https://doi.org/10.5194/essd-2026-577
Preprint. Discussion started: 29 September 2026

Open Access Earth System Science Data Discussions

A spatially and temporally disaggregated inland flood dataset with flood metrics (2000-2024)

Nicole J. Keeney[1] & Frances V. Davenport[1]

[1]Department of Civil & Environmental Engineering, Colorado State University, Fort Collins, 80521, USA

Correspondence to: Frances Davenport (f.davenport@colostate.edu)

Abstract. Floods are some of the most widespread and expensive natural disasters globally, yet current flood research suffers from a lack of detailed historical data on flood events and their impacts at fine spatial and temporal scales. We created a spatially and temporally disaggregated global flood dataset over 2000-2024 by integrating flood disaster records from EM-DAT with satellite-based flood detection and gridded population data. Starting with 4,073 inland flood events, we disaggregate each event to the admin1-month scale (first-level administrative regions by calendar month), generate satellite-derived flood maps using MODIS surface reflectance imagery, and combine these maps with population density data to calculate direct flood-exposed population. The resulting dataset contains 23,334 admin1-month flood records across 2,375 unique administrative regions in 177 countries, with satellite-derived metrics including flooded area and flooded population. Flooded population estimates based on the satellite flood maps are correlated with reported impact metrics, validating our approach despite known limitations in the flooded pixel detection algorithm.

1 Introduction

Floods are some of the most common and damaging natural hazards (CRED, 2025). Flood impacts are far-reaching, and include fatalities, damage to critical infrastructure, increasing disease transmission, disrupting livelihoods, displacing people, and slowing or even reversing development progress (Allaire, 2018; Merz et al., 2021). Although floods can have large impacts in countries of all development levels, they can be particularly devastating in regions with high socioeconomic vulnerability and lower capacity to recover (Doocy et al., 2013; IPCC, 2023; UN Office for Disaster Risk Reduction, 2025; Wing et al., 2022). Additionally, climate change, land use change, and population growth have compounding effects on flood risk and have already altered patterns of flooding and flood impacts on a global scale (Blöschl et al., 2019; Merz et al., 2021; Winsemius et al., 2016). Understanding the changing dynamics and regional disparities of flood impacts around the globe is of urgent interest in order to support climate adaptation, flood risk management, and climate financing efforts.

Historical flood records serve as an important tool for global flood impact studies, which enable researchers, policy-makers, and non-governmental agencies to quantify and plan for differences and changes in flood dynamics between and within

https://doi.org/10.5194/essd-2026-577
Preprint. Discussion started: 29 September 2026

countries. Although some countries maintain detailed flood records, global-scale studies are dependent on coarse global flood records that limit fine-scale analysis. The Emergency Events Database (EM-DAT; Delforge et al., 2025) is often used for global hazard analyses due to its long record and global coverage. However, while EM-DAT fills an important role in providing global-scale data, one of its limitations is incomplete or inconsistent spatial and temporal information across events (Delforge et al., 2025). For example, some EM-DAT events have start and end dates reported, whereas others only have month or year information available. Similarly, some EM-DAT events have location detail at the county-level (or equivalent), whereas other events are only reported at the country- or state-level (or equivalent). For all events, data on human or economic impacts are only provided at the country-level. As a result, many previous global studies using EM-DAT are limited to analysing country-level patterns (e.g. Donatti et al., 2024). Analyses that require more detailed flood data, such as econometric methods to disentangle causal effects of hazard, exposure, and vulnerability on flood impacts, have typically been restricted to individual countries or regions with more complete datasets (e.g. Davenport et al., 2021, Parida et al, 2022). Overall, data quality limitations have been identified as a barrier to studying socioeconomic impacts of floods and other natural hazards (Mahecha et al., 2020), and improved global flood records at a sub-national scale and consistent temporal resolution would fill an important gap for flood impact analysis.

Some prior efforts have been made to improve the level of spatial detail associated with EM-DAT events. At a global scale, the Geocoded Disasters (GDIS) dataset maps EM-DAT events from 1960-2018 to polygon boundaries from the Database of Global Administrative Areas (GADM) based on the 'Location' variable in EM-DAT (Rosvold & Buhaug, 2021). The Geo-Disasters dataset (Teber et al., 2025) performs a similar analysis, mapping EM-DAT hazard events from 1990-2023 to polygon boundaries from the Global Administrative Unit Layers (GAUL) dataset. In both cases, mapping hazard events to geocoded regions makes it possible to analyse EM-DAT events in combination with other spatially explicit meteorological or socio-economic datasets. At the same time, most floods only affect part of an administrative region, so political boundaries inevitably provide imprecise representations of the true flood event boundary.

Satellite-derived flood maps can provide more granular geospatial information about flood area and extent that human-built flood records like EM-DAT lack. MODIS spectral radiance imagery has been previously used to detect the presence of flood water (Sakamoto et al., 2007, Islam et al., 2010, Brakenridge and Anderson, 2006). Tellman et al. (2021) use MODIS imagery to determine the extend of flooding for 913 flood events included in the Dartmouth Flood Observatory, finding that the flood-exposed population has grown by more than 20% over two decades. This type of analysis illustrates how access to satellite-derived flood extent and flood metrics can enable researchers to perform finer scale analyses of floods and flood impacts. Additionally, because of the short revisit time of MODIS imagery (with two overpasses per day), satellite-derived flood maps can help provide more detailed information about the timing and duration of flooding.

https://doi.org/10.5194/essd-2026-577
Preprint. Discussion started: 29 September 2026

Open Access Earth System Science Data Discussions

To address current limitations in flood records, we compile a dataset of inland flood events from 2000–2024 by integrating EM-DAT hazard records with MODIS-derived flooded pixel detection. We first perform extensive preprocessing of the EM-DAT flood records: non–geo-encoded locations are manually geocoded; missing start and end dates are infilled to the first and last day of the month to define detection windows; and multi-region events are disaggregated into sub-events by first administrative ("admin1") unit. For floods spanning multiple months, we also split events into by calendar months to allow for more temporal granularity. Finally, each of the disaggregated admin1-month events is matched to its corresponding administrative boundary to provide spatial bounds for the flooded pixel algorithm. We adapt the flooded pixel algorithm presented in Tellman et al. (2021) to generate a flooded pixel map for each admin1-month event. We then calculate the population exposed to floodwater by overlaying gridded population data and inundated pixel maps. The resulting data products include (i) a repository of satellite-derived flood maps for each successfully mapped admin1–month event and (ii) a tabular admin1–month flood dataset with event-level flood metrics, including flooded population, flooded area, and flooded area normalized by total admin1 area. The dataset includes the original EM-DAT event identifier, allowing re-linkage to EM-DAT records for downstream impact analysis, as well as geo-encoded location information to enable matching to the admin1 boundary shapefiles for additional geospatial analyses.

