Journal of Hydrology 617 (2023) 128975

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Research papers

Developing reliable urban flood hazard mapping from LiDAR data

José M. Bodoque [a,*], Estefanía Aroca-Jiménez [a], Miguel Á. Eguibar [b], Juan A. García [c]

[a] Department of Mining and Geological Engineering, University of Castilla-La Mancha, Toledo, Spain
[b] Institute for Water and Environmental Engineering (IIAMA), Department of Hydraulic Engineering and Environment, Universitat Politècnica de València, Valencia, Spain
[c] Department of Business Administration, University of Castilla-La Mancha, Talavera de la Reina, Spain

ARTICLE INFO

This manuscript was handled by C. Corradini, Editor-in-Chief

Keywords:
Urban areas
Digital surface model
LiDAR dataset
Hydraulic modelling
Flood hazard

ABSTRACT

Digital surface models (DSMs) are crucial in providing accurate urban flood hazard maps. The ubiquitous availability of LiDAR data (where accessible) makes constructing geometrically sound DSMs feasible. However, little attention has been paid to developing approaches for producing geometrically consistent DSMs. Herein is described an application-driven procedure for creating a geometrically robust DSM (DSM1). Two further DSMs were created, one for portraying streets using breaklines as ancillary information (DSM2) and the other through direct interpolation of LiDAR data (DSM3). The geometrical correctness and vertical accuracy of these DSMs were examined qualitatively and quantitatively by plotting longitudinal profiles and cross-sections onto major runoff pathways and determining statistical error. The effect of these DSMs' geometric consistency on flood hazard maps was also evaluated. For this, hydraulic outputs from DSM1 were used as a benchmark to compare hydraulic outputs from DSM2 and DSM3. This comparison was conducted at two spatial resolutions: i) at the total area flooded using the F statistic; and ii) at the pixel level by employing global indices and category-level indices extracted from a confusion matrix. Our findings revealed that: 1) DSM1 defined the most geometrically coherent configurations for runoff pathways; 2) in urban areas with the most densely packed streets and buildings, DSM2 and particularly DSM3 featured the most unrealistic geometric representations of the urban domain, displaying fake water flow barriers and lower than real runoff pathway cross-sections; and 3) the geometric quality of the DSMs created had a significant impact on flood hazard maps reliability (i.e., the disagreement in flood hazard categories between DSM2 and DSM3 and DSM1 varied from 28% to 82%). These findings can be very valuable in achieving further reductions and better flood risk management.

1. Introduction

Between 2000 and 2019, 7348 natural hazard-related disasters caused 1.23 million deaths and USD 2.97 trillion in economic losses, affecting more than 4 billion people worldwide (UNDRR, 2020). During this period, climate-related disasters increased, with floods accounting for 44 % of all disasters (WMO, 2021). Flood risk, on the other hand, is expected to rise in the coming decades owing to climate change-induced increases in the frequency and magnitude of floods, along with the projected global growth of the exposed population and assets (Kundzewicz et al., 2014). Accordingly, flood risk mitigation is deemed to be one of the key societal challenges to be addressed during this century (O'Donnell and Thorne, 2020).

Coping with this challenge requires good urban flood risk management and in-depth knowledge of flood hazard along river basins and coastal areas (Tsakiris, 2014; Pasquier et al., 2019). Such an appraisal is determined by the type of flooding (e.g., fluvial, flash floods, or pluvial) and the probability or return period of the flood event to be assessed. As an example, the EU Floods Directive considers the following scenarios: i) low probability, or extreme events (500-year flood); ii) medium probability (100-year flood); and iii) high probability, where appropriate (10-year flood). Further, flood magnitude is considered (Prinos et al., 2008; Mudashiru et al., 2021). In the context outlined above, flood hazard maps are essential for reporting vulnerabilities and risks (Masood and Takeuchi, 2012). This mapping also facilitates the implementation of successful risk management strategies based on preventive and preparedness actions to avoid, minimize, transfer, share, or accept flood hazards (di Baldassarre et al., 2009; Surminski and

* Corresponding author.
E-mail addresses: josemaria.bodoque@uclm.es (J.M. Bodoque), estefania.aroca@uclm.es (E. Aroca-Jiménez), meguibar@hma.upv.es (M.Á. Eguibar), juan.garcia@uclm.es (J.A. García).

https://doi.org/10.1016/j.jhydrol.2022.128975
Received 18 October 2022; Received in revised form 6 December 2022; Accepted 10 December 2022
Available online 15 December 2022

Thieken, 2017).

Flood hazard assessment is a complex process that requires determining its magnitude. Ideally, this encompasses the specification of flood depth, flood wave velocity, its duration, and sediment and debris loading (Scawthorn et al., 2006). In practice, the above analysis is simplified by considering flood extent, water depth or, eventually, a combination of the two parameters above and flow velocity (van Alphen et al., 2009). In urban areas, hazard assessment and cartographic depiction rely on two-dimensional (2D) hydrodynamic models, since in this environment it is not feasible to assume a one-dimensional (1D) flow (Vojinovic and Tutulic, 2009). Upstream boundary conditions of 2D hydrodynamic models are generally provided by flow quantiles established through hydrological modelling, or flood frequency analysis, depending on the available data. Other mandatory inputs are land cover, to estimate roughness, and topography portrayed as digital surface models (DSMs), which capture both natural and built/artificial urban environment features (Bodoque et al., 2016).

DSM topographic data affects 2D hydrodynamic modelling because it influences the urban geometric layout of structured or non-structured meshes utilized for numerical computations (Yalcin, 2020). The availability of LiDAR data since the end of the last century offers a fast and cost-effective way to get detailed DSMs of urban areas (Priestnall et al., 2000). LiDAR data, by virtue of the very high density of recorded points, allows modelling, and flood mapping with a horizontal spatial resolution of 1 to 2 m and a Z-precision of a few decimetres (Costabile et al., 2015). However, producing DSMs from LiDAR data for flood hazard modelling and mapping is not straightforward. This is because LiDAR data alone cannot sufficiently capture topographic breaks, especially in complicated geometric configurations, such as urban areas (Turner et al., 2013).

To make the above possible, LiDAR data must be coupled with 3D breaklines that complement mass points, reinforcing topographic breakdowns inside triangulated irregular networks, TINs (Liu, 2008). The best way to identify breaklines is to employ local knowledge, orthophotos, and cadastral maps as data sources and GIS tools to handle this information systematically (Bodoque et al., 2016). Disaggregating LiDAR data into ground (terrain) and non-ground (man-made structures) is another important stage in creating urban DSMs. This stage is challenging since the algorithms used to classify LiDAR data are not totally reliable. Consequently, elements such as cars, canopies, or flowerpots are occasionally misclassified as ground, causing DSMs to block water flow (Noh et al., 2018).

Few studies have focused on the need for geometrically correct DSMs to generate reliable flood hazard maps that may subsequently be employed in urban flood risk management. To the best of our knowledge, there are hardly any studies on how to develop geometrically sound urban DSMs that display preferential drainage pathways and urban features that obstruct water flow (i.e., buildings and walls). In this regard, the papers by Meesuk et al. (2015) and Noh et al. (2018) serve as examples. In addition, there does not seem to be any research examining the effects of using DSMs that are not geometrically robust on the hydraulic performance of flooded areas in terms of water depth and flow velocity, which commonly dictate the layout of flood hazard maps.

To fill this knowledge gap, this study aims to evaluate how the geometric consistency of DSMs obtained from LiDAR data impacts on flood hazard outcomes resulting from 2D hydrodynamic modelling. To begin, a procedure for obtaining geometrically sound urban DSMs was devised. The resulting DSM (henceforth denoted as DSM1) was compared to two other DSMs (henceforth designated as DSM2 and DSM3) built under different assumptions. This was done to determine the differences in geometric configuration and flood hazard outcomes deployed by each DSM. To test geometric soundness, longitudinal profiles and cross-sections were created for the preferential drainage pathways during flooding (i.e., understood here as the 'streets-buildings' system). Vertical accuracy was also measured. This was done by computing discrepancies between DSM pixel values and 2,088 points taken using Differential Global Positioning System (DGPS) with a vertical accuracy lower than the LiDAR data used here. In addition, it was assessed how the geometric imprecision of DSMs influences the mapping of flood hazard. For this purpose, hydraulic outputs from DSM1 were utilized as a benchmark to compare with DSM2 and DSM3 outputs. This comparison was conducted by calculating the F statistic (Cook and Merwade, 2009) and global and category-level indices through a confusion matrix (Congalton, 1991; Congalton and Green, 1999).

2. Study site and data employed

Navaluenga is located in central Spain, on the banks of the Alberche River, between the Sierra del Valle (Gredos's eastern mountain range) and the Sierra de la Paramera (40°24′30″ N; 4°42′17″ W; 761 m.a.s.l., Fig. 1).

Navaluenga has 2135 inhabitants (data for 2021), however, this figure can rise to 20,000 during the summer. 4,311 dwellings are estimated, with 3392 being second homes. The Alberche River flows 70 km from 1,800 m.a.s.l. to Navaluenga. This river drains 717 km^2 up to this village, with a time of concentration (T_c) of 8.5 h. Its hydrological regime is hardly affected by human intervention, as there is no important hydraulic infrastructure (e.g., reservoirs) upstream of Navaluenga that regulates river water levels or the variability of river flows. Within Navaluenga and its surroundings, several torrents flow into the Alberche River. Among these, the Chorrerón Stream ($T_c \approx 3$ h) stands out. It is an ephemeral stream, completely channelled in its urban reach, which crosses Navaluenga perpendicularly from north to south (Fig. 1). Since at least the late Middle Ages, Navaluenga has been flooded by the Alberche river and the Chorrerón Stream. The 1990 s and 2000 s witnessed the most recent floods, which caused economic losses and endangered locals (Díez-Herrero, 2001).

The DSMs used here were built from the LiDAR data provided by the National Geographic Institute of Spain (IGN). Raw LiDAR data were collected by airborne LiDAR systems with a density of 0.5 points.m^{-2} and an altimetric precision (Z) of 15 cm. The reference system for this topographic dataset is the ETRS89/UTM zone 30 N (compatible with WGS 84). Using 2,088 differential GPS measurements provided by the Spanish General Directorate of Cadastre (SEC), the elevation accuracy of the DSMs was tested.

The study area was divided into 4 zones (Fig. 1): i) the total area of the study site, which includes the urban and *peri*-urban areas of Navaluenga (zone 1); ii) the whole urban and *peri*-urban floodable zone (zone 2); iii) the streets included within the urban floodable zone (zone 3); and iv) the floodable urban area with a higher density of streets and buildings (zone 4). Geometric consistency and elevation accuracy of DSMs were investigated in zones 1, 2, 3, and 4. In zones 2, 3 and 4, the results of hydraulic and flood hazards from each DSM were examined and compared.

3. Methodology

The methodological approach deployed here was divided into four stages (Fig. 2): (1) developing geometrically reliable DSMs for urban areas. This section provides a detailed description of the procedure designed to obtain a geometrically reliable DSM (DSM1), as well as the way in which the DSMs (DSM2 and DSM3) to be compared with DSM1 were obtained; (2) evaluation of the suitability of DSMs as topographic input to hydraulic models. The qualitative tools used to evaluate the geometric correctness of the DSMs were detailed, as was the statistical approach adopted to evaluate their vertical accuracy; (3) 2D hydraulic modelling and flood hazard analysis. It explains how the 2D hydrodynamic model was designed and implemented, and the approach employed to obtain flood hazard maps; and (4) comparison of flood hazard maps. The quantitative indicators used to compare the flood hazard maps displayed by each DSM, both at the level of the total flooded area and the flood hazard category displayed in each pixel, are

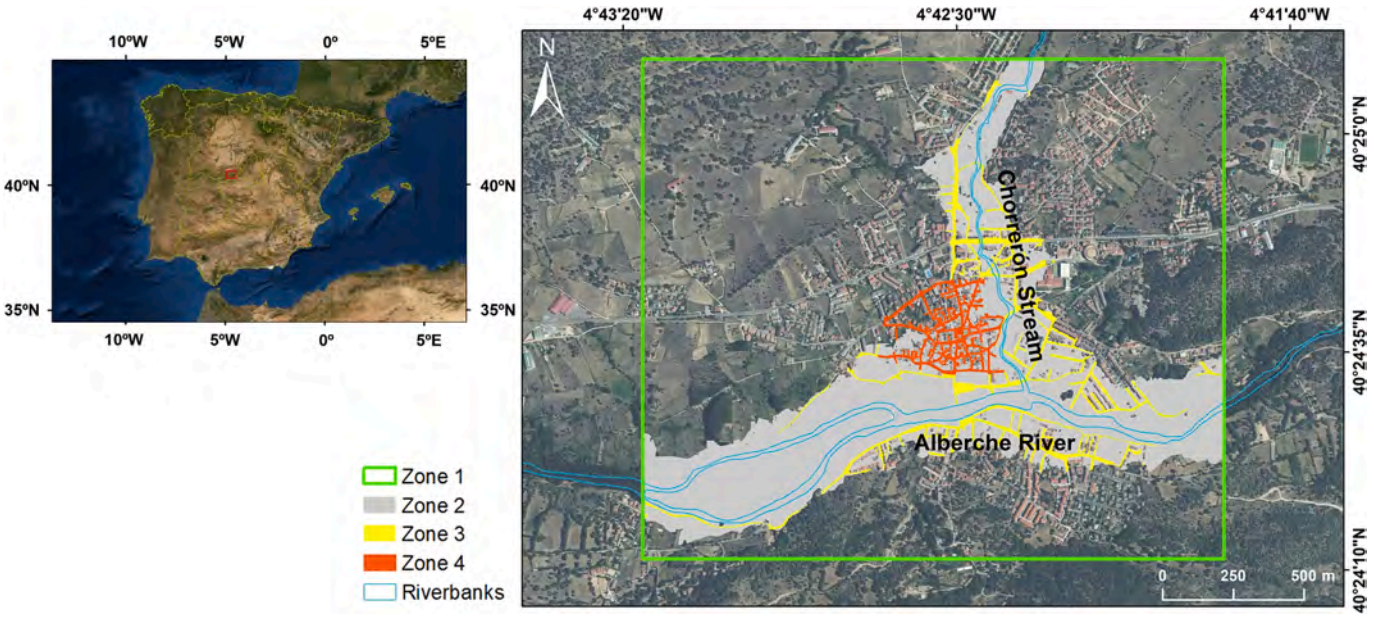

Fig. 1. Location of the case study. This figure depicts the zones into which the case study was subdivided to achieve the objectives of the research conducted here.

