Advances in Water Resources 102 (2017) 13–28

Contents lists available at ScienceDirect

Advances in Water Resources

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

Geomorphic classifiers for flood-prone areas delineation for data-scarce environments

Caterina Samela[a,*], Tara J. Troy[b], Salvatore Manfreda[a]

[a] Università degli Studi della Basilicata, Potenza, 85100, Italy
[b] Lehigh University, Bethlehem, PA 18015, USA

ARTICLE INFO

Article history:
Received 23 February 2016
Revised 19 January 2017
Accepted 19 January 2017
Available online 25 January 2017

Keywords:
Flood susceptibility
Data-scarce regions
Linear binary classifiers
Basin geomorphology
DEM
USA

ABSTRACT

Knowing the location and the extent of the areas exposed to flood hazards is essential to any strategy for minimizing the risk. Unfortunately, in ungauged basins the use of traditional floodplain mapping techniques is prevented by the lack of the extensive data required. The present work aims to overcome this limitation by defining an alternative simplified procedure for a preliminary floodplain delineation based on the use of geomorphic classifiers. To validate the method in a data-rich environment, eleven flood-related morphological descriptors derived from remotely sensed elevation data have been used as linear binary classifiers over the Ohio River basin and its sub-catchments. Their performances have been measured at the change of the topography and the size of the calibration area, allowing to explore the transferability of the calibrated parameters, and to define the minimum extent of the calibration area. The best performing classifiers among those analysed have been applied and validated across the continental U.S. The results suggest that the classifier based on the Geomorphic Flood Index (GFI), is the most suitable to detect the flood-prone areas in data-scarce regions and for large-scale applications, providing good accuracies with low requirements in terms of data and computational costs. This index is defined as the logarithm of the ratio between the water depth in the element of the river network closest to the point under exam (estimated using a hydraulic scaling function based on contributing area) and the elevation difference between these two points.

1. Introduction

Historically, man has always preferred to settle and live near the water. This tendency has not changed throughout time, and today nineteen of the twenty most populated agglomerations of the world (Demographia World Urban Areas, 2016) are located along watercourses or at the mouth of a river. On one hand, these locations are advantageous from many points of view. On the other hand, they expose significant populations and economic assets to a certain degree of flood hazard. The intensive urbanization may increase the occurrence of floods and, at the same time, the concentration of human population and socio-economic activities in areas exposed to risk may magnify the possibility that floods turn into a disaster (e.g., Hirabayashi et al., 2013; Jongman et al., 2012; Jonkman, 2005).

Knowing the areas exposed to flood hazard is essential to any strategy for minimizing the risk, which can then reduce loss of life, costly damages and disruption. Increasingly strict laws have been passed in recent years in many countries around the world (e.g. Floods Directive 2006 in Europe, Biggert-Waters Flood Insurance Reform Act 2012 of FEMA in the U.S.), yet a complete hydraulic flood hazard map is still lacking in many countries.

What is preventing such a mapping? It is not a lack of tools; in fact, there are numerous models to perform hydrologic/hydraulic analyses with different level of accuracy (e.g., Horritt and Bates, 2002; De Wrachien and Mambretti, 2011). Rather, the problem lies in the limited availability of adequate data for flood hazard studies, which require the fusion of hydrologic observations, historical information about past flood events, estimates of surface roughness and boundary conditions, and topographical surveys showing all natural and manmade features for the channel of interest and the adjacent areas. Unfortunately, gauging station networks are not widespread throughout the world and this scarcity is generally most pronounced in developing countries in Asia, Africa, or South America (e.g., Herold and Mouton, 2011), which furthermore suffer weak coping capacities due to limited resources to spend on flood protection. Hydraulic models have further drawbacks, being

* Corresponding author.
E-mail address: caterina.samela@unibas.it (C. Samela).

http://dx.doi.org/10.1016/j.advwatres.2017.01.007
0309-1708/

computationally-intensive, costly and difficult to apply over large study areas. Consequently, hydraulic hazard assessment over large, unstudied areas poses a significant challenge.

The determination of floodplain maps in data-scarce environments is a research topic of increasing interest, often related to the aim of producing global flood hazard maps. Several researchers have addressed the prediction of floods at ungauged sites by operating a transfer of hydrologic information using regionalization methods (e.g. Blöschl and Sivapalan, 1997; Merz and Blöschl, 2008; Padi et al., 2011). Some of them proposed simplified global scale models of surface water flows based on hydrological routing schemes driven by regional or global climate models (Pappenberger et al., 2012; Herold and Mouton, 2011).

An alternative to the typical hydraulic modelling approach is to use a basin's geomorphology. Specifically, one may obtain information about flood hazard exposure by analysing the basin's morphology that is the result of the complex interplay of many variables, such as climate, hydrology, geology, sediment load, vegetation, valley dimensions, and human activity. In particular, hydrological extremes and floods accelerate the processes of erosion, transport and deposition, and over long periods of time are able to shape and form geomorphic features (de Matauco et al., 2011).

Given this underlying assumption, several researchers investigated the mutual causal relationship that exists between floods and floodplain shape and extent, formulating simplified methods based on the characterization of the basin geomorphology (e.g., Williams et al., 2000; Noman et al., 2001; Gallant and Dowling, 2003; Mehlhorn et al., 2005; Dodov and Foufoula-Georgiou, 2006; Nardi et al., 2006; Manfreda et al., 2011; Degiorgis et al., 2012; Jalayer et al., 2014; De Risi et al., 2014; Manfreda et al., 2015; Samela et al., 2015).

Manfreda et al. (2014a) selected three DEM (Digital Elevation Model)-based approaches among the mentioned studies (Nardi et al., 2006; Manfreda et al., 2011; and Degiorgis et al., 2012), and carried out a comparative analysis on two sub-catchments of the Tiber River in Central Italy. The work emphasized the main characteristics of the investigated methodologies, revealing that methods based on morphological features provided a better description of the flooded areas. This study confirmed the ability of a single feature in characterizing the flood susceptibility, such as the difference in elevation (H) between the location under exam and the downstream river node to which the site is hydrologically connected. In general, this study provided a preliminary investigation about the role played by some morphologic features on flood exposure and it can be regarded as the starting point of the present research.

1.1. Study aims

In the introduced framework, the present study represents a further step forward, both in the number of descriptors investigated and the extent and relevance of the study area. The essential idea was learning from existing DEM-based procedures and defining a flood susceptibility mapping approach based on the analysis of the morphological signatures of the surface.

Two fundamental research questions guided this work:

- What are the most predictive geomorphological attributes for flood inundation processes?
- Is it possible to use such descriptors to map the flood exposure in data-scarce environments?

This study answers these questions with the overall aim of proposing an innovative method to identify preliminarily, but efficiently, the extent of floodplains in data-scarce environments and for large-scale analyses. This tool intends to be practical and cost-effective compared to the conventional engineering approach and based on information easily available worldwide (e.g., DEMs).

The identification of flood-prone areas was made through the use of linear binary classification techniques that have proven to be an appealing tool characterized by simple requirements in terms of input data, costs, and computational times. In particular, starting from morphological attributes of a basin related to hydrologic surface processes, eleven linear binary classifiers have been used to delineate flood-prone areas based solely on the information contained in DEMs. They include some traditional morphologic features (e.g., slope, contributing area, distance to the nearest channel, topographic convergence, etc.) as well as composite indices formulated with the specific aim to represent a metric of flood hazard. These indices of new formulation try to link some morphological characteristics of a basin to an estimate of the water depth calculated as a function of the contributing area, in order to take into account the essential role assumed by the flow depth in the flood inundation process.

Several applications and tests have been performed, changing the dominant topography of the training area and the scale of analysis, in order to investigate the relationship between basin characteristics and calibrated boundaries of classification (the thresholds of the linear binary classifiers), and to observe if the performances of a calibrated classifier are preserved in a different topographical context. To evaluate the reliability of each classifier, we adopted standard performance measures, such as the sensitivity, the specificity, and the Type I and Type II errors of the classification, which have been used to draw the Receiver Operating Characteristics (ROC) curves (Fawcett, 2006).

The behaviour of the selected classifiers has been analysed in a comprehensive study that involved ten sub-catchments of the Ohio River basin, the entire Ohio River basin, and then the water resources regions within the continental U.S. This allowed to identify the best performing descriptors, and to give a recommendation of a minimum extent of the calibration area required to perform a solid flood-prone areas identification.

2. Study area: the Ohio River basin

In this article, we focused our analysis on a large basin in the U.S., the Ohio River basin (Fig. 1). The Ohio River main stem is the fourth largest by discharge in the United States (mean annual discharge at Cairo, IL: 8733 m^3/s), and it is formed by the confluence of the Allegheny and Monongahela rivers in the Appalachian Mountains in Pittsburgh, Pennsylvania. From there, it flows roughly west-southwest for most of its length, approximately 1575 km, until it joins the Mississippi River near the city of Cairo, Illinois. The drainage area of the Ohio River is about 529,000 km^2.

The climate is temperate in the northern and mountainous regions and humid continental in the southeast, with abundant rainfall, cool moist winters, and warm summers. Precipitation for the basin is relatively evenly distributed throughout the year; the mountainous regions and the south receive more than 1200 mm per year, and the central and low-lying northern regions receive less than 1000 mm per year (Benke et al., 2011). The glacial history of the region has played an important role in controlling landforms and surface deposits. The elevations in the basin range from 100 m at the basin outlet to over 1700 m in the Appalachian Mountain headwaters (mean DEM grid cell elevation).

Floods have been a common occurrence on the Ohio River. Following the huge flood of 1937, 78 tributary flood-control dams were built. The coordinated use of these reservoirs allows water managers to provide up to half the flow of the river during droughts and to maintain a 3-m minimum pool depth throughout the river's length (Reid, 1991).

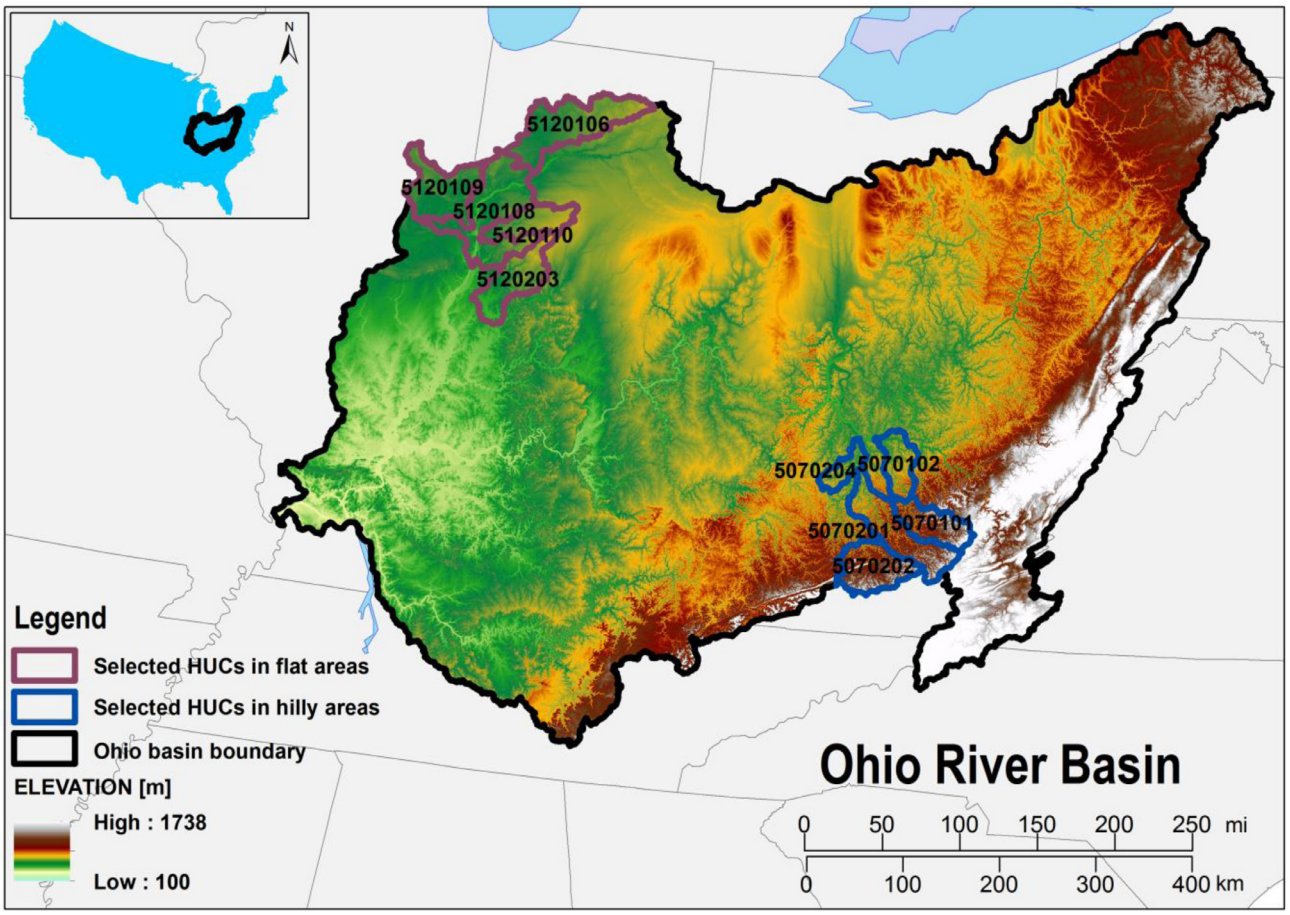


Fig. 1. The map shows the location of the Ohio River basin within the U.S., and the 90-m resolution DEM adopted to characterize its morphology (HydroSHEDS DEM-Void). The investigated Hydrologic Cataloguing Units (HUCs) are also represented: the HUCs in the flat region (Wabash basin) are depicted in violet, the HUCs in the hilly tributary (Big Sandy-Guyandotte region) are depicted in blue. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3. Datasets

