Hydrological Sciences Journal

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/thsj20

Geomorphic flood hazard mapping: from floodplain delineation to flood hazard characterization

Andrea Magnini, Michele Lombardi, Armir Bujari, Pietro Mattivi, Iuliia Shustikova, Simone Persiano, Marco Patella, Gabriele Bitelli, Paolo Bellavista, Francesco Lo Conti, Antonio Tirri, Stefano Bagli, Paolo Mazzoli & Attilio Castellarin

To cite this article: Andrea Magnini, Michele Lombardi, Armir Bujari, Pietro Mattivi, Iuliia Shustikova, Simone Persiano, Marco Patella, Gabriele Bitelli, Paolo Bellavista, Francesco Lo Conti, Antonio Tirri, Stefano Bagli, Paolo Mazzoli & Attilio Castellarin (2023) Geomorphic flood hazard mapping: from floodplain delineation to flood hazard characterization, Hydrological Sciences Journal, 68:16, 2388-2403, DOI: 10.1080/02626667.2023.2269909

To link to this article: https://doi.org/10.1080/02626667.2023.2269909

Published online: 14 Nov 2023.

HYDROLOGICAL SCIENCES JOURNAL
2023, VOL. 68, NO. 16, 2388–2403
https://doi.org/10.1080/02626667.2023.2269909

Geomorphic flood hazard mapping: from floodplain delineation to flood hazard characterization

Andrea Magnini[a], Michele Lombardi[b], Armir Bujari[b], Pietro Mattivi[a], Iuliia Shustikova[a,c], Simone Persiano[a,d], Marco Patella[b], Gabriele Bitelli[a], Paolo Bellavista[b], Francesco Lo Conti[e], Antonio Tirri[e], Stefano Bagli[f], Paolo Mazzoli[f] and Attilio Castellarin[a]

[a]Department of Civil, Chemical, Environmental and Materials Engineering (DICAM), University of Bologna, Bologna, Italy; [b]Department of Computer Science and Engineering (DISI), University of Bologna, Bologna, Italy; [c]Strategic Risk Consulting, WTW, London, UK; [d]Catastrophe Risk Modeling & Mitigation, UnipolSai Assicurazioni S.p.A, Bologna, Italy; [e]Leithà, Unipol Group, Milan and Bologna, Italy; [f]GECOsistema Srl – Geographic Environmental COnsulting, Cesena, Italy

ABSTRACT

Recent studies show how geomorphic descriptors, retrieved from digital elevation models (DEMs), can be used for flood hazard mapping. As they strictly depend on the accuracy of the input DEMs and reference flood hazard maps used for training, DEM-based flood hazard models may display severe inconsistencies. Our study shows the application of two advanced DEM-based models to a large study area, and presents two main innovative points. First, the delicate tasks of appropriately selecting the input DEM and flood hazard map are specifically addressed with newly defined methods. Second, the ability of DEM-based models to exploit their natural features to enhance flood hazard mapping over the study region is investigated. Our results show (a) the benefits of considering multiple geomorphic descriptors, (b) the potential of DEM-based models for completing the information of imperfect reference flood hazard maps, and (c) the advantages of continuous representation of hazard over binary flood maps.

ARTICLE HISTORY
Received 16 May 2023
Accepted 19 September 2023

EDITOR
K. Soulis

ASSOCIATE EDITOR
(not assigned)

KEYWORDS
geomorphic index; DEM; remote sensing; inundation scenario; machine learning; Italy

1 Introduction

Several studies show that damages caused by inundation events are steadily increasing worldwide (e.g. Jongman *et al.* 2014). This is due to multiple reasons, as population growth and the urbanization phenomenon are strictly linked to the exposure and vulnerability components of flood risk (Domeneghetti *et al.* 2015, Requena *et al.* 2017), while climate change can locally increase flood occurrence (see e.g. Sharma *et al.* 2018, Koutsoyiannis 2020).

In recent years, research on flood hazard and risk modelling has proposed data-driven, simplified methods, that aim to avoid using computationally expensive hydrodynamic models that require large amounts of data (Sampson *et al.* 2015). Data-driven methods make use of geomorphic descriptors that are easily retrievable from digital elevation models (DEMs), and some available flood-hazard reference information (i.e. flood hazard maps from hydraulic models or records of historical events). After a training phase (i.e. model calibration) based on the reference maps, the model links the geomorphic descriptors to the flood hazard. Thus, the model can be efficiently applied to any area where a DEM is available and the relationship between flood hazard and descriptors is assumed to be the same. Nowadays, several DEM-based approaches are reported in the literature, using geomorphic descriptors in univariate (Noman *et al.* 2001, Manfreda *et al.* 2014, 2015, Samela *et al.* 2017, Manfreda and Samela 2019) and multivariate models, either alone (Gnecco *et al.* 2017, Marchesini *et al.* 2021, NOAA Office of Water Prediction 2023) or combined with non-geomorphic information (Wang *et al.* 2015, Arabameri *et al.* 2019, Costache *et al.* 2020, Faridani *et al.* 2020).

Indeed, the availability of input DEMs and reference flood hazard maps for the application of DEM-based methods is rapidly increasing. The steady evolution of remote sensing techniques allows the production of a variety of DEMs with different data acquisition methods, resolution and vertical accuracy. Meanwhile, there is a large array of flood hazard maps that are based on different modelling techniques, that have varying resolution and accuracy. Nevertheless, the reliability of the flood-hazard maps and the accuracy of the DEMs is variable and geographically heterogeneous, and thus, effectively selecting the most appropriate datasets is not straightforward (Tavares da Costa *et al.* 2019, Lindersson *et al.* 2021).

The present research shows the application, comparison and discussion of two DEM-based models for flood-hazard mapping, one univariate (i.e. a single DEM-based geomorphic index is used) and one multivariate (a variety of indices is used; see also Magnini *et al.* 2022). We select Italy as the study area, given its remarkable variability of hydrological, climatic, and geomorphological characteristics.

CONTACT Andrea Magnini andrea.magnini@unibo.it Department of Civil, Chemical, Environmental and Materials Engineering (DICAM), University of Bologna, Via Aurelio Saffi, Bologna 40131, Italy

The initial phase of our study consists of the analysis of the available datasets for selecting the most appropriate input DEM and reference flood-hazard map. For both, an effective framework for the selection is proposed, exploiting EU-Hydro (Gallaun *et al.* 2019) as the reference river network. Second, as a baseline model we develop a univariate DEM-based flood hazard model using the geomorphic flood index (GFI; see Manfreda *et al.* 2014). For the multivariate model, we consider six geomorphic descriptors, which we identified based on a literature review (e.g. Manfreda *et al.* 2015), and we combine them to model flood hazard through a decision tree classifier (see e.g. Magnini *et al.* 2022). After the calibration over specific fractions of the study area, the two models are applied to the whole region. Finally, we validate both models by referring to independent information (i.e. datasets that were not used for training or calibrating the DEM-based models). Specifically, this consists of three inundation extents produced by the same number of recent flood events and delineated on the basis of remote sensing data (local validation), as well as a synthetic inundation scenario generated as the envelope of several levee-breaching simulations along a 300 km branch of the major Italian river by means of a two-dimensional (2D) hydrodynamic numerical model.

The methods adopted for setting up the models are inspired by previous literature (see e.g. Tavares da Costa *et al.* 2019, Magnini *et al.* 2022), but the nature and goal of our application and discussion are new. Differently from what has been done in relevant previous studies (see e.g. Tavares da Costa *et al.* 2020, Magnini *et al.* 2022), we do not limit our assessment of models' performance to the similarity of their output to the target flood map in test areas that differ from training ones. Instead, we perform a detailed assessment of flood DEM-based susceptibility maps in areas where the calibration flood map may be inhomogeneous or inaccurate – which is a common situation in practice. This is done by referring to validation sources (e.g. detailed output of hydrodynamic model runs) that are an alternative to the available target flood maps. The main research question here is whether a trained DEM-based flood hazard map may resolve heterogeneities and inconsistencies that are present in the target flood map used for training.

Usually, DEM-based methods are used at large scales to obtain binary maps, which delineate the maximum flood extent associated with a given return period, and they are considered auxiliary tools of the more accurate hydraulic models. However, by nature DEM-based flood hazard models can efficiently handle secondary river networks (see e.g. Nardi *et al.* 2019) and can produce a spatially continuous and highly homogeneous characterization of flood hazard. Some authors (e.g. Costache *et al.* 2020, Avand *et al.* 2022, Deroliya *et al.* 2022) exploited multivariate DEM-based approaches for a spatially continuous estimation of flood susceptibility. Nevertheless, their models were calibrated on a number of independent inundation events, instead of a coherent flood hazard map with a given return period. In the present study, we investigate the unexplored potential of these features of DEM-based approaches for enhancing the flood hazard information with respect to pre-existing calibration maps covering the whole study area. The univariate and multivariate models are compared and discussed in great detail: we consider not only the classical binary outputs (i.e. maps distinguishing flood-prone and non-flood-prone areas) but also the spatially continuous characterization of flood hazard naturally offered by geomorphic methods (e.g. the values of GFI over the entire study area). This enables us to further explore the reasons for the models' performance and discuss a better way to exploit their potential. Moreover, the presentation of a baseline reference framework to address the selection of the most appropriate DEM and reference hazard map is an innovative and useful element for future studies on the topic.

The paper is organized as follows: Section 2 describes the set-up of the DEM-based models. Section 3 presents the study area, evaluation metrics and validation datasets. The methods for the selection of the input DEM and target flood hazard map are described in Sections 4 and 5, respectively. Section 6 shows the results for the selection of DEM and reference hazard map, the application of the models to the study area and their validation. In Section 7, the methods and results are discussed.

2 Methods for DEM-based modelling

The univariate and multivariate DEM-based models are set up with open-source software (Van Rossum *et al.* 1995, Tarboton 2003, Pedregosa *et al.* 2011, GRASS Development Team, 2019) using the same input DEM and target flood hazard map, but different calibration (or training) strategies. These methods are described below.

2.1 Univariate DEM-based flood hazard modelling

The univariate DEM-based model is set up according to the methods described in the literature (e.g. Tavares da Costa *et al.* 2019), which consist of three consecutive phases: (1) computation of the selected geomorphic index (the GFI in this case); (2) definition of a calibration area that includes only the pixels close to the river network; and (3) calibration of the model – that is, finding the best threshold value for fitting the target flood hazard map.

We select the GFI [-], as according to reliable scientific publications (e.g. Manfreda *et al.* 2015, Samela *et al.* 2017) it is one of the most efficient DEM-based flood hazard modelling methods. It is defined as follows:

$$GFI = ln(h_r/HAND) \quad (1)$$

where $h_r \cong bA_r^n$ (in metres); A_r is the contributing area in the hydrologically nearest stream section, and coefficient b and exponent n can be appropriately estimated via calibration or taken from the literature (Nardi *et al.* 2006). The height above the nearest drainage (HAND) is defined as the vertical difference (in metres) between a given cell and the closest cell belonging to the river network (Rennó *et al.* 2008). The latter is the first cell of the river network encountered along the flow direction.

The calibration area is defined with a fixed 200 m buffer zone around the flood-prone areas of the target hazard map. This method allows an easy and efficient selection of the pixels

of the domain that are close to the main river network and that are appropriate to successfully train the model (Gnecco *et al.* 2017, Tavares da Costa *et al.* 2019, Marchesini *et al.* 2021, Magnini *et al.* 2022). The 200 m buffer has been adopted based on some preliminary sensitivity analysis.

