Flood Forecasting in Large River Basins Using FOSS Tool and HPC

Upasana Dutta [1,2,*], Yogesh Kumar Singh [1], T. S. Murugesh Prabhu [1], Girishchandra Yendargaye [1], Rohini Gopinath Kale [1], Binay Kumar [1], Manoj Khare [1], Rahul Yadav [1], Ritesh Khattar [3] and Sushant Kumar Samal [4]

[1] Centre for Development of Advanced Computing, C-DAC Innovation Park, Panchawati Road, Panchawati, Pashan, Pune 411008, Maharashtra, India; yogesh@cdac.in (Y.K.S.); murugeshp@cdac.in (T.S.M.P.); girishy@cdac.in (G.Y.); rohinik@cdac.in (R.G.K.); binay@cdac.in (B.K.); manojk@cdac.in (M.K.); rahuly@cdac.in (R.Y.)

[2] Department of Environmental Science, Savaitribai Phule Pune University (SPPU), Ganeshkhind Road, Ganeshkhind, Pune 411007, Maharashtra, India

[3] Sinchai Bhawan, Ganga Flood Control Commission, Department of Water Resources, RD & GR, Ministry of Jal Shakti, Govt. of India, Patna 800015, Bihar, India; riteshkhattar@gmail.com

[4] Hydrological Observation Circle, Central Water Commission (CWC), Mahanadi Bhawan, Department of Water Resources, RD & GR, Ministry of Jal Shakti, Govt. of India, Bhoi Nagar, Bhubaneshwar 751002, Odisha, India; sehocbbshwar-cwc@nic.in

* Correspondence: upasanad@cdac.in

Abstract: The Indian subcontinent is annually affected by floods that cause profound irreversible damage to crops and livelihoods. With increased incidences of floods and their related catastrophes, the design, development, and deployment of an Early Warning System for Flood Prediction (EWS-FP) for the river basins of India is needed, along with timely dissemination of flood-related information for mitigation of disaster impacts. Accurately drafted and disseminated early warnings/advisories may significantly reduce economic losses incurred due to floods. This study describes the design and development of an EWS-FP using advanced computational tools/methods, viz. HPC, remote sensing, GIS technologies, and open-source tools for the Mahanadi River Basin of India. The flood prediction is based on a robust 2D hydrodynamic model, which solves shallow water equations using the finite volume method. The model is open-source, supports geographic file formats, and is capable of simulating rainfall run-off, river routing, and tidal forcing, simultaneously. The model was tested for a part of the Mahanadi River Basin (Mahanadi Delta, 9225 sq km) with actual and predicted discharge, rainfall, and tide data. The simulated flood inundation spread and stage were compared with SAR data and CWC Observed Gauge data, respectively. The system shows good accuracy and better lead time suitable for flood forecasting in near-real-time.

Keywords: flood; prediction; large river basin; FOSS; HPC; ANUGA Hydro; Mahanadi River Basin

1. Introduction

The Indian subcontinent is regularly affected by floods that have a significant impact on life and property. Despite their regular occurrence, floods are difficult to predict, especially in India, for a multitude of reasons. The vast and interconnected network of rivers in the Indian subcontinent renders the data, which is minimal in most cases, and coarse, that flood prediction at finer resolution tends to be extremely difficult. There is also a dearth of models suiting Indian conditions or having provisions to modify them based on user requirements to improve accuracy. Moreover, most of the river basins of India are quite large, and as such, do not suit most flood modelling software because of their immense computational requirements.

The current study addresses the following issues: forecasting floods, using open-source models, and modelling flood prediction for large river basins. Flood forecasting, despite posing significant challenges to hydrologists, is a widely used method to manage

floods. Integrated meteorological and hydrological modelling, improved data collection through satellite observations, and advanced data processing algorithms have improved the reliability of forecasts. However, the solutions have limitations of cost-effectiveness and a lack of user-interactive software interface [1].

Forecasting floods requires parameters that should be as close as possible to replicating real-world scenarios. The crucial parameters include topography and rainfall and other area-specific parameters, like river discharge, surface roughness, etc.

Topographical data represented by Digital Elevation Models (DEM) is crucial for flood prediction. This data gives a near-accurate representation of the underlying features of the flood plain and river basin areas and has a key role in facilitating flood studies. Since the pioneering work of Miller and Laflamme [2], DEMs have been indispensable in their scientific applications. DEMs are primarily created using remote sensing techniques [3], covering a large spatial area at a lower cost. Open-access global DEMs are a boon to the scientific and academic communities. However, due to their poor resolution (>30 m), significant vertical errors hamper the accurate estimation of flood hazards. Remote sensing techniques include photogrammetry [4], airborne and space-borne Interferometric Synthetic Aperture Radar (InSAR) and Light Detection and Ranging (LiDAR). The use of high-resolution (~1 m) LiDAR data improves the accuracy of floodplain mapping, as it represents the slope and obstacles accurately.

During the design and management of flood control systems, hydrologists are also often faced with the challenge of predicting the peak discharge and the magnitude of rainfall generated run-offs for watersheds. The magnitude and time variation of rainfall has been more difficult to predict, mainly due to the inherently stochastic nature of the rainfall events. Associated run-offs from these events are thus expected to be just as difficult to predict. Most inputs for environmental models involve spatial variations, and hydrological models are particularly sensitive to the spatial variation of rainfall events [5,6]. Reliable and accurate representation of rainfall events is, thus, vital for hydrological modelling and flood forecasting. Rainfall–run-off models are sensitive to inputs from precipitation data, and if these inputs do not characterise the true precipitation correctly, then there is no empirical or physical model that could produce accurate streamflow simulations [7–10].

Flood modelling in large-sized basins (like the Mahanadi Basin) at sub-daily time scales or in the order of a few hours is challenging since the peak discharges tend to occur as a result of a localised rainfall event, and time taken to reach the flood peak may be inadequate for raising timely warnings based on real-time rainfall observations [11]. In such cases, potential flood warnings can be only given based on rainfall forecasts. Although the accuracy of rainfall forecasts has improved with new technologies and methods, rainfall forecasts still are the main source of uncertainty in flood forecasting, which limits the usability of hydrological models in operational applications, especially in the Indian region.

Hydrological modelling of large river basins is a challenging and complex issue. There are various parameters at play, and the large areas over which they need to be computed makes the modelling particularly difficult. Parallel computing technology provides a good computational means for the construction of national-level or basin-level flash flood warning systems having high resolution or local-level warning analysis [12,13]. High-Performance Computing (HPC) techniques and resources allow a higher spatial and temporal resolution to be used during simulation runs. HPC also allows handling of modelling over larger areas with ease. High-performance computers with capacities at the order of teraflops and petaflops prove useful while running simulations on such big areas at moderate resolutions. Flood simulation is a time-critical activity for flood forecasting; obtaining results sooner has been proven to be better and can only be achieved using high-performance computers [14].

Flood forecasting tools are available both commercial-off-the-shelf (COTS) and Free and Open source (FOSS). In this paper, the use of open-source software and tools are discussed. FOSS provides the flexibility of handling parameters as compared to COTS.

This paper describes the use of a few open-source tools and software that helped in flood simulation and analysis.

The model presented in this paper is developed for the Mahanadi River Basin and is meant to be applicable to all river basins of India, keeping flood-ravaged basins as a priority. In this paper, flood simulation for a part of the Mahanadi River Basin is done using an open-source hydrodynamic model (ANUGA Hydro) using freely available GIS tools for visualisation and analysis. This paper also reports the use of open-source tools/packages for data preparation, design, and set up of flood-simulation models and HPC-based analysis of the simulation results.

