FOCA: a new quality-controlled database of floods and catchment descriptors in Italy

Pierluigi Claps[1], Giulia Evangelista[1], Daniele Ganora[1], Paola Mazzoglio[1], Irene Monforte[1]

[1] Department of Environment, Land and Infrastructure Engineering, Politecnico di Torino, Torino, 10129, Italy

5 *Correspondence to*: Pierluigi Claps (pierluigi.claps@polito.it)

Abstract. We here present FOCA (Italian FlOod and Catchment Atlas), the first systematic collection of data of Italian river catchments for which discharge historical time series are available. Hydrometric information, including annual maximum peak discharge and average daily annual maximum discharge, is complemented by several geomorphological, climatological, extreme rainfall, land cover and soil-related catchment attributes. All hydrologic information derives from the most recently
10 released datasets of discharge and rainfall measurements. This paper provides a description of all the raw data and the algorithms used to build the basin attributes dataset, to enhance reproducibility and transferability of the analysis. We also describe the approaches adopted to solve problems encountered during the Digital Elevation Model elaboration in areas characterized by complex morphology. Details about the data quality-control procedure developed to detect and correct errors is also reported. One of the main novelties of FOCA with respect to other national-scale datasets is the inclusion of a rich set
15 of geomorphological attributes and of extreme rainfall features for a large set of basins covering a wide elevation and area range. With this first nationwide data collection (available at https://doi.org/10.5281/zenodo.10446258), a wide range of environmental applications, with particular reference to flood studies, can be undertaken on the Italian territory.

1 Introduction

The availability of complete, updated and quality-controlled hydro-geomorphological information over nationwide areas is a
20 key need in a vast range of applications, from hydrological modelling to hydraulic simulations, and rainfall/runoff analyses. Hydro-geomorphological catchment information, called catchments attributes, can also provide a comprehensive description of the landscape and on how the catchment stores and transfers water (Addor et al., 2017).
On a global scale, considerable effort to provide hydrological, climatological and land cover information on a global scale is represented by the work of Linke et al. (2019), who created the HydroATLAS database by mapping key hydrological variables
25 at a 500-meter grid resolution. In this latter work, however, few hydrological and climatic attributes have been considered, and geomorphological catchment descriptors are completely missing. In Europe, the importance of river basin catalogs as a support to water resources management was remarkably stressed by the publication of the Water Framework Directive in 2000. River catchments that cross administrative boundaries can suffer the lack of incomplete and non-uniform information. To tackle this shortcoming, in early 2000 an initiative undertook the creation of a pan-European river and Catchment Database (Vogt et al.,
30 2007). In recent years, in several European countries considerable efforts have been dedicated to produce similar datasets at a national-scale level. Relevant examples are the datasets built in the framework of the CAMELS (Catchment Attributes and MEteorology for Large-sample Studies) initiative, as in the case of Great Britain (Coxon et al., 2020), France (Andréassian et al., 2021; Delaigue et al., 2022), Switzerland (Höge et al., 2023) and Germany (Loritz et al., 2022). Similar datasets have been released also in other continents, developed within the same framework, as the United States (Addor et al., 2017), Chile
35 (Alvarez-Garreton et al., 2018), Brazil (Chagas et al., 2020) and Australia (Fowler et al., 2021). Most of these datasets are now linked together, in order to build a "Caravan" (Kratzert et al., 2023). Previous notable examples are the Atlases and the datasets developed in Switzerland (https://hydrologicalatlas.ch/), Austria (Fürst et al., 2009) and Canada (Arsenault et al., 2016). In recent years, other large-scale datasets were developed independently on the CAMELS framework, such as those related to

China (Hao et al., 2021) and the one related to the areas of the upper Danube up to the Austrian-Slovakian borders and some nearby catchments (Klinglet et al., 2021).

In Italy, as of today, only partial-coverage datasets (both in terms of spatial extent and number of variables) are available. We can mention the datasets developed over North-West of Italy (see e.g. Barbero et al., 2012; Gallo et al., 2013), North-East of Italy (Crespi et al., 2021), North-Central Italy (Pavan et al., 2019) or some studies aimed at mapping few variables conducted over entire Italy (ISPRA, 2005a; ISPRA, 2005b; Claps et al., 2008; Crespi et al., 2018; Braca et al., 2021). Among the above examples, the database realized within the CUBIST (Characterisation of Ungauged Basins by Integrated uSe of hydrological Techniques) project (Claps et al., 2008) was the only attempt to build a multi-variable national data collection, while all the other collections focus only on one specific topic (rainfall, or discharge, or geomorphology). A complete and updated database of hydro-geomorphological variables related to the main catchments of Italy is therefore still missing. One of the reasons behind this gap is the dismantlement of the National Mareographic and Hydrographic Service (Servizio Mareografico e Idrografico Nazionale, SIMN) that has led to a federated management of the national monitoring network by 21 different administrative agencies. Nonetheless, the high network density available for rain and stream gauges throughout the country can allow to compile nation-wide catalogs, also taking advantage of the considerable number of local studies performed using systematic samples of hydrological measurements.

A recent effort to collect relevant hydrological information with a national coverage is represented by the work of Claps et al. (2020a, 2020b, 2020c), in which annual maximum peak discharges (hereafter named peak discharges for brevity) and average daily annual maximum discharges (hereafter named daily discharges for brevity) from 1911 to 2016 for 631 Italian catchments are published, together with some basic geomorphological catchment attributes. This collection is mainly addressed to integrate previously untracked flood discharge measurements, whose historical records were not available in the Hydrological Yearbooks but only in special publications. In this paper we aim to substantially enhance the existing catchment and peak discharge catalogues by complementing the hydrological data with a large set of climatological, geomorphological, soil and land use attributes, many of them computed from recently-released databases. All data sources used comply with the following criteria: (a) nationwide coverage; (b) consistency in data quality (i.e., no regional or local biases); (c) adequacy of the original resolution in relation to the type of information. All the data sources used in this work can be downloaded from public repositories, except for a few variables that are not easily accessible. In these cases, the original data sources were included in the dataset created in this work, together with the mean catchment descriptors, to allow replicability. The result of this work is FOCA (Italian FlOod and Catchment Atlas), the first national-scale catchment attributes dataset in Italy (Claps et al., 2023).

The paper is structured as follows: in Section 2 the database history and the rationale used to select catchments to be included in the database are described; in Section 3 and 4 we present the different categories of raw data, that are partly already quality-controlled, which only required some adjustments and substantially concern digital elevation models and streamflow measures, and partly processed data, deriving from the aggregation and manipulation of rainfall and soil information. In Sections 3 and 4 we also provide an overview of all the geo-morpho-climatic attributes computed for each catchment, together with the methodologies used for their definition and the algorithms used for their evaluation, to grant replicability and thus allowing researchers to perform the same study or extend this one on other catchments. In Section 5 the main characteristics of the FOCA dataset are summarized, and some concluding remarks are presented.

