USING OPEN DATA CUBE ON THE CLOUD TO INVESTIGATE FOOD SECURITY BY MEANS OF CROPLAND CHANGES IN DJIBOUTI

J.A.Wardle [1], V. Sagan [1]*, F. Mohammed[1]

[1] Saint Louis University, Geospatial Institute, Saint Louis, United States of America - (julee.wardle, vasit.sagan, faten.mohammed)@slu.edu

Commission III, WG III/10

ABSTRACT:

Addressing hunger is one of the greatest unsolved challenges Humanity has ever faced. Africa is one of the most fragile ecosystems strongly affected by numerous factors ranging from climate change, increasing population, decreasing water resources, and undeveloped hydrological infrastructure. These factors make it exceptionally vulnerable to food insecurity. The purpose of this study was to establish the feasibility and methodology of using Open Data Cube (ODC) and conventional machine learning algorithms to determine the extent of decrease in cropped area in the desert climate of Djibouti, the smallest Horn of Africa country (by landmass) over a thirty-year period. The research question was answered using Landsat 5, 7, and 8 imagery taken during the month of June from 1990, 2000, 2010, and 2020 then classified through machine learning algorithms - including decision tree and random forest. The data acquisition, analysis, and modeling were completed in an Open Data Cube environment using a cloud-based user computational platform running completely in-browser, and all necessary software was provided as part of the environment. The research identified a decreasing trend in vegetative areas but was limited in determining whether the vegetative areas were purely agricultural cropland in nature or included native vegetation. While the research reveals a concerning decline in total vegetation over the thirty-year period, the lack of other data variables (such as weather and climate patterns) provides too narrow a picture to determine causation. Several areas for further research are outlined.

1. INTRODUCTION

Addressing hunger is one of the greatest unsolved challenges Humanity has ever faced. Hunger has many causes including geopolitical volatility, extreme weather events, and changes in diet such as moving to a more meat-based rather than vegetation-based diet (Wheeler and Von Braun 2013). The United Nations' Food and Agricultural Organization (FAO) broadly defines food security as "(i) the availability of sufficient quantities of food of appropriate quality, supplied through domestic production or imports; (ii) access by individuals to adequate resources (entitlements) for acquiring appropriate foods for a nutritious diet; (iii) utilization of food through adequate diet, clean water, sanitation, and health care to reach a state of nutritional well-being where all physiological needs are met; and (iv) stability, because to be food secure, a population, household or individual must have access to adequate food at all times (UNFAO 1996). Climate and food security are unavoidably intertwined. Global environmental threats to food security include climate change and variability, loss of biodiversity, and environmental pollution (UNFAO 1996).

Africa is one of the most fragile ecosystems strongly affected by numerous factors ranging from climate change, increasing population, decreasing water resources, and undeveloped hydrological infrastructure. These factors make it exceptionally vulnerable to food insecurity.

The Horn of Africa (HOA) region has seen famine, even in recent history, and is considered one of the most food-insecure regions of the world (Qu and Hao 2018). This region suffers from ongoing conflict, long-term poverty, and ecosystem deterioration from desertification, fuelwood scarcity, land degradation,

biodiversity loss, and human-induced droughts (Thrupp and Megateli 1999).

The rapid growth in satellite technology over the last few decades has improved monitoring agencies' ability to track agricultural/ vegetative changes in regions where food stability has traditionally been an area of concern. Satellite remote sensing is the leading technology to provide comprehensive information about different earth systems, particularly in monitoring global vegetation health and trends (Qu and Hao 2018).

There are various methods commonly used to monitor agricultural productivity including in-situ-based methods, optical remote sensing methods, thermal remote sensing methods, microwave remote sensing methods, combined remote sensing methods, and synergy between in-situ and remote sensing-based methods (Hazaymeh and Hassan 2016). Optical and thermal are considered passive, while microwave is an active sensing method. Among passive remote sensing methods, hyperspectral imaging (HSI) has the potential as a non-invasive and non-destructive tool for monitoring vegetative health (Jones 2010). This method captures and stores an object's spectroscopy information in a spectral cube, which contains spatial information and hundreds of contiguous wavelengths in the third dimension.

