Variables Description for

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

Sai Kiran Kuntla[1] and Manabendra Saharia[1,2]

[1]Department of Civil Engineering, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India

[2]Yardi School of Artificial Intelligence, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India

Contents of this file

Table 1: Information provided in the metadata of the database.

Field	Description
GaugeID	A unique ID is assigned to each gauge station, such as "INDOFLOODS-gauge-1", for easy identification and representation among various subsets of this database.
Warning Level	The level of the streamflow (mtr) above which the streamflow is classified as "Flood".
Danger Level	The level of the streamflow (mtr) above which the streamflow is classified as a "Severe Flood".
Station	Name of the station
Latitude and Longitude	The geographical coordinates of the gauge station.
River Name	The name of the river on which the gauge station is installed. The tributary and sub-tributary name(s) are also provided wherever available.
Basin	The name of the Basin in which the gauge station lies. Names of the basin are as per the records of CWC.
Start date	Date of the oldest record of streamflow data available based on which the flood events in INDOFLOODS are extracted. All dates conform to ISO 8601 (YYYY-MM-DD), the international standard for representing dates and times.
End date	Date of the latest record of streamflow data available based on which the flood events in INDOFLOODS are extracted. All dates conform to ISO 8601 (YYYY-MM-DD), the international standard for representing dates and times.
Level Entries	Number of stream level data entries available between the "Start date" and "End date"
Streamflow Entries	Number of stream discharge data entries available between the "Start date" and "End date"
Privacy	Indicates if the source streamflow data is classified or unclassified using the terms "Open" and "Restricted". Unclassified is open to all users by the CWC and free

	of charge. In comparison, the classified data is restricted and provided only on request.
Source catchment area	The upstream catchment area of the gauge station as per the documents of CWC.
Shapefile catchment area	The upstream catchment area of the gauge station, which was computed using the shapefile developed and provided through the INDOFLOODS database.
Area variation	The percentage of variation between “Source catchment area” and “Shapefile catchment area”.
Shapefile reliability	“Safe” if the “Area variation” is less than 25 percent, and “Caution” if the “Area variation” is more than 25 percent. Also, a shapefile is classified as “Caution” if the “Source catchment area” is unavailable.

Table 2: Description 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”.

Table 3: List of catchment-scale variables provided in the database.

Catchment Characteristics	**Description**	**Reference**
Geomorphology		
Stream order (SO)	Strahler stream order, numerical measure of river's branching complexity	(Strahler, 1952)
Drainage area (DA)	The surface area of the catchment	
Catchment magnitude (CM)	The number of first order streams	(Melton, 1957)
Maximal flow length (MFL)	the length along the longest watercourse from the mouth to the head of the channel	(Mueller, 1968)
Downvalley length (DVL)	The straight distance from the river cell of interest to the basin mouth	(Mueller, 1968)
Catchment relief (CR)	The elevation difference between the highest point on the drainage divide and the mouth	(Costa, 1987)
Catchment length (CL)	The maximal length of the line from a basin mouth to a point on the perimeter equidistant from the basin mouth in either direction around the perimeter	(Gregory & Walling, 1968)
Catchment perimeter (CP)	The length of outer boundary of the watershed that enclosed its area	(Schumm, 1956)
Sinuosity index (SI)	SI = MFL/DVL	(Wolman & Miller, 1960)
Form factor (FF)	$FF = DA/CL^2$	(Horton, 1945)
Relief ratio (RR)	RR = CR/CL	(Schumm, 1956)

Elongation ratio (ER)	$ER = 2/(CL \times (DA/\pi)^{0.5})$	(Schumm, 1956)
Circularity ratio (CR)	$CR = 4\ \pi DA/CP^2$	(Miller & Summerson, 1960)
Lemniscates value (LV)	$LV = CL^2/DA$	(Chorley, 1957)
Drainage texture (DT)	DT = Total number of streams of all order/CP	(Horton, 1945)
Drainage density (DD)	DD = Total length of streams of all orders/DA	(Horton, 1945)
Compactness coefficient (CC)	$CC = 0.2841\ (CP/DA^{0.5})$	(Gravelius, 1914)
Wandering ratio (WR)	MFL/CL	(Smart & Surkan, 1967)
Fitness ratio (FR)	MFL/CP	(Melton, 1957)
Channel frequency (CF)	CF = Total number of streams of all orders/DA	(Horton, 1932)
Drainage intensity (DI)	CF/DD	(Faniran, 1968)
Infiltration number (IN)	CF x DD	(Faniran, 1968)
Ruggedness number (RN)	CR x DD	(Strahler, 1964)
No. of streams by order	Number of first, second, third fourth, fifth, sixth, seventh, and eighth-order streams in the catchment.	(Horton, 1945)