2 Data and Methods

2.1 Datasets

2.1.1 EM-DAT inland flood events

We analyse events included in the EM-DAT disaster database. EM-DAT is an international hazard dataset that provides information on historical natural disaster events (Delforge et al., 2025). Events within the EM-DAT database meet at least one of these criteria: at least 10 deaths, at least 100 people affected, or an emergency declaration or call for international assistance. Event-specific variables within the EM-DAT database include event start and end dates, location information, the number of people affected, economic damages, as well as other details or descriptions of the event. We analyze all inland floods from 2000-2024, overlapping with the availability of MODIS imagery which starts in 2000. The subset that we analyze consists of 4073 events in total. Of these, 17.3% are classified as flash floods, 49.3% are classified as riverine floods, and 33.4% are classified as general floods.

We use the total people affected from EM-DAT as a comparison with our satellite-derived estimates. From the EM-DAT documentation, the number affected is equal to the total of people injured, made homeless, or impacted by the event in some other way. Fatalities are not included in the total number affected and are instead reported in a separate field. However, the number of fatalities is typically much lower than the total number affected. The total population affected is reported for roughly 90% of inland flood events but is missing for the remaining ~10%.

https://doi.org/10.5194/essd-2026-577
Preprint. Discussion started: 29 September 2026

2.1.2 Global Administrative Unit Layers dataset

We obtain boundaries for sub-national administrative regions from the 2015 Global Administrative Unit Layers (GAUL) dataset (Food and Agriculture Organization of the United Nations, 2015). We analyze the level 1 region (typically referred to as states or provinces) and level 2 regions (typically referred to as counties, districts, or municipalities). We use GAUL as opposed to other administrative boundary datasets such as GADM because many events in EM-DAT include location codes corresponding to GAUL regions. In our analysis, the GAUL region boundaries are used to define the extent of each first-level administrative region when generating satellite flood maps and to calculate the total area of each administrative region.

2.1.3 MODIS satellite imagery

We analyze daily imagery from the Moderate Resolution Imaging Spectroradiometer (MODIS) sensor. There are two MODIS sensors, one aboard NASA's Terra satellite and one aboard the Aqua satellite. Terra and Aqua were launched in 1999 and 2002, respectively. Multiple previous flood or inundation mapping studies have used MODIS imagery (e.g. Sakamoto et al., 2007, Islam et al., 2010, Fayne et al., 2017, Tellman et al., 2021). MODIS has a spatial resolution of 250m and has produced daily imagery with global coverage starting in February 2000, and twice-daily global imagery starting in 2002. The MODIS sensor measures 36 spectral bands; as described below, we analyze the short-wave-infrared, near-infrared, and red bands to perform flooded pixel detection. We access and process the MODIS imagery using the Google Earth Engine API.

2.1.4 Global population density data

We use population density estimates from the NASA Gridded Population of the World (GPW) dataset (CIESIN, 2018). The GPW dataset contains global, gridded population density estimates every 5 years at 1km spatial resolution. We use bilinear regridding to regrid the GPW population data to the higher-resolution 250m grid of the MODIS flood maps. Grid cell areas for each of the flood-map grid cells are calculated using the "xemsf" Python package (Zhuang et al., 2025). We calculate population in each grid cell by multiplying population density by cell area.

2.2 Creation of disaggregated flood event catalog and flood maps

2.2.1 Infilling of flood location and date information

As a result of being a human-compiled database and the challenges inherent in quantifying flood impact data, many of the EM-DAT event entries include missing fields, especially for event dates, locations, and number of people affected. Generating satellite-derived flood maps for an event requires knowing the dates and general location of the event. When these fields are incomplete, we infill the data using the follow steps. For flood events missing a start day, we assign the start day as the first day of the reported starting month. (This step is applied for 278 floods, or 6.8% of the total.) For events missing an end day, we assign the end day as the last of day of the reported ending month. (This step is applied for 268

https://doi.org/10.5194/essd-2026-577
Preprint. Discussion started: 29 September 2026

Open Access Earth System Science Data Discussions

floods, or 6.6% of the total.) We drop floods without a reported start year or month or end year or month from the analysis, as these are considered too subjective for infilling.

We also infill missing location information when possible. EM-DAT location information is provided in four separate fields: "Country," "River Basin," "Location", and "Admin Units". The "Location" field is a narrative description of affected locations, such as regions, cities or towns. The "Admin Units" field includes administrative codes that map directly to GAUL level 1 or level 2 administrative boundaries. Not all events have level 2 admin unit information, so we limit our analysis to admin 1 regions (i.e. state or province), which is the smallest administrative level reported for a majority of events. Events with level 2 administrative units listed are matched to the corresponding admin1 administrative units using. Most events have affected admin1 codes reported, but the "Admin Units" field is missing for 290 events (7.1% of total). In particular, the "Admin Units" field was missing for many events in 2022-2023 and all events in 2024. When the "Admin Units" field was empty, we identified level 1 units manually based on information contained in the "Location" string (275 events). When the "Location" field was also missing, we identified location information from news reports (15 events in total). Some administrative regions were renamed or rezoned after 2015. However, the sub-national boundaries in the GAUL dataset use the 2015 region definitions. For flood events after 2015 that occurred in regions that were renamed or rezoned, we remap the event to the pre-2015 boundaries, so that our final event catalog uses a consistent set of admin 1 definitions across all years. After infilling, the final flood event catalog contains 4,073 inland flood events, which span 2,375 unique admin1 regions across 177 countries.

