D1G1TAL AgATHON 2020, Washington State University
Landuse classification and monitoring using open-source satellite imagery

Overview

Digital Agriculture can benefit from applications based on free open-source satellite imagery. With satellite imagery you are observing the Earth from space, but the observations are not limited to what your eyes see—they relate to different parts of the electromagnetic spectrum. Different objects on the Earth's surface reflect or emit energy in varying wavelengths. Characteristic patterns in wavelengths reflected or emitted by certain types of objects can give us insights about the properties of objects being observed. For example, when identifying vegetation from space, information obtained in the red and near infrared (and sometimes green and blue) wavelengths of the electromagnetic spectrum is important. Thus, multiple spectral indices (e.g. Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI)) have been created based on characteristic combinations of measurements at these wavelengths. Observing how these indices vary in time can also provide us useful information. For example, the progression of vegetation indices over the year can indicate crop planting, growth and harvest cycles. Additionally, because these cycles vary from crop to crop, this information can be used to distinguish among various crop types. There are numerous applications using spectral indices for land cover mapping, land use identification, and monitoring.

Problem Statement

How can we facilitate landuse classification and monitoring strategies that leverage open-source satellite imagery and machine learning models? This problem statement deals with regional-scale questions rather than on-farm decisions (other themes in the D1G1TAL AgATHON). You will have two sets of data to work with -(a) open source satellite imagery and (b) ground truth data (field boundaries with associated information about the crop grown in a particular year, whether the field was irrigated or not, and the irrigation technology type).

Some example challenges are given here for you to consider. You are free to choose any other challenge that fits within a Digital Agriculture theme using this data as well.

Example challenge 1: Automated crop labeling/identification
Accurate crop labeling offers multiple direct and indirect benefits for Washington State's growers and commodity associations. For example, annual crop acreage statistics are of interest to grower associations and commodity groups for marketing and planning decisions, and state agencies such as the Washington State Department of Agriculture (WSDA) are often approached for this information. Locations and distribution of crops can help growers determine market

saturation and potential growing areas. Additionally, decision support tools for growers can be better tailored to location-specific environmental conditions by having a more accurate picture of where specific crops are grown. This information is also critical for the risk assessment process of pesticide registrations, Endangered Species Act Biological Evaluations (Environmental Protection Agency, National Marine Fisheries Service, United States Fish and Wildlife Service), and essential to minimize impacts to pesticide users and growers in the state. Spatially-explicit crop labelling is also an important input in several hydrological and agricultural applications that inform regional-scale water resources planning decisions.

This data are currently collected by WSDA through a laborious process of drive-by surveys. Therefore automating the crop labeling/identification process will reduce costs, increase spatial coverage and user accessibility and provide accurate/timely crop acreage statistics and trends. Given that the time series signature of vegetation indices could be unique to crops (Figure 1), this information can be used to train, build and apply a model for automated crop labeling.

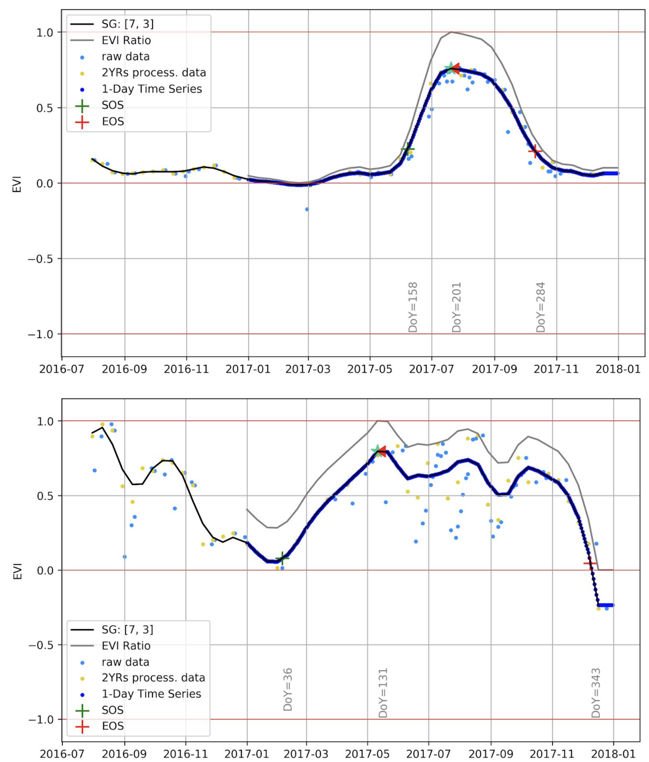

Figure 1. Time series of the Enhanced Vegetation Index for two different crops.

Example challenge 2: Irrigation technology identification

Along with spatially-explicit crop labeling, the type of irrigation technology used in an area is another key input to regional-scale hydrology and agricultural studies. Looking at the earth from high above, we see some fun different shapes such as circles or rectangles or packman. These boundaries of fields can provide clues for the irrigation type. For example, a center pivot irrigation type will be “seen” as a circle from space. Counting the circles can give estimates of total acreage under center pivot irrigation. This is an example where you may not need time series information of vegetation indices. Analysis of fields in just a couple of snapshot images at an appropriate time of the year may be enough.

Other example challenges:

- Distinguishing spatial extents of dryland versus irrigated agriculture.
- Automated crop phenology identification

Dataset Description

Two types of data are provided. The first is open-source satellite imagery and the second is the ground-truth data (field polygon boundaries for 2017/2018 with information about the crop grown in the field for that year and the irrigation type) for machine learning. A irrigation type of "none" corresponds to no irrigation or dryland (rainfed) farming practices. Details for accessing the data are provided in a separate section below.