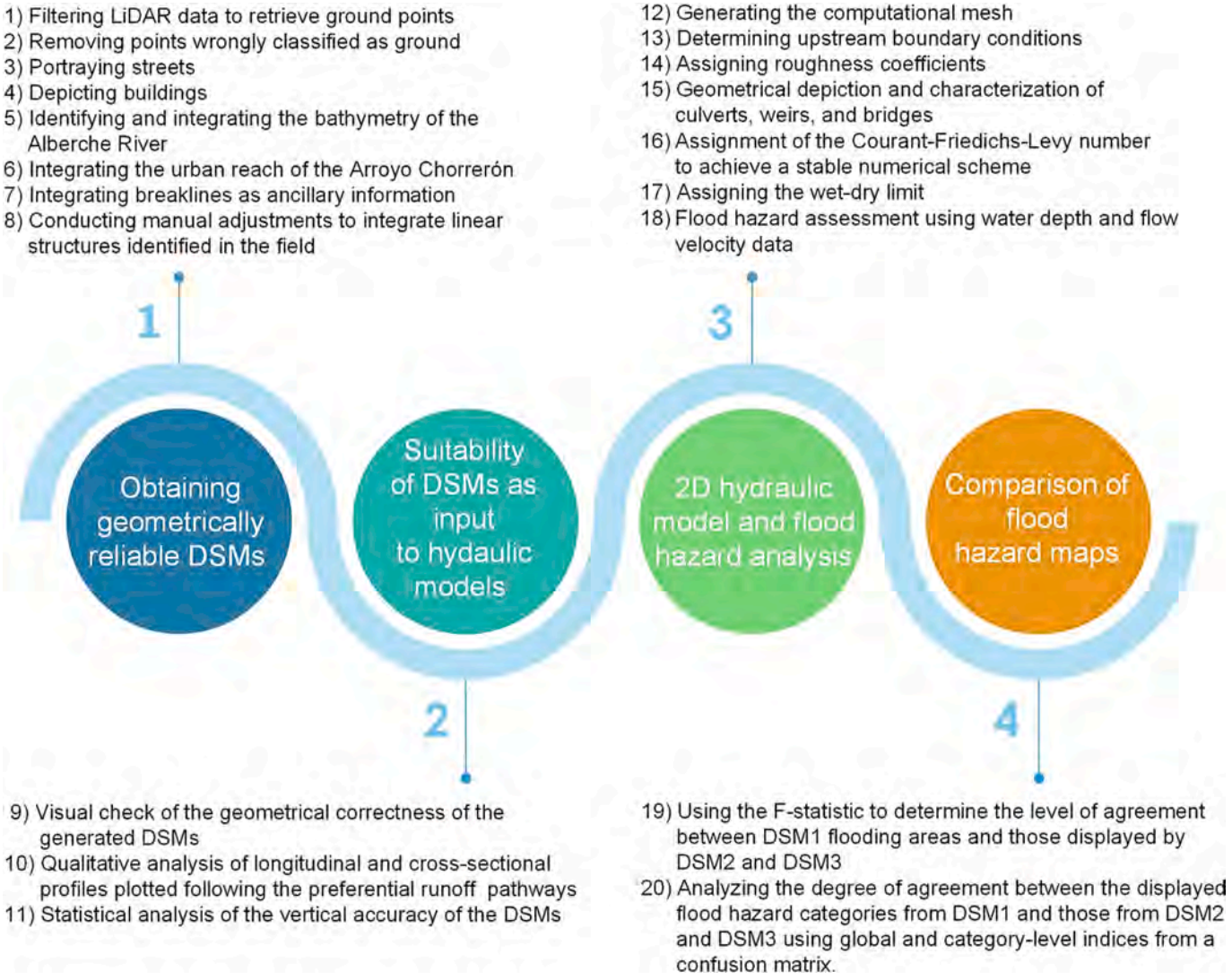

Fig. 2. Sequence of stages and sub-stages comprising the methodological strategy adopted for this research.

described.

3.1. Obtaining geometrically reliable DSMs for urban areas

When river floods affect urban areas, streets are the main runoff pathways. Accordingly, LiDAR point cloud processing was mainly focused on these areas. The first task was to filter LiDAR data to select those points that were automatically classified by the IGN as ground points. In complex urban settings, outdoor furniture (e.g., flowerpots, awnings) and other elements such as vehicles can be mistakenly categorized as ground points, which generate false barriers to the water flow once DSM is generated. To remove these "fake elements", two filters were sequentially applied: i) points with height differences of more than 0.30 m above the average were deleted; ii) points with slope differences of more than 10 % compared to the average were also omitted. Both filters were applied, taking different buffers into account. Specifically, 2 m were estimated for the filter based on the difference in height and between 1 and 5 m for the filter depending on slope.

For portraying streets, buildings were deemed to represent their outside limits. Therefore, the SEC-provided block layer was converted to points. The natural neighbour's interpolation approach was then used to get elevation from the LiDAR ground data for the points on the border of the block layer. As a frame for the street representation, the IGN-provided centre axis of each street was employed. Every 2 m, parallel lines to these central axes were drawn and converted into points. Similarly, to what was done with the block layer, the LiDAR ground data was used to assign elevation values to these points. To appropriately depict dwellings, "buildings"-designated LiDAR data points were retrieved. These points were then clipped using the SEC-provided building layer. As a proxy for roof height, the maximum point elevation per building was then extracted. The collected points were linked to the 3D points of the buildings, allowing for their accurate geometrical depiction.

Regarding the Alberche river, riverbanks were manually digitalized using the available orthophoto as a reference. Bathymetry was also collected through a field survey. To this end, a differential GPS (Trimble 5700) was used to obtain an average density of 0.3 points.m^{-2}. Regarding the Chorrerón Stream, streambanks were manually digitalized. Additionally, a couple of lines were drawn in parallel to the streambanks, matching the streambed. LiDAR points were used to assign surface elevation using the natural neighbour's interpolation method.

Once all topographic data were filtered and processed, a TIN was created. So, LiDAR and bathymetry points were included as mass points, whereas line features (e.g., riverbanks, street boundaries) were added as breaklines (hard lines) to enforce slope breaks in the TIN (Arrighi and Campo, 2019). At this phase, manual adjustments were also made to the TIN to integrate fieldwork-identified structures (e.g., walls, levees, embankments) not detected by the original LiDAR data. Errors detected in the TIN were also fixed. Last, a DSM was created (DSM1) by linearly interpolating the TIN and using a 2-metre cell size to match the point density of the original LiDAR data. Additionally, two other DSMs were produced. In the first (DSM 2), streets were portrayed by assigning breaklines using the average LAS z-value to produce 3D features. The second (DSM3), was obtained by direct interpolation of LiDAR data. This was done to determine how different the geometry exhibited by DSM2 and DSM3 is from that of DSM1, as well as to quantitatively define how the differences in the geometric arrangement of the DSMs affect the flood hazard maps that each DSM deploys. DSM1, DSM2 and DSM3 have been made freely available via the open-access repository Zenodo: https://doi.org/10.5281/zenodo.5171778 (Aroca-Jiménez et al., 2021).

3.2. Evaluation of the suitability of DSMs as topographic input to hydraulic models

The suitability of DSMs as topographic input to 2D hydrodynamic modelling was tested using both qualitative and quantitative criteria, with the latter based on statistical sampling error approaches focused on vertical accuracy. In the first approach, the geometric consistency of streets and channelling along the urban reach of the Chorrerón Stream was visually checked for each DSM. In a second phase, the conclusions of the previous analysis were corroborated by the layout of longitudinal profiles and cross-sections of the 'streets-buildings' system and the Chorrerón Stream.

A Differential Global Positioning System (dGPS) dataset covering Navaluenga and the area around it (zone 1) was used to test the vertical accuracy of DSMs. 2,088 points with vertical and horizontal accuracies $<$0.15 m were used to evaluate DSM vertical accuracy. For the other zones (all included within zone 1), the point samples were 546 (zone 2), 209 (zone 3), and 55 (zone 4). Vertical accuracy was then evaluated by computing the difference between the pixel values of the DSMs and the corresponding dGPS measurements.

The statistical distribution of errors throughout the four zones was captured by plotting histograms normalized by their respective means so the standard deviation could be visually contrasted. The error sample was further described using metrics of centrality (mean and median), dispersion (standard deviation and range) and measures of shape (kurtosis and skewness). Vertical accuracy was determined using the root mean square error (RMSE) and the mean absolute error (MAE). The coefficient of determination (R^2) was also used to assess the DSMs' goodness-of-fit to the sample of dGPS z-values.

3.3. 2D hydraulic modelling and flood hazard analysis

Hydraulic modelling was performed using the Iber two-dimensional hydrodynamic model (Bladé et al., 2014). The computational mesh was generated employing the RTIN (Regular Triangular Irregular Network) method, with maximum and minimum triangle lengths of 2 m and 1 m, respectively. Once produced, the node height of the mesh elements was adjusted from the DSMs. To do so, a threshold tolerance of 0.1 m was set, which is consistent with the altimetric accuracy of the LiDAR data used (0.15 m).

Upstream boundary conditions for the Alberche River and the Chorrerón Stream assumed steady-state subcritical flow conditions for the 500- and 25-year floods (Bodoque et al., 2016; Bodoque et al., 2020; Díez-Herrero, 2001). As a downstream boundary condition, it used a subcritical regime. The discharges used were 2006 $m^3.s^{-1}$ (500-year flood) and 856 $m^3.s^{-1}$ (25-year flood) for the Alberche River, and 167 $m^3.s^{-1}$ (500-year flood) and 10 $m^3.s^{-1}$ (25-year flood) for the Chorrerón Stream. The roughness coefficient was estimated from national land cover mapping at a scale of 1:25,000, and then allocated, in accordance with Chow (1959) the equivalent value of Manning's n to each land use/land cover unit.

Culverts and weirs operating in the study area (Fig. 3) were geometrically characterized and included as internal conditions in the 2D hydrodynamic model. The culverts' start and end locations, as well as their dimensions (width and height for rectangular culverts and diameter for circular culverts), were examined. All culverts were given a Manning's n value of 0.018. A weir with a discharge coefficient of 1.7 was deemed indicative of the Alberche River's urban reach. Also included in the computational mesh were the two bridges present in the reach (Fig. 3). Their geometry was replicated by subdividing the mesh. First, the coordinates of the upper chords' upstream ends and deck width were established. The relative coordinates and top and bottom elevations of the decks, as well as the relative location and width (crosswise to the flow direction) of the bridge piers, were to be determined next. Furthermore, the model was only affected by the bridge piles and abutments if the water level did not reach the bridge deck. When the deck height was exceeded, the coefficients for free or submerged pressure flow, as well as the weir coefficients, with values of 0.6, 0.8, and 1.7, were added.

Before launching the simulation, a second-order numerical approach was considered (Cea and Bladé, 2015). A Courant-Friedrichs-Lewy

Fig. 3. Location on orthophoto of the hydraulic infrastructures of Navaluenga and photographic detail of each one of them.

number of 0.45 was employed to achieve a stable numerical scheme. The wet-dry limit, which is the depth in metres below which an element is considered dry and is therefore left out of the hydrodynamic computation, was set at 0.01 m.

A flood hazard analysis was given by the product of water depth and velocity rasters derived from the 2D hydrodynamic modelling (Díez-Herrero et al., 2009). In this research, the following flood hazard categories were considered: i) high flood hazard (HFH); ii) medium flood hazard (MFH); iii) low flood hazard (LFH); and iv) non-flooded areas (NFA).

3.4. Comparison of flood hazard maps

The objective of this evaluation was to compare DSM1 (reference model) flood hazard maps with those from DSM2 and DSM3. The following scenarios were examined for this purpose: i) scenario1. 500-year flood generated by the Alberche River and the Chorrerón Stream; ii) scenario 2. 25-year flood induced by the Alberche River and the Chorrerón Stream; iii) scenario 3. 500-year flood triggered by the Chorrerón Stream; and iv) scenario 4. 25-year flood driven by the Chorrerón Stream. Comparative analyses were conducted in zones 2, 3 and 4 (see their description at the end of Section 2).

Two approaches were used to address the above comparison. First, using the F statistic (Cook and Merwade, 2009), the difference between the flooding areas of DSM1 and DSM2 and DSM3 was assessed. This statistic was determined by the equation:

$$F = 100.\left(\frac{Aop}{Ao + Ap - Aop}\right) \quad (1)$$

where *Ao* is the reference inundation area (derived from DSM1), *Ap* provides the area flooded by DSM2 and DSM3, and *Aop* is the overlapping flooding area for DSM1, DSM2, and DSM3. A score of 100 represents perfect agreement between DSM1, DSM2 and DSM3 flooding areas, whereas the lower the *F,* the greater the disagreement between DSM1 with DSM2 and DSM3 flooding areas.

Then, using pixel-level statistics, the agreement between the flood hazard categories from DSM1 and those from DSM2 and DSM3 was checked. A confusion matrix was devised for this purpose (Congalton, 1991; Congalton and Green, 1999). It included values showing the degree of similarity between paired observations provided by the set of flood hazard categories derived from DSM1, or the reference dataset (*RD*). Moreover, the set of DSM2 and DSM3 flood hazard categories, which were designated as the dataset under control (*DUC*), were considered.

A confusion matrix is a *MxM* (rows × columns) matrix, where *M* represents the flood hazard categories under consideration. *RH* (reference hazard) refers to RD classes, whereas *EH* relates to *DUC* classes (estimated hazard). The diagonal cells of the confusion matrix include values that correspond to well-defined flood hazard categories from DSM2 and DSM3 (*EHs* coincide with *RHs*). These cells were labelled as *CP* (i.e., pixels where there is agreement on flood hazard categories). Off-diagonal cells include values related to errors of omission and commission, which are referred to as *EC* (i.e., error or mismatch pixels in flood hazard categories) (Fig. 4).