2.2.2 Defining disaggregated flood events

Next, we disaggregate flood events into separate admin 1 regions and months. The motivation for this disaggregation is to create a dataset with a consistent spatial and temporal resolution, and to provide more spatiotemporal granularity. Disaggregated "admin1-month events" each correspond to one calendar month and one admin 1 region. If an event spans more than one month, the dates of each disaggregated event reflect only the event duration within each calendar month. For example, an inland flood that occurs in one admin 1 region starting on March 23 and ending on April 3 is divided into two admin1-month events: the first event would span March 23-31 and the other would span April 1-3. Likewise, floods occurring across more than one admin 1 region are divided into separate admin1-month events. We assign a unique ID to each of the admin1-month events a unique ID that includes the calendar month, admin1 code, and original EM-DAT ID (Fig. 1).

https://doi.org/10.5194/essd-2026-577
Preprint. Discussion started: 29 September 2026

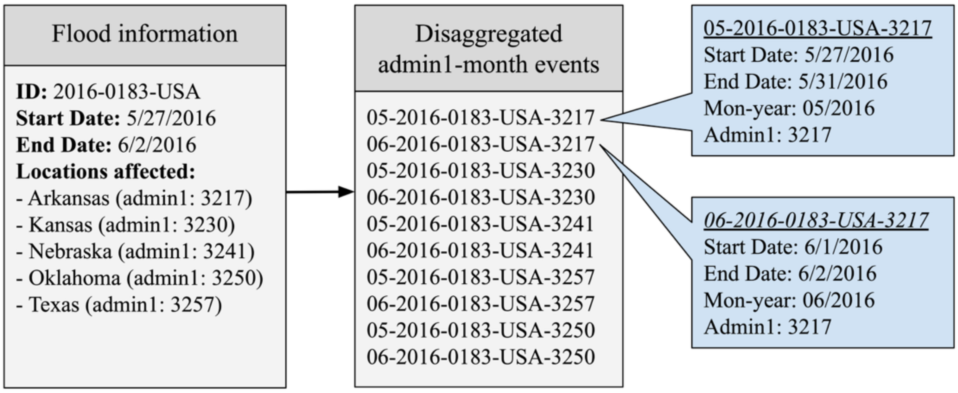

Figure 1: Illustration of the process to disaggregate a flood that spans multiple months and administrative regions.

2.2.3 Satellite-derived flood maps

Satellite imagery allows us to augment human-reported event records with additional fine-scale flood extent information. We use imagery from the MODIS sensors to produce a 250m-resolution map of flooded pixels during each admin1-month flood event. We adapt the flooded pixel detection algorithm described in Tellman et al. (2021) and made available through open-source code (CloudToStreet, 2021). We summarize the algorithm and our modifications below. The reader is referred to Tellman et al., (2021) for a more detailed description.

First, we use Google Earth Engine's (GEE) Python API to obtain surface reflectance imagery from the MODIS sensors. Imagery from the Terra sensor is available for the full study period, while imagery from the Aqua sensor is available starting in July 2002. Using the short-wave-infrared, near-infrared, and red bands, the water detection algorithm classifies each pixel as water (assigned a value of 1) or non-water (assigned a value of 0) using fixed, empirically-derived thresholds. We apply the water detection to all MODIS pixels overlapping the admin1 region associated with the event. For the flooded pixel detection, we analyze all imagery within the dates of the event, plus a one-day buffer before and a one-day buffer after the event. In other words, the number of days included for flood detection ranges from 3 days for an admin1-month event that lasts only one day, or up to 33 days for a 31-day event. (The number of included days is capped at 33 days because admin1-month events last one month or less by definition.) For comparison, the original algorithm from Tellman et al. (2021) uses two buffer days before and two buffer days after the start of each event; we reduce this to only one buffer day to minimize temporal overlap in the detection for admin1-month events that span two adjacent months.

For each day within the admin1-month event, we classify flooded pixels using all imagery from a 3-day window centered on that day. Because there are up to two overpasses per day from Terra and Aqua combined, each daily classification includes up to six observations for each pixel. To classify a pixel as having water, if three or more out of the six observations indicate water presence. We lower this threshold from 3 to 2 (i.e., a pixel must be flagged as water in at least 2 of the observations

https://doi.org/10.5194/essd-2026-577
Preprint. Discussion started: 29 September 2026

rather than 3) for events occurring before the Aqua data is available. The final steps in the algorithm are to apply a terrain slope mask to reduce false positives caused by terrain shadows and to apply a permanent water mask to separate floodwater from pre-existing water bodies.

We combine the daily flood maps for each admin1-event into a single flood map. In the final flood map, flooded pixels correspond to those that were flooded on any day within the event. Additional bands within the flood map report the number of flooded days for each pixel, as well as the number of cloud-free images available for each pixel within the event duration (Table 1).

Table 1. Description of bands stored in each event-specific flood map

Band	Description
Band 1 (flooded)	Binary flood mask (1=flooded, 0=not flooded)
Band 2 (duration)	Number of days pixel was flooded during the event
Band 3 (clear_views)	Number of cloud-free observations during the event
Band 4 (clear_perc_scaled)	Percent of cloud-free observations during the event (0-100)

2.2.4 Calculating flood-exposed population

After generating flood maps, we calculate the total population within flooded pixels for each admin1-month flood event. To do this, we mask the regridded population data with the corresponding flooded pixel map, and sum the total population across all flooded pixels. Because the GPW population data is only available every 5 years, we use the population density data from the most recently available preceding year—e.g., events in 2007 use population estimates from 2005.

2.2.5 Summary of final dataset

The final dataset includes 23,432 disaggregated admin1-month inland flood events across 2,375 unique administrative regions over 2000-2024. These sub-events originate from 4,073 flood events in the EM-DAT dataset. MODIS-derived flood maps were successfully generated for 17,476 admin1-month events (a discussion of missing flood maps is provided in *Section 3.1*). A description of the variables included for each admin1-month are listed in Table 2. Each admin1-month event is associated with an EM-DAT event identifier so that information in this database can be rejoined with additional data fields from EM-DAT. We also provide a set of quality flags for each admin1-month event to keep track of steps or errors generated during the data processing pipeline. These flags are described in Table 3 and are used to keep track of whether spatial or temporal infilling was performed for the event and whether errors were encountered when generating satellite flood maps and flooded population estimates.

https://doi.org/10.5194/essd-2026-577
Preprint. Discussion started: 29 September 2026

205 **Table 2. Description of variables provided for each disaggregated flood event**