3.1. Digital Elevation Models

A simple analysis of the basin morphology is possible using a Digital Elevation Model (DEM). Recently, a variety of DEMs, including Advanced Spaceborne Thermal Emission and Reflection Radiometer-Global Digital Elevation Model (ASTER GDEM) (Reuter et al., 2009), Shuttle Radar Topography Mission (SRTM) (Farr et al., 2007), and Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010), have been produced and offered freely for research purposes, followed by newly released improved versions (e.g. NASA SRTM v3, CGIAR-CSI SRTM v4.1, DLR/ASI SRTM X-SAR DEM, and ASTER GDEM v2). DEMs cover most of the populated regions of the world, with a minimum spatial resolution of 1 arc-second for SRTM and for ASTER, and 7.5 arc-seconds for the GMTED2010, and several studies have tested and evaluated their accuracy in different parts of the world (e.g. Hirt et al., 2010; Yamazaki et al., 2012; Rexer and Hirt, 2014; Athmania and Achour, 2014).

For the present study case, the DEMs of HydroSHEDS (Hydrological Data and Maps Based on Shuttle Elevation Derivatives at Multiple Scales) developed by the Conservation Science Program of the World Wildlife Fund have been used. This product is based primarily on elevation data obtained during NASA's Shuttle Radar Topography Mission (SRTM) with ancillary data sources (see Lehner and Döll, 2004). SRTM data are provided in near-global coverage from 56°S to 60°N. In its original release, the SRTM data contain regions with missing data point (voids or anomalies) and a large number of sinks or depressions. A void-filling procedure has been applied to remove all unnatural sinks, producing a continuous elevation surface called the void-filled elevation model (DEM VOID). Furthermore, in order to establish continuous flow for hydrological applications, a hydrologically conditioned elevation model (DEM-CON) has been produced by applying a sequence of conditioning procedures. Information about quality and shortcomings of HydroSHEDS data can be found in Lehner et al. (2008). The processing steps of generating HydroSHEDS are detailed in the data set's technical documentation (Lehner et al., 2006).

HydroSHEDS provides its data layers at multiple resolutions. In the present work, DEM-Void and DEM-CON with a 3 arc-second resolution have been used; the area was then regridded with a cell size of 90 m.

3.2. Standard flood maps

The standard flood hazard classification adopted in this work for calibration and validation purposes is the National Flood Hazard Layer (NFHL) 2015 of the Federal Emergency Management Agency (FEMA). The hazard classification identifies the 1-percent-annual-chance flood event (return period of $T = 100$ year), the 0.2-percent-annual-chance flood event ($T = 500$ year), and the areas of minimal flood hazard. Flood hazard areas are identified as Special Flood Hazard Areas (SFHA), defined as the areas that will be inundated by the flood event with a probability of exceedance of

1%. They have been determined using several methods of analysis, which include approximate methods, detailed hydraulic analyses, and specific analyses for coastal floodplains that have additional hazards associated due to storm-induced velocity wave action. Moderate flood hazard areas are defined as the area between 1 and 0.2% annual exceedance probability of inundation (T = 500 year). FEMA defines areas of minimal flood hazard as those outside the limits of the 0.2% annual exceedance probability. The hazard in such areas is low and tends to zero.

Although the 100-year flow is usually used as a standard in creating for flood risk maps, we use the 500-year flood in order to have a larger number of flooded points for the analysis.

This dataset has limitations, as one cannot know the method for each floodplain designation, but it is the standard floodplain designation method in the U.S. The map contains several gaps in different zones of the U.S. Moreover, the portions of the flood map designed as inundated because of storm surge in coastal areas or defined as “undetermined area”, “open water”, “area not included’ have not been considered for calibration and validation.

In this work, we initially analysed the cataloguing units within the Ohio River basin. Subsequently, the method was applied over the major geographic areas, excluding Alaska, Hawaii and Caribbean region. In order to divide the study area into successively smaller sub-catchments, we referred to hydrologic units identified by the United States Geological Survey (USGS) and identified by a unique hydrologic unit code (HUC) (Seaber et al., 1987).

4. Methodology

4.1. Linear binary classification and performance measures

A linear binary classification was performed to distinguish between flooded or not flooded areas. Linear binary classifiers seek to separate the data set in two classes using a linear boundary, which in this application is represented by a threshold value.

In the present work, we adopted several morphological features to perform the linear binary classification. In order to apply thresholds that cover the whole range of values it is convenient to standardize them through normalization. Specifically, the values were scaled and translated using a min-max normalization so that all data assumed a value between −1 and 1. Consequently, normalized thresholds, τ, were applied and iteratively changed over the portion of the basin used for calibration. Every threshold represents the boundary that divides the normalized values in two classes, and allows distinguishing flood-prone and non-flood-prone areas. Therefore, for every threshold we had a potential flood-prone area binary map that was compared to a standard flood map assumed as “gold standard truth”.

The evaluation of the performances for each descriptor first required the derivation of the number of cells identified as true positives (TP), false positives (FP), false negatives (FN) and true negatives (TN) predicted with respect to the standard flood map (Gold Standard Truth). Given these values, the standard metrics used to identify the errors and the correct predictions are:

- True positive rate or Sensitivity, which defines how many correct positive results occur among all positive samples available during the test: $R_{TP} = TP/(TP+FN)$.
- False negative rate (Error Type II), which defines how many incorrect negative results occur among all positive samples available during the test: $R_{FN} = FN/(TP+FN) = 1 - R_{TP}$.
- True negative rate or Specificity, which defines how many correct negative results occur among all negative samples available during the test: $R_{TN} = TN/(TN+FP)$.
- False positive rate (Error Type I), which defines how many incorrect positive results occur among all negative samples available during the test: $R_{FP} = FP/(TN+FP) = 1 - R_{TN}$.

The optimal normalized threshold for each descriptor was computed by minimizing the sum of the false positive rate and the false negative rate ($R_{FP}+R_{FN}$), thereby assigning equal weights to the two rates. In addition, the performances of the classifiers have been compared using the Receiver Operating Characteristics (ROC) curves and the Area Under the ROC Curve (AUC) (Fawcett, 2006). The value of the AUC ranges from 0.5 (completely random classifier) to 1.0 (perfectly discriminating classifier).

From the operational point of view, the standard flood hazard map was converted into a binary map, where the value 0 represents the areas not prone to floods (area of minimal flood hazard) and the value 1 represents the flood-prone areas. The optimal thresholds of each morphologic feature were first calibrated at the sample scale, then applied to extend the hazard information at the basin scale, and finally validated using the standard flood hazard maps for the study area. We exclusively focused on the use of classifiers based on a single descriptor. It has been observed that classifiers based on linear combinations of features may provide a limited improvement in performance, increasing severely the computational costs (Degiorgis et al., 2012; Manfreda et al., 2014a).

4.2. Morphological descriptors

Among the numerous features that describe the morphology of a basin, we identified the potential indicator of flood hazard exposure on the basis of their physical meaning, looking for the ones that have a role in the dynamics of flood propagation. They may refer to the tendency of water to accumulate in certain locations of the basin or the tendency of the gravitational force to move the water downhill. According to this criterion, we selected eleven morphological descriptors presumed to be good candidates as indicator of flood hazard exposure. Among them, we considered the five single morphological features analysed in the work of Degiorgis et al. (2012), with the addition of four composite indices that have been formulated with the specific aim to represent a metric for the flood hazard (Manfreda et al., 2014b; 2015; Samela et al., 2015). These indices link some morphological characteristics of a basin to an estimate of the water depth calculated as a function of the contributing area, in order to take into account the essential role assumed by the flow depth on floodplain delineation. Finally, along with these new indices, two additional topographic indices taken from literature have been also used. The selected descriptors are briefly described below:

Single features

1. upslope contributing area, A [m^2]: upslope portion of the watershed that contributes to water runoff to the point under exam.
2. surface curvature, $\nabla^2 H$ [-]: Laplacian of the elevation;
3. local slope, S [-]: maximum slope among the eight possible flow directions that connect the cell under exam to the adjacent cells;
4. flow distance to the nearest stream, D [m]: hydrologic distance from the location under exam to the nearest element of the reference drainage network;
5. elevation difference to the nearest stream, H [m]: difference between the elevation of the cell under exam and the elevation of the final point of the above-identified path.

Composite indices

6. ln (h_l/H): this index compares in each point of the basin a variable water depth h_l with the elevation difference H above defined. h_l is calculated for each basin location (‘l’ stands for ‘lo-

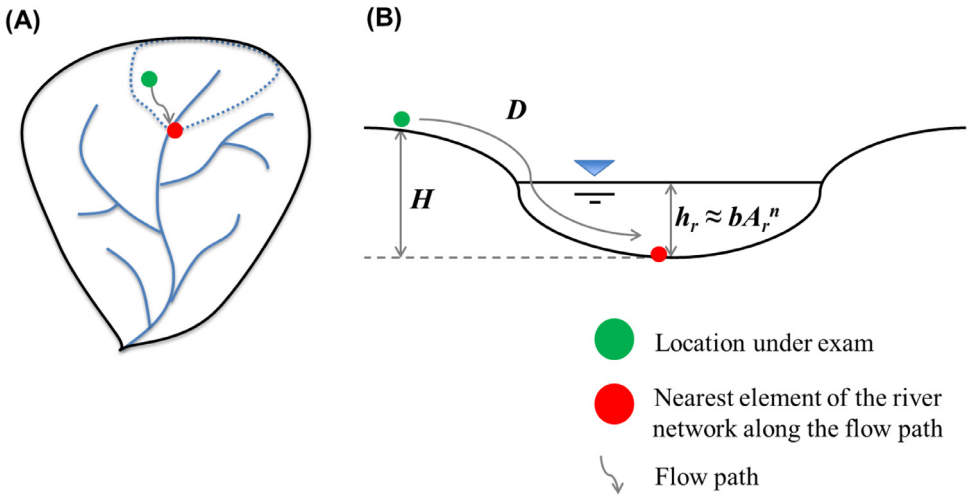


Fig. 2. Description of the Geomorphic Flood Index (GFI), $\ln(h_r/H)$. Representation of the parameters H and h_r in plan (A) and cross-section (B).

cal') assuming a scaling relationship with the local contributing area A_l, which typically follows a power law:

$$h_l \approx bA_l^n \quad (1)$$

where h_l is the derived water depth [m], A_l is the contributing area at the point of interest [km^2], b is a scale factor, and n is the exponent (dimensionless) that have been calibrated for the study area. Therefore, h_l is an estimate of the water level in the location under exam due to the amount of water that flows from the upslope part of the basin and that is drained by the examined point.

7. Geomorphic Flood Index (GFI), $\ln(h_r/H)$: this index compares in each point of the basin a variable water depth h_r with the elevation difference H. In this case, h_r is computed as a function of the contributing area A_r in the nearest point of the drainage network hydrologically connected to the point under exam ('r' stands for 'river') (see Fig. 2):

$$h_r \approx bA_r^n \quad (2)$$

By taking into account an estimate of the water level in the nearest element of the drainage network h_r, we are considering the nearest river as the hazard source.

8. $(h_r\text{-}H)/\tan(\alpha_d)$: change between water depth h_r and the elevation difference H divided by a surrogate of the hydraulic gradient represented by the downslope index.
9. $(h_r\text{-}H)/D$: change between water depth h_r and the elevation difference H divided by the distance D.
10. modified topographic index, TI_m (Manfreda et al., 2011): this index is defined as:

$$TI_m = ln\left[\frac{a_d{}^n}{\tan(\beta)}\right] \quad (3)$$

where a_d [m] = drained area per unit contour length; $\tan(\beta)$ = local gradient; n = exponent (n < 1).