A confusion matrix provides a thorough assessment of agreement and error distribution between flood hazard categories. However, understanding it requires indices to summarize confusion matrix information. In this research, the overall accuracy (OA) served as one of the global indices (Story and Congalton, 1986). It is the ratio between the total number of successfully categorized elements (cells of the main diagonal) and the total number of matrix cells. Moreover, the Kappa coefficient, KC (Congalton et al., 1983) was employed, which is the difference between the agreement percentage supplied by the main matrix diagonal and the chance of agreement derived from the marginal values (row and column totals; Eq. (2)). According to Landis and Koch (1977), kappa values under 0.2 imply minor, 0.2–0.4 fair, 0.4–0.6 moderate, 0.6–0.8 substantial, and 0.8–1.0 virtually perfect agreement.

$$Kappa(K) = \frac{N.\sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r}(x_{i+}.x_{+i})}{N^2 - \sum_{i=1}^{r}(x_{i+}.x_{+i})} \quad (2)$$

where:

$$N = CPpixels + ECpixels \quad (3)$$

$$\sum_{i=1}^{r} x_{ii} = \sum CPpixels \quad (4)$$

$$\sum_{i=1}^{r}(x_{i+}.x_{+i}) = (R1.C1) + (R2.C2) + (R3.C3) + (R4.C4) \quad (5)$$

Along with the global indices described above, errors of commission,

DSM2 / DSM3 (Dataset under control, DUC)	DSM2/DSM3 Hazard Categories – Estimated Hazard (EH)		DSM1 (Reference dataset, RD) – DSM1 Hazard Categories – Reference Hazard (RH): Low Hazard	Medium Hazard	High Hazard	Non-flooded	Row total
	Flooded	Low Hazard	Correctly Categorized Hazard	Underestimation	Underestimation	Omission Error (Overestimation)	R1
		Medium Hazard	Overestimation	Correctly Categorized Hazard	Underestimation	Omission Error (Overestimation)	R2
		High Hazard	Overestimation	Overestimation	Correctly Categorized Hazard	Omission Error (Overestimation)	R3
	Non-flooded		Commission Error (Underestimation)	Commission Error (Underestimation)	Commission Error (Underestimation)	Correctly Categorized No Hazard	R4
	Column total		C1	C2	C3	C4	

CP pixels
EC pixels

Fig. 4. Confusion matrix used for comparison between the flood hazard categories obtained from the reference model (DSM1) and the flood hazard categories obtained from the dataset under control (DSM2 and DSM3).

or user accuracy (UA), and errors of omission, or producer accuracy (PA), were also used as category-level indices (Congalton and Green, 1999). The user's accuracy was calculated by dividing the number of correctly categorized pixels by the total number of pixels in each flood hazard category (row total of the confusion matrix). Also, the producer's accuracy resulted from dividing the number of correctly identified pixels in each category by the number of reference pixels utilized for that category (column total of the confusion matrix).

4. Results

4.1. Geometric consistency and vertical accuracy of the DSMs developed

DSM1 defined urban-like geometry. Thus, streets and the Chorrerón Stream cross-sections revealed consistent rectangular shapes (Figs. 5 and 6). In DSM2, street cross-sections coherently represented their geometry, but unlike DSM1, the angle between streets and buildings was not orthogonal (Fig. 5). In the Chorrerón Stream, cross-sections 1 and 2 were v-shaped, defining smaller cross-sectional areas than the real ones (Fig. 6). In DSM3, street cross-sections also displayed v-shapes and saw-like shapes, in which high and low points alternated without transition, defining 6-metre elevation differences (Fig. 5, see cross-sections 1 and

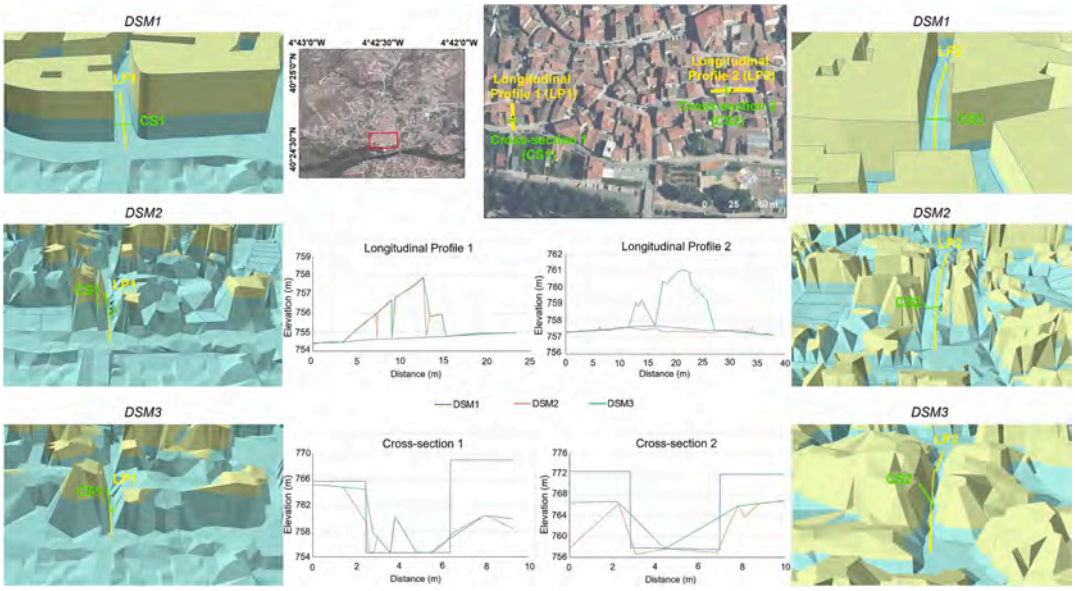

Fig. 5. Comparison between DSM1 and DSMs 2 and 3 for two cross-sections and two longitudinal profiles located on two streets within the 500-year flood prone area.

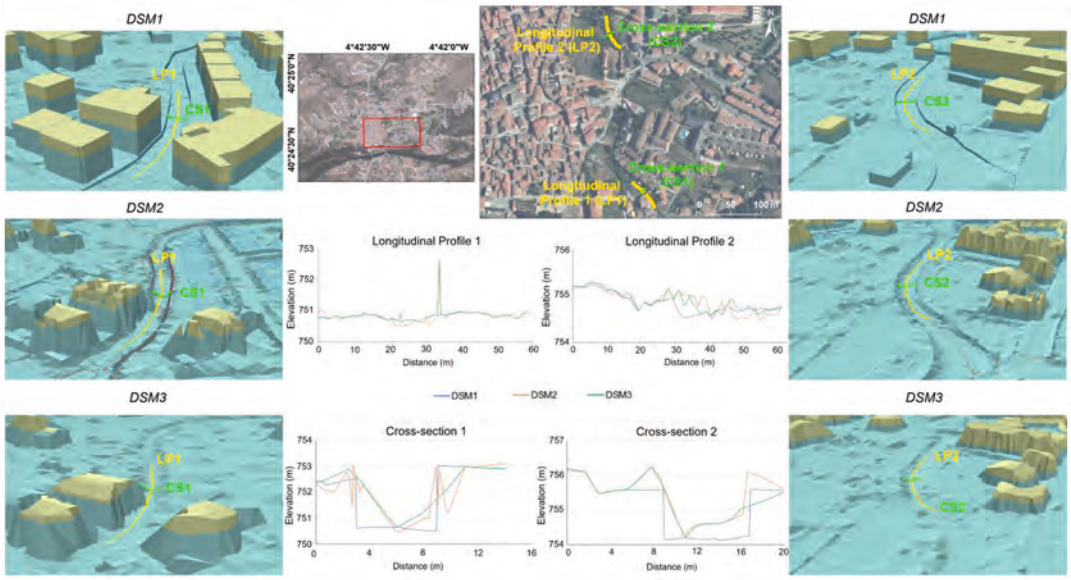

Fig. 6. Comparison between DSM1 and DSMs 2 and 3 for two cross-sections and two longitudinal profiles located in the Chorrerón Stream.

2). In the Chorrerón Stream, DSM3 cross-sections were like DSM2 ones (Fig. 6).

Regarding longitudinal profiles, DSM1 defined uniform slopes (Fig. 5). However, DSM2′s profile pattern was irregular. Thus, longitudinal profile 2 had a consistent slope, but longitudinal profile 1 had height variances of up to 3 m (Fig. 5). In DSM3, longitudinal profiles displayed erratic morphologies that do not match street patterns in cities (Fig. 5). In the Chorrerón Stream, longitudinal profiles from the three DSMs showed more resemblance than in streets. Thus, they displayed uneven patterns with altimetric fluctuations below half a meter, but sometimes exceeding 1.5 m (Fig. 6, see longitudinal profile 1).

Concerning the vertical accuracy of the DSMs, Table S1 of the supplementary data displays the vertical accuracy of the DSMs by comparing them to cadastral data over the four zones. For DSM1, the four zones had coefficients of determination values above 0.99. DSM2 had comparable results. For DSM3, the coefficient of determination in

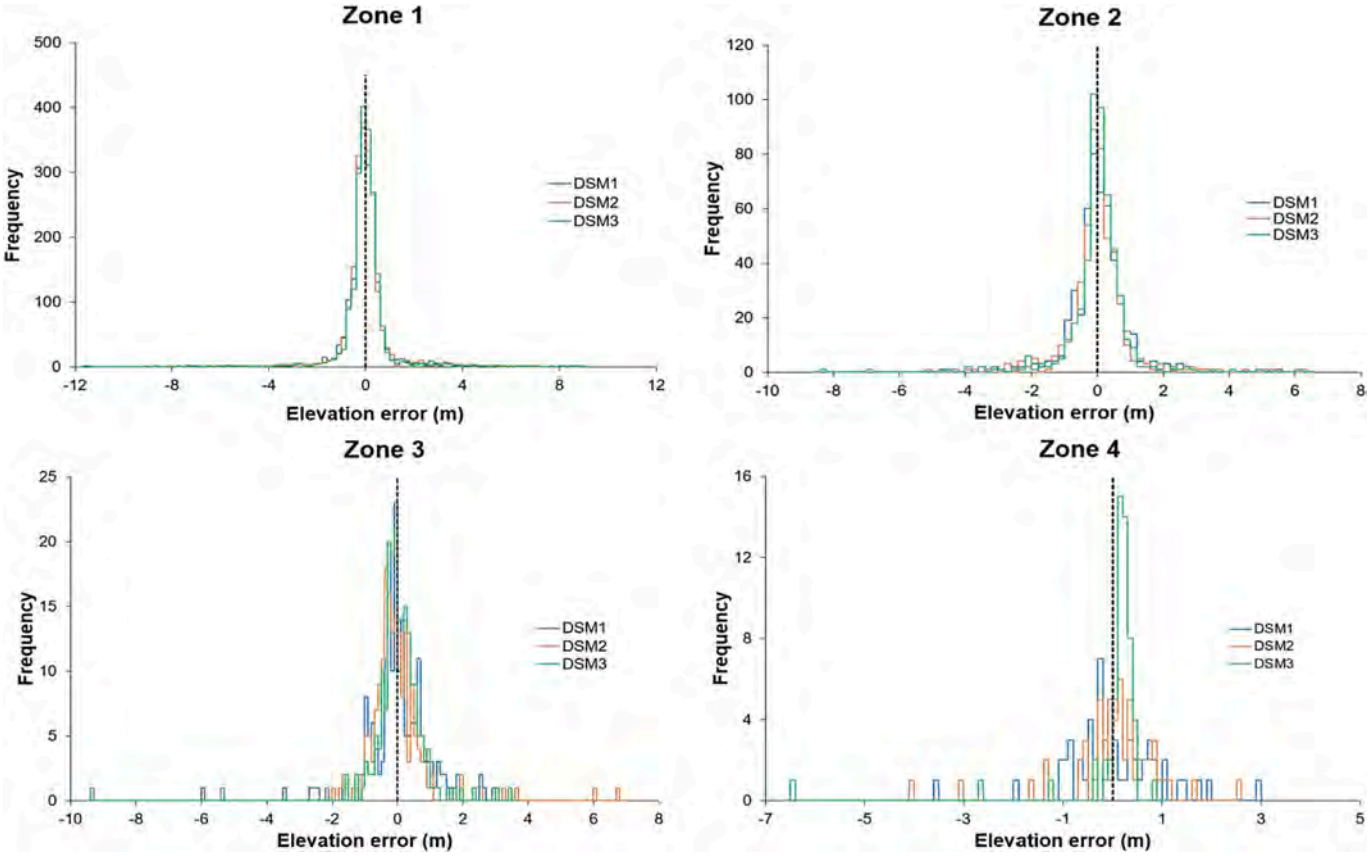

Fig. 7. Histograms of the normalized elevation errors estimated in the three DSMs.

zone 4 decreased below 0.95, suggesting a worse representation of this model in streets with higher complexity. Based on centrality measures, DSM1 had higher mean and median errors in Zones 1–3 than DSM2 and DSM3. In zone 4, however, DSM1 had lower mean and median errors than DSM3. DSM2 performed best in all four zones. Alternative error measures, such as RMSE or MAE, tended to confirm these findings, although RMSE identified DSM1′s advantage over the other two DSMs in zone 4.

Dispersion and shape metrics of elevation errors show that DSM1 had the lowest elevation error range in zones 1 and 4, and second lowest in zones 2 and 3 (very close to DSM2). DSM1 had the lowest standard error and standard deviation in zone 4, indicating reduced elevation error fluctuation. Concerning kurtosis and skewness, DSM1′s relative elevation errors displayed a closer-to-normal distribution (Fig. 7).

4.2. Hydraulic outcomes

Fig. 8, in relation to zone 2, depicts water depths, flow velocities, and Froude numbers for DSM1, DSM2, and DSM3 resulting from the 500-year flood. Upstream of the Chorrerón Stream-Alberche River junction, the three DSMs produced similar hydraulic outcomes. However, in the junction's vicinity and downstream, hydrodynamic response changed significantly.