Column Name	Description
mon-yr-adm1-id	Admin1-month event ID in the format {MM}-{YYYY}-{EM-DAT ID}-{ISO}-{ADMIN1 CODE}
id	ID of original EM-DAT event ("DisNo." column in original EM-DAT records)
mon-yr	Month and year of event in MMM-YY format
start_date	Start date of event in M/D/YY format (set to first day of month for events with missing start day)
end_date	End date of event in M/D/YY format (set to last day of month for events with missing end day)
ISO	3-letter (alpha-3) code representing a specific country, e.g., "BEL" for Belgium
adm1_code	Global Administrative Unit Layers (GAUL) administrative 1 region code corresponding to admin1-month event location
flooded_population	Total population in the flooded area (missing data set to NaN)
flooded_area	Total flooded area in km^2 (missing data set to NaN)
flooded_area_norm	Flooded area normalized by total admin1 area (missing data set to NaN)
flooded_population_norm	Flooded population normalized by total admin1 population (missing data set to NaN)
flags	Semi-colon separated list of quality flags generated during data processing pipeline

Table 3. Description of quality flags generated during data processing pipeline

Flag	Description
1	Start day originally NaN and was filled with first day of the month
2	End day originally NaN and was filled with last day of the month
3	Flooded pixel map was not generated because Terra Surface Reflectance data not available before 2000-02-25
4	Unspecified Earth Engine internal processing error
5	Missing GPW population data
6	Coordinate mismatch error between floodmap and GPW population data
7	Admin units manually infilled from Location string
8	Admin units manually infilled from literature
9	Missing start/end month and/or year
10	Missing location information; could not be infilled

https://doi.org/10.5194/essd-2026-577
Preprint. Discussion started: 29 September 2026

11	Unable to be disaggregated for unspecified reasons
12	No flooded pixels detected by MODIS water detection algorithm

3 Results

3.1 Validation of the satellite-derived flooded pixel estimates

In Fig. 2, the flooded population estimates based on satellite flood maps are compared to the total number of people affected reported in EM-DAT. This comparison provides some validation of whether the satellite-based flooded population estimates reflect human-reported flood impacts. Socioeconomic impacts in EM-DAT are only reported at the original country-level event scale, so to make this comparison, we combine the monthly satellite-derived flood maps corresponding to each EM-DAT event into a single flood map of the total flood extent during the event (for events spanning one month or less, this is equivalent to the monthly flood map). We then overlap this flood map with the gridded population data to calculate the total population exposed to flooding during the event. We find that the affected population estimates based on satellite flood maps and the EM-DAT reported number of people affected are positively correlated (Fig. 2). While this relationship is statistically significant, the correlation is somewhat low and the flooded population estimates based on satellite flood maps are typically lower than those reported in EM-DAT. This low correlation and the difference between the EM-DAT and satellite-derived values likely stem from multiple causes. First, while EM-DAT provides the most globally comprehensive data on flood impacts, the reported values are only rough estimates, and as such, can easily over- or under-estimate the true number affected. Second, flood impacts inevitably extend beyond the detected flooded area. For example, people outside of directly flooded areas are still impacted by damage to bridges, roads, utilities or other infrastructure, or may experience impeded access to necessities such as food, water, or medical treatment. Our calculation, which only accounts for people residing in the flooded zone, would therefore be an underestimate of the true population affected.

https://doi.org/10.5194/essd-2026-577
Preprint. Discussion started: 29 September 2026

Earth System Science Data
Open Access
Discussions

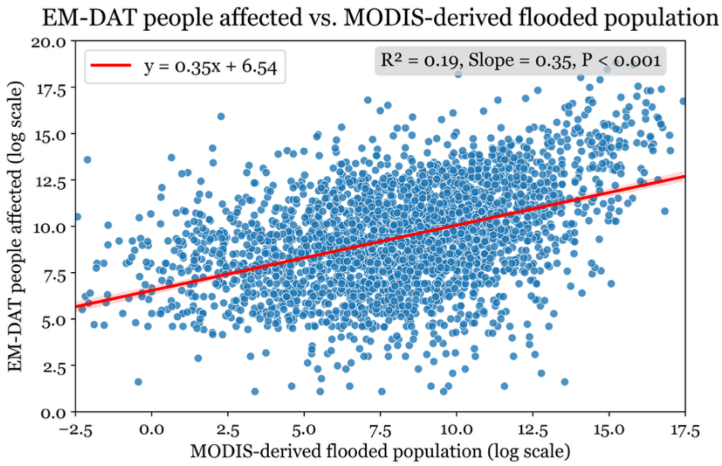

Figure 2: Relationship between the event-level reported number of people affected in EM-DAT and satellite-derived population affected. For each EM-DAT event, the flooded pixel maps for all corresponding admin1-month events were combined. The satellite-derived flooded population was then calculated from this overall flooded pixel map. The regression is based on 2955 flood events with reported population affected in EM-DAT and for which satellite flood maps were successfully generated.

Lastly, the water detection algorithm is known to underestimate the full extent of flooding, especially for shorter floods. Tellman et al. (2021) found multiple reasons that flooded pixels could be undetected, such as cloud cover during the event, inaccurate event dates that don't included the true flood period, issues detecting flash floods that may occur in between MODIS revist times, or urban floods where flooded streets are smaller than the 250m resolution of MODIS imagery. In our analysis, zero flooded pixels were detected for 24% of the admin1-month events. These events are often shorter in duration compared, and have a median duration of only 5 days compared to events with detected floodwater, which have a median duration of 13 days(Fig. 3). For these reasons combined, we expect that the true flood extent is larger than the satellite-derived flooded area.