Last but not least, validation of forecasts is as important as the forecast itself. Forecasts are validated in various ways, such as conducting ground-truthing or using satellite-based information. During the flood season, ground-truthing often becomes difficult, as most of the area gets inundated and is not accessible. Owing to its all-weather and day-and-night capabilities and wider spatial coverage, satellite-based data such as Synthetic Aperture Radar (SAR) data becomes very useful in such conditions to estimate the flood inundation [15–17]. SAR-based techniques for flood detection include thresholding-based methods [18], image segmentation [19], statistical active contouring [20], rule-based classification [21], and data fusion approaches.

2. Materials and Methods

The study area selected is the Mahanadi River Basin. It is the fourth largest river basin of India and a vital water resource for the State of Odisha. Out of the overall catchment area of Mahanadi of 145,600 sq km, 65,600 sq km of the area is in Odisha, which comprises 42.15% of the geographical area of the state and nearly 4.3% of the total geographical area of the country (Figure 1). Mahanadi river's annual flows account for 59.16 billion m^3 from the catchment in Odisha, according to the India-WRIS data (Water Resources Information System, Govt. of India). The work mentioned in this paper has been conducted over an area of 9225 sq km out of the total 145,600 sq km area of the Mahanadi basin, covering the deltaic region of the basin.
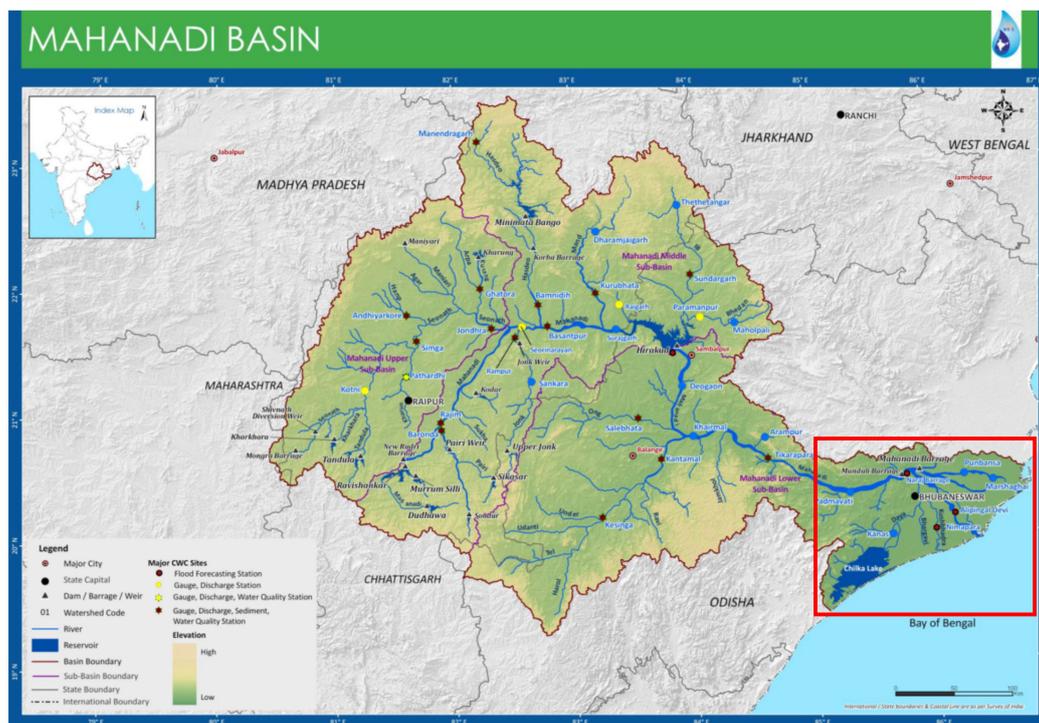
Figure 1. Study area: Mahanadi River Basin and Mahanadi delta region (red bounded box). Source www.india-wris.nrsc.gov.in (accessed on 15 March 2020).

Climatologically, the area is sub-tropical, hot, and humid. The average annual rainfall is 1572 mm, over 70% of which is precipitated during the southwest monsoon between mid-June and mid-September. After navigating a lengthy distance of more than 800 km, the Mahanadi River starts building up its delta plain from Naraj, where the whole Mahanadi branch forms its distributary system splitting in the delta plain area. Devi River is its principal distributary. As per a study carried out by Jahannathan et al., the Mahanadi River arcuate delta system was formed in a tectonic downwarp of the Gondwana graben, which is believed to be a failed arm of the triple junction on the eastern Indian coast passive margin [22]. The ridges and depressions in the area are affected by the presence of faults. Coastal and offshore areas recorded new basin development during the Tertiary. Two sets of major structural/fracture trends and a number of lineaments are identified in the delta plain area [23]. The undivided Mahanadi River at its delta head at Naraj carries an annual average discharge of 48,691 million cubic metres of water with a monsoonal component amounting to ca. 41,000 million cubic metres.

To assess the yearly flooding of the Mahanadi River Basin area, flood forecasting was carried out for the highly inundated deltaic region of the basin.

The following paragraphs detail the model, tools, and parameters used to carry out the same.

2.1. Model and FOSS Tools

Open Source Software ANUGA hydro and Q-GIS provide flexibilities to simulate real-world conditions [24].

2.1.1. ANUGA Hydro

ANUGA Hydro is a free and open-source software package with high capability to model hydrodynamic shallow water equations, which makes it suitable for predicting hydrological disasters such as riverine flooding, storm surges, and tsunamis. A flood inundation model for the delta region within the Mahanadi River Basin was designed using this software. The model routes the water from various sources such as rainfall, upstream discharge, and coastal tides over the topography of the study area, taking into account the spatial variability of the roughness coefficient. The model mimics the properties of the physical basin within the computational environment. The parameters that affect the water flow are set within the computational environment to predict flood inundation. Most of the ANUGA components are written in the object-oriented programming language Python. Computationally intensive components are written for efficiency in C routines working directly with Python NumPy structures. It also has an ANUGA viewer for graphical 3D rendering for animating the output files. It has additional viewing options using Crayfish (QGIS), which is extensively used in the current study.

2.1.2. QGIS

QGIS is a free and open-source cross-platform desktop geographic information system application. It has support to view, edit, and analyse the geospatial data. QGIS also has many plugins that proved invaluable during the data preparation. Once the simulation was carried out, the output was analysed using the Crayfish plugin available in QGIS, which is compatible with the output generated by ANUGA Hydro. Crayfish is a very versatile plugin, and it not only allows the viewing and analysis of the flood inundation maps but also provides information regarding other derived flood parameters.

2.2. High-Performance Computing Platform

Domain decomposition technique is being used in ANUGA to parallelise the code. PyMetis, which is a Python wrapper for the Metis graph partitioning, is used to divide the mesh for parallel computation.

PARAM Brahma Supercomputing facility setup at IISER Pune was used to achieve speedup in performance for a total computational domain of area 9225 sq m, with a mesh size of 900 sq m.

PARAM Brahma system is based on processor Intel Xeon Platinum 8268 with a total peak performance of 850 TFLOPS. The cluster consists of 189 nodes (2 + 4 + 4 + 179) connected with BullSequana XH2000 HDR 100 Infini Band interconnect network. The system uses the Lustre parallel file system. It consists of 2 Master nodes, 4 Login nodes, 4 Service nodes and 179 CPU nodes. The CPU nodes are workhorses of PARAM Brahma. Specifications of PARAM Brahma is shown in Figure 2. All the CPU intensive activities are carried on these nodes. Users can access these nodes from the login node to run interactive or batch jobs. However, the full capacity of the system was not used, and only 60 nodes (48 cores per node) were used for the simulation. Experimental simulation runs were also carried out to check the performance and time taken with an increase in the number of nodes. Some of the nodes have higher memory, which were also used to carry out memory intensive parts of the model run.