Thus, DSM2 and DSM3 disrupted water flow patterns, as evidenced by the sudden depth and Froude number fluctuations noticed in the Alberche River. DSM3 water depths ranged from 3.56 to 7.39 m. In DSM2, water depths were between 3.91 and 9.20 m, although the longitudinal profile was more homogeneous. Sharp changes in DSM3′s longitudinal water depth profile implied that the Froude number shifted from less than 1 to greater than 1. These changes in the Froude number also occurred in DSM2 but were less pronounced. In contrast, DSM1 hydraulic outputs revealed a homogeneous longitudinal water depth profile (7.09 to 9.35 m), which is why changes in the Froude number were not significant (See A in Fig. 8).

As regards the Chorrerón Stream (zone 3), the hydraulic response obtained from DSM1 was more homogenous. Thus, higher, and more uniform water flow velocities (0.71–1.50 m/s) were attained. In contrast, DSM2 and DSM3 exhibited much fewer uniform velocities in the Chorrerón Stream, as velocity was noticeably reduced. Thus, DSM2 and DSM3 displayed minimum velocities of 0.40 m/s and 0.18 m/s, respectively (see B in Fig. 8). Flooding downstream of the Navaluenga's national road (see E in Fig. 8) affected the urban area differently in the three DSMs (see C in Fig. 8). So, water flow follows the street network in DSM1, whereas in DSM2 and DSM3, water flow discontinuities were identified.

Fig. 9 illustrates the flood hydraulic performance of the three DSMs at the Chorrerón-Alberche confluence during the 500-year flood (zone 3). Again, DSM1 had more consistent hydraulic outputs, with the main channel's physical limits readily discernible (see A in Fig. 9) and floods matching the building perimeters. DSM2, and particularly DSM3, had erratic depths and velocities. Bridges in DSM3 lacked water-flow spans, preventing water from travelling through. As a result, DSM3 bridges functioned similarly to weirs crossing the river channel (see B in Fig. 9), limiting the river's capacity to evacuate floods. This caused a rise in depths upstream, resulting in lateral overflows and faking floods (see C in Fig. 9). When the bridge deck was exceeded, downstream velocities rose, and a critical regime was generated (see D in Fig. 9). This increase in velocities was less noticeable in DSM1 since there was only a small acceleration while crossing the bridge spans. DSM2 exhibited a hydraulic performance comparable to DSM1 (see E in Fig. 9).

The right bank of the Chorrerón Stream, upstream of its Alberche River junction, had the most changes (zone 4). Depending on the DSM, the flood extent in this area (see D in Fig. 8) varied substantially. Almost all of zone 4′s streets were flooded in DSM1 and DSM2. However, DSM2

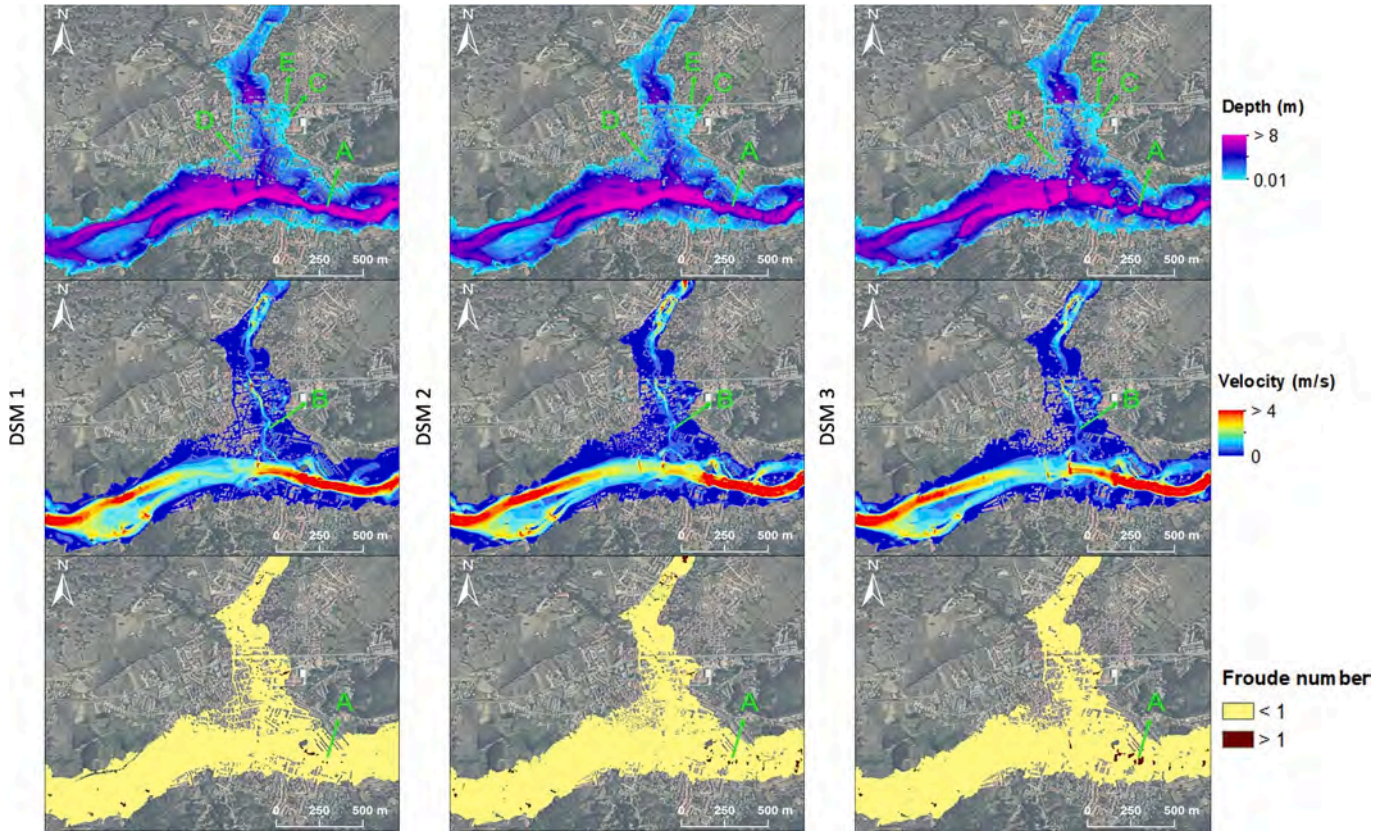

Fig. 8. Outputs of water depths, flow velocities and Froude numbers obtained after considering the 500-year flood as the upstream boundary condition. The capital letters included in this figure show some locations where these hydraulic outputs show relevant differences in each of the DSMs considered.

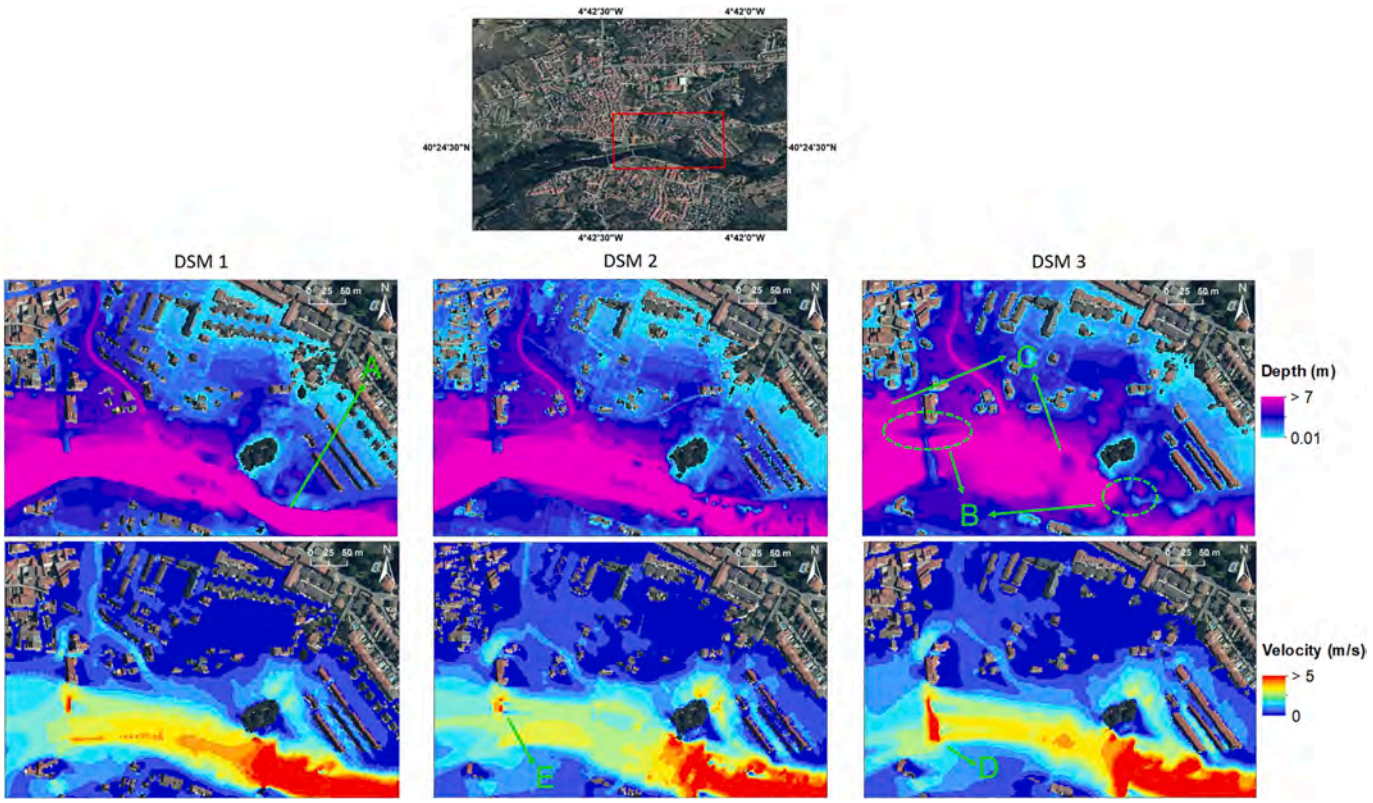

Fig. 9. Hydraulic outputs (water depth and flow velocity) at the junction of the Chorrerón Stream and the Alberche River (zone 3). The capital letters indicate the locations where these hydraulic outputs vary significantly depending on the DSM used.

results misrepresented the flooded area and buildings. DSM3′s flood extent for zone 4 was significantly lower than that of DSM1 and DSM2.

Fig. 10 details the differences described above for the 500-year flood. It shows two benchmark cross-sections (CS1 and CS2) located in zone 4. In CS1, DSM1 had depths close to 4.0 m, while DSM3 depths were 1.5 m to 3.0 m. In DSM2, CS1 depths ranged from 0.50 to 3.86 m. In CS2, there was no water flow through DSM2 and DSM3, while in DSM1 there was a water flow across CS2 deploying depths between 0.44 m and 0.72 m.

Regarding flow velocity, the results obtained in CS1 for all three DSMs were close to 0 m/s. DSM1 velocities in CS2 ranged from 0.54 to 0.71 m/s; however, DSM2 and DSM3 did not deploy flow through CS2. In DSM1 and DSM2, flow velocities were less than 1 m/s; yet, in some streets of DSM1, the flow accelerated continuously and in an orderly manner, reaching 0.37 m/s to 1.28 m/s. In DSM2, water flow was sped up in several streets to velocities comparable to DSM1. In contrast to DSM1, DSM2 displayed a discontinuous and chaotic maximum velocity

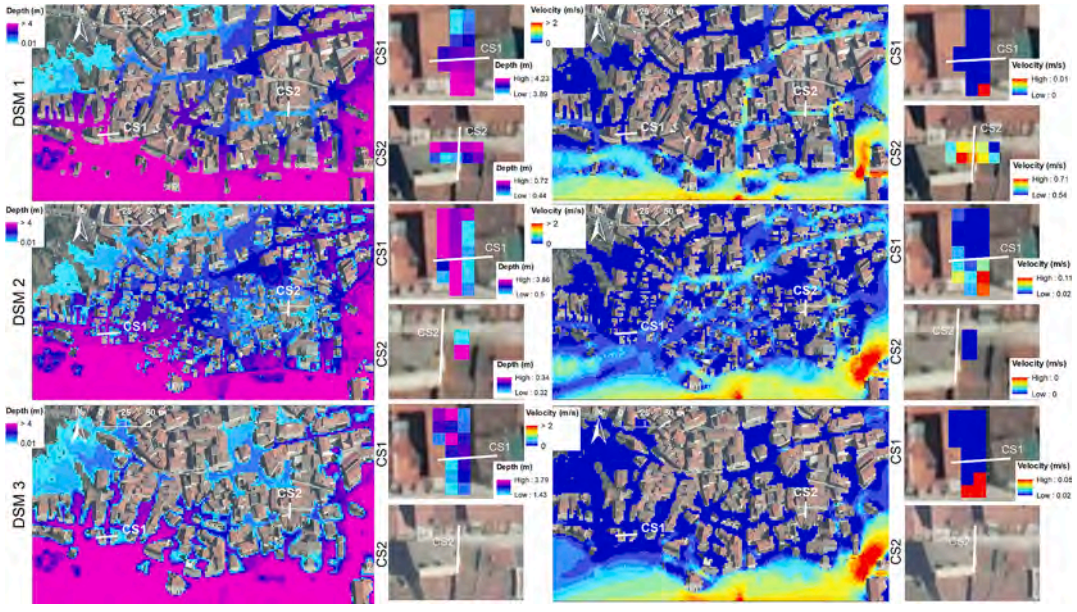

Fig. 10. Water depths and flow velocity patterns determined from DSM1, DSM2, and DSM3 in zone 4 and during the 500-year flood. The description focuses on how these hydraulic outputs behave at cross-sections CS1 and CS2.

pattern. DSM3 performed differently from DSM1 and DSM2 because it mostly deployed flow velocities around 0 m/s.