The calibration of the model consists of two phases: first, the geographical domain is divided into hydro-climatic districts; second, for each in turn, the best threshold for the GFI is sought by means of optimization of the true skill statistic (TSS; see Youden 1950) objective function (see Equation 2).

$$\mathrm{TSS} = \frac{\mathrm{TP}}{\mathrm{TP}+\mathrm{FN}} + \frac{\mathrm{TN}}{\mathrm{TN}+\mathrm{FP}} - 1 \tag{2}$$

where TP, TN, FP, FN represent the number of true positive, true negative, false positive and false negative cases, respectively. Pixels whose GFI value is higher than the threshold are identified as susceptible to floods.

2.2 Multivariate DEM-based flood hazard modelling

In our study, multivariate DEM-based flood hazard modelling strictly relies on the methods adopted by Magnini *et al.* (2022). First, geomorphic descriptors of the study area are retrieved from a selected DEM. Second, a calibration area is defined with a 2 km buffer radius (based on Magnini *et al.* 2022) around the target flood-prone areas (see Section 2.1). Third, a decision tree is trained within the calibration area and, finally, it is applied to the whole domain (i.e. even areas outside of the 2 km buffer) to produce a final flood susceptibility map. Since the nature of the multivariate model is very different from the univariate one (see Section 2.1), it should not be surprising that the optimal buffer distance is different in the two cases.

We consider six of the seven geomorphic descriptors adopted by Magnini *et al.* (2022). The descriptor that has been discarded (i.e. the modified topographic index, see Manfreda *et al.* 2008) is the one with the lowest influence in the models of Magnini *et al.* (2022), while the others are very common in the literature (e.g. Manfreda *et al.* 2015, Samela *et al.* 2017, Khosravi *et al.* 2018) and provide a rather complete description of the topography and hydrology of the study area:

(1) Elevation of the ground surface [m];
(2) Local slope (tan(β)) [-], estimated for each cell as the maximum slope among the eight possible flow directions and computed as the ratio between the vertical and the horizontal differences;
(3) Horizontal distance from the nearest stream (D), defined as the length [m] of the path that hydrologically connects each cell to the nearest cell of the river network;
(4) The abovementioned HAND [m] (see Section 2.1);
(5) The abovementioned GFI [-] (see Section 2.1);
(6) An alternative version of the GFI, hereinafter referred to as the local geomorphic flood index (LGFI) [-], defined as in Equation (3).

$$LGFI = ln(h_l/HAND) \tag{3}$$

where the water depth h_l is computed as for the h_r [m] for the GFI, but with reference to the contributing area of the considered pixel instead of the one of the nearest river section.

Decision trees are supervised machine learning techniques very commonly used for modelling hydraulic processes (Mosavi *et al.* 2018). They consist of hierarchical structures in which the input dataset is successively split into sub-groups according to some fixed attributes. Each of the final sub-groups is associated with a certain probability (or p value) of assuming a specific value of the output variable; if a threshold is assigned to the p value, a classification is obtained (for more details, see Hastie *et al.* 2009). In this study, the threshold is 0.5, as the default value. The versatility, simplicity and interpretability of decision trees make them very useful tools to solve different types of problems.

3 Application

The present section describes the study area, the metrics adopted to evaluate the results, and the validation datasets.

3.1 The study area: Italy

The selected study area consists of the whole Italian peninsula and islands, that amounts to approximately 301 × 10^3 km^2 (see Fig. 1). The overall extent and the large variety of geographical and climatic conditions make Italy an interesting and complex study area for large-scale flood-hazard modelling. In fact, the length of the peninsula and mostly mountainous hinterland make the climate highly diverse, ranging from humid sub-tropical to humid continental and oceanic (see e.g. climate classification in Cui *et al.* 2021). The Alps extend from the northwest to the northeast, and are the highest mountain range (i.e. with 12 peaks higher than 3500 m a.s.l., and the highest peak at 4809 m a.s.l.). The Apennines stretch from north to south along almost the whole peninsula (i.e. about 1300 km) and have lower peaks compared to the Alps (i.e. the highest peak is 2900 m a.s.l.). The largest plain, located in the north, is the floodplain of the Po River, which is the main river in Italy (~650 km in length, with a discharge of ~1500 m^3/s).

Italy is an interesting case study for flood modelling, as a large portion of its territory is subject to floods: 5.4% has a high probability hazard, which corresponds to 2.4 million people exposed; while 14% has a low probability hazard, corresponding to 12.2 million people (Trigila *et al.* 2021).

3.2 Performance metrics for binary tests

We use several performance metrics that vary between 0 and 1, where 1 is the optimal value: TSS (see Equation 2), accuracy (ACC), precision (or predicted positive value, PPV), recall (or true positive ratio, TPR), and the harmonic mean of precision and recall (F1). All of them depend on TP, TN, FP and FN (already defined in Section 2.1):

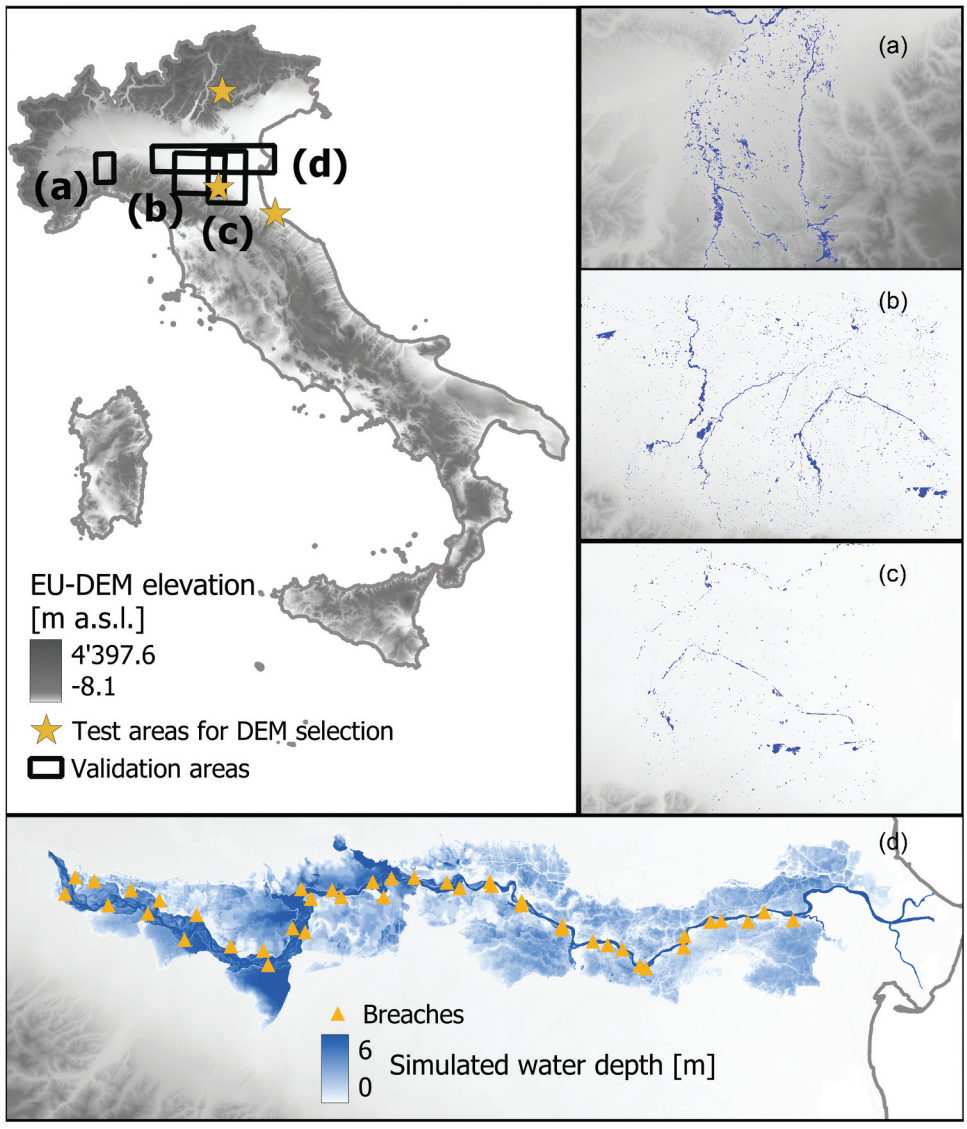

Figure 1. EU-DEM (i.e. European DEM, Garcia 2015) in Italy (top left panel); validation datasets: inundation maps associated with events (a) AL21/10/19, (b) BO20/11/19, and (c) BO21/11/19; catastrophic inundation scenario along the middle lower portion of the Po River in terms of maximum simulated water depths (d).

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

$$PPV = \frac{TP}{TP + FP} \quad (5)$$

$$TPR = \frac{TP}{TP + FN} \quad (6)$$

$$F1 = 2 \cdot \frac{PPV \cdot TPR}{PPV + TPR} \quad (7)$$

3.3 Validation datasets

After being trained (i.e. calibrated) in the calibration areas (i.e. 200 m and 2 km buffer areas, as described in Section 2), the DEM-based models are applied to the whole study region (i.e. even outside the buffer areas). The validation is performed by comparing the model outputs with new information, consisting of two types of datasets: (1) three inundation maps delineated from satellite data, and (2) one envelope flood hazard map obtained from the merger of several 2D hydrodynamic simulations. These all are associated with a return period that is approximately the same as, or lower than the reference flood hazard map. Based on the return period and the location, which corresponds to one of the most flood-susceptible areas in Italy, the four datasets can be adopted for an effective validation.

Concerning the first validation dataset, two recent flood events are selected: the inundation event that occurred between 19 and 24 October 2019 in Alessandria Province (AL, Piedmont region), and the one that occurred between 15 and 19 November 2019 in Bologna province (BO, Emilia-Romagna region). These events are described by three Sentinel-1 synthetic aperture radar (SAR) images; the corresponding inundation extents are delineated, within the present study, through an expressly developed change detection method, partially derived from Canty (2019), and

successfully validated with ground truth data. Finally, three inundation maps, named AL 21/10/19, BO 20/11/19 and BO 21/11/19, are obtained and used for the validation of the target and modelled flood maps (see panels (a–c) of Fig. 1).

The second validation dataset consists of a single flood hazard map obtained from a series of simulations run with the 2D hydrodynamic model LISFLOOD-FP (Bates and De Roo 2000, Neal *et al.* 2012, Shustikova *et al.* 2020). The flood hazard map with 50 m horizontal resolution was produced by merging 42 model simulations into one map (i.e. each simulation is associated with a return period higher than 200 years but lower than 500 years; see e.g. Domeneghetti *et al.* 2015). Each simulation represents hypothetical levee breaches on both banks of the river (21 on the right bank and 21 on the left bank; see panel (d) of Fig. 1).

4 Analysis of available DEMs

Nowadays, DEMs can be developed from a variety of surveying techniques, each characterized by specific advantages and disadvantages. Above all, the onset of modern satellite remote sensing techniques allowed the creation of freely distributed DEMs with global or semi-global coverage. However, the accuracy of the DEMs is often affected by inevitable errors associated with the techniques and algorithms used for creating the DEMs, and the characteristics of the terrain, such as morphology and land cover (Mukherjee *et al.* 2013, Thomas *et al.* 2014). Producers of global and national DEMs carry out quality assessments on their products, yet usually, only the global root mean square error (RMSE; see Equation 9) is provided as a measure of accuracy, which gives no information about the accuracy over specific areas of interest or geomorphological contexts. Thus, choosing the right terrain model for a study can be difficult without performing a specific data quality analysis (Thomas *et al.* 2014, Patel *et al.* 2016, Florinsky *et al.* 2018, Tavares da Costa *et al.* 2019).