CPU only Compute Nodes : 140	
2* Intel Xeon Platinum 8268	Total Cores = 6720 cores
Cores = 48, 2.9 GHz	
Memory= 192 GB, DDR4 2933 MHz	Total Memory=26880 GB
SSD = 480 GB (local scratch) per node	

CPU only Compute Nodes with High memory : 39	
2* Intel Xeon Platinum 8268	
Cores = 48, 2.9 GHz	Total Cores = 1872 cores
Memory= 768 GB, DDR4 2933 MHz	Total Memory=29952 GB
SSD = 480 GB (local scratch) per node	

Figure 2. Specifications of PARAM Brahma.

2.3. Simulation Parameters and Data Preprocessing

The data was sourced from various public and government domains, and as such, required considerable pre-processing and comparative analyses.

- Rainfall: Reliable and precise depictions of rainfall are vital for any hydrological modelling and flood forecasting. Multiple rainfall data sets were used for an accurate representation of rainfall for running the flood forecasting model. It included both global datasets, as well as indigenous data. For the real-time data, Global Precipitation Model (GPM) and Indian Meteorological Department (IMD) rainfall products were used. Global Forecasting System (GFS) and IMD products were used for the forecasted rainfall. These are summarised in Figure 3. Rainfall data was provided as an input for the entire domain. All the datasets used were in a grid format with different grid sizes, as mentioned below. Station rainfall data, which was obtained from CWC, were calculated over specific grids points using the Harversian formula and interpolation technique (Inverse Distance Weighted-IDW) written in Fortran programming lan-

guage. The outputs were saved in control points (header information) and binary files (for data). Finally, using the Climate Data Operator tool, this binary file was converted to netCDF format as per desired grid size. The daily/3 hourly data from 1 June 2020 to 31 October 2020 was used for the study. Depending on availability, IMD daily data and GPM 3 hourly data was used for previous day rainfall, and IMD forecasted and GFS forecasted 3 hourly data was used as predicted rainfall for current plus 2-day simulation.

- River discharge: Among the several real-time hydrological inputs for hydrological modelling, researchers [25–28] have found that river discharge, which is one of the uncertain variables, has to be considered for calculating the actual inflow of water released from dams and barrages into the flood plain. Considering the impact of discharge and controlled river discharge acquired from CWC data at two barrage sites, Naraj and Jobra (Mahanadi), located upstream of the delta region, were included in the simulation (see Figure 4).

- Topographic data: Topography is perhaps the key factor for the assessment of flood extent [29], but typically flood models use limited DEMs and focus more on exploring the uncertainty associated with other hydraulic parameters [30]. Largely, the quality of flood predictions does not necessarily increase with the higher resolution of DEMs. Also, too much detail can yield spurious results, which may not represent the uncertainties in making flood predictions [31,32]. To accurately represent the topographic information of the area, DEMs were sourced from both commercial and open-source platforms. In this study, 1 m LIDAR data (7550 sq km) and 30 m ALOS Prism data (for the remaining 1675 sq km) was used. The general elevation of the Mahanadi delta region ranges from 0 to 260 m approximately, as illustrated in Figure 4. A comparative analysis was carried out between all open-domain DEMs available before selecting ALOS Prism. LIDAR 1 m data, provided by CWC Delhi, covers an area of 7550 sq km (part of Delta region). For the remaining part of the delta region, various open-source DEMs were explored including, SRTM (86.453 m), ASTER (28.818 m), and ALOS (29.061 m). After comparison with SOI benchmarks (from topographic sheets) and running various simulations, comparisons with bathymetric data obtained from the ground survey done by CWC and comparing results with actual inundation extent (compared with SAR inundation output) and water level data (obtained from daily observation done by CWC), ALOS prism product was selected for the remaining part of the delta. Further DEM value extraction and merging were carried out using GDAL [33] library. The simulation was carried out such that other parameters were kept constant, and only the DEM was changed.

- Surface roughness: To understand the complete characteristics of a terrain, an effective roughness value needed to be incorporated into the model. The roughness value often changes spatially along the river and flood plain depending upon the riverbed material and its surrounding features. It is important to sufficiently represent the actual roughness characteristics of the floodplain and channel in order to reduce the uncertainties involved in the flood's travel time over the domain. The preliminary selection of Manning's roughness in the study area was based on land use-land cover characteristics (base map prepared using Sentinel-1 January 2016 imagery which was further improved using January 2020 Sentinel-1 imagery and terrain properties of the area) as presented in the referred publications [34,35]. The hydraulic model in this study was simulated for different values of channel roughness to study the uncertainties associated with it.

- Tidal data: Tidal height was used as one of the boundary conditions. As the simulation was carried out for the delta region of the Mahanadi River Basin, the effect of tidal water was also taken into consideration. The data was sourced from the Survey of India, which is in the open domain. The tidal data was then converted into hourly data using Rule-of-Twelfth, an approximation of the sinusoidal curve fitting. It is

basically a simplified method to estimate intermediate times and heights between high and low water without having to refer to tidal curves or graphs.

- SAR data: The Interferometric Wide (IW) swath mode C Band Ground Range Detected (GRD) datasets from Sentinel-1A satellite were used for the study (refer to Data Availability Statement at the end for details). IW swath mode is the main acquisition mode over land. For the study, we used intensity VV polarisation data covering the delta region of the Mahanadi basin. Some basic pre-processing has already been incorporated in the level-1 GRD dataset used here. For further processing, we used ESA's Sentinel Application Platform (SNAP) version 8.0 64-bit. For data processing, we first performed noise removal from the data (thermal noise removal) followed by orbital file calibration (Speckle Filtering). The refined Lee filter with 3×3 window size was applied to all calibrated data to reduce the inherent SAR speckle noise and improve the signal-to-noise ratio. Subsequently, GRD border noise removal was carried out on the data. To get the true pixel values of the image representing the radar backscattering of the reflecting surface, radiometric calibration to sigma nought (dB) was done using sensor calibration parameters. After calibration to sigma nought, the data were clipped/sub-set for the study area. In the GRD imagery provided by ESA, geometric distortions due to terrain effects are not considered for all areas. Therefore, the data were further subjected to correction to improve the geolocation accuracy. SRTM 1-arcsecond global data were used to overcome this issue and projected to World Geodetic System-1984 (WGS-84) Coordinate System geographical coordinates. Due to the active nature of the SAR system, all imagery is acquired in slant-looking geometry, which elevates the ground due to the presence of hills and valleys, and in turn, the travelling time of the signal is distorted, causing geometric shifts. To correct this error, Range–Doppler terrain correction is applied to the imagery.

Data Provider	IMD (Delhi)	GFS (Global)	GFS (India)	Station data (Mahanadi River Basin)	IMD Station + Satellite Data (India)	IMD Station data (India)	IMERG (Global)
Format	.nc	.grib	.grib	.xls	.grd	.grd	.tif
Resolution	3 km	25 km	25 km	25 km/10 km	25 km	25 km	10-11 km
Duration	72 hrs	360 hrs	240 hrs	24 hrs	24 hrs	24 hrs	24 hrs
Availability	12 PM	9.30 AM	2 PM	2:30 PM	12 PM	3 PM	10 AM
Priority	1st	2nd	3rd	1st	2nd	3rd	4th

