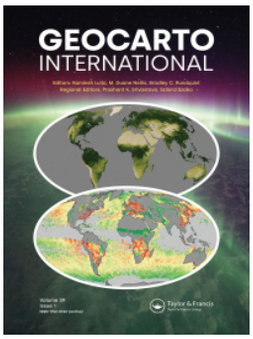

Geocarto International

ISSN: 1010-6049 (Print) 1752-0762 (Online) Journal homepage: www.tandfonline.com/journals/tgei20

Assessment of open-source and proprietary generated digital elevation models and building footprints for urban flood modelling in Cartagena De Indias, Colombia

Ambreen Masud , Maria Valasia Peppa , Yady Tatiana Solano-Correa , Jon Philip Mills & Cat Button

To cite this article: Ambreen Masud , Maria Valasia Peppa , Yady Tatiana Solano-Correa , Jon Philip Mills & Cat Button (2026) Assessment of open-source and proprietary generated digital elevation models and building footprints for urban flood modelling in Cartagena De Indias, Colombia, Geocarto International, 41:1, 2681279, DOI: 10.1080/10106049.2026.2681279

To link to this article: https://doi.org/10.1080/10106049.2026.2681279

GEOCARTO INTERNATIONAL
2026, VOL. 41, NO. 1, 2681279
https://doi.org/10.1080/10106049.2026.2681279

Taylor & Francis
by informa

RESEARCH ARTICLE

OPEN ACCESS Check for updates

Assessment of open-source and proprietary generated digital elevation models and building footprints for urban flood modelling in Cartagena De Indias, Colombia

Ambreen Masud[a], Maria Valasia Peppa[a], Yady Tatiana Solano-Correa[b], Jon Philip Mills[a] and Cat Button[c]

[a]Newcastle University, Department of Civil and Geospatial Engineering, Newcastle upon Tyne, UK; [b]Universidad Tecnológica de Bolívar, Escuela de Transformación Digital, Campus Tecnológico, Cartagena de Indias, Colombia; [c]Newcastle University, School of Architecture, Planning and Landscape, Newcastle upon Tyne, UK

ABSTRACT

Urban hydrological models are critical for flood risk management. However, the availability of high-resolution topographic data for reliable outputs remains challenging in data-scarce cities. Therefore, determining data quality in producing reliable information is fundamental. We evaluated open-source and proprietary topographic data for use in the two-dimensional hydrological model City Catchment Analysis Tool (CityCAT). We modelled 12 scenarios using combinations of open-source and light detection and ranging (LiDAR)-generated datasets, validating the results with community-generated flood risk maps. The findings show high agreement with scenarios using LiDAR-derived digital elevation models (DEMs) (bootstrapped Spearman's $\rho \approx 0.90$). However, open-source building footprints performed better, demonstrating that both are necessary for reliable urban flood risk mapping. As LiDAR is costly with limited access, we urge for publicly available high-resolution datasets for low- and middle-income countries (LMICs) disproportionately impacted by climate change. Therefore, we address this gap by focusing on a Latin American context with Cartagena de Indias (Colombia) as a case study.

ARTICLE HISTORY
Received 4 December 2025
Accepted 21 May 2026

KEYWORDS
Urban; LiDAR; flood risk; flood model; data-scarce

1. Introduction

Climate-induced precipitation patterns and extreme rainfall events have caused pluvial flooding to become commonplace globally, with dense urban areas vulnerable to changing weather patterns, leading to substantial economic and physical damage, and loss of human life (Kundzewicz et al. 2014; Zscheischler et al. 2018). Flooding is the most common and consistent natural threat in cities, and is projected to intensify and increase over the next decade (UN Habitat 2022). Thus, implementing appropriate flood preparedness measures is critical for managing and mitigating flood impacts (Etinay et al. 2018). Addressing this requires a principal understanding of risk using quality information, which would aid policy makers and disaster relief agencies in developing flood management plans for creating more resilient communities and cities (UNDRR 2015).

Computational flood models play a key role in providing flood information through hazard maps, risk quantification, vulnerability assessments and forecasting of future events (Prashar et al. 2023). A crucial component of model usage is topographic and elevation data. However, model accuracy and reliability are dependent on the quality of input data (Cea and Costabile 2022). Whilst high-income countries continue developing their data resources and technological capabilities, many low- and middle-income countries (LMICs) still lack up-to-date information (Sibandze et al. 2025). This hinders flood management efforts for rapidly developing cities with increasing populations (Löwe et al. 2017).

Open-source global spatial datasets such as Open Street Map (OSM) serve as proxies for alternative data sources when government data is limited or unavailable, particularly during disaster emergencies (Bozza et al. 2016). The availability of free satellite imagery, 30 m Digital Elevation Model (DEM)s such as NASA's

CONTACT Ambreen Masud a.masud2@newcastle.ac.uk
Supplemental data for this article can be accessed online at https://doi.org/10.1080/10106049.2026.2681279.

Shuttle Radar Topography Mission (SRTM), and building footprints generated with Artificial Intelligence (AI) create opportunities to use open-source topographic data in evidencing flood hazards, addressing relief efforts in data-scarce cities (Thakuri et al. 2022; Schumann 2024). Despite this, the spatial and temporal resolutions of open-source datasets are unable to capture the required detail for flood mapping in urban topographies. Moreover, data completeness in global datasets remains uneven, with some locations being better represented than others (Azizian and Brocca 2020; Muthusamy et al. 2021), thus requiring careful assessment over local contexts to determine reliable flood hazard information (Bernhofen et al. 2022; Gevaert et al. 2024).

Airborne light detection and ranging (LiDAR) sensors offer more precise data, generating 3D point information with X, Y planimetric coordinates and Z elevation (Muhadi et al. 2020). LiDAR is a valuable resource for environmental applications such as flooding because of its capability to generate high-resolution topographic data, from 1 m spatial resolution DEMs (Walker et al. 2025), to centimetre-scale features such as buildings, roads and vegetation (Hung et al. 2018; Yoshida et al. 2023). However, use cases of LiDAR are most often found in high-income countries. Whilst increasingly available, LiDAR data is still costly and less available in countries most vulnerable to disaster risk (Nkwunonwo et al. 2020; Abdelmoneim et al. 2023). The use of LiDAR for flood risk research in Latin American contexts is still disproportionately under-represented. However, flood research using LiDAR is steadily increasing in Brazil and Mexico (Pinos and Quesada-Román 2021).

This research addresses the need for improved data for flood preparedness in generating resilient cities, focusing on a Latin American context in Cartagena de Indias, Colombia, to narrow the knowledge gap. Cartagena experiences recurring flooding, partly due to its geography on the Caribbean coastline, yet increasingly due to climate change and dense urbanisation (OECD 2022). More frequent and intense precipitation is exacerbating flood risk, causing devastation and disruption to lives and livelihoods each year. Whilst flood management strategies are urgently required, these are undermined by outdated data, which hinders the understanding of flood hazards and vulnerability, leaving the city and its citizens inadequately prepared for flooding.

Therefore, this research aims to assess the quality of existing open-source geospatial data and compare this with proprietary geospatial data in producing reliable and updated flood exposure information to support communities, local policy makers and disaster relief agencies. To achieve our aim, we evaluated and prepared existing open-source global and local datasets for use in the 2D hydrodynamic model CityCAT, and then created a DEM and building footprints using airborne LiDAR for use as model data inputs. We simulated multiple scenarios using combinations of data and then validated the scenarios against both community- and government-generated flood risk maps. We then evaluated all other scenarios against a baseline scenario best aligned with the community flood risk maps. Using the root mean square error (RMSE), mean absolute error (MAE) and bootstrapped Spearman's rank correlation coefficient, we quantified the differences between each scenario to determine which datasets improved flood risk information in Cartagena.

2. Study area: Cartagena de Indias, Colombia

Located to the North-West of Colombia on the Caribbean coast (Figure 1), Cartagena de Indias is vulnerable to climate change, facing rising sea levels and variable and increasing precipitation patterns that impact the Caribbean region (OECD 2022). Cartagena is prone to annual flooding during heavy rainfall months in the wet season, typically between April–June and August–November, the severity of which is impacted by the El Niño Southern Oscillation (ENSO) seasonal weather patterns (Calil et al. 2017). With an estimated 2024 urban population of 903,895 inhabitants in an area of 82 km^2 (DANE 2025), a complex network of canal channels and antiquated drainage systems, and high levels of poverty and socio-economic disparities (Espinosa et al. 2020), Cartagena faces many challenges in adapting to changing weather patterns. Figure S1 (Supplementary material) demonstrates daily recorded rainfall in Cartagena, showing greater rainfall spread throughout the year in recent years. Between 2014 and 2024, Cartagena saw 110 days of over 30 mm of daily rainfall, with a maximum daily rainfall of 222 mm. Figure 2 demonstrates the levels of flooding that local residents can face.

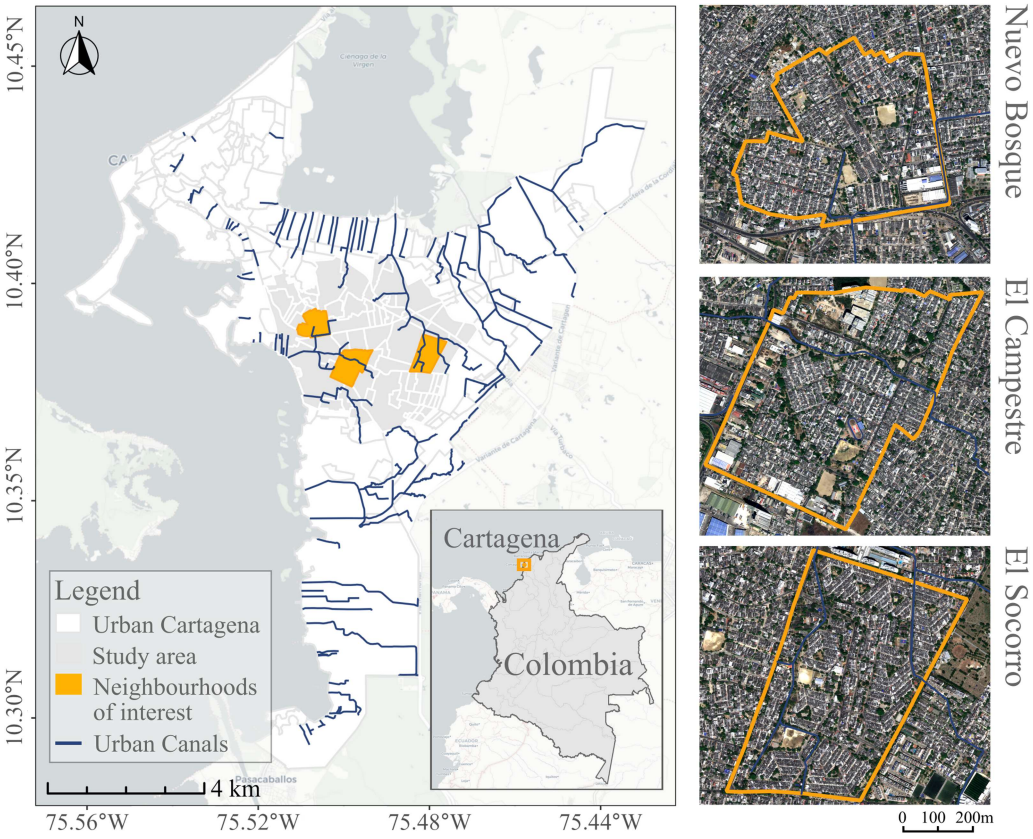

Figure 1. Study area location, highlighting neighbourhoods whose residents were research stakeholders. Satellite imagery ©Airbus Defence and Space (2024).

Figure 2. Images of flooding captured by local residents of Cartagena in October 2023.

Renewed urbanisation plans under the current administration demonstrate a drive for proactive flood preparedness (Secretaría de Planeación 2024). However, risk reduction strategies are undermined by political, social and economic challenges (Cohen et al. 2025). Tourism is a key economic driver. However, it also contributes to spatial disparities between touristic and economically advantaged zones in the west of

the city where there is concentrated development, and the 'rest of the city', where there is greater socio-economic disadvantage. Irrespective, all zones and populations are vulnerable to recurrent flooding, disrupting daily life, damaging property and risking people's lives. Yet, existing flood information is heavily reliant on news reports post-event, lacking adequate flood preparedness information before and during events.