Satellite Imagery

A harmonized Landsat - Sentinel data product for multiple years of data is available as an Azure open dataset that can be accessed via APIs. The Harmonized Landsat Sentinel-2 (HLS) product includes data from the Landsat-8 (NASA USGS) mission and Sentinel-2 (European Space Agnecy) mission, aligned to a common tiling system at 30m resolution. The timeframe of data available is 2013 to the present for Landsat 8 and 2015 to the present for Sentinel-2. HLS is administered by the National Aeronautics and Space Administration (NASA). The Lansdat 8 data product (L30) is at a 16 day temporal resolution and Sentinel data product (S30) is at a 5 day (or less) temporal resolution. We recommend using the S30 data product for this hackathon, but you are free to use L30 as well, especially if you are interested in using the thermal band.

Sentinel-2 imagery corresponds to multi-spectral optical data with measurements corresponding to 13 bands in the visible, near infrared and short wave infrared part of the electromagnetic spectrum, at a 10 - 60 m spatial resolution (depending on the wavelength), and a 5-day or less temporal resolution. For example images and spectral indices generated with Sentinel 2 data, go to https://gisgeography.com/sentinel-2-bands-combinations/. Remember that what you have in the Azure open dataset is NOT the raw Sentinel 2 imagery, but harmonized Sentinel 2 imagery resampled to a 30 m resolution.

A description of the spectral bands available in the Azure open dataset can be found here:

Harmonized Landsat Sentinel-2 on Azure

If you have prior experience with the native Landsat and Sentinel products you might notice that the band names might be relabelled in both L30 and S30. So pay attention to that aspect.

Depending on your question, you might need to calculate relevant indices from these bands. For example, if you are interested in vegegation indices you might calculate NDVI or EVI as follows.

VI	Mathematical Formula	Harmonized Landsat-Sentinel-2
NDVI	$\frac{NIR - RED}{NIR + RED}$	$\frac{B09 - B04}{B09 + B04}$
EVI	$2.5 \times \frac{NIR - RED}{NIR + 6 \times RED - 7.5 \times BLUE + 1.0}$	$2.5 \times \frac{B09 - B04}{B09 + 6 \times B04 - 7.5 \times B02 + 1.0}$

There are numerous indices useful for different purposes. If you want to look at a comprehensive list, you can check out this link.
http://www.harrisgeospatial.com/docs/broadbandgreenness.html
https://www.tandfonline.com/doi/full/10.1080/24751839.2019.1694765

Once you create a time series of indices you might consider smoothing the data as well.

Cropland Data
We provide a training dataset that includes spatial polygons (for georeferencing) of crop fields labeled with the type of crop that is grown and the type of irrigation technology that is used. If the type of irrigation technology is “none”, it is indicative of dryland (rainfed) irrigation. This dataset was created by the Washington State Department of Agriculture through drive-by surveys.

We have provided you with shapefiles of fields from two counties in Washington State - Grant and Yakima - surveyed in 2017 or 2018. Please make sure you use the appropriate year’s dataset with the appropriate year of satellite imagery (use the LstSrvD column for this). Below we describe what each column in the data represents.

1. ID: This is an ID that labels each field uniquely.
2. CropTyp: Is the type of crop grown in the field. Please note that crop type is a label created by humans who drive by the fields. So, if a field is double-cropped, it will have a label that corresponds to the crop that was being grown at the time of the

survey/driving-by (early or late in the year). Furthermore, these labels are created by visual inspection. So, they may include human error.
3. Irrigtn: The irrigation system used in a given field. “None” implies no irrigation system (dryland agriculture).
4. LstSrvD: The date when the field was last surveyed. If you are using satellite imagery for 2018 you can only ground truth fields that were surveyed in 2018. This columnprovides that information. Just keep in mind that some data was surveyed in 2017 and some in 2018. Just make sure you use the appropriate year of satellite imagery.

Other columns (Acres, IntlSrD, DataSrc, Notes, TRS, county, RtCrpTy, ExctAcr, Shp_Lng, Shap_Ar, CropGrp) are either self-explanatory or not important or both.

Spatial extent

The spatial extent for which data is provided include the Grant and Yakima Counties in Washington State.

How do you access the data on Azure?

Instructions for accessing the satellite imagery and ground truth dataset are provided as “notebooks” in the “Tutorials” section of the TEAMS group.
Tutorials → Topic Setup → Topic 1 Land use classification and monitoring

1. **Harmonized Landsat Sentinel-2 data** (Sample notebooks for data access)
2. **Accessing WSDA data**

```
import os, uuid
from azure.storage.blob import BlobServiceClient, BlobClient, ContainerClient, __version__
url = "https://sidml17885946100.blob.core.windows.net/wsdadata"

sas_token =
"?sv=2019-12-12&ss=bfqt&srt=co&sp=rwdlacupx&se=2020-10-08T14:07:44Z&st=2020-09-25T06:07:44
Z&spr=https&sig=2nguWk%2Bke3e08JSn0iC0MIxQZ%2FEaYSj0jpZ7%2F6qijZA%3D"
container_client = ContainerClient.from_container_url(url,sas_token)
blobs_list = container_client.list_blobs()

for blob in blobs_list:
print(blob.name + '\n')
   blob_client = container_client.get_blob_client(blob.name)
   output_file_name = os.path.basename(blob.name)
withopen(f"./{output_file_name}", "wb") as my_blob:
    download_stream = blob_client.download_blob()
    my_blob.write(download_stream.readall())
```

What tools might be helpful?

This information can be found in the Tutorial section on TEAMS.