Figure 3. Prioritisation of rainfall products as used in the simulation.
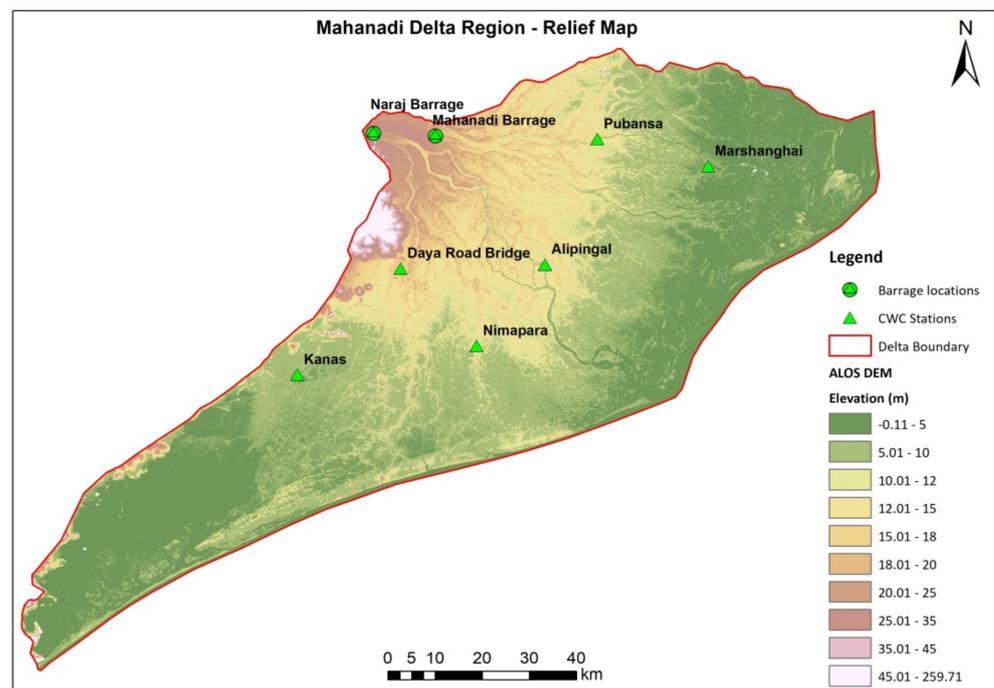
Figure 4. Relief Map of Mahanadi Delta Region showing barrages and gauge locations.

After this pre-processing, a georeferenced, radiometrically calibrated, and speckle-removed image was obtained. In the image, in general, the darkened areas during monsoon season (low values) represented either high moisture or water due to flooding (This did not include rivers, lakes, and ponds, which are permanent water bodies). Image thresholding was carried out to get a better distinction between land and water masses. Finally, Band math's SNAP tool was used to generate a binary image (water masking). The end product was exported to the Google Earth KMZ file for visualisation and for cross-checking with simulated outputs for the same date and time.

2.4. Simulation Setup

The simulation was carried out using ANUGA Hydro. The model setup was made such that it represents the study area as accurately as possible.

Efficient flood forecasting depends on various processes performed on time and in a correct sequence; for example, precise data collection, correct data format conversion (rainfall, tide, discharge etc.), efficient use of HPC (C-DAC NSM HPC clusters), and on-time result generation. To address all these requirements, a multi-layered architecture is detailed in this paper, as shown in Figure 5. The variation of these parameters from basin to basin and their effects on the efficiency of the model encouraged proposing this layered architecture. The proposed architecture involves three layers: Data Collection and Preprocessing Layer (DCPL), Water Flow Computation on HPC Layer (WFCH), and the topmost layer called the Post Processing and Data Dissemination Layer (PPDDL). A specific set of functionalities are performed by every layer independently, and their outputs are used in the higher layers. Priority and availability-based collection of roughness data, rainfall data (Station/IMD/GFS), tide data, and inlet/outlet discharge data are done at DCPL from various sensors and manual resources. Pre-processing involves discharge data conversion, tide data conversion, roughness data, DEM data extraction etc. After pre-processing, the extracted rainfall/tide/discharge/roughness/DEM data is used by WFCH for performing further computation to forecast floods in any geographical area effectively. Inundation, water level extraction, and data dissemination over the portal are done at PPDDL. This layer manages alerts and notifications as per user requirements.
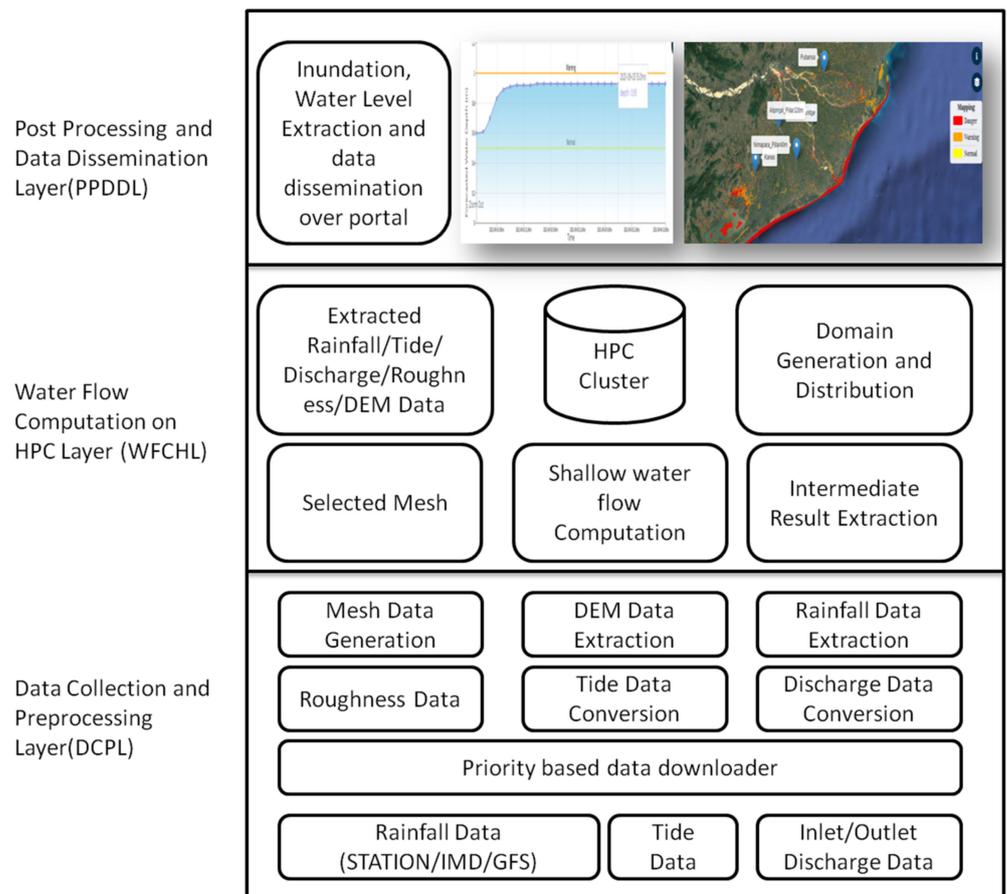
Figure 5. High-level design of model setup.

Flood simulation using ANUGA hydrodynamic model also required the following mandatory inputs apart from the ones detailed under simulation parameters:

- Mesh resolution: ANUGA generates a mesh, which discretises the study area into small elements, within which the Shallow Water Wave equations are run, to estimate the flood depth. The maximum mesh resolution was set at 900 sq. m, which yielded 15,777,513 mesh elements (triangles).
- Boundary condition was defined for the computational domain area to allow the model to understand the behaviour of the flow of water at its edges. Tidal heights are set as a time boundary condition at the edges of the seaside, and the rest of the edges were set as a reflective boundary condition.

The model supports different solver algorithms with varying polynomial order. Based on the recommendation by the model developers and trial simulations, Discontinuous Elevation (DE1) was chosen, which gives the required accuracy, optimum time of computation, and numerical stability. DE1 invokes the numerical solver, discontinuous elevation of polynomial order 2, which gives stability when using sudden bed gradients and tidal forcing in the model.