2 Presentation of historical catalogs and rationale for catchment selection

Since early '900, the Italian monitoring network was managed by the SIMN. This institution was in charge of data collection, validation, and publication in a series of Hydrological Yearbooks with yearly updates. For each year, two standardized documents (Part I and Part II) were usually published. The first volume (Part I) was dedicated to temperature and precipitation; the second volume contained information and elaborations related to daily and monthly average discharges.

To retrieve an entire time series measured by a gauging station, the consultation of all the Yearbooks was necessary, due to this separation. All the mentioned Yearbooks are available as images (in gif format) on http://www.bio.isprambiente.it/annalipdf/ (last access: 23 October 2023). To facilitate the consultation of discharge data, this information was then processed by the SIMN at monthly time resolution and grouped by individual gauging station, with the aim of providing several years of data in one data sheet for each station. This summarized information, available up to 1970 and complemented with some basic catchment information, became the "Pubblicazione n°17" (Publication n°17); it also included the annual peak values that were not directly available in the Yearbooks. This document was updated about every 10 years from 1934, and several issues are still available, by progressively including the data acquired up to 1970 (Servizio Idrografico, 1980).

Unfortunately, the update of this publication was interrupted, as about thirty years ago the SIMN was dismantled and 21 different local agencies became responsible of the management of the hydrographic services. Interruption of the publication of the "Pubblicazione n°17" was particularly negative for the peak discharge data. They have indeed been published only in the "Pubblicazione n°17", as the maxima of the instantaneous discharge and have never been available in the Hydrological Yearbooks. This means that peak discharges measured after 1970 were only available, for many years, only in the regional hydrographic offices. The main effort to recover these unpublished data was done by the GNDCI (Gruppo Nazionale per la Difesa dalle Catastrofi Idrogeologiche) in the VAPI (VAlutazione delle PIene) project (http://www.idrologia.polito.it/gndci/Vapi.htm, last access: 23 October 2023) subsequently integrated by the CUBIST project. Within the CUBIST project (Claps et al., 2008) a first, major, advancement in the creation of a national-wide database was achieved. More specifically, historical information available only in printed format were digitized and merged with more recent ones. Moreover, each catchment closed by the gauging stations included in such collection was characterized from a climatological and geomorphological point of view by means of the computation of key catchment attributes.

In the following years, some regional-scale works were conducted to update this systematic collection but without a nation-wide coordination. For example, in North-West Italy, a first follow-up work was the "Catalogo delle portate massime annuali al colmo del bacino occidentale del Po" (Barbero et al., 2012) that includes a considerable number of additional observations from the gauging stations operated by the regional environmental agency as well as additional gauging station managed by other public/private bodies, such as ENEL (a major Italian energy provider) and CNR (Italian Research Council). In this publication, a systematic and extensive work of data validation was performed to convert into discharge values the hydrometric stage values already recorded by the SIMN but never processed with adequate rating curves. This collection summarizes the information of 140 catchments with at least 5 years of data of peak or daily discharges and contains some basic information of the catchments. Unfortunately, this publication does not contain all the attributes evaluated within the CUBIST project. This regional collection was thus improved with the release of the "Atlante dei bacini imbriferi piemontesi" (Gallo et al., 2013), that contains geomorphological attributes of about 200 gauged river catchments located over the same region. In this case, however, the peak discharges were not reported as not available for all the basins. Similar works were also carried out over other limited areas, as required by local studied on flood frequency analysis (e.g., Rossi and Caporali, 2010).

Such collections of flood datasets were recently homogenized, integrated and updated by including the most recent data acquired all over Italy, releasing a comprehensive catalogue of floods ("Catalogo delle Piene dei Corsi d'acqua Italiani"; Claps et al., 2020a, 2020b, 2020c). However, in this case, only basic geomorphological information was included for the considered basins.

The catchment selection criterion adopted in the present work stems from the purpose to improve the work by Claps et al. (2020a, 2020b, 2020c). The 631 chosen catchments are those for which peak or daily discharges are available, and therefore all included in Claps et al. (2020a, 2020b, 2020c). Details on the different sources of the historical discharge time series and their integration are given in Section 4.4.

3 Data and catchment geomorphological attributes

3.1 The Digital Elevation Model

As of today, the only national-scale high-resolution Digital Elevation Model (DEM) available for the whole of Italy is the
125   TINITALY/01 (Tarquini et al., 2007). Despite its very high spatial resolution (10 m) that makes its use quite interesting, especially for the delineation of small catchments, it presents the drawback of being obtained by merging separate DEMs of single administrative regions, thus not allowing to work with the same accuracy level nationwide. To overcome this drawback, we adopted the DEM from the Shuttle Radar Topography Mission (SRTM) at 30 meters spatial resolution (Farr et al., 2007). Processing of DEM and, in general, all the other catchment attributes, has been performed with open-source software, namely
130   GRASS GIS and R. In this work, the original SRTM DEM has been re-projected into the WGS84/UTM zone 32N coordinate system by means of a bicubic interpolation and resampled to obtain integer cells, i.e., the data format requested by the *r.basin* GRASS GIS add-on (Di Leo and Di Stefano, 2013) used to derive most of the geomorphological attributes, as will be discussed in the following sub-sections. DEMs usually contains pits (i.e, elevation values way lower than those of nearby pixels, that are errors due to the resolution of the data) that should be filled to ensure a proper delineation of catchment boundaries and drainage
135   networks. Thus, the pit filling procedure was carried out using the *r.hydrodem* GRASS GIS add-on (Lindsay et al., 2005). The *r.basin* add-on algorithm requires a series of information to be executed. The input parameters required by *r.basin* for the attribute extraction routine are presented in the following sub-sections.

3.2 Catchment boundaries

To determine the basin boundaries, it is first necessary to generate the drainage network from the depitted DEM. This step can
140   be done only once with the r.basin command, which adopts the following steps: (i) calculation of drainage directions with the Multiple Flow Direction (MFD) algorithm; (ii) calculation of the flow accumulation, i.e. the Total Contributing Area (TCA) map; (iii) estimation of stream network, after specifying a threshold value that defines the minimum drainage area required to initiate a channel. In this work threshold values equal to 0.02, 0.1 and 1 km^2 are used to extract the stream network for basin areas smaller than 1 km^2, between 1 and 10 km^2 and larger than 10 km^2, respectively. These values have been identified
145   considering recommendations derived from several research works that investigated the spatial resolution sensitivity of catchment geomorphologic properties and the effect on hydrological simulation, such as Montgomery and Foufoula-Georgiou (1993), Yang et al. (2001) and Beighley and Gummadi (2011). These works highlighted that the river networks generated with larger threshold areas tend to lose relevant information and led us not to use a fixed threshold value, that could be problematic in small basins.

