Mapping Cropland Extent in Google Earth Engine

The scripts in this repo show how I used Google Earth Engine to estimate the total area of land fallowed in California's Central Valley in 2015 relative to 2010 as an attempt to see what impact the ongoing drought has had on agriculture. The following shows how I approached the project and could be used as a detailed tutorial for someone interested in using Google Earth Engine for a similar project.

Table of Contents

Introduction

Here, we present an approach for cropland extent mapping using the Sentinel-2 and Landsat-8 archive within Google Earth Engine. The generic methodology is capable of handling massive high resolution (30-m or better) time-series (e.g., repeated every 10-16 days) satellite datasets with support of cloud-based Google Earth Engine (GEE) computing platform. In order to support the continent level mapping efforts, this research specifically aims at: (1) integrating data from two state-of-art operational satellite sensors that include 30-m Landsat-8 Operational Land Imager (OLI) and 10-m to 20-m Sentinel-2 multi-spectral instrument (MSI) data; (2) incorporating two machine learning algorithms, the random forest (RF) and vector support machines (SVMs), that are supervised classifications using extensive reference training data, and (3) ingesting textural information from sub-meter to 5-meter very high spatial resolution imagery (VHRI) using Hierarchical Image Segmentation (HSEG) to improve the pixel-based classification result. You can go to check the dynamic web portal of our project.

Setting Up Google Earth Engine

Cloud-based geo-spatial computing platforms and satellite image inventory offer opportunities for mapping croplands to meet the spatial and temporal requirements of broad applications. Google Earth Engine is a tool for analyzing geospatial information. It stores global satellite imagery from the past 40+ years in an organized fashion, facilitating large-scale data analysis. It's a cloud-based platform that uses Google's computational infrastructure for parallel processing, so it can process geospatial data much faster than an ordinary personal computer. You can use Earth Engine either through the Explorer (a high-level point-and-click-only GUI) or through the Playground (a more low-level IDE for writing custom scripts), and it has APIs for JavaScript and Python. Google Earth Engine is currently in beta (as of December 2015), so to access its features, you must fill out the form at https://earthengine.google.com/signup/ and be accepted as an Earth Engine Tester for research, education, and nonprofit usage.

Methdology

A comprehensive overview of the methodology is shown, which consists of following :

