Data

These data were produced as part of the study, "Satellite imaging reveals increased proportion of population exposed to floods," Nature (2021), doi: 10.1038/s41586-021-03695-w. Each flood map represents the maximum observed flood extent during the time period of the event using the MODIS sensor on board NASA's Aqua and Terra satellites. For a full description of the methods used to generate these maps, please refer to the original paper.

You can access the database as an image collection in Google Earth Engine here: https://developers.google.com/earth-engine/datasets/catalog/GLOBAL_FLOOD_DB_MODIS_EVENTS_V1. You can also download the data as a GeoTiff from https://global-flood-database.cloudtostreet.ai. When downloading, the file comes as a zip file that contains a folder with the GeoTIFF, this metadata document, a license file, and a json that contains a dictionary of the properties associated with the image (matches the properties of the Google Earth Engine images). The zip file naming convention is

DFO_*eventID*_From_*startDate*_to_*endDate*.zip

where *eventID* is the unique ID for the flood event as catalogued in the DFO flood database, *startDate* is the start date of the event in YYYYMMDD format, and *endDate* is the end date of the event in YYYYMMDD format. The GeoTIFF in the zipped folder has the same naming convention but with a ".tif" extension instead of ".zip".

Description of GeoTIFF Bands

Each event map is provided as a GeoTIFF image in WGS 84 Geographic Coordinate system with a pixel resolution of 0.0022458 degrees (250 meters). This GeoTIFF contains 5 bands:

Band 1: Name: "flooded", Min: 0, Max: 1. Maximum extent of flood water during an event.

- 1 – area of surface water
- 0 – no water

Band 2: Name: "flood_duration", Units: days, Min: 0, Max: 65535. Duration of surface water during an event in days. Pixel values indicate the number of composite days a pixel's area was considered water during an event. 3-day MODIS composites were used.

Band 3: Name: "clear_views", Units: days, Min: 0, Max: 65535. Number of cloud-free observations in days between the start and end day of each event. Cloud coverage is determined by the MODIS Quality Assurance band ('state_1km').

Band 4: Name: "clear_perc", Units: %, Min: 0, Max: 100. Percentage of clear view observations during a given flood event. This is equivalent to the 'clear_views' band but normalized to the number of MODIS images per flood event. Cloud coverage is determined by the MODIS Quality Assurance band ('state_1km').

Band 5: Name:"jrc_perm_water", Min: 0, Max: 1. Permanent water determined by the JRC Global Surface Water dataset using the 'transition' band. Resolution is maintained as the original 30-meter resolution of the JRC dataset.
- 1 – permanent water
- 0 – non-water

Description of Algorithms

Standard DFO Algorithm

This method allows for the detection of discrete flood events from daily MODIS imagery. In particular, the DFO algorithm is able to avoid a common misclassification of cloud shadows and hill shade areas as water due to their similar spectral signatures. In this dataset, we apply a modified version of the DFO algorithm by 2 or 3-day composites of images that maintain stationary elements (water) and eliminate mobile elements (cloud shadows) between daily images. These 3-day composites reduce the coincidence of cloud shadows across scenes but miss rapid or flash flood events that are highly transient, creating more false negatives. This method provides a relatively accurate approach for observing flood events with an 83% overall accuracy (Tellman et al. 2021).

Otsu modified DFO Algorithm

The DFO flood detection method described above utilizes two thresholds to identify water pixels within an image including: a ratio of NIR and RED bands (NIR/Red Ratio) and a threshold of the SWIR band with standard values of 0.70 and 675, respectively. Using these standard thresholds for this analysis, misclassifications can occur, highlighting the need for adjustments to these thresholds.

To determine optimal thresholds for distinguishing land from water a technique known as Otsu thresholding was used (Otsu 1975). In short, the Otsu threshold determines the interclass variance within a land/water histogram to find the threshold of the greatest interclass variance. The maximum interclass variance indicates the optimal threshold for detection of water versus land. To determine the highest interclass variance and thus the optimal threshold, different thresholds were tested in a stepwise fashion (~100 thresholds) to identify where the interclass variance peaks. Otsu thresholding is known to work best when the water class represents a significant portion of pixels within the histogram and are not obfuscated by clouds. This technique has been successful employed in several studies including detection of watercourses in the Murray-Darling Basin in Australia using Landsat-8 (Donchyts et al. 2016), river delineation of the Brahmaputra River in India using Landsat-5 (Yang et al. 2014) , and surface water detection in the Yangtze River Basin in China (Li et al. 2013). See the original article for our method to apply and estimate Otsu thresholds (Tellman et al. 2021). The image property “threshold_type” indicates if Otsu or standard thresholds were used to produce the flood map. This property can be found either attached to the Google Earth Engine asset or in the JSON file containing a dictionary of properties in the zip archive with the GeoTIFF (see below).

