

INDOFLOODS: A Comprehensive Database for Flood Events in India Enhanced with Catchment Attributes

Sai Kiran Kuntla^a and Manabendra Saharia^{a,b}

KEYWORDS:

Hydrologic cycle;
Flood events;
Streamflow;
Hydrology;
Hydrometeorology;
Databases

ABSTRACT: Despite floods causing significant loss of life and property, datasets to characterize flooding events in developing countries such as India are not widely available, hampered by limited hydrometric records. Current flood databases are limited to continental river basins, case studies, and government reports, which do not represent the diversity of flooding in terms of climate, basin morphometry, and triggering mechanisms. This first-of-its-kind flood event database called Indian Observational Flood Events Database (INDOFLOODS) is developed using a unique approach of combining long-term station discharge observations with official flooding thresholds for warning and danger water level. Flooding information include start and end time, peak flood level and discharge and its date of occurrence, flood volume, event duration, time to peak, and recession time. Along with metadata such as upstream catchment area, coordinates, shapefiles, and river and tributary names, the database is augmented with a large number of geomorphological, climatological, event-scale precipitation, land-cover, soil, lithology, and anthropogenic characteristics derived at the catchment scale. Preliminary data analysis based on envelope curves shows that the magnitude of extreme floods in India is higher than that reported in the United States. While every dataset has limitations, this collation of flooding events with a plethora of causative hydrogeomorphic factors in a standardized format will be a major asset for the community and serve as an example for how inconsistent data records in developing countries can be turned into useful flood event databases for data-driven studies. This large sample database is expected to cater to a wide range of applications advancing flood research and management, such as trend analysis, hazard and severity assessment, calibration and validation of hydrological and hydraulic models, and developing new metrics for impact assessment.

SIGNIFICANCE STATEMENT: Indian Observational Flood Events Database (INDOFLOODS) is the first comprehensive flood event database for India and represents a significant advancement in flood research and management. By integrating long-term flood event data with extensive catchment characteristics, it provides a standardized dataset crucial for broad applications in hydrometeorology. The preliminary observations reveal the severity of Indian floods compared to that in the United States, highlighting the need for targeted flood mitigation strategies. INDOFLOODS not only enhances our understanding of flood dynamics in diverse climatic and geographical settings but also exemplifies how systematic data collation can transform fragmented records into a valuable resource for scientific and practical applications, supporting better flood preparedness and response strategies.

DOI: 10.1175/BAMS-D-24-0008.1

Corresponding author: Manabendra Saharia, msaharia@iitd.ac.in

Supplemental information related to this paper is available at the Journals Online website: <https://doi.org/10.1175/BAMS-D-24-0008.s1>.

Manuscript received 7 January 2024, in final form 10 January 2025, accepted 17 January 2025

© 2025 American Meteorological Society. This published article is licensed under the terms of the default AMS reuse license. For information regarding reuse of this content and general copyright information, consult the AMS Copyright Policy (www.ametsoc.org/PUBSReuseLicenses).

1. Introduction

The frequency of riverine floods and their resulting destruction to the economy continues to rise across the world (Kuntla 2021; Newman and Noy 2023). For instance, in India alone, 113 943 human casualties were reported, and 53 billion U.S. dollars of damage was caused to crops, houses, and public utilities by floods between 1953 and 2017 [Central Water Commission (CWC) 2022]. Besides, the flood risk is expected to increase in the country in the near future due to a warming climate, environmental degradation, and urban sprawl. Hence, much more attention is required to mitigate its adverse effects by advancing our knowledge of flood characteristics and drivers. However, a major limitation hindering comprehensive flood modeling and characterization studies in India is the unavailability of a holistic database describing the physical characteristics of past flood events, such as flood level, discharge, and duration, as well as factors influencing it.

Gourley et al. (2013) developed a unified event-based flood database for flash floods across the United States, which has found wide usage in the hydrometeorology community. It consists of 98 668 events that exceeded action stages that correspond to bank full conditions as defined by the NWS (2019) at 2948 gauge locations. Nielsen et al. (2015) used this database to analyze the meteorological and climatological characteristics associated with TORFF events (tornadoes and flash floods together) over the United States. Smith and Smith (2015) used the dataset to identify the flashiest watersheds in the United States based on the frequency of floods with discharge peaks of more than $1 \text{ m}^3 \text{ s}^{-1} \text{ km}^{-2}$. Saharia et al. (2017a) used it to develop flashiness as an index and identify the flash flood-prone areas, and Kumar et al. (2024) studied the flash flood recovery in the United States. Many such efforts exist to develop flood event databases useful in computational hydrology in the United States (Huang et al. 2021; Shen et al. 2017) and Europe (Gaume et al. 2009). However, such large-sample observed flood event databases are rare in the public domain from underdeveloped or developing countries like India. At the same time, global databases and inventories like the Emergency Disasters Database (EM-DAT), the Dartmouth Flood Observatory (DFO), and the International Flood Network (IFNET) are primarily based on surveys or reports and focus heavily on the impacts of events. As a result, they tend to document only major floods that have significant impacts. Moreover, they often lack consistent information on key physical characteristics of floods, such as flood levels, time to peak, discharge, and duration. Consequently, this limitation restricts their utility in computational hydrology and prevents them from capturing the full spectrum of flood events, including variations in geographical location, flood type, triggering mechanisms, and catchment hydrogeomorphology. Similarly, Saharia et al. (2021a) compiled a list of past flood events in the country from printed publications of the Indian Meteorological Department (IMD). However, this dataset does not incorporate hydrogeomorphic characteristics and contains no information regarding the associated physical flood characteristics, making it less useful for computational hydrology and subject to many limitations as global databases.