The simulation model computes the water level and momentum in each mesh element present in the domain based on an implicit finite volume scheme. Though the internal time step is decided by the model, based on Courant–Friedrichs–Lewy (CFL) condition, the output can be written to an external file at user-specified intervals (yield step). Based on user observations, this was set at 3-hour intervals. The simulation duration was kept at 5 days. The forecast logic is elucidated as follows:

If the current day is 31 August 2020, the simulation runs from 0600 h IST on 30 August 2020 to 0600 h IST on 4 September 2020. Day 1 simulation is from 0600 h IST on 30 August

2020 to 0600 h IST on 31 August 2020, which uses actual discharge, rainfall, and tidal data. Day 2 to day 5 simulation is from 0600 h IST on 31 August 2020 to 0600 h IST on 4 September 2020, which uses predicted discharge, rainfall and tidal data. Since the forecast system (EWS-FP) is set to operate for the entire flood season (August to October), daily simulations are run for a 5-day duration perpetually, with overlapping runs for day 1.

This strategy helps to minimise forecast error for longer durations, since day 1 is run with actual data. Likewise, to account for the previous state of the water spread, a 'hot start' method is implemented. So, for the 31 August 2020 simulation, a hot start with the water spread state obtained from the previous day's simulation is used as an initial condition, that is, water levels at 0600 h IST on 30 August 2020 are set as initial water levels to the corresponding mesh elements. The model setup is illustrated in the flowchart below (Figure 6):
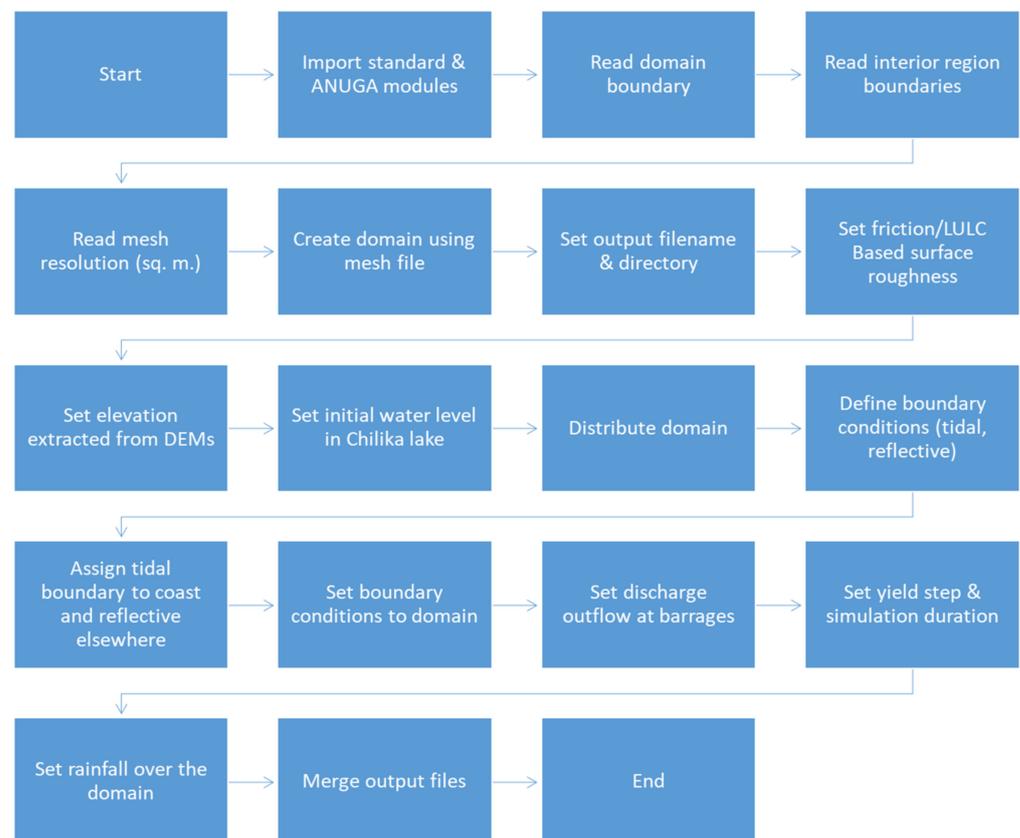
Figure 6. Detailed model setup flowchart.

3. Results

3.1. Model/Simulation

- The model result for a 5–day simulation of the delta region was obtained within a span of 3 h 26 min on 60 Nodes of PARAM Brahma. The simulated inundation output within Mahanadi delta region for 31 August 2020 is shown in Figure 7 below.
- For an area of 9225 sq. km with a maximum mesh resolution of 900 sq. m, the model had 15,777,513 mesh elements (triangles). The 5–day simulation was carried out on PARAM Brahma HPC (Architecture). Performance statistics with different nodes are shown below in Figure 8.
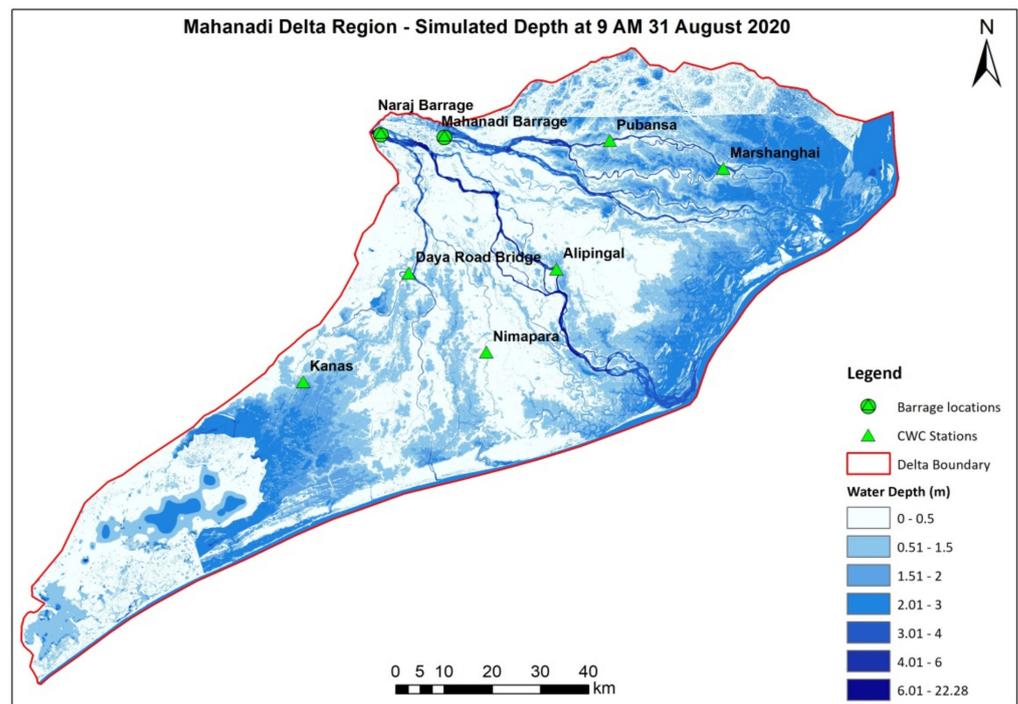
Figure 7. Simulated output of Mahanadi delta region with discharge and gauge locations.
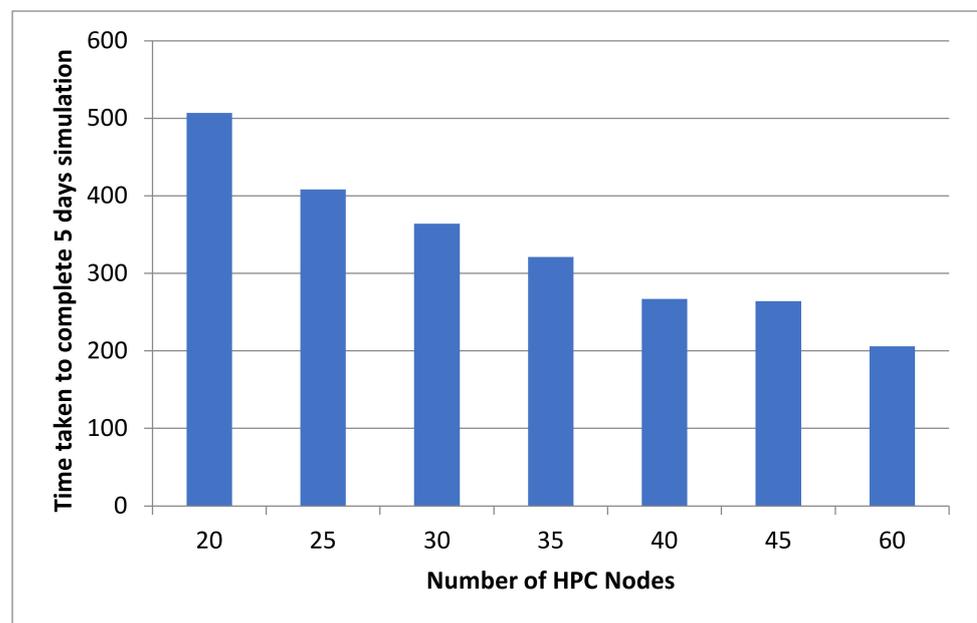
Figure 8. Graphical representation of benchmarking of nodes (number) vs. time (in minutes) on PARAM Brahma.