For this reason, seven DEMs with a spatial resolution finer than 100 m, obtained with various techniques and covering the whole of Italy, are here considered and assessed. They are: SRTMGL1 (i.e. Shuttle Radar Topography Mission Global 30m, see Rodríguez *et al.* 2006, Farr *et al.* 2007), ASTER GDEM (i.e. Advanced Spaceborne Thermal Emission and Reflection Radiometer Global Digital Elevation Model, see (Tachikawa *et al.* 2011, Abrams 2016, Gesch *et al.* 2016), ALOS (i.e. Advanced Land Observing Satellite) AW3D30 (i.e. ALOS World 3D 30m, see Tadono *et al.* 2016, Takaku and Tadono 2017), TINITALY (i.e. Triangular Irregular Network Italy, see Favalli 2004, Tarquini *et al.* 2012), EU-DEM (i.e. European DEM, Bashfield and Keim 2011, Tøttrup 2014, Garcia 2015), HydroSHEDS DEM (Lehner and Grill 2013) and MERIT (i.e. Multi-Error-Removed Improved-Terrain) DEM (Yamazaki *et al.* 2017). The spatial resolution of these DEMs is: 1 arc second (≈ 30 m) for STRMGL1, ASTER GDEM and ALOS AW3D30; 10 m for TINITALY; 25 m for EU-DEM and 3 arc second (≈ 90 m) for HydroSHEDS DEM and MERIT DEM.

The seven DEMs are tested over three areas of interest (see top left of Fig. 1), each with different morphological and land-cover characteristics. These are, from north to south, Valsugana (a valley in a predominantly mountainous region in northern Italy), the territory of Bologna Municipality and surroundings (an area characterized by flatland, hills and urban zones), and the territory of Rimini and surroundings (a coastal urban area on the Adriatic Sea). In these areas, the national DEMs are compared with the high-resolution DEMs obtained by the airborne LiDAR surveys performed for the Special Remote Sensing Plan (see Costabile 2010), carried out by the Italian Ministry of the Environment and for Protection of the Land and Sea, which we assume here as ground truth. These reference DEMs have slightly different characteristics for each area of interest, with the resolution varying in a 1–2 m range, and vertical accuracy being between 15 and 30 cm. A series of tests are carried out to evaluate the vertical accuracy of the DEMs in respect to the reference LiDAR for the three areas, considering different terrain slopes (i.e. <5°, 5–10°, 10–30° and >30°), land cover types (i.e. urban, forest, low vegetation and crop fields, bare land) and HAND values (i.e. HAND < 3 m and HAND < 5 m). When performing these tests, the LiDAR datasets are resampled to the resolution of the tested DEM with a bilinear method. It consists of computing each pixel of the output file as the weighted average of the four closest pixels of the input (Netravali and Hasskell 1995).

In addition to the vertical accuracy, hydraulic consistency is also evaluated by comparing the topographic wetness index (or TWI, see Beven and Kirkby 1979) with the EU-Hydro photo-interpreted river network dataset (Gallaun *et al.* 2019), made available by the Copernicus Programme from the European Union (https://www.copernicus.eu/en). This choice is derived from preliminary analyses, which pointed out the strong agreement between the EU-Hydro dataset and the actual Italian stream network. The TWI (see Equation 8) is directly computed from the DEMs via the specific catchment area (SCA, defined as the contributing area per unit width of contour, in metres), which represents the tendency of a pixel to receive water, and the slope φ, which represents the tendency to drain. Thus, it is strongly related to the water flow direction (Mattivi *et al.* 2019) and can be used as a qualitative evaluation of hydraulic consistency.

$$\mathrm{TWI} = \ln\left(\frac{\mathrm{SCA}}{\tan(\varphi)}\right) \tag{8}$$

The vertical accuracy of the models is obtained by computing the residuals, $e(x,y)$, between the two DEMs (i.e. $e(x,y) = f'(x,y) - f(x,y)$, where $f'(x,y)$ is the surface of the DEM under analysis and $f(x,y)$ is the surface of the reference DEM). We refer to widely used performance metrics to quantify the accuracy of each DEM; in particular, we consider the linear error with 90% confidence (LE90), and the RMSE (Equation 9).

$$RMSE = \frac{\sqrt{\sum_{i=1}^{N} e(x,y)^2}}{N} \tag{9}$$

where N is the sample size – in this case, the number of pixels of the reference geographical area.

5 Analysis of available reference flood-hazard maps

In recent years, the large demand for reliable flood hazard maps of different scale (e.g. national, continental, global) has accelerated the development of various flood modelling methods and

frameworks (e.g. Yamazaki *et al.* 2011, Pappenberger *et al.* 2012, Winsemius *et al.* 2013, Sampson *et al.* 2015, Dottori *et al.* 2016). As a result, plenty of flood hazard maps are available, but the comparison of different large-scale models is not straightforward (Ward *et al.* 2015, Trigg *et al.* 2016, Lindersson *et al.* 2021), making the selection of a target map critical.

Here, we consider the two hazard maps used by Magnini *et al.* (2022) and we compare them in Italy based on their capillarity in representing flood hazard along the national river network. Indeed, having target flood hazard information for a greater number of minor streams, which have lower accumulation area, can lead to more effective training of DEM-based models (e.g. see Magnini *et al.* 2022). The first map is the European-scale map made available by the European Joint Research Centre (JRC; see Alfieri *et al.* 2014), with a 500-year return period and 100 m resolution, that was developed as a component of the Copernicus European Flood Awareness System (EFAS, www.efas.eu). This map results from a cascading model simulation approach, which is composed of a distributed hydrological model for derivation of peak flows and flood hydrographs with selected return period, and 2D hydraulic simulations for a large number of river sections with source area larger than 500 km^2.

The second map is provided by the Italian Institute for Environmental Protection and Research (ISPRA; see ISPRA 2018) to fulfil the EU Floods Directive of the European Commission (2007/60/EC). It refers to a return period of about 500 years and is the merger of different hazard maps produced by local authorities. Resulting from local maps obtained with different methodologies, the ISPRA flood hazard map has notable heterogeneity: detailed flood hazard mapping characterizes some regions (e.g. see the northwestern region of Fig. 2), while mapping is sparser in others (e.g. see the northeastern portion of the study area in Fig. 2).

The evaluation of these two maps is carried out by considering the overlap with the EU-Hydro river network at a national scale, that is selected as reference (as for the DEM selection phase). Two steps are needed: first, the EU-Hydro shapefile is converted to a raster file with the same resolution and dimension of the considered hazard map. Second, we compute the ratio between the number of flood-prone pixels falling on the river network and the total number of river-network pixels. This is done separately for the different Strahler orders, and is used as a measure of capillarity and completeness of the target datasets for training DEM-based flood hazard models.

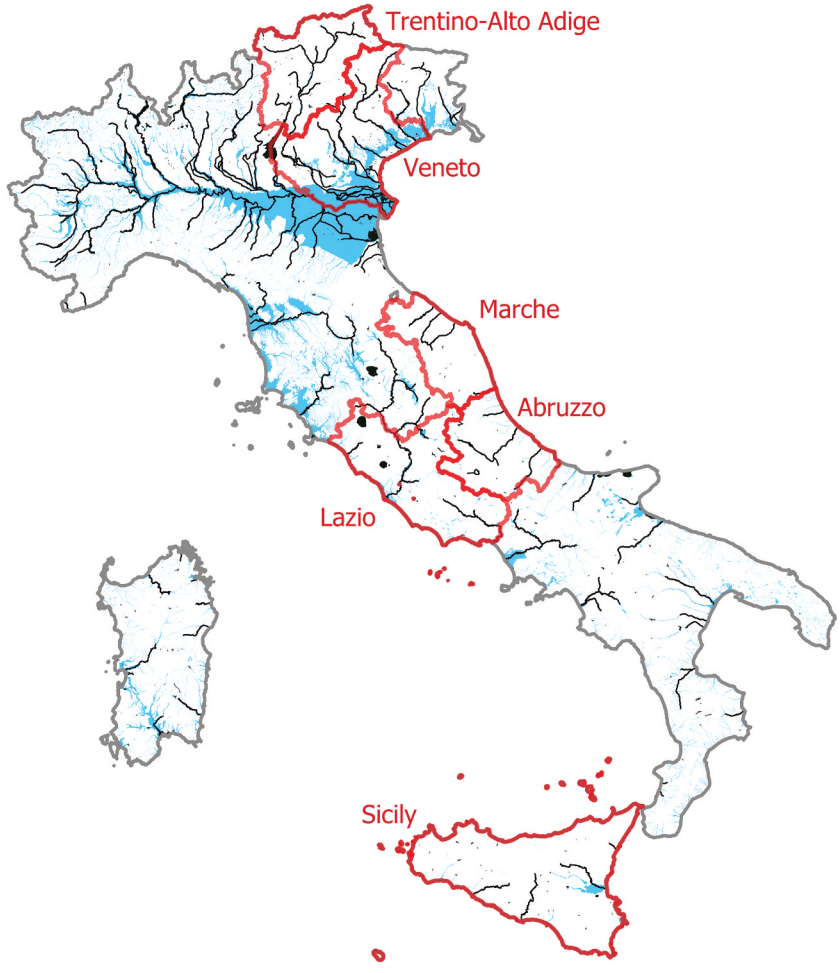

Figure 2. ISPRA (i.e. Istituto Superiore per la Protezione e Ricerca Ambientale) flood hazard map with 500-year return period (shaded light blue); in black: lakes and major rivers (Strahler order ≥ 5), from the EU-Hydro dataset (© European Union, Copernicus Land Monitoring Service 2021, European Environment Agency (EEA)). The six Italian administrative regions mentioned in Section 7.1 are outlined in red.

6 Results and validation

6.1 Results for DEM selection

Metrics for the evaluation of the vertical accuracy of the considered DEMs (i.e. RMSE and LE90; see Section 4) in the test areas (i.e. Valsugana, Bologna and Rimini, from north to south in Fig. 1) are computed with reference to the LiDAR measurements (see Section 4) and reported in Table 1. The three DEMs with the lowest errors are TINITALY, SRTM and MERIT for Valsugana; SRTM, HydroSHEDS, and MERIT for Bologna; and EU-DEM, HydroSHEDS and MERIT for Rimini. It can be noted also that the vertical accuracy for EU-DEM and SRTM is very similar in Bologna. Metrics computed for the different terrain slopes, land cover types and HAND values are not reported for the sake of brevity, as they confirm what is seen in Table 1.

The analysis of TWI clearly shows that the hydraulic consistency of TINITALY is very low, while the other DEMs have similar performance. Thus, TWI is useful to exclude inappropriate DEMs, but not for numerically ranking the best ones. For the sake of brevity, a detailed example of only the comparison between TINITALY and EU-DEM within the Bologna test area is shown (Fig. 3).

Given the combination of good performance in flat areas (i.e. Bologna and Rimini), the good hydraulic consistency, and the good resolution, the EU-DEM is selected as the most appropriate DEM for the present study. This decision is also supported by the nature of the EU-DEM, which was produced by hydraulic conditioning on the EU-Hydro river network (Bashfield and Keim 2011), the same dataset largely used for analysing the results of the present study.