11. downslope index, DW_i (Hjerdt et al., 2004): it represents an estimate of the hydraulic gradient and is defined as:

$$\tan(a_d) = \frac{d}{L_d} \quad (4)$$

L_d [m] is the distance that a parcel of water has to travel along its flow path to lose a certain amount of potential energy d[m]. The downslope index was calculated in an iterative way, changing the value of d in the range $5 \div 40$ m. The best performance was obtained for d = 5 m, therefore it was chosen as optimal d value in the present case.

The features related to the hydrological paths, such as A_l, A_r, D, H, were derived using the hydrologically conditioned DEM, while the others were determined using the DEM-Void. The DEM-CON was also used to identify the drainage network on the basis of an area-slope criterion proposed by Giannoni et al. (2005). In particular, the channel is assumed to start from locations where the quantity AS^k (where A is the upslope contributing area, S is the local slope and k = 1.7) exceeds a given threshold – set equal to 10^5 m^2 – while its path to the outlet is identified by following the maximum slope direction. The threshold adopted for the network identification was chosen according to previous application on DEMs of the same data source and same resolution (Manfreda et al., 2014a,b).

5. Application and results

To answer the main research questions and understand what are the most meaningful geomorphological attributes as concerns the flood-prone areas identification, we tested the sensitivity of the selected classifiers to changes in the input data in terms of:

i. Dominant topography of the training area;
   - Are the thresholds identified for the various indices related to basin characteristics, such as topography?
   - How robust are the thresholds and how transferrable?
ii. Size of the training area;
   - What percentage of a basin's area requires calibration?

The linear binary classification was carried out on the Ohio River basin and its sub-catchments, following the procedure illustrated in Section 4. The broader range of morphological characteristics of this basin allowed the selection of several calibration areas, in order to study the differences in the obtained results. Furthermore, its almost complete flood hazard classification carried out by FEMA allowed the validation of the procedure over the entire basin area.

5.1. Calibration of the hydraulic scaling function

Among the selected descriptors, four indices take into account an estimate of the water level calculated for each basin location assuming a scaling relationship with the contributing area, which typically follows a power law, as showed in Eqs. 1 and 2. Parameters b and n have been calibrated for the study area, exploiting streamflow observations at 486 gauging stations within the Ohio River basin using a historical period length of 80 years (1934–2013). The gauge data were obtained from the USGS database

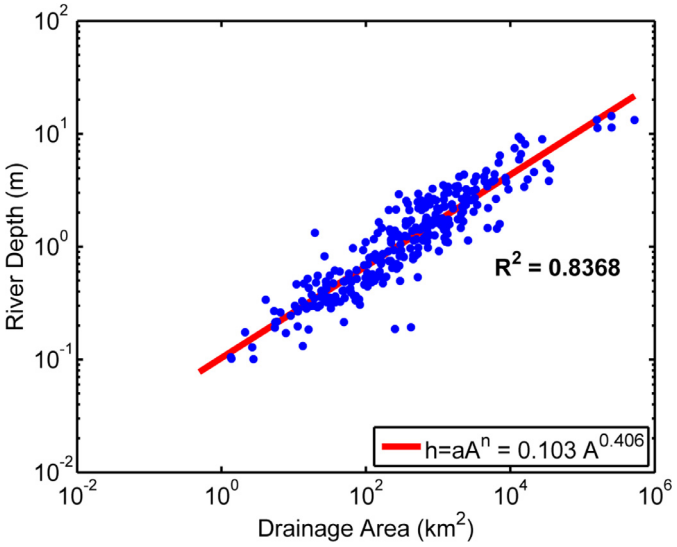


Fig. 3. The graph shows the USGS measurements in the Ohio Basin (blue dots) and the hydraulic scaling relationship (red line) calibrated using the stream gauge information. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

(URL: http://nwis.waterdata.usgs.gov/nwis/inventory) along with field measurements of width, area, and stage. Among them, we searched for the condition of the channel completely filled, with the aim to calibrate the hydraulic scaling function valid during flood events. Thus, the discharge data have been used to calculate the 95th percentile flow (roughly 2-year flood). For those measurements within 10% of the 95th percentile, we calculated the average depth. For the sites matching the search criteria, we used the stream gauge information for that flow to carry out a linear regression in log-log space between the water level, h, and the contributing drainage area, A, of the same station to estimate the coefficient and the exponent of the scaling law. The relationship obtained for the Ohio River Basin is presented in Fig. 3.

It worth noting that for the purposes of this kind of application the most important parameter of the scaling relation is the exponent n. Its value has some variation among different basins and ranges approximately between 0.24 and 0.45 (e.g. Gupta and Mesa, 2014; Leopold et al., 1964).

5.2. Test sensitivity of thresholds

The analysis of the geomorphic descriptors over the Ohio River basin started with investigating the sensitivity of the thresholds to basin topography. Therefore, we selected a flat and a hilly major tributary of the Ohio River, identified five smaller sub-basins in each, and calibrated the classifiers in each sub-basin individually to match the FEMA flood-prone areas. We used the estimated thresholds for each morphological descriptor to identify flood-prone areas of other sub-catchments within the major tributary. This allowed us to evaluate how the thresholds and the performances vary within sub-basins characterized by similar topography, and among basins characterized by very different topography.

As hilly tributary, we chose the Big Sandy–Guyandotte River basin, one of the steepest portions of the Ohio basin based on slope. Five sub-basins were identified in it, with drainage areas between 1000 and 4000 km^2 and an average slope between 18 and 34% (identified with dark blue contours in Fig. 1).

For the flat tributary, the Wabash River was used, which is the largest northern tributary (drainage area of 103,500 km^2). The five sub-basins chosen for calibration are depicted with violet contours in Fig. 1; they vary between 2198 and 6030 km^2, with an average slope between 2 and 4%.

More detailed information about the selected sub-basins are given in the Supplementary Material (SM) in Tables 1 and 2.

5.2.1. Calibration

The calibration was performed using individually five HUCs for the hilly area and five for the flat area. A comparison among the different descriptors was obtained drawing the ROC curves for each basin. For the sake of brevity, we report two representative examples referring to the steepest and flattest catchments studied (Fig. 4). Fig. 4A and B shows the results for the Upper Guyandotte River basin that has a mean slope of 34%. In these graphs, one can appreciate how the curve of $\ln(h_r/H)$, namely the GFI, outperforms the others, having the highest area under the curve (AUC = 0.92). The graphs in Fig. 5 allow a visual comparison of the results, showing the true positive rates (A) and the false positive rates (B) obtained in calibrating the eleven linear binary classifiers in the five hilly HUCs. For these hilly basins, the best performing features are the GFI and the elevation difference to the nearest channel, H. In the calibration phase, H was capable of correctly identifying the flood-prone areas with a true positive rate that ranges from 84–87% and an overestimation error between 16 and 26%. The GFI generally had better performances with a true positive rate between 85 and 92% and a false positive rate between 17 and 23%. The specific values obtained for each sub-basin are reported in Table 3 in the SM.

We repeated the procedure for the flat sub-basins and an example of the ROC curves obtained for the flattest one (Vermilion River basin, mean slope 2%) is reported in Fig. 4C and D. Here, one can clearly appreciate that the best performing predictor is the flow path distance, D, followed by the GFI. In this case, D was capable of correctly identifying the flood-prone areas with a R_{TP} that ranges from 74–86%, and an overestimation error between 18 and 29% (see Fig. 6). The GFI, achieved slightly lower performances, with a R_{TP} between 69 and 79% and a R_{FP} between 16 and 34%. Regarding the classifier based on the slope, S, we should not be deceived by what is apparently a good result in terms of R_{TP}, because this feature produced a rate of overestimation of about 100%. The results in terms of Sensitivity (R_{TP}) and Error Type I (R_{FP}) are depicted in the bar graphs of Fig. 6 considering all descriptors and basins studied.

The three classifiers identified as the best performing ones (D, H, GFI) were followed in terms of performances by $\ln(h_l/H)$. The specific values of the performance measures are reported in Table 4 in the SM.

5.2.2. Validation

After the calibration of the classifiers, we wanted to observe the parameter transferability within areas having a dominant topography similar to the training areas. Therefore, the binary classification was applied within the same tributary, using as test set the union of the five selected sub-basins, both for the flat and the hilly case. Then, the outcomes were validated using as performance measures the same multi-comparison statistics previously mentioned. Based on the previous results, only the four best performing classifiers (D, H, GFI, $\ln(h_l/H)$) have been evaluated.

The scatterplots in Fig. 7 show the performances of the classifiers obtained during calibration (x-axis) and validation (y-axis). Fig. 7A and B displays the true positive rates and false positive rates in hilly areas, Fig. 7C and D in flat areas. This plot demonstrates the potential transferability of a calibrated threshold of a given descriptor within a similar geomorphological context.

In the hilly tributary, descriptors D, H and $\ln(h_l/H)$ produced larger variability than the GFI around the 1:1 line. The GFI lost almost no performance in terms of R_{TP}. In terms of R_{FP}, all classi-

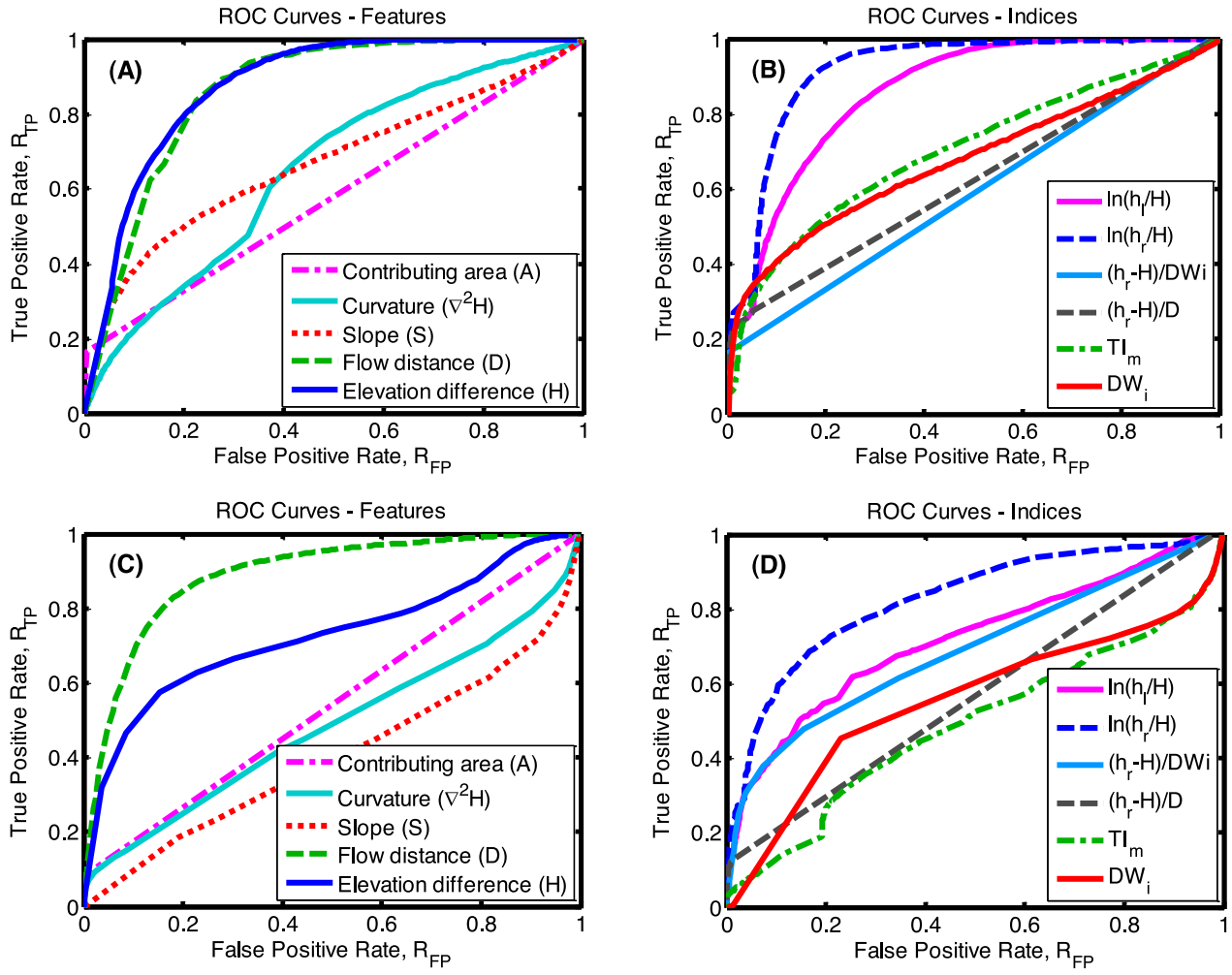


Fig. 4. Receiver Operating Characteristics curves (ROC) obtained for: (A) the five single features and (B) the composite indices for the steepest investigated HUC in Big Sandy – Guyandotte River basin (HUC5070101 - mean slope 34%); (C) the five single features and (D) the composite indices for the flattest investigated HUC in Wabash River basin (HUC5120106 - mean slope 2%).

fiers provided similar values in calibration and validation (differences are generally lower than 10%), apart from D, which in one case had an increase in R_{FP} of 14% and in another had a reduction in R_{FP} of 20%.