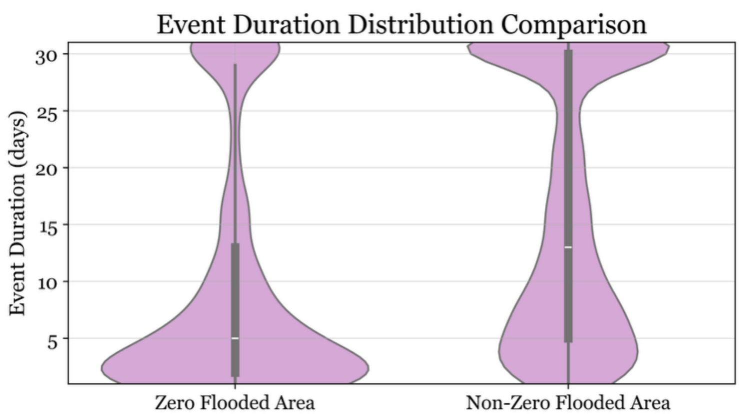

Figure 3: Distribution of admin1-month flood event duration across floods with and without satellite-detected floodwater.

https://doi.org/10.5194/essd-2026-577
Preprint. Discussion started: 29 September 2026

Open Access Earth System Science Data Discussions

3.2 Summary of global flood patterns

In this section, we summarize the characteristics of floods at the disaggregated admin1-month scale. Across admin1 regions, the number of total months with reported floods over 2020-2024 ranges from 1 to 105 (Fig. 4). Most admin1 regions experienced 10 or fewer floods over this period (mean = 9.8, median = 6). A much smaller number of regions are affected by more than 25 events over the study period (Fig. 5). The 20 admin1 regions with the most frequent floods are located in China (8 regions), Colombia (9 regions), India (2 regions) and Pakistan (1 region). The maximum of 105 months with reported flooding occurred in Sichuan Province, China.

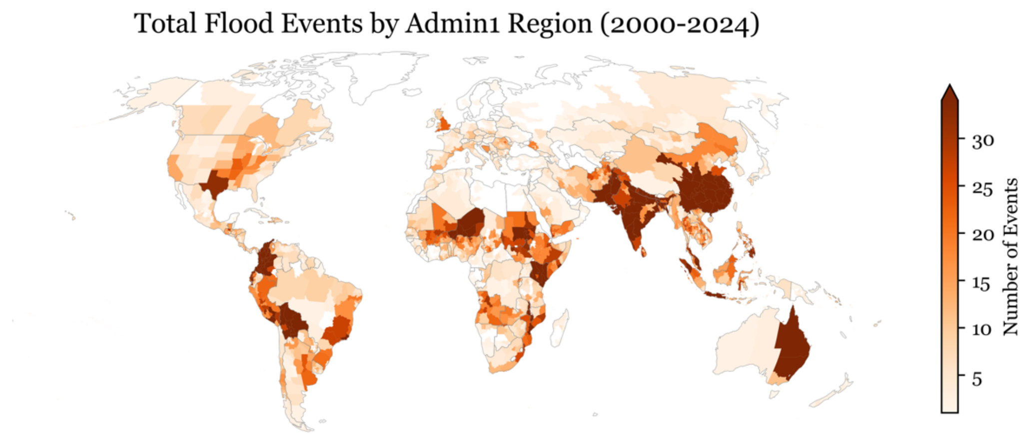

Figure 4: Map of total number of admin1-month flood events (2000-2024) by admin 1 region

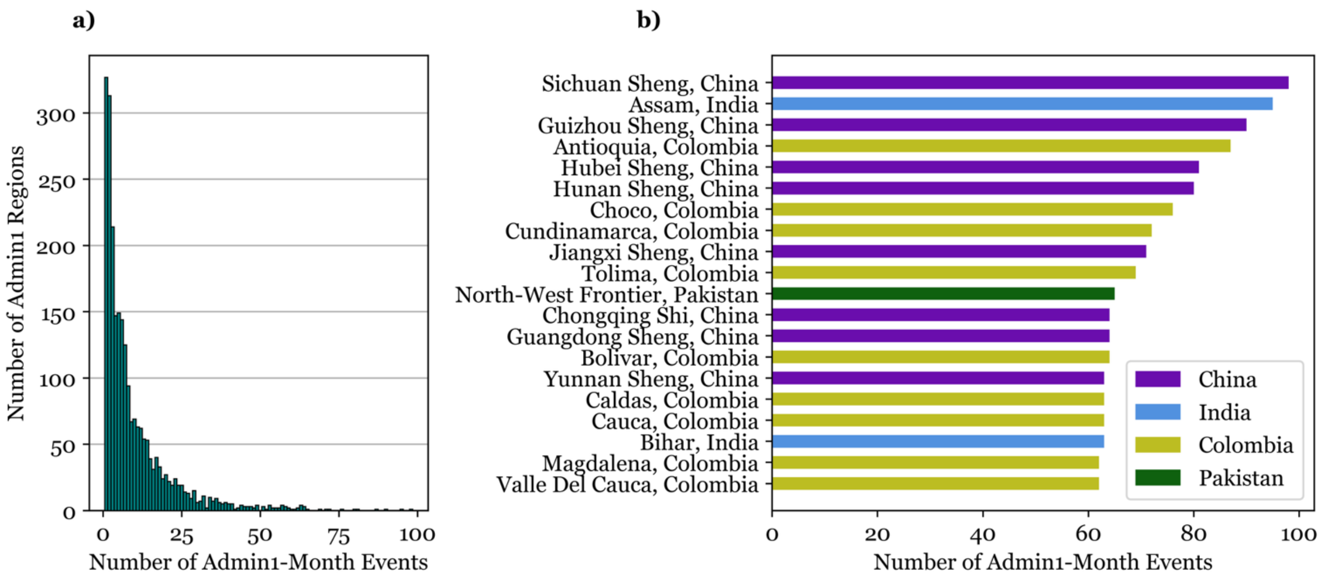

Figure 5: a) Distribution of the number of months with reported flooding per admin1 region. b) Top 20 admin1 regions with the highest number of months with reported flooding.

https://doi.org/10.5194/essd-2026-577
Preprint. Discussion started: 29 September 2026

Many admin1 regions with the largest average flood extent are located in Australia, Canada, central and east Asia, and the Amazon (Fig. 6a). However, flood extent is somewhat a function of the size of the admin1 region itself, and, by definition, large admin1 regions can have larger floods in terms of total area than small admin1 regions. When we normalize flood area by the total admin1 area, we find locations where floods affect a larger percentage of the state (Fig 6b). Multiple areas of the globe have admin1 regions that experience floods that are high in absolute area and normalized flood extent, including the U.S. Midwest, smaller states in South America and Sub-Saharan Africa, Southeast Asia, and some larger states in Siberia. The normalized flooded area is greater than 1.0 for 10.7% of admin1-month events (2,466 of 23,112). This result can be traced to the raster clipping process: when selecting MODIS pixels within a particular admin1 boundary, any pixels that intersect the boundary are included in their entirety. As a result, the total area of the included pixels can exceed the area of the original admin1 polygon, which can cause the flooded area to slightly exceed the polygon area if many of the pixels are flooded. This effect is greatest in very small admin1 regions where boundary pixels constitute a large fraction of the total area.