Large-sample studies have emerged as the fourth paradigm of hydrology, where data-driven methods are being used to discover new and robust conclusions on processes and relationships hindered by conventional catchment or region-specific studies (Addor et al. 2020; Peters-Lidard et al. 2017). However, due to the lack of large-sample event-based flood observational data, most hydrological studies across the globe, including India, use alternative approaches to identify and characterize floods or restrict to catchment-specific studies. The most common practice in flood characterization studies is to take annual peak discharge as the indicator of floods (Blöschl et al. 2017; Do et al. 2020; Jain et al. 2017; Jena et al. 2014). However, based on this approach, all the values accounted as floods may not be flooding, and it may miss actual flood events. For example, in flood-rich years, only one of the largest observed values in the time series values along the year is considered a flood, leaving the remaining actual floods. Also, a small river flow will be selected as a flood in flood-poor years, but it is not. Moreover, at all the gauge stations, floods may not occur. However, the abovementioned method does not consider this point and returns flood data (their annual peak discharge) at all the locations. The other most common approaches are floods identified based on indices, peak over thresholds, streamflow return periods, and simulation results, which are limited in their own ways. This study develops the flood database based on the operational definition of floods, which makes the conclusions directly applicable to field hydrologists.

Hence, considering operational applications in mind, a methodology based on field threshold definitions of floods has been employed in this study to develop a comprehensive event-based flood database called the Indian Observational Flood Events Database (INDOFLOODS), which catalogs the physical characteristics of observed flood events such as flood duration, volume, and peak water level. Additionally, the database provides general metadata, event-scale precipitation, multiple catchment characteristics, and shapefiles of the upstream catchment boundaries in an analysis-ready format for broader usage.

2. Data sources

The event-based flood database is developed using data from three primary sources. Extensive preprocessing was undertaken to standardize these datasets before compilation. The metadata, including general information about gauge stations such as geographical coordinates, station name, river and basin names, and upstream catchment area, was obtained from CWC published reports. The CWC is responsible for collecting and compiling streamflow data on all major rivers in the country. Observational discharge data and their corresponding level over the gauge stations were sourced from the India Water Resources Information System (India-WRIS; <https://indiawris.gov.in/>). India-WRIS provides data that mostly belong to the continental basins in the peninsular region free of charge to the public, with restrictions on large downloads, and requires extensive curation and cleaning before use. The data from transboundary basins, such as Ganga, Brahmaputra, and Indus, are provided only upon request. Last, official flood thresholds, including warning and danger levels as defined by the CWC for each gauge station, are extracted from the flood forecasting portal of the CWC. The thresholds are fixed by the CWC in consultation with the state governments and local authorities based on their experience, historical data, and ground surveys (NDMA, Government of India 2008).

After meticulously fetching and processing these datasets, we identified 214 gauge stations where all the required data were available and where floods were observed based on flooding thresholds. The locations of these gauge stations are plotted in Fig. 1, and the catchments of these gauge stations cover all the primary climatic classes, landscapes, and topography in the country. Moreover, it is the only database in the country with such a critical representativeness based on observed streamflow datasets.

3. Data structure and processing

The database has four components:

a. Metadata. Metadata contains the general information of all the gauge stations of the database. It includes a unique gauge ID assigned, the official CWC-provided corresponding flooding thresholds, the geographical coordinates, and basin and river names on which the gauge station lies, the start and end date of the corresponding gauge station streamflow records, upstream catchment area, etc. The complete list of all the fields in the metadata file and their detailed description is provided in Table S1 in the online supplemental material.

b. Flood events. The inset in Fig. 1 illustrates how the flood events are defined, and their variables are derived from the time series stream water level data utilizing flooding thresholds. Whenever the water level (stage) of the stream reaches or exceeds the “warning level,” we consider it as the start date of a flood event. When the water drops below the same threshold, that day is regarded as the end date of the respective event. Warning level is usually above the normal water level of the stream course. It provides an early warning for preparedness. The peak water level between the start and end date is the “peak flood level” for that event. If the peak flood level is above the “danger level,” then the corresponding flood event is classified as a “severe flood.” Otherwise, “flood.” On the other hand, the peak discharge between the exact dates is the “peak discharge” of the flood event. Danger level depicts the stage of the river, which, if crossed by flood water, will start causing damage to the nearby