Whilst lack of adequate flood information is a problem throughout the city, areas outside of Cartagena's touristic zones are under-represented. Therefore, to aid inclusion and improve local preparedness information, this research focuses on neighbourhoods in the south of the city, outside of Cartagena's touristic zones. The overall area examined is 15.6 km^2, focusing on the neighbourhoods of Nuevo Bosque, El Campestre and El Socorro, which are highlighted in Figure 1. El Campestre and El Socorro frequently appear in news reports declaring emergencies following heavy rainfall events (El Universal 2022; El Espectador 2023; El Heraldo 2024), and whilst they are more prone to flooding, Nuevo Bosque also experiences flooding to the south of the neighbourhood. All three neighbourhoods are densely populated (Table 1), have canals in or bordering them and have drainage issues. Local community leaders and local residents of these neighbourhoods were research stakeholders, all of whom urge for greater institutional support to address the problem of flooding.

3. Methodology

To determine the availability of existing datasets for flood model inputs, the proposed workflow included an initial audit of available open-source data completeness. This included assessing global, national and local government datasets. The topographic information required over Cartagena was rainfall, buildings, elevation and green areas, as these were critical data inputs for CityCAT, the two-dimensional hydro-dynamic urban flood model used in this research (Glenis et al. 2013). Table 2 demonstrates the datasets audited in this research. We also generated additional DEM and building data, with a spatial resolution of between 1 and 4 m, with privately obtained LiDAR data (Cuatro Conceptos 2024). The methodological approach is illustrated in Figure 3.

This study simulated 12 scenarios for 5th November 2022, using a combination of different processed DEM, rainfall, building and green area datasets. This was the heaviest rainfall day from rain gauge data during a period of national flood emergency, as declared by the Colombia National Unit for Disaster Risk Management (UNGRD) (World Health Organization 2023). Numerous flood-related emergencies occurred during this period, affecting more than 37 neighbourhoods and 12,000 people (ReliefWeb 2022; Semana 2022). Comparing the sensitivity of data inputs is necessary in determining the reliability of the data outputs, as it can determine whether at-risk places and communities are included or excluded in flood risk management strategies.

Previous studies have demonstrated the value of community co-creation to capture lived flood experiences through observational data, which in turn includes communities and aids in the credibility of modelled outputs in data-scarce areas, thus advancing local hydrological knowledge (Assumpção et al. 2018; See 2019; Mandal et al. 2026). This research builds on similar work by comparing modelled scenario outputs with community flood risk maps created with community members living in the local neighbourhoods of El Socorro, Nuevo Bosque and El Campestre during social cartography sessions held in 2023. These were also compared with local government flood risk maps, which, whilst obtained from the Secretaría de Planeacíon in 2023, were generated much earlier, though the exact date is unknown owing to inconsistent metadata.

Each simulated scenario was given a label made up of the input data used whilst processing that scenario. Figure 4 assists with understanding the flood model outputs in Figure 6. A label is provided for

Table 1. Information on study neighbourhoods (Secretaría de Planeación 2023).

Neighbourhood	Neighbourhood area size (ha)	Population size (2018 Census)	Population density per hectare (ha)
El Campestre	59	10,306	174
El Socorro	52	10,609	204
Nuevo Bosque	43	8851	205

Table 2. Information on the datasets audited in this research.

Data type	Dataset	Spatial resolution, data type, data size	Access type	Source
DEM	Advanced Land Observing Satellite Phased Array Type L-band Synthetic Aperture Radar (ALOS PALSAR)	12.5 m, Raster, 69.8 MB	Open	JAXA (2014)
	ALOS World 3D – 30	30 m, Raster, 123 KB	Open	Takaku et al. (2014)
	NASADEM	30 m, Raster, 123 KB	Open	NASA JPL (2020)
	Government	12 m, Raster, 7.3 MB	Open	Secretaría de Planeación (2023)
	Airborne LiDAR	≈13 pts/m^2, 3D point cloud, 15.9 GB	Restricted	Cuatro Conceptos (2024)
Buildings	OSM Buildings	-, Vector, 3.78 MB	Open	OpenStreetMap (2024)
	Microsoft Global Building Footprints	-, Vector generated from Bing satellite imagery, 9.3 MB	Open	Microsoft (2023)
	Google Open Buildings	-, Vector generated from 50 cm satellite imagery, 32.0 MB	Open	Google (2024)
	Government Urban Buildings	-, Vector, 35.4 MB	Open	Secretaría de Planeación (2023)
	Airborne LiDAR	≈13 pts/m^2, 3D point cloud, 16 GB	Restricted	Cuatro Conceptos (2024)
Green spaces	OSM	-, Vector, 1.58 MB	Open	OpenStreetMap (2024)
	Government Green Zones	-, Vector, 0.8 MB	Open	Alcaldia Mayor de Cartagena de Indias (2024)
Rainfall	Daily Precipitation Rafael Núñez International Airport Weather Station	-, Tabular (CSV), 2.01 MB	Open	IDEAM (2025)

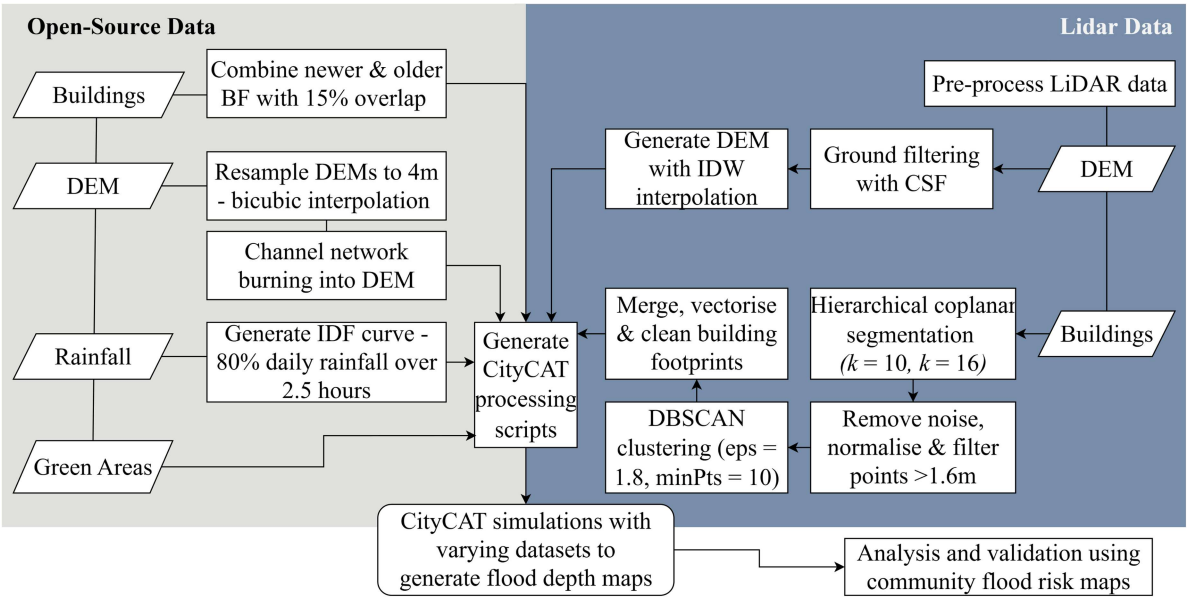

Figure 3. Workflow detailing the open-source (grey) and LiDAR (blue) methodological approach taken to process data for use in CityCAT flood simulations.

each scenario. This label is made up of a scenario number and corresponding codes that signify the data type and input used. For example, in *S2: B-Gov/D-AW3D/G-Gov*, S2 refers to scenario 2, *B-* refers to building data, and *Gov* signifies the use of local government building footprints (BF). This is the same as for the other codes in the label. For example, *D-* refers to DEM, and *AW3D* signifies the use of ALOS World 3D. Across all scenarios, the data inputs for green areas are the same (*G-Gov*).

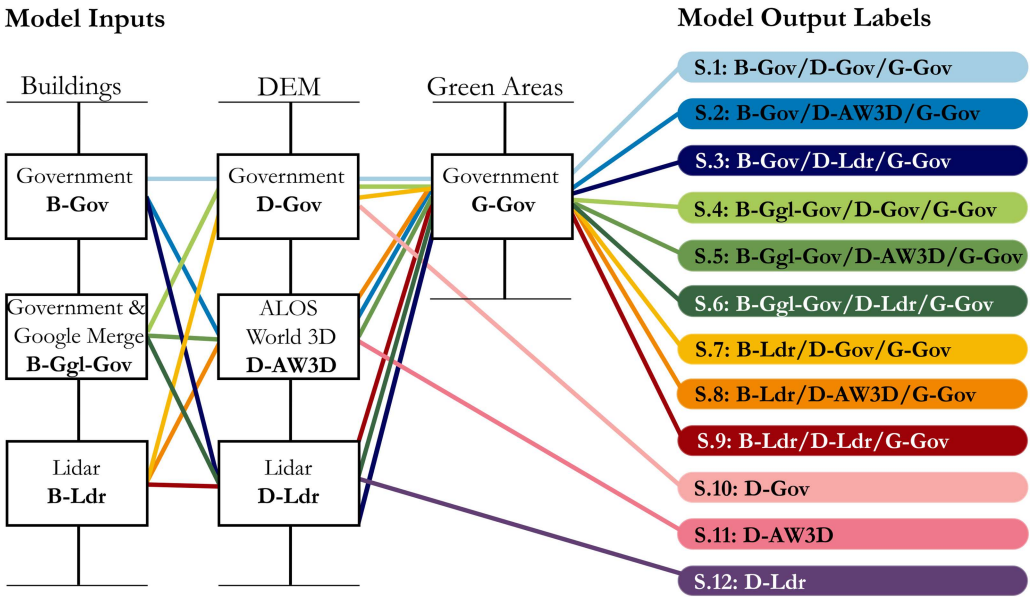

Figure 4. Schematic showing scenario labels to assist in comprehending Figure 6.

3.1. CityCAT flood model and data inputs

City Catchment Analysis Tool (CityCAT) is a physics-based two-dimensional hydrodynamic urban flood model that simulates water flow and depth using shallow water equations, obtaining a solution using finite volume methods with shock-capturing schemes (Glenis et al. 2013). This captures flood wave propagation and handles wetting and drying processes of the urban terrain to obtain flood depth (Toro 2013). The model generates a computational grid using a DEM and polygon-based vector layers of urban features such as buildings and green areas. The building footprints are treated as solid exteriors using the 'building hole' approach and excluded from the grid to produce realistic overland flow paths, whilst green areas are used to calculate infiltration using an iterative Green-Ampt method (Glenis, Kutija, and Kilsby 2018). Manning's roughness coefficients were assigned using the model's standard values, which automatically differentiate values between permeable green areas and impermeable surfaces. This introduces uncertainty, as local roughness measurements were not available. Furthermore, whilst CityCAT can account for drainage, obtaining such data is challenging in many contexts, especially in data-scarce urban environments such as Cartagena (Bertsch et al. 2022).

Despite these uncertainties, we used CityCAT for its advancement in representing water flows after high-rainfall events in dense urban environments, efficiently capturing buildings and natural drainage systems with a low computational load. Other two-dimensional flood models, such as HEC-RAS, could also be applied to this analysis (Brunner et al. 2015); however, we chose to take advantage of institutional links with CityCAT developers for any technical support. Furthermore, whilst CityCAT use cases have been researched in Europe and the USA (Guerreiro et al. 2017; Iliadis et al. 2023), less research has occurred in LMIC contexts where access to complete or high-resolution data is often limited. We sought to broaden model use, whilst applying more advanced techniques to a Latin American context.

3.2. Flood modelling with open-source data

3.2.1. Rainfall

Cartagena experiences variable local rainfall patterns, yet obtaining rain gauge data is difficult, as gauges are not spatially distributed throughout the city. Daily precipitation records are available from the National Institute of Hydrology and Metrology, Instituto de Hidrología, Meterología y Estudios Ambientales (IDEAM), and their rain gauge is located at the Rafael Núñez Airport in the North-West of Cartagena (IDEAM 2025). This area falls outside the study area, therefore does not always capture short-duration

intense localised rainfall, which is typical in Cartagena (Mouthon-Bello et al. 2022). This introduces uncertainty in the model; however, available precipitation data from a nearby weather station were preferred as an indicator of rainfall distribution over simulated global atmospheric datasets and are a consistent constraint in research relying on rain gauge data in Cartagena (Gonzalez-Alvarez et al. 2018).