150   To proceed toward delineation, a unified determination of the basin outlet coordinates is necessary, as the official coordinates, taken from hydrometric stations, do not necessarily coincide with stream locations automatically determined from the DEM. So, the coexistence of two sets of coordinates, the real and the operational DEM-based ones, must be properly accounted for in the creation and management of the dataset. The second set of coordinates, the DEM-based ones, are evaluated on the basis of the reprojected SRTM DEM: in case a new DEM will be used in future works, this set of points needs to be evaluated again.
155   Even though the *r.basin* algorithm is able to automatically work with input outlet coordinates not exactly overlapping with the DEM-based river network, through snapping to the closest point belonging to the network, in some cases the relocation may fail and a manual repositioning of the outlets is thus required. Compared to the real coordinates, the required adjustments are in some cases of the order of a few kilometers. This adjustment is needed to obtain a river network that matches with the reference one provided by the Istituto Superiore per la Ricerca e la Protezione Ambientale (ISPRA, available on
160   http://www.sinanet.isprambiente.it/it/sia-ispra/download-mais/reticolo-idrografico/view, last access: 23 October 2023).

As expected, the procedure for the catchment boundary delineation is generally very accurate where the differences in elevation are quite marked (i.e., the alpine areas). Thus, the only manipulation required is the repositioning of the catchment outlet, in

our case always positioned on the gauging stations. However, in flat areas this method does not always provide the real drainage directions and it is therefore necessary to manually force the DEM to correct the Total Contributing Area (TCA) map built by

165   the *r.basin* procedures, a practice commonly known as *stream-burning* (Lindsay, 2016). This operation is performed individually, when needed, by comparing the unconstrained river network produced by *r.basin* with the reference one provided by ISPRA. This quality checking has been iteratively repeated by carving rivers or inserting artificial barriers to force the stream to the correct path. The extent of this operation i.e., location and length of the carved and/or the walled portions, are unique for each area: for this reason, this operation requires an individual assessment. The procedure described above is

170   depicted in the flow chart in Figure 1.

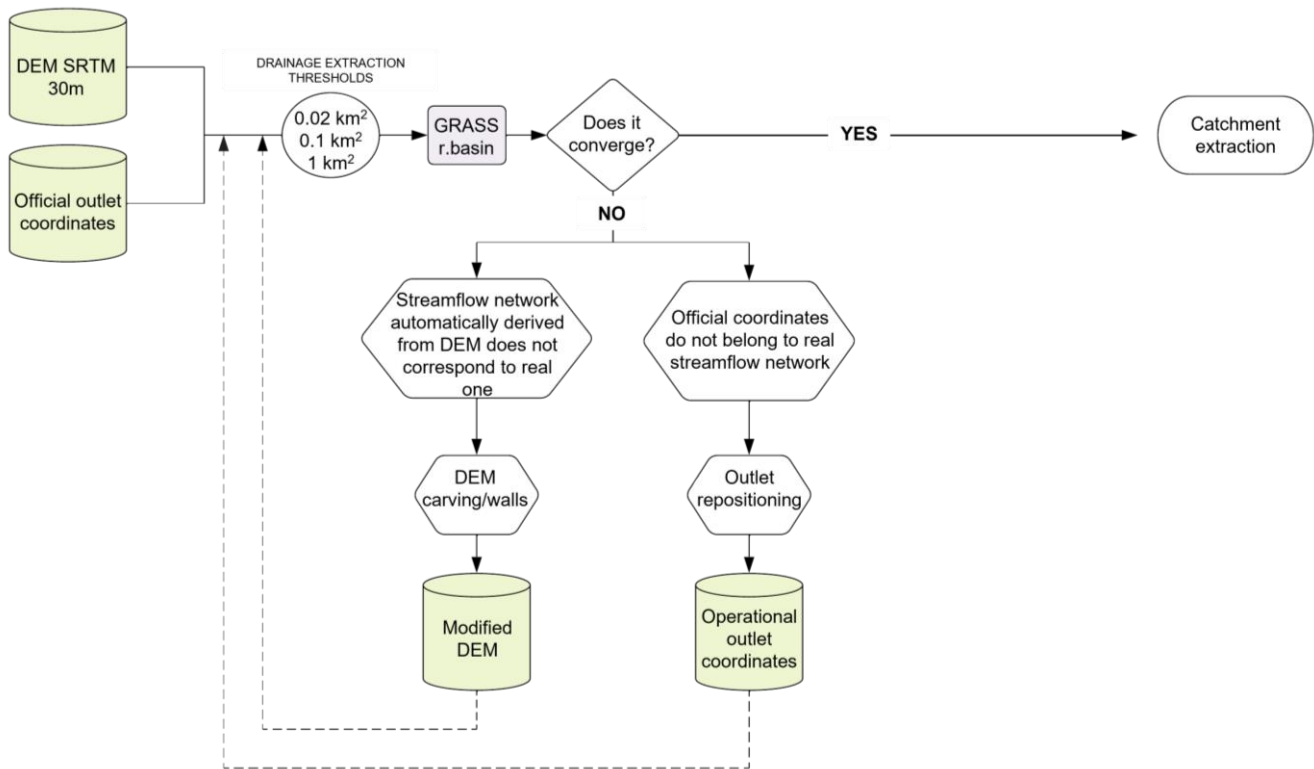



Figure 1. Visual representation of the workflow adopted to delineate catchment boundaries.

175   The catchment boundaries resulting from the delimitation are finally made available in vector format in the WGS84 UTM32 N (EPSG 32632) coordinate system. The catchment boundaries were used as masks to clip several layers of climatological and soil-related attributes. Moreover, the provision of this geographic information will allow users to expand this database by computing other descriptors using possible new gridded datasets.

3.3 Geomorphological catchment attributes

180   Geomorphoclimatic information, combined with soil characteristics, is essential to try to characterize how the catchment stores and transports water, both on and below the surface. In this work, we have first selected from the literature a large set of morphological attributes that can be directly obtained by processing the 30-meter resolution STRM DEM. Automatic procedures adapted around the GRASS add-on *r.basin* is used for this computation, as *r.stat*, *r.slope.aspect, r.stream.stats* (Jasiewicz, 2021) and *r.accumulate* (Cho, 2020) functions, obtaining 61 geomorphological attributes (see Table 1 for a

185   complete description).

Table 1. List of geomorphological attributes, with a brief description and an indication of the algorithm/add-on used for their computation. All the attributes are computed by processing the SRTM DEM at 30 m resolution with the r.basin add-on, that takes advantage of other the GRASS GIS algorithms mentioned at the beginning of Section 3.3.