3.2. Validation of Simulation

It was found that the method adopted in this study can clearly distinguish between water and land surface and can also give a fairly accurate understanding of the flood area extent (quantitative measurement) using Sentinel-1 SAR data. As far as methodology is concerned, it is simple to implement, especially with the availability of the SNAP tool, which is freely made available. SAR images from the start of the flood season, i.e., June 2020 to the end of October 2020. Figure 9 shows a visual comparison of simulated output with SAR output overlaid over Google Maps. In general, the extent of flood

showed similar patterns on both simulated and SAR data; however, the extent is at a lower side on SAR as compared to the simulated output. The flooding maps obtained over different dates also show significant variability in both spatial and temporal extents, with the floods mainly occurring along the riverbanks (on sites like Kanas and Nimapara), becoming more extensive as it spreads towards the coastal part due to the additional effect of tidal water (near Marshaghai where tidal water encroaches inside up to almost 20–25 km). The mapping results illustrate the increase in the entire flood extent from 15 July to 10 September 2020.
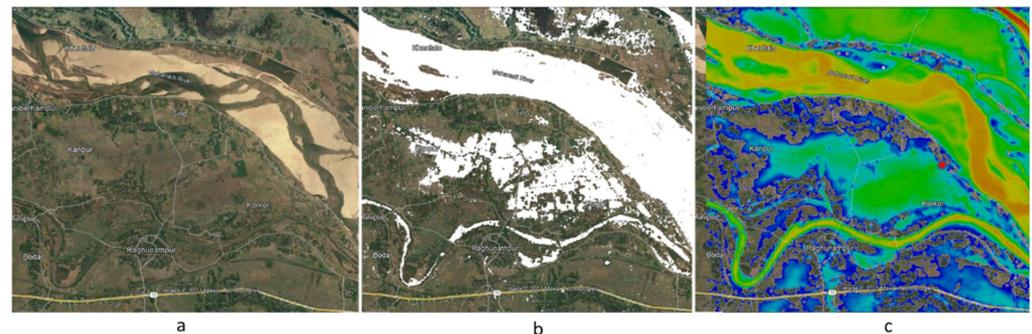
Figure 9. Simulated inundation comparison with SAR data. (**a**) Google Earth Image of the field survey location (**b**) Extracted SAR Inundation—9 AM 31 August 2020 (**c**) Simulated Inundation—9 AM 31 August 2020 (red dot indicates field survey location).

4. Conclusions and Summary

The Mahanadi River Basin chosen for the study has topological complexities. During the monsoon season, which roughly lasts for about 3 months, the basin accommodates a large amount of water, which inundates most parts of the delta region. The basin deals with the monsoon rainfall, as well as increased dam release from the Hirakud dam upstream and occasional storm surges from the Bay of Bengal. However, the Mahanadi River Basin has been deemed a non-classified basin according to CWC, which means that the hydrological data pertaining to this river basin is available in the open domain, as per Govt. of India policy. This provided an advantage for conducting the study. For certain parameters, although the data was available, it was sparse and archaic. Daily hydrological and meteorological data were provided by the respective authorities for daily and predicted simulation. Other parameters had consistency issues, which were handled during pre-processing of the data. Most of the data were subjected to comparative analysis to select the best fit.

4.1. Rainfall Data

The selection of data and prioritisation of the source of data was based both on availability and reasonable accuracy for both predicted and real-time data.

4.2. Flood Inundation Extent and Water Level Verification

Simulated outputs were compared with ground data for verification purposes. It was found to be almost the same with a difference of 0.1–0.5 m difference in the water level at certain flood forecasting sites. The ground data is manually collected daily by the state water resources department. The difference was attributed to the use of DEM, which was a 1 m LIDAR acquired in the year 2005–06. The deposition and erosion of the river bed may have contributed to this difference. Although not conclusive, it is a plausible deduction.

Four gauge observations (see Figure 6) were compared with the simulated flood (Alipingal, Daya Road Bridge, Nimapara, Pubansa), as shown in Table 1 below. Alipingal and Pubansa, which are on the main river, show the best overall accuracy. However, at the Daya Road site, the results were poor, with both false negatives and positives at the higher end, so it is reflected in the kappa value and overall accuracy. The inaccuracy could be due

to the uncertainty in the river bathymetry. Similar to flood extent, water level comparison has also shown the same trend, with higher accuracy at Alipingal and Pubansa and lesser at the Daya Road site.

Table 1. Comparison of Observed and Simulated Stage in four gauge observations sites.

Date/Time	Nimapara		Daya Rd. Bridge		Alipingal		Pubansa	
	Stage Observed	Stage Simulated	Stage Observed	Stage Simulated	Stage Observed	Stage Simulated	Stage Observed	Stage Simulated
31 August 2020 06:00	10.140	8.124	16.180	14.439	11.330	9.688	11.460	10.650
31 August 2020 09:00	10.060	8.119	16.140	14.413	11.310	9.626	11.380	10.641
31 August 2020 12:00	10.080	8.118	16.160	14.438	11.310	9.695	11.400	10.648
31 August 2020 15:00	10.100	8.118	16.160	14.421	11.320	9.651	11.440	10.647
31 August 2020 18:00	10.100	8.121	16.180	14.432	11.330	9.667	11.460	10.647
31 August 2020 21:00	10.120	8.129	16.180	14.440	11.330	9.684	11.460	10.650
31 August 2020 23:00	10.120	8.126	16.180	14.440	11.330	9.687	11.460	10.653

4.3. Digital Elevation Model and Flood Progression

A comparison was done based on reasonable accuracy and computational requirements needed for processing the same. We found that 1 m LIDAR and 30 m ALOS provided a fair inundation pattern, almost mimicking the ground observations. However, it was noticed that the use of LIDAR, although it showed a lesser inundation area, it had captured minute differences in topography accurately. ALOS, compared to LIDAR, showed more spread of inundation. This led the authors to reassert that topographical data plays a pivotal role in determining the correctness of model outputs such as flood progression and flood inundation. It is, therefore, essential to understand the difference in the inundation area obtained by different DEMs before choosing the appropriate DEM for the study.