We adapted daily rainfall records in the flood model simulations based on research by Mouthon-Bello et al. (2022) and Castilla Villalba et al. (2023), who found that 80% of daily rain falls within the first 2.5 h of a rainfall event in Cartagena. On 5th November 2022, the Rafael Núñez Airport rain gauge recorded 128.1 mm of daily precipitation. We adapted this method to create an intensity–duration–frequency (IDF) curve, generating a storm profile of 80% of rainfall depth (102 mm) over a 2.5-h time period, using scale invariance to downscale the rainfall intensity over 5-min intervals. To address rainfall uncertainty, we performed a sensitivity analysis by modifying the adapted rainfall amount by ±20%. We found that different rainfall amounts vary in flood extent and depth; however, spatial flood path patterns remain consistent (Figure S2, supplementary material). Whilst this does not resolve the lack of localised rain gauge data, it demonstrates model consistency.

3.2.2. Digital elevation models

For improved precision in high-resolution urban flood models such as CityCAT, DEMs with a spatial resolution of 5 m or less are required, as larger cell sizes misrepresent flow between buildings (Iliadis et al. 2023). We tested different open-source DEMs, assessing 12.5 m spatial resolution Advanced Land Observing Satellite Phased Array Type L-band Synthetic Aperture Radar (ALOS PALSAR) (JAXA 2014), 30 m NASADEM (NASA JPL 2020) and 30 m ALOS World 3D (AW3D30) (Takaku et al. 2014). Among the open-source DEMs evaluated, AW3D30 performed best and was therefore selected for further use in model simulations (see Figure S3 supplementary material for results demonstrating simulated DEM differences). A 12-m DEM was also obtained from the local mayor's planning office (Secretaría de Planeación 2023). Owing to missing metadata, it was difficult to determine the government DEM origins; however, it was included within this study as an official dataset.

To process the DEMs for use in CityCAT and compare open-source DEMs against the LiDAR-generated 4 m DEM, we increased the spatial resolution of all open-source DEMs to 4 m using polynomial bicubic interpolation. We acknowledge that resampling a DEM has the potential to introduce large errors due to the loss of topographic information, and without panchromatic data, upsampling does not increase spatial detail (Wechsler 2007). However, we wanted to assess the usefulness of available open-source DEMs in high-resolution flood models when requiring critical flood risk information in cities with limited data availability. The ALOS PALSAR DEM was initially upscaled to 20 m and then downscaled to 4 m to aid in reducing noise and artefacts. We then burnt the channel network into all open-source DEMs using channel vector data, approximating measurements reported by Naufal et al. (2024) due to a lack of data. As CityCAT addresses buildings within the computational grid, they were not filtered at this stage.

3.2.3. Buildings

Using accurate and up-to-date BF is a crucial element as missing buildings in flood simulations are unlikely to characterize flood risk in those areas. The government BF was made available by the local government planning office (Secretaría de Planeación 2023). These are illustrated in Figure 5 (map a). Owing to missing metadata, it is unknown when the government BF were created. We analyzed the dataset against satellite imagery acquired in 2024, finding that whilst there was overall coverage spatially distributed across the city, there were mismatching BF with the satellite image. The missing footprints are likely due to different temporal information between the BF and imagery, and the city's rapid expansion.

Open-source alternatives were available; however, they required further inspection to determine applicability. Volunteer-generated map OSM generated large amounts of missing data over the study area when compared with government BF (Figure 5, map b), which is a common issue across LMICs (Herfort et al. 2023; OpenStreetMap 2024). AI-generated datasets by Microsoft and Google (with a 75% confidence level) were also inspected (Microsoft 2023; Google 2024). Microsoft's BF were large in area size and not reflective of buildings in Cartagena. Whilst average building area size and coverage were largely improved with Google's BF, missing BF remained (Figure 5, maps c-d).

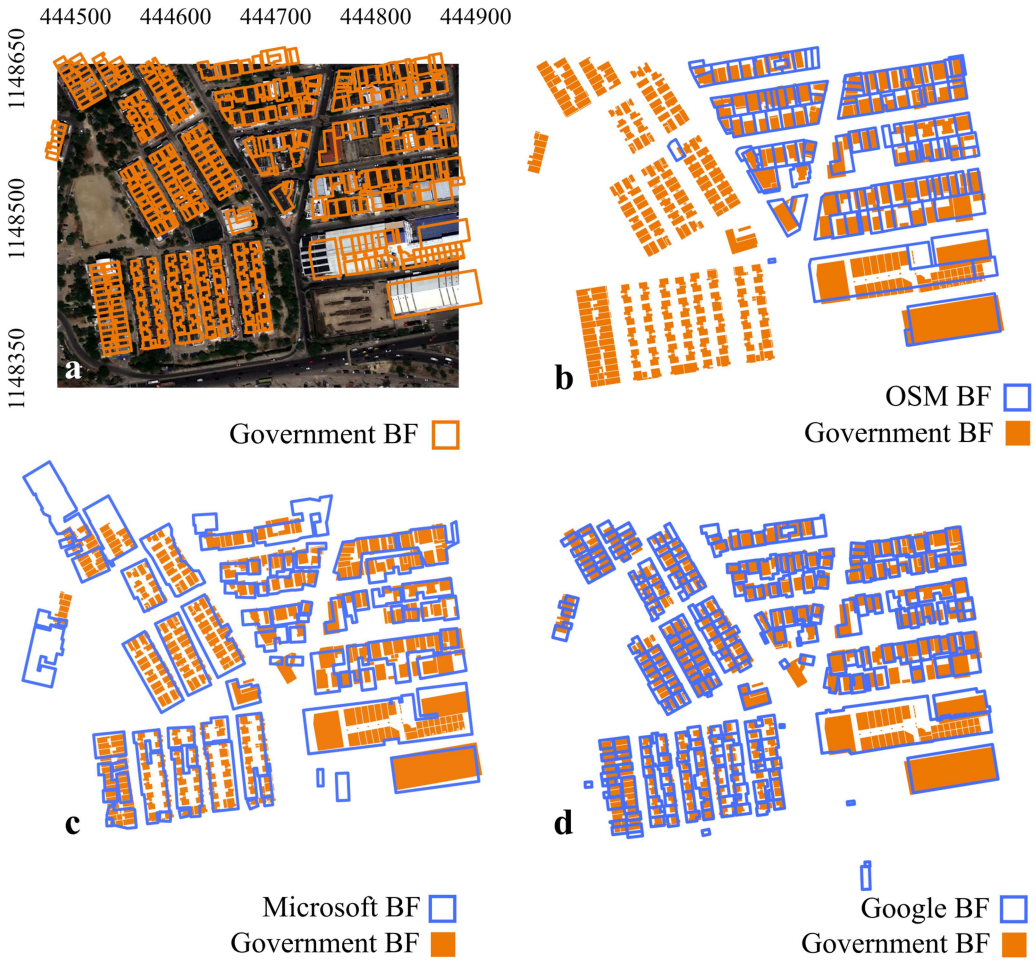

Figure 5. Comparison of open-source building footprints available over Cartagena, using satellite imagery as a baseline of comparison.

We determined that the government BF, despite not being the most up-to-date, was more credible as an official dataset, though Google BF was useful as a proxy for missing data. Thus, merging both datasets was one method of overcoming data incompleteness with an accurate building size. However, despite both being reprojected to the WGS 84/UTM zone 18N (Northern Hemisphere) coordinate reference system, we found a systematic spatial offset, with median residuals of 2.3 m. This is likely due to the use of centimetre-resolution satellite imagery in the processing of the Google BF, as well as the different methods in producing each dataset (Sirko et al. 2021).

To merge the datasets and create a more up-to-date BF, we first examined the offset with an overlap analysis. This was to determine more recently developed BF in the Google dataset, which are missing from the government BF. There was a median overlap of 72% between the BF of both datasets, with an upper quartile of 83%. We then compared BFs with less than 28% overlap against Google satellite imagery and found that buildings in the Google BF with less than 15% overlap only overlapped on the corners of buildings in the government dataset. These residual buildings with slightly overlapping corners were merged with the government BF, creating a new dataset of updated BF in Cartagena. This dataset was then used in CityCAT. We also used the government BF in simulations to compare model outputs on each BF dataset.

3.2.4. Green areas

Data on green areas was found on the OSM and the local government's geospatial portal, MIDAS (Alcaldia Mayor de Cartagena de Indias 2024). Observations of satellite imagery determined gaps between the two

datasets. The OSM included newer green spaces that were not observed in MIDAS, but the OSM data were incomplete. As Cartagena is densely urbanised with few green spaces, the use of the MIDAS dataset was deemed sufficient, as any buildings identified in green areas would be prioritised by the model.

3.3. Generating data inputs with LiDAR

Airborne point cloud data was acquired from the Colombian-based company Cuatro Conceptos (Cuatro Conceptos 2024). The point cloud covered a 45.84 km^2 area over Cartagena and was generated on 17th and 18th June 2019. The dataset was large, with an average density of 13 points per m^2 and 23 LAS files. LiDAR processing was carried out in 2 distinct stages: firstly, to generate elevation data and then to extract BFs, using LasTools in QGIS 3.4.1 and LidR in RStudio 4.2.1 (Roussel et al. 2020).

3.3.1. Data cleaning and pre-processing

The LiDAR dataset was provided raw, therefore required pre-processing, including:

- reprojecting files to the WGS 84/UTM zone 18 N coordinate reference system;
- X and Y scale transformation;
- unclassifying rogue points to 0.

The data also had large height (Z) outliers, displaying values from −113 to 533 m. To reduce these, we analyzed point density and neighbouring points to determine parameters for a statistical outlier removal (SOR) noise removal algorithm, determining 15 nearest neighbors (k), with a maximum distance threshold of 2 standard deviations, as suitable for this dataset. This reduced outliers, but further refinement was required, as the height values still demonstrated outliers. Owing to Cartagena's location as a coastal city and the prevalence of channels, Z values below 0 were expected. Further inspection of minus values showed a median average of −7 m. High-value outliers were also investigated, often displaying cranes and power lines. These were separated from high-rise buildings and removed. After cleaning the data, minimum and maximum height values were set to −7 and 115 m, respectively.

3.3.2. LiDAR-generated DEM

A 4 m resolution DEM was derived from the point cloud data for use in CityCAT. The ground points were classified using the cloth simulation filtering (CSF) algorithm, which inverts the point cloud so that the ground is facing up, placing a cloth over this to determine nodes identified as ground (Zhang et al. 2016). It works well in flat terrains; therefore, it is suited to the topography of Cartagena. Table S1 (supplementary material) displays the default CSF parameters, which were tested and then fine-tuned, evaluating against satellite imagery to reduce misclassification. The DEM was then derived using the ground points and interpolated using a K-nearest neighbor classification with inverse distance weighting (IDW).

The algorithm determines the nearest points (k) within a given radius of unsampled locations in the point cloud data, allocating weights to each point by calculating the inverse distance raised to a given power (p). Whilst IDW is a commonly used algorithm and regarded for its simplicity and computational speed, there is also a critique of accuracy (Bater and Coops 2009). However, Guo et al. (2010) found that IDW performs well when creating 5–10 m spatial resolution DEM's, whereas the use of IDW at higher spatial resolutions (1 m or less) produces more inaccuracies. Nearest neighbor algorithms have also demonstrated strong performance in urban areas, so combining the traditional IDW with a K-nearest neighbor (KNN) classifier can improve ground point detection (Polat et al. 2015). Sensitivity analyses were performed to determine suitable values of k (10) and p (2), before formatting the DEM to ascii format ready for use in CityCAT.

3.3.3. LiDAR-generated buildings

The provided LiDAR data was also used to experiment with generating BF to overcome missing data. Whilst there have been significant advances in LiDAR segmentation, differentiating between urban features within close proximity, such as trees, cars and buildings, remains a challenging task (Karsli et al. 2024). Cartagena exemplifies this with trees of varying heights tightly woven throughout the city's

urban fabric. We address this using a coplanar approach, whereby neighbouring points on the same two-dimensional plane are segmented and identified as rooftops, and then use the density-based clustering algorithm density-based spatial clustering of applications with noise (DBSCAN), as proposed by Ester et al. (1996) to identify individual buildings.