Traditional remote sensing imagery in the past has been expensive, unwieldy to store, manage, and access, and required highly specialized skills and software to analyze. Data cubes are

---

* Corresponding author

a newer approach to tackling these organizational and analytic issues. An image data cube stores large collections of temporal, analysis-ready, multispectral Earth observation data that enables fast access and analysis from a variety of web and desktop applications (Kopp, Becker et al. 2019).

Open Data Cube is a non-profit, open-source project motivated by the need to better manage satellite data and is freely available to the public on GitHub. It is a set of Python libraries and PostgreSQL database that helps analysts work with geospatial raster data in a common analytical framework composed of a series of data structures and tools (ODC 2021).

Data cubes are being increasingly used to assess cropland dynamics from a spatio-temporal perspective. For example, Digital Earth Australia, an Open Data Cube (ODC) initiative, offers several high dimensional statistical products of value for land use classification and is being used for change detection and machine learning in land use classification, especially over areas that undergo large changes in cover within a year, like the irrigated croplands in Western Australia (Wellington, Renzullo et al. 2021). Earth Observation Data Centre for Water Resources Monitoring in Austria designed a data cube of Sentinel-1 data that has been used for rice mapping in the Mediterranean, Europe-wide vegetation monitoring, and soil moisture retrieval in Italy (Wagner, Bauer-Marschallinger et al. 2021).

Digital Earth Africa is another ODC initiative, funded in part by the Helmsley Charitable Trust and supported by Digital Earth Australia (Africa 2020). It was established to improve quality of life on the African continent by translating Earth observations into knowledge that will aid long-term development. The DE Africa platform is used by African governments, industry, and academic institutions to track remotely sensed changes across the continent, especially in flooding, drought, soil and coastal erosion, agriculture, forest cover, land use and land cover change, water availability, and quality, and changes to human settlements (Africa 2020).

DE Africa maintains robust United States Geological Survey (USGS) Landsat and European Commission (EC) and European Space Agency (ESA) Copernicus Sentinel-2 satellite imagery archives over the continent as well as provides a cloud-based user computational platform in the form of a sandbox, operating in a Jupyter Lab environment (DEA 2020). It is all made freely available for users working on any type of African geospatial challenges.

The purpose of this research was to establish the feasibility and methodology of using Open Data Cube and conventional machine learning algorithms to determine the extent of decrease in total cropped area in the desert climate of Djibouti, the smallest Horn of Africa country (by landmass) over a thirty-year period (Figure 2).

2. DATA

The data used in this study was primarily USGS Landsat and ESA Sentinel-2 imagery, all provided as "Analysis Ready Data" (ARD) on the DE Africa platform (Africa 2020) using the Committee on Earth Observing Satellites (CEOS) ARD specifications. CEOS ARD data is processed to a minimum set of requirements and organized to allow rapid analysis with a minimal amount of additional user effort and interoperability both through time and with other datasets (CEOS 2016). Using data in the ARD format significantly decreased time spent accessing, pre-processing, and organizing the data for the study.
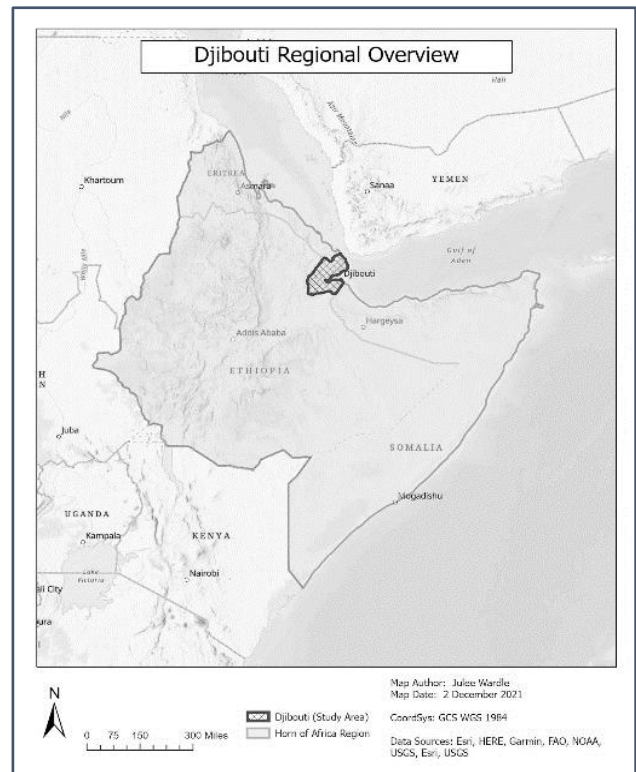

