THESIS

A SPATIALLY AND TEMPORALLY DISAGGREGATED TWENTY FIRST CENTURY GLOBAL FLOOD RECORD FOR FLOOD IMPACT ANALYSIS

Submitted by

Nicole Keeney

Department of Civil & Environmental Engineering

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Fall 2025

Master's Committee:

Advisor: Frances Davenport

James Hurrell
Ryan Smith

ABSTRACT

A SPATIALLY AND TEMPORALLY DISAGGREGATED TWENTY FIRST CENTURY GLOBAL FLOOD RECORD FOR FLOOD IMPACT ANALYSIS

Given that floods are one of the most widespread and costly disasters, there is significant attention on understanding the hydrologic and socioeconomic drivers of these events. However, current research efforts are limited by a lack of detailed, comprehensive, and global-scale data on historical flood events and impacts. In this study, we combine flood damage records with climate reanalysis data and satellite-based flood detection to create a spatially and temporally disaggregated 21st century flood dataset. We start with 4073 inland floods from 2000-2024 contained in the EM-DAT international hazard database. We then disaggregate each event into sub-national administrative levels (e.g., states or provinces) and, for longer duration events, by calendar month to enable analysis at finer spatial and temporal scales. For each event, we derive flooded pixel maps from MODIS satellite imagery. By combining these high-resolution flood maps with gridded population density data, we calculate the number of people exposed to flooded areas in each state or province over time. Lastly, we use climate reanalysis data to extract historical precipitation data at the administrative region and month level in order to characterize the meteorological conditions of each flood. The result is a spatially and temporally consistent dataset of global flood characteristics and impacts to enable more granular impact analysis. In our exploratory analyses, we use the dataset to investigate how historical precipitation contributes to observed flood impacts—both in terms of people affected and economic damages—

across different regions and through time. We find a statistically significant relationship between monthly extreme precipitation and flood impacts using a global-scale fixed-effects panel regression model. In future work, this model could serve as the basis for attributing flood impacts to underlying changes in the precipitation distribution.

TABLE OF CONTENTS

ABSTRACT ii
1. Introduction 1
2. Materials & Methods 3
2.1. EM-DAT flood event catalog 4
2.2. Satellite-derived flood maps and flood-exposed population 7
2.3. Constructing a flood impact panel dataset at the admin1-month level 9
2.4. Panel regression model 11
3. Results & Discussion 13
3.1. Patterns of flood occurrence and impact in EM-DAT 13
3.2 Validation of satellite flood map detection and event disaggregation results 15
3.3 Global patterns of flood impacts at the admin-1 level 18
3.4. Relationship between monthly precipitation and flood impacts 23
4. Conclusions 28
References 31

1. Introduction

Flooding is one of the deadliest and costliest natural disasters. While floods typically occur over the course of days or months, disastrous flood events can have negative impacts lasting years or even decades in affected regions. The effects of floods are widespread, causing fatalities and disease, disrupting livelihoods, damaging or destroying infrastructure, displacing populations, and stalling or reversing development gains, particularly in areas with existing vulnerabilities and limited capacity to respond or recover (Doocy et al., 2013; IPCC, 2023; Lynch et al., 2025; UN Office for Disaster Risk Reduction, 2025; Wing et al., 2022).

The combined—and often compounding—effects of climate change, population growth, and land use change are altering flood patterns and their societal and economic impacts worldwide (Blöschl et al., 2019; Merz et al., 2021; Winsemius et al., 2016). Climate change is increasing the intensity and frequency of extreme precipitation (Diffenbaugh et al., 2017; Dong et al., 2021; Min et al., 2011; Trenberth, 2011), which is likely to affect the intensity and frequency of flooding. However, there are unresolved uncertainties and disagreements in climate projections regarding the magnitude, spatial patterns, and even direction of change in extreme precipitation worldwide (IPCC, 2023), as well as substantial uncertainties in the drivers of flood impacts in different regions (Arnell & Gosling, 2016; Dottori et al., 2018; Sharma et al., 2018).

Further, understanding how changes in extreme precipitation translate into flood impacts remains very difficult to quantify globally. Characterizing this relationship is critical for risk management, planning, allocating disaster relief and compensation, and guiding cost-benefit analyses for climate adaptation. Flood impacts, such as the damage

from an event or the number of people affected, arise due to the physical characteristics of the flood itself interacting with patterns of socioeconomic exposure and vulnerability of the affected region (Intergovernmental Panel On Climate Change (IPCC), 2023). As a result, recent trends in flood impacts reflect both changing climate risk as well as underlying patterns of population growth, development, and adaptation.

Fixed-effects panel regression, a statistical tool borrowed from econometrics, has become a useful tool to isolate the influence of climate variables on a particular outcome. Unlike cross-sectional data, which includes observations across space at a single point in time, or time-series data, which includes observations over time for a single location, panel data includes repeated observations across multiple locations over time. By statistically controlling for potential confounders across both space and time, panel models can quantify causal relationships between variables that would otherwise be difficult to estimate with standard regression or or trend analyses. There is a growing body of research using this methodology in climate impact analyses that supports the utility of this approach, including modelling temperature effects on agricultural yields (Schlenker & Roberts, 2009), tropical cyclone effects on economic activity (Hsiang & Jina, 2014) and on mortality (Young & Hsiang, 2024), and heat wave effects on mortality (Guo et al., 2017; Xu et al., 2025). Davenport et al (2021) used fixed-effects panel regression to quantify how flood damages in the United States are affected by precipitation, showing that around a third of recent flood damages could be attributed to recent changes in precipitation. While panel regression models could theoretically be used to extend this analysis globally, a critical limitation is the lack of standardized global data on flood hazards and flood impacts.

The goals of this study are two-fold. First, we aim to overcome current data limitations by combining multiple data sources with satellite detection methods to create a spatially and temporally disaggregated global flood impact dataset for 2000-2024. This integrated dataset is organized by month and administrative level 1 (i.e., states or provinces), providing a standardized panel dataset on meteorological conditions, flood occurrence, and flood impacts. Second, we apply panel regression analysis to this integrated dataset to quantify how extreme precipitation influences both economic damages and population affected by floods. The results from these regression models can ultimately be used to inform climate change impact attribution and adaptation analyses.

2. Materials & Methods

To support our analysis of the relationship between extreme precipitation and flood impacts, we create a global panel dataset of inland floods from 2000–2024 at the scale of subnational regions and months. First, we disaggregate large and long-duration floods from the EM-DAT database, map flooded extent for each event with MODIS imagery, and overlay population density to estimate the total population exposed to each event. Next, we compute population-weighted flood impacts and combine these with precipitation measurements from a climate reanalysis dataset to produce a standardized panel dataset at a consistent spatial and temporal scale. Finally, we use the dataset to explore global patterns of flood extent and impacts and apply panel regression analyses to assess how extreme precipitation has influenced these outcomes in the 21st century.

2.1. EM-DAT flood event catalog

We analyze inland flood events from the EM-DAT disaster database, a standardized, international hazard dataset that includes event information for floods and other natural and anthropogenic hazards (Delforge et al., 2025). For an event to be included in the EM-DAT database, it must meet one of the following three criteria: ten or more deaths (including dead and missing), 100 or more people affected, and any call for international assistance or an emergency declaration. The dataset is provided in tabular format and includes numerous variables such as economic damages, the number of people affected, location information, and other relevant details for each hazard event. For our analysis, we subset the EM-DAT database to include all inland floods from 2000-2024. This subset consists of 4073 total events, of which 49.3% are classified as riverine floods, 17.3% as flash floods, and 33.4% as general floods. We focus on inland floods rather than coastal floods because the environmental drivers of the latter are more often oceanic (i.e. coastal surge) rather than atmospheric (i.e. precipitation).

We consider two impact variables from the EM-DAT records: total damages in U.S. dollars adjusted for inflation and total number of affected people. In EM-DAT, total damages for events up until 2023 have already been adjusted to 2023 U.S dollar equivalent. We use the same adjustment procedure as EM-DAT to calculate 2023 equivalent damages for all events in 2024 based on the Organization for Economic Cooperation and Development (OECD) Consumer Price Index (Organisation for Economic Co-operation and Development, n.d.). Total affected people is defined as the combined number of people injured, made homeless, or otherwise impacted by the event. Notably, this number does not include fatalities, which are recorded in a separate

field and excluded from our population-weighted impact estimates due to their comparatively low magnitude relative to the total number people affected.