6.2 Results for selection of reference flood hazard map

Quantification of the agreement between the considered flood hazard maps (i.e. the ISPRA and JRC maps) and the EU-Hydro river network over the study area is reported in Table 2. It is evident that flood-prone areas comprise a much higher portion of the river network in the ISPRA map than in the JRC map. This advantage is more significant for stream segments with lower Strahler orders, which is taken as evidence that, due to the threshold source area of 500 km^2, the JRC map neglects a significative number of minor streams. As a result, the ISPRA map is selected as the target flood hazard map.

6.3 Reproduction of target hazard maps

Panels (a) and (c) of Fig. 4 represent the standardized GFI (i.e. rescaled GFI values so that the maximum is equal to 1 – the highest susceptibility of being flooded, and the minimum is equal to 0 – the lowest susceptibility), and the p value

Table 1. Metrics for vertical accuracy of the considered DEMs. Higher values (corresponding to worse accuracy) are marked with darker coloured cells.

		Valsugana		Bologna		Rimini	
DEM	Resolution [m]	RMSE [m]	LE90 [m]	RMSE [m]	LE90 [m]	RMSE [m]	LE90 [m]
TINITALY	10	8.236	12.141	4.511	6.475	2.832	4.412
EU-DEM	25	18.722	26.865	3.968	6.091	2.041	3.185
SRTM	30	15.764	21.083	3.203	4.665	2.239	3.537
ASTGTM	30	16.621	24.577	6.769	9.885	5.683	8.917
AW3D30	30	12.93	20.597	5.264	8.329	2.967	4.431
HydroSHEDS	90	30.166	44.597	3.355	4.549	2.108	3.362
MERIT	90	15.043	21.955	3.466	4.598	1.98	3.074

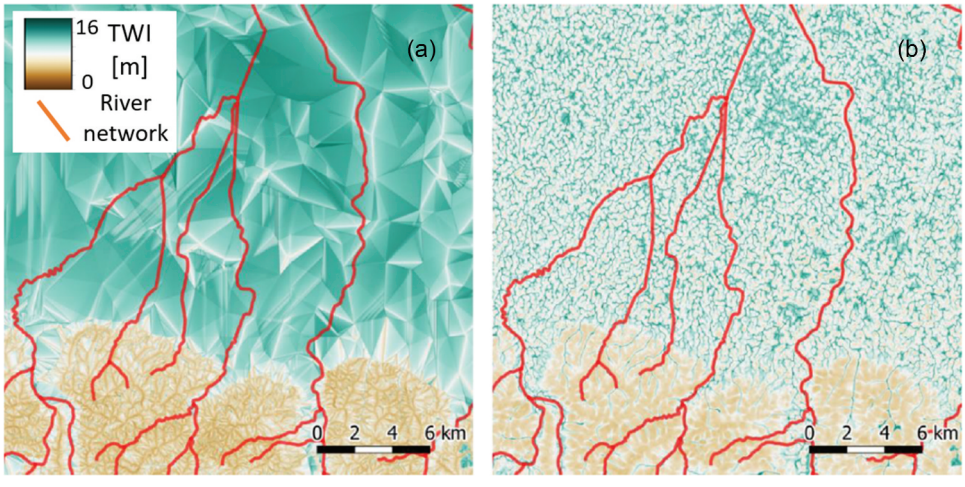

Figure 3. Focus on part of the Bologna test area. TWI (i.e. Topographic Wetness Index) computed from (a) TINITALY and (b) EU-DEM; the river network from the EU-Hydro dataset is marked in red (© European Union, Copernicus Land Monitoring Service 2021, European Environment Agency (EEA)).

Table 2. Percentage of overlap between the EU-Hydro river network and the considered reference flood hazard maps (i.e. the one from ISPRA - Istituto Superiore per la Protezione e Ricerca Ambientale, and the one from JRC - Joint Research Center). A darker colour means a higher percentage, which in turn means better agreement between the two datasets.

	Strahler order							
	1	2	3	4	5	6	7	8
Overlap with ISPRA map	10.30%	23.10%	37.60%	51.70%	64.80%	58.10%	99.60%	97.50%
Overlap with JRC map	0.02%	2.90%	3.20%	14.70%	58.80%	69.40%	90.60%	91.40%

computed by the decision tree (varying from 1, the highest probability for a pixel to be classified as "floodable," and 0, the lowest probability; see Section 2.2). These are the starting points from which the univariate and multivariate models derive their binary output maps (panels (b) and (d), respectively, of Fig. 4). Both figures show strong similarity between

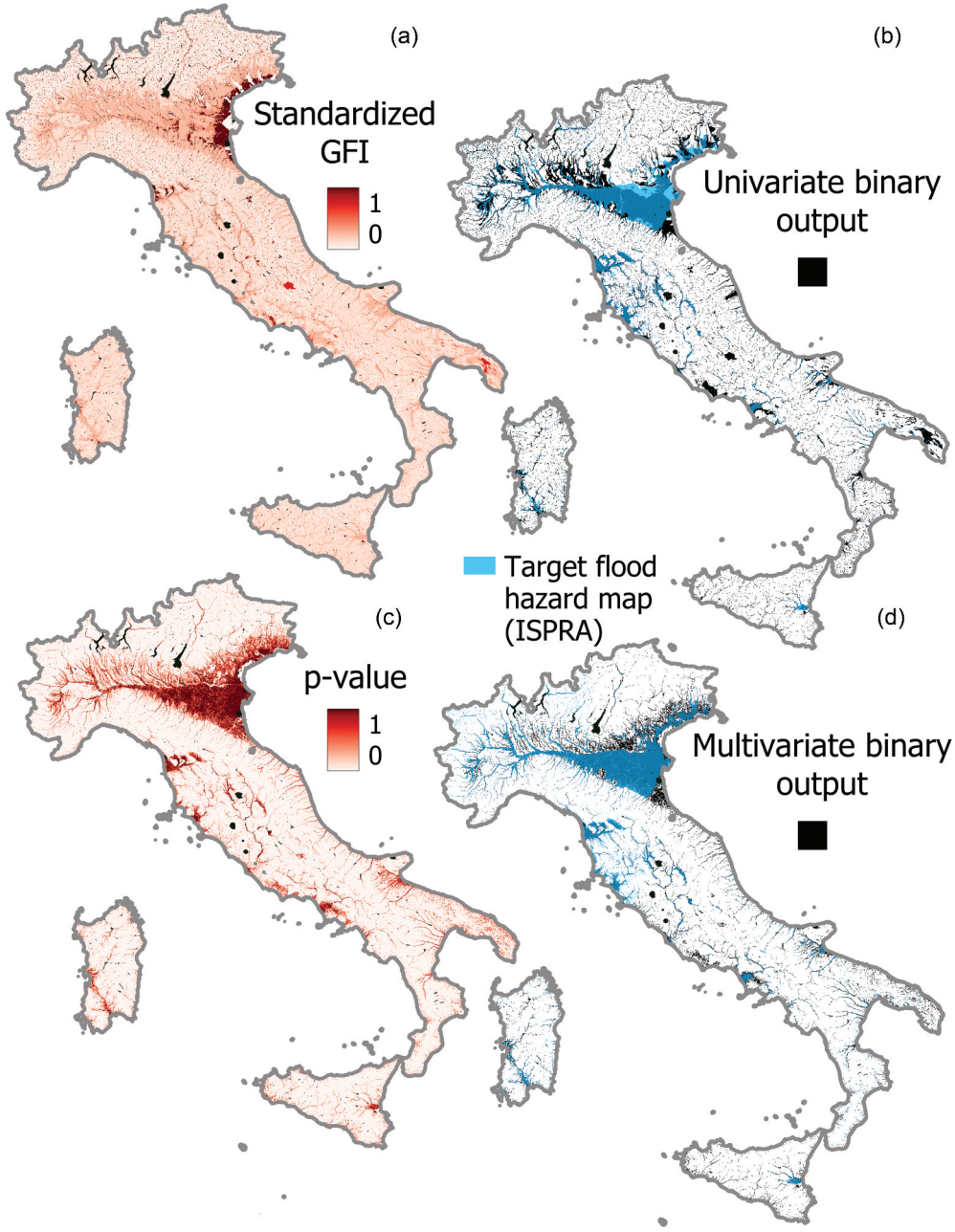

Figure 4. Output of DEM-based models. Univariate model: (a) standardized GFI (i.e. Geomorphic Flood Index) values (red scale); (b) binary flood hazard map (black). Multivariate model: (c) *p* value (red scale); (d) binary flood hazard map (black). In (b) and (c), blue shading represents the target ISPRA hazard map; dark blue shading identifies areas of overlap between the target and model maps.

Table 3. Performance metrics for the DEM-based hazard maps computed for testing pixels located inside a 200 m buffer area around the target flood hazard map. The highest and lowest values for each column are marked in bold and italic, respectively.

Model	TSS	ACC	PPV	TPR	F1
Univariate	*0.528*	*0.762*	*0.830*	**0.754**	*0.790*
Multivariate	**0.596**	**0.781**	**0.905**	*0.705*	**0.792**

the target and the output maps (i.e. dark blue areas in panels (b) and (d) of Fig. 4), which is particularly evident for the multivariate model.

Table 3 reports the performance metrics of the two DEM-based binary maps. It is evident that the multivariate method (second line) leads to better metrics (except for the TPR), but the difference from the univariate one (first line) is low, suggesting that the two binary outputs are very similar.

An additional evaluation is performed by computing the overlap between the flood-prone areas of the DEM-based binary maps and the EU-Hydro river network (Table 4), as was done for the selection of the reference ISPRA map (Section 7.2). The univariate model detects flood susceptibility on a significantly higher portion of the river network than in the ISPRA map. The exception is Strahler order 8, where the overlap drops to 70.72%, suggesting inaccuracy of the model. The increase in the overlap for the multivariate model is lower but still significant for Strahler orders from 3 to 6. For the others, the percentage of flood-prone areas over the river network is substantially the same as in the reference ISPRA map.

6.4 Validation against historical inundations and envelope flood hazard map

In Table 5 the agreement between the DEM-based models and the validation maps is shown, by means of the overlapping flood-susceptible areas. Considering the inundation maps (see the first three sections, from the left, in Table 5), it is evident that the output map from the multivariate model has the lowest agreement with the validation dataset. However, with the exception of the AL 21/10/19 event, which is more poorly represented in all the maps, the overlapping percentage is always more than 85%. Differently, the envelope validation map has higher agreement with the multivariate model (i.e. 94.45%) than with the univariate (77.19%).

Regarding the values assumed by the standardized GFI and the multivariate p value over the areas detected by the validation maps (Fig. 5), higher median values of the latter are observed (i.e. p value higher than 0.6, vs. standardized GFI lower than 0.4). It is again evident here that the event that occurred in Alessandria is more critical to model than the one in Bologna, as both the p value and the standardized GFI are partially under the classification threshold over the area.

The continuous indices and binary output maps of the DEM-based models can be better examined with a specific focus on the validation areas of the AL 21/20/19 event and the synthetic scenario (illustrative examples in Figs. 6 and 7, respectively). From this analysis, two points emerge: (1) the multivariate model leads to discontinuous floodplain delineation, while sharp floodplain boundaries are produced by the univariate one (panels (a–c) of Figs. 6 and 7); and (2) the multivariate model is more efficient in characterizing flood hazard outside the main river network (panels (d) and (e) of Figs. 6 and 7).

7 Discussion

7.1 Selection of the input DEM

On the importance of the selection of the appropriate DEM for calibrating a DEM-based flood hazard model, several studies have already been carried out (e.g. Tavares da Costa *et al.* 2019). Here, a framework for performing this operation is given. The method presented is valuable as it considers several characteristics of the DEMs: vertical accuracy, hydraulic consistency and resolution. In fact, even if the vertical accuracy is the simplest way to evaluate a DEM, the calibration of a DEM-based model requires also

Table 4. Percentage of overlap between the EU-Hydro river network, reference ISPRA flood hazard and binary outputs from DEM-based models. A darker colour means a higher percentage (see Table 2).