Figure 1. Djibouti's location in the Horn of Africa.

There were some initial difficulties acquiring imagery as the 30-year span in the study required data from several different satellites. Landsat 5 was initially used for the 1990 and 2010 calculations. Landsat 7 was used for 2000 and Landsat 8 for the 2020 calculations. Landsat 7 had significant issues due to the Scan Line Corrector failure in 2003 (Masek 2021). The failure made the images over Djibouti in 2010 nearly unusable, so Landsat 5 was used as its' mission continued through 2013 (Rocchio). Landsat 5 has different bands than the later satellites which had to be accounted for in coding the scripts.

One type of ARD provided by DE Africa are GeoMAD cloud-free composites. GeoMAD stands for Geomedian and Median Absolute Deviations and each composite contains a "representative, multi-spectral image for every pixel of the African continent. The result is a comprehensive dataset that can be used either to generate true-colour images for visual inspection of the landscape, or the full spectral dataset can be used to develop more complex algorithms" (Africa 2022). The GeoMAD products used extensively in this study were "gm_ls8_annual" (composite using Landsat-8 imagery, available for the years 2013 – 2020) and "gm_ls5_ls7_annual" (composite combining both Landsat-5 and Landsat-7 imagery, available for the years 1984 – 2012) (Africa 2022).

Other datasets included Djibouti country vector boundaries acquired through ArcGIS Online (Esri 2020) used primarily for situational awareness. The source data for the country boundaries and attribute data was from the 2019 World Factbook (CIA 2022). Training data for the models were created using ArcGIS Pro 2.8 then uploaded to the DE Africa sandbox environment.

Accessing and analyzing the data through the ODC significantly decreased processing time and both software and hardware

requirements, enabling the study to be completed on a standard computer, without relying on a traditional remote sensing computer laboratory for hardware and software requirements.

3. METHOD

The United States Department of Agriculture (USDA) defines cropland as "areas used for the production of adapted crops for harvest" and recognizes two subcategories: cultivated and non-cultivated. Cultivated cropland being row crops or close-grown crops and other cultivated cropland such as hay land or pastureland that is in a rotation with row or close-grown crops. Non-cultivated cropland is permanent hay land or horticultural cropland such as orchards (Agriculture 2022).

The technical workflow of this study (Figure 3) covered data processing, feature engineering, modelling, and final calculations. For data processing the necessary pre-processing steps of radiometric calibration, pan-sharpening, and co-registration were completed by DE Africa's GeoMAD product line. All imagery acquisition and processing, feature engineering, and modelling were completed in DE Africa OCD's JupyterLab environment. The vector data processing, particularly the cropland/not cropland shapefile creation, and final raster calculations were completed in ArcGIS Pro 2.8. The hand-crafted features including various spectral and textural indices were calculated, stacked into multitemporal image cubes, then clipped by country boundary shapefile to produce individual image chips representing each year.

Due to the sheer size of the imagery cubes encompassing the entire country of Djibouti and the limitations of the ODC computing environment, the study boundary was decreased to the area surrounding Djibouti city. This region had the most visible cropland on satellite imagery and was most effective in producing differentiating results over the 30-year period.