Given the nature of a human-compiled database and the complex realities of characterizing flood and impact data, the EM-DAT database has missing or incomplete data. We aim to maximize the total number of events in our analysis by infilling missing values for event dates and location when appropriate. If the flood start day is missing, we infill that value with the first day of the month (we apply this step for 278 events; 6.8% of total). Similarly, if the end date is missing, we infill that value with the last day of the month (applied for 268 events; 6.6% of total). If a flood is missing a start/end month or year, it is deemed too subjective for manual infilling and is dropped from our subset.

We also address missing spatial information. EM-DAT provides location data through four fields: "Country," "River Basin," "Location" (narrative descriptions of affected places), and "Admin Units" (structured administrative codes and names). The "Admin Units" field maps directly to the 2015 Global Administrative Unit Layers (GAUL) dataset, which defines polygon boundaries for administrative regions worldwide at level 1 (states/provinces) and level 2 (counties/districts/municipalities) (Food and Agriculture Organization of the United Nations, 2015). While some event records contain level 2 location information (county/district/municipality level), many events only contain level 1 information. For this reason, we focus our analysis at the admin 1 (state/province) level. For most events, we use the "Admin Units" field to directly match flood events to affected level 1 regions, but 290 events (7.1% of total) lack this information. We manually infilled missing entries using the "Location" string (275 events) or news reports (15 events) when location strings were also absent. This infilling

was particularly necessary for recent years, as many 2022-2023 events and all 2024 events lacked data in the "Admin Units" field. Some administrative regions underwent renaming or rezoning after 2015, when administrative regions in the GAUL dataset are defined. To address this, post-2015 flood events in renamed or rezoned regions were remapped to their 2015-equivalent boundaries for consistency. The infilled flood event catalog includes 4073 inland flood events across 2375 unique admin1 regions across 177 countries. A substantial subset of events in EM-DAT also contain empty fields for one or more flood impact variables. Missing flood impact information is addressed in a later step when we apply the panel regression analysis (see Section 2.4 *Panel Regression Model*).

Following initial data infilling steps, flood events are disaggregated into separate months and admin 1 regions using the location and date information. The goal of the disaggregation step is to produce a dataset with a consistent spatial (admin 1) and temporal (monthly) resolution, which is a requirement of fixed-effect panel regression methods. Disaggregated events are referred to hereafter as admin1-month events. During the event disaggregation process, each unique admin code is extracted from the "Admin Units" field, and second-level administrative regions are mapped to their corresponding first-level region using the GAUL database. Events are disaggregated so that each admin1-month event corresponds to a single month and a single administrative 1 region. For events that span multiple months, the start and end dates for each admin1-month event is adjusted to the portion of the event that falls within that month. For example, an event occurring within a single administrative 1 region with a start date of May 27 and an end date of June 2 would be split into two admin1-month events: the first spanning May 27-31 and the other spanning June 1-2. Similarly, events

covering multiple regions are split so that each admin1-month event represents only one administrative 1 region. Each admin1-month event is then assigned a unique ID linking it to the corresponding month, administrative 1 code, and original event ID, as shown in Figure 1.

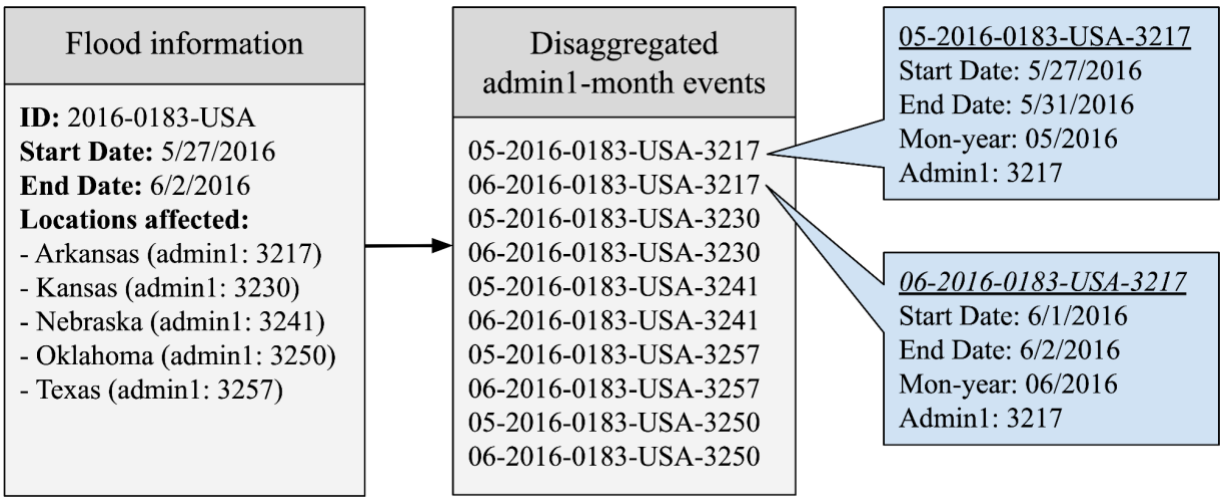

Fig. 1. Example of the disaggregation process for a 2016 flood that spans several months and administrative regions in the southwestern United States

2.2. Satellite-derived flood maps and flood-exposed population

Satellite imagery allows us to provide additional fine-scale flood information that is not available in human-reported event records. For each admin1-month event, we use MODIS satellite imagery to generate a high-resolution map of flooded pixels using a modified version of the inundated flood detection algorithm outlined in Tellman et al. (2021) and further detailed in their open source code repository (CloudToStreet, 2021). Here, we provide a summary of the algorithm, outline our modifications, and discuss its limitations. For more comprehensive details about the algorithm, we refer the reader to the Tellman manuscript.

In brief, the algorithm works as follows. First, Google Earth Engine's (GEE) Python API is used for retrieving surface reflectance imagery from NASA's MODIS sensors aboard the Terra (entire study duration) and Aqua (after July 2002) satellites. The water detection algorithm then utilizes the short-wave-infrared, near-infrared, and red bands from the surface reflectance data, along with fixed, empirically-derived threshold values, to classify each pixel as either water (1) or non-water (0). The water detection algorithm is applied across all 250m MODIS pixels within the flood event geometry, defined as the bounding box of the GAUL polygon of the corresponding admin 1 region. Dates covering the full admin1-month event duration plus a one-day buffer before and after the event are retrieved. This means that the number of days used in the detection algorithm is as short as 3 days for an event with a single day duration, or as long as 33 days for an event with 31-day (full month) duration, with the buffer days on either side. Detection never exceeds 33 days because admin1-month events are, by definition, limited to a maximum of one month.

Our modified multi-day composite method uses a 3-day window centered on the retrieval date of the image. With two satellite overpasses per day, each single-day composite incorporates up to six observations. A pixel is classified as water if at least three of the six observations indicate water presence. Tellman et al. (2021) found that using a multiday composite reduced the rate of false positives. The algorithm then applies a terrain slope mask and permanent water mask to distinguish floodwater from permanent surface water and to further reduce false positives due to terrain or cloud shadows. The final flood map for a given admin1-month event combines all daily composites from the event into a single image, where a pixel is classified as inundated if it was identified as inundated on any day during the event.

Population density data are sourced from NASA's Gridded Population of the World (GPW) dataset, which provides global estimates of population density at 1km resolution in 5-year intervals (Center For International Earth Science Information Network-CIESIN-Columbia University, 2018). We reproject this population density dataset to the finer spatial grid and projection (250m, EPSG:4326) of the satellite-derived flood maps using a bilinear reprojection method so the two datasets can be used together. Grid cell areas are then calculated with the Python package "xemsf" (Zhuang et al., 2025), and population counts are derived by multiplying population density by cell area. Finally, for each admin1-month event, we compute the total population within the flood zone by masking the population raster with the corresponding flood map and summing the population count of the flooded cells. For each admin1-month event, the population density raster corresponding to the closest preceding GPW release year is used—e.g., events occurring in 2003 use population data from the year 2000.

2.3. Constructing a flood impact panel dataset at the admin1-month level

To compare flood impacts on a consistent admin1-month scale, we distribute the impacts of floods that span multiple administrative regions and/or months across the corresponding admin1-month events. For most floods, we are able to compute population-weighted impacts to better capture the uneven distribution of impacts across different months and administrative regions. This process relies on three data sources: the satellite-derived flood maps for each admin1-month event, gridded global population density estimates, and the EM-DAT impact data.

Three approaches are used to allocate impacts across admin1-month events for each flood, depending on the results of the satellite-derived flood maps:

Case 1: All admin1-month events have non-zero flood maps (i.e. flooded pixels were detected in every admin 1 region in every month). Population-weighted impacts are calculated by multiplying each event's fractional flooded population by the impact variables from that flood (approach #1). More specifically, fractional population is defined as the satellite-derived flooded population for that admin1-month event divided by the total satellite-derived flooded population across all admin1-month events for the flood.