	Strahler order							
	1	2	3	4	5	6	7	8
Overlap with ISPRA map	10.30%	23.10%	37.60%	51.70%	64.80%	58.10%	99.60%	97.50%
Overlap with univariate map	40.60%	69.70%	82.81%	91.05%	92.14%	84.64%	98.61%	70.72%
Overlap with multivariate map	9.38%	22.03%	45.43%	71.41%	78.84%	73.64%	95.77%	95.95%

Table 5. Overlap between binary flood hazard maps (target: reference flood hazard map, ISPRA; univariate: GFI DEM-based model; multivariate: decision tree DEM-based model) and validation maps (i.e. observed inundation extents retrieved from satellite data, inundation scenario from 2D hydrodynamic modelling). The highest and lowest values for each column are marked in bold and italic, respectively.

	Observed inundation events							
	AL 21/10/19		BO 20/11/19		BO 21/11/19		Synthetic inundation scenario	
Model	Overlap area [km^2]	%	Overlap area [km^2]	%	Overlap area [km^2]	%	Overlap area [km^2]	%
Target	**16.16**	**76.79**	**37.43**	**98.59**	**14.93**	**99.60**	**3347.62**	**99.72**
Univariate	14.63	69.51	33.52	88.30	14.46	96.43	*2591.19*	*77.19*
Multivariate	*12.06*	*57.30*	*33.09*	*87.17*	*13.25*	*88.36*	3170.73	94.45

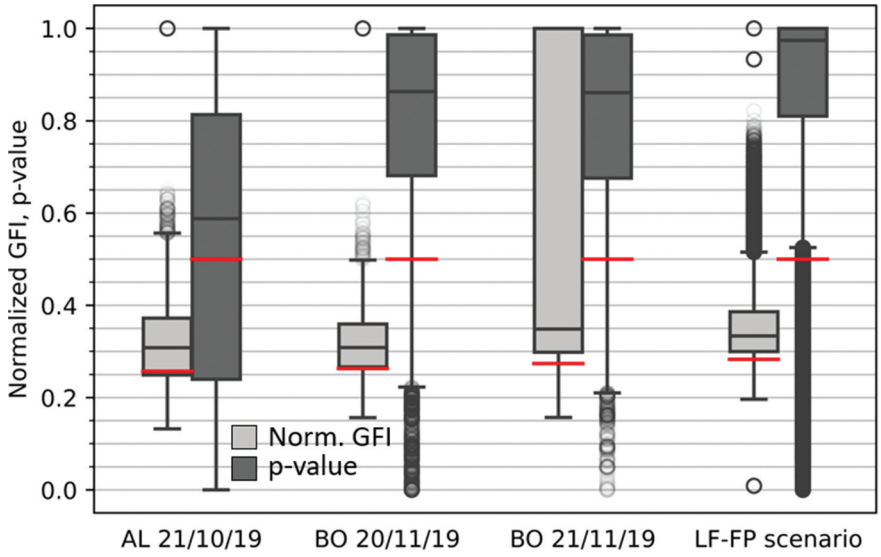

Figure 5. Box plot of standardized GFI (univariate model, light grey) and *p* value (multivariate model, dark grey) within the four inundated areas used in validation. Red lines indicate the thresholds for the decision tree classification (i.e. 0.5 at national level) and for the GFI classification (i.e. 0.265, 0.260, 0.249, 0.283 for AL 21/10/19, BO 20/11/19, BO 21/11/19 and the 2D envelope area, in this order).

good reliability of the river network extracted. Also, a DEM with higher vertical accuracy in flat areas, where the flood hazard is generally more difficult to estimate, has to be preferred over one with higher accuracy in mountainous contexts.

Indeed, one of the critical aspects of the proposed method is that some ground-truth data for evaluating vertical accuracy of DEMs may not be available. In this case, the selection should be based not only on the horizontal resolution but also on the hydraulic consistency. As an example, the river network or some geomorphic descriptors can be derived from the considered DEMs and compared with a reference river network (see Tavares da Costa *et al.* 2019). Alternatively, studies showing that a specific DEM has good performance over the target area can be followed.

7.2 Selection of the reference flood hazard map

An objective evaluation of the most appropriate reference flood hazard map is very difficult, as each map has its own advantages and weaknesses (see Section 5). In this study, a framework for quantitative selection is given.

We are aware that the coverage of river network extent and the accuracy of flood hazard estimation are two extremely different concepts. However, it is evident that in the case of the present study, there are several areas of Italy where the minor streams are susceptible to floods with a return period of 500 years according to the ISPRA map (see e.g. the extreme northwestern spot of Italy), while in JRC modelling they are not. Thus, it is reasonable to assume that these minor streams are accurately modelled in the ISPRA map, and that they do not appear as flood-susceptible in the JRC map due to the source area threshold (see Section 5). Thus, evaluating the portion of modelled river network that is covered by flood-prone areas can be an effective way to quantify the comprehensiveness of the hydraulic modelling, and is used here as a selection method for the reference hazard map.

This approach can be suitable for large study areas, as in the present study, but for medium- to small-scale applications other aspects of the reference maps should also be considered (e.g. the way hydraulic structures were modelled). Finally, the proposed method requires a river network dataset to be available to assume as reference, which should be rather feasible, thanks to open-source online resources such as EU-Hydro. Nevertheless, in a case where no dataset is available, a user-defined river network can be obtained by manual extraction from satellite data.

7.3 DEM-based modelling

The discussion of the DEM-based models is divided into two parts: (1) the comparison between univariate and multivariate modelling; and (2) the comparison between two possible ways to define flood susceptibility: binary mapping (floodplain delineation) and spatially continuous representation (flood hazard characterization).

7.3.1 Univariate vs multivariate modelling

When comparing the binary maps obtained across the entire study area (Table 3), the multivariate model achieves slightly better metrics than the univariate one. Focusing on smaller scales (i.e. single inundation events), the univariate model performs slightly better (Table 5, Observed inundation events). However, the univariate model produces a significantly worse prediction of the inundation scenario at the river branch scale (Table 5, Synthetic inundation scenario). On one hand, better overlap between inundation events and the univariate binary map confirms the validity of GFI. In fact, while its computation is straightforward, its accuracy for inundation susceptibility can be locally very significant. In particular, this is true when the thresholding is performed through a watershed-wise

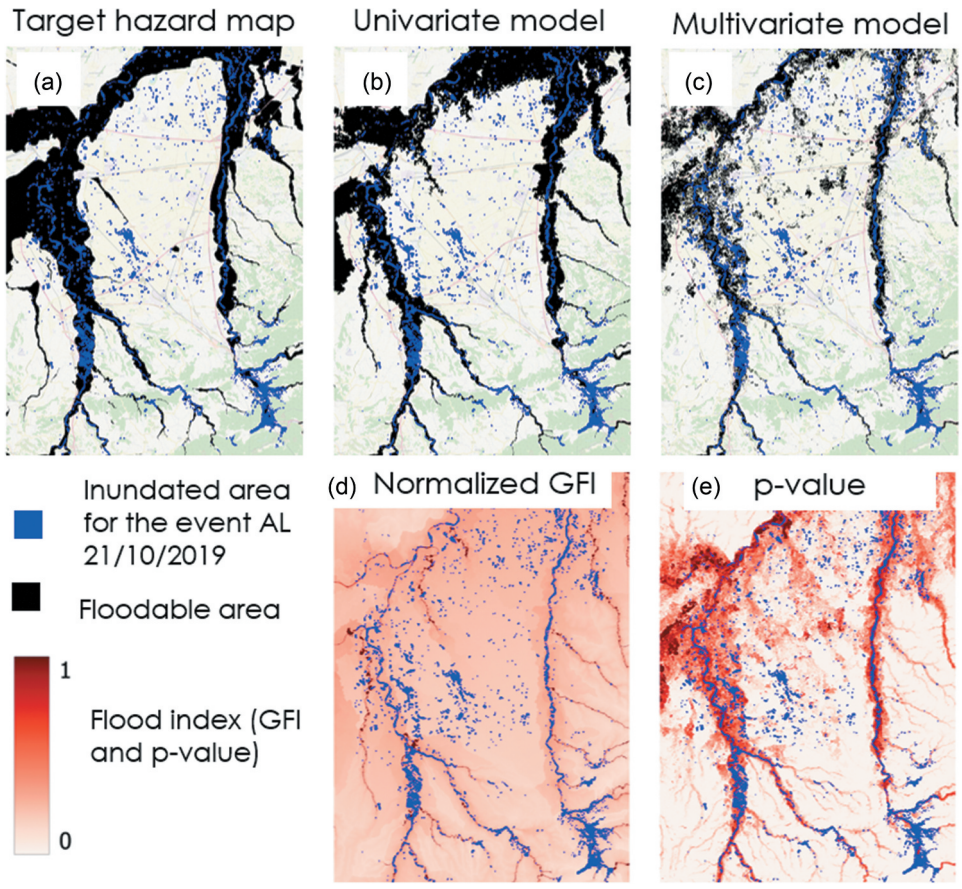

Figure 6. (a–c) Comparison of the floodable area (black) according to the target flood hazard, and DEM-based binary outputs with observed inundated areas (blue); (d, e) standardized GFI values and *p* value (colour scale) compared with inundated areas (blue) for the AL 21/10/19 event.

strategy, as in this case. On the other hand, the multivariate model trained nationwide shows higher accuracy in reproducing large inundation scenarios obtained through hydrodynamic modelling (Fig. 7). This is a clear indication that considering a variety of morphological descriptors as opposed to a single index leads to a better delineation of the envelope of all possible inundation events, which is the true objective of geomorphic floodplain delineation (on this point, see also the dark blue areas in panels (b) and (d) of Fig. 4).

Numerous discontinuities (i.e. isolated non-floodable pixels or small pixel clusters in floodable areas and vice versa) can be identified in the multivariate binary map (upper right corner of Figs. 6 and 7) by looking at specific areas in more detail. These discontinuities result from the combination of the *p* value thresholding (*p* value = 0.5) that is required in the production of a binary map, with the nature of an approach based on decision trees. In fact, as this multivariate approach is pixel-based, it does not explicitly enforce spatial coherence of the output. These isolated pixels are responsible for the lower metrics obtained while validating the multivariate flood-hazard map against the inundated areas for specific flood events (see the columns titled "Observed inundation events" in Table 5). In contrast, discontinuities are not present in the univariate flood-hazard map, due to the hydrological topologic consistence that characterizes GFI. In fact, the contributing area increases and the elevation above the nearest river-pixel decreases moving downstream, implying a monotonic increase of GFI in the same direction. Nevertheless, the GFI's local descriptiveness of the actual susceptibility of a pixel of being inundated may vary from region to region. Thus, relying only on GFI can lead to improved accuracy in mountainous areas and upper river segments (panel (b) in Fig. 6), and simultaneously to very significant and spatially broad inconsistencies in large predominantly flat areas (see e.g. the area north to the Po River in panel (b) of Fig. 7). These findings are in line with the literature, where consistency of univariate models is found to be higher in floodplain areas unaltered by humans (e.g. Nardi *et al.* 2018) and influenced by river stream order (Annis *et al.* 2019).