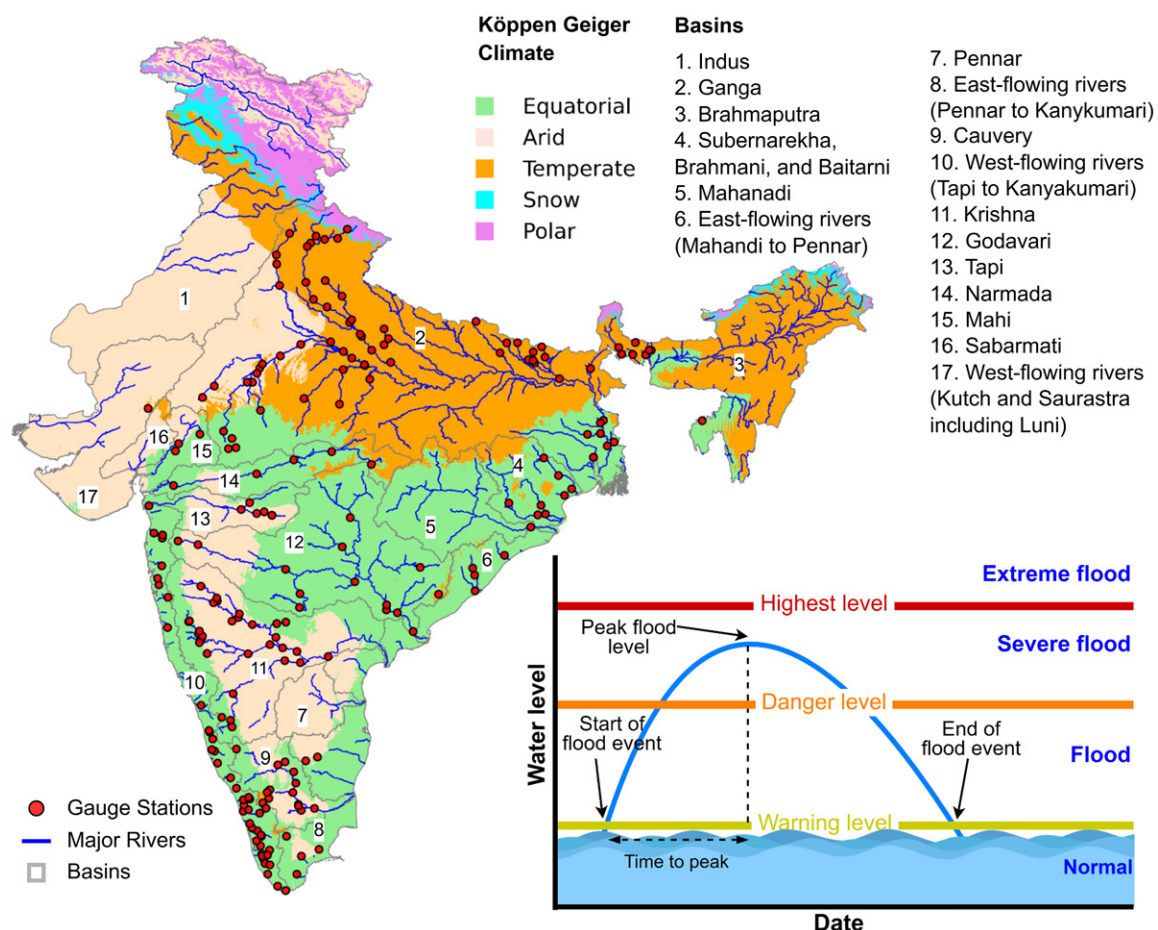


FIG. 1. The geographical location of the gauge stations where floods are observed and whose data are available in the INDOFLOODS database. The base layer of the plot showcases the Köppen–Geiger climate (Kottek et al. 2006), major river networks, and basins in the country. The inset in the lower-right corner illustrates how floods are extracted from the time series of water level data.

areas, human settlements, and infrastructure. Evacuation and relief measures are initiated at this level (CWC, Ministry of Jal Shakti, Government of India 2020; GFCC, Government of India. 2004). The number of days between and including the start and end date of the event is considered “event duration.” While the time taken to reach the peak from the start date is the “time to peak,” and the time taken to drop below the same threshold from the peak is the “recession time.” At the same time, the total discharge during the event is provided as “flood volume” in the database. All the variables embedded with each flood event in the database are listed in Table 1. Following the above procedure, the database contains 8342 flood events, including 5525 floods and 2817 severe floods, observed over 214 gauge stations over 62 years between 1959 and 2020 across the country.

A yellow bulletin is issued by the CWC when flood levels exceed the warning level but remain below the danger level. At the same time, a red/orange bulletin is disseminated during severe floods, i.e., when water crosses the danger level (CWC, Ministry of Jal Shakti, Government of India 2020). All three bulletins are released to the concerned authorities for necessary action and to the public on social media and the official dedicated websites of CWC and the National Disaster Management Authority (NDMA).

c. Catchment characteristics. The database aims to offer data on catchment variables upstream of the gauge station to extend the utility of the flood event data for a holistic, comprehensive characterization and various other applications. Catchment characteristics include geomorphology (28 variables), climatology (19 variables), event-scale precipitation (10 variables), anthropogenic (7 variables), soil, lithology, land cover, and Köppen–Geiger climate variables. Table S2 tabulates all the catchment characteristics provided in the database.

d. Catchment shapefiles. To facilitate further analysis of the flood events, the shapefiles of the upstream catchments are also provided. This will allow users to extend the influencing

TABLE 1. Definition of flood event variables contributing to the database.