The respective portrayals of the Chorrerón Stream 500-year flood by the three DSMs were comparable (see Fig. S1 of the supplementary data). For the 25-year flood, however, the DSMs differed significantly. Fig. 11 illustrates these differences and as shown in Fig. 10, two cross-sections (CS1 and CS2) were used as a benchmark for comparison. Around CS1, DSM2 and DSM3 had a significant right-bank overflow, but DSM1 had essentially none. In CS2, DSM1 depths revealed a clear contrast between channel flow (1.72 m) and bank flooding (a few centimetres). This pattern was also noticeable in DSM2 and DSM3, although downstream of CS2, flooded urban area depths were 1.3 m, whereas in DSM1 the depth was 0.8 m (see B in Fig. 11). DSM1 had the best water flow continuity, reducing lateral overflows into surrounding streets. This meant, for example, that a street was flooded according to DSM3 and to some extent DSM2, but not DSM1 (see A in Fig. 11). Concerning flow velocities, maximum values were obtained in DSM1 (CS1 = 1.64 m/s; CS2 = 1.11 m/s) while the lowest ones were reported in DSM3 (CS1 = 0.96 m/s; CS2 = 0.71 m/s). The simulated flow velocities in DSM2 ranged from 0.68 m/s (CS2) to 1.02 m/s (CS1).

The hydraulic outputs of water depth and flow velocity displayed by each DSM, as well as the resulting flood hazard maps, are available for unrestricted download in the open access Zenodo repository: https://doi.org/10.5281/zenodo.5171778 (Aroca-Jiménez et al., 2021).

4.3. Level of agreement on flooding outcomes elicited from DSM1, DSM2 and DSM3

4.3.1. Matching degree of flooded areas

For scenario 1, in zone 2, DSM2 and DSM3 set for higher matching with DSM1, since the F-statistic in both DSMs was more than 87 %. In zone 3, the F-statistic defined values of 80 % in DSM3 and 86 % in DSM2. In zone 4, DSM3 scored badly (F statistic = 64 %), while DSM2 matched similarly in zones 2 and 3 (F statistic = 82 %). For scenario 2, in zone 2, the F-statistic varied between 82 % (DSM3) and 86 % (DSM2). In zone 3, the F-statistic dropped to 68 % (DSM3) and 74 % (DSM2). In zone 4, the matching level of DSM2 was the worst (F statistic = 59 %), while the agreement degree of DSM3 also declined (F statistic = 62 %).

For scenario 3, DSM2 performed better than DSM3. So, in zone 2, the F-statistic ranged between 71 % (DSM3) and 79 % (DSM2). In zone 3, DSM3 delivered an F-statistic of 68 %, while the F-statistic provided by DSM2 was 76 %. In zone 4, the F-statistic corresponding to DSM3 decreased to 36 %, while the match percentage supplied by DSM2 was 66 %. For scenario 4, in zone 2, the F-statistics ranged between 63 % (DSM3) and 72 % (DSM2), whereas in zones 3 and 4, the F-statistics were 48 % and 62 %, respectively (Table 1).

4.3.2. Degree of agreement between flood hazard outcomes at the pixel level

The confusion matrices can be consulted in Tables S2–S5 of the supplementary data. They contain estimates of global indices (i.e., OA and KC) and category-level indices (i.e., UA and PA).

4.3.2.1. Scenario 1. Zone 2 had the highest OA, with 72 % in DSM2 and DSM3 and a KC of 0.46 in both DSMs (moderate level of agreement). Zone 4 had the lowest OA, with 31 % (DSM3) and 33 % (DSM2), with KC of 0.06 (DSM3) and 0.11 (DSM2), showing a minor level of agreement. In zone 3, the OA for DSM2 and DSM3 was 59 % and 57 %, respectively. The KC for DSM2 and DSM3 were 0.40 and 0.37 (moderate and fair level of agreement).

Regarding the HFH category, zone 2 had the best agreements. PA and UA specified values exceeding 90 % for the HFH category in DSM2 and DSM3. In zone 3, DSM2 produced 83 % and 84 % concerning UA and PA, whereas DSM3 delivered 78 % (UA) and 84 % (PA). In zone 4, DSM2 indicated that the UA was 53 % and the PA was 65 %. The results of DSM3 were 70 %(UA) and 50 % (PA).

For the remaining flood hazard categories in zone 2, DSM2 recorded UAs ranging from 41 % (MFH) to 43 % (LFH) and PAs from 40 % (LFH) to 52 %. (MFH). DSM3 defined UAs from 42 % (MFH) to 43 % (LFH) and PAs from 40 % (LFH) to 48 %. (MFH). In zone 3, DSM2 and DSM3 reported agreements like those of zone 2. Compared to zones 2 and 3, DSM2 agreements for LFH and MFH categories were poorer in zone 4, as UAs were 25 (MFH) to 38 % (LFH) and PAs were 33 % (LFH) to 41 % (MFH). Regarding DSM3, LFH agreements in zone 4 were like those in zones 2 and 3. Lower values were recorded for the MFH category (UA = 25 %; PA = 26 %), while the HFH category reported UA and PA outcomes of 70 % and 50 %, respectively.

These disagreements resulted in different flood hazard categories in DSM1, DSM2, and DSM3. Thus, DSM1 and DSM3 overlapped flood hazard categories in 72 % of pixels in zone 2, 57 % in zone 3, and 31 % in zone 4. In zones 2, 3, and 4, where flood hazard categories did not

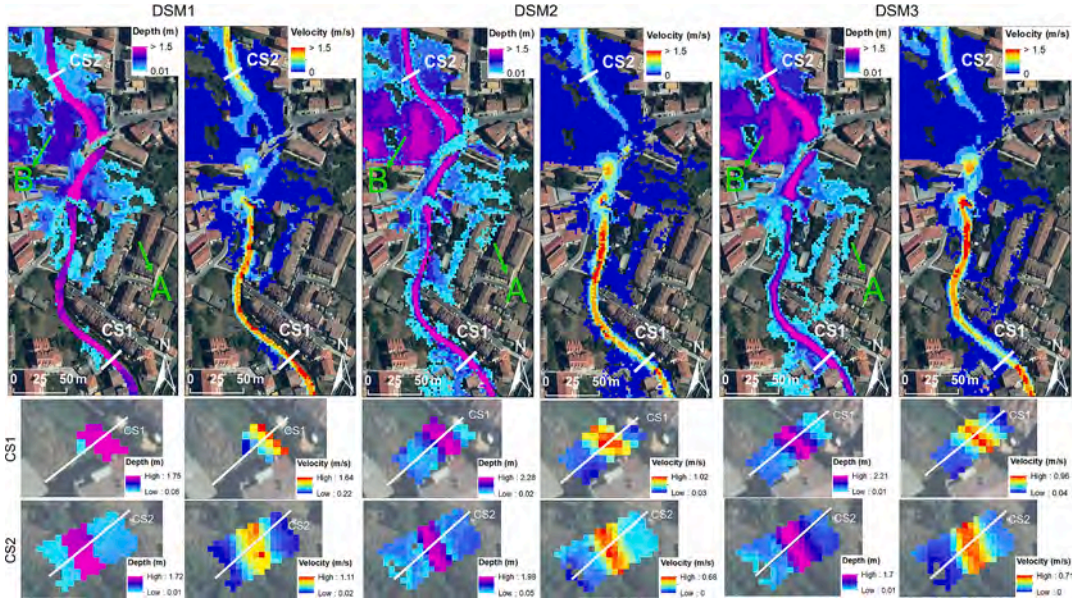

Fig. 11. Pattern that follows water depths and flow velocities from DSM1, DSM2, and DSM3 in the Chorrerón Stream and for the 25-year flood. The focus of the description is on how these hydraulic outputs behave at the cross-sections CS1 and CS2.

Table 1
Comparative analysis of the flooded areas derived from the three DSMs using the F-statistic as a quantitative metric.

				DSM1	DSM2	DSM3
500-year flood	Alberche River and Chorrerón Stream (scenario 1)	Flood-prone area (m^2)	Zone 2	828,325	841,531	850,094
			Zone 3	100,344	96,189	100,183
			Zone 4	18,887	17,108	14,121
		F statistic (%)	Zone 2		87.18	87.51
			Zone 3		85.82	80.18
			Zone 4		81.51	64.04
	Chorrerón Stream (scenario 3)	Flood-prone area (m^2)	Zone 2	292,029	309,608	324,402
			Zone 3	45,089	43,451	42,053
			Zone 4	8,498	9,946	3,962
		F statistic (%)	Zone 2		78.96	71.34
			Zone 3		76.02	67.89
			Zone 4		66.49	35.75
25-year flood	Alberche River and Chorrerón Stream (scenario 2)	Flood-prone area (m^2)	Zone 2	528,214	525,436	572,239
			Zone 3	37,475	37,230	48,772
			Zone 4	4,570	4,485	5,041
		F statistic (%)	Zone 2		86.30	82.42
			Zone 3		74.47	67.72
			Zone 4		58.63	62.16
	Chorrerón Stream (scenario 4)	Flood-prone area (m^2)	Zone 2	79,702	88,473	96,095
			Zone 3	7,187	10,234	10,234
			Zone 4	1,011	1,018	858
		F statistic (%)	Zone 2		71.74	63.49
			Zone 3		47.64	48.91
			Zone 4		62.84	60.45

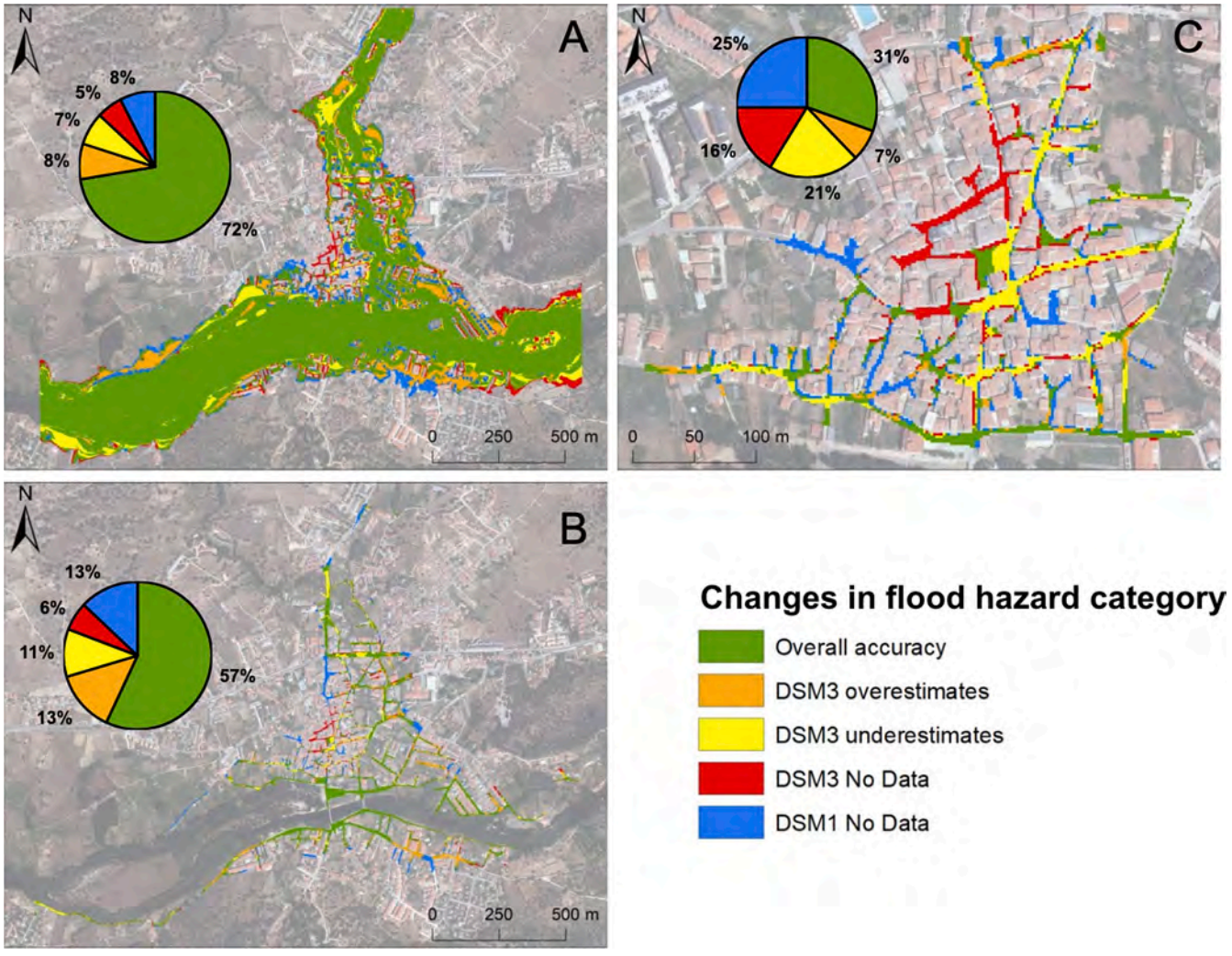

Fig. 12. Level of coincidence for scenario 1 in flood hazard categories mapped from DSM1 and DSM3. The capital letters included in the figure correspond to zone 2 (A); zone 3 (B) and zone 4 (C).

overlap, pixels mirrored those of locations where neither DSM1 nor DSM3 deployed floods, i.e., 13 % in zone 2, 19 % in zone 3, and 41 % in zone 4 (Fig. 12). In addition, 7 % of the pixels in zone 2, 11 % in zone 3, and 21 % in zone 4 underestimated the flood hazard category provided by DSM1. Conversely, 8 % of the pixels in zone 2, 13 % in zone 3, and 7 % in zone 4 overestimated the flood hazard category provided by DSM1. In scenario 1, DSM2 yielded findings that were comparable with DSM3 (Fig. S2 of the supplementary data).

4.3.2.2. Scenario 2. DSM2 gave comparable OAs and KCs for zones 2, 3, and 4 as in scenario 1. DSM3 showed similar outcomes for zones 2 and 3 as in scenario 1. In zone 4, DSM3 showed more agreement (OA = 42 %; KC = 0.17; minor agreement) than scenario1. Comparing DSM2 with DSM3, it was found that DSM3 reported OAs between 42 % and 64 % and KCs between 0.17 (minor agreement) and 0.42 (moderate agreement) for zones 2, 3, and 4, like those derived from DSM2 (OAs between 39 % and 71 % and KCs between 0.11 (minor agreement) and 0.52 (moderate agreement).