Total length of streams by order (SL)	Total length of first, second, third, fourth, fifth, sixth, seventh, and eighth-order streams in the catchment.	(Horton, 1945)
Mean length of streams by order (MSL)	Mean length of first, second, third, fourth, fifth, sixth, seventh, and eighth-order streams in the catchment.	(Strahler, 1964)
Stream length ratio (SLR)	Ratio between the total length of streams of a given order to the number of streams of the next higher order.	(Horton, 1945)
Bifurcation ratio (BR)	Ratio between the number of streams of a given order to the number of streams of the next higher order.	(Schumm, 1956)
Climatological		
Annual mean temperature (Bio1)	The annual mean temperature	(Fick & Hijmans, 2017)
Mean diurnal range (Bio2)	The mean of the monthly temperature ranges (monthly maximum minus monthly minimum)	
Isothermality (Bio3)	Isothermality quantifies how large the day-to-night temperatures oscillate relative to the summer-to-winter (annual) oscillations. Bio3= (Bio2/Bio7)x100	
Temperature seasonality (Bio4)	The amount of temperature variation over a given year (or averaged years) based on the standard deviation (variation) of monthly temperature averages.	
Maximum temperature of	The maximum monthly temperature occurrence over a given year (time-series) or averaged span of years (normal).	

warmest month (Bio5)		
Minimum temperature of coldest month (Bio6)	The minimum monthly temperature occurrence over a given year (time-series) or averaged span of years (normal).	
Temperature annual range (Bio7)	A measure of temperature variation over a given period. Bio7=Bio5-Bio6	
Mean temperature of wettest Quarter (Bio8)	This quarterly index approximates mean temperatures that prevail during the wettest season.	
Mean temperature of driest quarter (Bio9)	This quarterly index approximates mean temperatures that prevail during the driest quarter.	
Mean temperature of warmest quarter (Bio10)	This quarterly index approximates mean temperatures that prevail during the warmest quarter.	
Mean temperature of coldest quarter (Bio11)	This quarterly index approximates mean temperatures that prevail during the coldest quarter.	
Annual precipitation (Bio12)	This is the sum of all total monthly precipitation values.	
Precipitation of wettest month (Bio13)	This index identifies the total precipitation that prevails during the wettest month.	

Precipitation of driest month (Bio14)	This index identifies the total precipitation that prevails during the driest month.	
Precipitation seasonality (Bio15)	This is a measure of the variation in monthly precipitation totals over the course of the year. This index is the ratio of the standard deviation of the monthly total precipitation to the mean monthly total precipitation (also known as the coefficient of variation) and is expressed as a percentage.	
Precipitation of wettest quarter (Bio16)	This quarterly index approximates total precipitation that prevails during the wettest quarter.	
Precipitation of driest quarter (Bio17)	This quarterly index approximates the total precipitation that prevails during the driest quarter.	
Precipitation of warmest quarter (Bio18)	This quarterly index approximates the total precipitation that prevails during the warmest quarter.	
Precipitation of coldest quarter (Bio19)	This quarterly index approximates the total precipitation that prevails during the coldest quarter.	
Anthropogenic		
Road density	The average road density in the catchments was computed based on the Global Roads Inventory Project (GRIP) dataset. It includes all road types.	(Meijer et al., 2018)
Population Count	The total population in the catchments is computed based on Gridded Population of the World (GPW) database for the year 2010.	(CIESIN, 2016a)

Population density	The average population density (number of persons per square kilometer) in the catchments is computed based on the Gridded Population of the World (GPW) database for the year 2010.	(CIESIN, 2016b)
Urban percentage	The urban area percentage in the catchments for the year 2015 is based on the GHS - Settlement Model grid (SMOD). All the urban clusters in the GHS-SMOD: class 30: urban centre, class 23: dense urban, class 22: semi-dense urban, and class 21: "suburban or per-urban, are considered as urban for computing urban area percentage out of total.	(Pesaresi & Freire, 2016)
Gross Domestic Product (GDP)	The average GDP (PPP - Purchasing Power Parity) and GDP per capita (PPP) in the catchments is provided for the years 1990, 1995, 2000, 2010, 2015. GDP is given in 2011 international US dollars.	(Kummu et al., 2018)
Human Development Index (HDI)	The average HDI in the catchments is provided for the years 1990, 1995, 2000, 2010, 2015. HDI is a composite index of average achievement in key dimensions of human development (dimensionless indicator between 0 and 1).	(Kummu et al., 2018)
Nighttime light	The average nighttime light in the catchment is provided for the year 2010. Nighttime light represents light visible at night generated by human activity, including settlements, gas flaring, or agricultural fires.	(Doll, 2008)
Event-scale precipitation		