Finally, the increase in the extension of flood-prone areas in the DEM-based outputs with respect to the reference ISPRA map should be discussed. This is a major task, as no method exists to evaluate geomorphic-based flood hazard information where no reference is available from hydraulic models. However, it is quite reasonable to expect areas close to rivers to be flood-susceptible when considering a 500-year return period (see also Section 7.2); thus, the overlap between the flood-prone areas and the river network is considered in the present study. At a national scale, only the overlap for the DEM-based outputs can be compared with the one for the reference map (Table 4), while at a regional scale, the JRC map can also be used (Table 6). In fact, in some regions the JRC has a greater overlap with the EU-Hydro river network than the ISPRA map, showing inconsistencies of the latter. In these

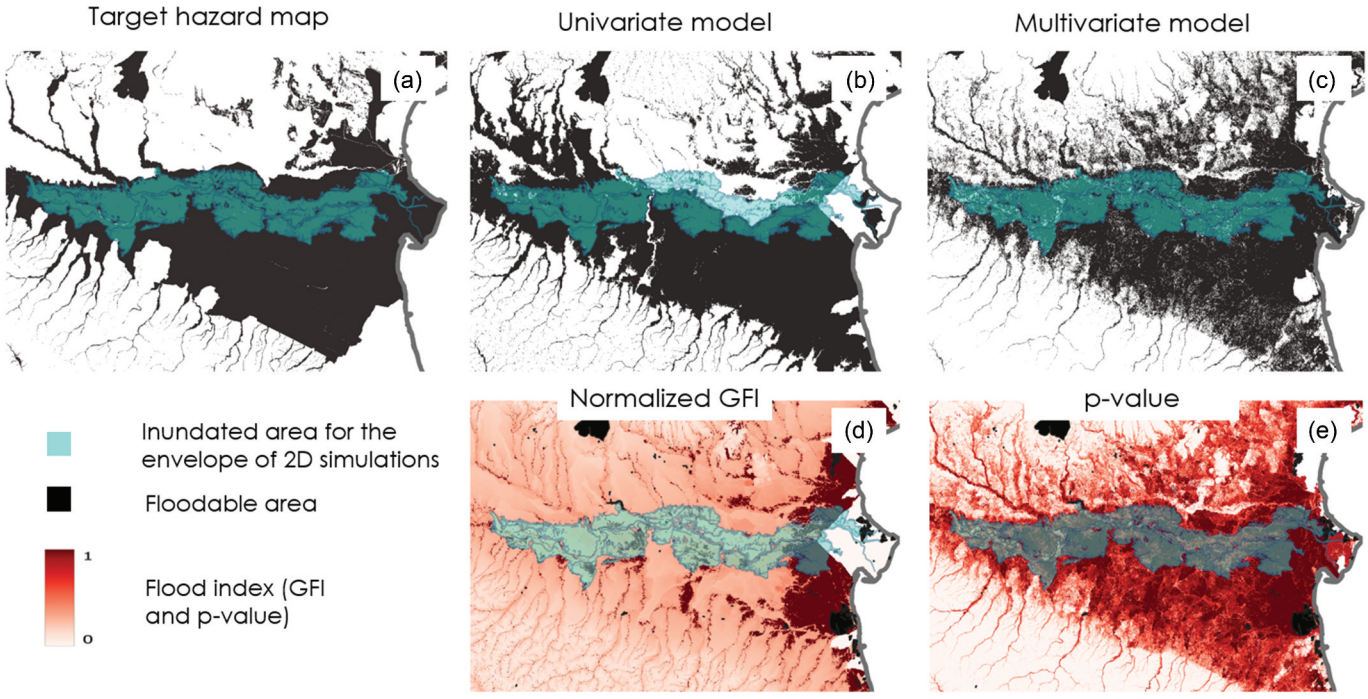

Figure 7. Comparison between the envelope of the synthetic inundation scenario (transparent light blue) and binary flood hazard maps [(a) target map, (b) univariate model, and (c) multivariate model] and continuous flood-susceptibility indices [(d) standardized GFI of the univariate model, and (e) *p* value of the multivariate model].

regions (Fig. 2), the overlap values for the DEM-based models and JRC map are very similar, and both are higher than those of the ISPRA map (Table 6). This shows that geomorphic approaches can effectively use the reference information collected elsewhere to accurately predict flood susceptibility where an exact reference is not available. Globally (Table 4), it is possible to consider as an advantage the increase in overlap for high Strahler orders (i.e. 3, 4, 5 and 6) observed for the multivariate model, as it represents an advance with respect to the probable inaccuracy of the reference ISPRA map. In contrast, the huge overlap for minor streams (i.e. Strahler orders 1, 2 and 3) for the univariate model should be considered with suspicion. These findings seem to confirm that DEM-based models can be useful to complete flood hazard information where it is already available, but with some inconsistencies (see also Lindersson *et al.* 2021).

7.3.2 Floodplain delineation vs flood hazard characterization

To conclude, concerning the informativeness of a binary geomorphological flood hazard modelling, as opposed to a continuous representation of flood hazard, it is worth comparing the standardized GFI values and the corresponding *p* values from the multivariate approach. By looking at panels (a) and (c) of Fig. 4, and panels (e) and (f) of Figs. 6 and 7, all adopting the same colour scale, it is possible to observe that both the indices assume higher values in floodplains and with greater river proximity, correctly estimating flood susceptibility in the entire spatial domain. However, the differences in the strength of the regional patterns produced by the two modelling approaches are rather striking. On one hand, the *p* values show a smooth and gradual decrease moving further away from the river network, and higher values in floodplains relative to standardized GFI values. On the other hand, very high standardized GFI values can be found in well-defined areas and very close to the river network. Then, the univariate flood susceptibility decreases rather abruptly when moving from the river network to its immediate proximity, having an overall variability between ~0.2 and 0.0 from a river floodplain to a mountain peak. This is confirmed by the box plots of Fig. 5, showing that the inundated areas of the studied flood events are characterized by lower standardized GFI values and narrower ranges relative to the *p* values associated with the multivariate model. An additional confirmation is the low GFI thresholds (red lines in Fig. 5) obtained in calibration for the four considered areas, and the corresponding wider floodplains delineated through the univariate flood hazard model.

The above considerations can support a more efficient and effective use of DEM-based flood hazard modelling products. In fact, most of the scientific literature focuses on the capability of these models to reproduce binary target flood-hazard maps. However, these models produce a great deal of information on susceptibility to inundation that is mostly lost if binary mapping, which we could also refer to as "floodplain delineation," is preferred to a continuous representation of flood susceptibility, which instead is closer to a "flood hazard mapping" in the strict sense. Just a few studies (e.g. Costache *et al.* 2020, Avand *et al.* 2022, Deroliya *et al.* 2022) show the application of DEM-based approaches for obtaining spatially continuous flood-susceptibility maps, but their models are trained on a pool of single inundation events. Thus, their analyses do not extensively focus on the information gain with respect to floodplain delineation, and rather show how their approaches can combine sparse information into a coherent output. In contrast, we train our models on binary flood hazard maps, and examine how these methods can improve information on flood hazard maps that is already available.

Table 6. Percentage of overlap between the EU-Hydro river network, reference ISPRA flood hazard, JRC map, and binary outputs from DEM-based models. A darker colour means a higher percentage (see Table 2). Only regions and Strahler orders where overlap for JRC is significantly higher than for ISPRA are reported. When regions do not have rivers of Strahler order 6, values are missing.

Region	Map	Strahler 3	Strahler 4	Strahler 5	Strahler 6
Trentino-Alto Adige	ISPRA	3.34%	10.21%	16.38%	5.05%
	JRC	5.92%	4.85%	46.32%	71.34%
	Univariate	57.61%	73.46%	87.89%	92.46%
	Multivariate	16.08%	39.28%	52.47%	53.73%
Veneto	ISPRA	11.03%	4.37%	16.41%	20.50%
	JRC	10.62%	3.37%	49.66%	55.44%
	Univariate	54.46%	65.95%	72.71%	65.02%
	Multivariate	29.70%	48.05%	63.41%	59.13%
Marche	ISPRA	2.52%	2.21%	0.19%	
	JRC	3.16%	19.92%	47.55%	
	Univariate	87.47%	92.65%	92.73%	
	Multivariate	60.39%	80.16%	74.62%	
Lazio	ISPRA	14.78%	35.56%	76.87%	48.63%
	JRC	4.90%	22.54%	79.57%	83.57%
	Univariate	72.47%	86.91%	98.72%	91.30%
	Multivariate	48.41%	72.10%	90.69%	84.72%
Abruzzo	ISPRA	23.36%	34.12%	49.23%	66.34%
	JRC	2.01%	5.65%	61.13%	87.04%
	Univariate	74.03%	85.97%	92.59%	95.90%
	Multivariate	35.59%	64.65%	72.39%	89.31%
Sicily	ISPRA	10.17%	35.59%	19.18%	
	JRC	1.01%	16.90%	78.50%	
	Univariate	84.28%	93.31%	97.90%	
	Multivariate	59.39%	82.19%	84.16%	

In this context, continuous representation of flood susceptibility can be used for the entire study domain, or limited to a buffer of delineated floodplains (i.e. binary maps), proving a graphical representation of the uncertainty of the binary map itself. Based on the outcomes of the present study, when a continuous spatial representation of flood hazard is considered, a simple multivariate approach to geomorphic flood hazard modelling seems to be associated with much higher potential and informativeness of a univariate modelling adopting a single, and yet very effective, morphological index.

Overall, the analyses above show the effectiveness of a new application of DEM-based models. So far, most of the literature describes how they can mimic pre-existing flood hazard maps (e.g. Manfreda *et al.* 2014, Nardi *et al.* 2018), derive spatially continuous estimation of flood susceptibility from multiple single measurements of inundation events (e.g. Costache *et al.* 2020, Avand *et al.* 2022), or predict flood hazard over data-scarce regions (e.g. Magnini *et al.* 2022). Differently, we gave evidence that DEM-based modelling can be used for enhancing incomplete or inexact information contained in national target flood hazard maps. This is possible thanks to their natural capabilities to produce spatially homogeneous and continuous flood-susceptibility maps, which seem to gain robustness when using a variety of morphometric indices instead of a single one. So far, the main limitations of geomorphic approaches – in particular in coastal and flat areas, as observed by other authors (e.g. Lindersson *et al.* 2021) – remain unsolved. Also, evaluating the goodness of the additional information produced by DEM-based models with respect to the reference maps (e.g. minor streams) is very challenging. In these cases, appropriate use of global flood hazard maps and ancillary datasets (e.g. JRC flood maps and the EU-Hydro river network) can be resolutive.

8 Conclusions

Recent scientific literature reports on a number of simplified and efficient methods for flood hazard modelling that combine the topographic information contained in DEMs with that of selected reference flood hazard maps, and are therefore often referred to as DEM-based (e.g. Manfreda and Samela 2019). Several models have been investigated, relying on single geomorphic descriptors (i.e. univariate models; see Noman *et al.* 2001, Manfreda and Samela 2019) or selected blends of them (i.e. multivariate models; see Degiorgis *et al.* 2012, Marchesini *et al.* 2021, Magnini *et al.* 2022), that in some cases are combined with additional information which is not retrievable from DEMs (e.g. see Arabameri *et al.* 2019, Costache *et al.* 2020). These methods are usually considered promising auxiliary/complementary tools to the traditional 2D hydrodynamic models for mapping riverine flood hazard. Their accuracy is lower than that of 2D inundation models when considering complex geometries (e.g. hydraulic structures) or flood dynamics (e.g. breaches), and yet their straightforward implementation and low computational intensity make them

particularly suitable for application to large-scale and high-resolution domains (Nardi *et al.* 2019).