Variable	Description
EventID	It is the unique ID corresponding to the flood event in an extensible format such as INDOFLOODS-gauge-118-10, where “INDOFLOODS-gauge-118” is the gauge ID (from metadata) of the gauge station where the respective flood event has been observed, and “10” is for the serial number of the event at that station. This schema makes it flexible enough to incorporate future events and connect them with metadata and other datasets.
Start date	Start date of the flood event—when the flow exceeds the flooding threshold (warning level). All dates conform to ISO 8601 (YYYY-MM-DD), the international standard for representing dates and times.
End date	End date of the flood event—when the flow dropped below the flooding thresholds (warning level).
Peak flood level	Peak flood level corresponding to that event.
Peak flood level date	The corresponding date of occurrence of the peak flood level.
Number of peak flood level	The number of times (days) the flow has touched the peak flood level in the same event.
Peak discharge	The magnitude of the peak discharge corresponding to that event.
Peak discharge date	The corresponding date of occurrence of the “peak discharge.”
Flood volume	The total discharge of the flood event. It equals the sum of daily discharge over all the days of a flood event.
Event duration	The total number of days the flood lasted.
Time to peak	The total number of days it took to reach the peak of flood level from start time.
Recession time	The total number of days it took from the peak of flood level to end time.
Flood type	Specifies whether the event is a flood or severe flood. If the peak flood level of the corresponding event is above the danger level, it is a severe flood. Otherwise, flood.

characteristics in the database with an ever-increasing variety of geospatial datasets. Each shapefile is designated with its corresponding unique ID as per the metadata (GaugeID). A consistent methodology has been employed to delineate catchment boundaries for every gauge station. An algorithm has been developed to automate the process of shapefile generation for catchment boundaries, taking the geographical coordinates of gauge stations as pour points/outlets for the respective catchments. This task is complicated since the precision of the supplied geographic coordinates of the gauge stations is inaccurate but crucial. For instance, if the coordinates of a particular station do not fall on a river network, the generated catchment based on the Flow Direction Map (FDR) would be incorrect. Hence, we have utilized the source catchment area (from CWC documents; provided in the metadata of INDOFLOODS) and raster upstream catchment area of HydroSHEDS (Lehner et al. 2008) as a proxy to search for the compatible coordinates to extract the most likely catchment boundary shapefile. The steps of the shapefile generation algorithm are as follows:

- Step 1: The algorithm uses the source catchment area for every gauge station and performs a nearest neighbor search based on the available geographic coordinates in the metadata over the gridded catchment area and FDR maps of HydroSHEDS and returns the geographic coordinates of the pixel whose catchment area value is close to the source value within 1.5-km distance. This ensures that the geographic coordinates fall on a river network itself.
- Step 2: Sixteen gauge stations (7%) out of the total in the INDOFLOODS do not have source catchment area information. For such stations, the algorithm searches for the maximum catchment area in the 1.5-km vicinity range of given geographical coordinates and returns its geographic coordinates.
- Step 3: If the geographical coordinates are relocated in either of the cases, the new coordinates are used. Otherwise, the same old coordinates are used as the outlet/pour point of the catchment to derive the shapefile of the corresponding catchment based on the FDR map.

Since the quality of shapefiles is critical for determining the catchment-scale variables, the quality of the delineated catchment boundary is also made available to the users in the metadata, as tabulated in Table S1. The catchment area of the generated shapefiles and source catchment area from CWC are compared to find the area variability between them. If the area variability is below 25%, we classify it as “safe.” Otherwise, “caution.” Since quality assessment can only be done using the catchment area, we have classified the quality of shapefiles of 16 catchments whose source catchment area information is unavailable as caution. Per these criteria, 87% ($n = 186$) of the shapefiles provided in INDOFLOODS are safe. The quality indicator should be considered based on the intended user application before using these shapefiles and any accompanying information. These generated shapefiles themselves were used to derive catchment-scale variables discussed in section 3c.

4. Applications

Being able to connect upstream catchment characteristics to flood events can lead to wide research applications. This section provides a few examples of the applications of the database exploring first-order relationships between flood events and a few catchment characteristics.

a. Analyzing the dependence of flood event characteristics on catchment characteristics.

Figure 2 assesses how the mean flood peak discharge and duration are correlated with the geomorphology of the catchments. Figure 2a shows a gradual increase in flood peak discharge as the basin magnitude and drainage texture increase. Higher basin magnitude,

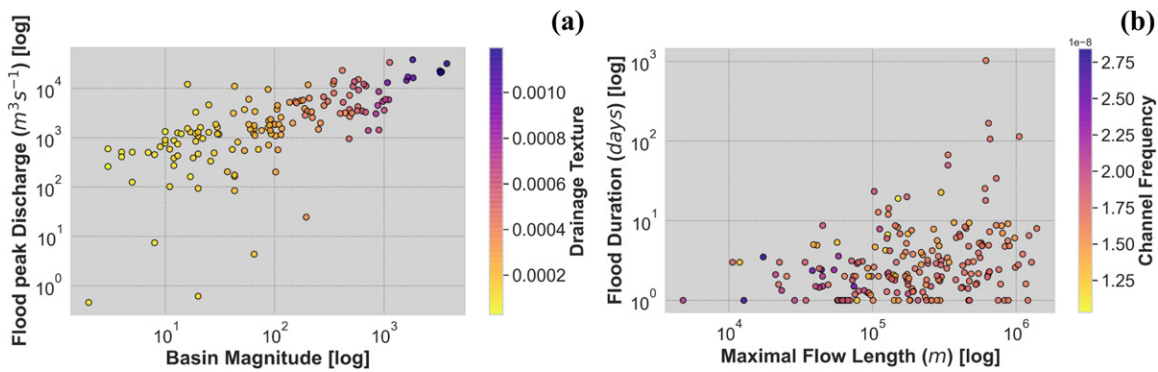


FIG. 2. Dependence of flood event characteristics: (a) peak discharge and (b) duration on geomorphology of the catchments.

which denotes the number of first-order streams in the catchment (Melton 1957), makes catchments more favorable to produce higher flood peak discharges as first-order streams converge all the overland flow water to the downstream, amplifies the flow in main channel downstream, and increases the peak discharge at the outlet. On the other hand, drainage texture, which is the ratio of the total number of stream segments of all orders in the catchment to the perimeter of the catchment, provides valuable information about the complexity and connectivity of the drainage network within a catchment, which influences its hydrological response to precipitation events, including peak discharge behavior. Higher values of drainage texture indicate a more well-developed drainage network, allowing water to be routed efficiently through the network, leading to faster and higher runoff responses at the outlet.

