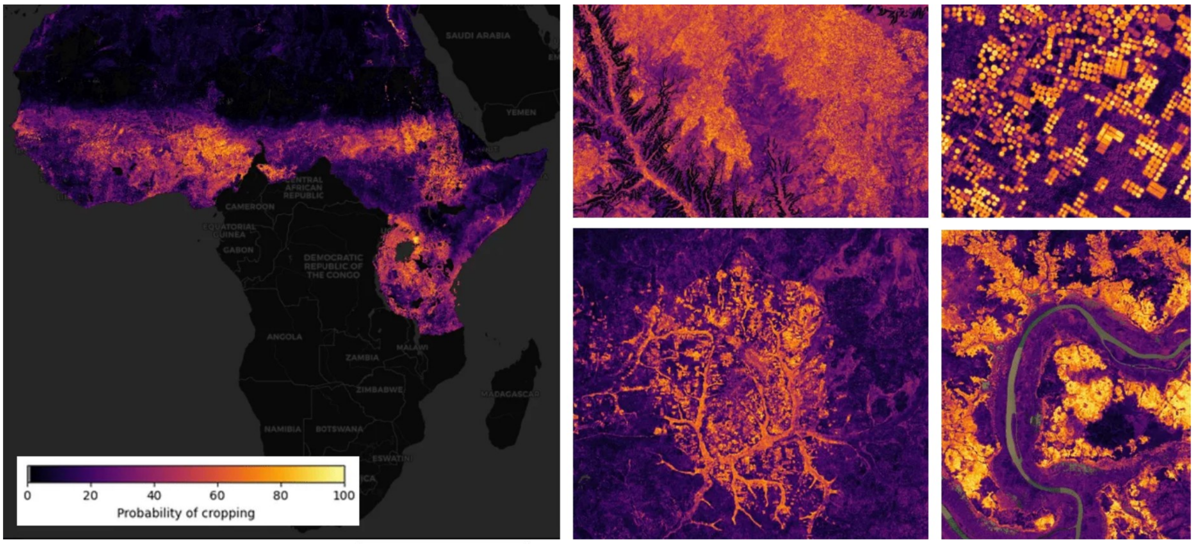
SCALING ML PREDICTIONS WITH CLOUD COMPUTING AND THE OPEN DATA CUBE

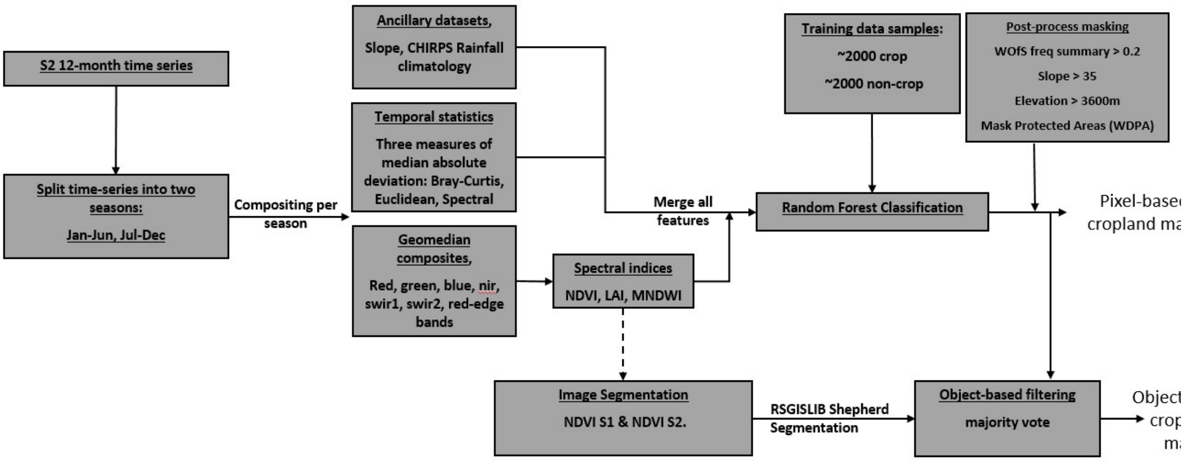
The cropland extent maps were scaled using the Open Data Cube's Statistician (https://github.com/opendatacube/odc-stats) library. Statisticia framework of tools for generating, via coud computing, statistical summaries of large collections of Earth Observation Imagery managed in a Open Datacube Instance. Using Statistician and AWS's Kubernetes platform, machine learning predictions over billions of pixels can be run i hours. The figure below provides an example of a cropland extent run where ~6 TB of RAM and ~3,000 CPUs (on AWS's servers in Capetov SA) were utilised to complete the machine leanring classification pipeline for Sahel Africa is 1.5 hours.