Our study builds on a large-scale application of advanced univariate and multivariate DEM-based models to present an in-depth discussion of their potential and main limitations. The univariate model relies only on the GFI (Manfreda *et al.* 2014), which is one of the most accurate and versatile indices (e.g. see Samela *et al.* 2017). The multivariate model uses a data-driven blend of various DEM-based indices, including GFI. We selected Italy as study area, due to its wide spectrum of morphological and hydrological conditions.

There are two main innovative elements in the present study. The first is an effective framework for selecting the appropriate input DEM and reference flood hazard map, relying on the open-source EU-Hydro river network dataset (Gallaun *et al.* 2019). In fact, these represent major tasks for successfully setting up DEM-based models (see Trigg *et al.* 2016, Tavares da Costa *et al.* 2019) that are not fully addressed in the literature. The second is that the models are applied to resolve heterogeneities and inconsistencies of the same reference map used for training, instead of used in geographical extrapolation (e.g. Magnini *et al.* 2022). Thus, we analyse the ability of DEM-based models of handling the secondary river network, and producing spatially continuous and homogeneous characterization of flood hazard. These advantages are naturally offered by these methods, but as of yet only partially exploited (e.g. see Deroliya *et al.* 2022), and not fully discussed. Accordingly, we validate the two models against independent information – that is, remotely sensed inundated areas during three different flood events, and a synthetic catastrophic inundation scenario obtained as the envelope of several 2D hydrodynamic simulations.

In brief, the discussion of our results points out five main conclusions: (1) a univariate DEM-based model can be rather accurate if a meaningful geomorphic descriptor is considered (e.g. GFI) and a good calibration strategy is adopted; (2) our multivariate approach proved to be more accurate than the selected univariate one in reproducing the target information; (3) multivariate DEM-based models can be used to complete flood hazard information (with reasonable uncertainty) where the reference map is incomplete or inaccurate; (4) the spatially continuous representation of flood susceptibility (flood hazard mapping) should always be preferred to a binary representation (floodplain delineation) as it provides a wealth of information, e.g. on the uncertainty and descriptiveness of the simplified DEM-based model; and (5) in the case where spatially continuous flood susceptibility maps are to be prepared, multivariate approaches (e.g. p value from a decision-tree classifier) seem to be preferable to univariate ones (e.g. GFI alone) due to their higher descriptiveness and information content.

In conclusion, we suggest a more effective use of national flood hazard maps obtained from DEM-based models: we should not limit the application of the model to mimicking the national target flood hazard maps, which are often heterogeneous and inexact. Rather, we should exploit the training information contained in national target flood hazard maps to produce spatially homogeneous and continuous flood-susceptibility maps, possibly using a variety of morphometric indices instead of a single and very informative one (QGIS Development Team 2021).

Acknowledgements

The authors gratefully acknowledge the use of free and open-source software, in particular Python, Scikit-learn, QGIS, GRASS GIS, and TauDEM. Finally, the authors sincerely thank the reviewers (Luciano di Baldassarre and two anonymous referees), the editor Kate Hill, and the co-editor Konstantinos Soulis.

Authors' contribution

Andrea Magnini: conceptualization, methodology, investigation, software, writing – original draft. Michele Lombardi: conceptualization, coding. Armir Bujari: software, writing – original draft. Pietro Mattivi: investigation, validation. Iuliia Shustikova: methodology, writing – original draft. Simone Persiano: methodology, conceptualization. Marco Patella: methodology. Gabriele Bitelli: investigation, validation, writing – original draft. Paolo Bellavista: software. Francesco Lo Conti: funding acquisition, resources, conceptualization. Antonio Tirri: funding acquisition, resources, conceptualization. Stefano Bagli: methodology. Paolo Mazzoli: methodology. Attilio Castellarin: supervision, conceptualization, methodology, writing – original draft.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by Leithà S.r.l. – Unipol Group, under [grant nos. 66CAS17219, REP. 172/2019], and Autorità di bacino distrettuale del fiume Po, under [grant nos. L241CASTELLARIN12820, REP.128 PROT.3431].

ORCID

Andrea Magnini http://orcid.org/0000-0002-0704-1759
Simone Persiano http://orcid.org/0000-0002-9857-738X

Data availability statement

DEM analysis and set-up of the multivariate flood hazard model was performed using four sequential Python codes (Python version > 3.6). They are available according the Creative Commons Attribution 4.0 International Public License at the following URL: https://github.com/albadmin/hydro_risk_unibo. Part of the code (i.e. computation of the river network from a DEM) is hidden for reasons of intellectual property restriction clauses, but it can be easily replaced by common computation routines (e.g. assigning a user-defined threshold to the flow accumulation file that is produced by the code). The authors of the codes are Andrea Magnini, Armir Bujari, and Michele Lombardi. The input DEM (i.e. EU-DEM; see Bashfield and Keim 2011) can be downloaded free of charge from https://land.copernicus.eu/imagery-in-situ/eu-dem/eu-dem-v1.1. The other DEMs considered for the study are available at the following websites:SRTM: http://www2.jpl.nasa.gov/srtm/; ASTER GDEM: https://asterweb.jpl.nasa.gov/gdem.asp; AW3D30: https://www.eorc.jaxa.jp/ALOS/en/aw3d30/index.htm; TINITALY: http://tinitaly.pi.ingv.it/; EU-DEM: https://www.eea.europa.eu/data-and-maps/data/copernicus-land-monitoring-service-eu-dem; HydroSHEDS: https://www.hydrosheds.org/; MERIT

DEM: http://hydro.iis.u-tokyo.ac.jp/~yamadai/MERIT_DEM/. The official flood hazard by the Italian Institute for Environmental Protection and Research (ISPRA; see ISPRA 2018) is freely provided at the website https://idrogeo.isprambiente.it/app/page/open-data.

References

Abrams, M., 2016. ASTER global DEM version 3, and new ASTER water body dataset. *ISPRS - The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B4, 107–110. doi:10.5194/isprsarchives-XLI-B4-107-2016.

Alfieri, L., *et al*., 2014. Advances in Pan-European flood hazard mapping. *Hydrological Processes*, 28, 4067–4077. doi:10.1002/hyp.9947.

Annis, A., *et al*., 2019. Investigating hydrogeomorphic floodplain mapping performance with varying DTM resolution and stream order. *Hydrological Sciences Journal*, 64, 525–538. doi:10.1080/02626667.2019.1591623.

Arabameri, A., *et al*., 2019. A comparison of statistical methods and multi-criteria decision making to map flood hazard susceptibility in Northern Iran. *The Science of the Total Environment*, 660, 443–458. doi:10.1016/j.scitotenv.2019.01.021.

Avand, M., *et al*., 2022. DEM resolution effects on machine learning performance for flood probability mapping. *Journal of Hydro-environment Research*, 40, 1–16. doi:10.1016/j.jher.2021.10.002.

Bashfield, A. and Keim, A., 2011. Continent-wide DEM creation for the European union. *In*: *34th international symposium on remote sensing of environment - the GEOSS Era: towards operational environmental monitoring*. Available from: https://www.scopus.com/inward/record.uri?eid=2-s2.0-84879743325&partnerID=40&md5=1df8bbc919aa976e95574d2e71cedbf4.

Bates, P.D. and De Roo, A.P.J., 2000. A simple raster-based model for flood inundation simulation. *Journal of Hydrology*, 236, 54–77. doi:10.1016/S0022-1694(00)00278-X.

Beven, K.J. and Kirkby, M.J., 1979. A physically based, variable contributing area model of basin hydrology/Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant. *Hydrological Sciences Bulletin*, 24, 43–69. doi:10.1080/02626667909491834.

Canty, M.J., 2019. *Image analysis, classification and change detection in remote sensing: with algorithms for Python*. 4th ed. Boca Raton, FL: CRC Press/Taylor & Francis Group.

Costabile, S., 2010. Geoportale Nazionale: il piano straordinario di telerilevamento per l'ambiente. *GEOmedia*, 14 (3).

Costache, R., *et al*., 2020. Novel hybrid models between bivariate statistics, artificial neural networks and boosting algorithms for flood susceptibility assessment. *Journal of Environmental Management*, 265, 110485. doi:10.1016/j.jenvman.2020.110485.

Cui, D., *et al*., 2021. A 1 km global dataset of historical (1979–2013) and future (2020–2100) Köppen–Geiger climate classification and bioclimatic variables. *Earth System Science Data*, 13, 5087–5114. doi:10.5194/essd-13-5087-2021.

Degiorgis, M., *et al*., 2012. Classifiers for the detection of flood-prone areas using remote sensed elevation data. *Journal of Hydrology*, 470–471, 302–315. doi:10.1016/j.jhydrol.2012.09.006.

Deroliya, P., *et al*., 2022. A novel flood risk mapping approach with machine learning considering geomorphic and socio-economic vulnerability dimensions. *Science of the Total Environment*, 851, 158002. doi:10.1016/j.scitotenv.2022.158002.

Domeneghetti, A., *et al*., 2015. Evolution of flood risk over large areas: quantitative assessment for the Po River. *Journal of Hydrology*, 527, 809–823. doi:10.1016/j.jhydrol.2015.05.043.

Dottori, F., *et al*., 2016. Development and evaluation of a framework for global flood hazard mapping. *Advances in Water Resources*, 94, 87–102. doi:10.1016/j.advwatres.2016.05.002.

Faridani, F., *et al*., 2020. Estimating flood characteristics using geomorphologic flood index with regards to rainfall intensity-duration-frequency-area curves and CADDIES-2D model in three Iranian Basins. *Sustainability*, 12, 7371. doi:10.3390/su12187371.

Farr, T.G., *et al*., 2007. The shuttle radar topography mission. *Reviews of Geophysics*, 45, RG2004. doi:10.1029/2005RG000183.

Favalli, M., 2004. Digital elevation model construction from structured topographic data: the DEST algorithm. *Journal of Geophysical Research*, 109, F04004. doi:10.1029/2004JF000150.

Florinsky, I.V., Skrypitsyna, T.N., and Luschikova, O.S., 2018. Comparative accuracy of the AW3D30 DSM, ASTER GDEM, and SRTM1 DEM: a case study on the Zaoksky testing ground, Central European Russia. *Remote Sensing Letters*, 9, 706–714. doi:10.1080/2150704X.2018.1468098.

Gallaun, H., *et al*., 2019. *EU-hydro - river net user guide 1.3*. European Environment Agency. Available from: https://land.copernicus.eu/user-corner/technical-library/eu-hydro_user_guide.pdf.

Garcia, G.J.C., 2015. *EU-DEM upgrade - documentation EEA user manual*. Indra Sistemas S.A., Mar Egeo 4, Polígono Industrial n° 1 SAN FERNANDO DE HENARES, 28830, MADRID. Available from: https://land.copernicus.eu/user-corner/technical-library/eu-dem-v1-1-user-guide.

Gesch, D., *et al*., 2016. Validation of the ASTER global digital elevation model version 3 over the conterminous United States. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B4, 143–148. doi:10.5194/isprs-archives-XLI-B4-143-2016.

Gnecco, G., *et al*., 2017. Supervised and semi-supervised classifiers for the detection of flood-prone areas. *Soft Computing*, 21, 3673–3685. doi:10.1007/s00500-015-1983-z.

GRASS Development Team, 2019. *Geographic Resources Analysis Support System (GRASS) Software, version 7.6*. Open Source Geospatial Foundation. Available from: https://grass.osgeo.org.

Hastie, T., Tibshirani, R., and Friedman, J., 2009. *The elements of statistical learning*. Springer New York, NY: Springer Series in Statistics. doi:10.1007/978-0-387-84858-7.