Case 2: All events have either zero detected flooded pixels or the flood maps are unable to be generated due to internal GEE errors. In this case, the impact variables are divided equally across all admin1-month events (approach #2).

Case 3: Some admin1-month events have flooded pixels while others do not. Here, a hybrid approach is used in order to maintain information from the non-zero flood maps (approach #3). In this hybrid method, a small baseline fraction (5%) of the total impact is allocated equally among all events with zero flooded pixels (following approach #2), while the remaining 95% is distributed among events with detected flooding, weighted by their flooded population (following approach #1).

Altogether, for 55% of floods we allocate impacts to admin1-events using approach #1, for 19% of events using approach #2, and 26% of events using approach #3.

To account for differences in economic conditions, population density, and geographic area across admin1 regions and years, we normalize flood impacts to enable comparisons across regions. GDP-standardized economic damages are calculated as the damages divided by the admin1 region's yearly GDP, expressed as a percentage. The computation utilizes the mean GDP per capita for each year and admin1 region, which

we extract from the global gridded yearly GDP per capita dataset at 5 arc-minute resolution from Kummu et al. (2025). Because the Kummu dataset is only available through 2022, we use the 2022 GDP values for events in 2023 and 2024. Similarly, the total people affected are expressed as a percentage of the total admin1 population (using GPW-derived population count), and total flooded area is expressed as a percentage of the total area of the admin1 region (using GAUL-derived areas).

Historical precipitation data enable us to add a climatological dimension to our flood dataset. For each admin1 region, we calculate the area-averaged daily mean and 75th percentile precipitation using the open-source Python package "exact_extract" (Baston, 2025). These calculations use daily 0.1° resolution data from GloH2O's Multi-Source Weighted-Ensemble Precipitation (MSWEP) dataset for 2000–2024 (Beck et al., 2019), a bias-corrected product that combines satellite retrievals and reanalysis precipitation data, as well as gauge observations for data pre-2020. For each of the precipitation variables, we compute the mean and standard deviation within each admin1 region across the 2000-2024 time period, and use these values to transform each monthly value into a standardized anomaly.

2.4. Panel regression model

To quantify the relationship between monthly precipitation and flood impacts, we use a fixed-effects panel regression model adapted from Davenport et al. (2021) for two normalized impact variables: total people affected and economic damages. Fixed effects refer to unobserved, and often unmeasurable, differences between panel groups. In our case, we include fixed effects that account for differences in both space and time across administrative regions. These differences encompass variations in socioeconomic

conditions, infrastructure, political climate, reporting practices, average weather, and other factors that might influence actual and reported flood impacts. Year and month fixed effects are used to control for long-term trends, interannual variability, and seasonal patterns in both socioeconomic factors and climate within each admin1. The strength of the fixed-effects model lies in its ability to absorb unobserved differences between groups, isolating the causal relationship between the independent variable (monthly precipitation) and the dependent variable (economic damages or people affected) as captured by the model coefficients. We apply a log-linear and a log-quadratic model, as defined below:

$$\ln(y_{itm}) = \beta P_{itm} + \delta_{it} + \mu_{im} + \varepsilon_{itm}$$

$$\ln(y_{itm}) = \beta_1 P_{itm} + \beta_2 (P_{itm})^2 + \delta_{it} + \mu_{im} + \varepsilon_{itm}$$

where y_{itm} is the normalized flood impact in admin1 i during month m of year t, P_{itm} is the standardized precipitation anomaly for the same admin1, month, and year, δ_{it} is the admin1-year fixed effect controlling for interannual differences within each admin, μ_{im} is the admin1-month fixed effect controlling for seasonal differences and allowing comparison of flood impacts across wet and dry seasons, and β is the regression coefficient. The error term ε_{itm} captures unobserved factors affecting flood impacts, clustered at the country level. By using a log-transformed dependent variable, our regression model captures the expected *percent change* in people affected or economic damages due to an increase in precipitation. Models are implemented in code using the Python package “pyfixest” (The PyFixest Authors, 2025).

Our panel dataset used in this analysis provides admin1-month indexed data covering all unique admin1-month combinations from the event database, which includes months in which no events were recorded. When multiple flood events occur within the same admin1-month, impact variables for each event are summed, resulting in a single aggregate observation for that admin1-month. Because the logarithm of zero is undefined, admin1-months with missing impact data or without a reported flood must either be dropped from the analysis or infilled with a non-zero value. Following Davenport et al. (2021), we use a percentile-based approach to infill impacts during these months with small values. For admin1-months with recorded flood events but missing impact data, we assign these impact variables the 5th percentile values for GDP-standardized damages and normalized total people affected calculated from the subset of events with complete data (0.00138 percent of annual GDP and 0.01689 percent of total population, respectively). For admin1-months with no recorded flood event, we infill impact variables with the 2nd percentile values calculated from the subset of months with complete data (0.00038 percent of annual GDP and 0.00343 percent of total population, respectively). In essence, this infilling approach accounts for the fact that admin1-months without reported impacts in EM-DAT may still have experienced some minor flooding.

3. Results & Discussion

3.1. Patterns of flood occurrence and impact in EM-DAT

Across the subset of inland flood events from 2000–2024, floods affected a total of 1.84 billion people and resulted in 135,491 fatalities. The average number of people

affected per flood is 503,159 and the median is 9,512, with values reported for 90% of floods. The average number of fatalities per flood is 46 and the median is 12, with values reported for 73% of floods. In inflation-adjusted U.S dollars, the average economic damages per flood are $865 million and the median is $95.3 million, with values reported for 28% of floods. Floods in Sub-Saharan Africa and West Asia have the highest incidence of missing economic damage data (90% and 84%, respectively) relative to the number of recorded events. The largest number of flood events are recorded in China, Indonesia, and India (220, 203, and 198 events). Figure 2 shows the temporal distribution of these impacts aggregated by year. There were generally higher numbers of people affected concentrated in the earlier period (2000-2010), higher economic damages in later years (post-2010), and a relatively consistent number of events per year across the study period.

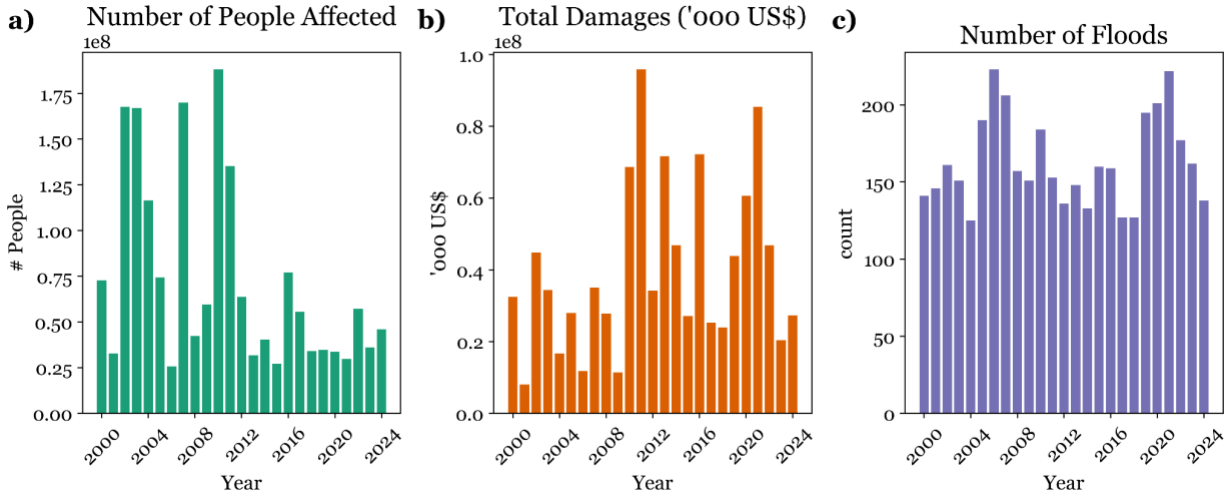

Fig. 2. Annual aggregated impacts from the EM-DAT catalog. **(a)** Number of people affected. **(b)** Total damages adjusted (thousands of US$). **(c)** Number of flood events per year.