Within the ODC, we ran conventional machine learning algorithms, including decision tree classifier, random forest classifier and support vector machine classifier. The training data consisted of a shapefile depicting 348 features of cropland / non-cropland areas. The areas were specifically selected to capture the spectral variability of both classes. Class imbalance was a concern as there is so little cropland in Djibouti, but the two classes were kept as equal as possible. The imagery used in the models and the indices was from late June to early July each year to be consistent within seasonal time frame.

To reduce the model complexity (and improve performance) we reduced the total bands from 12 down to 5 and reduced the cropland / non-cropland features from the initial 629 to 348. To improve accuracy of the final output, we calculated three band indices and added them to the model: Normalized Difference Vegetation Index (NDVI), Built-up Index (BUI), and Modified Normalized Difference Water Index (MNDWI).

The output from the models and the indices were exported from the ODC sandbox and imported to ArcGIS Pro for further analysis. Using a raster calculator, we compared the differences between the span of years for each model and the indices.

The ODC provided the means to manage, store and analyze the sheer quantity of multitemporal, multispectral image cubes for Djibouti during the 30-year period. Using DE Africa's cloud-based user computational platform significantly reduced computational time on the models. The sandbox provided 4 cores and 32G of memory. It ran completely in-browser, and all necessary software was provided as part of the environment, so no additional installation or configuration was required.

4. RESULTS AND DISCUSSION

The results of the processing and analysis showed both increases and decreases in cropped land throughout the thirty-year period., but the overall pattern between 1990 and 2020 was a decreasing trend.

The decision tree model accuracy ranged from 92% to 94%. Support vector machine models were 91% for both years. The random forest models were the most accurate with a consistent 98% accuracy.

The initial results of the decision tree model (Figure 4) were much too inclusive given the nature of decision tree models to overfitting (L. Breiman 1984) and despite our efforts at balancing the dataset prior to fitting. The 2020 results did show a decreasing trend in overall vegetation (Figure 5) but was still overly inclusive of all vegetation.
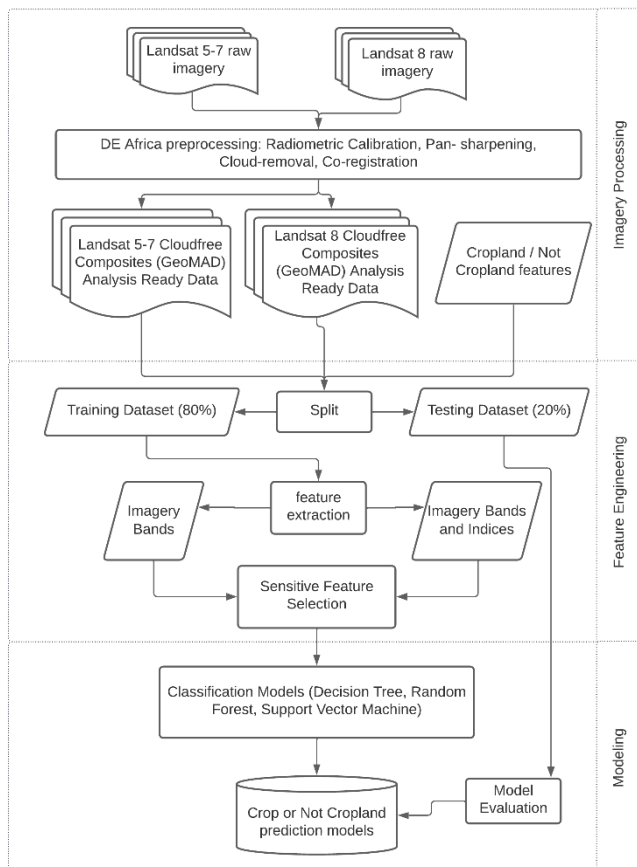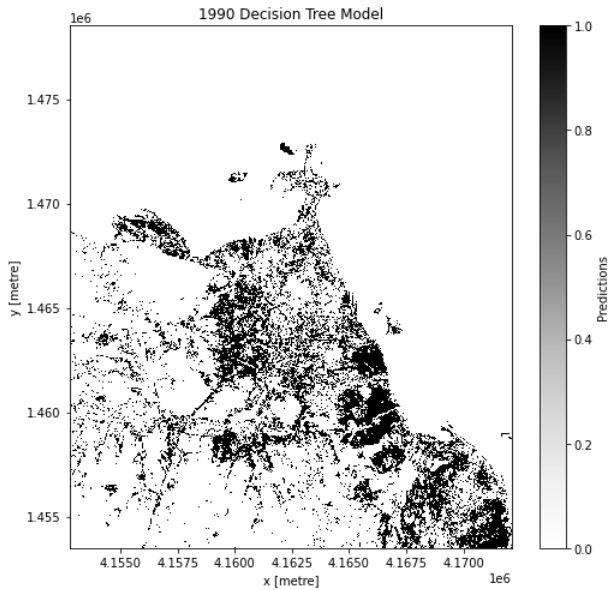