In the flat tributary, the classifier based on D provided good results during validation for both R_{TP} and R_{FP}. For H and ln(h_l/H), the R_{TP} decreased of about 12–17% for two cases, and improved in the others. The GFI showed a limited variability, but larger than what was observed in the hilly site. For the overestimation error (R_{FP}), most of the classifiers provided similar results in the validation phase, except for one point related to the descriptors H, GFI and ln(h_l/H).

All computed performance measures are reported in the SM in Table 5 and Table 6, for the hilly and flat tributaries, respectively.

5.3. Transferability of parameters: how robust and how transferable are the thresholds?

The previous analysis highlighted that optimal thresholds are sensitive to the characteristic topography of the basin. The normalized threshold of the twelve investigated sub-basins (five HUCs in flat areas plus the union of the HUCs, five HUCs in hilly areas plus the union of the HUCs) display that some descriptors have more stable behaviour than others (see Table 1). This is also confirmed by Table 2, where the range of variability is quantified by the coefficients of variation of the thresholds. Some descriptors, such as A, (h_r - H)/DW_i, (h_r - H)/D, and D, have almost constant thresholds across all the examined sub-basins. The thresholds of the composite indices GFI and ln(h_l/H) show a low variability. The threshold of H, DW_i and TI_m display similar values among the same group of catchments, but show a noticeable change in the transition from flat surfaces to hilly areas. The slope, S, is the most variable descriptor. It is so dependent on the topography that the normalized thresholds even change from positive values in the flattest basins, to negative values in basins with moderate/high slope.

To understand whether or not these thresholds are location specific or if they are applicable in different contexts, we swapped the normalized thresholds between flat and hilly basins, determining the sensitivity, specificity and the Type I and Type II errors. We took the optimal thresholds found in the Wabash basin for the four best performing descriptors and used them to detect the flood-prone areas in Big Sandy-Guyandotte basin areas, and vice versa.

When the thresholds calibrated in flat lands were used to delineate the floodplains in hilly areas, the performances of the single features H and D improved in terms of correct identification of flood-prone areas. However, they suffered a severe increase of the overestimation error, around 20% for D and over 50% for H. Instead, the GFI seemed to be more consistent in its performance, both in terms of R_{TP} and R_{FP}, and in one case it achieved a reduction of

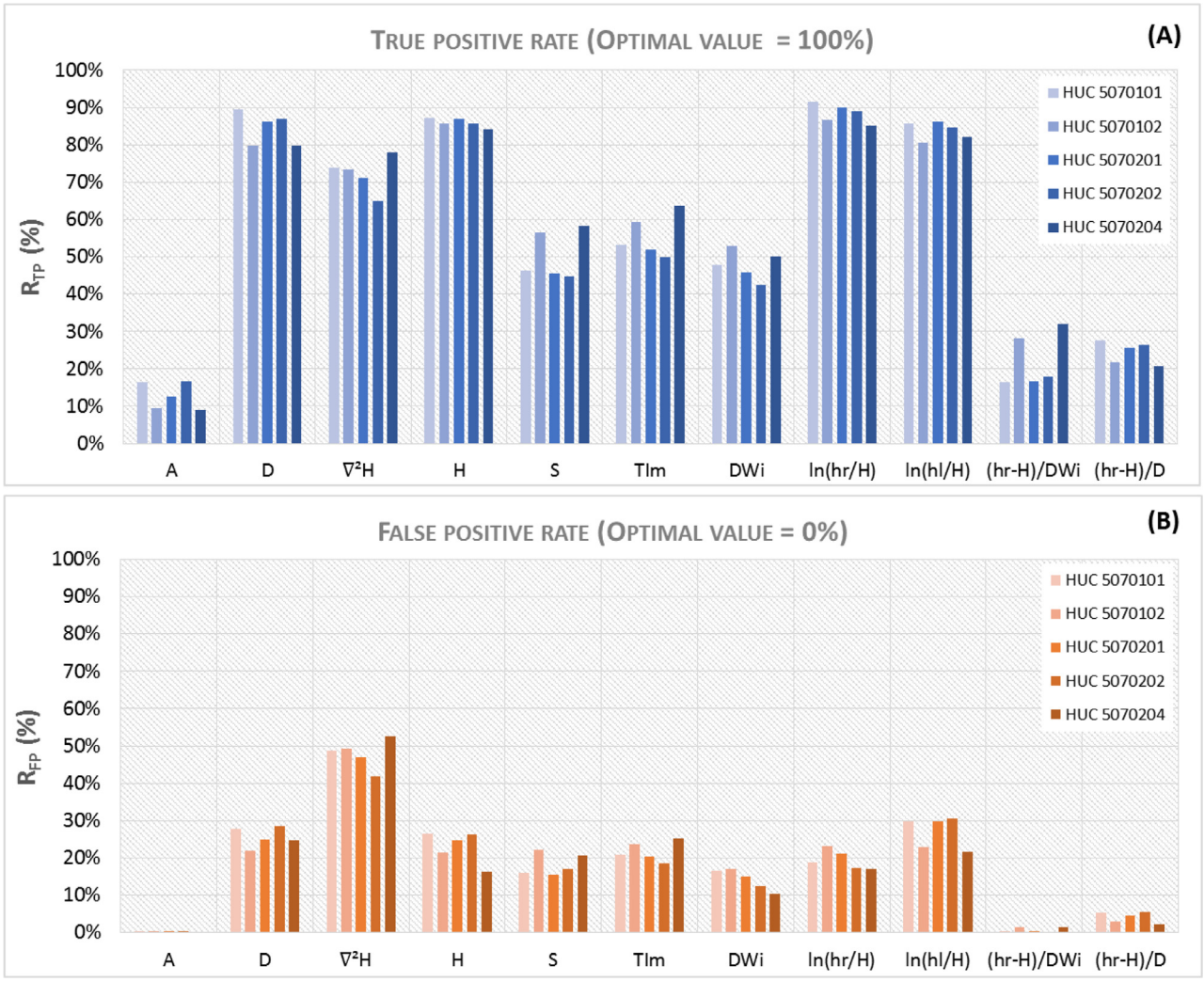


Fig. 5. (A) True positive rates and (B) false positive rates obtained for the eleven morphological descriptors after calibrating the optimal threshold in hilly areas.

the overestimation of about 20%. Fig. 8A and B shows the difference in terms of R_{TP} and R_{FP} obtained during the calibration in flat areas and the validation in hilly areas (more detailed results are reported in the SM in Table 7).

Conversely, when the thresholds calibrated in hilly areas were used to detect the flood-prone areas in flat lands, H suffered a total loss of its ability to predict the flood exposure. The performance of D decreased by about 20% in terms of correct identification of flood-prone areas. Again, the GFI provided almost constant performance in terms of R_{TP}, but with an increase of R_{FP} that ranges from 8% to 34%. Fig. 8C and D shows the difference in terms of R_{TP} and R_{FP} obtained for the calibration of the thresholds in hilly areas, and their validation in flat areas (Table 8 in the SM provides the values of the computed performance measures).

These results suggest that D and GFI are more transferrable than the others in areas with diverse topography, maintaining the ability to identify the flood-prone areas with performances and errors similar to those obtained during calibration. These tests also highlight the shortcomings of the elevation difference to the nearest channel, H. It seemed initially a good candidate when used in basins with a marked topography, and according to previous studies (Degiorgis et al., 2012; Manfreda et al., 2014). However, when it is applied in flat lands, it loses its ability to predict the flood exposure.

5.4. What percentage of a basin's area requires calibration?

The significant advantage of this methodology is the possibility to realize a fast delineation of the floodplains at the catchment scale, starting from the study of a small portion of the basin used for calibration. It is important to answer this question: "How small can this calibration area be?" For this purpose, an analysis has been performed partitioning the Ohio River basin into grids with cell sizes defined as a percentage of the total Ohio drainage area. Each element of these grids was used to calibrate the optimal normalized threshold for the four best performing descriptors: D, H, GFI, $\ln(h_l/H)$. Specifically, the assigned cell sizes, in terms of percentage of the Ohio basin, are: 4% (15,114 km^2), 3% (11,336 km^2), 2% (7559 km^2), 1% (3779 km^2), 0.5% (1890 km^2), 0.3% (1133 km^2), 0.1% (378 km^2), 0.08% (302 km^2), 0.06% (226 km^2), 0.04% (152 km^2), 0.02% (76 km^2), 0.01% (37 km^2). For example, using the grid with a cell size of 0.01% of the Ohio River Basin (37 km^2), and excluding the elements that do not have the presence of the river network, we obtained 6617 elements. We used each of these elements as calibration area, obtaining 6617 thresholds for each descriptor. We repeated this procedure using the other percentages assigned.

To investigate the variability of the thresholds when the calibration area is reduced from the bigger square elements to the smaller ones, we used this population of thresholds to calculate the coefficients of variation (CV) reported in Fig. 9. The graph clearly

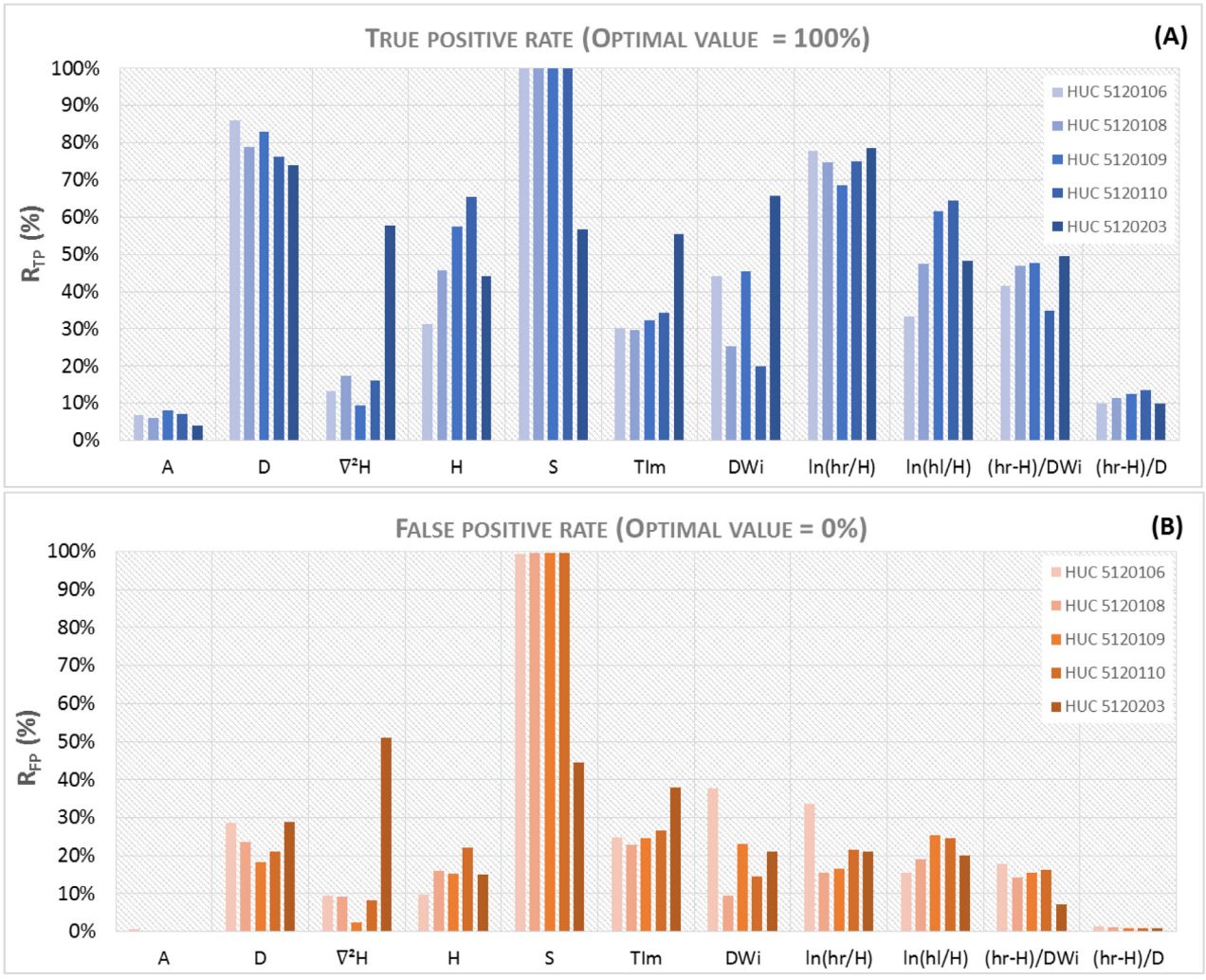


Fig. 6. (A) True positive rates and (B) false positive rates obtained for the eleven morphological descriptors after calibrating the optimal threshold in flat areas.