Data availability and quality issues in the EM-DAT database pose some limitations and challenges in our analyses. First, there are documented geographic and

accounting biases in both data quality and coverage in the EM-DAT database. Reporting of economic damages tends to be more complete in wealthier or more developed regions. Additionally, across all regions, smaller, lower-impact disasters are less likely to be recorded, resulting in a thresholding bias. As a result of these biases, EM-DAT likely underestimates the true number and severity of historical floods, especially in regions with limited reporting, such as Sub-Saharan Africa, and for smaller, less severe events. As a result, our panel dataset carries these biases from the source data, such that there is likely an underrepresentation of flood impacts in regions with poorer data quality or coverage. That being said, the fixed effects in the panel regression model controls for average differences in flood impacts across admin1 regions, which partially reflects differences in reporting. The impact of data quality on our analysis is discussed in more detail in *Section 3.4.* .

3.2 Validation of satellite flood map detection and event disaggregation results

We compare our satellite-derived flooded population estimates to the total number of people affected during each event as reported in EM-DAT. This comparison allows us to evaluate how well satellite-based flooded population estimates reproduce reported flood impacts. The satellite-derived flooded population estimates (aggregated across all admin1-month events per flood) are positively correlated with the total number of people affected per flood as reported by EM-DAT ($R^2 = 0.20$, slope = 0.35, $p < 0.001$), indicating a statistically significant relationship between independent satellite-based exposure metrics and reported disaster impacts (Fig. 3). The satellite-derived flooded population estimates are consistently lower than the EM-DAT total affected population, likely due to the following two factors. First, flood impacts extend

far beyond the flood zone. People not directly located in inundated areas can still be impacted by floods due to destruction of key infrastructure like bridges, roads, or hospitals, or impeded access to food, water, or medicine. Second, the water detection algorithm likely underestimates the true extent of flooding, especially during short-duration flooding (as detailed below), which would correspondingly reduce the satellite-derived flooded population estimates.

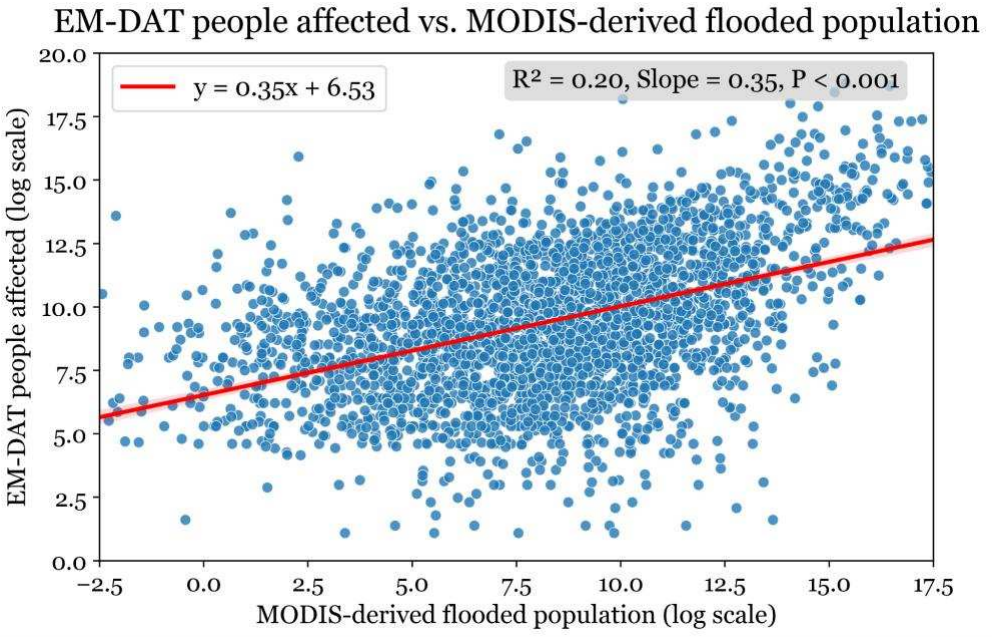

Fig. 3. Regression line on a log scale illustrating the relationship between the total number of people affected per flood and satellite-derived flooded population estimates. For each flood, flooded pixel estimates were generated for individual admin1-month events and then summed to provide a total value at the same flood-level scale of the EM-DAT reported impacts.

There are several limitations to the water detection algorithm that can result in misclassifications of pixels. Tellman et al. (2021) found that the water detection algorithm frequently underdetected flooded pixels for a number of reasons, including extreme cloud cover during the event, incorrect start and end dates reported in the event catalogue that don't actually encompass the full flood period, and challenges in

detecting short-duration floods (such as flash floods) and urban flooding where flooded streets or channels may be much smaller than the MODIS 250m resolution. Using our modified water detection algorithm, 24% of the admin1-month events produce flood maps with no detected floodwater. Events with zero flooded pixels detected are predominantly short in duration, with a median of only 5 days (versus 13 days for events with detected flooded pixel maps), confirming the algorithm's limitations with detecting floodwaters for short duration events (Fig. 4). Given these limitations, we anticipate that the flooded pixel counts for each event represent conservative estimates of the true flooded extent.

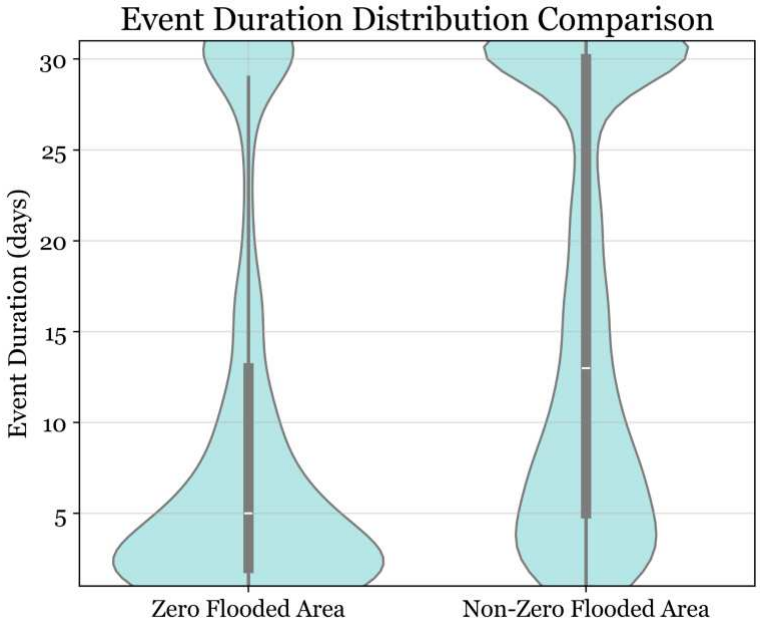

Fig. 4. Comparison of event duration for floods with and without successful flooded pixel detection, showing that events without detectable flooding are predominantly shorter in duration (note that, by definition, the maximum monthly flood duration is 30 or 31 days).

3.3 Global patterns of flood impacts at the admin-1 level

This section presents a summary and analysis of global patterns of flood frequency, distribution, and event impacts at the disaggregated admin1-month scale. As shown in Fig 5, the total number of months with reported flooding in each region ranges from 1 to 105, with the majority of regions experiencing fewer than 10 events throughout the study duration (mean = 9.8, median = 6). The distribution is right-skewed, with a small number of regions disproportionately affected by high event frequencies. The top 20 admin1 regions by event frequency are located in China (8 regions), Colombia (9 regions), India (2 regions) and Pakistan (1 region), with Sichuan Province in China recording the maximum frequency of 105 admin1-month events (Fig. 6).

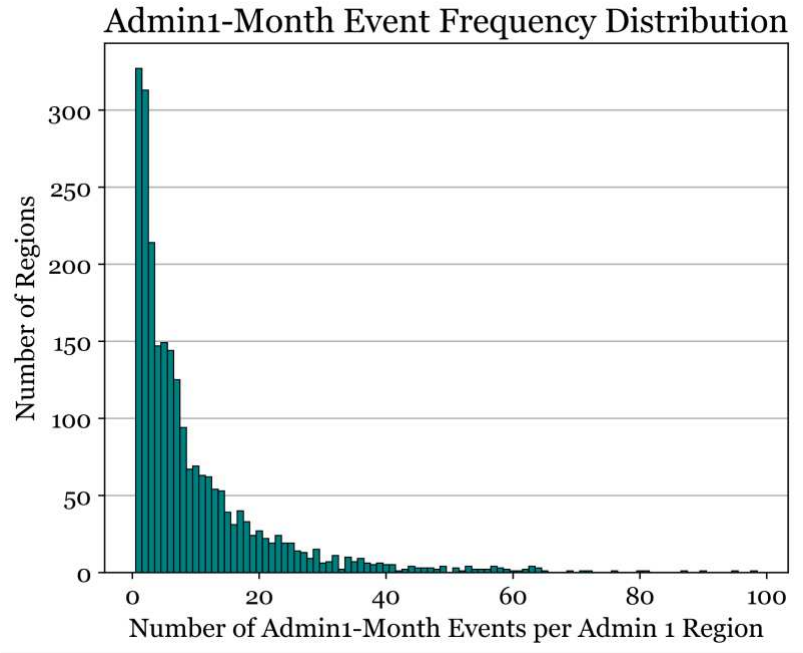

Fig. 5. Histogram of number of admin1-month events per admin1 region, showing a right-skewed distribution.

