Radiant MLHub
EARTH IMAGERY FOR IMPACT

Dataset Documentation

Dataset Name: NASA Flood Extent Detection Dataset

Description

This dataset contains synthetic aperture radar (SAR) raster imagery for various flood events acquired from the European Space Agency's Sentinel-1A and Sentinel-1B missions, providing C-Band dual-polarized imagery that spans geographical areas of interest in the United States and Bangladesh. The main emphasis was on the labeling of open water areas where specular reflection of the radar signal off of the relatively still, flat open water surface results in reduced backscatter, low amplitude, and an overall darkened appearance within the image. The labels for the water surface reflectance are also provided in GeoTiff rasterized file format in scenes aligned with the SAR source raster imagery.

Images were processed to a radiometric and terrain-corrected (RTC) image of the radar amplitude, then converted to a grayscale image for visual analysis using the Hybrid Pluggable Processing Pipeline or "HyP3" system which takes the Sentinel archive and uses standard SAR image processing procedures to get to a consistent method of generating the VV / VH amplitude or power imagery. In normal conditions, ponds, lakes, and rivers will appear dark and usually include crisp edges where water adjoins the nearby vegetation and topography. Following heavy rains and flooding, additional dark features occur and often include expanded, flooding growth of dark regions along the normal water areas or standing water in fields or other topographic features where ponding of water is likely.

Citation

Gahlot, S., Gurung, I., Molthan, A., Maskey, M., & Ramasubramanian, M. (2021) "Flood Extent Data for Machine Learning", Version 1.0, Radiant MLHub, https://doi.org/10.34911/rdnt.ebk43x

License

CC-BY-4.0 https://creativecommons.org/licenses/by/4.0/

Creator(s)

The University of Alabama Huntsville

Contact

Iksha Gurung, Shubhankar Gahlot, Muthukumaran Ramasubramanian, Andrew Molthan, Manil Maskey

Location and boundaries

Overall Location Method

- ☐ Ground collection only
- ☐ Ground collection with boundary drawn using imagery
- ☐ Ground collection with spatial buffer added
- ☑ Boundary drawn from imagery
- ☐ Other _____
- ☐ Unknown

Radiant MLHub
EARTH IMAGERY FOR IMPACT

GeoLocation Device

- [ ] Industrial grade GPS (List model) _____
- [ ] Retail grade GPS
- [ ] Mobile Phone GPS
- [x] N/A
- [ ] Unknown

Ground Boundary Method (Details explained in Appendix A)

- [ ] Live/Continuous point capture of walk-around
- [ ] Manual point capture of walk-around
- [ ] Manual point capture of polygon boundaries (not whole field)
- [ ] Manual point capture for later image annotation
- [ ] Manual point capture for spatial buffer within field
- [ ] Manual point capture while looking at but not in field, with heading recorded
- [ ] Other _____
- [x] Unknown

Imagery used (Skip if no imagery used)

Sensor: Sentinel-1B and Sentinel-1B

Date(s): 2017-01-08 to 2019-12-27

List scenes used in Appendix B

Imagery Annotation methods

- [ ] Boundaries drawn based on a single ground point captured
- [ ] Boundaries drawn/edited based on multiple ground points captured
- [ ] Buffer validated from ground point captured
- [ ] Boundary drawn without ground reference data (Include description of methods in Appendix C)
- [x] Pixels annotated without ground reference data (Include description of methods in Appendix C)
- [ ] Unknown

Boundary inclusion

- [ ] Captured polygon includes the entire field/area
- [ ] Captured polygon includes only a sample of the field/area
- [x] N/A

Classification

Classification Type

- ☐ Land cover
- ☐ Crop type
- ☑ Other: Flood Extent

Classes/fields used

- 255 – Water
- 0 – No Water
- 15 – No Data

Image Referenced Classification

Describe methods used in Appendix C

Appendix C: Describe how boundaries and classes were determined without ground reference data

Flood domain experts reviewed the data before it was provided to Earth science students for labeling. Imagery for the various flood events were made available in the ImageLabeler tool developed by the NASA-IMPACT team. Multiple dates of post-event scenes were generated as grayscale imagery enhanced to focus on the contrast of dark, open water features for visual identification. Detailed polygons were drawn for suspected water areas and vetted through discussion with other analysts and project team members with additional SAR imagery expertise. Areas that were “dark” in the SAR images and might not have been water bodies were particularly challenging to examine. Alternate data sources were used to make sure that they were permanent water bodies. These polygons represent the open water class as expert labels and were used to classify open water pixels relative to vegetation and other classes in the image.

Appendix D: List all top-level classes or the classification guidance used

Because the label GeoTiffs were rasterized from grayscale SAR imagery, they have a full range of values between 0 and 255, however only three distinct values are used in the labeling process (as listed above in greater detail here):

- 255 – pixels in the raster image labeled as “water”
- 0 – pixels in the raster image labeled as “no water”
- 15 – pixels in the raster image labeled as “no data”