Figure 3. Overall workflow for the study

Figure 4. Initial results of the 1990 decision tree model.

Figure 5. Decision tree model shows less vegetation in 2020.

To refine the decision tree results, we calculated three indices (NDVI, BUI, and MNDWI) and determined the importance of each variable (band) to the model. Once we discovered the five most important features (Figure 6), we ran the models specifying only that subset of features. The decision tree models for all four years, while they did show a decreasing trend in overall vegetation, showed minimal evidence of differentiation between wild vegetation and cropland.

Figure 6. Bands in order of importance.

The results of the Random Forest classifier also showed a decreasing trend in overall vegetation from 1990 to 2020 (Figure 7 and Figure 8) and were slightly less inclusive of wild vegetation as compared to the decision tree model. Despite the increase in accuracy, it was still not precise enough to purely differentiate between just cropland and all other vegetation.

Figure 7. Random forest results for 1990.

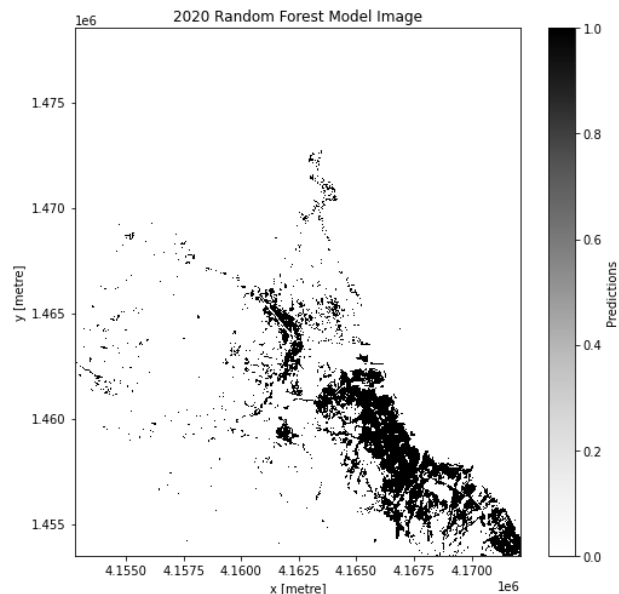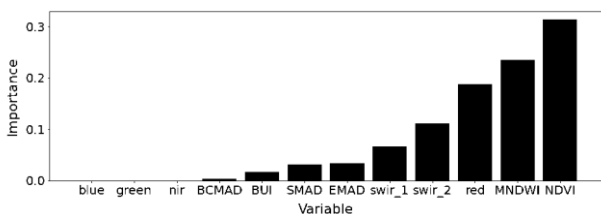

Figure 8. Random forest for 2020 showing less vegetation.