Table 1
Values of the optimal normalized thresholds obtained in calibrating the linear binary classifiers over the selected sub-basins in hilly and flat areas.

HUCs	A	D	∇^2H	H	S	TI_m	DW_i	$\ln(h_r/H)$	$\ln(h_l/H)$	$(h_r\text{-}H)/DW_i$	$(h_r\text{-}H)/D$
5120106	−0.995	−0.880	0.095	−0.630	0.325	−0.270	0.235	−0.570	−0.600	−0.985	−0.975
5120108	−0.995	−0.905	0.095	−0.595	0.575	−0.295	0.235	−0.515	−0.645	−0.995	−0.995
5120109	−0.995	−0.905	0.105	−0.615	0.395	−0.265	0.235	−0.480	−0.645	−0.990	−0.995
5120110	−0.995	−0.915	0.095	−0.595	0.790	−0.325	0.235	−0.525	−0.645	−0.995	−0.995
5120203	−0.995	−0.895	0.060	−0.580	−0.925	−0.465	0.240	−0.585	−0.685	−0.995	−0.995
Union 512	−0.995	−0.895	0.095	−0.615	0.788	−0.315	0.235	−0.550	−0.645	−0.990	−0.995
5070101	−0.995	−0.945	0.080	−0.775	−0.865	−0.655	0.125	−0.625	−0.760	−0.990	−0.995
5070102	−0.995	−0.935	0.080	−0.840	−0.885	−0.620	0.080	−0.590	−0.705	−0.990	−0.995
5070201	−0.995	−0.945	0.080	−0.785	−0.865	−0.655	0.115	−0.635	−0.760	−0.990	−0.995
5070202	−0.995	−0.945	0.095	−0.780	−0.865	−0.630	0.090	−0.610	−0.760	−0.990	−0.990
5070204	−0.995	−0.915	0.075	−0.865	−0.915	−0.600	0.045	−0.550	−0.690	−0.990	−0.995
Union 507	−0.995	−0.935	0.080	−0.815	−0.878	−0.645	0.090	−0.615	−0.735	−0.990	−0.990

demonstrates how the elevation difference, H, is affected by a substantial variability in the estimated thresholds. Even using training areas of about 15,114 km^2 (4% of the Ohio River basin), the thresholds already produce a coefficient of variation of almost 82%. This confirms that the parameter H is sharply influenced by the choice of the calibration area, which can lead to very different results.

The other three descriptors have much smaller CVs. In particular, the parameter D shows the smallest variability; considering the thresholds obtained for the 6617 training areas corresponding to 0.01% of the area of Ohio, we obtain a CV of only 8.9%. $\ln(h_l/H)$ has a CV value slightly higher than D. GFI shows a limited variability, with a CV lower than 10% for areas of 2% (corresponding to about 7500 km^2) of the total basin area.

For all the descriptors, we observed that by reducing the training area by thousands of square kilometres, and moving from 15,114 km^2 to half of it (7559 km^2), the CV maintains similar values. Then, the CV starts increasing steadily between 7559 km^2 (2%) and 1133 km^2 (0.3%), and a noticeable rise occurs for limited reductions of the area. Between 0.1% (378 km^2) and 0.01% (37 km^2), the

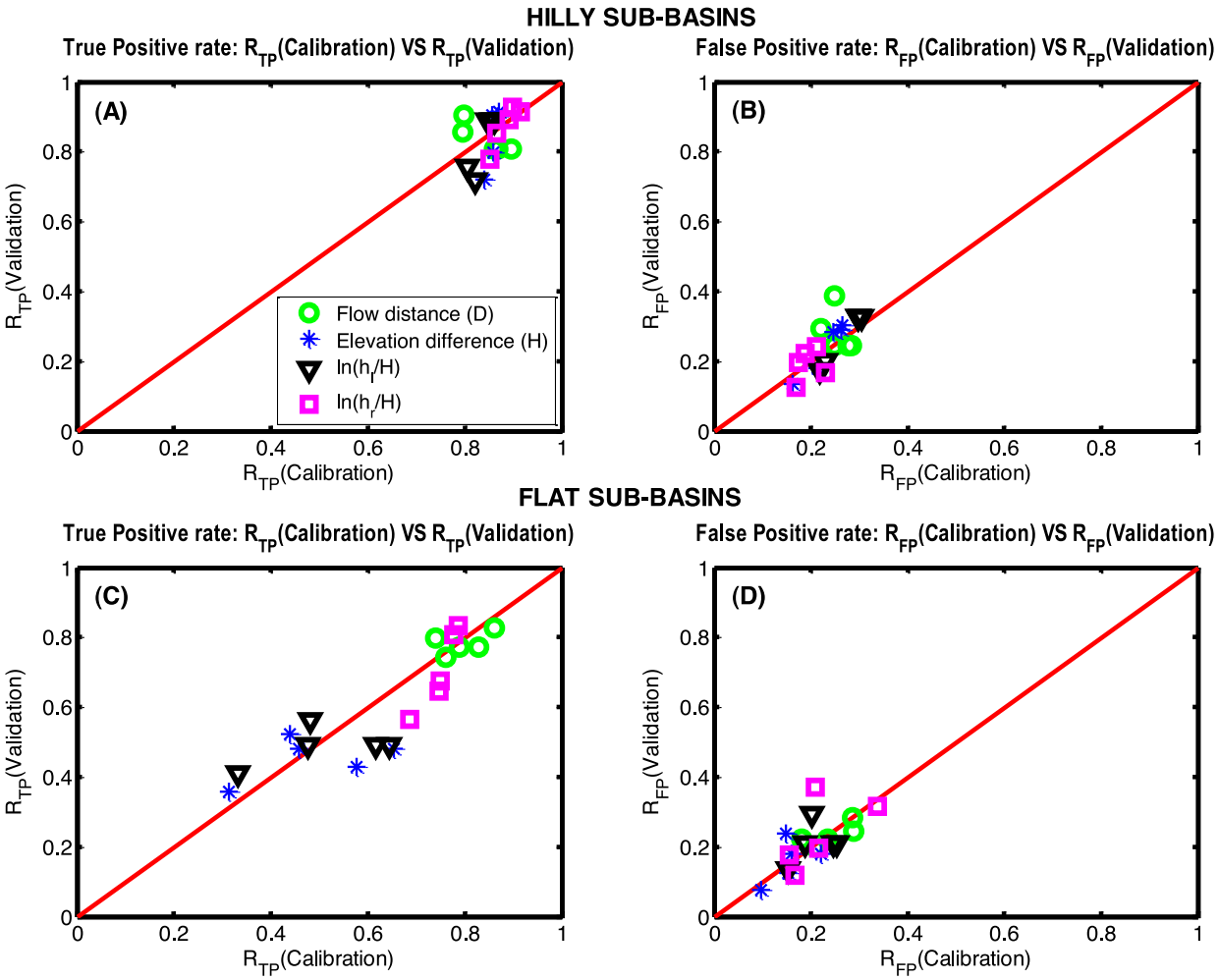


Fig. 7. Scatter plots representing the performances of the four best classifiers obtained during calibration (x-axis) and validation (y-axis). The distance from the 1:1 line gives an idea of the relative change in performance. Graphs provide: (A) true positive rates $\boldsymbol{R_{TP}}$ in hilly areas; (B) false positive rates $\boldsymbol{R_{FP}}$ in hilly areas; (C) the true positive rates $\boldsymbol{R_{TP}}$ in flat areas; (D) false positive rates $\boldsymbol{R_{FP}}$ in flat areas.

Table 2
Coefficients of variation calculated for the thresholds obtained in the 12 analysed sub-basins.

Thresholds	Coefficient of variation
A	0%
$(h_r\text{-}H)/DW_i$	0%
$(h_r\text{-}H)/D$	1%
D	2%
$\ln(h_l/H)$	8%
GFI	8%
$\nabla^2 H$	14%
H	16%
TI_m	36%
DW_i	48%
S	275%

CV rises sharply for extremely small reductions of the calibration area (see Fig. 1 in the SM).

In conclusion, the selection of a calibration area reduced up to 2% of the total area does not involve a significant increase in the CV of the calibrated thresholds and, within this range, one can expect to estimate a threshold value close to the optimal one. This result suggests not adopting a calibration area lower than 2% of the area of interest in order to accurately delineate flood-prone areas.

5.5. Application and validation of the procedure across U.S

As the ultimate test, the classifiers based on the four descriptors we are focusing on (D, H, GFI, $\ln(h_l/H)$) were used to identify the flood-prone areas across the continental U.S. The aim was to compare the performances of the selected classifiers for a large-scale application, in order to have further evidence that helps to decide which one is more suitable for the scope.

To cover the continental U.S., the FEMA National Flood Hazard Layers (NFHL) of 48 states were merged together. Considering the significant extent of the study area, the analyses were performed by sub-dividing the continental U.S. according to the eighteen water resources regions identified by USGS (http://water.usgs.gov/GIS/huc_name.html). The normalization of the descriptors, however, was executed across the entire U.S. in order to avoid abrupt and discontinuous delimitations of the flood-prone areas along the boundaries of the regions.

Within each region, training areas having a size of about 5% of that region have been randomly selected and used to calibrate the linear binary classifiers. Table 3 reports the calibration results for the water resources regions. Then, each classifier was applied over the entire region, and the resulting flood-prone areas maps were validated by comparison with the FEMA floodplain maps. To summarize the results of the validation for the different regions, we calculated how many times each classifier provided better perfor-

Table 3
Values of the optimal normalized thresholds and relative performance measures obtained in calibrating the linear binary classifiers in the eighteen water resources regions identified within the continental U.S. by the U.S. Geological Survey. The normalization of the classifiers has been executed considering the minimum and maximum values of the descriptors across all U.S.

D

Hydrologic unit code	Region name	τ	R_{TP}	R_{FN}	R_{TN}	R_{FP}	$R_{FP}+R_{FN}$	AUC
01	New England	−0.9822	62.5%	37.5%	74.2%	25.8%	63.3%	72.7%
02	Mid Atlantic	−0.9870	67.3%	32.7%	79.8%	20.2%	52.9%	78.6%
03	South Atlantic-Gulf	−0.9750	65.5%	34.5%	78.9%	21.1%	55.6%	78.3%
04	Great Lakes	−0.9633	67.0%	33.0%	70.3%	29.7%	62.7%	74.4%
05	Ohio	−0.9821	70.0%	30.0%	78.2%	21.8%	51.8%	81.1%
06	Tennessee	−0.9831	79.5%	20.5%	70.5%	29.5%	50.0%	81.4%
07	Upper Mississippi	−0.9727	66.2%	33.8%	74.4%	25.6%	59.4%	76.8%
08	Lower Mississippi	−0.9563	46.8%	53.2%	67.7%	32.3%	85.5%	59.3%
09−10	Missouri-Souris-Red-Rainy	−0.9732	77.3%	22.7%	66.7%	33.3%	56.1%	78.8%
11	Arkansas-White-Red	−0.9722	72.1%	27.9%	68.6%	31.4%	59.3%	76.9%
12	Texas-Gulf	−0.9708	67.7%	32.3%	72.9%	27.1%	59.4%	76.1%
13	Rio Grande	−0.9819	71.9%	28.1%	72.6%	27.4%	55.5%	77.5%
14	Upper Colorado	−0.9886	78.1%	21.9%	64.4%	35.6%	57.5%	76.2%
15	Lower Colorado	−0.9789	59.4%	40.6%	67.3%	32.7%	73.3%	67.7%
16	Great Basin	−0.9641	83.9%	16.1%	46.5%	53.5%	69.6%	69.6%
17	Pacific Northwest	−0.9875	56.9%	43.1%	68.2%	31.8%	74.9%	68.3%
18	California	−0.9400	60.2%	39.8%	53.0%	47.0%	86.9%	58.9%