Attribute sub-category	Attribute	Notation	Units	Description
Altimetric and geometrical	Area	area	km^2	Catchment area computed by multiplying the area of a single pixel by the number of pixels within the catchment boundary.
	Mean Elevation	elev_mean	m a.s.l.	Catchment mean elevation.
	Maximum Elevation	elev_max	m a.s.l.	Catchment maximum elevation.
	Minimum Elevation	elev_min	m a.s.l.	Catchment minimum elevation.
	Aspect	aspect	°	Mean of the angle of exposure on the horizontal plane of each cell of the catchment. The adopted convention is that North is 0° and the aspect is computed clockwise.
	Hypsographic curve	elev_x	m a.s.l.	Elevation values of the hypsographic curve (i.e., the curve that defines the distribution of catchment areas located within a specific elevation range). Each x corresponds to a different percentage of area (2.5, 5, 10, 25, 50, 75, 90, 95 and 97.5%).
	Geographic centroid	x_g y_g	m	Coordinates of the pixel nearest to the centroid of the geometric figure resulting from the projection of the catchment on the horizontal plane.
	Length of the orientation vector	dir_length	km	Length of the segment linking the catchment centroid to the outlet.
	Orientation	orient	°	Angle of the orientation vector with respect to North.
	Mean slope 1	slope1	%	Mean slope value calculated averaging the slope map.
	Mean slope 2	slope2	%	Angle at the base of the right-angled triangle whose base is the square root of the catchment area and twice the median elevation of the catchment (relative to the closing section) as height. This slope is calculated with respect to a catchment of square shape equivalent to the real one and does not consider its actual shape, which can be elongated.
Horton Ratios	Horton-Strahler numbers	HS_num_u HS_length_u HS_area_u HS_slope_u	-	4 sets of $u = 3$ vectors (each corresponding to a Horton order), containing respectively: the number of streams of a given order [-], the average length of the streams of a given order [km], the average contributing area for each order [km^2] and the average slope of the streams of each order [%]. Slopes are calculated as ratio of difference of elevation between ends of the segment to its length.
	Area ratio	R_a	-	Ratio of the average area drained by streams of a given order $u+1$ and streams of order u.
	Bifurcation ratio	R_b	-	Ratio of number of stream branches of a given order u to the number of streams branches of the next order $u+1$.
	Length ratio	R_l	-	Ratio of average length of streams of two adjacent orders u and $u+1$.
	Slope ratio	R_s	-	Ratio of average slope of streams of two adjacent orders u and $u+1$.
Streamflow network	Total stream length	TSL	km	Total length of the river network, obtained by summing the length of all its segments.
	Drainage density	drain_dens	km^{-1}	Ratio between the total stream length and the catchment area.
	Length of main channel	LMC	km	Length of the longest succession of segments that connect a source to the outlet of the catchment.
	Length of longest drainage path	LLDP	km	Path included between the outlet and the furthest point from it, placed on the catchment boundary and identified by following the drainage directions. For most of its length the longest drainage path overlaps the main channel.
	Topological diameter	topo_d	-	Number of confluences found on the main channel.
	Mean hillslope length	MHL	km	Average of the distances (measured following the drainage directions) of all the pixels not belonging to the hydrographic network, starting from the first pixel of the hydrographic network into which they drain.
	Mean slope of longest drainage path	LLDP_slope	%	Calculated as $$LLDP\_slope = \frac{1}{topo\_d}\sum\frac{\Delta z_i}{L_i}\cdot 100 \qquad (1)$$ where topo_d is the topological diameter, L_i is the length of the i-th segment into which $LLDP$ is divided and Δz_i is the corresponding elevation difference
Shape factor and amplitude function	Shape Factor	shape_f	-	Ratio of catchment area to the square of the length of the main channel.
	Elongation Ratio	elong_r	-	Ratio of the diameter of a circle of the same area as the catchment to the maximum drainage path length.
	Circularity Ratio	circ_r	-	Ratio between the catchment area to the area of the circle having the same circumference as the perimeter of the catchment.
	Compactness coefficient	c_c	-	Ratio of catchment perimeter to the diameter of the circle having the same area of the catchment.
	Width function characteristics	width_mean width_var width_skw width_kur	-	Frequency distribution of the distances of each cell of the catchment, along the drainage path, to the outlet. The first four statistical moments (mean, variance, skewness and kurtosis) of this function were calculated as well as the percentiles vector

	width_x	containing the distance to the outlet that includes pixel percentages of x = 5%, 10%, 15%, 30%, 40%, 50%, 60%, 70%, 85% and 95%.

The provision of this set of 61 different attributes for more than 600 catchments nationwide represents one of the strengths of the FOCA dataset. Comparing to other relevant examples that we found in the literature, such as the CAMELS or the LamaH-CE datasets, one realizes that they include only a limited number of geomorphological descriptors, focusing mainly on climatic, hydrologic, land cover, soil and geological indices. As an example, in the CAMELS developed for the U.S. (Addor et al., 2017) or for Brazil (Chagas et al., 2020), only basic topographic characteristics related to outlet coordinates, catchment area, mean elevation and mean catchment slopes are reported. In the LamaH-CE dataset (Klingler et al., 2021) some additional geomorphological attributes were included, such as the median basin elevation, the range of catchment elevation, the elongation ratio, the horizontal distance from the farthest point of the catchment to the corresponding gauge and the drainage density. The inclusion of other descriptors in LamaH-CE was motivated by the need of knowing the shape of the catchments and the stream network influence on runoff formation. We believe that such a small number of descriptors included in previous works can be insufficient to fully characterize areas with complex river networks. Here we stress the need to carefully characterize the catchments also from a geomorphological point of view, especially when working with a high percentage of catchments located in mountainous areas.

To further highlight the importance of this latter point, we can underline that in the Italian context, that is the focus of this work, non-standard geomorphological parameters have already demonstrated to be significant in several applications of statistical hydrology, as for example the regionalization of the flood frequency curves developed by Laio et al. (2011) over NW Italy. In that case, the proposed regression models consider as covariates geomorphological parameters like the area, the mean elevation, the length of the longest drainage path, the length of the orientation vector, and the catchment outlet coordinates. In a revised version of the methodology (Ganora et al., 2014), also the minimum elevation and the shape factor resulted as significant covariates. In Ganora et al. (2023) a non dimensional flood reduction function to estimate design hydrographs in ungauged catchments can be obtained with a multiple linear regression model that includes the longest drainage path length and slope of the catchment, the average catchment elevation and the fourth statistical moment of the width function. Moving to a broader context, the literature suggests that among the factors that govern catchment hydrological response in the process of rainfall-runoff transformation, the catchment geomorphological features have been well represented (see e.g. Nagy et al., 2021; Ravazzani et al., 2019). The synthesis of basin response function from physical basin characteristics becomes crucial in ungauged catchments (Singh et al., 2013). Finally, most of the formulations available in the literature to estimate response time when no runoff data is available typically contain a characteristic length and the slope of the catchment or of the main channel (e.g. Chow, 1962; Kirpich, 1940; Sheridan, 1994), which are not easily available in the other above-mentioned databases.