Support Vector Machine classifier had similar results to Random Forest but with a lower accuracy score (91%) , clearly showing a decreasing trend (Figure 9 and Figure 10) in overall vegetation in the region, but unable to fully differentiate between cropland and other vegetation.
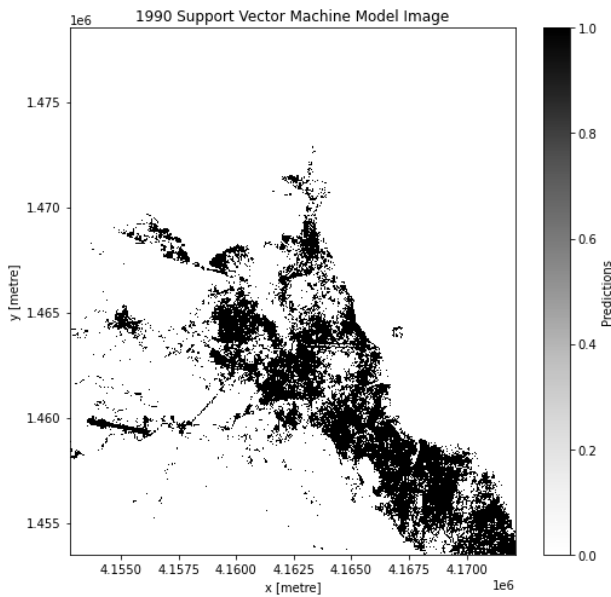
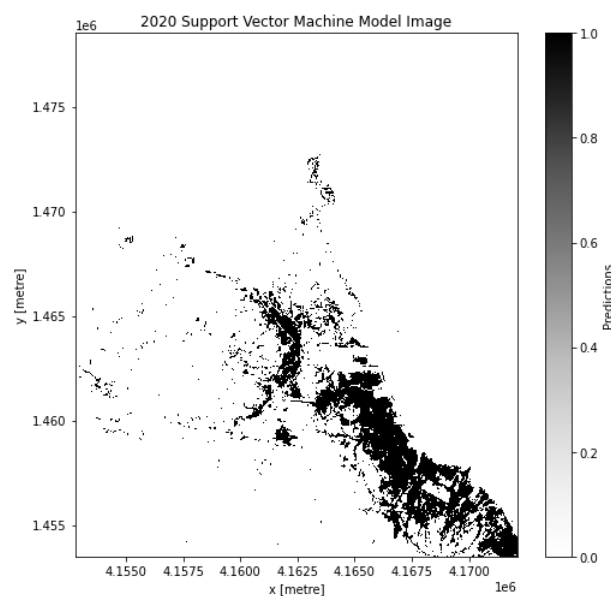Figure 9. Support vector machine results for 1990.

Figure 11. NDVI results for 1990 in greyscale.

Figure 10. Support vector machine shows decrease for 2020.

Figure 12. NDVI results for 2020 in greyscale.

The NDVI results initially appeared to show a slight increase in vegetation from 1990 – 2020. However, after further inspection, the apparent change was likely due to the differences in Landsat satellite eras as the initial imagery for 1990 was Landsat 5 and the latest imagery in 2020 was Landsat 8. While difficult to determine in the greyscale output (Figure 11 and Figure 12) once the data was imported into ArcGIS Pro and the differences calculated, there was still a decreasing trend in total vegetative area.
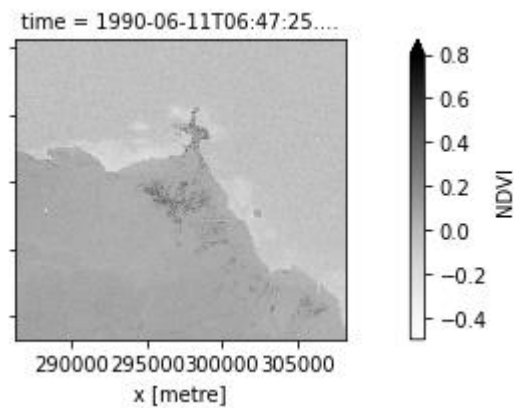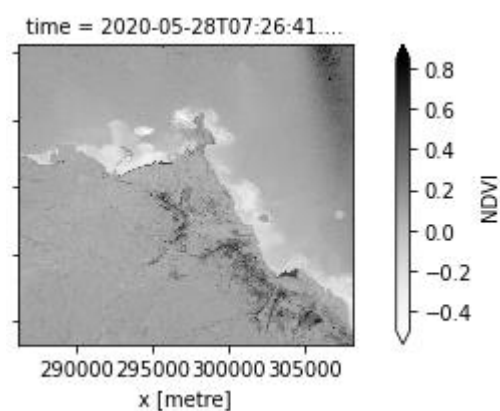
All the models and NDVI were able to show the decrease in vegetation very well, but the none of the models were sensitive enough to differentiate between actual cropland and normal vegetation (which is still very sparse in the region).