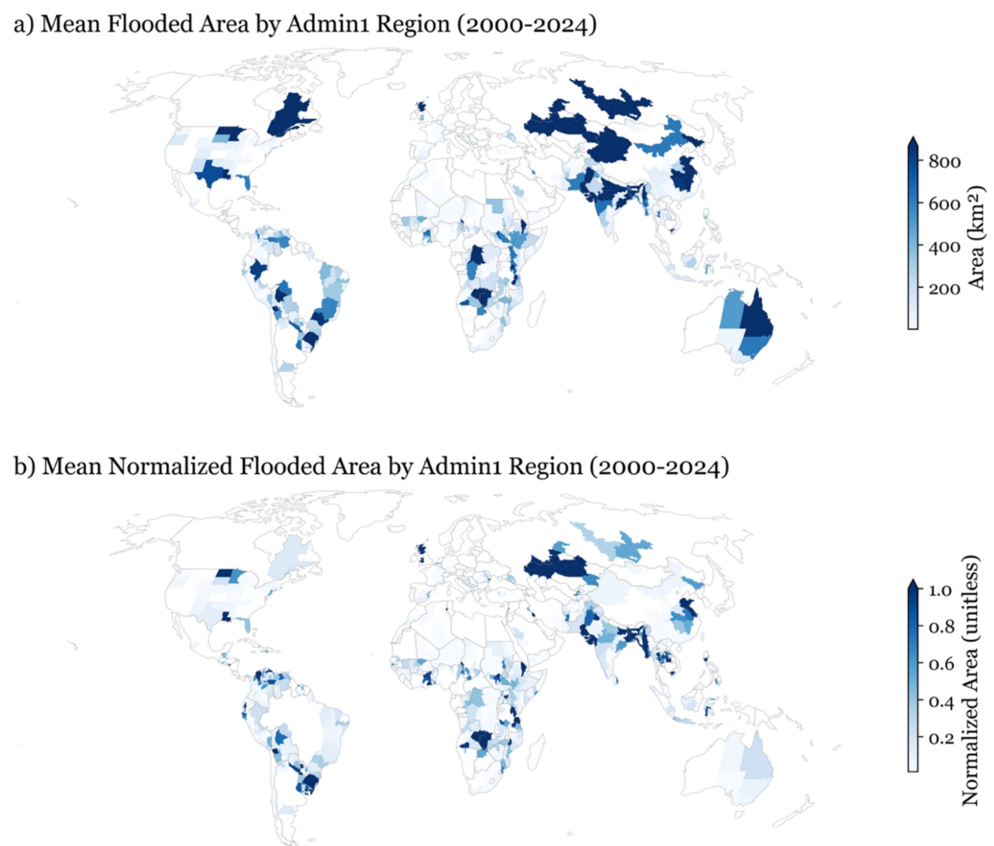

Figure 6: a) map of average admin1-month flood area (2000-2024) by admin 1 region. b) same as a, but with flood area normalized by total admin1 area

https://doi.org/10.5194/essd-2026-577
Preprint. Discussion started: 29 September 2026

Average flooded population per admin1-month event shows patterns that are distinct from those for average flooded area. The regions with the highest people affected per flood occur mostly in South and Southeast Asia, including across the Indian subcontinent, Bangladesh, and Indonesia. In some of these locations, the average admin1-month event affects more than 80,000 people on average. These patterns indicate both high flood occurrence and high population density in these regions. Central and eastern China, central Africa, and isolated areas of South America also include admin1 regions with high values of average population affected per event.

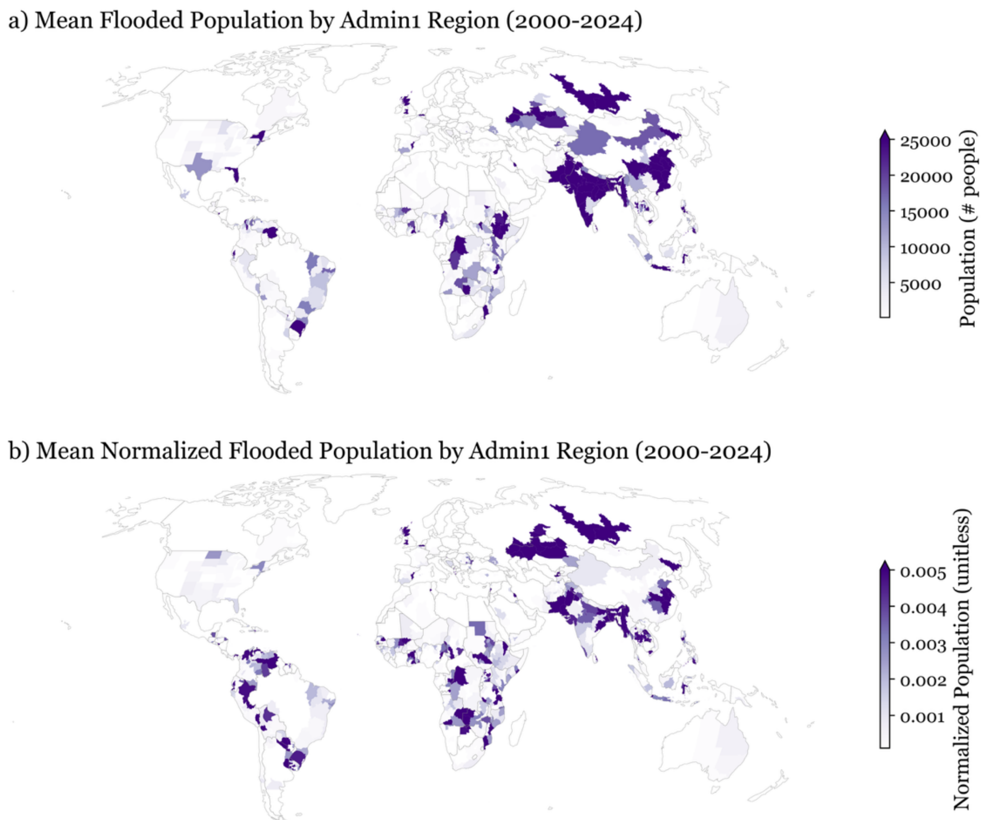

Figure 7: a) Map of average satellite-derived flooded population per admin1-month flood event by admin 1 region. b) Map of average satellite-derived flood population, normalized by the total admin1 population.