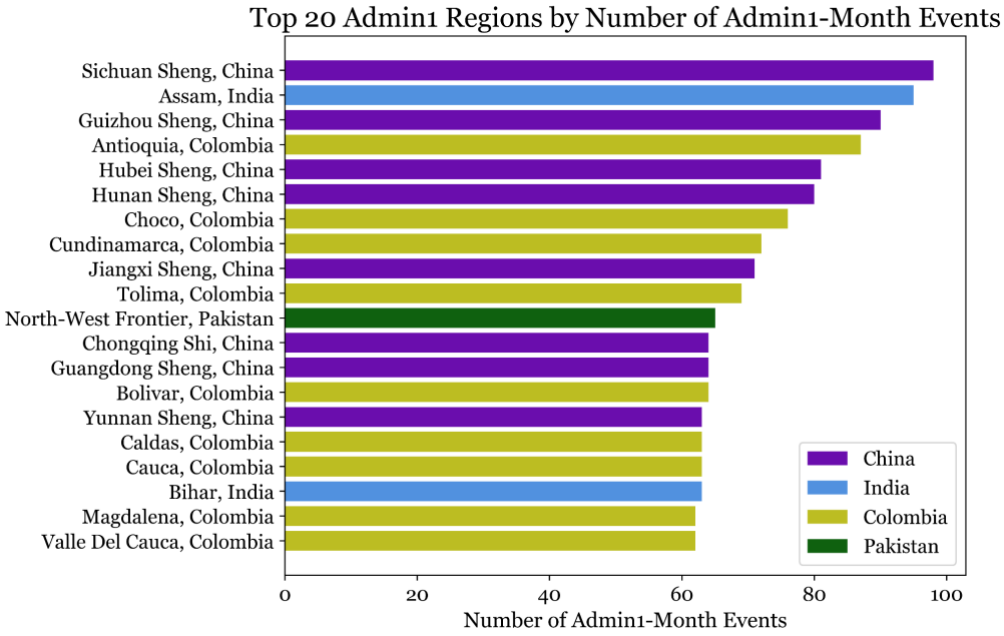

Fig. 6. Top 20 admin1 regions by event frequency, colored by country.

The global distribution of average satellite-derived flooded area per admin1-month event reveals distinct geographical clustering, with the highest absolute values concentrated in Australia, Canada, central Asia, east Asia, and the Brazilian Amazon (Fig. 7). These patterns must be interpreted considering that admin1 regions vary substantially in size globally, with large regions naturally accommodating larger absolute flood extents than small regions. When the flooded area is normalized by the area of the admin1 region, the data indicates regions where floods disproportionately affect large percentages of state territory. This normalization reveals hotspots in smaller admin1 regions, including parts of central Africa, Mexican coastal states, and large portions of Southeast Asia.

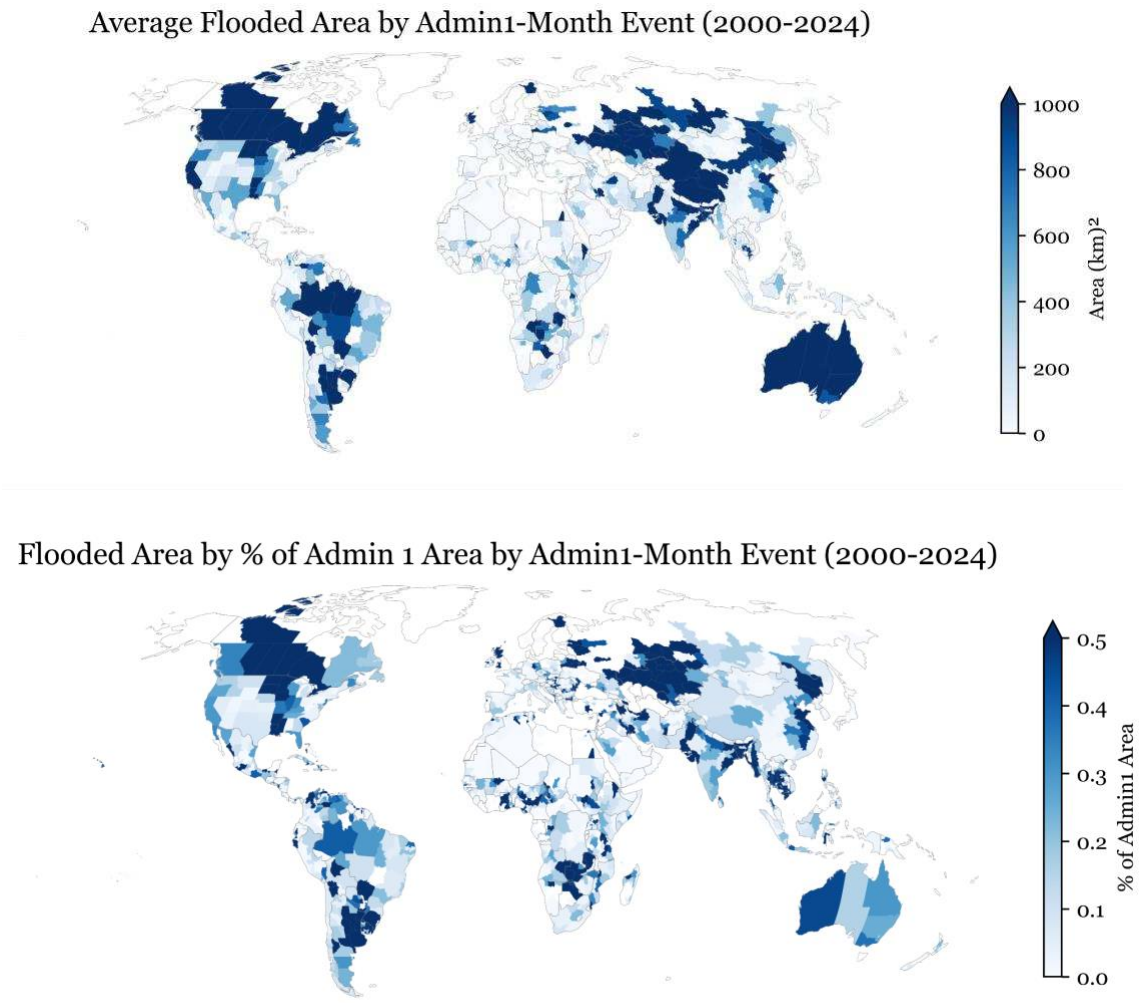

Fig. 7. Average flooded area by admin1 region (top figure) and average percentage of the admin1 population that is flooded (bottom figure).

Global patterns of average number of people affected by each admin1-month event are markedly different from those for flooded area, with absolute impacts concentrated in Asia, central Africa, and parts of South America, where some regions experience average impacts exceeding 80,000 people per admin1-month flood event (Fig. 8). When normalized by admin1 population, the global distribution reveals additional high impact regions across large portions of Africa and along the western and northern coastlines of South America, indicating regions where floods affect large

percentages of the population despite lower absolute numbers. Notably, some regions with high values for average flooded areas– such as Australia and Canada– do not appear as hotspots for population impacts. This could be due to lower population densities in flood-prone regions, more robust flood preparedness infrastructure, and/or more effective disaster response in the immediate aftermath of flooding that reduce the number of people affected.