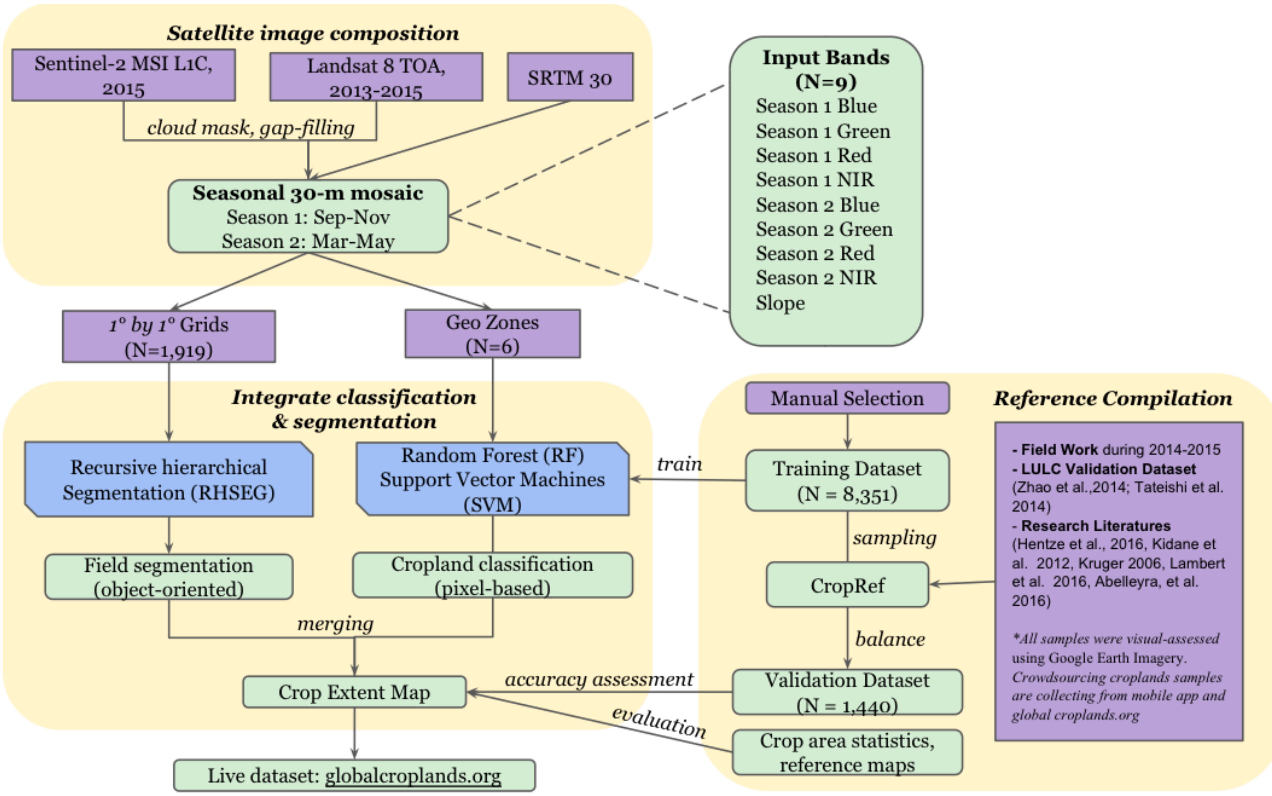

Figure 1. Overview of integrated cropland extent mapping methods. The study integrates pixel-based classification involving the random forests and support vector machines (SVMs) with Recursive Hierarchical Image Segmentation (RHSEG)

Supervised (RF) & Support Vector Machines (SVM) Classifier

The random forest (RF) classifier uses bootstrap aggregating (bagging) to form an ensemble of trees by searching random subspaces from the given data (features) and the best splitting of the nodes by minimizing the correlation between the trees. It is more robust, relatively faster in speed of classification, and easier to implement than many other classifiers (Pelletier, 2016). The RF classification is an ensemble classification, which refers to a new approach that uses not only one, but also many classifiers. Accurate land cover classification and better performance of the random forests model have been described by many researchers.

Though the RF classification model was tuned through training sample editing, errors were identified in some areas, due to the over-fitting error in RF. To reduce the time spent on correcting such kind of issue, Support vector machines was applied when necessary. SVM projects raw input data into a higher dimensional space to increase the separability between different classes when they cannot be appropriately separated by a linear hyperplane. This transformation is realized through different kernel functions and training samples, which cause more scatter after projection into a higher dimensional space. Linear, polynomial, radial basis (RBF) and sigmoidal functions are the most commonly-used core functions, and the performance of the classification model is strongly influenced by different kernel functions.

Training Dataset Collection

TBD

Recursive Image segmentation (RHSeg)

Image segmentation is a principal function that splits an image into separated regions or objects depending on parameters specified. A group of pixels having similar spectral and spatial properties is considered an object in the object-based classification prototype. Hierarchical classification strategies have also been tested by several researchers with a series of per-class classifiers to minimize the effect of spectral confusion among different land cover classes. A recursive hierarchical segmentation (RHSeg) segmentation algorithm (Tilton, 2012) was applied to the imagery. Noting that some image scenes had a significant percentage of water pixels or were masked out due to clouds, we realized that more consistent results would be obtained by selecting results from the RHSeg segmentation hierarchy based on merging thresholds instead of the number of regions. Merge thresholds of 7.5 and 15.0 were selected based on the analysis of several representative images. This RHSEG segmentation results gives detailed agriculture field borders at 30-m resolution image and is ready to be integrated with pixel-based classification results.