H

Hydrologic unit code	Region name	τ	R_{TP}	R_{FN}	R_{TN}	R_{FP}	$R_{FP}+R_{FN}$	AUC
01	New England	−0.9789	82.8%	17.2%	69.1%	30.9%	48.1%	83.8%
02	Mid Atlantic	−0.9810	85.4%	14.6%	76.2%	23.8%	38.4%	89.2%
03	South Atlantic-Gulf	−0.9914	60.1%	39.9%	80.7%	19.3%	59.2%	75.3%
04	Great Lakes	−0.9922	73.8%	26.2%	61.8%	38.2%	64.4%	75.0%
05	Ohio	−0.9808	80.8%	19.2%	82.4%	17.6%	36.8%	89.0%
06	Tennessee	−0.9819	83.6%	16.4%	83.7%	16.3%	32.7%	90.3%
07	Upper Mississippi	−0.9904	60.1%	39.9%	75.2%	24.8%	64.8%	71.5%
08	Lower Mississippi	−0.9854	46.5%	53.5%	70.0%	30.0%	83.5%	61.1%
09−10	Missouri-Souris-Red-Rainy	−0.9859	71.0%	29.0%	74.8%	25.2%	54.3%	79.3%
11	Arkansas-White-Red	−0.9883	68.2%	31.8%	81.0%	19.0%	50.7%	81.7%
12	Texas-Gulf	−0.9904	75.1%	24.9%	66.9%	33.1%	58.0%	77.2%
13	Rio Grande	−0.9881	82.5%	17.5%	71.7%	28.3%	45.8%	82.4%
14	Upper Colorado	−0.9818	87.0%	13.0%	76.4%	23.6%	36.6%	88.6%
15	Lower Colorado	−0.9782	85.8%	14.2%	79.5%	20.5%	34.7%	88.2%
16	Great Basin	−0.9773	94.5%	5.5%	66.6%	33.4%	38.8%	84.8%
17	Pacific Northwest	−0.9750	87.9%	12.1%	70.5%	29.5%	41.6%	85.4%
18	California	−0.9654	95.3%	4.7%	55.9%	44.1%	48.9%	79.3%

$\ln(h_r/H)$

Hydrologic unit code	Region name	τ	R_{TP}	R_{FN}	R_{TN}	R_{FP}	$R_{FP}+R_{FN}$	AUC
01	New England	−0.5372	79.2%	20.8%	74.5%	25.5%	46.3%	83.8%
02	Mid Atlantic	−0.5310	80.7%	19.3%	81.8%	18.2%	37.5%	88.5%
03	South Atlantic-Gulf	−0.4941	75.6%	24.4%	85.0%	15.0%	39.4%	87.3%
04	Great Lakes	−0.5100	87.6%	12.4%	35.0%	65.0%	77.4%	64.6%
05	Ohio	−0.5221	81.8%	18.2%	82.0%	18.0%	36.2%	90.0%
06	Tennessee	−0.5225	85.8%	14.2%	82.2%	17.8%	32.0%	91.1%
07	Upper Mississippi	−0.4712	73.7%	26.3%	79.0%	21.0%	47.3%	81.5%
08	Lower Mississippi	−0.5038	63.6%	36.4%	54.7%	45.3%	81.6%	61.4%
09−10	Missouri-Souris-Red-Rainy	−0.4559	71.2%	28.8%	79.0%	21.0%	49.8%	82.2%
11	Arkansas-White-Red	−0.4739	77.1%	22.9%	77.2%	22.8%	45.7%	84.4%
12	Texas-Gulf	−0.4754	68.4%	31.6%	78.3%	21.7%	53.3%	80.2%
13	Rio Grande	−0.4913	77.1%	22.9%	79.7%	20.3%	43.2%	84.6%
14	Upper Colorado	−0.5209	89.2%	10.8%	80.3%	19.7%	30.6%	91.3%
15	Lower Colorado	−0.5243	79.8%	20.2%	72.7%	27.3%	47.5%	82.9%
16	Great Basin	−0.5373	88.6%	11.4%	69.5%	30.5%	41.9%	83.8%
17	Pacific Northwest	−0.5230	94.3%	5.7%	80.6%	19.4%	25.1%	91.7%
18	California	−0.5471	90.3%	9.7%	74.0%	26.0%	35.7%	86.3%

$\ln(h_l/H)$

Hydrologic unit code	Region name	τ	R_{TP}	R_{FN}	R_{TN}	R_{FP}	$R_{FP}+R_{FN}$	AUC
01	New England	−0.6656	78.1%	21.9%	73.0%	27.0%	48.9%	83.5%
02	Mid Atlantic	−0.6675	83.3%	16.7%	76.3%	23.7%	40.5%	86.8%
03	South Atlantic-Gulf	−0.6245	66.9%	33.1%	65.9%	34.1%	67.2%	72.4%
04	Great Lakes	−0.4917	61.5%	38.5%	55.2%	44.8%	83.4%	60.7%
05	Ohio	−0.6663	79.3%	20.7%	79.9%	20.1%	40.8%	87.8%
06	Tennessee	−0.6638	84.4%	15.6%	75.8%	24.2%	39.8%	88.3%
07	Upper Mississippi	−0.6169	66.1%	33.9%	73.9%	26.1%	59.9%	75.8%
08	Lower Mississippi	−0.5767	45.2%	54.8%	70.7%	29.3%	84.1%	61.2%

(*continued on next page*)

Table 3 (*continued*)

09–10	Missouri-Souris-Red-Rainy	−0.6341	75.3%	24.7%	71.2%	28.8%	53.5%	79.1%
11	Arkansas-White-Red	−0.6367	70.7%	29.3%	77.4%	22.6%	51.9%	81.1%
12	Texas-Gulf	−0.6292	74.4%	25.6%	67.9%	32.1%	57.7%	77.5%
13	Rio Grande	−0.6413	83.6%	16.4%	67.6%	32.4%	48.9%	82.8%
14	Upper Colorado	−0.6705	90.7%	9.3%	81.5%	18.5%	27.7%	90.0%
15	Lower Colorado	−0.6729	84.4%	15.6%	67.8%	32.2%	47.8%	82.2%
16	Great Basin	−0.6709	83.6%	16.4%	72.1%	27.9%	44.3%	84.4%
17	Pacific Northwest	−0.6875	83.9%	16.1%	70.4%	29.6%	45.7%	84.3%
18	California	−0.7100	90.8%	9.2%	48.1%	51.9%	61.1%	74.1%

mances with respect to the others, and expressed this number as a relative frequency. A visual representation of this outcome is reported in Fig. 10. The full validation values obtained in each region are provided in the SM in Table 9 and Table 10.

The ensemble of the computed performances endorses the classifier based on GFI as the most satisfactory among those analysed. Its sensitivity (R_{TP}) outstrips the other descriptors in eleven of the eighteen regions analysed (R_{TP} ranges between 57% and 89%), its specificity (R_{TN}) exceeds the other descriptors in fifteen regions, and the minimization of the total error provided by the GFI achieves the lowest values in all the analysed regions.

A pictorial representation of the flood-prone areas identified using the best performing classifier is reported in Fig. 11. Panel 1 in the upper part of Fig. 11 shows the areas exposed to flood hazard in the continental U.S. according to the linear binary classifier based on the GFI (depicted in dark blue). To better appreciate the results, in the same figure we also reported four subplots with a comparison at a smaller scale between the FEMA flood map and the flood-prone areas identified using the GFI. Panel 2.A and 2.B focus on the confluence between Mississippi and Missouri River, often involved in extensive floods. Image 3.A and 3.B show the state of Colorado, which had severe flooding in September 2013 that caused the death of 8 people and $2 billion of damages.

The large-scale maps make it difficult to visually recognize the overestimation and/or underestimation in the GFI flood-prone areas, but they do demonstrate that the GFI produces a realistic description of the flood-prone areas also in large and flat valley bottoms where other classifiers have proven to be misleading. Furthermore, they may help to provide information about the flood haz-

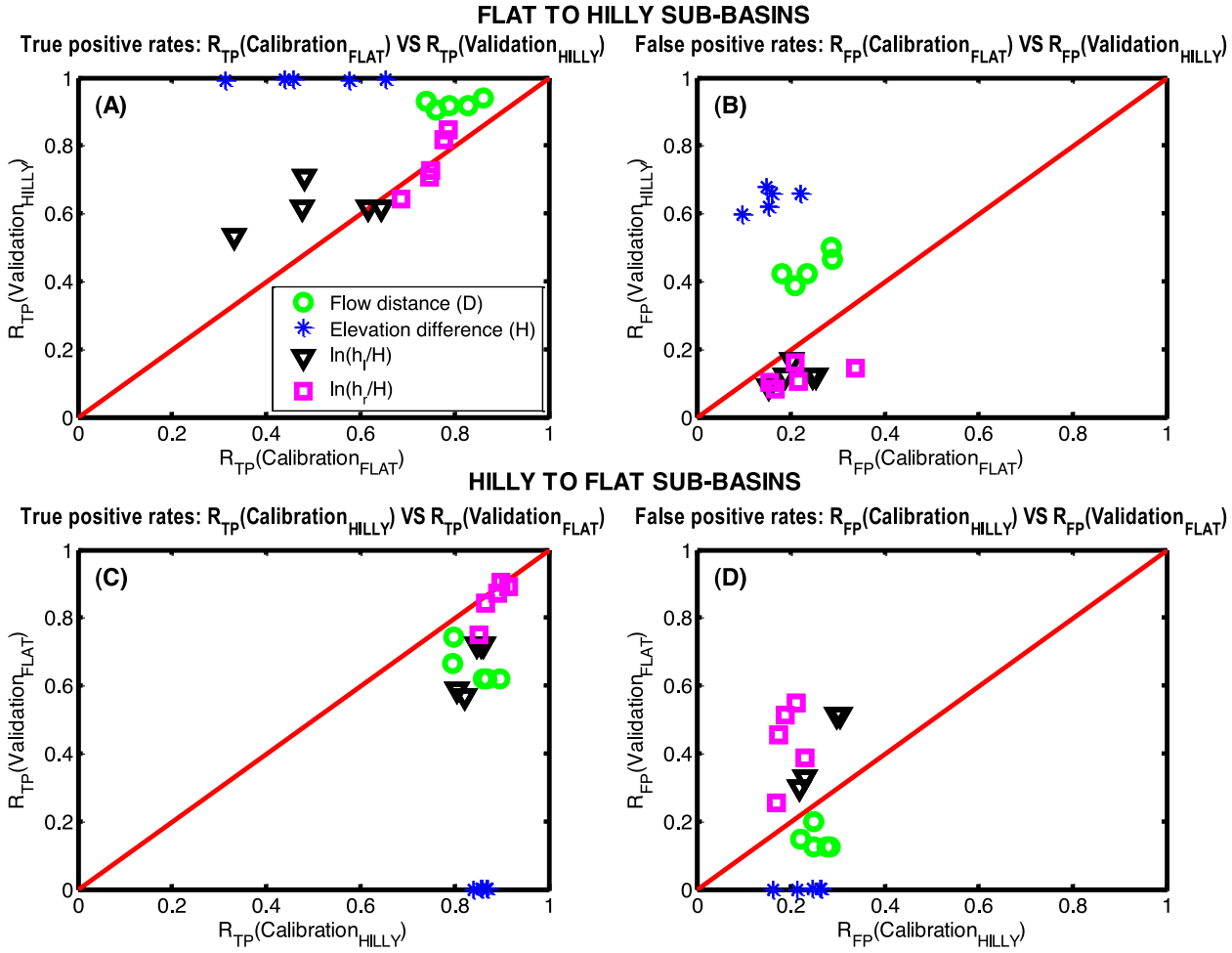
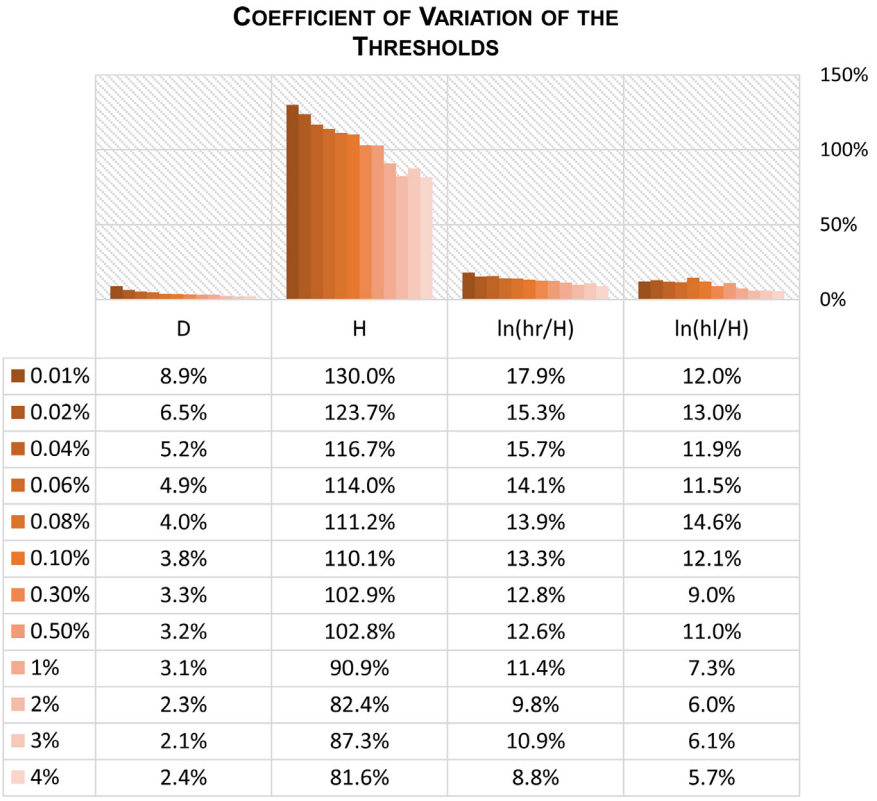


Fig. 8. Scatter plots representing the different performances of the four best classifiers obtained during calibration (x-axis) and validation (y-axis). The distance from the 1:1 line gives an idea of the relative change in performances. Graphs provide: (A) True positive rates and (B) false positive rates for thresholds calibrated in flat areas and applied in hilly areas; (C) true positive rates and (D) false positive rates for thresholds calibrated in hilly areas and validated in flat areas.