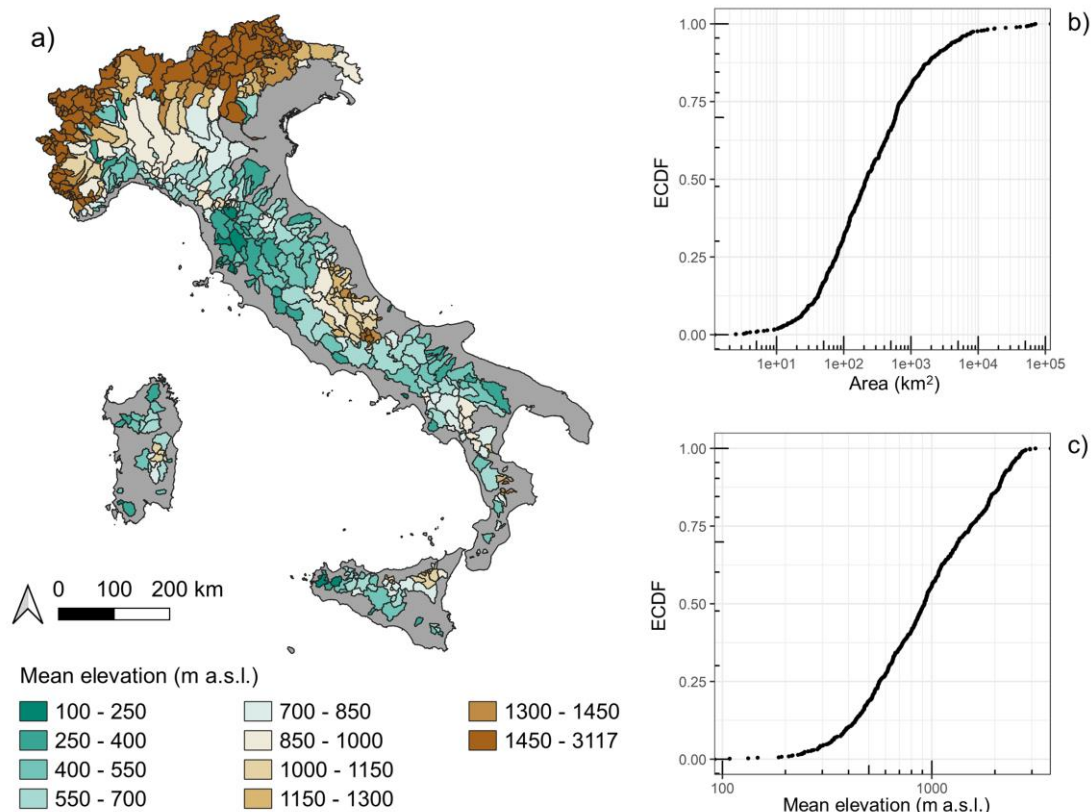
The catchments included in the FOCA dataset cover a wide variety of morphological features and a considerable elevation range (Figure 2a). Figure 2 shows also the empirical cumulative distribution function (ECDF) of some key catchment attributes, such as the area (Figure 2b) and the mean catchment elevation (Figure 2c). Figure 2 also reveals the limited number of catchments with outlets close to the sea. However, there is a reason behind this absence: the 631 chosen catchments that we considered are those for which peak or daily discharges are available, and therefore all included in the "Catalogo delle Piene dei Corsi d'acqua Italiani". This imply that all the "grey" areas are ungauged and could be included in the dataset only if gauging stations will be installed in the near future.

230

Figure 2. Mean elevation of the 631 catchment (a) and empirical cumulative distribution function (ECDF) of some key catchment attributes: area (b), mean elevation (c). Administrative boundaries: GADM v3.6.

3.4 Quality check of geomorphic data

To provide a robust set of catchment features, we have checked the consistency between this new dataset and the previously

235 published one (Claps et al., 2020a, 2020b, 2020c) in terms of catchment areas, which were mostly based on the values published in the Hydrologic Yearbooks and in the "Pubblicazione n°17". In order to set the level of accuracy of the procedures that we used, a 10% maximum deviation of the difference (positive or negative) was considered acceptable. All the catchments for which the discrepancy is greater than 10% have been individually re-examined, as described in Section 3.2. In some cases, the discrepancy was found to derive by a small shift of the catchment outlets upstream or downstream a confluence, thus including

240 or excluding relevant sub-catchments.

Another important check has been made analyzing the difference between the length of main channel (LMC) of each catchment and the length of the longest drainage path (LLDP). Even though a strong relationship between these two attributes is well known, this connection was not always found in our data: for some catchments the two lengths were significantly different, even after a thorough manual check. A 3-km different threshold (corresponding to the value that select the upper 5% of the

245 catchments with marked differences) was used to identify those catchments for which the difference in length between LMC and LLDP should be further investigated. For these catchments, manual inspections highlighted a drawback in the GIS procedure that produced LMC and LLDP measures with unrealistic discrepancies. Two different cases were observed. In one case, the resulting main channel shapefile consisted of a polygonal chain made up by multiple features that needed to be merged in one. In the other case, multiple LLDP that differed from each other no more than one hundred meters (2-3 pixels) were

250 identified for the same catchment. In this last case, one LLDP is manually chosen, and the other ones are removed. We also observed that the two situations could also occur simultaneously.

To reach a general feeling on the consistency of these two attributes some scaling laws of the drainage network can be employed. One of the best-known is the Hack's law equation (Hack, 1957; Eq. (2)), which reads:

$$LLDP = \alpha \, area^\beta \tag{2}$$

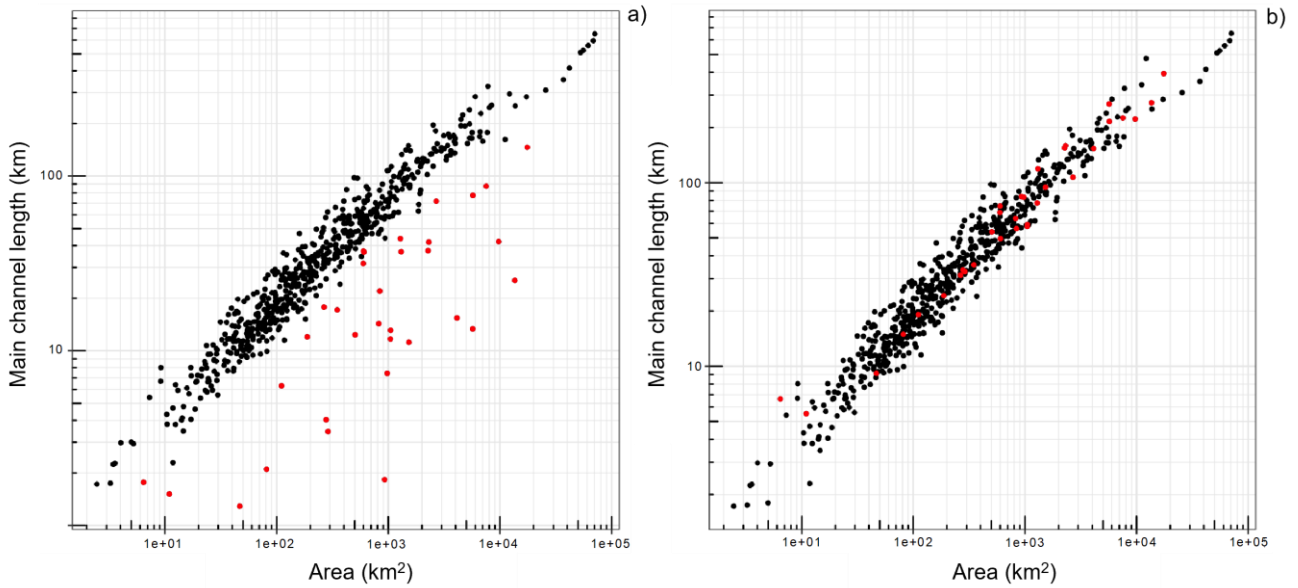
255 where the coefficients α and β vary depending on the study area. A log-log scatter plot between the main channel lengths and the basin areas for all the 631 catchments is displayed in Figure 3, where panels (a) and (b) refer to the results before and after the quality control, respectively. By means of the above-mentioned comparisons, we were able to double-check consistency between the two attributes, recognizing errors up to the order of about 20 km that we checked and corrected.

260

Figure 3. Scaling law between main channel lengths and basin areas for the 631 catchments before (a) and after (b) the quality-control procedure on LLDP and LMC.

4. Data and catchment attributes concerning soil, vegetation and climate

This section describes the work done and the datasets used to evaluate area-averaged catchment attributes concerning soil,
265 vegetation, climate and related extremes (both for rainfall and discharges).

For all the attributes, a spatial average value is provided, with an additional assessment of the spatial variability in terms of the (spatial) coefficient of variation. On the other hand, the temporal variability is not considered here. The only exception is the coefficient of variation of the NDVI and rainfall regimes, that will be introduced later, whose description is reported in Table 4. The spatial coefficient of variation can be thus considered as a proxy of the uncertainty associated to the use of a mean value
270 for the entire catchment. This statistic is an additional information that is not usually reported in the CAMELS datasets but represents an added-value here, especially in a nation with such a complex geomorphology and climate as Italy.

All the attributes that will be hereafter described are affected by different sources of uncertainty that may or may not be quantified. To avoid that resampling of the original data (available at different spatial resolutions) to a standard specific resolution could negatively influence the results, in this work we used input data with an adequate resolution in relation to the
275 type of information, so we avoided the resampling. Moreover, when dealing with raster obtained through spatial interpolation of at-site values, the spatialization method could introduce a certain level of uncertainty at the local scale, that could be quantified in pixels where a measurement is present. For the input data that we used, information about measurement or interpolation uncertainty is not always available since we used several raster maps produced by other research groups. This is the case of land cover, NDVI and climatological data. Information about uncertainty of the soil data that we used are available
280 in Hengl et al. (2017). When referring to datasets produced by the authors involved in the creation of FOCA (discharge and rainfall extremes, see Sections 4.2 and 4.3), some additional information is provided.

4.1 Soil, land cover and NDVI catchment attributes

In this sub-section, different rasterized information is analyzed to provide area-averaged values. The spatial resolution of the raw data range from 100 to 1000 m and data are not resampled at a unique resolution before computing the average values.

Soil descriptors included in the FOCA dataset can provide information connected to an area-averaged estimate of the soil permeability conditions. They are the mean areal values of the Curve Number (Soil Conservation Service, 1972) and of the saturated hydraulic conductivity.

The Curve Number is an empirical parameter used to evaluate the portion of total rainfall that becomes net rainfall in a flood event. The values used in this study were taken from a national scale cartography produced by Carriero (2004) at 250 m resolution, consistently with the work of Ganora et al. (2013). Since this dataset is currently not available online, it was included in the dataset also in the form of a raster map to grant replicability and thus allowing researchers to perform the same study or extend this one on other catchments. Three types of Curve Numbers are defined, which vary according to the antecedent wetness conditions (i.e. dry, average or wet) of the soils. For each Curve Number type, we provided the mean value and the (spatial) coefficient of variation.

To derive a feature that approximates soil permeability features we started from soil texture fractions maps derived from SoilGrids (Hengl et al., 2017; available on: https://soilgrids.org/, last access: 23 October 2023). These cartographies map the spatial distribution of soil properties across the globe at 250 m spatial resolution, at seven standard depths, from 0 cm to 200 cm. SoilGrids maps are based on over 230,000 soil profile observations from the WoSIS (World Soil Information Service) Database (Batjes et al., 2009). Soil texture information were derived from these maps averaged over the first 30 cm of depth, a value consistent with the hydrological purposes of this work. Based on sand and clay content, we derived the saturated hydraulic conductivity using the pedo-transfer function proposed by Saxton et al. (1986; Eq. (3)).

The 7 soil descriptors (6 related to the Curve Number and one to the saturated hydraulic conductivity) that we have considered are listed in Table 2.

Table 2. List of soil attributes.

Soil attributes				
Attribute sub-category	Attribute	Notation	Units	Description
Soil	Curve Number	CN1, CN1_cv CN2, CN2_cv, CN3, CN3_cv	-	Empirical parameter developed by the Soil Conservation Service (1972) and used to predict direct runoff, whose value is between 0 and 100. According to the antecedent moisture condition that refers to the preceding wetness condition of soils, CN is divided into three classes, namely CN1, CN2, and CN3. CN2 is the average condition, while CN1 and CN3 represent the lowest (dry soil) and highest (saturated soil) runoff potentials, respectively.
	Saturated Hydraulic Conductivity	k	cm/d	Computed from sand and clay content of SoilGrids maps at 250 m resolution as follow: $$k_s = 24e^{\left[12.012 - 7.55 \cdot 10^{-2}\, s + \frac{(-3.895 + 3.671 \cdot 10^{-2}\, s - 0.1103c + 8.7546 \cdot 10^{-4}\, c^2)}{0.332 - 7.251 \cdot 10^{-4}\, s + 0.1276 \log(c)}\right]} \qquad (3)$$ where s is the sand content (%) and c is the clay content (%).

Figure 4 provides a snapshot map of the catchment-averaged soil parameter values. The spatial distribution resulting for the saturated hydraulic conductivity (Figure 4a) illustrates significantly different conditions in the soil properties moving from North to South of Italy and reflects the high clay content characterizing the soils of the Apennine basins. The same level of difference is not visible in the CN2 values (Figure 4b), essentially because they reflect the geologic and land-use (not soil) information.