We pre-processed the non-ground points identified in Section 3.3.2, filtering points above 0 to remove outliers. The point cloud was then normalized by subtracting a 1 m DEM for coarse-resolution building footprints. Elevation values ≤1.6 m were removed to eliminate artefacts such as cars and low-level vegetation (Ma 2005). Duplicated points from overlapping tiles were also removed before identifying building shapes.

The coplanar algorithm in the LidR package determines clusters of dominant planes using covariance eigenvalue tests and then uses weighted Gaussian kernels in a Hough spherical accumulator (Limberger and Oliveira 2015; Roussel et al. 2020). The algorithm was run twice on the original dataset because of variable densities in areas of overlapping tiles. Between 4 and 25 k nearest neighbours were tested, determining an initial run of 10 nearest neighbours to identify smaller buildings, and a second run of 16 nearest neighbours to identify larger buildings. The remaining isolated points were removed using conservative SOR parameters ($k = 2$, $m = 4$).

Coplanar detection of isolated building points was performed, but did not identify individual buildings. For this purpose, DBSCAN was used with values of 1.8 ε and 10 (minPts). A range of parameters was determined using a shoulder method and then tested against satellite imagery, which is critical, as DBSCAN can be highly sensitive (Schubert et al. 2017). Clusters identified as noise were removed, and the remaining clusters were polygonized. DBSCAN was used on both coplanar datasets before they were fused together. To create a final vector BF dataset, post-processing included filling gaps within the BF, removing polygons with an area size less than 20 m^2, as these are unlikely to be buildings, and deleting polygons which intersected with roads. Manual cleaning was also required. To improve the linearity of the building shape, buildings were simplified and orthogonalized in the QGIS. These were then used for processing in CityCAT.

3.4. Quantifying scenario differences and uncertainties

In the absence of quantifiable flood maps for validation, no one scenario was deemed accurate. Instead, the scenario that best aligned with the community flood risk maps was used as a baseline scenario to compare all other scenarios and understand the differences between them. We used the mean absolute error (MAE) to measure the average absolute differences between all scenarios against the baseline, treating values as a comparative interpretation rather than absolute differences, and explored this further with the root mean square error (RMSE).

Furthermore, bootstrapping is well used in hydrological studies as a statistical technique to measure uncertainty, particularly in the absence of ground-truth data (Hirsch et al. 2015; Faghih et al. 2017; Couasnon et al. 2020; Landwehr et al. 2024). Bootstrapping uses random samples of observed data, never using the same sample twice, calculating uncertainty for each sample and then measuring the average across all samples (Tibshirani and Efron 1993). In this study, we used bootstrapping of 1000 samples with Spearman's rank correlation coefficient to measure spatial uncertainty between the baseline scenario and all other scenarios.

4. Results and discussion

We assessed the influence of open-source and proprietary geospatial data in producing improved flood exposure information using a two-dimensional hydrodynamic urban flood model in a data-scarce context. Through a comparative analysis with community flood risk maps, model outputs demonstrate varying results with greater spatial flood extent differences when using open-source global DEMs. The use of a LiDAR-derived DEM captured improved flood information at a local scale. These differences are due to the native spatial resolution and micro-topographic detail of each dataset. Despite this, AW3D demonstrated that pre-processing the DEM through burning the channels network can also improve the simulated flows. When assessing BF, the LiDAR-generated BF distort flood estimates, as buildings are

too large, but the open-source BF provides flood information in between the buildings, which is necessary in dense urban areas. Using a combination of open source and proprietary data can work well in the absence of official data, but auditing existing information and assessing data quality is critical to generating improved flood maps and making informed decisions, particularly in data-scarce contexts.

4.1. Flood model scenario maps

The 12 generated flood scenario maps are displayed in Figure 6. Each scenario in each row in Figure 6 uses the same building data input during simulation. For example, row 1 shows S1, S2 and S3, which used *B-Gov*, indicating government buildings. Each figure column used the same DEM. For example, column 1 shows S1, S4, S7 and S10, which use *D-Gov*, indicating the government DEM.

Scenarios that use the LiDAR DEM demonstrate natural flows of floodwater, as the largest flood depths are concentrated along existing channels. This is illustrated best in S10 (Gov DEM), S11 (AW3D) and S12 (LiDAR DEM), where only a DEM was simulated to determine whether elevation data in the absence of building data could provide reliable information in data-scarce contexts. Scenarios using AW3D also demonstrate floodwater representation along channels. However, some flood zones identified by local communities are not captured, and sections of the channel are missing flooded areas. This is likely a limitation of the vector-based channel burning in the DEM. Obtaining accurate channel measurements could improve this, but the data is not publicly available. Although, AW3D performed better than scenarios using the government DEM, which was inconsistent with the results of the community reports. The LiDAR-generated DEM captured more variation in topography, coarse urban features and the channel network. Figure 7 demonstrates this, highlighting the loss of topography when resampling the open-source DEMs. This reinforces the trade-off between the spatial resolution available and the need for fine-scale detail when capturing flood risk in complex urban terrain.

Across scenarios using the same DEM (Figure 6, Columns 1, 2 and 3), patterns of spatial variations of flooding are similar. However, the scenarios without buildings (S10, S11 and S12) show an average spatial extent of 15.63 km, which is approximately 39% greater than that when buildings are included, as demonstrated in the statistical analysis in Table 3. Overall flood extents in S1 to S9, which include buildings show between 9 and 11 km^2 of flood extent. This highlights that whilst the LiDAR DEM can demonstrate spatial flood patterns, DEMs alone are insufficient in capturing flood risk dynamics.

Figure 8 shows the flood depth distribution with maximum depth values displayed above each scenario, corresponding with the colour-coded scenarios in Figures 4 and 6. In Figure 8, we see similar patterns of flood depth distribution in S3, S6, S9 and S12 (LiDAR DEM), compared to all other scenarios. The embedded box plots demonstrate that across all scenarios, the majority of the flooded cells were low risk (<30 cm). However, the maximum flood depths ranged between 3.78 and 4.86 m across the scenarios, demonstrating that dangerously high flood risk was concentrated in specific areas.

In Table 3, we used the median absolute deviation (MAD) and interquartile range (IQR) to understand the dispersion of flood depth data. We found that the scenarios with building data (S1–S9) demonstrated greater flood depth values than scenarios without buildings, with medium- to high-risk flood pixels (>30 cm) forming between 11% and 14% of the overall spatial extent. Adding buildings in the flood model simulations showed higher flood depths in the 90th percentile, suggesting that buildings intensify the accumulation of flooding, increasing flood depth.

4.1.1. Neighbourhood level flood risk

Scenarios were further explored at the neighbourhood level to determine the influence of data inputs alongside government and community flood risk maps. Figure 9 illustrates this with scenarios S1–S3 (Gov BF) in the neighbourhood of Nuevo Bosque. There is a misalignment between the government flood risk map and flood risk as reported by local communities. Of the generated flood maps, S3 (LiDAR DEM) demonstrated flood patterns that were more closely aligned with community perceptions of flood risk, particularly in the South and centre of Nuevo Bosque. However, some areas deemed as medium to high risk by communities were not evident in S3. This is in part due to incomplete government data with missing building footprints in scenarios S1–S3. When alternative BF were used, high flood risk pixels were modelled as demonstrated in Figure 10, highlighting the importance of up-to-date BF.

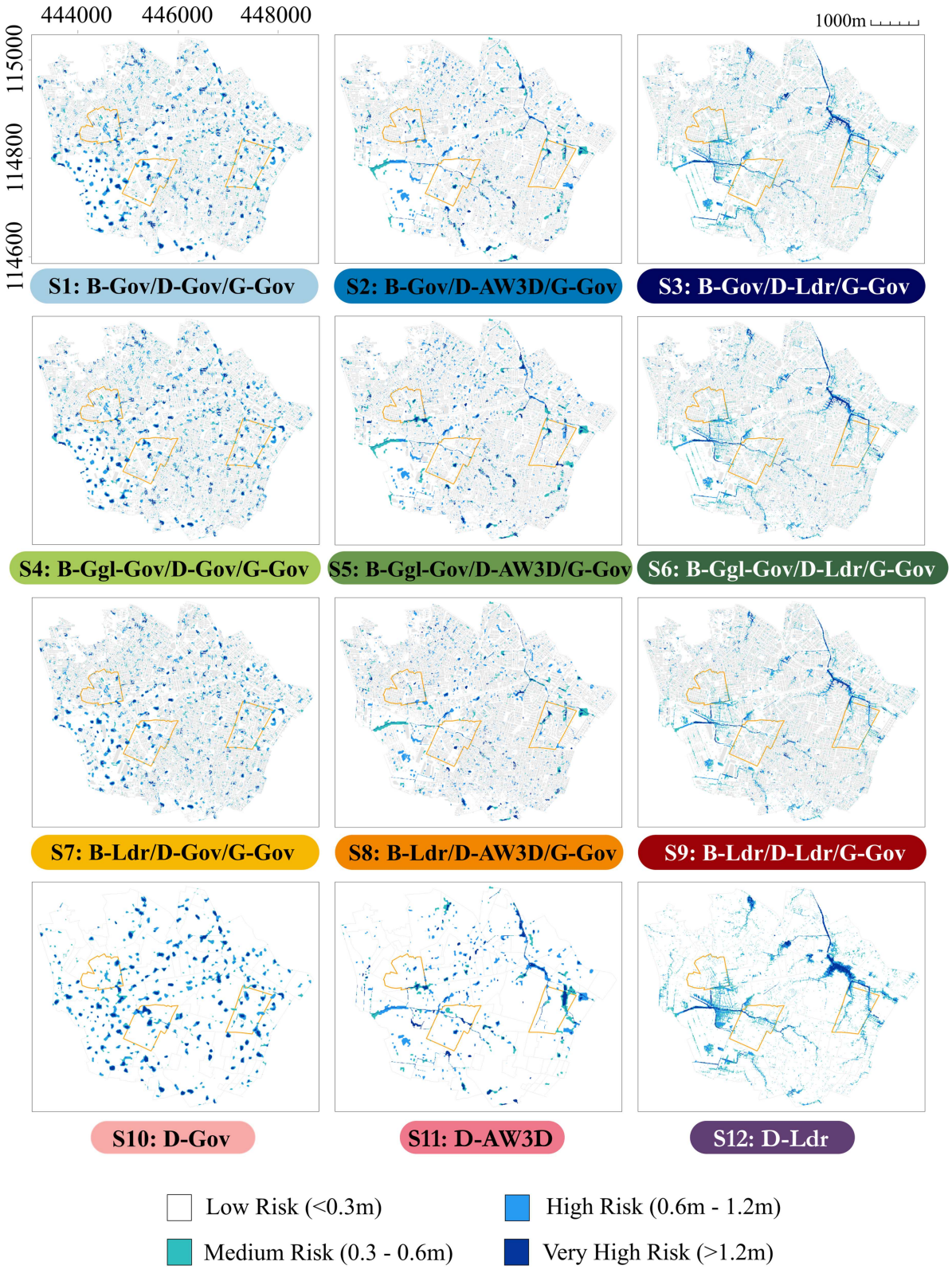

Figure 6. Each scenario demonstrates simulated flood risk. The plot labels indicate data inputs in the simulation (refer to Figure 4 for scenario label information).

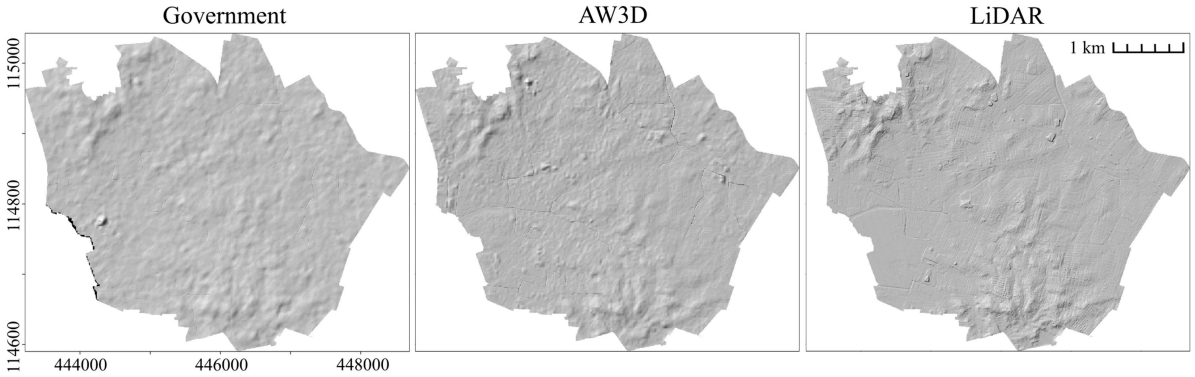

Figure 7. Hillshade comparison of the 4 m resampled government and AW3D DEMs, and the 4 m generated LiDAR DEM.