	D	H	ln(hr/H)	ln(hl/H)
0.01%	8.9%	130.0%	17.9%	12.0%
0.02%	6.5%	123.7%	15.3%	13.0%
0.04%	5.2%	116.7%	15.7%	11.9%
0.06%	4.9%	114.0%	14.1%	11.5%
0.08%	4.0%	111.2%	13.9%	14.6%
0.10%	3.8%	110.1%	13.3%	12.1%
0.30%	3.3%	102.9%	12.8%	9.0%
0.50%	3.2%	102.8%	12.6%	11.0%
1%	3.1%	90.9%	11.4%	7.3%
2%	2.3%	82.4%	9.8%	6.0%
3%	2.1%	87.3%	10.9%	6.1%
4%	2.4%	81.6%	8.8%	5.7%

Fig. 9. Coefficients of variation of the optimal thresholds computed as a function of the training area (expressed as a percentage of the Ohio River basin) for the four best-performing descriptors.

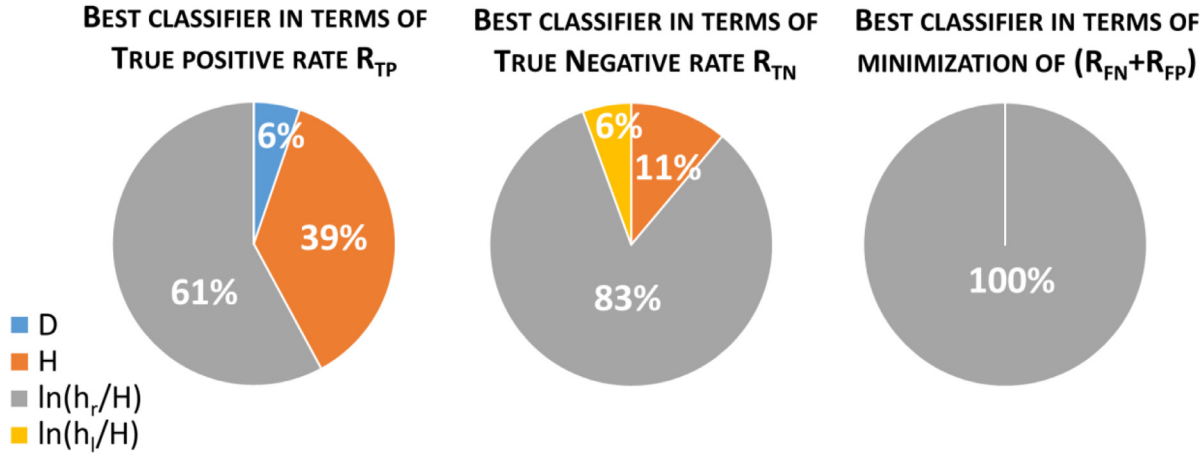


Fig. 10. . Pie charts demonstrating how many times each selected classifier outperforms the others in terms of true positive rate, true negative rate, and in minimizing the total error over the continental U.S. The numbers have been estimated as the relative frequency calculated by applying and validating the procedure over the eighteen water resources regions identified by the U.S. Geological Survey.

ard exposure in those portions of the country where FEMA maps are lacking. This happens, in particular, in central and western U.S., characterized by extensive voids in the FEMA flood hazard graduation.

6. Discussion

The ensemble of tests carried out depict a comprehensive description of the potential of different morphological indices in revealing the flood hazard exposure, identifying the most reliable and stable ones. Results allow delineating a strategy for a preliminary delineation of flood-prone areas and suggesting the reference area necessary for calibration.

Initially, the proposed procedure was applied using eleven classifiers over two regions: the first one included five sub-catchments within the Ohio River basin with a marked topography, the second one had five sub-catchments characterized by a flat topography. In spite of the exhaustive investigation conducted, very few classifiers were useful to identify the flood-prone areas. The main outcomes of this analysis can be summarized in the following points:

i. The composite index GFI achieved better performances in every hilly catchment with respect to any other classifier, and outperformed D in terms of every computed performance measure (R_{TP}, R_{TN}, $R_{FP}+R_{FN}$, AUC). Such performances were preserved in extending and validating the procedure outside the calibration

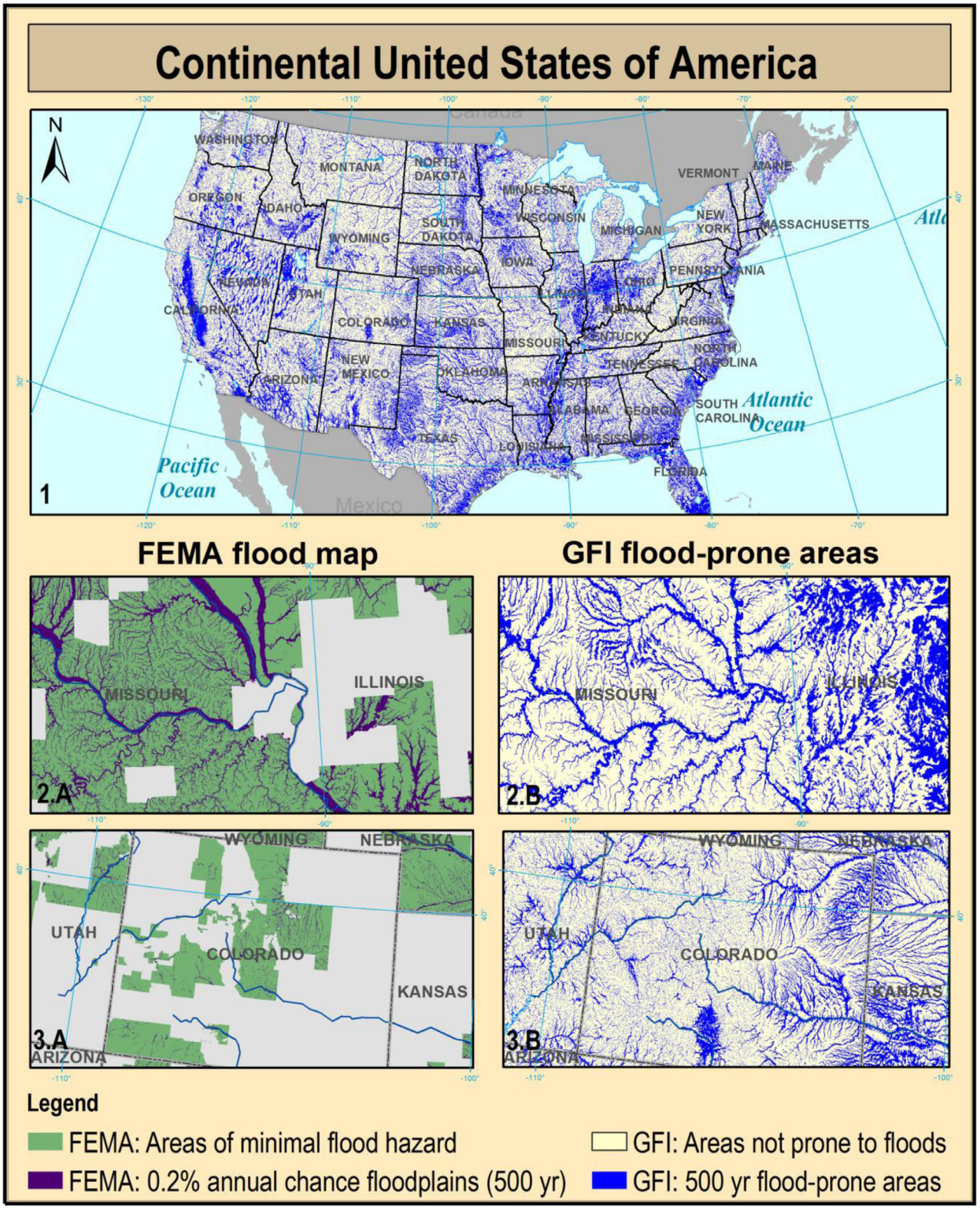


Fig. 11.. Map 1 shows the flood-prone areas identified in the continental U.S. according to the linear binary classifier based on the Geomorphic Flood Index for a return period of 500-year (depicted in dark blue). The following two couples of images provide a more detailed plot of the flood-prone areas identified using this approach in two areas characterized by gaps in the FEMA floodplain map (turned into a binary map). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

area with a single exception, where the classifier D had a better R_{TP}, but not the other statistics.

ii. In flat areas, the performances of the two classifiers (GFI and D) were similar, with D being slightly better than GFI in three of the five basins analysed. Similarly to hilly areas, also in flat areas the validation confirmed the good performances of these classifiers in detecting the flood-prone areas.

iii. The thresholds of both GFI and D are characterized by a limited variability, with coefficients of variation of 8% and 2% respectively. This allows a better transferability to other basins.

The observed performances allow to state that the most reliable classifier is the index named GFI, followed by the flow path distance to the nearest channel, D.

The suitability of these descriptors was supported by verifying their classification ability in areas with topography completely different from the training area. The GFI and D maintained their skill when thresholds calibrated in flat lands were used to delineate the floodplains in hilly areas. Specifically, the GFI regularly outperformed D in the minimization of the total error and the R_{TN}, and D was superior in terms of the correct identification of the areas exposed to floods (R_{TP}), despite being affected by a sharp rise in overestimation error. Conversely, when the thresholds calibrated in hilly areas were used to detect the flood-prone areas in flat lands, we dealt with the opposite situation: even though the performance of D decreased of about 20% in terms of correct identification of the flood-prone areas, it had the minimum values of the total error $R_{FP}+R_{FN}$, while the GFI kept the higher performance in terms of R_{TP}.

From the operative point of view, the application of the methodology requires a preliminary calibration. The challenge is to identify a training area that may provide a good estimate of the threshold parameter. In particular, in searching the minimum area required for the calibration of the classifiers, it was observed that the optimal thresholds for the two mentioned classifiers (GFI and D) show a small variability as long as the training area is larger than 2% of the total area of the basin. In order to test the impact of small training areas on the methodology, the narrowing of the training area was pushed to an extent of 0.01% of the Ohio River basin, returning coefficients of variation of the thresholds of 18% for the GFI, and 9% for D.

Finally, the classifiers have been tested over the continental U.S. with the intent to evaluate their transferability and to quantify the errors associated to their use, considering that their performances are certainly influenced by the local morphology and also by the specific climate of the selected basin. The large-scale validation over the continental U.S., a territory extremely variable in terms of climatic, hydrologic and topographic conditions, decrees that the GFI is consistently the best classifier among those analysed.

The results of this type of procedure usually depend on the quality and resolution of the DEMs used for the terrain analysis. By providing high-resolution DEMs and detailed hydraulic flood hazard maps for the calibration, the accuracy of the modelled floodplain extent may be increased. Nevertheless, the challenge is to tackle the definition of flood-prone areas when limited data is available, and this explains the adoption of readily and freely available DEMs with moderate resolution. In this regard, a mention is due to a previous study carried out over the Bradano River basin (Manfreda et al., 2015). In this application, the performances of the GFI were observed changing the input data in terms of resolution of the adopted DEM (90 m versus 10 m) and reference hydraulic map used for calibration (1-D versus 2-D hydraulic models). Also in this experiment, the GFI seemed to be consistent and largely insensitive to these changes, providing the same optimal threshold for both the input datasets. This induce to think that this index represents a solid and reliable indicator of flood exposure, whose performances are not strongly influenced by changes in terms of dominant topography of the investigated basin, resolution of the DEM adopted, and hydraulic method used to derive the standard flood map adopted.