ISPRA, 2018. Landslides and floods in Italy: hazard and risk indicators – summary report 2018, ISPRA Reports 287/bis/2018.

Jongman, B., *et al*., 2014. Increasing flood exposure in the Netherlands: implications for risk financing. *Natural Hazards and Earth System Sciences*, 14, 1245–1255. doi:10.5194/nhess-14-1245-2014.

Khosravi, K., *et al*., 2018. A comparative assessment of decision trees algorithms for flash flood susceptibility modeling at haraz watershed, northern iran. *Science of the Total Environment*, 627, 744–755. doi:10.1016/j.scitotenv.2018.01.266.

Koutsoyiannis, D., 2020. Revisiting the global hydrological cycle: is it intensifying? *Hydrology and Earth System Sciences*, 24, 3899–3932. doi:10.5194/hess-24-3899-2020.

Lehner, B. and Grill, G., 2013. Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems: global river hydrography and network routing. *Hydrological Processes*, 27, 2171–2186. doi:10.1002/hyp.9740.

Lindersson, S., *et al*., 2021. Global riverine flood risk – how do hydrogeomorphic floodplain maps compare to flood hazard maps? *Natural Hazards and Earth System Sciences*, 21, 2921–2948. doi:10.5194/nhess-21-2921-2021.

Magnini, A., *et al*., 2022. Machine-learning blends of geomorphic descriptors: value and limitations for flood hazard assessment across large floodplains. *Natural Hazards and Earth System Sciences*, 22, 1469–1486. doi:10.5194/nhess-22-1469-2022.

Manfreda, S., *et al*., 2014. Investigation on the use of geomorphic approaches for the delineation of flood prone areas. *Journal of Hydrology*, 517, 863–876. doi:10.1016/j.jhydrol.2014.06.009.

Manfreda, S., *et al*., 2015. Flood-prone areas assessment using linear binary classifiers based on flood maps obtained from 1D and 2D hydraulic models. *Natural Hazards*, 79, 735–754. doi:10.1007/s11069-015-1869-5.

Manfreda, S. and Samela, C., 2019. A digital elevation model based method for a rapid estimation of flood inundation depth. *Journal of Flood Risk Management*, 12. doi:10.1111/jfr3.12541.

Manfreda, S., Sole, A., and Fiorentino, M., 2008. Can the basin morphology alone provide an insight into floodplain delineation? *In*: D. Proverbs, C.A. Brebbia, and E. Penning-Roswell, eds. *Flood recovery, innovation and response I*. London, England: WITpress, 47–56. doi:10.2495/FRIAR080051.

Marchesini, I., *et al*., 2021. Data-driven flood hazard zonation of Italy. *Journal of Environmental Management*, 294, 112986. doi:10.1016/j.jenvman.2021.112986.

Mattivi, P., *et al*., 2019. TWI computation: a comparison of different open source GISs. *Open Geospatial Data, Software and Standards*, 4, 6. doi:10.1186/s40965-019-0066-y.

Mosavi, A., Ozturk, P., and Chau, K., 2018. Flood prediction using machine learning models: literature review. *Water*, 10, 1536. doi:10.3390/w10111536.

Mukherjee, S., *et al*., 2013. Evaluation of vertical accuracy of open source Digital Elevation Model (DEM). *International Journal of Applied Earth Observation and Geoinformation*, 21, 205–217. doi:10.1016/j.jag.2012.09.004.

Nardi, F., *et al*., 2018. Hydrologic scaling for hydrogeomorphic floodplain mapping: insights into human-induced floodplain disconnectivity. *River Research and Applications*, 34, 675–685. doi:10.1002/rra.3296.

Nardi, F., *et al*., 2019. GFPLAIN250m, a global high-resolution dataset of Earth's floodplains. *Scientific Data*, 6. doi:10.1038/sdata.2018.309.

Nardi, F., Vivoni, E.R., and Grimaldi, S., 2006. Investigating a floodplain scaling relation using a hydrogeomorphic delineation method: hydrogeomorphic floodplain delineation method. *Water Resources Research*, 42. doi:10.1029/2005WR004155.

Neal, J., Schumann, G., and Bates, P., 2012. A subgrid channel model for simulating river hydraulics and floodplain inundation over large and data sparse areas: floodplain inundation over large areas. *Water Resources Research*, 48. doi:10.1029/2012WR012514.

Netravali, A.N. and Hasskell, B.G., 1995. *Digital pictures: representation, compression and standards*. 2nd. New York: Plenum Press.

NOAA Office of Water Prediction, 2023. *Inundation mapping*. Available from: https://github.com/NOAA-OWP/inundation-mapping. Version:4.3.11.7.

Noman, N.S., Nelson, E.J., and Zundel, A.K., 2001. Review of automated floodplain delineation from digital terrain models. *Journal of Water Resources Planning and Management*, 127, 394–402. doi:10.1061/(ASCE)0733-9496(2001)127:6(394).

Pappenberger, F., *et al*., 2012. Deriving global flood hazard maps of fluvial floods through a physical model cascade. *Hydrology and Earth System Sciences*, 16, 4143–4156. doi:10.5194/hess-16-4143-2012.

Patel, A., Katiyar, S.K., and Prasad, V., 2016. Performances evaluation of different open source DEM using Differential Global Positioning System (DGPS). *Egyptian Journal of Remote Sensing and Space Sciences*, 19, 7–16. doi:10.1016/j.ejrs.2015.12.004.

Pedregosa, F., *et al*., 2011. Scikit-learn: machine learning in Python, arXiv [preprint]. *Journal of Machine Learning Research*, 12, arxiv:1201.0490.

QGIS Development Team, 2021. *QGIS geographic information system*. QGIS Association. Available from: https://www.qgis.org.

Rennó, C.D., *et al*., 2008. HAND, a new terrain descriptor using SRTM-DEM: mapping terra-firme rainforest environments in Amazonia. *Remote Sensing of Environment*, 112, 3469–3481. doi:10.1016/j.rse.2008.03.018.

Requena, A.I., *et al*., 2017. A bivariate trend analysis to investigate the effect of increasing urbanisation on flood characteristics. *Hydrology Research*, 48, 802–821. doi:10.2166/nh.2016.105.

Rodríguez, E., Morris, C.S., and Belz, J.E., 2006. A global assessment of the SRTM performance. *Photogrammetric Engineering & Remote Sensing*, 72, 249–260. doi:10.14358/PERS.72.3.249.

Samela, C., Troy, T.J., and Manfreda, S., 2017. Geomorphic classifiers for flood-prone areas delineation for data-scarce environments. *Advances in Water Resources*, 102, 13–28. doi:10.1016/j.advwatres.2017.01.007.

Sampson, C.C., *et al*., 2015. A high-resolution global flood hazard model: a high-resolution global flood hazard model. *Water Resources Research*, 51, 7358–7381. doi:10.1002/2015WR016954.

Sharma, A., Wasko, C., and Lettenmaier, D.P., 2018. If precipitation extremes are increasing, why aren't floods? *Water Resources Research*, 54, 8545–8551. doi:10.1029/2018WR023749.

Shustikova, I., *et al*., 2020. Levee breaching: a new extension to the LISFLOOD-FP model. *Water*, 12, 942. doi:10.3390/w12040942.

Tachikawa, T., *et al*., 2011. Characteristics of ASTER GDEM version 2. *In*: *2011 IEEE international geoscience and remote sensing symposium. Presented at the IGARSS 2011–2011 IEEE international geoscience and remote sensing symposium*. Vancouver, BC: IEEE, 3657–3660. doi:10.1109/IGARSS.2011.6050017.

Tadono, T., *et al*., 2016. Generation of the 30 M-MESH global digital surface model by ALOS prism. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, XLI-B4, 157–162. doi:10.5194/isprs-archives-XLI-B4-157-2016.

Takaku, J. and Tadono, T., 2017. Quality updates of 'AW3D' global DSM generated from ALOS PRISM. *In*: *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Presented at the 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. Fort Worth, TX: IEEE, 5666–5669. doi:10.1109/IGARSS.2017.8128293.

Tarboton, D.G., 2003. Terrain analysis using digital elevation models in hydrology. *In*: *23rd ESRI international users conference*. San Diego, California, 6–9 July.

Tarquini, S., *et al*., 2012. Release of a 10-m-resolution DEM for the Italian territory: comparison with global-coverage DEMs and anaglyph-mode exploration via the web. *Computers & Geosciences*, 38, 168–170. doi:10.1016/j.cageo.2011.04.018.

Tavares da Costa, R., *et al*., 2019. A web application for hydrogeomorphic flood hazard mapping. *Environmental Modelling and Software*, 118, 172–186. doi:10.1016/j.envsoft.2019.04.010

Tavares da Costa, R., *et al*., 2020. Predictive modeling of envelope flood extents using geomorphic and climatic-hydrologic catchment characteristics. *Water Resources Research*, 56 (9): e2019WR026453.

Tavares da Costa, R., Mazzoli, P., and Bagli, S., 2019. Limitations posed by free DEMs in watershed studies: the case of River Tanaro in Italy. *Frontiers of Earth Science*, 7, 141. doi:10.3389/feart.2019.00141.

Thomas, J., *et al*., 2014. Sensitivity of digital elevation models: the scenario from two tropical mountain river basins of the Western Ghats, India. *Geoscience Frontiers*, 5, 893–909. doi:10.1016/j.gsf.2013.12.008

Tøttrup, C., 2014. *EU-DEM statistical validation*. DHI GRAS Geocenter Denmark, Øster Voldgade 10, DK-1350, Copenhagen. Available from: https://ec.europa.eu/eurostat/documents/7116161/7172326/Report-EU-DEM-statistical-validation-August2014.pdf.

Trigg, M.A., *et al*., 2016. The credibility challenge for global fluvial flood risk analysis. *Environmental Research Letters*, 11, 094014. doi:10.1088/1748-9326/11/9/094014.

Trigila, A., *et al*., 2021. Dissesto idrogeologico in Italia: pericolosità e indicatori di rischio - Edizione 2021. ISPRA, Reports 356/2021.

Van Rossum, G. and Drake Jr, F.L., 1995. *Python reference manual, Centrum voor Wiskunde en Informatica Amsterdam*. Available from: https://ir.cwi.nl/pub/5008 [Accessed 11 May 2023].

Wang, Z., *et al*., 2015. Flood hazard risk assessment model based on random forest. *Journal of Hydrology*, 527, 1130–1141. doi:10.1016/j.jhydrol.2015.06.008.

Ward, P.J., *et al*., 2015. Usefulness and limitations of global flood risk models. *Nature Climate Change*, 5, 712–715. doi:10.1038/nclimate2742.

Winsemius, H.C., *et al*., 2013. A framework for global river flood risk assessments. *Hydrology and Earth System Sciences*, 17, 1871–1892. doi:10.5194/hess-17-1871-2013.

Yamazaki, D., *et al*., 2011. A physically based description of floodplain inundation dynamics in a global river routing model: floodplain inundation dynamics. *Water Resources Research*, 47. doi:10.1029/2010WR009726.

Yamazaki, D., *et al*., 2017. A high-accuracy map of global terrain elevations: accurate global terrain elevation map. *Geophysical Research Letters*, 44, 5844–5853. doi:10.1002/2017GL072874.

Youden, W.J., 1950. Index for rating diagnostic tests. *Cancer*, 3 (1), 32–35. doi:10.1002/1097-0142(1950)3:1<32::aid-cncr2820030106>3.0.CO;2-3.