Integration of pixel-based (RF, SVM) classification and hierarchical segmentation

Both pixel-based classification and object-based segmentation have their strengths and limitations; so combining the classification results of the two approaches will optimize precision and accuracies of cropland mapping. Per-pixel classification has several limitations. For example, the pixel's square shape is arbitrary in relation to patchy or continuous land features of interest, and there is a spectral contamination among neighboring pixels (known as the modulation transfer function). As a result, pixel-based classification may generate a large number of misclassified pixels (the "salt-and-pepper effect") due to the spectral confusion between land cover types and spectral diversity within the same land cover type. However, pixel-based classification provide fast classification results and easy to seamlessly implement over large areas. Object-based segmentation such as RHSeg can preserve membership information about whether pixels come from the same field. However, HSEG requires is complex to implement over large areas and requires much larger computation resources. Nevertheless, integrating the two approaches will result in best classification results and is a worthwhile pursuit. Thereby, we integrated the the pixel-based classification results with segmented objects, reassigning the classification of all pixels in a region segment to the class with the plurality vote across the region segment. The region is assigned to a "Non-Cropland" class if the plurality vote percentage is lower than a user set threshold, which will vary depending on the field size of the landscape.

Figure 2. The example of (a) a true color sub-meter to 5-m Google Earth Imagery is used for reference and compared

with: (b) combined results from the pixel-based random forest and support vector machine classifier, (c) the object-based RHSeg image segmentation result with edge-based processing window artifact elimination, and (d) the merged results with RHSeg segmentation result with pixel-based Random Forest and support vector machine classification.

How to use the code

Run #1 - #4 in GEE platform and #5 in HPC

- 01-sampling.js create training dataset and validate dataset as well.
- 02-classifying.js supervised classification based on sentinel-2 and landsat 8 seasonal stack
- 03-segment.js export sentinel-2 and landsat 8 seasonal stack to geotiff, as input for Rhseg
- 04-mergestack.js export supervised-classifcation as well as segmentation for merging
- 05-merge.sh merge classifcation and segmetnation, for final results

See also the comments in the scripts while running them.

How to use Rhseg for segmentation

- Go to Rhseg homagepage for downloading.
- Install Rhseg program on your work computing, following Rhseg manual
- Unzip combineff.zip, put it under the same folder with Rhseg, compile the combineff program

Contact

If you have any questions, feel free to contact me

More details about our project and product can be found at: croplands.org

Reference

- Xiong, J.; Thenkabail, P. S.; Gumma, M. K.; Teluguntla, P.; Poehnelt, J.; Congalton, R. G.; Yadav, K.; Thau, D. Automated cropland mapping of continental Africa using Google Earth Engine cloud computing. ISPRS Journal of Photogrammetry and Remote Sensing 2017, 126, 225–244.
- Teluguntla, P.; Thenkabail, P. S.; Xiong, J.; Gumma, M. K.; Congalton, R. G.; Oliphant, A.; Poehnelt, J.; Yadav, K.; Rao, M. N.; Massey, R. Spectral Matching Techniques (SMTs) and Automated Cropland Classification Algorithms (ACCAs) for Mapping Croplands of Australia using MODIS 250-m Time-series (2000-2015) Data. International Journal of Digital Earth 2017, 1–36.
- Teluguntla, P.; Thenkabail, P.; Xiong, J.; Gumma, M. K.; Giri, C.; Milesi, C.; Ozdogan, M.; Congalton, R.; Yadav, K. CHAPTER 6 - Global Food Security Support Analysis Data at Nominal 1 km (GFSAD1km) Derived from Remote Sensing in Support of Food Security in the Twenty-First Century: Current Achievements and Future Possibilities. In Remote Sensing Handbook(Volume II):Land Resources Monitoring, Modeling, and Mapping with Remote Sensing.; Thenkabail, P. S., Ed.; CRC Press: Boca Raton, London, New York., 2015; pp. 131–160.
- Tilton, J. C.; Tarabalka, Y.; Montesano, P. M.; Gofman, E. Best Merge Region-Growing Segmentation With Integrated Nonadjacent Region Object Aggregation. IEEE Transactions on Geoscience and Remote Sensing 2012, 50, 4454–4467.
- Pelletier, C.; Valero, S.; Inglada, J.; Champion, N.; Dedieu, G. Assessing the robustness of Random Forests to map land cover with high resolution satellite image time series over large areas. Remote Sensing of Environment 2016, 187, 156–168.