Dictionary of image properties:
The properties available for each image are available in the table below. These metadata can be linked the corresponding GeoTiff using the "id" which is in the file name as described above. These properties are attached to the Google Earth Engine assets, and are also contained as a dictionary in a json file included in the zip archive when you download it from the website. This json file has the naming convention "DFO_*eventID*_properties.json" where *eventID* is the unique ID from the database (matches the property "id").

Name	Type	Description
id	INT	Unique catalog ID of flood event that aligns with the Dartmouth Flood Observatory (DFO).
cc	STRING	Three-letter ISO country codes (in a list) for countries for which flood water was detected in watersheds intersecting the DFO event polygon.
countries	STRING	Country names (in a list) for countries for which flood water was detected in watersheds intersecting the DFO event polygon.
dfo_centroid_x	DOUBLE	Centroid longitude of DFO polygon estimating the location of a flood event (DFO database).
dfo_centroid_y	DOUBLE	Centroid latitude of DFO polygon estimating the location of a flood event (DFO database).
dfo_country	STRING	Primary country of flooding (DFO database).
dfo_other_country	STRING	Secondary country of flooding (DFO database).
dfo_displaced	INT	Estimated total number of people left homeless or evacuated after a flood event (DFO database).
dfo_main_cause	STRING	The main cause of a flood event in the DFO database. Not normalized.
dfo_severity	DOUBLE	Severity of a flood event (DFO database): 1 - large flood events, significant damage to structure or agriculture, fatalities, and/or 5-15 year reported interval since the last similar event 1.5 - very large events: >15 year but <100 year recurrence interval 2 - extreme events: recurrence interval >100 years)
dfo_dead	INT	Estimated fatalities due to a flood event (DFO database).
dfo_validation_type	STRING	Primary source of flood event confirmation (DFO database). Not normalized.
glide_index	STRING	GLobal IDEntifier Number.
gfd_country_code	STRING	Comma-separated list of two-letter FIPS country codes for countries intersecting with the watershed used as the area of interest in the water detection algorithm.

gfd_country_name	STRING	Country names (in a list) for countries intersecting with the watershed used as the area of interest in the water detection algorithm.
composite_type	STRING	Number of days used for compositing in water detection algorithm.
threshold_type	STRING	Type of thresholds used to classify water/ non-water in water detection algorithm - "otsu" or "standard".
threshold_b1b2	DOUBLE	Threshold value applied to the b2b1-ratio used in the water detection algorithm.
threshold_b7	DOUBLE	Threshold value applied to band 7 (SWIR) used in the water detection algorithm.
otsu_sample_res	DOUBLE	Spatial resolution (in m) of the reducer used to build MODIS mosaic from which to then sample and estimate an otsu threshold (only available for flood events which used an otsu and not the default threshold).
slope_threshold	DOUBLE	Value used to mask steep areas from water detection algorithm to minimize error from terrain shadows.

References:

Coltin, B., S. McMichael, T. Smith, and T. Fong. 2016. Automatic boosted flood mapping from satellite data. *International Journal of Remote Sensing* 37(5):993–1015.

Donchyts, G., J. Schellekens, H. Winsemius, E. Eisemann, and N. van de Giesen. 2016. A 30 m Resolution Surface Water Mask Including Estimation of Positional and Thematic Differences Using Landsat 8, SRTM and OpenStreetMap: A Case Study in the Murray-Darling Basin, Australia. *Remote Sensing* 8(5):386.

Li, W., Z. Du, F. Ling, D. Zhou, H. Wang, Y. Gui, B. Sun, and X. Zhang. 2013. A Comparison of Land Surface Water Mapping Using the Normalized Difference Water Index from TM, ETM+ and ALI. *Remote Sensing* 5(11):5530-5549.

Otsu, N. 1975. A threshold selection method from gray-level histograms. *Automatica* 11(285–296):23–27.

Pekel, J.-F., A. Cottam, N. Gorelick, and A. S. Belward. 2016. High-resolution mapping of global surface water and its long-term changes. *Nature*.

Tellman, B., J.A. Sullivan, C. Kuhn, A.J. Kettner, C.S. Doyle, G.R. Brakenridge, T. Erickson, D.A. Slayback. 2021. Satellites observe increasing proportion of population exposed to floods. *Nature*.

Yang, K., M. Li, Y. Liu, L. Cheng, Y. Duan, and M. Zhou. 2014. River Delineation from Remotely Sensed Imagery Using a Multi-Scale Classification Approach. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 7(12):4726–4737.