Though dynamic variables like the amount of rainfall that triggered the flood event, the number of pouring days, and the subsequent flood volume are some of the primary factors that influence flood duration at a location, the geomorphology of the catchment could also affect the inundation time as shown in Fig. 2b. It illustrates the dependence of flood duration on maximal flow length, which corresponds to the length along the most extended stream from the head of the channel to the outlet, and the channel frequency, which determines the density of channels in the catchment. Lower channel frequency suggests a less interconnected drainage network and is generally associated with flat terrain, where streams are more widely spaced and water accumulates longer. Hence, catchments with lower channel frequency may lead to longer lag times between rainfall input and peak discharge at the outlet, contributing to extended flood durations. On the other hand, the catchments with higher channel frequency have a greater number of streams per unit area, leading to more efficient runoff conveyance in a shorter time, resulting in shorter flood durations. Similarly, shorter travel distances within the catchment with lower maximal flow length may result in shorter flood durations and vice versa.

b. Establishing the envelope curves for the extreme floods against the basin area. Envelope curves are popular in hydrology studies to provide an upper bound on the magnitude of the extreme floods that can be expected based on the catchment area (Abdullah et al. 2019; Castellarin 2007; Kuntla et al. 2022; Saharia et al. 2017b). Figure 3 represents the envelope curves of the extreme floods reported in the database. These curves serve as a fundamental tool for hydrologists and engineers to identify patterns in flood behavior across different catchment sizes, which are crucial for estimating flood magnitude, designing infrastructure to withstand peak flows, and developing flood control measures and urban drainage systems. Envelope curves are based on a power-law equation [Eq. (1)] and plotted on a log–log diagram:

$$Q = \alpha A^\beta. \quad (1)$$

Here, Q ($\text{m}^3 \text{s}^{-1} \text{km}^{-2}$) is the peak specific discharge, A (km^2) is the catchment area, α [$\text{m}^3 \text{s}^{-1} \text{km}^{-2(1+\beta)}$] is the reduced discharge, and β is the scaling coefficient. The reduced discharge can serve as an indication of flood magnitude by reducing the dependence of the catchment area on the analysis. Calculating the β value involves fitting a regression line between $\log(Q)$ and $\log(A)$ (Castellarin 2007). This β value represents the rate of change in peak specific discharge concerning alterations in the catchment area. A β value closer to zero signifies a smaller change in peak specific discharge with variations in the catchment area.

In Fig. 3, the envelope curves of India are accompanied by envelope curves of extreme floods reported globally (Kuntla et al. 2022) and in the United States (Saharia et al. 2017b) for comparison. The reduced discharge α for India is observed to be 875.9, and the scaling coefficient β is -0.43 . These values help estimate the peak specific discharge at ungauged stations in the country based on their catchment area. At the same time, comparing these values with those observed globally ($\alpha = 4130.6$, $\beta = -0.58$) (Kuntla et al. 2022) and in the United States ($\alpha = 108$, $\beta = -0.47$) (Saharia et al. 2017b) reveals that the flood magnitudes in India are much higher than those in the United States but not the highest globally. In addition, the rate of change in flood magnitudes with unit catchment areas in India is lower than that of the United States and globally.

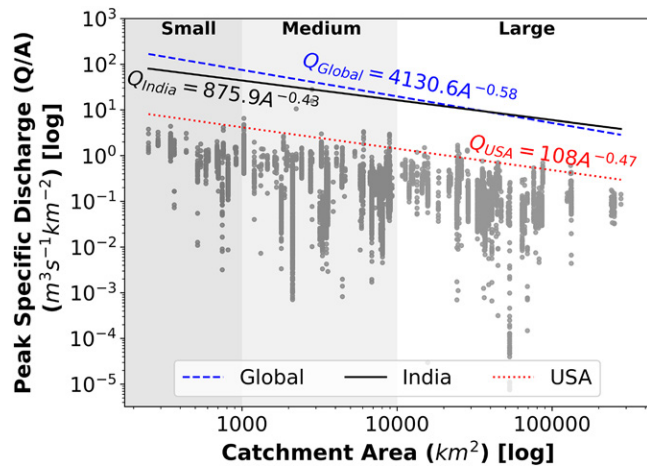


FIG. 3. Peak specific discharge as a function of the catchment area for all the 8342 flood events, along with the envelope curves for extreme floods. The envelope curves for observed extreme floods over globally and continental United States are taken from Kuntla et al. (2022) and Saharia et al. (2017b), respectively.

c. Preliminary observations of antecedent precipitation patterns that contribute to flood events.