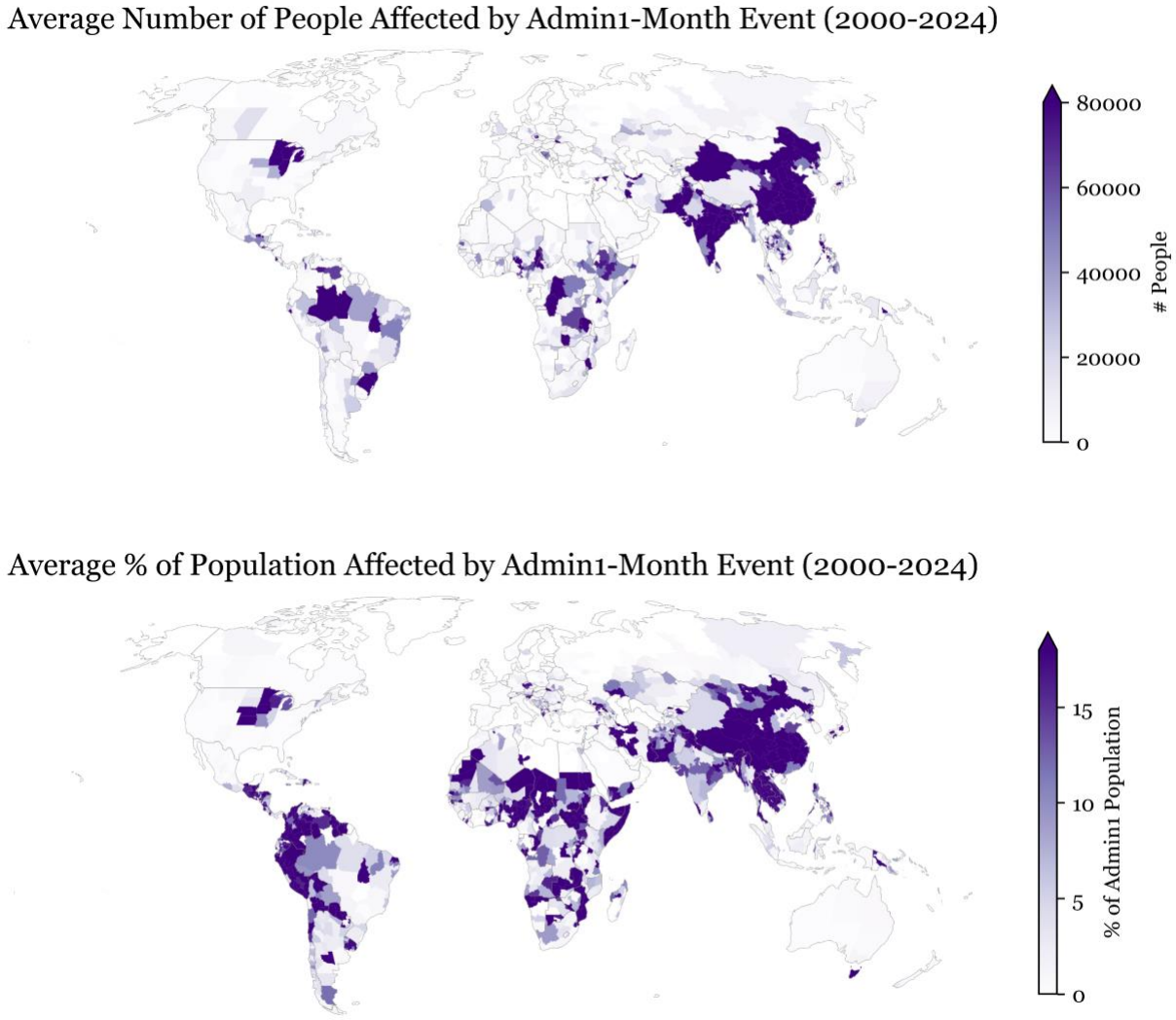

Fig. 8. Average number of people affected by admin1 region (top figure) and average percentage of the admin1 population affected (bottom figure).

Average economic impacts from flooding (Fig. 9) show that absolute damages are concentrated primarily in high-income developed economies including the United States, Canada, Europe, Australia, and Japan, as well large portions of China. Average damages for some admin1 regions in these areas can exceed $350 million per admin1-month flood event. When standardized by regional GDP, the spatial distribution shifts to reveal economic vulnerability hotspots across South America, parts of Africa, and Southeast Asia, with some admin1 regions experiencing average damages of more than 5% of GDP. This normalization demonstrates that while wealthy regions experience the highest absolute economic damages from floods, developing regions face disproportionate economic burdens relative to their economic capacity. Some regions, including the US Midwest and eastern China, emerge as hotspots in both metrics, indicating that floods in these areas generate both high absolute damages and disproportionate economic impacts relative to regional economic output.

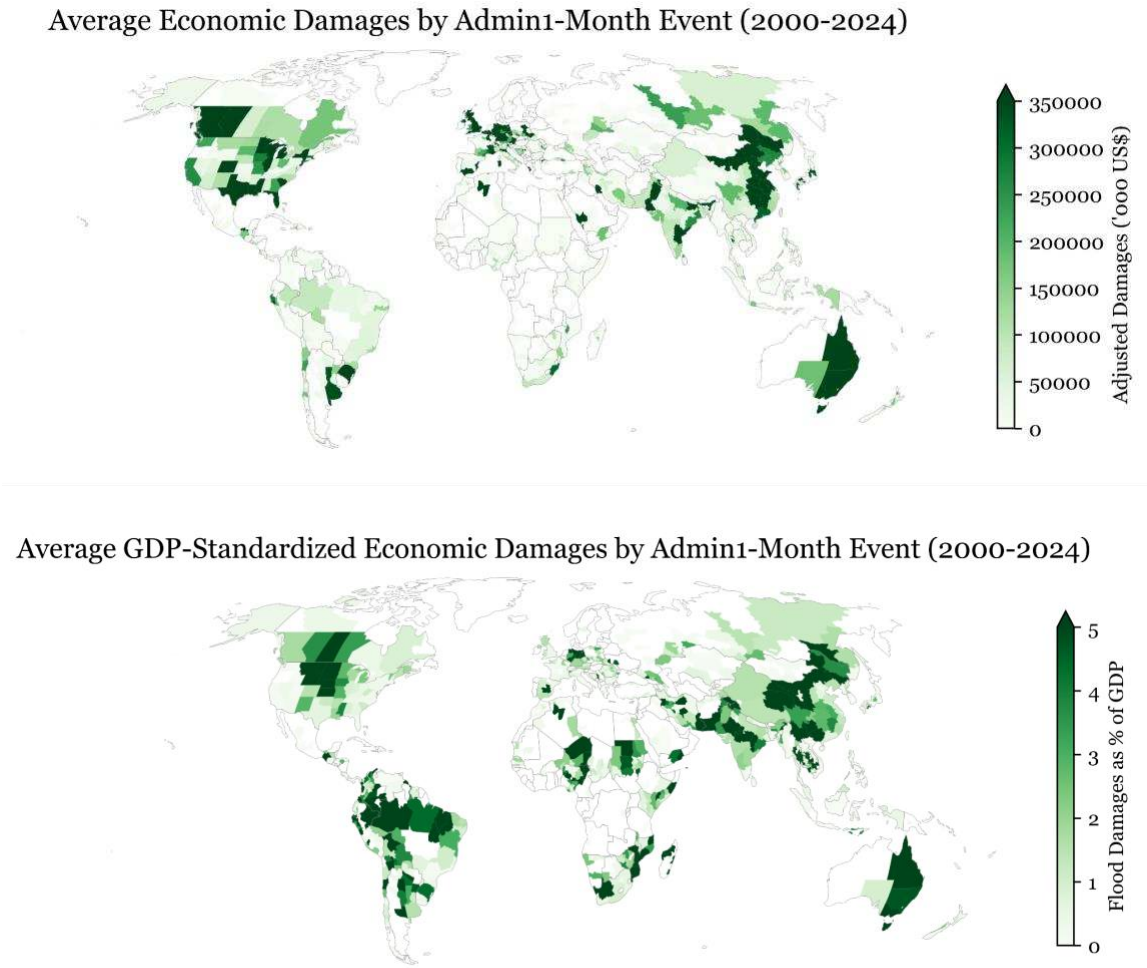

Fig. 9. Average economic damages by admin1 region (top figure**)** and average percentage of the admin1 population that is flooded (bottom figure).

3.4. Relationship between monthly precipitation and flood impacts

We analyze historical precipitation anomalies at the admin1-month level across all months, and during months with reported flooding. Figure 10 shows how flooding and flood impacts are distributed across different monthly precipitation anomalies (measured in standard deviations above/below average). Results show that months in which a flood occurs exhibit higher precipitation anomalies (Fig. 10b) and that more severe impacts concentrate at extreme precipitation levels (Fig. 10c-d). Additionally,

flood probability increases steadily from near zero at negative precipitation anomalies to approximately 0.25 at the highest positive anomalies (Fig. 10e).