At the flood hazard category level, DSM2 and DSM3 provided similar outcomes to those of scenario 1 in zone 2 and for the HFH category. For the remaining flood hazard categories in zone 2, DSM2 produced a stronger agreement (UAs and PAs were between 49 % and 57 %) than under scenario 1 (UAs and PAs were between 40 % and 43 %), whereas DSM3 performed similarly in scenario 2 as it did in scenario 1. In zone 3, DSM2 revealed a greater agreement in the LFH category for scenario 2 (UA = 54 % and PA = 46 %) than for scenario 1 (UA = 46 % and PA = 42 %). In MFH and HFH, DSM2 scored lower in scenario 2 (UAs and PAs between 39 % and 84 %) than in scenario 1 (UAs and PAs between 51 % and 84 %). DSM3, in zone 3, provided similar agreements in scenarios 1 and 2. In the MFH category, UA in scenario 2 (27 %) was lower than in scenario 1 (46 %).

DSM2 reported higher percentages of UA and PA in scenario 2 than scenario 1 for zone 4. In HFH, the estimated percentage of PA in scenario 2 (53 %) was lower than in scenario 1 (64 %). DSM3 showed that LFH and MFH UAs and PAs were higher in scenario 2 than in scenario 1. For HFH, the UA estimated for scenario 2 (51 %) was lower than scenario 1 (70 %), while the PA was higher in scenario 2 (69 %).

Compared to DSM2 and DSM3, DSM2 reported higher PAs and UAs in zone 2 and for all hazard categories. For zone 3, the same conclusion was obtained as for zone 2, except that the percentage of PA related to HFH was higher in DSM3 (89 %) than in DSM2 (73 %). In zone 4, DSM3 exhibited greater PA and UA in the LFH category than DSM2. DSM2 performed better in the MFH (UA = 49 %; PA = 43 %) than DSM3 (UA = 41 %; PA = 38 %). In the HFH category, DSM3 deployed a greater PA than DSM2 (69 % vs 53 %), whereas DSM2′s UA (88 %) was higher than DSM3′s (51 %).

In zones 2 and 3, DSM2 had the largest proportion of matches with DSM1 (71 % and 50 %, respectively). In zone 4, this percentage was reduced to 39 %. Most pixels (14 % in zone 2, 26 % in zone 3, and 40 % in zone 4) corresponded to locations where DSM2 did not record flooding but DSM1 did, or vice versa. The remaining percentages (15 %

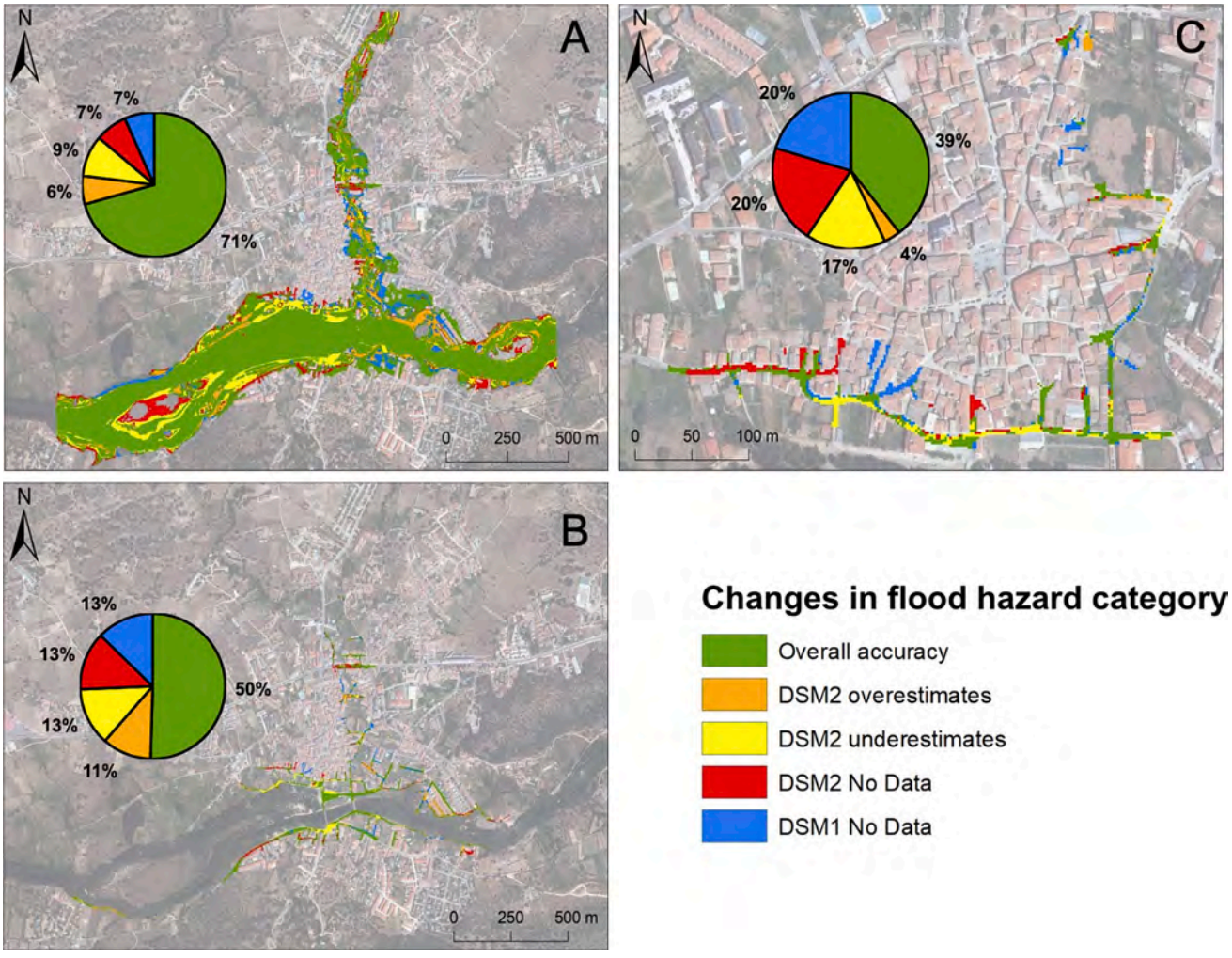

Fig. 13. Level of coincidence for scenario 2 in flood hazard categories mapped from DSM1 and DSM2. The capital letters included in the figure correspond to zone 2 (A); zone 3 (B) and zone 4 (C).

in zone 2, 24 % in zone 3, and 21 % in zone 4) pertain to places where DSM2 overestimated or underestimated flood hazard categories compared to DSM1 (Fig. 13). Zone 2 had the largest proportion of DSM3 matches (64 %) compared to zones 3 and 4, each of which had 42 %. The most mismatched pixels in DSM1 and DSM3 belonged to locations where DSM1 did not record flooding but DSM3 did (13 % in zone 2, 28 % in zone 3, and 24 % in zone 4), while places where DSM1 reported flooding but DSM3 did not do so represented 5 % in zones 2 and 3 and 13 % in zone 4. In zone 2, 18 % of DSM3 pixels overestimated or underestimated the flood hazard categories that DSM1 displayed. This percentage rose to 25 % in zone 3 and 21 % in zone 4 (Fig. S3 of the supplementary data).

4.3.2.3. Scenario 3. OA in zone 2 was 63 % in DSM2 and 51 % in DSM3. DSM2 had a KC of 0.46 (moderate agreement) in this zone, whereas DSM3 yielded a KC of 0.32 (fair agreement). The estimated percentages of OAs in zone 3 were 55 % (DSM2) and 47 %. (DSM3). In both cases, estimated KCs of 0.39 (DSM2) and 0.29 (DSM3) indicated fair agreement. In zone 4, OAs were 46 % in DSM2 and 18 % in DSM3. In DSM2, the reported KC was 0.23 (fair agreement), but in DSM3, it was 0.03 (minor agreement).

In zone 2, DSM2 and DSM3 reported UAs and PAs ranging from 81 % to 88 % when HFH was considered. DSM2 provided UAs and PAs between 52 % and 60 % for the remaining flood hazard categories, whereas DSM3 varied between 32 % and 50 %. In zone 3, UAs and PAs for DSM2 and DSM3 HFH category fluctuated between 70 % and 81 %. DSM2 displayed more agreement than DSM3 for the other flood hazard categories, with UAs and PAs defining outcomes comparable to zone 2. DSM3 yielded a higher UA than DSM2 when considering HFH in zone 4: 82 % vs 54 %. The PA was higher in DSM2 than in DSM3: 69 % vs 55 %. DSM2 UAs and PAs for MFH and LFH categories were between 49 % and 61 %, whereas DSM3 outputs were between 12 % and 41 %.

Due to disagreements between DSM1 and DSM3, match percentages in zones 2 and 3 were lower (51 % and 47 %, respectively) than between DSM1 and DSM2 (63 % and 55 %). In zone 4, matching between DSM1 and DSM3 was much lower, with 18 % of pixels identifying the same flood hazard category, compared to 46 % for DSM1 and DSM2 (Fig. 14 and Fig. S4 of the supplementary data). For zone 2, mismatches were caused by pixels that did not record flooding in DSM3 (10 %) but were inundated in DSM1 or pixels that did not report flooding in DSM1 (19 %) but were flooded in DSM3. In zones 3 and 4, non-overlap was owing to the absence of floods in DSM3 (19 % and 56 % of pixels, respectively). Regarding DSM2, 13 % of zone 2 pixels that did not match with DSM1 were due to the absence of floods in DSM1. This proportion fell to 11 % in zone 3, whereas 14 % reported flooding in DSM1 but not DSM2. 24 % of non-overlapping pixels in zone 4 showed flooding in DSM2 but not DSM1 (Fig. 14 and Fig. S4 of the supplementary data).

In zone 2, around 9 % of pixels in DSM2 and DSM3 underestimated the flood hazard category described by DSM1, whereas 8 % and 11 % of the remaining pixels in DSM2 and DSM3 overestimated it. In zone 3, 12 % of DSM2 pixels underestimated the flood hazard category compared to DSM1. When DSM3 outcomes were analysed, this percentage increased to 13 %. 8 % of DSM2 and DSM3 pixels overestimated the flood hazard category. In zone 4, 4 % of the DSM3 pixels overestimated the flood hazard category compared to DSM1, increasing to 12 % when

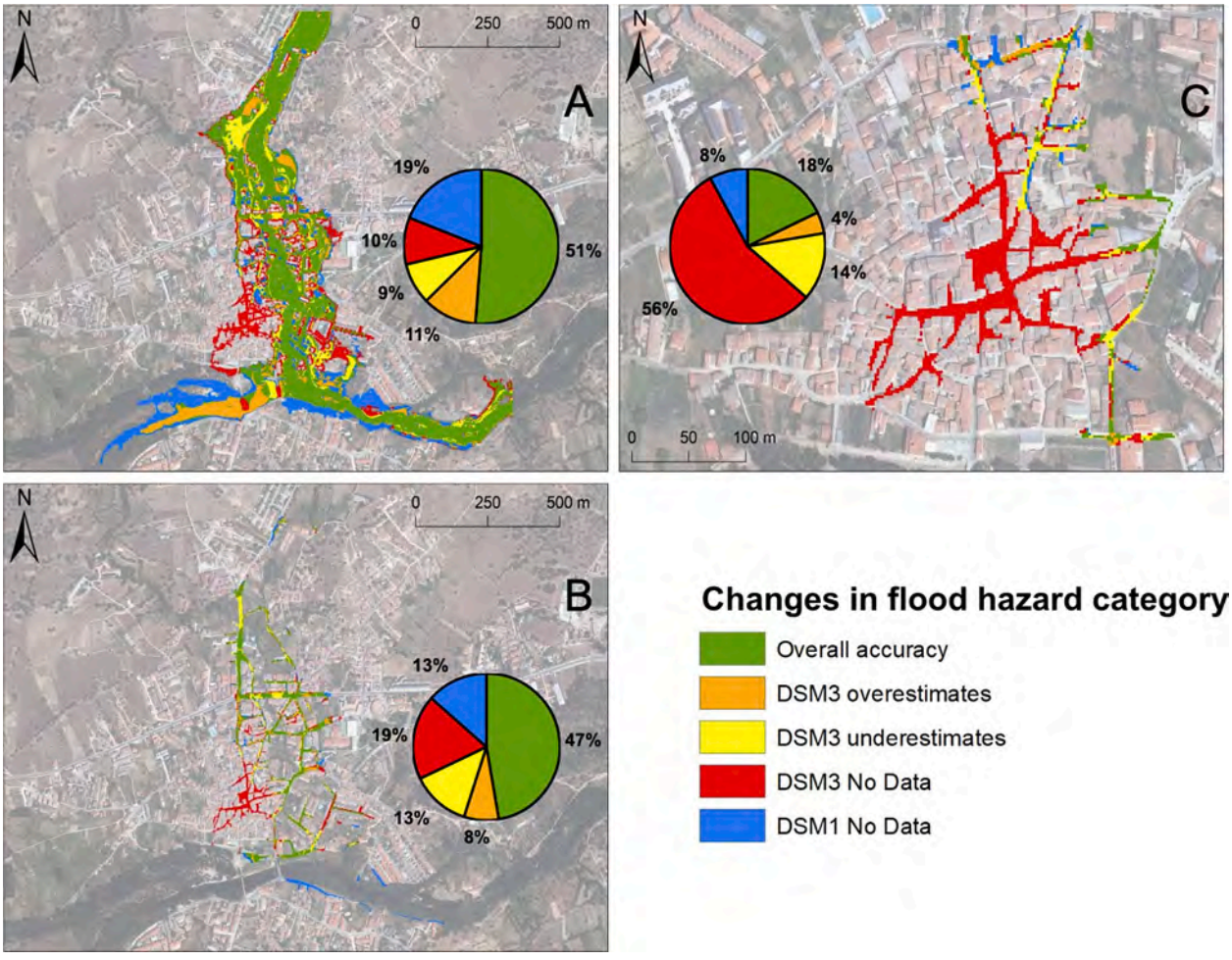

Fig. 14. Level of coincidence for scenario 3 in flood hazard categories mapped from DSM1 and DSM3. The capital letters correspond to zone 2(A); zone 3 (B) and zone 4 (C).

DSM2 results were reviewed. Besides, 8 % of DSM2 pixels underestimated the flood hazard category provided by DSM1, increasing to 14 % with DSM3 (Fig. 14. and Fig. S4 of the supplementary data).