Precipitation plays a fundamental role in triggering floods, as its intensity, duration, and timing significantly affect runoff and water accumulation in catchments. For instance, a thorough investigation of event-scale precipitation variables provided, including cumulative daily precipitation 1–10 days before the flood event, gives valuable insights into flooding drivers and flood generation mechanisms. Figure 4 summarizes the daily precipitation before all flood events in the database. Preliminary observations of event-scale precipitation variables highlight that multiday precipitation is responsible for most of the flooding in the country, as the gradual increase in cumulative rainfall is observed before the majority of the flood events. Nevertheless, it is also observed that 1 day of precipitation, just a day before the flood start date, without any

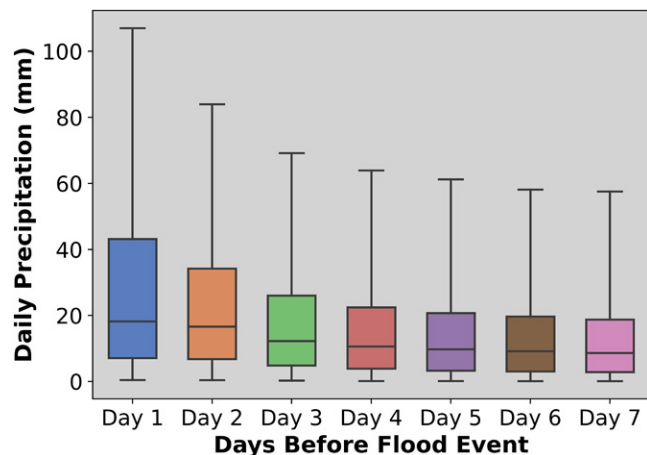


FIG. 4. The box plots showing the distribution of daily precipitation before flood events. The box spans the interquartile range (i.e., 25th and 75th percentiles), with the bar in the middle representing the median. The whiskers are the two vertical lines outside the box extended until the 5th and 95th percentiles.

precipitation before that (>1 mm in the catchment), has led to a few floods in the country. In addition, in 33.5% of flood events, the maximum daily precipitation is observed just a day before the flood event compared to the past 10 days before the respective flood events. These observed patterns suggest a complex interplay between short-term intense rainfall events occurring just before the flood event and the cumulative effect of rainfall over several days leading up to the event. Besides, a recent study confirms that the spatial organization of rainfall influences the basin response on par with geomorphology and climatology (Saharia et al. 2021b). However, more comprehensive studies are required to arrive at detailed conclusions.

d. Preliminary observations on the dependence of flood events on catchment anthropogenic characteristics. Flood characteristics such as peak flood level, discharge, and event duration exhibit notable relationships with catchment anthropogenic features, including urban area percentage, Human Development Index (HDI), and population density. As depicted in Fig. 5, Spearman correlations reveal that densely populated urban areas tend to experience prolonged flood durations. This trend suggests that increased urbanization, often accompanied by inadequate drainage systems and slower water recession, contributes to extended flood events. Conversely, a weak negative correlation (-0.11) is observed between HDI and flood duration, potentially indicating that regions with higher HDI benefit from better infrastructure and flood management practices, thereby mitigating flood impacts. HDI, widely used by international organizations to assess the socioeconomic development status of an area (Kummu et al. 2018), provides a useful lens for understanding how development levels influence flood resilience.

Interestingly, a negative correlation observed between urban area percentage and population density with peak flood level and discharge highlights the possible role of water management practices in urbanized catchments. For instance, this could be attributed to the higher water demand in urban environments and densely populated catchments, where water storage reservoirs are often managed to accommodate seasonal flooding during the monsoon. Furthermore, urbanization often alters the natural drainage patterns and reduces flood peaks (Kuntla et al. 2023). However, the weak nature of these correlations in the analysis, ranging between -0.29 and 0.17 , suggests that anthropogenic factors alone cannot fully explain the observed variations in flood characteristics.

These observations emphasize the complexity of flood generation mechanisms, necessitating a multidimensional approach that considers both anthropogenic and natural catchment characteristics. By integrating factors such as geomorphology, land use, precipitation, soil properties, and hydroclimatology, researchers and policymakers can better understand the interactions driving flood events. This knowledge is crucial for devising targeted flood management strategies, particularly in rapidly urbanizing regions, to reduce disaster risks and foster resilience.

5. Summary

This study describes the newly developed flood event database for India enhanced with a multitude of catchment hydrogeomorphic characteristics. This first-of-its-kind database involved extensive processing of large samples of stream-flow records, flooding thresholds, and numerous catchment characteristics derived from satellites, models,

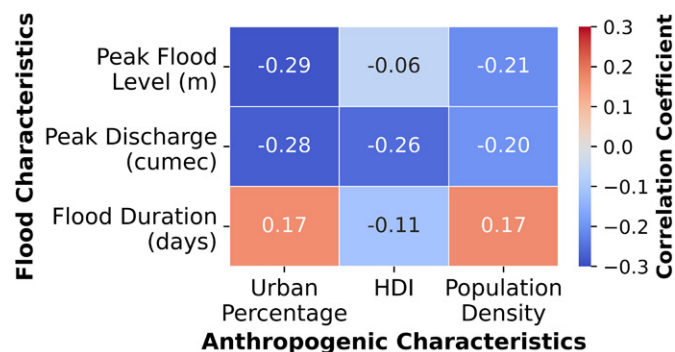


FIG. 5. Heatmap showing Spearman correlations between peak flood level, discharge, and duration with selected anthropogenic characteristics.

and ground sources. The database is designed as an analysis-ready resource to support a broad spectrum of flood research and management applications, including flood hazard and risk assessment and calibration and validation of hydrological and hydraulic models.