1 day precipitation (T1d)	Daily precipitation a day before the flood start date. Units: mm The corrected mean of Em-earth daily precipitation probabilistic estimates available at 0.1° spatial resolution is used to derive event-scale precipitation over the catchments for all the flood events.	(Tang et al., 2022)
2 days to 10 days precipitation (T2d – T10d)	Cumulative 2 days, 3days, 4 days, 5days, 6 days, 7 days, 8 days, 9 days, and 10 days daily precipitation before the flood start date. Units: mm	
Other catchment characteristics		
Lithology type	Catchment lithology if one single lithology type is present over more than 50% of the catchment area. Otherwise, 'No dominant class'.	(Hartmann & Moosdorf, 2012)
Soil type	Catchment soil class (WRB) if one single soil class is present over more than 50% of the catchment area. Otherwise, 'No dominant class'.	(Hengl et al., 2017)
Koppen-Geiger climate type	Catchment climate (major groups of Koppen-Geiger system) if one Climate type is present over more than 50% catchment area. Otherwise, 'No dominant class'.	(Kottek et al., 2006)
Land cover	Catchment land cover (U.N. Classification System for 2015) if one single land-cover type is present over more than 50% catchment area. Otherwise, 'No dominant class'.	The Climate Change Initiative Land Cover (CCI-LC)

The climatological variables provided correspond to average values in the upstream catchment. For example, the annual precipitation (Bio12) variable gives the average annual precipitation in the catchment.

Data sources:

- Worldclim: https://www.worldclim.org/data/bioclim.html
- Precipitation: https://doi.org/10.20383/102.0547
- Global Distributed Basin Characteristics (GDBC): https://figshare.Com/s/6cd00491b850bad716d7
- Land Cover: https://maps.elie.ucl.ac.be/CCI/viewer/
- Koppen-Geiger climate: https://koeppen-geiger.vu-wien.ac.at/present.htm
- Road density: https://www.globio.info/download-grip-dataset
- Nighttime light: https://ngdc.noaa.gov/eog/dmsp/downloadV4composites.html#AXP
- Urban area percentage: https://human-settlement.emergency.copernicus.eu/download.php?ds=smod
- Population density: https://doi.org/10.7927/H49C6VHW
- Population count: https://doi.org/10.7927/H4JW8BX5
- GDP and HDI: https://doi.org/10.1038/sdata.2018.4

References

Chorley, R. J. (1957). Illustrating the Laws of Morphometry. *Geological Magazine*, *94*(2), 140–150. Scopus. https://doi.org/10.1017/S0016756800068412

CIESIN, C. for I. E. S. I. N. (2016a). *Gridded Population of the World (GPW), v4: Population Count*. Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H4JW8BX5

CIESIN, C. for I. E. S. I. N. (2016b). *Gridded Population of the World (GPW), v4: Population Density*. Socioeconomic Data and Applications Center (SEDAC). https://doi.org/10.7927/H49C6VHW

Costa, J. E. (1987). Hydraulics and basin morphometry of the largest flash floods in the conterminous United States. *Journal of Hydrology*, *93*(3), 313–338. https://doi.org/10.1016/0022-1694(87)90102-8

Doll, C. N. H. (2008). *CIESIN Thematic guide to night-time light remote sensing and its applications*. Center for International Earth Science Information Network (CIESIN), Columbia University. https://ciesin.columbia.edu/binaries/web/sedac/thematic-guides/ciesin_nl_tg.pdf

Faniran, A. (1968). The index of drainage intensity—A provisional new drainage factor. *Aust J Sci*, *31*(9), 328–330. Scopus.

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12), 4302–4315. https://doi.org/10.1002/joc.5086

Gravelius, H. (1914). *Flusskunde*. Goschen'sche Verlagshandlung.