Table 3. Comparison of each scenario, showing overall flooded extent in km^2 and a statistical analysis of the spread of flood depth through observing the maximum, median, the median absolute deviation (MAD) and interquartile range (IQR).

	Overall extent (km2)	Flooded extent >0.3 m (%)	Depth max (m)	Depth median (m)	Depth MAD (m)	Depth IQR (m)	Depth IQR p90 (m)
S1	10.93	12.95	4.24	0.013	0.009	0.067	0.448
S2	10.94	10.46	4.63	0.026	0.018	0.067	0.324
S3	10.94	11.79	4.60	0.012	0.011	0.106	0.353
S4	10.32	13.93	3.77	0.014	0.010	0.088	0.486
S5	10.34	11.62	4.60	0.029	0.020	0.079	0.377
S6	10.34	12.44	4.81	0.014	0.012	0.120	0.369
S7	9.45	13.32	4.70	0.013	0.009	0.072	0.455
S8	9.47	10.81	4.58	0.026	0.018	0.066	0.348
S9	9.68	11.72	4.77	0.012	0.011	0.103	0.351
S10	15.63	9.06	4.36	0.009	0.004	0.017	0.235
S11	15.64	6.85	4.85	0.018	0.011	0.035	0.164
S12	15.64	9.31	4.59	0.007	0.006	0.053	0.276

When observing scenarios S4–S6 (merged Google and Gov BF) over El Socorro, a similar story to Nuevo Bosque was observed in that modelled flood risk in S6 (LiDAR DEM), which displayed spatial patterns similar to the community flood risk map (Figure 11). We also observe similar patterns in S5 (AW3D), which is likely due to the two major channels running throughout the neighbourhood. In El Socorro, there were greater discrepancies between the government and community flood risk maps, where areas highlighted as flood risk zones by communities were not exhibited in official data. The outputs in S5 (AW3D) and S6 (LiDAR DEM) displayed a combination of risk demonstrated on the community and government flood risk maps. This suggests that the model could be over- and under-representing and requires cautious interpretation of the results. Despite this, the modelled flood extent showed closer alignment with the experiences of local communities when using the LiDAR DEM.

Furthermore, whilst the community in El Socorro stated that flood levels are dangerously high during heavy rainfall periods, flood depth was not quantified. Therefore, modelled flood depths which increased by approximately 21% in S6 (LiDAR DEM) and S5 (AW3D) than S4 (ALOS PALSAR) should be interpreted as indicative rather than absolute and treated with caution until quantifiable validation data are available. Ideally, this would include surveying flood extent and water levels after a flood event using high-precision equipment such as a GNSS or a total station. However, in practice, this approach can be dangerous and difficult to implement. Thus, the community flood risk maps provide locally relevant information to compare against the simulated outputs.

When considering the impact of BF on the scenarios, there were greater overall flooded pixels in S3 (Gov BF) and S6 (merged Google and Gov BF), whilst S9 (LiDAR BF) showed 9% less flood extent but greater flood depth. This is likely due to the building size. Table 4 presents a comparison of BF across the different datasets. The median average of building size demonstrates that the government BF is 44% smaller than the LiDAR BF. This impacted the model flood depth and extent, with a larger BF under-estimating the flood extent.

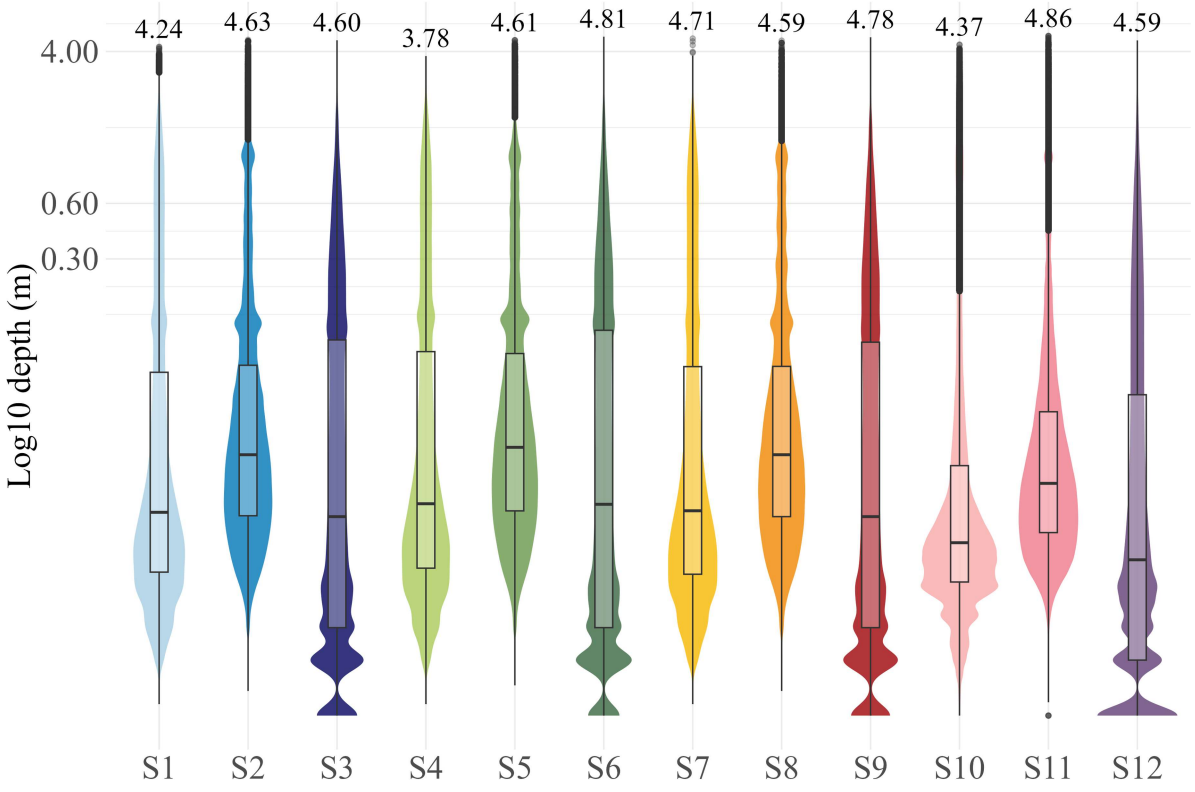

Figure 8. Comparison of flood depth distribution (m) for each scenario, shown on a Log^{10} scale (y-axis).

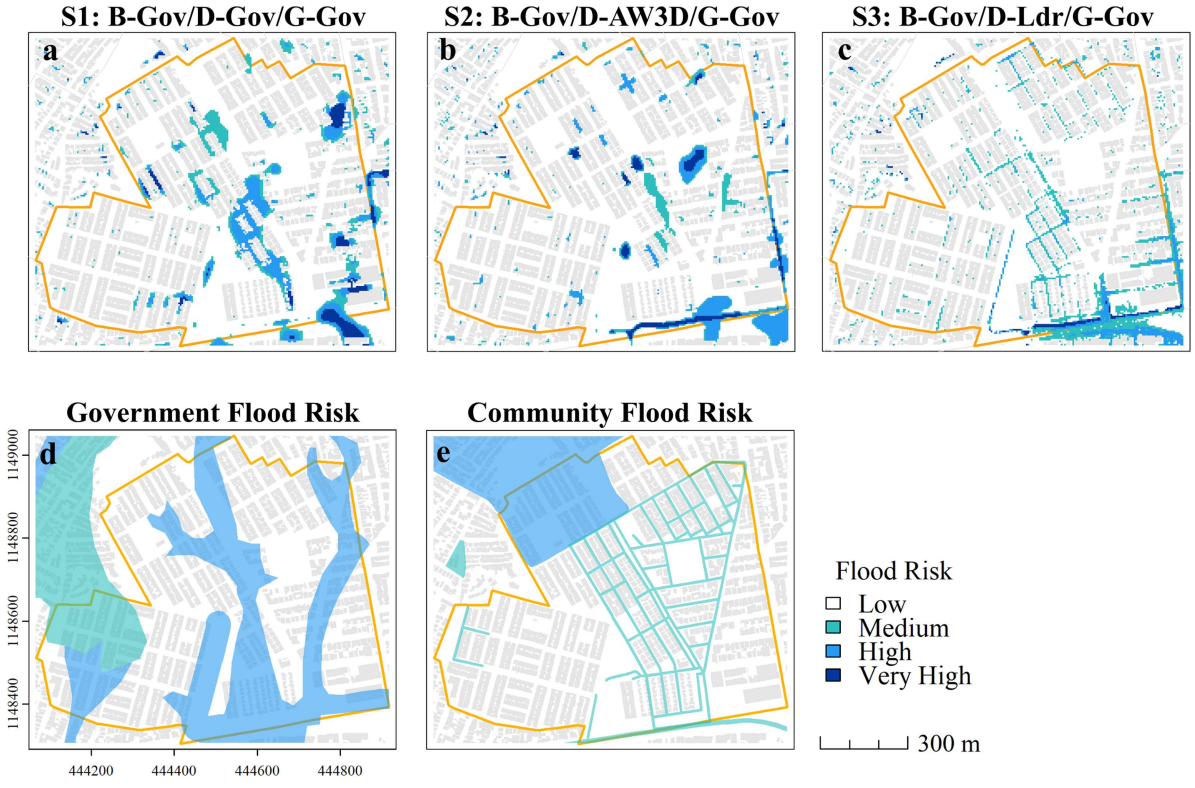

Figure 9. Flood risk maps in the neighbourhood of Nuevo Bosque. Maps include the simulated model outputs (a–c), the local government Secretaría de Planeación (2023) flood risk map (d) and co-created neighbourhood risk maps (e).

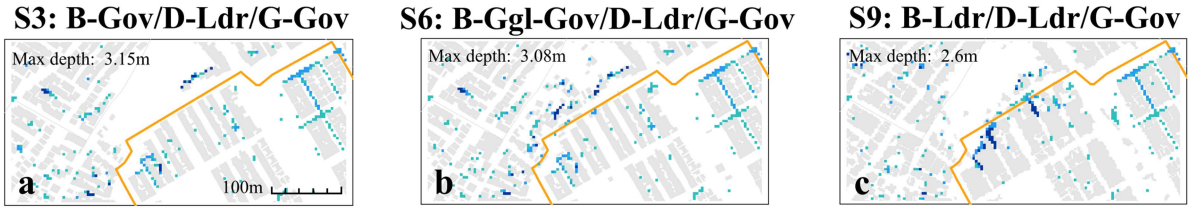

Figure 10. Flood maps showing the impact of missing building footprints on simulated flood model scenarios S3, S6 and S9.