This article also provides key insights into flood behavior, demonstrating the first-order dependence of flood peak discharge and flood duration on a few geomorphological characteristics. For example, basin magnitude and drainage texture are observed to be proportional to peak discharge, while an increase in maximal flow length and channel frequency may increase and decrease the flood duration, respectively. Additionally, envelope curves established for extreme floods revealed that the magnitude of floods observed in India surpasses that reported in the United States but remains lower than that observed globally. The rate of change in peak specific discharge per change in a unit catchment area in India is found to be lower than that reported in the United States and globally.

The inclusion of socioeconomic and demographic data provides the capability to evaluate the differential vulnerabilities and resilience in flood-affected communities. Furthermore, the availability of catchment characteristics and shapefiles of catchment boundaries ensures the database's flexibility, allowing users to expand its scope for addressing diverse research needs and conducting in-depth analyses of flood events. Overall, INDOFLOODS is expected to significantly advance flood research and management in India and globally. We expect this database to be an ongoing community effort that will aid future data collection efforts, initiate thinking toward unique techniques for valuable data development, and promote open science.

Acknowledgments. This research was conducted in the HydroSense lab (<https://hydrosense.iitd.ac.in/>) of IIT Delhi, and the authors acknowledge the IIT Delhi High Performance Computing facility for providing computational and storage resources. Dr. Manabendra Saharia gratefully acknowledges financial support for this work through grants from the Ministry of Earth Sciences/IITM Pune Monsoon Mission III (RP04574), the Ministry of Earth Sciences DeepINDRA project (RP04741), and DST IC-IMPACTS (RP04558). The authors gratefully acknowledge the Central Water Commission (CWC), the National Water Informatics Centre (NWIC), and the Ministry of Jal Shakti (MoJS) for providing the streamflow datasets used in this study. The authors thank Dr. Jonathan J. Gourley (NSSL/NOAA) for commenting on an early version of the manuscript. The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper. Sai Kiran Kuntla: conceptualization, methodology, formal analysis, data wrangling and curation, and writing—original draft. Manabendra Saharia: conceptualization, funding acquisition, and writing—review and editing.

Data availability statement. INDOFLOODS has been publicly released and is freely accessible at <https://doi.org/10.5281/zenodo.14584654>.