4.3.2.4. Scenario 4. For scenario 4, neither DSM2 nor DSM3 showed pixel flooding in zone 4. In this scenario, DSM2 performed worse in zones 2 and 3 than it did in scenario 3. Consequently, OAs ranged from 44 % to 54 % and KCs were between 0.06 (minor) and 0.32 (fair), while in scenario 3, OAs ranged from 55 % to 63 % and KCs were between 0.39 (fair) and 0.46. (moderate). DSM3 provided comparable results in zone 2 as scenario 3, i.e., OA specified 51 % and KC output was 0.27. (fair). In zone 3, DSM3 scored poorer than in scenario 3, with an OA of 34 % and a KC of 0.04 (minor) compared to 47 % (OA) and 0.29 (KC–fair) for scenario 3. Comparing the performance of DSM2 and DSM3, DSM2 performed better in zones 2 and 3 (the average OA was 49 %, while KC defined values within the moderate agreement range) than DSM3, which deployed OAs ranging from 34 % (zone 3) to 51 % (zone 2), while KC defined a value of 0.04 (minor) in zone 3 and 0.27 (fair) in zone 2.

DSM2 performed better than DSM3 for the HFH category, with UAs and PAs between 28 % and 76 % compared to 14 % and 61 % for DSM3. These percentages were lower than those projected for the same category of flood hazard in scenario 3 (UASs and PAs ranged between 80 % and 88 %). For the LFH and MFH categories, DSM2 generated better results in zone 2 (UAs and PAs ranged from 59 % to 71 %) than scenario 3 (UA and PA set percentages ranged from 52 % to 60 %). In zone 3, DSM2 showed similar performance in scenarios 3 and 4 for LFH and MFH, defining PAs and UAs with average percentages of 57 % and 53 %, respectively.

Regarding the LFH category, DSM3 performed worse in scenario 3 (PA = 34 %; UA = 32 %) than in scenario 4 (PA = 60 %; UA = 57 %) in zone 2. Also, in zone 2, scenario 4 (PA = 70 %; UA = 57 %) produced a better agreement than scenario 3 (PA = 50 %; UA = 39 %) when the MFH category was addressed. In zone 3 and for the MFH category, the PA was greater in scenario 4 (PA = 66 % vs 55 % in scenario 3), while the UA was higher in scenario 3 (54 % vs 32 % in scenario 4). In Zones 2 and 3 and for LFH and MFH, DSM2 showed better agreement with DSM1 (UAs and PAs between 41 % and 71 %) than DSM3 (UAs and PAs between 32 % and 70 %).

When the disagreements described in the previous paragraphs are presented spatially (Fig. 15 and Fig. S5 of the supplementary data), DSM2 and DSM3 exhibited similar OAs in zone 2 compared to DSM1 (between 51 % and 54 %). Where there was disagreement, around 19 % of the flood hazard projections from DSM2 and DSM3 were not represented in DSM1. In contrast, 11 % of pixels were flooded in DSM1 but not in DSM2 or DSM3. On the other hand, roughly 8 % of pixels in DSM2 and DSM3 underestimated the flood hazard category in DSM1, while about 10 % overestimated it. In zone 3, DSM2 outperformed DSM3, defining an OA of 44 % as opposed to 34 % for DSM3. In addition, 23 % and 32 %, respectively, of the pixels flooded by DSM2 and DSM3 were not flooded by DSM1. In contrast, whereas 10 % of pixels in DSM2 overestimated the flood hazard category relative to DSM1 outcomes, this percentage increased to 18 % when DSM3 outcomes were evaluated. Moreover, 3 % of DSM3 pixels underestimated the flood hazard category, as opposed to 6 % of DSM2 pixels. Compared to DSM1, 17 % of DSM2 pixels and 13 % of DSM3 pixels were not flooded, while pixels that were flooded in DSM2 and DSM3, but not in DSM1, accounted for 23 % and 32 %, respectively.

5. Discussion

Highly accurate DSMs can be generated utilizing LiDAR datasets, which offer high point spatial density and vertical accuracy. Accordingly, to acquire trustworthy topographic data on a large scale, LiDAR data is suggested (Hohenthal et al., 2011). However, like any discrete measurement technique, LiDAR data has object space ambiguities, which is relevant in urban settings, where buildings and streets, along with the floodplain, represent a complicated landscape that can only be accurately depicted by employing ancillary information, e.g., through the addition of breaklines to a triangulated irregular network, TIN (Bodoque et al., 2016). Another drawback of LiDAR data lies in the errors caused by automated point cloud classification techniques. These inaccuracies may lead to elements such as flowerpots, awnings, or vehicles being wrongly classified as ground (Shan and Sampath, 2005). This may generate false obstructions to water flow, which leads to errors

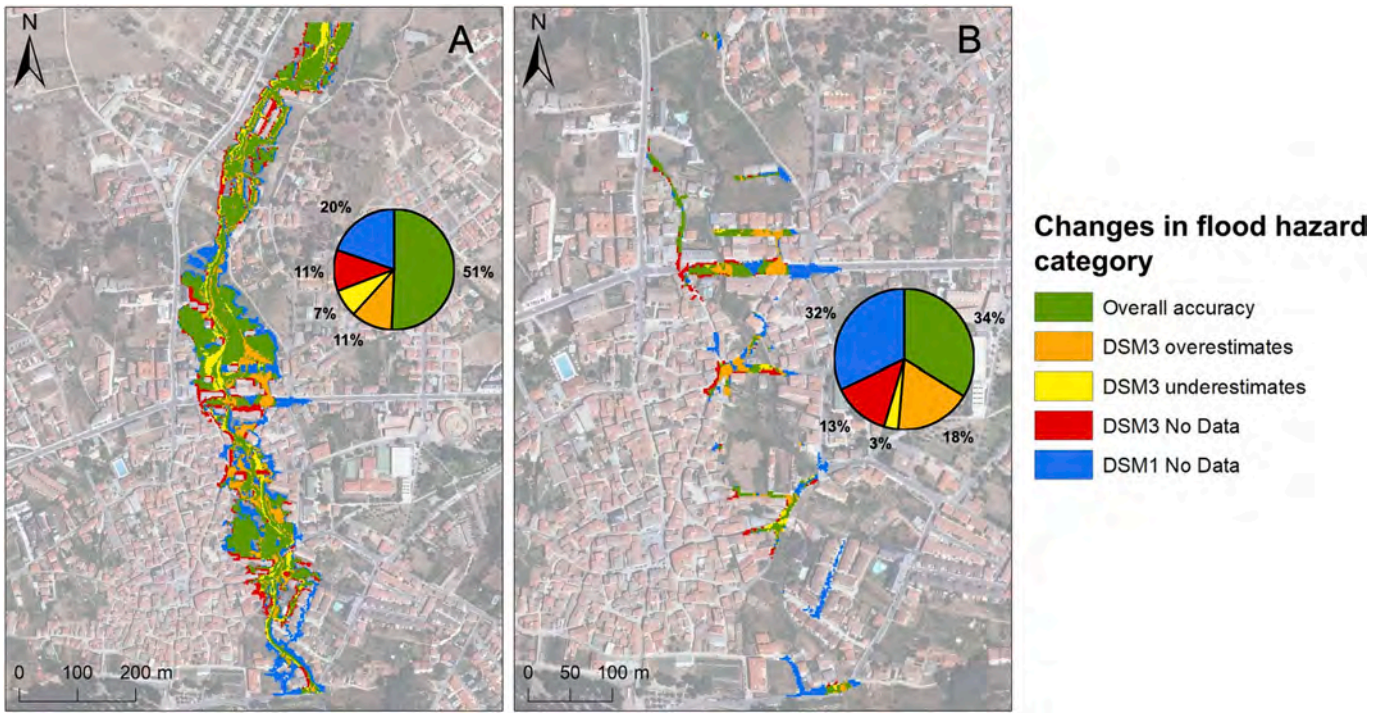

Fig. 15. Level of coincidence for scenario 4 in flood hazard categories mapped from DSM1 and DSM3. The capital letters correspond to zone 2(A) and zone 3 (B).

in flood hazard assessments. Even though LiDAR data has these flaws, it usually does not indicate in research papers what pre-processing tasks must be undertaken before incorporating it into 2D hydrodynamic models to create flood hazard maps.

According to our knowledge, the research presented here is the first to examine the detrimental impact of adopting geometrically inconsistent DSMs for urban areas on flood hazard assessment. Besides, few studies give procedures for obtaining DSMs of urban areas, and there are only a few papers in the literature that describe the added value of DSMs using LiDAR data in urban flood analyses. For instance, LiDAR data has been used to provide distributed assessments of urban water level and extent (Neal et al., 2009; Fewtrell et al., 2011). Schubert et al. (2008) used semi-automated approaches for geospatial data processing to enable mesh generation, building representation, and resistance parameter estimation. Other papers have focused on applying and testing numerical schemes with different surface friction configurations on high-resolution DSMs derived from terrestrial LiDAR (Ozdemir et al., 2013); on investigating the sensitivity of flood damage assessments concerning the quality of DSMs and building representation (Arrighi and Campo, 2019); or on the extraction of building facade openings (e.g., windows, doors) to assess flood risk (Feng et al., 2022). Fewer published papers address the issue of how to build DSMs, which allow the characterization of flow paths beneath urban objects such as elevated roads and overpasses (Noh et al., 2018), or how to construct DSMs that properly portray the surface of urban areas (Meesuk et al., 2015).

The purpose of this research was to establish the requirements for obtaining geometrically consistent DSMs of urban areas to develop trustworthy flood hazard maps. To this end, an application-oriented approach to producing DSMs has been developed (referred to as DSM1 in this study), in which the primary drainage paths of urban areas at risk of river flooding (e.g., the street-building system) were accurately geometrically depicted. DSM1 was compared using qualitative and quantitative approaches to a model that represents urban streets but not buildings (DSM2) and a third model, DSM3, created using direct interpolation of LiDAR data. This comparison aimed to characterize the geometric quality and vertical accuracy of these DSMs. Moreover, differences found in flood hazard mapping resulting from each DSM were examined.

In this study, it was assumed that the success of employing direct interpolation of LiDAR data to construct geometrically consistent DSMs is dependent on the complexity of the surface being mapped. To test this assumption, the study area was divided into four zones, with zone 1 (which was only considered for assessing the vertical accuracy of DSMs) being the least complicated from a geometric standpoint since it includes non-urban and *peri*-urban areas, and zone 4 being the most complex because it has the most buildings and streets. Zones 2 and 3 exhibited features midway between zones 1 and 4. Zones 2, 3, and 4 were utilized to test the levels of agreement between DSM2 and DSM3 with DSM1 regarding flood hazard assessment. The impact of geometric inconsistency on flood hazard mapping was also evaluated. To do this, four scenarios were explored: i) Scenario 1 depicts the 500-year flood in both the Alberche and the Chorrerón Streams; ii) Scenario 2 portrays the 25-year flood in both the Alberche River and the Chorrerón Stream; iii) Scenario 3 captures the 500-year flood in the Chorrerón Stream; and iv) Scenario 4 illustrates the 25-year flood in the Chorrerón Stream.

The findings reported here demonstrate divergences in the geometric configurations of DSM2 and DSM3 compared to DSM1, which served as the benchmark. The differences were largest in zone 4, which contains the most buildings and streets and a more complicated geometric arrangement, and lowest in zone 2, which has more undeveloped and *peri*-urban land. Comparing the performance of DSM2 and DSM3 (Figs. 5 and 6), it was found that in DSM2, cross-sections streets often exhibited a smaller area than the real one, since the DSM's construction did not account for buildings. The geometric portrayal of DSM3 is even farther from reality since a qualitative analysis of the longitudinal and transversal profiles of streets (Figs. 5 and 6) revealed the presence of artificial barriers with a high capacity to disrupt flood wave propagation. Because neither DSMs considered adding breaklines to consistently portray the Chorrerón Stream channelling, the morphological patterns in DSM2 and DSM3 were comparable, with smaller cross-sections compared to the real ones. DSM1 showed transverse and longitudinal profiles for streets and the Chorrerón Stream that are consistent with Navaluenga's geometric configuration. These findings agree with Zhou et al. (2004), who showed that LiDAR data interpolation algorithms perform poorly when there are abrupt fluctuations in elevation.

Geometric inconsistency in DSM3, and to a lesser extent in DSM2, affected flood hazard maps since 2D hydrodynamic model outputs depend on proper site geometry (Aronica et al., 2012). The level of agreement was lower as the geometric discrepancy with DSM1 rose, both at the level of the total flooded area (calculated using the F-statistic) and at the level of the predicted flood hazard category for each pixel (determined from the indices extracted from the confusion matrices). Flood hazard outputs from DSM3 were less in accord with DSM1 than DSM2. This was due to DSM3 displaying a geometric arrangement that was farther from reality than DSM2. In preferential flow pathways (streets and the urban reach of the Chorrerón Stream), DSM2 and particularly DSM3 typically generated smaller cross-sections than the true ones by creating triangulations that defined artificial barriers that hinder or restrict the passage of the flood wave. Consequently, increased water depths and more varied velocity patterns were deployed, distorting flood magnitude. Accordingly, flood hazard may define distinct flood hazard categories depending on the DSM employed, as well as streets that flood in DSM2 and DSM3 but not flooding in DSM1 and vice versa. Flood hazard outputs from DSM2 and DSM3 showed a higher agreement with DSM1 when analysing 500-year flood scenarios, while such agreements were lower when examining 25-year flood scenarios. This is because the geometric inconsistency of DSMs has less of an effect on flood hazard outputs with increasing water levels during floods.

Flood hazard mapping is a critical tool for managing flood risk as it enables potentially flood-prone areas to be identified and analysed for their risk (Mudashiru et al., 2021). Flood hazard mapping may also be used to aid in the planning and development of flood protection and evacuation procedures (Musolino et al., 2020). Another use of flood hazard mapping is that it can be employed to educate the public about flood risk and how to protect themselves and their property in the event of a flood (Henstra et al., 2019). The findings presented here reveal that the geometric consistency of the DSMs used to adjust the node height of the elements comprising the computational mesh has a large impact on the outputs of 2D hydrodynamic models and, therefore, on flood hazard assessments. If there is no geometric consistency, the flooded area and flood hazard categories indicated in the maps may vary considerably from what is expected, jeopardizing the credibility of flood risk assessments and the efficacy of flood risk mitigation measures. Consequently, the development of geometrically consistent DSMs is a critical step in developing new tools to mitigate the possible effects of floods via risk management supported by hazard and risk mapping.