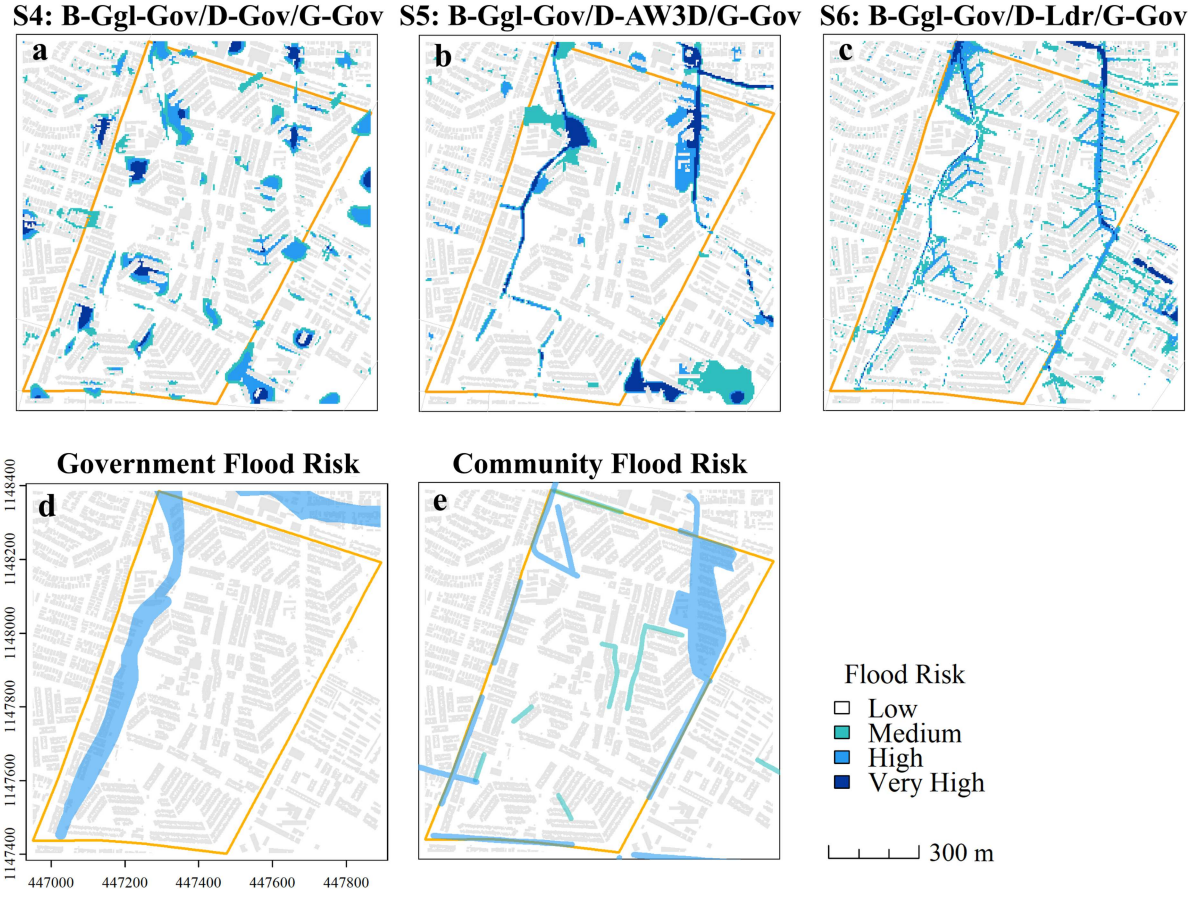

Figure 11. Flood risk maps in the neighbourhood of El Socorro. Maps include the simulated flood maps of S4, S5 and S6 (a–c), the Secretaría de Planeación (2023) flood risk map (d), and flood risk as experienced by local residents (e).

Table 4. Comparison of building footprint datasets used in scenario simulations, presenting the overall number of buildings, density and coverage, and average building size.

Building footprint dataset	Number of buildings	Mean building area (m^2)	Median building area (m^2)	Building density (km^2)	Building coverage (%)
Government	74,246	66	57	4756	31
Government & Google Merged	80,818	68	57	5177	35
LiDAR	19,586	319	102	1254	40

We present this in Figure 12, where the simulated maps a–c used the LiDAR DEM and different building footprints. In this figure, we highlight an area of flood risk to compare the co-created flood risk map with communities (map e) against the other flood maps. S9 (LiDAR BF) demonstrated lower flood risk than S3 (Gov BF) and S6 (merged Google and Gov BF). Whilst smaller buildings can be observed in

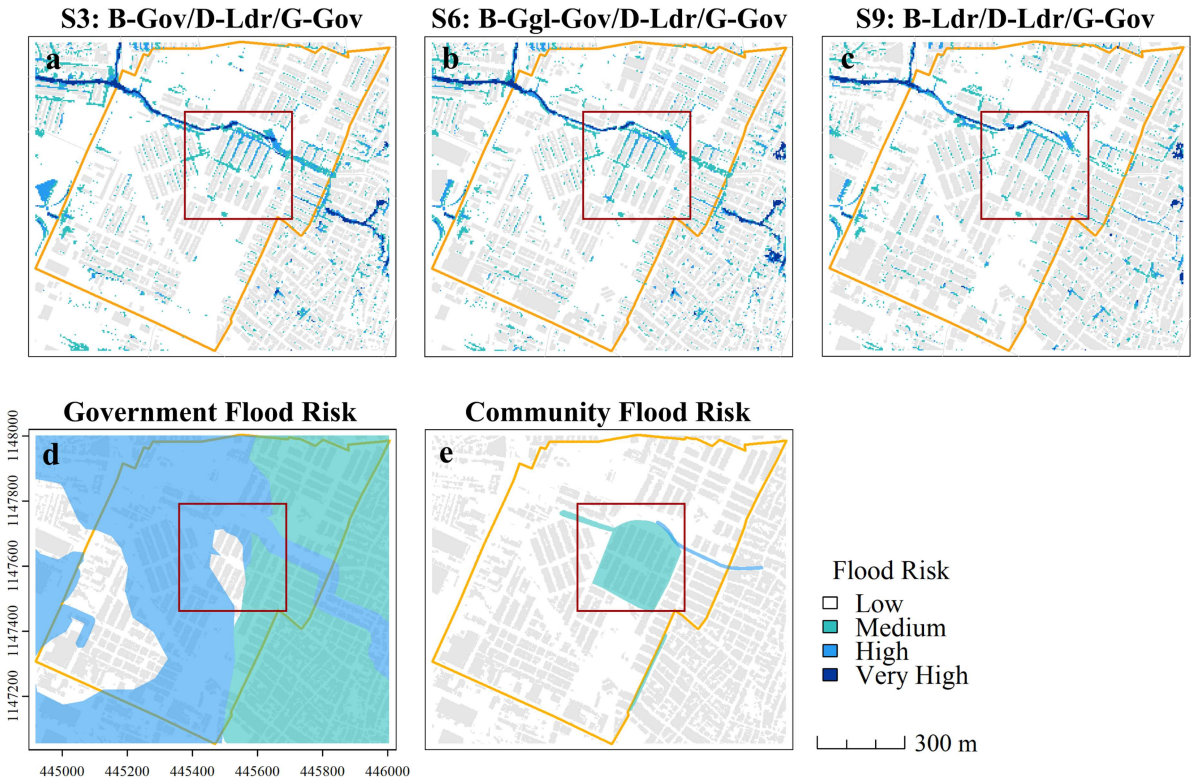

Figure 12. Maps showing flood risk in the neighbourhood of El Campestre. Maps (a–c) are the simulated maps of S3, S6 and S9, map (d) shows the local government Secretaría de Planeación (2023) flood risk and map (e) shows the neighbourhood flood risk map.

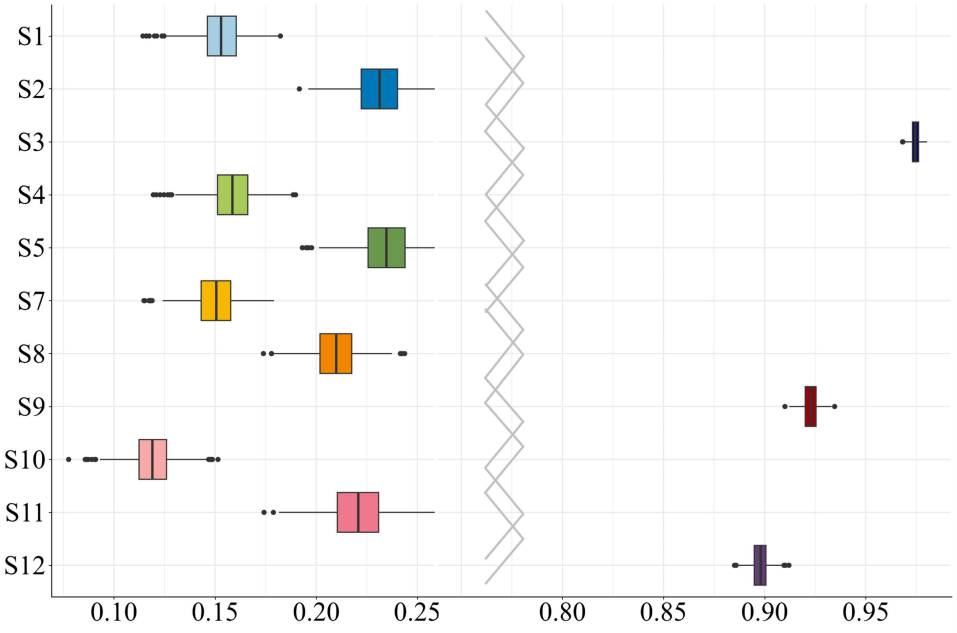

Figure 13. Confidence intervals using bootstrapped Spearman rank correlation across flood map scenarios.

Figure 11 Map C, the LiDAR-generated BFs were largely overestimated, merging closely connected buildings to form blocks of buildings rather than individual BF.

This highlights the importance of using BF, which reflects the existing terrain in flood model simulations. In the future, using approaches such as automatically assigning neighbourhoods with a region-growing approach or integrating LiDAR with very-high-resolution multispectral and hyperspectral imagery using deep learning approaches may improve building footprint accuracy (Hess et al. 2025; Rehman et al. 2025). Advanced processing techniques are not always available to practitioners; therefore, open-source alternatives remain valuable, and in this case, a combination of merged government and Google BF provided the updated BF to scale required for use in flood simulations.

4.2. Scenario comparisons

For the comparative analysis, we chose S6 (LiDAR DEM) as a baseline scenario to compare against as it demonstrated the closest alignment with flood risk areas identified by local communities and used the most up-to-date BF at the time of analysis, with the inclusion of the Google Open Buildings, which were produced in 2022. S3 (Gov BF) and S9 (LiDAR BF) were not chosen as the baseline, as S3 had missing BF, which impacted flood model outputs of areas at risk, as presented in Figure 10, and S9 had a reduced overall flood extent due to large buildings.

Figure S4 (supplementary material) demonstrates depth differences between S6 and all other scenarios, with MAE and RMSE to quantify divergence. Scenarios S3, S9 and S12 showed smaller differences with MAE values of 0.05 m or less, and RMSE values of 0.15 m or less. Similarities were expected in these scenarios, as S6 used the LiDAR DEM; however, S3 demonstrated the smallest difference, with an MAE of ≈0.01 m and an RMSE of 0.08 cm. All other scenarios demonstrated larger differences, with an RMSE of more than 0.32 m, showing greater divergence from S6.

Figure 13 presents the results of the bootstrapped Spearman's rank correlation coefficient. It shows a high correlation coefficient (ρ) score between the LiDAR DEM-based scenarios, meaning that there is greater similarity between these flood scenarios and S6. However, there were differences between each of these scenarios. With a tighter distribution and ρ-value of 0.97, S3 was most closely aligned with S6. S9 and S12 showed more variability, although the differences between the latter scenarios were small, with ρ-values of 0.90 and 0.92, respectively. This suggests that there are small differences between using the LiDAR BF and only using a LiDAR DEM. Scenarios using AW3D and the government DEMs showed ρ-values ranging between 0.11 and 0.23 with larger outliers, meaning that these scenarios are less consistent with local people's experiences of flooding in Cartagena.

4.3. Limitations and opportunities

This research has shown that flood exposure information can be improved with high-resolution spatial data. The differences were greater when using global DEMs, government, and AI-generated datasets, thus demonstrating that each dataset requires careful inspection and filtering before use in flood models. However, the absence of quantifiable validation data limits confidence in the simulated depth values. Furthermore, the rainfall sensitivity analysis (Figure S2, Supplementary material) demonstrated how the rainfall amount also introduces uncertainty in the magnitude of flood depth and extent. Therefore, estimates should be interpreted as approximate rather than exact values, and best-case scenarios must be evaluated cautiously rather than as definitive predictions. Obtaining spatially variable rainfall data would provide more realistic flood risk patterns, as would drainage information. However, without this information, data with meter-level precision can still improve flood risk understanding in data-scarce contexts. In this study, the LiDAR-derived DEM displayed spatial patterns of flood pathways broadly consistent with flooded areas identified by local communities, whilst demonstrating differences from official flood risk maps.

This is not unique to Cartagena; it is a common experience across many cities in LMICs. Open-source data is often used as a proxy in data-scarce cities, and whilst it is a valuable resource as can be seen with ALOS World 3D, this research has shown that it should be used alongside high-resolution datasets to preserve micro-topographic detail needed to model localized flooding. In Cartagena, the mayor's office is

keen to improve the city's resilience to climate change (Secretaría de Planeación 2024). Investing in high-resolution spatial data would help with this, albeit this is dependent on capacity, political will and prioritisation of competing demands. LiDAR is expensive, difficult to process due to the data size and requires high computational power. Collaborative relationships between researchers and local private companies can bridge these barriers, as was the case in this research, and enable vital disaster preparedness efforts (Ekeu-wei and Blackburn 2018).