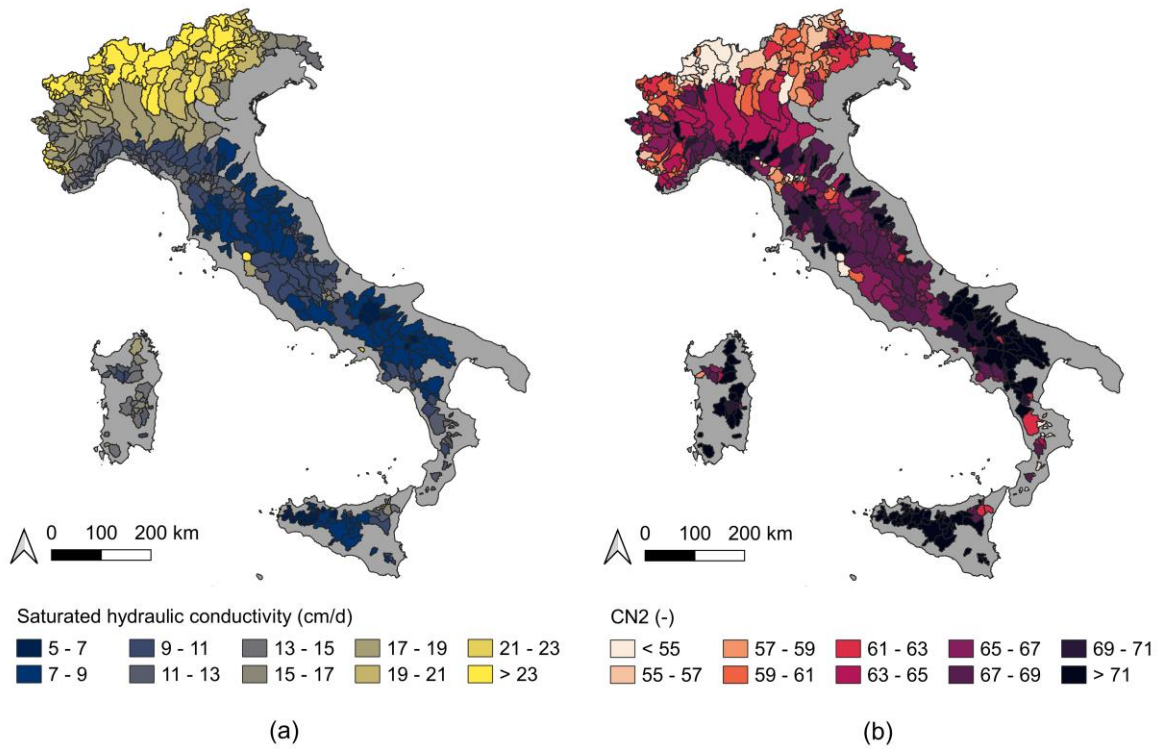



Figure 4. Mean areal values of the saturated hydraulic conductivity (a) and the Curve Number in average conditions, i.e. CN2 (b). Administrative boundaries: GADM v3.6.

Land cover characteristics at 100 m resolution are extracted from the 44 classes of the third level of CORINE Land Cover 2018 (available on: https://land.copernicus.eu/, last access: 23 October 2023). In particular, 5 land indices are considered here, as obtained by merging similar classes. The 5 land indices are outlined in Table 3. A map of one of the classes is provided in Figure 5a: it displays an overview of the percentage of urbanized areas, thus providing some information related to the anthropogenic settlement relevance within the catchments. As expected, the vegetation coverage is lower in high elevation and cold regions, i.e. the Alps, that are also regions with low percentages of urbanized areas.

We also computed multi-temporal indicators of the NDVI (Normalized Difference Vegetation Index) maps, whose data are provided by the Copernicus Land Monitoring Service (available on https://land.copernicus.eu/, last access: 23 October 2023). NDVI is an index that shows if the area under investigation contains live green vegetation and its overall health. We used the values of the Long Term Statistics (LTS) NDVI V3.0.1 of the Copernicus service, with a 1 km spatial resolution, that are NDVI mean observations over the period 1999-2019 for each of the 36 ten-daily periods of the year, resulting in 36 raster maps. These maps were used to compute the mean NDVI value, the (temporal) coefficient of variation of the NDVI and the spatio-temporal mean NDVI regime. The NDVI regime is intended as the diagram representing the multitemporal mean on 36 time intervals of 10 days. To synthetically characterize the latter, a Fourier series representation was used, which allows the shape of the regime to be described with fewer parameters (4, in total) than the 36 10-day average values of which it is made up. A more detailed description of the four parameters that describe the coefficients of the Fourier series that represent the NDVI regimes is reported in Appendix A.

After this data preparation, catchment boundaries were used to extract a total of 11 land cover and NDVI attributes, listed in Table 3. The mean NDVI is mapped in Figure 5b to provide some insights of the variation of the mean greenness of the biomes.

340

Table 3. List of land use and NDVI attributes.

				Land cover and NDVI attributes
Attribute sub-category	**Attribute**	**Notation**	**Units**	**Description**
Land cover	Corine Land Cover 1	clc1	%	Percentage, on the catchment area, of continuous and discontinuous urbanized areas (CORINE classes 111, 112).
	Corine Land Cover 2	clc2	%	Percentage, on the catchment area, of woods (311, 312, 313), arboreal vegetation, shrub vegetation, bushes (CORINE classes 324, 323, 321, 322).
	Corine Land Cover 3	clc3	%	Percentage, on the catchment area, of herbaceous vegetation, meadow pasture, special crops, olive groves, vineyards, arable land (CORINE classes 231, 222, 223, 221, 211,241, 243, 242, 142).
	Corine Land Cover 4	clc4	%	Percentage, on the catchment area, of non-vegetated areas (331, 333, 332, 334), mining areas, landfills, and construction sites (CORINE classes 131, 133), industrial and commercial areas, communication networks (CORINE classes 121, 122, 123, 124).
	Corine Land Cover 5	clc5	%	Percentage, on the catchment area, of humid areas (CORINE classes 411, 512, 521).
NDVI	NDVI	NDVI NDVI_cv	-	Indicator of the greenness of the biomes measured by satellite, whose value is between 0 and 1. It is defined as $$NDVI = \frac{REF_{nir} - REF_{red}}{REF_{nir} + REF_{red}} \qquad (4)$$ where REF_{nir} and REF_{red} are the spectral reflectances measured in the near infrared and red wavebands respectively. Mean value and its (temporal) coefficient of variation were computed.
	B1, B2, C1, C2	B1_NDVI, B2_NDVI, C1_NDVI, C2_NDVI	-	Mean values of the coefficients of the Fourier series representation of NDVI (see Appendix A).