4 Discussion and Conclusions

Our study combines over two decades (25 years) of EM-DAT flood disaster data with global population density data and satellite-derived flood maps. This dataset provides more spatially detailed flood map information that enhances recent efforts to add geo-encoded information to EM-DAT events (e.g. Rosvold & Buhaug, 2021; Teber et al., 2025) and covers a larger

https://doi.org/10.5194/essd-2026-577
Preprint. Discussion started: 29 September 2026

number of flood events compared to other satellite-based flood mapping studies (e.g. Tellman et al., 2021). Our flood maps also include flood duration information and have been separated by month for multi-month floods to provide more temporal granularity. In addition to event-specific flood maps, we provide summary flood statistics for each event, include flooded area and population exposed to flooding. The flooded population estimates based on satellite flood maps are significantly correlated with the reported affected population from EM-DAT, indicating that satellite-based flood impact metrics can provide useful supplementary information when reported impact metrics are missing.

While this dataset fills an important gap in historical global flood records by providing higher spatial resolution flood maps associated with major flood events, it is not without limitations. Our dataset is built on flood events reported in EM-DAT, and like many disaster datasets, EM-DAT is subject to reporting biases (Delforge et al., 2025; Gall et al., 2009). It is essential that subsequent flood impact studies take appropriate steps to handle missing data and account for temporal, geographic, or threshold biases present in the underlying data (Jones et al., 2023). In the future, one way to minimize missing event data could be to combine EM-DAT flood records with the global Dartmouth Flood Observatory and region-specific disaster reports. More generally, satellite-based flood detection could be used to identify the occurrence of floods that are missing from all major datasets. These efforts could be valuable directions for future work that would ultimately enable more comprehensive flood research.

Appendices

n/a

Code and data availability

All the code and output data products for this paper are openly available. The code is archived on both Zenodo (https://doi.org/10.5281/zenodo.17905101) and GitHub (https://github.com/nicolejkeeney/emdat-modis-flood-dataset) under the MIT open-source license. The satellite-based flooded pixel maps for each admin1-month event are available as GeoTiffs in a Dryad data repository at https://doi.org/ 10.5061/dryad.mpg4f4rdr (Keeney and Davenport, 2026). Note: DOI will be activated after publication, current dataset link for peer review: http://datadryad.org/share/LINK_NOT_FOR_PUBLICATION/1swIIXfUuKjvgZJkShMkoxhwQO7zy3kEvwaHOR38-74). In the Dryad repository, we also include the EM-DAT derived disaggregated flood record table with satellite-derived metrics at the admin1-month scale in csv format, as well as a reference table explaining data quality and processing flags.

The following datasets were used to generate the output data products. They are not made available in the code or data repositories but are openly available for download. The EM-DAT hazard dataset was made available by the Centre for

Research on the Epidemiology of Disasters in open access for non-commercial use. The dataset is actively maintained; the version used in this paper was downloaded in February 2025. MODIS satellite imagery and the permanent water mask used in the flooded pixel detection algorithm were derived from Google Earth Engine. Administrative level 1 and 2 geospatial boundaries and names are freely available as shapefiles via the GAUL dataset implemented by FAO within the CountrySTAT and Agricultural Market Information System. (AMIS) projects. The GPW dataset used for the flooded population estimates is sourced from NASA.

Supplement link

n/a

Author contributions

NK pre-processed the input data, developed the code, generated the data products and figures, and wrote the initial paper draft. FV conceived of the project, supervised the dataset generation, and contributed to paper writing and editing.

Competing interests

The contact author has declared that none of the authors has any competing interests.

Disclaimer

Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors. Views expressed in the text are those of the authors and do not necessarily reflect the views of the publisher.

Acknowledgements

We acknowledge the Center for Research on the Epidemiology of Disasters for maintaining the EM-DAT database and making it freely available for use. We acknowledge Cloud2Street for providing open-source code for the flooded pixel algorithm, as well as Google Earth Engine for the cloud compute resources for processing the satellite imagery. Portions of this work were performed on the Colorado State University Walter Scott Jr. College of Engineering high-performance computer, Cashew.

https://doi.org/10.5194/essd-2026-577
Preprint. Discussion started: 29 September 2026

Financial support

n/a

Review statement

n/a

References

Allaire, M. (2018). Socio-economic impacts of flooding: A review of the empirical literature. Water Security, 3, 18–26. https://doi.org/10.1016/j.wasec.2018.09.002

Blöschl, G., Bierkens, M. F. P., Chambel, A., Cudennec, C., Destouni, G., Fiori, A., et al. (2019). Twenty-three unsolved problems in hydrology (UPH) – a community perspective. Hydrological Sciences Journal, 64(10), 1141–1158. https://doi.org/10.1080/02626667.2019.1620507

Brakenridge, R., & Anderson, E. (2006). MODIS-based flood detection, mapping and measurement: The potential for operational hydrological applications. In J. Marsalek, G. Stancalie, & G. Balint (Eds.), Transboundary Floods: Reducing Risks Through Flood Management (NATO Science Series IV: Earth and Environmental Sciences, Vol. 72). Springer. https://doi.org/10.1007/1-4020-4902-1_1

Center For International Earth Science Information Network-CIESIN-Columbia University. (2018). Gridded Population of the World, Version 4 (GPWv4): Population Count, Revision 11 (Version 4.11) [Data set]. Palisades, NY: NASA Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4JW8BX5

CloudToStreet. (2021). cloudtostreet/MODIS Global Flood Database (v1.0.0). Retrieved from https://github.com/cloudtostreet/MODIS_GlobalFloodDatabase/releases/tag/v1.0.0

CRED. (2025). Disaster year in review 2024 (CRED Crunch Issue No. 78). Centre for Research on the Epidemiology of Disasters, UCLouvain. https://files.emdat.be/reports/2024_EMDAT_report.pdf

Davenport, F. V., Burke, M., & Diffenbaugh, N. S. (2021). Contribution of historical precipitation change to US flood damages. Proceedings of the National Academy of Sciences, 118(4), e2017524118. https://doi.org/10.1073/pnas.2017524118

https://doi.org/10.5194/essd-2026-577
Preprint. Discussion started: 29 September 2026