Working with local communities to understand flooding in the absence of official data is extremely valuable for producing improved flood knowledge (Wolff 2021). However, Landwehr, Dasgupta, and Waske (2024) argue that validating flood maps with communities only is insufficient. We acknowledge that there were broad uncertainties in this study, as validation was carried out qualitatively on a small scale. Despite this, communities are the first to experience the direct impacts of flooding; therefore, they have rich insight and experience to contribute to advancing critical flood knowledge (Paul et al. 2020). In this study, local knowledge advanced flood information of the existing landscape in Cartagena when data availability and quality were limited. This presents an opportunity for future work to embed the knowledge of those impacted by flooding more systematically through multimedia methods such as images, fusing these with hydrological information, for timely, reliable flood information (Starkey et al. 2017; Fehri et al. 2020). However, in-situ approaches also have uncertainties and challenges, including safety, which is important to consider in Cartagena (Fraisl et al. 2022).

Flood observations can also be complemented by synthetic aperture radar (SAR) imaging in the absence of ground truth data in data-scarce regions (Dasgupta et al. 2020). However, our early experimentation using open-source Sentinel-1 imagery has not yet produced reliable results. This may be due to double-bounce backscatter issues from Cartagena's dense urban fabric, which can mask channels (Schumann 2020), and the spatial resolution of Sentinel-1 (10 m). Studies using 2–3 m spatial resolution SAR have found improved accuracy in urban areas when integrating the data with flood model results (Mason et al. 2021). We will investigate this further in future work through leveraging relationships with potential commercial partners, demonstrating the challenges of data acquisition.

Inequalities in data hinder equitable progress. As high-income countries adopt more advanced technology and higher resolution data, the gap in knowledge systems continues to increase, potentially undermining Sustainable Development Goals (SDGs) progress (Zachariah et al. 2025). In order not to leave communities behind, updated and realistic flood information is necessary for preparedness efforts. This research has focused on improving the data quality in Cartagena with LiDAR, but future work will continue to broaden these efforts on two parallel approaches. Firstly, to continue working with local stakeholders to systematically develop flood knowledge. Secondly, to further investigate model uncertainties by comparing alternative widely used 2D hydrodynamic models such as LISFLOOD and HEC-RAS (Bates and De Roo 2000; Brunner et al. 2015). Ongoing work will focus on addressing the limitations mentioned, whilst continuing to develop further relationships with communities and commercial entities to create a system that embeds and improves flood information in a data-scarce city.

5. Conclusion

In assessing open-source and proprietary datasets for more accurate flood risk information in data-scarce cities, we showed that metre-level spatial resolution data is necessary, particularly when risk is located at the urban neighbourhood level, where most vulnerabilities occur. Whilst ALOS World 3D performed better in simulations than the government DEM, the LiDAR DEM outperforms, as there were larger differences generated by open-source DEMs when compared against community flood risk maps. However, AI-generated building data supplemented official data, which was outdated and therefore missing data. Through auditing and using suitable open-source data, and leveraging relationships with commercial partners to generate spatial datasets using LiDAR data, we generated new information for use in hydrological models. However, precise magnitude values in modelled outputs must be interpreted with caution, taking into account data and model uncertainties. In the absence of quantitative validation data, we carried out a comparative analysis of dataset differences and flood patterns using bootstrapping with Spearman's rank correlation coefficient to compare all scenarios against the one scenario that qualitatively demonstrated community experiences of flooding. As a result, we generated improved flood maps that

align closely with the lived experiences of local communities, creating much-needed flood information in Cartagena.

A similar workflow could aid local governments and disaster relief agencies working in areas where existing flood information is missing or outdated, to provide updated information, which in turn can aid relevant flood preparedness plans. This would also ensure that in disaster preparedness, all local communities are included. To the best of our knowledge, this research was a principal study in using LiDAR data for use in a 2D physics-based flood model in the context of Cartagena, contributing to the literature and narrowing the gap in flood research in Latin America. Further future work will include integrating community experiences with flood model outputs more quantitatively, using geolocated images via mobile phones for capturing flood risk and depth, which would aid in improving certainty and contribute to in-situ flood observations.

Acknowledgements

We extend our gratitude to Cuatro Conceptos S.A.S. for providing the LiDAR data, which was critical to achieving the research aims. We also thank the Universidad Tecnológica de Bolívar, who were invaluable in supporting the research activities. We are grateful to the Water Security and Sustainable Development Hub (UK Research and Innovation: ES/S008179/1) for facilitating the connection. We thank Christos Iliadis for his support in using CityCAT. We are deeply grateful to Ángel Barreto Mora, Freddy Agámez Gómez and Oscar Martínez Montero, the leaders of El Campestre, El Socorro and Nuevo Bosque, who led their community to share their lived experiences of flooding. Finally, we thank the anonymous reviewers and editors for their time and valuable suggestions, which aided in strengthening this research.

Author contributions

CRediT: **Ambreen Masud:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Software, Validation, Visualization, Writing – original draft, Writing – review & editing; **Maria Valasia Peppa:** Conceptualization, Methodology, Supervision, Validation, Writing – review & editing; **Yady Tatiana Solano-Correa:** Conceptualization, Methodology, Supervision, Validation, Writing – review & editing; **Jon Philip Mills:** Conceptualization, Funding acquisition, Methodology, Supervision, Validation, Writing – review & editing; **Cat Button:** Conceptualization, Supervision, Writing – review & editing.

Disclosure statement

The authors declare no conflict of interest associated with this research.

Funding

This research was funded by the UK's Engineering and Physical Sciences Research Council (EPSRC) Centre for Doctoral Training (CDT) in Geospatial Systems (Reference: EP/S023577/1).

Data availability statement

The data that support the findings of this study are available upon request from the corresponding author. However, the LiDAR data is not publicly available due to commercial restrictions.

References

Abdelmoneim H et al. 2023. Integrating multi-sensor observations and rainfall-runoff inundation modeling for mapping flood extents over the Nile river basin: example from the 2020 flooding in Sudan. Geocarto Int. 38(1):2197504. https://doi.org/10.1080/10106049.2023.2197504

Airbus Defence and Space. 2024. Pleiades Neo 0.3 m satellite imagery of Cartagena, Colombia.

Amayor de Cartagena de Indias. 2024. Mapa Interactivo Digital de Asuntos del Suelo (MIDAS). https://midas.cartagena.gov.co/

Assumpção T, Popescu I, Jonoski A, Solomatine D. 2018. Citizen observations contributing to flood modelling: opportunities and challenges. Hydrol Earth Syst Sci. 22(2):1473–1489. https://doi.org/10.5194/hess-22-1473-2018

Azizian A, Brocca L. 2020. Determining the best remotely sensed DEM for flood inundation mapping in data sparse regions. IJRS. 41(5):1884–1906. https://doi.org/10.1080/01431161.2019.1677968

Bater C, Coops N. 2009. Evaluating error associated with lidar-derived DEM interpolation. Comput Geosci. 35(2):289–300. https://doi.org/10.1016/j.cageo.2008.09.001
Bates P, De Roo A. 2000. A simple raster-based model for flood inundation simulation. J Hydrol. 236(1-2):54–77. https://doi.org/10.1016/S0022-1694(00)00278-X
Bernhofen M et al. 2022. The role of global data sets for riverine flood risk management at national scales. Water Resour Res. 58(4):e2021WR031555. https://doi.org/10.1029/2021WR031555
Bertsch R, Glenis V, Kilsby C. 2022. Building level flood exposure analysis using a hydrodynamic model. Environ Model Softw. 156:105490. https://doi.org/10.1016/j.envsoft.2022.105490
Bozza A et al. 2016. Potential of remote sensing and open street data for flood mapping in poorly gauged areas: a case study in gonaives, Haiti. Appl Geom. 8(2):117–131. https://doi.org/10.1007/s12518-016-0171-x
Brunner G, Piper S, Jensen M, Chacon B. 2015. Combined 1D and 2D hydraulic modeling within HEC-RAS.
Calil J, Reguero B, Zamora A, Losada I, Méndez F. 2017. Comparative coastal risk index (CCRI): a multidisciplinary risk index for latin america and the Caribbean. PLoS One. 12(11):e0187011. https://doi.org/10.1371/journal.pone.0187011
Castilla Villalba D, Rueda Chávez L, Chiquillo Padilla I, Estrada Mesa S, Bustillo Mendoza G. 2023. Análisis de Datos Meteorológicos - IDEAM: Precipitación, Humedad, Evaporación y Temperatura, Ejercicios de Hidrología.
Cea L, Costabile P. 2022. Flood risk in urban areas: modelling, management and adaptation to climate change. A review. Hydrology. 9(3):50. https://doi.org/10.3390/hydrology9030050
Cohen J et al. 2025. A politics of global datasets and models in flood risk management. Water Altern. 18(2):305–329. https://www.water-alternatives.org/index.php/alldoc/articles/vol18/v18issue2/784-a18-2-9/file
Cuatro Conceptos. 2024. Airborne LiDAR. https://cuatroconceptos.com/en/
Couasnon A et al. 2020. Measuring compound flood potential from river discharge and storm surge extremes at the global scale. NHESS. 20(2):489–504. https://doi.org/10.5194/nhess-20-489-2020
DANE. 2025. Proyecciones de Población y Estudios Demográficos (PPED): Serie municipal de población por área, para el periodo 2018-2042. https://www.dane.gov.co/index.php/estadisticas-por-tema/demografia-ypoblacion/proyecciones-de-poblacion
Dasgupta A, Thakur P, Gupta P. 2020. Potential of Sar-derived flood maps for hydrodynamic model calibration in data scarce regions. J Hydrol Eng. 25(9):05020028. https://doi.org/10.1061/(ASCE)HE.1943-5584.0001988
Ekeu-wei I, Blackburn G. 2018. Applications of open-access remotely sensed data for flood modelling and mapping in developing regions. Hydrology. 5(3):39. https://doi.org/10.3390/hydrology5030039
El Espectador. 2023. Inundaciones en Cartagena dejaron más de 10.000 familias afectadas. El Espectador. https://www.elespectador.com/colombia/cartagena/cartagena-e-inundaciones-portorrenciales-lluvias-10-mil-familias-afectadas-emergencias-noticias-colombia/
El Heraldo. 2024. Calles y casas inundadas: reporte de primeras lluvias del año en Cartagena. El Heraldo. https://www.elheraldo.co/bolivar/2024/04/22/calles-ycasas-inundadas-reporte-de-primeras-lluvias-del-ano-en-cartagena/
El Universal. 2022. Cartagena bajo el agua: inundaciones y canales desbordados por lluvias. El Universal. https://www.eluniversal.com.co/cartagena/2022/09/15/cartagena-bajo-el-aguainundaciones-y-canales-desbordados-por-lluvias/
Espinosa A, Madero Jirado M, Rodríguez-Puello G, Díaz-Canedo L. 2020. Ethnicity, space and human development in poor urban communities: commune 6 in cartagena de indias, Colombia. Cuadernos de Economía. 39:635–666. https://doi.org/10.15446/cuad.econ.v39n81.77333
Ester M et al. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. KDD. 96:226–231.
Etinay N, Egbua C, Murray V. 2018. Building urban resilience for disaster risk management and disaster risk reduction. Procedia Eng. 212:575–582. https://doi.org/10.1016/j.proeng.2018.01.074
Faghih M, Mirzaei M, Adamowski J, Lee J, El-Shafie A. 2017. Uncertainty estimation in flood inundation mapping: an application of non-parametric bootstrapping. River Res Appl. 33(4):611–619. https://doi.org/10.1002/rra.3108
Fehri R, Bogaert P, Khlifi S, Vanclooster M. 2020. Data fusion of citizen-generated smartphone discharge measurements in Tunisia. J Hydrol. 590:125518. https://doi.org/10.1016/j.jhydrol.2020.125518
Fraisl D et al. 2022. Citizen science in environmental and ecological sciences. Nat Rev Methods Primers. 2(1):64. https://doi.org/10.1038/s43586-022-00144-4
Gevaert C, Buunk T, Van Den Homberg M. 2024. Auditing geospatial datasets for biases: using global building datasets for disaster risk management. IJSTA. 17:12579–12590. https://doi.org/10.1109/JSTARS.2024.3422503
Glenis V, McGough A, Kutija V, Kilsby C, Woodman S. 2013. Flood modelling for cities using cloud computing. J Cloud Comput Adv Syst Appl. 2(1):7. https://doi.org/10.1186/2192-113X-2-7
Glenis V, Kutija V, Kilsby C G 2018. A fully hydrodynamic urban flood modelling system representing buildings, green space and interventions. Environmental Modelling & Software. 109:272–292. https://doi.org/10.1016/j.envsoft.2018.07.018
Gonzalez-Alvarez A, Coronado-Hernández O, Fuertes-Miquel V, Ramos H. 2018. Effect of the non-stationarity of rainfall events on the design of hydraulic structures for runoff management and its applications to a case study at gordo creek watershed in cartagena de indias, Colombia. Fluids. 3(2):27. https://doi.org/10.3390/fluids3020027
Google. 2024. Open Buildings. https://sites.research.google/gr/openbuildings

