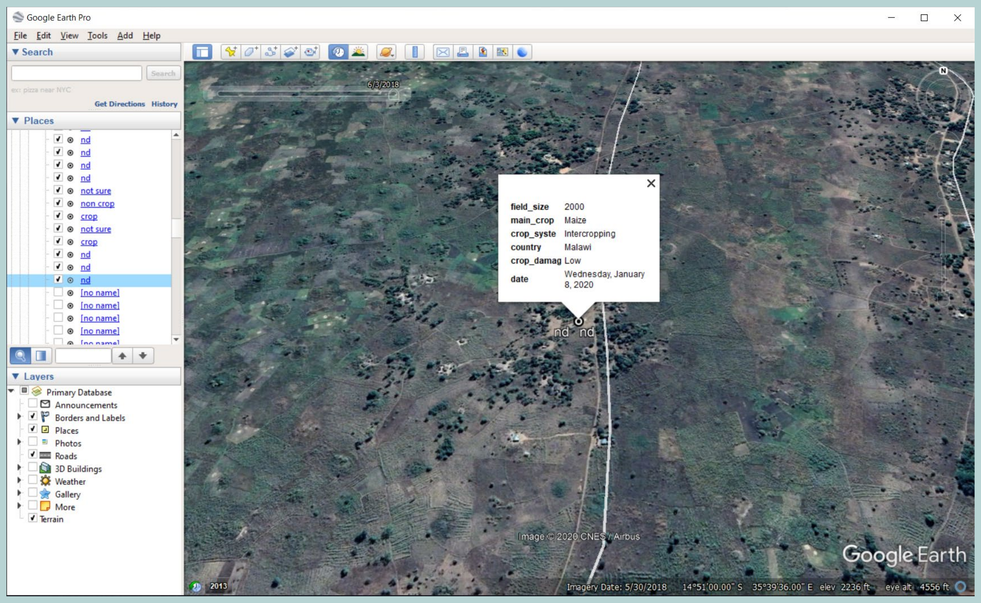
Identifying Erroneous Data Points for Crop Mapping in Kenya

Avi Grant
College Park Scholars – Science & Global Change Program
GIS
agrant19@terpmail.umd.edu
College Park Scholars Academic Showcase, April 30, 2021

Introduction

I worked on the Fall Army Worm dataset of ground truth data used for training models for the NASA Harvest Africa program in order to recognize crops in satellite images. A challenge the researchers have is that with many datasets there are some erroneous samples that are incorrectly located on buildings, roads, etc. instead of fields (likely due to GPS error).



Using Google Earth Pro to analyze and "clean" erroneous crop mapping data.

Site Information:

Name of Site: Department of Geographical Sciences, University of Maryland; NASA Harvest Africa Program. https://nasaharvest.org/initiatives/harvest-africa

Address: 4321 Hartwick Rd. College Park, MD

Your supervisor: Hannah Kerner, Assistant Research Professor at University of Maryland and Machine Learning Lead for NASA Harvest

The site mission: to improve outcomes for farmers, decrease hunger, and lessen food insecurity in sub-Saharan Africa.

Site goals: To train machine learning models to produce maps of where crops (and specific crop types) are growing using satellite data. These maps are critical for decision making and policymaking around agriculture and food security, but the main limitation to producing high-quality models and maps is a lack of ground reference data.

Issues Confronting Site:

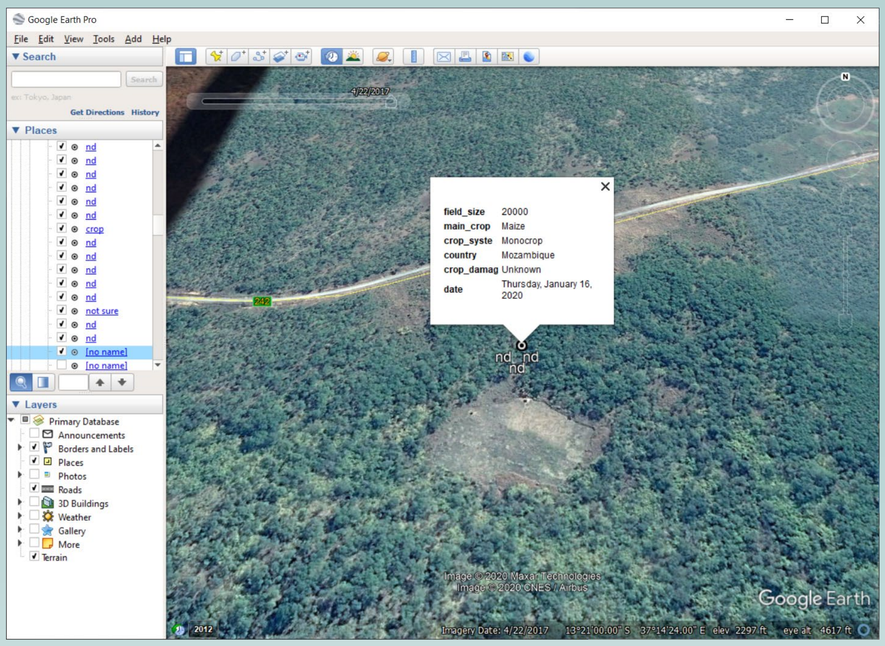
Food security in Africa is a major problem due to climate change and recent events, such as deforestation, droughts, etc. This research project seeks to help address these serious concerns.

Activities:

The models get confused by incorrect labels, so the research team wanted to identify and remove the erroneous points before using the datasets – I helped with this task manually by going through and analyzing the data by hand in Google Earth Pro (a GIS tool), renaming points as to whether they were "crop," "non-crop," "not sure," or "nd" (no data). I reviewed a subset of this FAW dataset, which in total has about 34,000 points.

Impact:

The points I manually cleaned will be used to validate the research team's outlier detections (i.e., to check that the outliers detected by the model are the same points I removed because they are not in fields/cropland).



Using Google Earth Pro to analyze and "clean" erroneous crop mapping data.

Discussion:

This paper describes the methods for crop mapping that researchers, including Ms. Kerner, are developing for Kenya. The paper details the need for crop mapping by explaining that "Spatial information about where crops are being grown, known as cropland maps, are critical inputs for analyses and decision-making related to food security and climate change. Despite a widespread need for readily-updated annual and in season cropland maps at the management (field) scale, these maps are unavailable for most regions at risk of food insecurity."[1]

Another document also gives some larger context into the Harvest Africa program which my research was a part of, specifically, that "Food security is one of the most pressing issues, if not the most pressing, faced by many African countries today. And events in recent years have increasingly strained food supplies for populations in sub-Saharan Africa." The NASA Harvest program that my research is assisting "seeks to strengthen food security by producing and distributing relevant and actionable information on agricultural conditions and production outlooks at national, regional, and global scales. In particular, the Harvest Africa initiative is spearheading the uptake and integration of EO data by national and regional agencies to support decision making and to benefit food security, agriculture, and human and environmental resilience in Africa."[2]

Future Work:

The researchers are currently developing outlier detection methods they can use for excluding erroneous outlier data automatically because manually doing so is tedious and time consuming.

Citations:
[1] Gabriel Tseng, Hannah Kerner, Catherine Nakalembe and Inbal Becker-Reshef. 2020. Annual and in-season mapping of cropland at field scale with sparse labels. Tackling Climate Change with Machine Learning workshop at NeurIPS'20: December 11th, 2020. https://www.climatechange.ai/papers/neurips2020/29/paper.pdf

[2] Nakalembe, C., C. Justice, H. Kerner, C. Justice, and I. Becker-Reshef (2021), Sowing seeds of food security in Africa, Eos, 102, https://doi.org/10.1029/2021EO153329. Published on 25 January 2021.

Acknowledgments:

Site Supervisor: Hannah Kerner; Drs. Holtz & Merck, College Park Scholars, Science & Global Change .