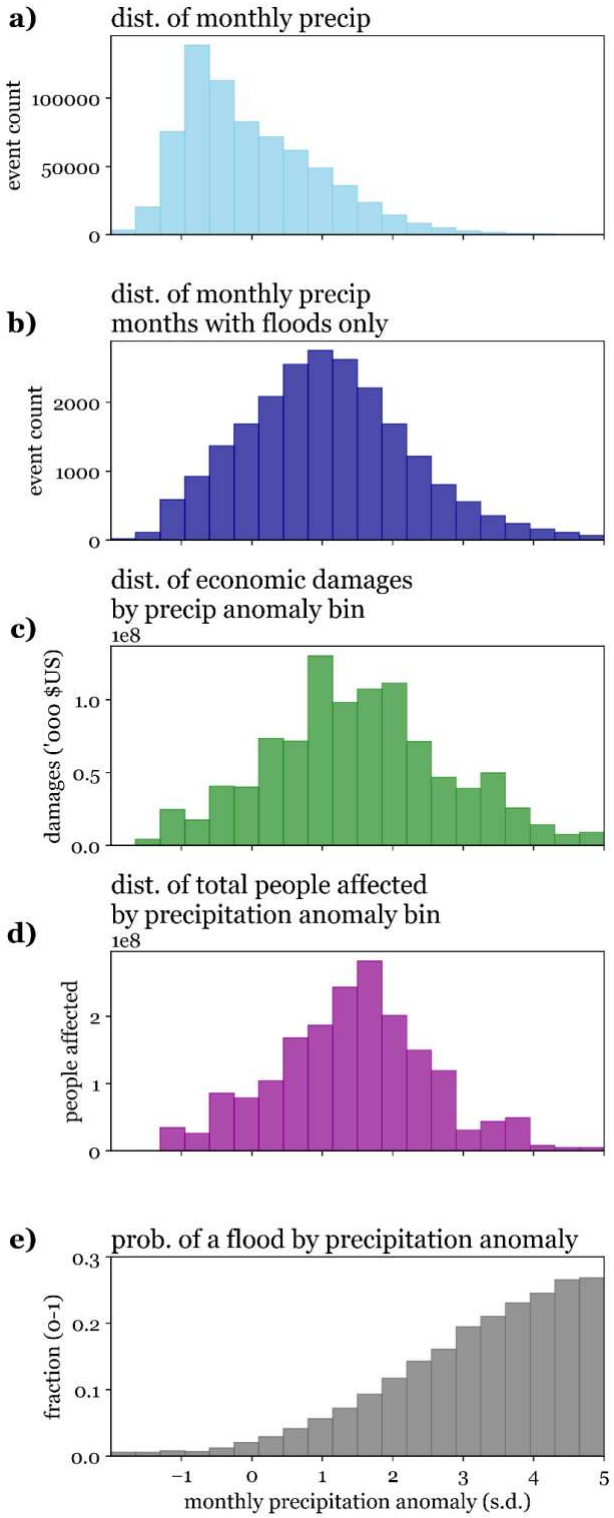

Fig. 10. Histograms showing the distribution of monthly precipitation anomalies across all months (a), monthly precipitation anomalies across months with recorded flood events only (b), total economic damages ('000 US$) by precipitation anomaly bin (c), total people affected by precipitation anomaly bin (d), and the probability of a flood by precipitation anomaly bin (e).

While Figure 10 provides an initial view of the relationship between precipitation extremes and flood damages, a more nuanced analysis must account for differences in climate, exposure, vulnerability, and reporting biases across regions and time. Using a panel regression with fixed effects to control for these factors, we present results from log-linear and log-quadratic models for both impact variables, as described below.

Across our global panel of admin1 regions from 2000 to 2024, higher than average monthly precipitation is significantly associated with increased normalized flood impacts (Table 1). Fig. 11 shows the regression lines of log-linear and log-quadratic models for each impact variable, each with a positive slope indicating higher impact at higher precipitation anomalies. Using a log-linear model, a one standard deviation increase in precipitation corresponds to a 10.9% increase in GDP-standardized damages. For our linear model of normalized total people affected, a one standard deviation increase in monthly precipitation corresponds to 20.4% more people affected. The values reported for quadratic percent increase varies with precipitation anomaly value. Going from a precipitation anomaly of zero to an anomaly of one standard deviation would increase expected damages by 9.5% and increase people affected by 17.9%. At lower precipitation anomalies, the effect of precipitation on flood impacts is smaller (where the quadratic bows out). However, at higher precipitation anomalies, the quadratic model suggests the effects of precipitation increases are much larger than predicted by the linear model.

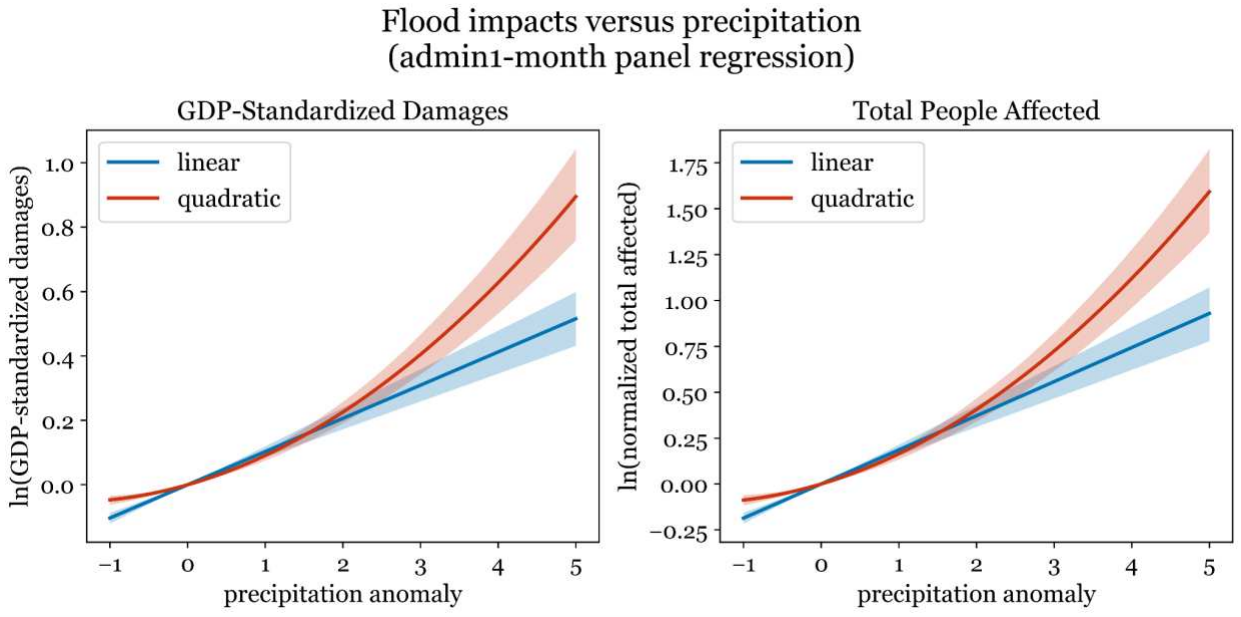

Fig. 11. Relationship between normalized flood damages (log-scale) and monthly precipitation at the admin1 level using linear and quadratic models. Shading indicates the 95% confidence interval estimated by bootstrapping countries.

Low within-model R^2 values for each model indicate that absolute variation in normalized flood impacts is mostly due to factors other than precipitation anomalies, even after controlling for regional and temporal differences. In other words, while the model captures a statistically significant causal effect of precipitation anomalies, most of the variation in flood impacts is driven by other unobserved factors, such as data quality and accuracy, local infrastructure conditions (including variations in infrastructure conditions across smaller scales than admin1 regions), and disaster response effectiveness. Nonetheless, precipitation remains a key, measurable contributor to flood impacts after accounting for seasonal, interannual, and regional fixed effects. Coefficient

values and model summary statistics are presented in Table 1.

	Linear		Quadratic	
	Damages (1)	People Affected (2)	Damages (3)	People Affected (4)
β	0.103***	0.186***		
	(0.011)	(0.021)		
$\beta 1$			0.069***	0.126***
			(0.009)	(0.018)
$\beta 2$			0.022***	0.039***
			(0.003)	(0.005)
Observations	712500	712500	712500	712500
S.E. type	by: country	by: country	by: country	by: country
R^2	0.252	0.242	0.254	0.244
R^2 Within	0.012	0.014	0.015	0.017

Significance levels: * p < 0.05, ** p < 0.01, *** p < 0.001. Format of coefficient cell: Coefficient (Std. Error)

Table 1. Panel regression details and descriptive statistics for all four models.

