

## Dataset Documentation

**Dataset Name:** Cloud to Street – Microsoft Flood and Clouds Dataset

### Description

The C2S-MS Floods Dataset is a dataset of global flood events with labeled Sentinel-1 & Sentinel-2 pairs. There are 900 sets (1800 total) of near-coincident Sentinel-1 and Sentinel-2 chips (512 x 512 pixels) from 18 global flood events. Each chip contains a water label for both Sentinel-1 and Sentinel-2, as well as a cloud/cloud shadow mask for Sentinel-2. The dataset was constructed by Cloud to Street in collaboration with and funded by the Microsoft Planetary Computer team.

### Citation

Cloud to Street, Microsoft, Radiant Earth Foundation. (2022). A global flood events and cloud cover dataset (Version 1.0). [Date Accessed]. Radiant MLHub.

<https://doi.org/10.34911/rdnt.oz32gz>

### License

Creative Commons Attribution-ShareAlike 4.0 (CC BY-4.0)

### Creator(s)

[Cloud to Street](#)

### Contact

Cloud to Street

[support@cloudtostreet.info](mailto:support@cloudtostreet.info)

### Tutorials (Optional)

[How to map floodwater from radar imagery using semantic segmentation - benchmark](#)

### 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

#### GeoLocation Device

- Industrial grade GPS (List model) \_\_\_\_\_
- Retail grade GPS
- Mobile Phone GPS



- 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 \_\_\_\_\_ N/A \_\_\_\_\_
- Unknown

## Imagery used (Skip if no imagery used)

Sensor: **Sentinel-1**

Date(s): 2016-08-12 to 2020-10-20

\*\*List scenes used in Appendix B

Sensor: **Sentinel-2**

Date(s): 2016-08-12 to 2020-10-22

\*\*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)
- 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
- N/A

## Classification

### Classification Type

- Land cover
- Crop type
- Other: water, cloud cover, cloud shadow

**Classes/fields used**

Describe in Appendix D

**Appendix B: List imagery scenes used for annotation (ideally also included in metadata)**

---

**Sentinel-1**

Images are derived from GRD (Ground Range Detection) backscatter product.

**Sentinel-2**

Images are derived from Level-1C (Top of Atmosphere) reflectance product.

Sensor	Valid Data Range	CRS	Band Count	Data Type	No Data Value
Sentinel-1	N/A	Matched to coincident Sentinel-2 image	2	float32	0
Sentinel-2	0 - 10,000	Based on source Sentinel-2 image	13	uint16	0

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

---

A label for water classes are provided for each Sentinel-1 and Sentinel-2. Labels are provided for each due to time differences in image capture, difference in detectable water between sensors, as well as cloud cover in Sentinel-2 imagery.

Sentinel-1 has a single label for water, LabelWater, stored as a GeoTIFF. Sentinel-2 has two labels provided (LabelWater & LabelCloud), each stored as a GeoTIFF.

Label Code	Classification Table	Data Type	No Data Value
LabelWater	0: Not Water 1: Water	uint8	255
LabelCloud	0: Background 1: Cloud	uint8	255

2: Cloud Shadow

### Appendix F: Additional details on preview label images

Preview images are provided as an option to quickly view the raw imagery and labels when doing model development / dataset exploration. There are preview images available for the raw data, named with a 3-5 letter abbreviation, and preview images for the labels. All preview labels are provided as PNG files, with an .aux.xml sidecar file that contains coordinate information.

Preview Image Abbreviation	Visualization Name	Sensor	Intended Usage
RFCC	Ratio False Color Composite	Sentinel-1	Water identification
RGB	True Color Composite	Sentinel-2	Cloud / Cloud Shadow Identification
CIR	Color Infra-red Composite	Sentinel-2	Cloud / Cloud Shadow Identification
SWIR	Shortwave Infra-red Composite	Sentinel-2	Water Identification
SWIRP	Shortwave Infra-red Composite (Purple)	Sentinel-2	Water Identification

Preview Image Abbreviation	Hex Table	Sensor(s)
LabelWater	#D3D3D3: Background #0000FF: Water #000000: NoData	Sentinel-1 / Sentinel-2
LabelCloud	#D3D3D3: Background #FCC90F: Clouds #2F4F4F: Cloud Shadows #000000: NoData	Sentinel-2