Guerreiro S, Glenis V, Dawson R, Kilsby C. 2017. Pluvial flooding in european cities—A continental approach to urban flood modelling. Water. 9(4):296. https://doi.org/10.3390/w9040296
Guo Q, Li W, Yu H, Alvarez O. 2010. Effects of topographic variability and lidar sampling density on several DEM interpolation methods. Photogramm Eng Remote Sens. 76(6):000–000. https://doi.org/10.14358/PERS.76.6.701
Herfort B, Lautenbach S, Porto de Albuquerque J, Anderson J, Zipf A. 2023. A spatiotemporal analysis investigating completeness and inequalities of global urban building data in OpenStreetMap. Nat Commun. 14(1):3985. https://doi.org/10.1038/s41467-023-39698-6
Hess M, Rheinwalt A, Bookhagen B. 2025. Refining point-cloud neighborhood construction for improved classification. Sci Remote Sens. 12:100325. https://doi.org/10.1016/j.srs.2025.100325
Hirsch R, Archfield S, De Cicco L. 2015. A bootstrap method for estimating uncertainty of water quality trends. Environ Model Softw. 73:148–166. https://doi.org/10.1016/j.envsoft.2015.07.017
Hung C-L, James L, Hodgson M. 2018. An automated algorithm for mapping building impervious areas from airborne LiDAR point-cloud data for flood hydrology. GISci Remote Sens. 55(6):793–816. https://doi.org/10.1080/15481603.2018.1452588
IDEAM. 2025. Instituto de Hidrología, Meteorología y Estudios Ambientales (IDEAM): Consulta y Descarga de Datos Hidrometeorológicos. http://dhime.ideam.gov.co/atencionciudadano/
Iliadis C, Glenis V, Kilsby C. 2023. Cloud modelling of property-level flood exposure in megacities. Water. 15(19):3395. https://doi.org/10.3390/w15193395
JAXA. 2014. "ALOS PALSAR High Resolution (12.5m) Digital Elevation Model (20062011)." Japan Aerospace Exploration Agency (JAXA), Earth Observation Research Center (EORC). https://www.earthdata.nasa.gov/data/projects/alos-palsar-rtc-project
Karsli B, Yilmazturk F, Bahadir M, Karsli F, Ozdemir E. 2024. Automatic building footprint extraction from photogrammetric and LiDAR point clouds using a novel improved-octree approach. J Build Eng. 82:108281. https://doi.org/10.1016/j.jobe.2023.108281
Landwehr T, Dasgupta A, Waske B. 2024. Towards robust validation strategies for EO flood maps. Remote Sens Environ. 315:114439. https://doi.org/10.1016/j.rse.2024.114439
Limberger F, Oliveira M. 2015. Real-time detection of planar regions in unorganized point clouds. Pattern Recognit. 48(6):2043–2053. https://doi.org/10.1016/j.patcog.2014.12.020
Löwe R et al. 2017. Assessment of urban pluvial flood risk and efficiency of adaptation options through simulationsA new generation of urban planning tools. J Hydrol. 550:355–367. https://doi.org/10.1016/j.jhydrol.2017.05.009
Ma R. 2005. DEM generation and building detection from lidar data. Photogramm Eng Remote Sens. 71(7):847–854. https://doi.org/10.14358/PERS.71.7.847
Mandal S et al. 2026. Bridging modelled and lived flood risk: a participatory GIS approach to community-based flood vulnerability mapping in north bengal, India. Global Earth Surf Process Change. 6:100017.
Mason D et al. 2021. Improving urban flood mapping by merging synthetic aperture radar-derived flood footprints with flood hazard maps. Water. 13(11):1577. https://doi.org/10.3390/w13111577
Microsoft. 2023. Microsoft Building Footprints. https://github.com/microsoft/GlobalMLBuildingFootprints
Mouthon-Bello J et al. 2022. Spatial variability study of rainfall in cartagena de indias, Colombia. J Water Land Dev (55). 138–149. https://doi.org/10.24425/jwld.2022.142316
Muhadi N, Abdullah A, Bejo S, Mahadi M, Mijic A. 2020. The use of LiDAR-derived DEM in flood applications: a review. Remote Sens. 12(14):2308. https://doi.org/10.3390/rs12142308
Muthusamy M, Casado M, Butler D, Leinster P. 2021. Understanding the effects of digital elevation model resolution in urban fluvial flood modelling. J Hydrol. 596:126088. https://doi.org/10.1016/j.jhydrol.2021.126088
Nasa J. 2020. NASADEM Merged DEM Global 1 arc second V001.
Naufal C, Solano-Correa Y, Marrugo A. 2024. Time series water body analysis through planet satellite imagery: a coastal urban case study. Geospatial Informatics XIV. 13037(7):28–35. https://doi.org/10.1117/12.3014198
Nkwunonwo U, Whitworth M, Baily B. 2020. A review of the current status of flood modelling for urban flood risk management in the developing countries. Sci Afr. 7:e00269. https://doi.org/10.1016/j.sciaf.2020.e00269
OECD. 2022. National Urban Policy Review of Colombia. OECD Publishing.
OpenStreetMap. 2024. OpenStreetMap. https://www.OpenStreetMap.org
Paul J et al. 2020. Applying citizen science for sustainable development: rainfall monitoring in Western Nepal. In: Frontiers in Water. 2, Publisher: Frontiers Media S.A. https://doi.org/10.3389/frwa.2020.581375
Pinos J, Quesada-Román A. 2021. Flood risk-related research trends in latin america and the Caribbean. Water. 14(1):10. https://doi.org/10.3390/w14010010
Polat N, Uysal M, Toprak A. 2015. An investigation of DEM generation process based on LiDAR data filtering, decimation, and interpolation methods for an urban area. Meas. 75:50–56. https://doi.org/10.1016/j.measurement.2015.08.008
Prashar N, Lakra H, Kaur H, Shaw R. 2023. Urban flood resilience: mapping knowledge, trends and structure through bibliometric analysis. Environ Dev Sustain. 26:8235–8265. Accessed 2023-06-01. https://doi.org/10.1007/s10668-023-03094-3. Accessed 2023-06-01.
Rehman M Z U, Mohammed Shamsul Islam S, Blake D, Ulhaq A, Janjua N 2025. Deep learning for land use classification: a systematic review of HS-LiDAR imagery. Artif Intell Rev. 58(9):272. https://doi.org/10.1007/s10462-025-11265-z

ReliefWeb. 2022. Natural Hazards Monitoring - 9 November. https://ReliefWeb.int/report/colombia/natural-hazards-monitoring-9-november-2022
Roussel J-R et al. 2020. lidR: An R package for analysis of airborne laser scanning (ALS) data. Remote Sens Environ. 251:112061. https://doi.org/10.1016/j.rse.2020.112061
Sirko W et al. 2021. Continental-scale building detection from high resolution satellite imagery. arXiv preprint arXiv:2107.12283.
Schubert E, Sander J, Ester M, Kriegel H-P, Xu X. 2017. DBSCAN revisited, revisited: why and how you should (Still) use DBSCAN. ACM Trans Database Syst 42(3):19–21. https://doi.org/10.1145/3068335
Schumann G-P. 2020. Grand challenges in microwave remote sensing. Front Remote Sens. 1:603650. https://doi.org/10.3389/frsen.2020.603650
Schumann G-P. 2024. Breakthroughs in satellite remote sensing of floods. Front Remote Sens. 4:1280654. https://doi.org/10.3389/frsen.2023.1280654
Secretaría de Planeación. 2023. Estadísticas de población del Distrito por géneros y edades: Proyecciones 2019–2024. In: Alcaldía Mayor de Cartagena de Indias. Departamento Nacional de Planeación (DNP).
Secretaría de Planeación. 2024. Plan de Desarrollo 2024-2027. Cartagena, Ciudad de Derechos. https://planeacion.cartagena.gov.co/acuerdo-2024-2027
Secretaría de Planeación. 2023. Plan de Ordenamiento Territorial Geodata. https://planeacion.cartagena.gov.co/
See L. 2019. A review of citizen science and crowdsourcing in applications of pluvial flooding. FrEaS. 7:44. https://doi.org/10.3389/feart.2019.00044
Semana. 2022. 96 emergencias atendidas en 37 barrios de Cartagena: esto es el reporte de las fuertes lluvias. https://www.semana.com/nacion/cartagena/articulo/96-emergencias-atendidas-en-37barrios-de-cartagena-esto-es-el-reporte-de-las-fuertes-lluvias/202216/
Sibandze P, Kalumba A, H. Aljaddani A, Zhou L, Afuye G. 2025. Geospatial mapping and meteorological flood risk assessment: a global research trend analysis. Environ Manage. 75(1):137–154. https://doi.org/10.1007/s00267-024-02059-0
Starkey E et al. 2017. Demonstrating the value of community-based ('citizen science') observations for catchment modelling and characterisation. J Hydrol. 548:801–817. https://doi.org/10.1016/j.jhydrol.2017.03.019
Takaku J, Tadono T, Tsutsui K. 2014. Generation of high resolution global DSM from ALOS PRISM. Int Arc Photogramm Remote Sens Spatial Inf Sci. 40:243–248. https://doi.org/10.5194/isprsarchives-XL-4-243-2014
Thakuri S et al. 2022. Opensource data alternatives and models for flood risk management in Nepal. Remote Sens. 14(22):5660. https://doi.org/10.3390/rs14225660
Tibshirani R, Efron B. 1993. An introduction to the bootstrap. Monogr Stat Appl Probab. 57(1):1–436.
Toro E F. 2013. Riemann solvers and numerical methods for fluid dynamics: a practical introduction. Springer Science & Business Media.
UN Habitat. 2022. World Cities Report 2022. https://unhabitat.org/wcr/
UNDRR. 2015. "Sendai Framework for Disaster Risk Reduction 2015–2030." United Nations Office for Disaster Risk Reduction (UNDRR). https://www.undrr.org/publication/sendai-framework-disaster-risk-reduction-2015-2030
Walker S, Wilkinson S, McVicar T, Castellazzi P, Khan S. 2025. Optimising sub-metre resolution 3D geomorphic change detection over large areas using multitemporal airborne laser scanning with Sentinel-1 InSAR and Sentinel-2 optical observations. Remote Sens Environ. 317:114522. https://doi.org/10.1016/j.rse.2024.114522
Wechsler S. 2007. Uncertainties associated with digital elevation models for hydrologic applications: a review. Hydrol Earth Syst Sci. 11(4):1481–1500. https://doi.org/10.5194/hess-11-1481-2007
Wolff E. 2021. The promise of a "people-centred" approach to floods: types of participation in the global literature of citizen science and community-based flood risk reduction in the context of the sendai framework. Progress Disaster Sci. 10:100171. https://doi.org/10.1016/j.pdisas.2021.100171
World Health Organization. 2023. Appeal Colombia. https://www.who.int/emergencies/funding/health-emergency-appeals/2022/2023/2023appeals/appeal-colombia
Yoshida K et al. 2023. Threedimensional numerical modelling of floods in river corridor with complex vegetation quantified using airborne LiDAR imagery. J Hydraul Res. 61(1):88–108. https://doi.org/10.1080/00221686.2022.2106596
Zachariah M et al. 2025. Growing exposure and uncertain rainfall trends highlight the critical need for climate resilience in Colombia and Venezuela.
Zhang W et al. 2016. An easy-to-use airborne LiDAR data filtering method based on cloth simulation. Remote Sens. 8(6):501. https://doi.org/10.3390/rs8060501
Zscheischler J et al. 2018. Future climate risk from compound events. Nat Clim Change. 8(6):469–477. https://doi.org/10.1038/s41558-018-0156-3