The panel dataset integrates many data sources across various processing steps, introducing uncertainty that weakens the detected relationship in regression models. Multiple sources contribute to this effect: missing and underreported data from EM-DAT, uncertainty introduced in our disaggregation and missing data infilling processes, missing satellite imagery due to cloud shadows, and errors from the water detection algorithm. Analysis published in Davenport et al. (2021) showed that missing data (i.e. unreported flood events) will result in a lower magnitude of the estimated model coefficients. The effect of missing data could explain some patterns in our regression model results. First, economic damage information is missing more often than information on total people affected. Accordingly, we find weaker effects of precipitation in the economic damages model compared to the people-affected model. Our economic damages model also has much lower coefficients than those reported for

the U.S. in Davenport et al. (2021). This difference could be partially explained by a much higher rate of missing data in our global dataset compared to the U.S. data. Our economic damages regression also shows lower R^2 than Davenport et al. (2021) reported, possibly because the U.S. generally has higher reporting accuracy and coverage, and because their data was pre-disaggregated to the state level while our global analysis required an additional disaggregation step which could introduce additional errors. Notably, random noise in the outcome variable will not bias the regression coefficients (i.e. will not lead to systematic under- or over-estimation) but will increase uncertainty in the model. Overall, it is likely that the combined effects of data errors reduce the panel coefficient magnitudes, generally underestimating the relationship between precipitation anomalies and impact variables.

4. Conclusions

While floods have affected human societies throughout history, challenges in the modern era require more thoughtful, science-backed, and equity-minded solutions to better manage the changing reality of flood risk. The compounding effects of intensifying patterns in extreme weather due to anthropogenic climate change, ongoing human and infrastructure development in flood-prone regions, and socioeconomic inequalities within and between countries are exacerbating existing vulnerabilities, highlighting the need for a better understanding of floods and their impacts to inform flood planning and risk management for the future. Our study integrates 25 years of disaster data from the EM-DAT disaster database with satellite-derived flood maps and gridded population estimates to characterize flood extent and impacts at a consistent

spatial and temporal resolution across the globe. Using this dataset, we analyze regional patterns in flood incidence and impacts from 2000-2024. To address ongoing questions in climate impact science about how extreme precipitation influences flood impacts, we also use this dataset to apply a fixed-effects panel regression model that controls for regional and temporal differences throughout the study period.

Our analysis reveals that absolute and normalized metrics tell different stories about flood vulnerability: high-income regions dominate in absolute economic damages and flooded area, while GDP-normalized damages expose disproportionate economic burdens in developing regions across South America, Africa, and Southeast Asia. Similarly, patterns of population impact show that regions with high flooded areas do not necessarily translate to high numbers of people affected, while smaller-scale floods can be highly destructive in densely populated regions or regions with limited resources to plan for or respond to floods.

Our panel regression model finds a statistically significant relationship between precipitation anomalies and flood impacts, albeit weakened by the uncertainties introduced by missing data and the required preprocessing steps to derive our panel dataset. Despite these limitations, our model demonstrates that flood impacts are intensified at higher precipitation anomalies, and provides a tool to use in climate impact attribution and prediction research..

This work presents a scalable and adaptable foundation for further analysis. Future directions of this work could further address data limitations and expand upon the panel regression modeling. Combining our analysis of EM-DAT with other flood hazard datasets—such as the Dartmouth Flood Observatory database or regional-scale disaster records — could reduce the incidence of missing and incomplete hazard data.

Additionally, the MODIS flooded pixel algorithm could be used to detect incidence of flooding that is unreported in these datasets, overall providing a more complete flood record. Our satellite-derived flooded population estimates showed a statistically significant correlation with the reported total population affected values. In the future, the satellite-derived flooded population could be used to predict total population affected for events missing this information.

For the panel regression model, one next step is to incorporate additional climate variables into the panel model, such as 75th or 90th percentile 5-day precipitation or antecedent soil moisture conditions, to evaluate the impact of other environmental variables on flood impacts. This analysis would provide a more complex understanding of how changing hydrometeorological conditions affect flood impacts. Finally, precipitation estimates from climate model projections could be used with our panel model to predict how flood impacts may change under different global warming levels and in different regions.

References

Arnell, N. W., & Gosling, S. N. (2016). The impacts of climate change on river flood risk at the global scale. *Climatic Change*, *134*(3), 387–401. https://doi.org/10.1007/s10584-014-1084-5

Baston, D. (2025). exactextract. Retrieved from https://github.com/isciences/exactextract/releases/tag/v0.2.2

Beck, H. E., Wood, E. F., Pan, M., Fisher, C. K., Miralles, D. G., Dijk, A. I. J. M. van, et al. (2019). MSWEP V2 Global 3-Hourly 0.1° Precipitation: Methodology and Quantitative Assessment. *Bulletin of the American Meteorological Society*, *100*(3), 473–500. https://doi.org/10.1175/BAMS-D-17-0138.1

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

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

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

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

Diffenbaugh, N. S., Singh, D., Mankin, J. S., Horton, D. E., Swain, D. L., Touma, D., et al. (2017). Quantifying the influence of global warming on unprecedented extreme climate events. *Proceedings of the National Academy of Sciences*, *114*(19), 4881–4886. https://doi.org/10.1073/pnas.1618082114

Dong, S., Sun, Y., Li, C., Zhang, X., Min, S.-K., & Kim, Y.-H. (2021). Attribution of Extreme Precipitation with Updated Observations and CMIP6 Simulations. *Journal of Climate*, *34*(3), 871–881. https://doi.org/10.1175/JCLI-D-19-1017.1

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

Dottori, F., Szewczyk, W., Ciscar, J.-C., Zhao, F., Alfieri, L., Hirabayashi, Y., et al. (2018). Increased human and economic losses from river flooding with anthropogenic warming. *Nature Climate Change*, *8*(9), 781–786. https://doi.org/10.1038/s41558-018-0257-z

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

Guo, Y., Gasparrini, A., Armstrong, B. G., Tawatsupa, B., Tobias, A., Lavigne, E., et al. (2017). Heat Wave and Mortality: A Multicountry, Multicommunity Study. *Environmental Health Perspectives*, *125*(8), 087006. https://doi.org/10.1289/EHP1026

Hsiang, S., & Jina, A. (2014). *The Causal Effect of Environmental Catastrophe on Long-Run Economic Growth: Evidence From 6,700 Cyclones* (No. w20352) (p. w20352). Cambridge, MA: National Bureau of Economic Research. https://doi.org/10.3386/w20352

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

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

Kummu, M., Kosonen, M., & Masoumzadeh Sayyar, S. (2025). Downscaled gridded global dataset for gross domestic product (GDP) per capita PPP over 1990–2022. *Scientific Data*, *12*(1), 178. https://doi.org/10.1038/s41597-025-04487-x

Lynch, V. D., Sullivan, J. A., Flores, A. B., Xie, X., Aggarwal, S., Nethery, R. C., et al. (2025). Large floods drive changes in cause-specific mortality in the United States. *Nature Medicine*, *31*(2), 663–671. https://doi.org/10.1038/s41591-024-

03358-z

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

Min, S.-K., Zhang, X., Zwiers, F. W., & Hegerl, G. C. (2011). Human contribution to more-intense precipitation extremes. *Nature*, *470*(7334), 378–381. https://doi.org/10.1038/nature09763

Organisation for Economic Co-operation and Development. (n.d.). Inflation (CPI) [Data set]. OECD. Retrieved from https://www.oecd.org/en/data/indicators/inflation-cpi.html

Schlenker, W., & Roberts, M. J. (2009). Nonlinear temperature effects indicate severe damages to U.S. crop yields under climate change. *Proceedings of the National Academy of Sciences*, *106*(37), 15594–15598. https://doi.org/10.1073/pnas.0906865106

Sharma, A., Wasko, C., & Lettenmaier, D. P. (2018). If Precipitation Extremes Are Increasing, Why Aren't Floods? *Water Resources Research*, *54*(11), 8545–8551. https://doi.org/10.1029/2018WR023749

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

The PyFixest Authors. (2025). pyfixest: Fast high-dimensional fixed effect estimation in Python. Retrieved from https://github.com/py-econometrics/pyfixest

Trenberth, K. (2011). Changes in precipitation with climate change. *Climate Research*, *47*(1), 123–138. https://doi.org/10.3354/cr00953

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

Wing, O. E. J., Lehman, W., Bates, P. D., Sampson, C. C., Quinn, N., Smith, A. M., et al. (2022). Inequitable patterns of US flood risk in the Anthropocene. *Nature Climate Change*, *12*(2), 156–162. https://doi.org/10.1038/s41558-021-01265-6

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

Xu, J., Fan, C., Su, X., & Han, H. (2025). An inference approach for assessing place-based vulnerability to heat mortality. *Humanities and Social Sciences Communications*, *12*(1), 1324. https://doi.org/10.1057/s41599-025-05368-9

Young, R., & Hsiang, S. (2024). Mortality caused by tropical cyclones in the United States. *Nature*, *635*(8037), 121–128. https://doi.org/10.1038/s41586-024-07945-5

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