6. Conclusions

This study provides an application-oriented procedure for constructing geometrically and vertically accurate urban DSMs (designated here as DSM1). Given the above, LiDAR data is especially valuable for producing consistent DSMs when supplemented with ancillary information such as break lines, which enables a precise portrayal of slope breaks common in urban settings (e.g., between streets and buildings). Through the F statistic and comprehensive and category-level indices produced from a confusion matrix, this work also looked into the effects of using DSMs that do not adequately portray urban geometry on flood hazard assessments. This research provides beyond a doubt that DSMs that are not geometrically consistent (DSM2 and DSM3 in this paper) have a detrimental impact on flood hazard estimates. Because

geometrically robust DSMs are a highly sensitive input for obtaining reliable flood hazard maps, a lack of geometric consistency in DSMs causes significant errors in the distribution of depths and velocities, as well as the spatial distribution of hazard categories and the area occupied by the flooding zone, which may lead decision-makers to take the wrong decisions during the flood risk management process. So, making sure that DSMs are geometrically consistent is important because flood hazard mapping helps protect people and economic assets from flooding while also letting communities keep control over urban growth in floodplains. As a result, spatial planners may restrict development to the best available sites, provided they have access to reliable and up-to-date data on the flood hazard in a given area.

CRediT authorship contribution statement

José M. Bodoque: Conceptualization, Methodology, Supervision, Writing – original draft, Funding acquisition. **Estefanía Aroca-Jiménez:** Methodology, Formal analysis, Data curation, Writing – review & editing. **Miguel Á. Eguibar:** Methodology, Formal analysis, Writing – review & editing. **Juan A. García:** Methodology, Formal analysis, Writing – review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All of the data from this research, including the DSMs, the hydraulic outputs, and the resulting flood hazard maps, can be freely downloaded from Zenodo (https://doi.org/10.5281/zenodo.5171778).

Acknowledgments

This research is part of the project CGL2017-83546-C3-1-R/AEI/FEDER, UE, funded by MCIN/AEI/10.13039/501100011033 and by "ERDF A way of making Europe".

Data availability statement

All of the data from this research, including the DSMs (DSM1, DSM2, and DSM3), the hydraulic outputs (i.e., water depths, flow velocities and Froude numbers) displayed by each DSM, and the resulting flood hazard maps, can be freely downloaded from Zenodo (https://doi.org/10.5281/zenodo.5171778).

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jhydrol.2022.128975.

References

Aroca-Jiménez, E., Eguibar, M.A., Bodoque, J.M., 2021. Comparison among three different Digital Surface Models and their respective hydraulic outcomes in the flood-prone urban area of Navaluenga Ávila. Zenodo, Spain.

Aronica, G.-T., Franza, F., Bates, P.-D., Neal, J.-C., 2012. Probabilistic evaluation of flood hazard in urban areas using Monte Carlo simulation. Hydrol. Process. 26 (26), 3962–3972. https://doi.org/10.1002/hyp.8370.

Arrighi, C., Campo, L., 2019. Effects of digital terrain model uncertainties on high-resolution urban flood damage assessment. J. Flood Risk Manag. 12 (S2), e12530.

Bladé, E., Cea, L., Corestein, G., Escolano, E., Puertas, J., Vázquez-Cendón, E., Dolz, J., Coll, A., 2014. Iber: herramienta de simulación numérica del flujo en ríos. Rev. Int. Métodos Numér. Cálc. Diseño Ing. 30 (1), 1–10.

Bodoque, J.M., Ballesteros-Cánovas, J.A., Stoffel, M., 2020. An application-oriented protocol for flood frequency analysis based on botanical evidence. J. Hydrol. 590, 125242. https://doi.org/10.1016/j.jhydrol.2020.125242.

Bodoque, J.M., Guardiola-Albert, C., Aroca-Jiménez, E., Eguibar, M.Á., Martínez-Chenoll, M.L., 2016. Flood damage analysis: First floor elevation uncertainty resulting from LiDAR-derived digital surface models. Remote Sens 8 (7), 604. https://doi.org/10.3390/rs8070604.

Cea, L., Bladé, E., 2015. A simple and efficient unstructured finite volume scheme for solving the shallow water equations in overland flow applications. Water Resour. Res. 51 (7), 5464–5486. https://doi.org/10.1002/2014WR016547.

Congalton, R.G., 1991. A review of assessing the accuracy of classifications of remotely sensed Data. Remote Sens. Environ. 37 (1), 35–46.

Congalton, R.G., Green, K., 1999. Assessing the accuracy of remotely sensed data: principles and practices. Lewis Publishers, Boca Raton, Florida, USA.

Congalton, R.G., Oderwald, R.G., Mead, R.A., 1983. Assessing landsat classification accuracy using discrete multivariate analysis statistical techniques. Photogramm. Eng. Remote Sens. 49 (12), 1671–1678.

Cook, A., Merwade, V., 2009. Effect of topographic data, geometric configuration and modeling approach on flood inundation mapping. J. Hydrol. 377 (1–2), 131–142. https://doi.org/10.1016/j.jhydrol.2009.08.015.

Costabile, P., Macchione, F., Natale, L., Petaccia, G., 2015. Flood mapping using LIDAR DEM. Limitations of the 1-D modeling highlighted by the 2-D approach. Nat. Hazards 77 (1), 181–204. https://doi.org/10.1007/s11069-015-1606-0.

Di Baldassarre, G., Castellarin, A., Montanari, A., Brath, A., 2009. Probability-weighted hazard maps for comparing different flood risk management strategies: A case study. Nat. Hazards 50 (3), 479–496. https://doi.org/10.1007/s11069-009-9355-6.

Díez-Herrero, A., 2001. Geomorfología e Hidrología Fluvial del Río Alberche. Modelos y S.I.G. para la Gestión de Riberas (Tesis Doctoral). Universidad Complutense de Madrid, Madrid.

A. Díez-Herrero L.L. Huerta M.L. Isidro A handbook on flood hazard mapping methodologies Publications of the Geological Survey of Spain (IGME) 2009 Madrid, Spain.

Feng, Y., Xiao, Q., Brenner, C., Peche, A., Yang, J., Feuerhake, U., Sester, M., 2022. Determination of building flood risk maps from LiDAR mobile mapping data. Comput. Environ. Urban Syst. 93, 101759 https://doi.org/10.1016/j.compenvurbsys.2022.101759.

Fewtrell, T.J., Duncan, A., Sampson, C.C., Neal, J.C., Bates, P.D., 2011. Benchmarking urban flood models of varying complexity and scale using high resolution terrestrial LiDAR data. Phys. Chem. Earth 36 (7–8), 281–291. https://doi.org/10.1016/j.pce.2010.12.011.

Henstra, D., Minano, A., Thistlethwaite, J., 2019. Communicating disaster risk? An evaluation of the availability and quality of flood maps. Nat. Hazards Earth Syst. Sci. 19 (1), 313–323. https://doi.org/10.5194/nhess-19-313-2019.

Hohenthal, J., Alho, P., Hyyppä, J., Hyyppä, H., 2011. Laser scanning applications in fluvial studies. Prog. Phys. Geogr. 35 (6), 782–809. https://doi.org/10.1177/0309133311414605.

Kundzewicz, Z.W., Kanae, S., Seneviratne, S.I., Handmer, J., Nicholls, N., Peduzzi, P., Mechler, R., Bouwer, L.M., Arnell, N., Mach, K., Muir-Wood, R., Brakenridge, G.R., Kron, W., Benito, G., Honda, Y., Takahashi, K., Sherstyukov, B., 2014. Flood risk and climate change: Global and regional perspectives. Hydrol. Sci. J. 59 (1), 1–28. https://doi.org/10.1080/02626667.2013.857411.

Landis, J.R., Koch, G.G., 1977. The measurement of observer agreement for categorical data. Biometrics 33, 159–174.

Liu, X., 2008. Airborne LiDAR for DEM generation: Some critical issues. Prog. Phys. Geogr. 32 (1), 31–49. https://doi.org/10.1177/0309133308089496.

Masood, M., Takeuchi, K., 2012. Assessment of flood hazard, vulnerability and risk of mid-eastern Dhaka using DEM and 1D hydrodynamic model. Nat. Hazards 61 (2), 757–770. https://doi.org/10.1007/s11069-011-0060-x.

Meesuk, V., Vojinovic, Z., Mynett, A.E., Abdullah, A.F., 2015. Urban flood modelling combining top-view LiDAR data with ground-view SfM observations. Adv. Water Resour. 75, 105–117. https://doi.org/10.1016/j.advwatres.2014.11.008.

Mudashiru, R.B., Sabtu, N., Abustan, I., Balogun, W., 2021. Flood hazard mapping methods: A review. J. Hydrol. 603, 126846 https://doi.org/10.1016/j.jhydrol.2021.126846.

Musolino, G., Ahmadian, R., Xia, J., Falconer, R.A., 2020. Mapping the danger to life in flash flood events adopting a mechanics based methodology and planning evacuation routes. J. Flood Risk Manag. 13 (4), e12627.

Neal, J.C., Bates, P.D., Fewtrell, T.J., Hunter, N.M., Wilson, M.D., Horritt, M.S., 2009. Distributed whole city water level measurements from the Carlisle 2005 urban flood event and comparison with hydraulic model simulations. J. Hydrol. 368 (1–4), 42–55. https://doi.org/10.1016/j.jhydrol.2009.01.026.

Noh, S.J., Lee, J.H., Lee, S., Kawaike, K., Seo, D.J., 2018. Hyper-resolution 1D–2D urban flood modelling using LiDAR data and hybrid parallelization. Environ. Modell. Softw. 103, 131–145. https://doi.org/10.1016/j.envsoft.2018.02.008.

O'Donnell, E.C., Thorne, C.R., 2020. Drivers of future urban flood risk. Philos. Trans. R. Soc. A 378 (2168), 20190216. https://doi.org/10.1098/rsta.2019.0216.

Ozdemir, H., Sampson, C.C., de Almeida, G.A.M., Bates, P.D., 2013. Evaluating scale and roughness effects in urban flood modelling using terrestrial LIDAR data. Hydrol. Earth Syst. Sci. 17 (10), 4015–4030. https://doi.org/10.5194/hess-17-4015-2013.

Pasquier, U., He, Y., Hooton, S., Goulden, M., Hiscock, K.M., 2019. An integrated 1D–2D hydraulic modelling approach to assess the sensitivity of a coastal region to compound flooding hazard under climate change. Nat. Hazards 98 (3), 915–937. https://doi.org/10.1007/s11069-018-3462-1.

Priestnall, G., Jaafar, J., Duncan, A., 2000. Extracting urban features from LiDAR digital surface models. Comput. Environ. Urban Syst. 24 (2), 65–78.

Prinos, P., Kortenhaus, A., Swerpel, B., Jiménez, J., 2008. Review of Flood Hazard Mapping. FLOODsite report T03-07-01.

Scawthorn, C., Asce, F., Blais, N., Seligson, H., Tate, E., Mifflin, E., Thomas, W., Murphy, J., Jones, C., 2006. HAZUS-MH Flood Loss Estimation Methodology. I: Overview and

Flood Hazard Characterization. Nat. Hazards Rev. 7 (2), 60–71. https://doi.org/10.1061/ASCE1527-698820067:260.
Schubert, J.E., Sanders, B.F., Smith, M.J., Wright, N.G., 2008. Unstructured mesh generation and landcover-based resistance for hydrodynamic modeling of urban flooding. Adv. Water Resour. 31 (12), 1603–1621. https://doi.org/10.1016/j.advwatres.2008.07.012.
Shan, J., Sampath, A., 2005. Urban DEM generation from raw lidar data: A labeling algorithm and its performance. Photogramm. Eng. Remote Sens. 71 (2), 217–226.
Story, M., Congalton, R.G., 1986. Accuracy Assessment: A User's Perspective. Photogramm. Eng. Remote Sens. 52 (3), 397–399.
Surminski, S., Thieken, A.H., 2017. Promoting flood risk reduction: The role of insurance in Germany and England. Earth Future 5 (10), 979–1001. https://doi.org/10.1002/2017EF000587.
Tsakiris, G., 2014. Flood risk assessment: Concepts, modelling, applications. Nat Hazards Earth Syst. Sci. 14 (5), 1361–1369. https://doi.org/10.5194/nhess-14-1361-2014.
Turner, A.B., Colby, J.D., Csontos, R.M., Batten, M., 2013. Flood modeling using a synthesis of multi-platform LiDAR data. Water 5 (4), 1533–1560. https://doi.org/10.3390/w5041533.
Undrr, 2020. The human cost of disasters: an overview of the last 20 years (2000–2019). UN Office for Disaster Risk Reduction, Geneva, Switzerland.
van Alphen, J., Martini, F., Loat, R., Slomp, R., Passchier, R., 2009. Flood risk mapping in Europe, experiences and best practices. J. Flood Risk Manag. 2 (4), 285–292. https://doi.org/10.1111/j.1753-318X.2009.01045.x.
Vojinovic, Z., Tutulic, D., 2009. On the use of 1D and coupled 1D–2D modelling approaches for assessment of flood damage in urban areas. Urban Water J. 6 (3), 183–199. https://doi.org/10.1080/15730620802566877.
Wmo, 2021. Atlas of Mortality and Economic Losses from Weather, Climate and Water Extremes (1970–2019). Technical Report, World Meteorological Organization, Geneva, Switzerland, WMO-No, p. 1267.
Yalcin, E., 2020. Assessing the impact of topography and land cover data resolutions on two-dimensional HEC-RAS hydrodynamic model simulations for urban flood hazard analysis. Nat. Hazards 101 (3), 995–1017. https://doi.org/10.1007/s11069-020-03906-z.
Zhou, G., Song, C., Simmers, J., Cheng, P., 2004. Urban 3D GIS From LiDAR and digital aerial images. Comput. Geosci. 30 (4), 345–353. https://doi.org/10.1016/j.cageo.2003.08.012.