The study was limited in that the training set only had two categories of area (cropland and not cropland). Cropland and desert vegetation have similar spectral signatures warranting a deeper study using ground truth ("in situ") collection to establish the difference. Combining multitemporal and multispectral image cubes with long term climate and weather trend data would show a more accurate picture of what was happening with crop output over the study period.

While the data analysis shows a decreasing trend in total cropland across the thirty-year study period, there are numerous variables to consider before drawing drastic conclusions. Even when the results of the analysis are completely accurate and reliable, it is still important to consider this is only a thirty-year glimpse into a country's vegetative health.

Lastly, the results showing a decreasing trend in healthy vegetative areas, while limited in scope as mentioned earlier, are worrisome enough to warrant deeper study of the region and vegetative health trends over time. Given the historical food security issues of the region, more study would be useful to develop a clearer picture of potential future agricultural needs of the country of Djibouti and its effect on the surrounding region.

REFERENCES

Africa, D. (2020). "Digital Earth Africa." from https://www.digitalearthafrica.org.

Africa, D. E. (2022). GeoMAD. D. Africa.

Agriculture, U. S. D. o. (2022). "Cropland." Natural Resources Conservation Service 2022, from https://www.nrcs.usda.gov/wps/portal/nrcs/main/national/landuse/crops/.

CEOS. (2016). "Committee on Earth Observation Satellites." from https://ceos.org/ard/.

CIA (2022). The World Factbook. C.-U. Govt. Washington DC.

DEA. (2020). "Agriculture and Food Security - DE AFrica." from https://www.digitalearthafrica.org/why-digital-earth-africa/agriculture-and-food-security.

Esri (2020). World Countries. Esri, Esri.

Hazaymeh, K. and Q. K. Hassan (2016). "Remote sensing of agricultural drought monitoring: A state of art review." AIMS Environmental Science 3(4): 604-630.

Jones, H. G. V., R.A. (2010). Remote Sensing of Vegetation: Principles, Techniques, and Applications. Oxford, UK, Oxford University Press.

L. Breiman, J. F., R. Olshen, and C. Stone. (1984). Classification and Regression Trees. Belmont, CA, Wadsworth.

ODC. (2021). "Open Data Cube." from https://www.opendatacube.org.

Qu, C. and X. Hao (2018). Agriculture drought and food security monitoring over the horn of Africa (HOA) from space. 2018 7th International Conference on Agro-geoinformatics (Agro-geoinformatics), IEEE.

Thrupp, L. A. and N. Megateli (1999). Critical Links: Food security and the environment in the Greater Horn of Africa, Citeseer.

UNFAO (1996). Rome declaration on world food security and World Food Summit plan of action. Rome, Italy, United Nations Food and Agriculture Organization.

Wagner, W., B. Bauer-Marschallinger, C. Navacchi, F. Reuß, S. Cao, C. Reimer, M. Schramm and C. Briese (2021). "A Sentinel-1 Backscatter Datacube for Global Land Monitoring Applications †." Remote Sensing 13(22): 4622-4622.

Wellington, M. J., L. J. Renzullo and M. Hosseini (2021). "High-Dimensional Satellite Image Compositing and Statistics for Enhanced Irrigated Crop Mapping." Remote Sensing 13(7): 1300.

Wheeler, T. and J. Von Braun (2013). "Climate change impacts on global food security." Science 341(6145): 508-513.