Delforge, D., Wathelet, V., Below, R., Sofia, C. L., Tonnelier, M., Van Loenhout, J. A. F., & Speybroeck, N. (2025). EM-DAT: the Emergency Events Database. International Journal of Disaster Risk Reduction, 124, 105509. https://doi.org/10.1016/j.ijdrr.2025.105509

Donatti, C. I., Nicholas, K., Fedele, G., Delforge, D., Speybroeck, N., Moraga, P., et al. (2024). Global hotspots of climate-related disasters. International Journal of Disaster Risk Reduction, 108, 104488. https://doi.org/10.1016/j.ijdrr.2024.104488

Doocy, S., Daniels, A., Murray, S., & Kirsch, T. D. (2013). The Human Impact of Floods: a Historical Review of Events 1980-2009 and Systematic Literature Review. PLoS Currents. https://doi.org/10.1371/currents.dis.f4deb457904936b07c09daa98ee8171a

Fayne, J. V., Bolten, J. D., Doyle, C. S., Fuhrmann, S., Rice, M. T., Houser, P. R., & Lakshmi, V. (2017). Flood mapping in the lower Mekong River Basin using daily MODIS observations. International Journal of Remote Sensing, 38(6), 1737–1757. https://doi.org/10.1080/01431161.2017.1285503

Food and Agriculture Organization of the United Nations. (2015). Global Administrative Unit Layers (GAUL) 2015 [Data set]. FAO. Retrieved from https://data.apps.fao.org/catalog/dataset/global-administrative-unit-layers-gaul-2015

Intergovernmental Panel On Climate Change (Ipcc). (2023). Climate Change 2022 – Impacts, Adaptation and Vulnerability: Working Group II Contribution to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change (1st ed.). Cambridge University Press. https://doi.org/10.1017/9781009325844

Gall, M., Borden, K. A., & Cutter, S. L. (2009). When do losses count? Six fallacies of natural hazards loss data. Bulletin of the American Meteorological Society, 90(6), 799–810. https://doi.org/10.1175/2008BAMS2721.1

IPCC. (2023). IPCC, 2023: Climate Change 2023: Synthesis Report. Contribution of Working Groups I, II and III to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team, H. Lee and J. Romero (eds.)]. IPCC, Geneva, Switzerland. (First). Intergovernmental Panel on Climate Change (IPCC). https://doi.org/10.59327/IPCC/AR6-9789291691647

Islam, A. S., Bala, S. K., & Haque, M. A. (2010). Flood inundation map of Bangladesh using MODIS time-series images. Journal of Flood Risk Management, 3(3), 210–222. https://doi.org/10.1111/j.1753-318X.2010.01074.x

https://doi.org/10.5194/essd-2026-577
Preprint. Discussion started: 29 September 2026

Jones, R. L., Kharb, A., & Tubeuf, S. (2023). The untold story of missing data in disaster research: A systematic review of the empirical literature utilising the Emergency Events Database (EM-DAT). Environmental Research Letters, 18(10), 103006. https://doi.org/10.1088/1748-9326/acfd42

Keeney, N. & F. Davenport: Disaggregated EM-DAT inland flood records with satellite-derived metrics at the admin1-month scale (2000-2024) [data set], Dryad, https://doi.org/10.5061/dryad.mpg4f4rdr, 2026.

Mahecha, M. D., Guha-Sapir, D., Smits, J., Gans, F., & Kraemer, G. (2020). Data challenges limit our global understanding of humanitarian disasters triggered by climate extremes. In J. Sillmann, S. Sippel, & S. Russo (Eds.), Climate Extremes and Their Implications for Impact and Risk Assessment (pp. 243–256). Elsevier. https://doi.org/10.1016/B978-0-12-814895-2.00013-6

Merz, B., Blöschl, G., Vorogushyn, S., Dottori, F., Aerts, J. C. J. H., Bates, P., et al. (2021). Causes, impacts and patterns of disastrous river floods. Nature Reviews Earth & Environment, 2(9), 592–609. https://doi.org/10.1038/s43017-021-00195-3

Parida, Y., Roy Chowdhury, J., Saini, S., et al. (2022). Role of income and government responsiveness in reducing the death toll from floods in Indian states. Scientific Reports, 12, 16978. https://doi.org/10.1038/s41598-022-21334-w

Rosvold, E. L., & Buhaug, H. (2021). GDIS, a global dataset of geocoded disaster locations. Scientific Data, 8, 61. https://doi.org/10.1038/s41597-021-00846-6

Sakamoto, T., Van Nguyen, N., Kotera, A., Ohno, H., Ishitsuka, N., & Yokozawa, M. (2007). Detecting temporal changes in the extent of annual flooding within the Cambodia and the Vietnamese Mekong Delta from MODIS time-series imagery. Remote Sensing of Environment, 109(3), 295–313. https://doi.org/10.1016/j.rse.2007.01.011

Teber, K., Weynants, M., Gans, F., & Mahecha, M. D. (2025). Geo-Disasters: Geocoding climate-related events in the international disaster database EM-DAT. Big Earth Data, 1–16. https://doi.org/10.1080/20964471.2025.2576274

Tellman, B., Sullivan, J. A., Kuhn, C., Kettner, A. J., Doyle, C. S., Brakenridge, G. R., et al. (2021). Satellite imaging reveals increased proportion of population exposed to floods. Nature, 596(7870), 80–86. https://doi.org/10.1038/s41586-021-03695-w

https://doi.org/10.5194/essd-2026-577
Preprint. Discussion started: 29 September 2026

UN Office for Disaster Risk Reduction. (2025). Global Assessment Report on Disaster Risk Reduction 2025: Resilience Pays: Financing and Investing for Our Future. Geneva: United Nations Office for Disaster Risk Reduction. Retrieved from https://www.undrr.org/gar/gar2025

Winsemius, H. C., Aerts, J. C. J. H., Van Beek, L. P. H., Bierkens, M. F. P., Bouwman, A., Jongman, B., et al. (2016). Global drivers of future river flood risk. Nature Climate Change, 6(4), 381–385. https://doi.org/10.1038/nclimate2893

Zhuang, J., Dussin, R., Huard, D., Bourgault, P., & Anderson Banihirwe. (2025, April 29). pangeo-data/xESMF: v0.8.10 (Version v0.8.10). Zenodo. https://doi.org/10.5281/ZENODO.4294774