4.4. Validation of Simulated Output with SAR Data

A flood mapping method must be able to capture floods in near-real-time, which means that it should operate rapidly with little computational cost, no tedious training, and minimal parameter adjustment processes, especially when using multiple images, in order to observe flood evolution patterns. Sentinel-1A images were used to find the extent of inundation during seasonal flood events in the study area. The thresholding method applied in the study has proven to be simple and fairly accurate for mapping floods using SAR data. The thresholding was carried out manually, but in the long term, while processing multiple data sets simultaneously, an automatic thresholding method can also be implemented. The C-band sentinel data seems to be working fine in the delta part of the study area, as the vegetation is predominantly paddy which does not have much of the double-bounce issue but mostly has normal volume scattering returns. Similar results were also obtained while using different SAR data while using the above mentioned pre-processing steps and thresholding for flood mapping under environmental and vegetative conditions. Since the area was dominated by paddy crops, VV polarised intensity data were used in this study, although the possibility of using HH/VH/full-polarised features and interferometric coherence is not ruled out to extract inundated areas. With the availability of both Sentinel-1 A and B and study areas falling near the equatorial region, a better

repeat cycle will be highly useful, especially during the monsoon season. However, in the delta region, where shuttle change in the water level inundates a huge area within a short span of time, the current temporal resolution of the sentinel data needs improvement. The inundation results matched very well with the available cloud-free optical data, with an advantage over SAR for vegetative areas where optical data does not capture water. In certain pockets, some false backscattering values were obtained mainly due to wet and flooded muddy soil resulting in a high dielectric constant value which increases the values of the SAR intensity image. These are often classified as no-water regions; however, there is water there.

The correctness of the output obtained from the model simulation was estimated by comparing the results with the inundation area obtained from Sentinel-1 data as shown in Table 2 below. For a quantitative assessment of accuracy, kappa coefficient, F-measure, which balances the producer's and user's accuracy by performing a weighted average computation [36] and probabilities of false positives (PFP) and false negatives (PFN), which denote the commission and omission errors, respectively, were analysed for the sites where ground observations were available.

Table 2. Quantitative comparisons and accuracy assessments for four validation sites. PFP and PFN refer to the probabilities of false positives and false negatives, respectively.

Sites	PFP (%)	PFN (%)	F-Measure (%)	Kappa	Overall Accuracy (%)
Alipingal	2.15	2.52	94.5	0.93	96.7
Daya Road	8.55	14.83	75.12	0.74	74.6
Nimapara	4.03	7.11	85.2	0.84	82.3
Pubansa	2.83	4.8	86.12	0.87	86.86

4.5. Open Source Tools

Indian river basins, due to being enormous, have notoriously complex behaviour of hydrological parameters. Open-source tools have the capability of handling flood predictions on such river basins with ease. Open-source tools provided the inclusion of parameters based on ground conditions, which also meant that its scalable to include every kind of river basin.

4.6. Flood Forecast Lead Time—HPC Performance Statistics

As demonstrated in the results section, it was noted that the computation speed improved with an increase in the number of nodes on the HPC system, leading to better flood forecast lead time. However, it reaches a point when the speedup does not have an increasing trend when applied to the same area size with an increase in the number of nodes. It reaches a certain stagnancy as far as computation time is concerned. More analysis, however, is required to acquire optimisation for HPC performance versus resolution of the simulated area.

Summary

Floods are an annual calamity affecting the Indian subcontinent. With vast areas underwater, and most of the time inaccessible, it becomes imperative to have information about the ground situation for the Relief & Response teams to act upon. Flood simulation and analysis is a very important and complicated application, and hence the tools and software that help accomplish that become significantly and singularly important, more so if the tools are malleable enough to be tweaked for existing ground conditions. Also, flood simulation for forecasting is a time-critical application and the sooner the results are obtained, the better. This can be achieved using high-performance computers and, clubbed with large-sized river basins, this becomes all the more important.

The paper aims to highlight these requirements and the methodologies and tools adopted that can be helpful when performing inundation prediction for large river basins

like the Mahanadi River Basin. This will help the flood managers take a holistic approach in handling floods. The largeness of the basins define the flood arrival time, and as such, a predictive application, as discussed in this paper, will provide enough lead time for disaster managers to work. The application, as discussed, has been kept flexible and scalable to adapt to all the flood-affected river basins of the country. It will have a significant impact on the current flood management scenario of the country.

Author Contributions: Conceptualisation, U.D., Y.K.S., T.S.M.P., B.K.; Methodology, U.D., Y.K.S., T.S.M.P.; Software, G.Y., T.S.M.P., R.G.K., R.Y.; Validation, Y.K.S., B.K., U.D., T.S.M.P., R.K., S.K.S.; Formal analysis, U.D., Y.K.S., T.S.M.P.; Investigation, U.D.; Resources, Y.K.S., U.D.; Data curation, Y.K.S., U.D., T.S.M.P., R.K., S.K.S.; Writing—original draft preparation, Y.K.S., U.D., T.S.M.P.; Writing— review and editing, U.D., Y.K.S., T.S.M.P., G.Y., B.K.; Visualization, G.Y.; Supervision, U.D., M.K.; Project administration, U.D.; Funding acquisition, U.D., Y.K.S.; Proof Reading, B.K., T.S.M.P. All authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement:

Data	Description	Source
PALSAR (DEM)	Resolution 30 m—Panchromatic Remote-sensing Instrument for Stereo Mapping (PRISM)	https://www.eorc.jaxa.jp/ALOS/en/aw3d30/data/index.htm accessed on 31 July 2020
ASTER (DEM)	Resolution 30 m—Thermal Emission and Reflection Radiometer	https://asterweb.jpl.nasa.gov/gdem.asp accessed on 3 August 2020
SRTM (DEM)	Resolution 90 m & 30 m—Shuttle Radar Topography	https://earthexplorer.usgs.gov accessed on 29 July 2020
LIDAR (DEM)	Resolution 1 m—data acquired in 2005	Central Water Commission (CWC), New Delhi
GFS (Rainfall)	Resolution 0.25 degree, 3 hourly data	ftp://ftp.ncep.noaa.gov/pub/data/nccf/com/gfs/prod/ accessed on 31 August 2020
GPM (Rainfall)	Resolution 0.1 degree, 3 hourly data	https://jsimpsonhttps.pps.eosdis.nasa.gov/imerg/gis/early accessed on 30 July 2020
IMD (Rainfall)	Resolution 4km (WRF) & 25 km (GFS), 3 hourly data	through ftp services
CWC (Rainfall)	Station data, daily data	through email
Discharge	CWC, 3 hourly data	through email
Tide	SOI	https://surveyofindia.gov.in/pages/tidal accessed on 15 August 2020
LULC	Sentinel optical data	https://search.asf.alaska.edu/ accessed on 31 June 2020
SAR	Sentinel microwave data	https://search.asf.alaska.edu/ accessed on 12 September 2020
Vector	Watershed Boundary	CWC
Historical Hydrological data	Discharge, Water level, Cross section, etc	WRIS, CWC https://indiawris.gov.in/wris/#/ accessed on 15 March 2020

