Springboard Capstone submission for 2021-2022 - Identifying Sand Shoals Using Satellite Derived Bathymetry Methodology

Created: 11/7/2021
Updated: 3/6/2022

Created by: Maggie Satterfield (amtsatterfield@gmail.com)

Background

Bathymetric (bathy) maps are topographic maps for the ocean floor. These maps show us where underwater features, such as sand shoals, may be located and are a valuable tool for engineering, scientific, geophysical, or environmental studies (National...2021). This capstone will reference and replicate methods used by Sawaga et al. in their 2019 case study - Satellite Derived Bathymetry Using Machine Learning and Multi-Temporal Satellite Images.

Problem Statement

Sand shoals are areas of oceanic topography where sand is consistently deposited and produces areas of shallow depth. The areas may change with low and high tide cycles potentially causing hazards to shipping operations. Generally these areas are mapped using LiDAR or sonic methods; however, lack of accessibility or lack of resources can make mapping these areas difficult. Using satellite imagery coupled with machine learning prediction models is a cost-effective reconnaise solution for areas where estimates of shoal or bottom depth are needed.

Data Source

Data will be downloaded from the United States Geologic Survey's LandsatLook web map application ([Explore — LandsatLook (usgs.gov)]). This is an open access application which allows users to download tiles of multi-spectral imagery from the Landsat 9 and earlier satellites. This multi-spectral imagery will be used as the test and train datasets for my ML model. Evaluation datasets will be either LiDAR derived bathymetric information or sonar bathymetric information depending on availability for the chosen test area/tiles. These data are generally available from a local, state, or federal agency. Only areas with this information available will be chosen for this capstone project.

Sawaga et al. designated 5 areas (tiles) where they extracted satellite imagery for training their ML model. Their training images met the following parameters:

- 20% or less cloud cover
- Spatial resolution greater than 30 meters
- Landsat-8 (LaSCR) - Surface reflectance, bands 1-7
- Dates: April 2013 - August 2018

Data Preprocessing and Storage

The high-level data preprocessing workflow I plan to follow is below:

1. Download the Oahu, HI Landsat 8 tile/image (~800 MB per tile)

2. Download LiDAR data for the tile
    a. Acquisition time frame should be similar
3. Create 5,000 random points on the image
4. Extract arrays of color-band values at random points and write to csv. This csv will be divided into test and train datasets for a random forest classifier.

This initial data will be used for training and proof of concept and will be stored on my local drive. A link to this data on Google Drive will be available in my project's GitHub repository.

Computational Resources

- Python 64-bit is recommended (https://www.youtube.com/watch?v=H9MnS2oFN7I)
- 1 tile TIF when stacked is about 800 MB
- QGIS for visualizing data during the processing (~2 GB disk space required for software)

Citations

National Ocean Service (NOS) Office of Coast Survey U.S. Bathymetric & Fishing Maps; Map Types [Internet]. NOAA. [cited 7 November 2021]. Available from https://www.ngdc.noaa.gov/mgg/bathymetry/maps/nos_intro.html

European Space Imaging (ESI); Short-wave Infared Imagery (SWIR) https://www.euspaceimaging.com/wp-content/uploads/2018/06/EUSI-SWIR.pdf

Random Forest Image Classification in Python https://www.youtube.com/watch?v=H9MnS2oFN7I