7. Conclusions

The overall objective of this work was to propose a simplified procedure able to provide reliable flood susceptibility maps in data-scarce regions and for large-scale applications. With this aim, we presented the results of a series of analyses carried out over the Ohio River basin, one of the largest basins in the U.S., and then extended across the continental U.S. Eleven flood-related morphological descriptors were tested, changing the size and topography of the training area to measure the performances of the different classifiers. The observed results seem to prove that the linear binary classification carried out using the Geomorphic Flood Index, GFI, exhibited a higher accuracy than the others in each test, and is the most suitable morphologic classifier for the preliminary mapping of the flood-prone areas in poorly gauged basins and unstudied areas. The second best performing classifier is based on the flow path distance to the nearest channel D, which is very reliable in flat areas. The classification ability of the other descriptors is unreliable for large-scale applications, where can likely occur variations of the topographical characteristics among different portions of the same basin.

The procedure proposed in this work can be a useful tool to have a preliminary, but efficient, delineation with simple data requirements, low costs and reduced computational times. It may be used to fill the gaps that exist in most of the standard flood maps, since in many cases the hydraulic modelling is applied only over portions of the river basins. It may be used also for flood mapping in ungauged basins, exploiting the mentioned reliability of the GFI and the principles described in the present paper. Therefore, this geomorphic method represents a complementary tool for flood hazard mapping that may be useful for river basin authorities, environmental agencies, insurance companies, etc., and can also be helpful for other kinds of applications. For instance, it may be used to identify the computational domain of 2D hydraulic simulations, helping to optimize the computational time. Furthermore, it may be used as ancillary data in flood mapping algorithms based on remote sensed data (D' Addabbo et al., 2016) to monitor flooded areas as a result of extreme weather events.

Acknowledgements

This work was supported by the grant PCBAS1 of Civil Protection of Basilicata.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.advwatres.2017.01.007.

References

Athmania, D., Achour, H., 2014. External validation of the ASTER GDEM2, GMTED2010 and CGIAR-CSI-SRTM v4.1 free access digital elevation models (DEMs) in Tunisia and Algeria. Remote Sens. 6 (5), 4600–4620. http://dx.doi.org/10.3390/rs6054600.

Benke, A.C., Cushing, C.E. (Eds.), 2011, Rivers of North America. Academic Press.

Blöschl, G., Sivapalan, M., 1997. Process controls on regional flood frequency: coefficient of variation and basin scale. Water Resour. Res. 33 (12), 2967–2980. http://dx.doi.org/10.1029/97WR00568.

D'Addabbo, A., Refice, A., Pasquariello, G., Lovergine, F.P., Capolongo, D., Manfreda, S., 2016. A bayesian network for flood detection combining SAR imagery and ancillary data. IEEE Trans. Geosci. Remote Sens. 54 (6), 3612–3625. http://dx.doi.org/10.1109/TGRS.2016.2520487.

Degiorgis, M., Gnecco, G., Gorni, S., Roth, G., Sanguineti, M., Taramasso, A.C., 2012. Classifiers for the detection of flood-prone areas using remote sensed elevation data. J. Hydrol. 470, 302–315. http://dx.doi.org/10.1016/j.jhydrol.2012.09.006.

de Matauco, A.I.G., Ojeda, A.O., Bea, E.D., 2011. Influence of catchment processes on fluvial morphology and river habitats. Limnetica 30 (2), 169–182 c Asociación Ibérica de Limnología, Madrid. Spain. ISSN: 0213-8409.

Demographia World Urban Areas: 12th annual edition: 2016.04 (http://demographia.com/db-worldua.pdf).

De Risi, R., Jalayer, F., De Paola, F., Giugni, M., 2014. Probabilistic delineation of flood-prone areas based on a digital elevation model and the extent of historical flooding: the case of Ouagadougou. Boletín Geológico Minero 125, 329–340.

De Wrachien, D., Mambretti, S., 2011. Mathematical models for flood hazard assessment. Int. J. Saf. Secur. Eng. 1 (4), 353–362. http://dx.doi.org/10.2495/SAFE-V1-N4-353-362.

Dodov, B.A., Foufoula-Georgiou, E., 2006. Floodplain morphometry extraction from a high-resolution digital elevation model: a simple algorithm for regional analysis studies. Geosci. Remote Sens. Lett. IEEE 3 (3), 410–413. http://dx.doi.org/10.1109/LGRS.2006.874161.

Farr, T.G., Rosen, P.A., Caro, E., Crippen, R., Duren, R., Hensley, S., ..., Alsdorf, D., 2007. The shuttle radar topography mission. Rev. Geophys. 45 (2). http://dx.doi.org/10.1029/2005RG000183.

Fawcett, T., 2006. An introduction to ROC analysis. Pattern Recognit. Lett. 27 (8), 861–874. http://dx.doi.org/10.1016/j.patrec.2005.10.010.

Gallant, J.C., Dowling, T.I., 2003. A multiresolution index of valley bottom flatness for mapping depositional areas. Water Resour. Res. 39 (12). http://dx.doi.org/10.1029/2002WR001426.

Giannoni, F., Roth, &, G., Rudari, R., 2005. A procedure for drainage network identification from geomorphology and its application to the prediction of the hydrologic response. Adv. Water Resour. 28 (6), 567–581. http://dx.doi.org/10.1016/j.advwatres.2004.11.013.

Gupta, V.K., Mesa, O.J., 2014. Horton laws for hydraulic–geometric variables and their scaling exponents in self-similar Tokunaga river networks. Nonlinear Processes Geophys. 21 (5), 1007–1025. http://dx.doi.org/10.5194/npg-21-1007-2014.

Herold, C., Mouton, F., 2011. Global flood hazard mapping using statistical peak flow estimates. Hydrol. Earth Syst. Sci. Discuss. 8 (1), 305–363. http://dx.doi.org/10.5194/hessd-8-305-2011.

Hirabayashi, Y., Mahendran, R., Koirala, S., Konoshima, L., Yamazaki, D., Watanabe, S., ..., Kanae, S., 2013. Global flood risk under climate change. Nat. Clim. Change 3 (9), 816–821. http://dx.doi.org/10.1038/nclimate1911.

Hirt, C., Filmer, M.S., Featherstone, W.E., 2010. Comparison and validation of the recent freely available ASTER-GDEM ver1, SRTM ver4.1 and GEODATA DEM-9S ver3 digital elevation models over Australia. Aust. J. Earth Sci. 57 (3), 337–347. http://dx.doi.org/10.1080/08120091003677553.

Hjerdt, K.N., McDonnell, J.J., Seibert, J., Rodhe, A., 2004. A new topographic index to quantify downslope controls on local drainage. Water Resour. Res. 40 (5). http://dx.doi.org/10.1029/2004WR003130.

Horritt, M.S., Bates, P.D., 2002. Evaluation of 1D and 2D numerical models for predicting river flood inundation. J. Hydrol. 268 (1), 87–99. http://dx.doi.org/10.1016/S0022-1694(02)00121-X.

Jalayer, F., De Risi, R., De Paola, F., Giugni, M., Manfredi, G., Gasparini, P., ..., Renner, F., 2014. Probabilistic GIS-based method for delineation of urban flooding risk hotspots. Nat. Hazards 73 (2), 975–1001. http://dx.doi.org/10.1007/s11069-014-1119-2.

Jongman, B., Ward, P.J., Aerts, J.C., 2012. Global exposure to river and coastal flooding: Long term trends and changes. Global Environ. Change 22 (4), 823–835. http://dx.doi.org/10.1016/j.gloenvcha.2012.07.004.

Jonkman, S.N., 2005. Global perspectives on loss of human life caused by floods. Nat. Hazards 34 (2), 151–175. http://dx.doi.org/10.1007/s11069-004-8891-3.

Lehner, B., Döll, P., 2004. Development and validation of a global database of lakes, reservoirs and wetlands. J. Hydrol. 296 (1-4), 1–22. http://dx.doi.org/10.1016/j.jhydrol.2004.03.028.

Lehner, B., Verdin, K., Jarvis, A., 2008. New global hydrography derived from spaceborne elevation data. EOS Trans. AGU 89 (10). http://dx.doi.org/10.1029/2008EO100001.

Lehner, B., Verdin, K., Jarvis, A., 2006. HydroSHEDS Technical Documentation. World Wildlife Fund, Washington, D. C. Available at https://hydrosheds.cr.usgs.gov/.

Leopold, L.B., Wolman, M.G., Miller, J.P., 1964. Fluvial Processes in Geomorphology. WH Freeman, San Francisco 1964.

Manfreda, S., Di Leo, M., Sole, A., 2011. Detection of flood prone areas using digital elevation models. J. Hydrol. Eng. 16 (10), 781–790. http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0000367.

Manfreda, S., Nardi, F., Samela, C., Grimaldi, S., Taramasso, A.C., Roth, G., Sole, A., 2014. Investigation on the use of geomorphic approaches for the delineation of flood prone areas. J. Hydrol. 517, 863–876. http://dx.doi.org/10.1016/j.jhydrol.2014.06.009.

Manfreda, S., Samela, C., Sole, A., Fiorentino, M., 2014b. Flood-prone areas assessment using linear binary classifiers based on morphological indices. In: Vulnerability, Uncertainty, and Risk Quantification, Mitigation, and Management, pp. 2002–2011. ASCE http://dx.doi.org/10.1061/9780784413609.201.

Manfreda, S., Samela, C., Gioia, A., Consoli, G., Iacobellis, V., Giuzio, L., Sole, A., 2015. Flood-prone areas assessment using linear binary classifiers based on flood maps obtained from 1D and 2D hydraulic models. Nat. Hazards 79 (2), 735–754. 2015 http://dx.doi.org/10.1007/s11069-015-1869-5.

Mehlhorn, J., Feyen, L., Banovsky, I., Menzinger, I., 2005. FRAT1.0-an example of applying the geomorphologic regression approach for detailed single location flood risk assessment. In: Geophysical Research Abstracts, 7, p. 07419.

Merz, R., Blöschl, G., 2008. Flood frequency hydrology: 1. Temporal, spatial, and causal expansion of information. Water Resour. Res. 44 (8). http://dx.doi.org/10.1029/2007WR006744.

Nardi, F., Vivoni, E.R., Grimaldi, S., 2006. Investigating a floodplain scaling relation using a hydrogeomorphic delineation method. Water Resour. Res. 42 (9). http://dx.doi.org/10.1029/2005WR004155.

National Flood Hazard Layer (NFHL), Federal Emergency Management Agency (FEMA), 2015 http://fema.maps.arcgis.com/home/item.html?id=cbe088e7c8704464aa0fc34eb99e7f30.

Noman, N.S., Nelson, E.J., Zundel, A.K., 2001. Review of automated floodplain delineation from digital terrain models. J. Water Resour. Plann. Manage. 127 (6), 394–402. http://dx.doi.org/10.1061/(ASCE)0733-9496(2001)127:6(394).

Padi, P.T., Di Baldassarre, G., Castellarin, A., 2011. Floodplain management in Africa: large scale analysis of flood data. Phys. Chem. Earth, Parts A/B/C 36 (7), 292–298. http://dx.doi.org/10.1016/j.pce.2011.02.002.

Pappenberger, F., Dutra, E., Wetterhall, F., Cloke, H.L., 2012. Deriving global flood hazard maps of fluvial floods through a physical model cascade. Hydrol. Earth Syst. Sci. 16 (11), 4143–4156. http://dx.doi.org/10.5194/hess-16-4143-2012.

Reid, R.L., 1991. Always a River: The Ohio River and the American Experience. Indiana Univ Pr..

Reuter, H.I., Nelson, A., Strobl, P., Mehl, W., Jarvis, A., 2009, July. A first assessment of Aster GDEM tiles for absolute accuracy, relative accuracy and terrain parameters, 5, pp. V–240. IEEE http://dx.doi.org/10.1109/IGARSS.2009.5417688.

Rexer, M., Hirt, C., 2014. Comparison of free high resolution digital elevation data sets (ASTER GDEM2, SRTM v2. 1/v4. 1) and validation against accurate heights from the Australian National Gravity Database. Aust. J. Earth Sci. 61 (2), 213–226. http://dx.doi.org/10.1080/08120099.2014.884983.

Samela, C., Manfreda, S., Paola, F.D., Giugni, M., Sole, A., Fiorentino, M., 2015. DEM-based approaches for the delineation of flood-prone areas in an ungauged basin in Africa. J. Hydrol. Eng.. http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0001272.

Seaber, P.R., Kapinos, F.P., Knapp, G.L., 1987. Hydrologic Unit Maps. US Government Printing Office.

Williams, W.A., Jensen, M.E., Winne, J.C., Redmond, R.L., 2000. An automated technique for delineating and characterizing valley-bottom settings. Environ. Monit. Assess. 64 (1), 105–114. http://dx.doi.org/10.1023/A:1006471427421.

Yamazaki, D., Baugh, C.A., Bates, P.D., Kanae, S., Alsdorf, D.E., Oki, T., 2012. Adjustment of a spaceborne DEM for use in floodplain hydrodynamic modeling. J. Hydrol. 436, 81–91. http://dx.doi.org/10.1016/j.jhydrol.2012.02.045.