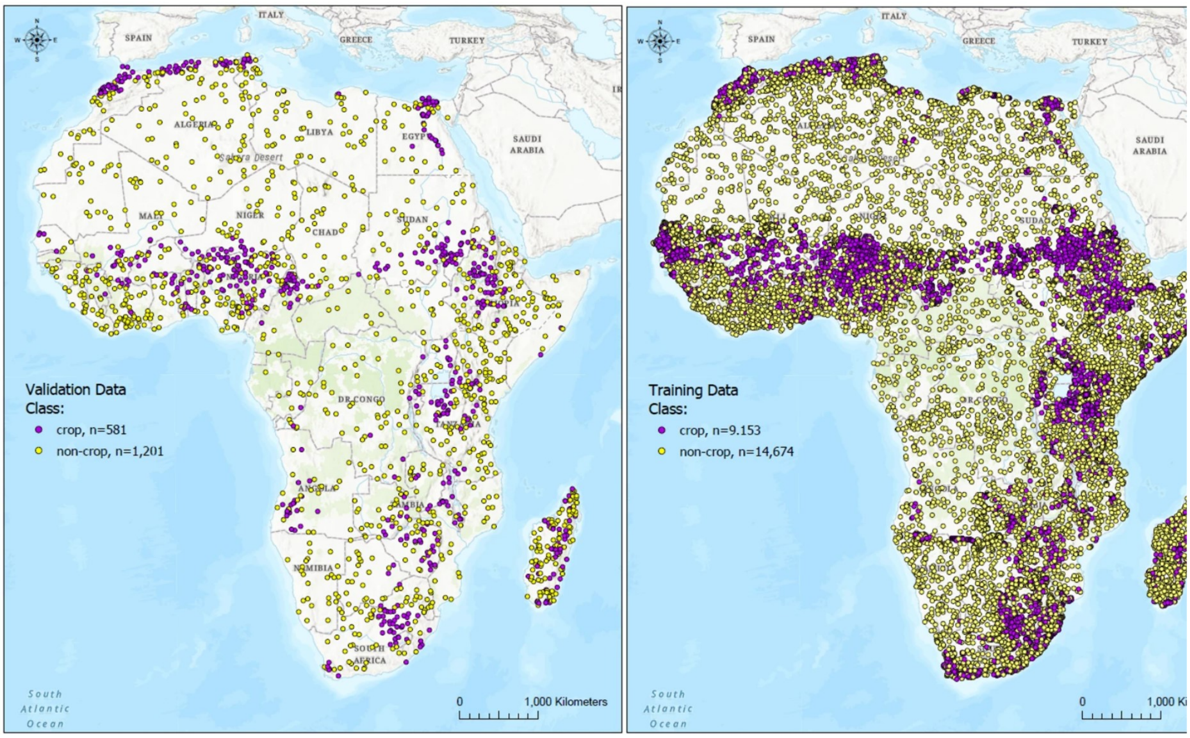







0
20
40
60
80
100
Probability of cropping


S2 12-month time series
Split time-series into two seasons:
Jan-Jun, Jul-Dec
Compositing per season
Ancillary datasets,
Slope, CHIRPS Rainfall climatology
Temporal statistics
Three measures of median absolute deviation: Bray-Curtis, Euclidean, Spectral
Geomedian composites,
Red, green, blue, nir, swir1, swir2, red-edge bands
Spectral indices
NDVI, LAI, MNDWI
Merge all features
Training data samples:
~2000 crop
~2000 non-crop
Random Forest Classification
Post-process masking
WOfS freq summary > 0.2
Slope > 35
Elevation > 3600m
Mask Protected Areas (WDPA)
Pixel-based cropland ma
Image Segmentation
NDVI S1 & NDVI S2.
RSGISLIB Shepherd Segmentation
Object-based filtering
majority vote
Object crop ma


Validation Data
Class:
crop, n=581
non-crop, n=1,201
0 1,000 Kilometers
South Atlantic Ocean
Training Data
Class:
crop, n=9.153
non-crop, n=14,674
0 1,000 Ki
South Atlantic Ocean

AUTHOR INFORMATION

Chad Burton is an Earth Observation scientist for Digital Earth Africa. He currently works as the project lead for the continental cropland extent service being produced by DE Africa. Prior to working for DE Africa Chad worked for Digital Earth Australia where he led an effort to use the Landsat satellite series to map 30 years of irrigation in Australia's most extensive food producing river basin. Prior to working for DE Australia, Chad worked as a research assistant in the ecosystems lab at Oxford's Environmental Change Institute where he researched the impact of the El Nino Southern Oscillation on the global carbon cycle. Chad has a keen interest in understanding and mapping environmental change to better inform environmental decision making. He has a Master's in Environmental Change and Management from the University of Oxford, and a Bachelor of Science (Honours) in Geosciences from Monash University.

ABSTRACT

A central focus for governing bodies in Africa is the need to secure the necessary food sources to support their populations. It has been estimated that the current production of crops will need to double by 2050 to meet future needs for food production. Higher level crop-based products that can assist with managing food insecurity, such as cropping watering intensities, crop types, or crop productivity, require as a starting point precise and accurate cropland extent maps indicating where cropland occurs. Current continental cropland extent maps of Africa are either inaccurate, have too coarse spatial resolutions, or are not updated regularly. An accurate, high-resolution, and regularly updated cropland extent map for the African continent is therefore recognised as a gap in the current crop monitoring services. Using Digital Earth Africa's Open Data Cube platform, and working in conjunction with multiple regional African geospatial institutions, we co-develop a 10 metre resolution cropland extent map over the African continent using a Random Forest machine learning classifier and an annual time-series of Sentinel-2 satellite images. Members of the regional African geospatial institutions (RCMRD, OSS, Afrigist, AGRHYMET, and NADMO) were instrumental in defining the specifications of the product, in developing and implementing a continental scale reference data collection strategy, and assisted with iterative model building. The cropland extent map comes packaged with three layers: a pixel-based classification, a pixel-based cropland probability layer, and an object-based segmentation filtered classification. All the components of Digital Earth Africa's cropland extent map: models, reference data, code, and results are open source and freely available online through Digital Earth Africa's mapping and analysis platforms.