Gregory, K. J., & Walling, D. E. (1968). The Variation of Drainage Density Within a Catchment. *International Association of Scientific Hydrology. Bulletin*, *13*(2), 61–68. https://doi.org/10.1080/02626666809493583

Hartmann, J., & Moosdorf, N. (2012). Global Lithological Map Database v1.0 (gridded to 0.5° spatial resolution). In *Supplement to: Hartmann, Jens; Moosdorf, Nils (2012): The new global lithological map database GLiM: A representation of rock properties at the Earth surface. Geochemistry, Geophysics, Geosystems, 13, Q12004, https://doi.org/10.1029/2012GC004370*. PANGAEA. https://doi.org/10.1594/PANGAEA.788537

Hengl, T., Jesus, J. M. de, Heuvelink, G. B. M., Gonzalez, M. R., Kilibarda, M., Blagotić, A., Shangguan, W., Wright, M. N., Geng, X., Bauer-Marschallinger, B., Guevara, M. A., Vargas, R., MacMillan, R. A., Batjes, N. H., Leenaars, J. G. B., Ribeiro, E., Wheeler, I., Mantel, S., & Kempen, B. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE*, *12*(2), e0169748. https://doi.org/10.1371/journal.pone.0169748

Horton, R. E. (1932). Drainage-basin characteristics. *Transactions, American Geophysical Union*, *13*, 350–361. https://doi.org/10.1029/TR013i001p00350

Horton, R. E. (1945). Erosional development of streams and their drainage basins; Hydrophysical approach to quantitative morphology. *GSA Bulletin*, *56*(3), 275–370. https://doi.org/10.1130/0016-7606(1945)56[275:EDOSAT]2.0.CO;2

Kottek, M., Grieser, J., Beck, C., Rudolf, B., & Rubel, F. (2006). World Map of the Köppen-Geiger climate classification updated. *Meteorologische Zeitschrift*, 259–263. https://doi.org/10.1127/0941-2948/2006/0130

Kummu, M., Taka, M., & Guillaume, J. H. A. (2018). Gridded global datasets for Gross Domestic Product and Human Development Index over 1990–2015. *Scientific Data*, *5*(1), 180004. https://doi.org/10.1038/sdata.2018.4

Meijer, J. R., Huijbregts, M. A. J., Schotten, K. C. G. J., & Schipper, A. M. (2018). Global patterns of current and future road infrastructure. *Environmental Research Letters*, *13*(6), 064006. https://doi.org/10.1088/1748-9326/aabd42

Melton, M. A. (1957). *An analysis of the relations among elements of climate, surface properties, and geomorphology*. New York : Department of Geology, Columbia University. http://archive.org/details/analysisofrelati00melt

Miller, O. M., & Summerson, C. H. (1960). Slope-Zone Maps. *Geographical Review*, *50*(2), 194–202. https://doi.org/10.2307/211507

Mueller, J. E. (1968). An Introduction to the Hydraulic and Topographic Sinuosity Indexes. *Annals of the Association of American Geographers*, *58*(2), 371–385. https://doi.org/10.1111/j.1467-8306.1968.tb00650.x

Pesaresi, M., & Freire, S. (2016). *GHS-SMOD R2016A - GHS settlement grid, following the REGIO model 2014 in application to GHSL Landsat and CIESIN GPW v4-multitemporal*

(1975-1990-2000-2015). European Commission, Joint Research Centre (JRC). http://data.europa.eu/89h/jrc-ghsl-ghs_smod_pop_globe_r2016a

Schumm, S. A. (1956). Evolution of drainage systems and slopes in badlands at Perth Amboy, New Jersey. *GSA Bulletin*, *67*(5), 597–646. https://doi.org/10.1130/0016-7606(1956)67[597:EODSAS]2.0.CO;2

Smart, J. S., & Surkan, A. J. (1967). The relation between mainstream length and area in drainage basins. *Water Resources Research*, *3*(4), 963–974. Scopus. https://doi.org/10.1029/WR003i004p00963

Strahler, A. N. (1952). Dynamic Basis of Geomorphology. *GSA Bulletin*, *63*(9), 923–938. https://doi.org/10.1130/0016-7606(1952)63[923:DBOG]2.0.CO;2

Strahler, A. N. (1964). *Quantitative Geomorphology of Drainage Basins and Channel Networks*. McGraw Hill.

Tang, G., Clark, M. P., & Papalexiou, S. M. (2022). EM-Earth: The Ensemble Meteorological Dataset for Planet Earth. *Bulletin of the American Meteorological Society*, *103*(4), E996–E1018. https://doi.org/10.1175/BAMS-D-21-0106.1

The Climate Change Initiative Land Cover (CCI-LC). (n.d.). ESA/CCI Viewer. Retrieved April 15, 2022, from https://maps.elie.ucl.ac.be/CCI/viewer/

Wolman, M. G., & Miller, J. P. (1960). Magnitude and Frequency of Forces in Geomorphic Processes. *The Journal of Geology*, *68*(1), 54–74. https://doi.org/10.1086/626637