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jain, S.K.; Mani, P.; Jain, S.K.; Prakash, P.; Singh, V.P.; Tullos, D.; Kumar, S.; Agarwal, S.P.; Dimri, A.P. A Brief review of flood forecasting techniques and their applications. *JRBM* **2018**, *16*, 329–344. [CrossRef]
2. Miller, C.L.; Laflamme, R.A. *The Digital Terrain Model—Theory & Application*; Photogrammetric Engineering, MIT Photogrammetry Laboratory: Cambridge, MA, USA, 1958; pp. 433–442.
3. Smith, M.J.; Clark, C.D. Methods for the visualization of digital elevation models for landform mapping. *Earth Surf. Process. Landf.* **2005**, *30*, 885–900. [CrossRef]
4. Uysal, M.; Toprak, A.S.; Polat, N. DEM generation with UAV Photogrammetry and accuracy analysis in Sahitler hill. *Measurement* **2015**, *73*, 539–543. [CrossRef]
5. Lopes, V.L. On the effect of uncertainty in spatial distribution of rainfall on catchment modelling. *Catena* **1996**, *28*, 107–119. [CrossRef]
6. Chaplot, V.; Saleh, A.; Jaynes, D. Effect of the accuracy of spatial rainfall information on the modeling of water, sediment, and NO$_3$—N loads at the watershed level. *J. Hydrol.* **2005**, *312*, 223–234. [CrossRef]
7. Beven, K. *Rainfall–Runoff Modelling: The Primer*; Wiley: Chichester, UK, 2001.
8. Kobold, M. *Hydrological Modelling of Floods. Advances in Urban Flood Management*, 1st ed.; Ashley, R., Garvin, S., Pasche, E., Vassilopoulos, A., Zevenbergen, C., Eds.; CRC Press: London, UK, 2007; pp. 433–452.
9. Chintalapudi, S.; Sharif, H.O.; Xie, H. Sensitivity of Distributed Hydrologic Simulations to Ground and Satellite Based Rainfall Products. *Water* **2014**, *6*, 1221–1245. [CrossRef]
10. Wang, L.; Meerveld, I.V.; Seibert, J. When should stream water be sampled to be most informative for event-based, multi-criteria model calibration? *Hydrol. Res.* **2017**, *48*, 1566–1584. [CrossRef]
11. Ferraris, L.; Rudari, R.; Siccardi, F. The uncertainty in the prediction of flash floods in the Northern Mediterranean environment. *J. Hydrometeorol.* **2002**, *3*, 714–727. [CrossRef]
12. Shang, Y.; Guo, Y.; Shang, L.; Ye, Y.; Liu, R.; Wang, G. Processing conversion and parallel control platform: A parallel approach to serial hydrodynamic simulators for complex hydrodynamic simulations. *J. Hydroinform.* **2016**, *18*, 851–866. [CrossRef]
13. Liu, R.; Wei, J.; Ren, Y.; Liu, Q.; Wang, G.; Shao, S.; Tang, S. HydroMP—A computing platform for hydrodynamic simulation based on cloud computing. *J. Hydroinform.* **2017**, *19*, 953–972. [CrossRef]
14. La Loggia, G.; Freni, G. A Parallel Flood Forecasting and Warning Platform Based on HPC Clusters. In Proceedings of the 13th International Conference on Hydroinformatics (HIC 2018), Palermo, Italy, 1 July 2018; Puleo, V., de Marchis, M., Eds.; EPiC Series in Engineering. EasyChair Publications: Manchester, UK, 2018; Volume 3, pp. 1232–1239.
15. Landuyt, L.; Van Wesemael, A.; Schumann, G.J.P.; Hostache, R.; Verhoest, N.E.C.; Van Coillie, F.M.B. Flood mapping based on synthetic aperture radar: An assessment of established approaches. *IEEE Trans. Geosci Remote Sens.* **2019**, *57*, 722–739. [CrossRef]
16. Hess, L.L.; Melack, J.M.; Simonett, D.S. Radar detection of flooding beneath the forest canopy—A review. *Int. J. Remote Sens.* **1990**, *11*, 1313–1325. [CrossRef]
17. Mason, D.C.; Davenport, I.J.; Neal, J.C.; Schumann, G.J.-P.; Bates, P.D. Near real-time flood detection in urban and rural areas using high-resolution synthetic aperture radar images. *IEEE Trans. Geosci. Remote Sens.* **2012**, *50*, 3041–3052. [CrossRef]
18. Inglada, J.; Mercier, G. A new statistical similarity measure for change detection in multitemporal SAR images and its extension to multiscale change analysis. *IEEE Trans. Geosci. Remote Sens.* **2007**, *45*, 1432–1445. [CrossRef]
19. Martinis, S.; Twele, A.; Voigt, S. Towards operational near real-time flood detection using a split-based automatic thresholding procedure on high resolution TerraSAR-X data. *Nat. Hazards Earth Syst. Sci.* **2009**, *9*, 303–314. [CrossRef]
20. Horritt, M.S.; Mason, D.C.; Luckman, A.J. Flood boundary delineation from synthetic aperture radar imagery using a statistical active contour model. *Int. J. Remote Sens.* **2001**, *22*, 2489–2507. [CrossRef]
21. Pradhan, B.; Tehrany, M.S.; Jebur, M.N. A new semiautomated detection mapping of flood extent from TerraSAR-X satellite image using rule-based classification and taguchi optimization techniques. *IEEE Trans. Geosci. Remote Sens.* **2016**, *54*, 4331–4342. [CrossRef]
22. Jagannathan, C.R.; Ratnam, C.; Baishya, N.C.; Dasgupta, U. Geology of the offshore Mahanadi Basin. *Pet. Asia J.* **1983**, *6*, 101–104.
23. Bharali, B.; Rath, S.; Sarma, R. A brief review of Mahanadi Delta and the deltaic sediments in Mahanadi Basin. *Mem. Geol. Soc. India* **1991**, *22*, 31–49.
24. ANUGA Hydro. Available online: https://github.com/GeoscienceAustralia/anuga_core (accessed on 6 September 2021).
25. Oegema, B.W.; McBean, E.A. Uncertainties in flood plain mapping. In *Application of Frequency and Risk in Water Resources*; Springer: Dordrecht, The Netherlands, 1987; Volume 106, pp. 293–303.

26. Pappenberger, F.; Matgen, P.; Beven, K.J.; Henry, J.B.; Pfister, L.; Fraipont, P. Influence of uncertain boundary conditions and model structure on flood inundation predictions. *Adv. Water Resour.* **2006**, *29*, 1430–1449. [CrossRef]

27. Merwade, V.; Olivera, F.; Arabi, M.; Edleman, S. Uncertainty in flood inundation mapping: Current issues and future directions. *J. Hydrol. Eng.* **2008**, *13*, 608–620. [CrossRef]

28. Di Baldassarre, G.; Montanari, A. Uncertainty in river discharge observations: A quantitative analysis. *Hydrol. Earth Syst. Sci.* **2009**, *13*, 913–921. [CrossRef]

29. Horritt, M.S.; Bates, P. Effects of spatial resolution on a raster based model of flood flow. *J. Hydrol.* **2001**, *253*, 239–249. [CrossRef]

30. Wechsler, S.P. Uncertainties associated with digital elevation models for hydrologic applications: A review. *Hydrol. Earth Syst. Sci.* **2007**, *11*, 1481–1500. [CrossRef]

31. Dottori, F.; Di Baldassarre, G.; Todini, E. Detailed data is welcome, but with a pinch of salt: Accuracy, precision, and uncertainty in flood inundation modeling. *Water Resour. Res.* **2013**, *49*, 6079–6085. [CrossRef]

32. Savage, J.T.S.; Bates, P.; Freer, J.; Neal, J.; Aronica, G. When does spatial resolution become spurious in probabilistic flood inundation predictions? *Hydrol. Process.* **2016**, *30*, 2014–2032. [CrossRef]

33. GDAL. GDAL—Geospatial Data Abstraction Library. Available online: https://gdal.org/api/python.html (accessed on 6 September 2021).

34. Arcement, G.J.; Verne, R.S. Guide for Selecting Manning's Roughness Coefficients for Natural Channels and Flood Plains. 1989. Available online: http://dpw.lacounty.gov/lacfcd/wdr/files/WG/041615/Guide%20for%20Selecting%20n-Value.pdf (accessed on 6 September 2021).

35. Barnes, H.H., Jr.; Roughness Characteristics of Natural Streams. US Geological Survey Water Supply Paper. 1849. Available online: https://pubs.usgs.gov/wsp/wsp_1849/pdf/wsp_1849.pdf (accessed on 6 September 2021).

36. Zhang, M.; Chen, F.; Tian, B. Glacial lake detection from gaofen-2 multispectral imagery using an integrated nonlocal active contour approach: A case study of the Altai mountains, northern Xinjiang province. *Water* **2018**, *10*, 455. [CrossRef]