References

- Abdullah, J., N. S. Muhammad, S. A. Muhammad, and P. Y. Julien, 2019: Envelope curves for the specific discharge of extreme floods in Malaysia. *J. Hydro-environment Res.*, **25**, 1–11, <https://doi.org/10.1016/j.jher.2019.05.002>.
- Addor, N., H. X. Do, C. Alvarez-Garretón, G. Coxon, K. Fowler, and P. A. Mendoza, 2020: Large-sample hydrology: Recent progress, guidelines for new datasets and grand challenges. *Hydrol. Sci. J.*, **65**, 712–725, <https://doi.org/10.1080/02626667.2019.1683182>.
- Blöschl, G., and Coauthors, 2017: Changing climate shifts timing of European floods. *Science*, **357**, 588–590, <https://doi.org/10.1126/science.aan2506>.
- Castellarin, A., 2007: Probabilistic envelope curves for design flood estimation at ungauged sites. *Water Resour. Res.*, **43**, W04406, <https://doi.org/10.1029/2005WR004384>.
- CWC, 2022: Flood damage statistics during 1953–2020 (3/38/2012-FHM/11-105). CWC, Ministry of Jal Shakti, Government of India, 40 pp., <https://cwc.gov.in/sites/default/files/flood-damage-data-merged.pdf>.
- CWC, Ministry of Jal Shakti, Government of India, 2020: Flood forecasting/hydrological observation. <http://cwc.gov.in/flood-forecasting-hydrological-observation>.
- Do, H. X., S. Westra, M. Leonard, and L. Gudmundsson, 2020: Global-scale prediction of flood timing using atmospheric reanalysis. *Water Resour. Res.*, **56**, e2019WR024945, <https://doi.org/10.1029/2019WR024945>.
- Gaume, E., and Coauthors, 2009: A compilation of data on European flash floods. *J. Hydrol.*, **367**, 70–78, <https://doi.org/10.1016/j.jhydrol.2008.12.028>.
- GFCC, Government of India, 2004: Flood management guidelines of GFCC 2004. Ganga Flood Control Commission, 72 pp., <https://gfcc.gov.in>.
- Gourley, J. J., and Coauthors, 2013: A unified flash flood database across the United States. *Bull. Amer. Meteor. Soc.*, **94**, 799–805, <https://doi.org/10.1175/BAMS-D-12-00198.1>.
- Huang, Z., H. Wu, R. F. Adler, G. Schumann, J. J. Gourley, A. Kettner, and N. Nanding, 2021: Multisourced flood inventories over the contiguous United States for actual and natural conditions. *Bull. Amer. Meteor. Soc.*, **102**, E1133–E1149, <https://doi.org/10.1175/BAMS-D-20-0001.1>.
- Jain, S. K., P. C. Nayak, Y. Singh, and S. K. Chandniha, 2017: Trends in rainfall and peak flows for some river basins in India. *Curr. Sci.*, **112**, 1712–1726, <https://doi.org/10.18520/cs/v112/i08/1712-1726>.
- Jena, P. P., C. Chatterjee, G. Pradhan, and A. Mishra, 2014: Are recent frequent high floods in Mahanadi basin in eastern India due to increase in extreme rainfalls? *J. Hydrol.*, **517**, 847–862, <https://doi.org/10.1016/j.jhydrol.2014.06.021>.
- Kottek, M., J. Grieser, C. Beck, B. Rudolf, and F. Rubel, 2006: World map of the Köppen-Geiger climate classification updated. *Meteor. Z.*, **15**, 259–263, <https://doi.org/10.1127/0941-2948/2006/0130>.
- Kumar, A., M. Saharia, and P. Kirstetter, 2024: Mapping a novel metric for flash flood recovery using interpretable machine learning. *J. Hydrol.*, **25**, 1863–1875, <https://doi.org/10.1175/JHM-D-23-0196.1>.
- Kummu, M., M. Taka, and J. H. A. Guillaume, 2018: Gridded global datasets for Gross Domestic Product and Human Development Index over 1990–2015. *Sci. Data*, **5**, 180004, <https://doi.org/10.1038/sdata.2018.4>.
- Kuntla, S. K., 2021: An era of Sentinels in flood management: Potential of Sentinel-1, -2, and -3 satellites for effective flood management. *Open Geosci.*, **13**, 1616–1642, <https://doi.org/10.1515/geo-2020-0325>.
- , M. Saharia, and P. Kirstetter, 2022: Global-scale characterization of streamflow extremes. *J. Hydrol.*, **615**, 128668, <https://doi.org/10.1016/j.jhydrol.2022.128668>.
- , S. Prakash, and G. Villarini, 2023: Precipitation inequality exacerbates streamflow inequality, but dams moderate it. *Sci. Total Environ.*, **912**, 169098, <https://doi.org/10.31223/X5VQ3J>.
- Lehner, B., K. Verdin, and A. Jarvis, 2008: New global hydrography derived from spaceborne elevation data. *Eos, Trans. Amer. Geophys. Union*, **89**, 93–94, <https://doi.org/10.1029/2008EO100001>.
- Melton, M. A., 1957: An analysis of the relations among elements of climate, surface properties and geomorphology. Dept. of Geology, Columbia University, Tech. Rep. 11, Project NR 389-042, 118 pp., <https://archive.org/details/analysisofrelati00melt>.
- NWS, 2019: Water Resources Services Program (NWSPD 10-9). National Weather Service, 95 pp., https://www.weather.gov/media/directives/010_pdfs_archived/pd01009023h.pdf.
- NDMA, Government of India, 2008: National disaster management guidelines: Management of floods. National Disaster Management Authority, 168 pp., <https://nidm.gov.in/PDF/guidelines/floods.pdf>.
- Newman, R., and I. Noy, 2023: The global costs of extreme weather that are attributable to climate change. *Nat. Commun.*, **14**, 6103, <https://doi.org/10.1038/s41467-023-41888-1>.
- Nielsen, E. R., G. R. Herman, R. C. Tournay, J. M. Peters, and R. S. Schumacher, 2015: Double impact: When both tornadoes and flash floods threaten the same place at the same time. *Wea. Forecasting*, **30**, 1673–1693, <https://doi.org/10.1175/WAF-D-15-0084.1>.
- Peters-Lidard, C. D., and Coauthors, 2017: Scaling, similarity, and the fourth paradigm for hydrology. *Hydrol. Earth Syst. Sci.*, **21**, 3701–3713, <https://doi.org/10.5194/hess-21-3701-2017>.
- Saharia, M., P.-E. Kirstetter, H. Vergara, J. J. Gourley, and Y. Hong, 2017a: Characterization of floods in the United States. *J. Hydrol.*, **548**, 524–535, <https://doi.org/10.1016/j.jhydrol.2017.03.010>.
- , ———, ———, ———, ———, and M. Giroud, 2017b: Mapping flash flood severity in the United States. *J. Hydrometeorol.*, **18**, 397–411, <https://doi.org/10.1175/JHM-D-16-0082.1>.
- , A. Jain, R. R. Baishya, S. Haobam, O. P. Sreejith, D. S. Pai, and A. Rafieeiniasab, 2021a: India flood inventory: Creation of a multi-source national geospatial database to facilitate comprehensive flood research. *Nat. Hazards*, **108**, 619–633, <https://doi.org/10.1007/s11069-021-04698-6>.
- , P.-E. Kirstetter, H. Vergara, J. J. Gourley, I. Emmanuel, and H. Andrieu, 2021b: On the impact of rainfall spatial variability, geomorphology, and climatology on flash floods. *Water Resour. Res.*, **57**, e2020WR029124, <https://doi.org/10.1029/2020WR029124>.
- Shen, X., Y. Mei, and E. N. Anagnostou, 2017: A comprehensive database of flood events in the contiguous United States from 2002 to 2013. *Bull. Amer. Meteor. Soc.*, **98**, 1493–1502, <https://doi.org/10.1175/BAMS-D-16-0125.1>.
- Smith, B. K., and J. A. Smith, 2015: The flashiest watersheds in the contiguous United States. *J. Hydrometeorol.*, **16**, 2365–2381, <https://doi.org/10.1175/JHM-D-14-0217.1>.