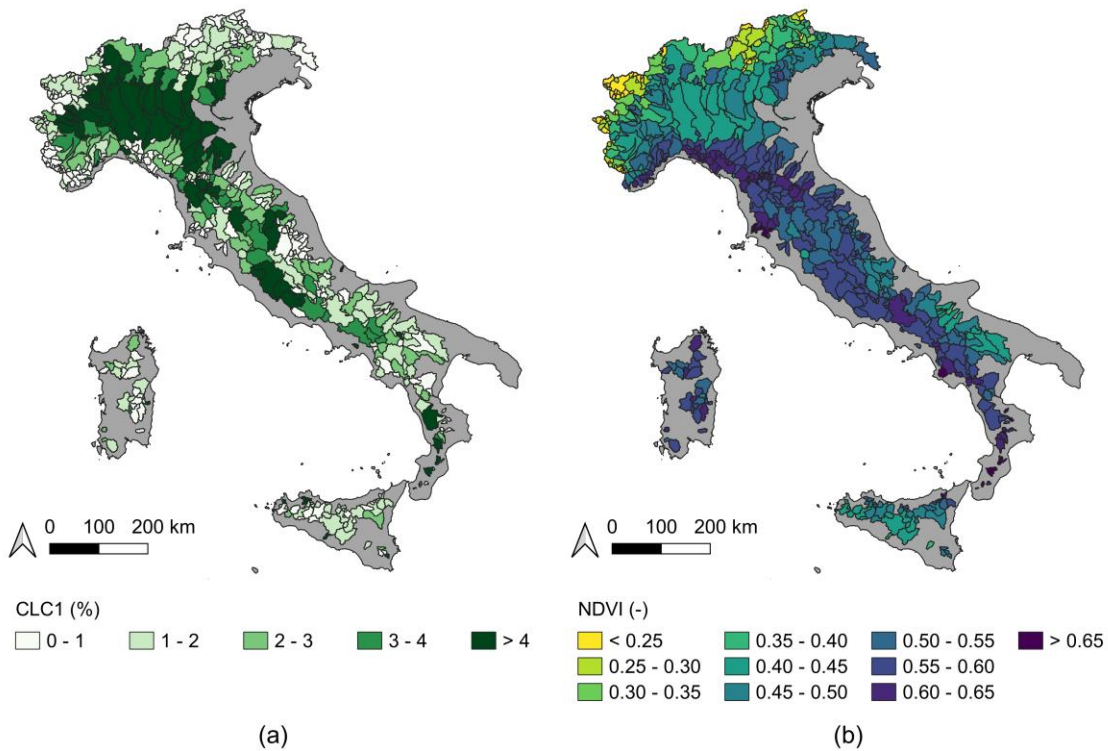



Figure 5. Spatial variability of the percentage of clc1 (a) and mean areal NDVI (b). Administrative boundaries: GADM v3.6.

4.2 Climatological catchment attributes

State-of-the-art national-scale datasets at 1 km resolution have been used for the evaluation of several climatological attributes, described below.

Mean monthly precipitation information is taken from the BIGBANG (Bilancio Idrologico GIS BAsed a scala Nazionale su Griglia regolare; Braca et al., 2021) 4.0 dataset, which covers the 1951-2019 period and is obtained by means of spatial interpolation at 1 km resolution of rain gauge measurements, integrating, only over limited areas and for certain years, the spatial interpolation produced by ARCIS (Archivio Climatologico per l'Italia Centro Settentrionale; Pavan et al., 2019). Mean

monthly temperature data are also derived from this dataset. Both mean monthly precipitation depths and mean monthly temperature data are processed to compute the mean coefficients of the Fourier series that approximate the precipitation and temperature regimes (4 for rainfall, 4 for temperature, see Appendix A).

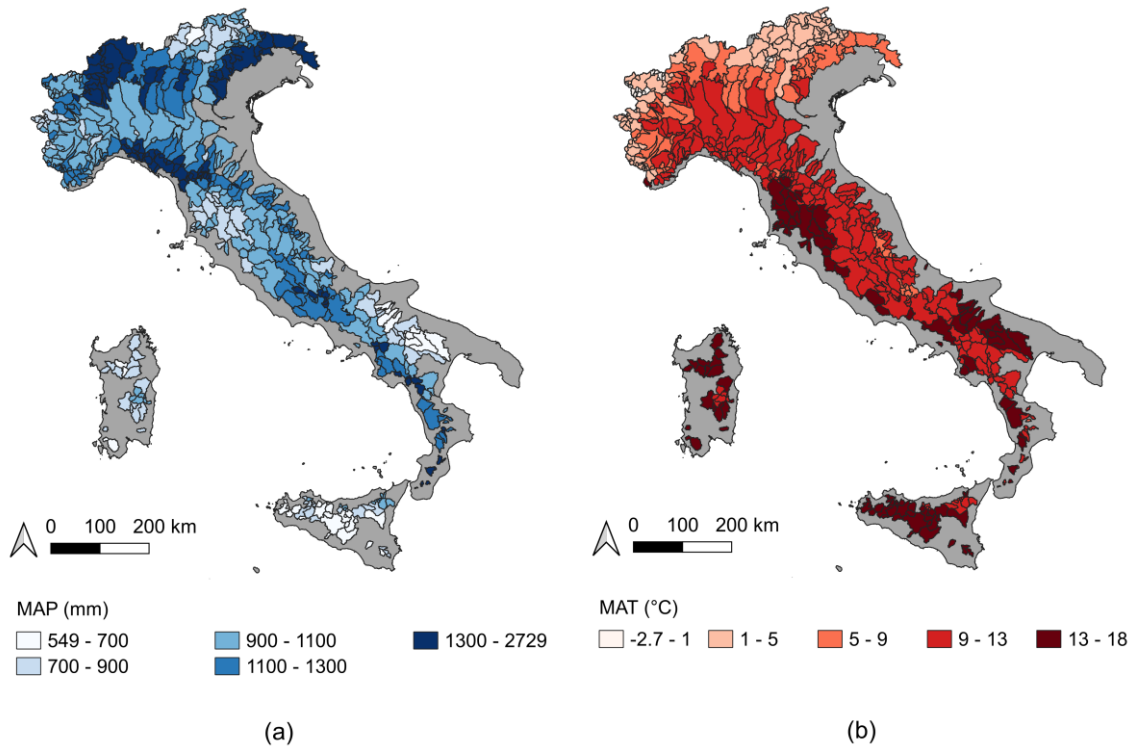
355 This dataset is also used to compute the Mean Annual Precipitation (MAP) and the Mean Annual Temperature (MAT). Catchment boundaries were used to clip the above-mentioned precipitation and temperature maps and to obtain spatial averages for the 14 climatological attributes listed in Table 4.

Table 4. List of climatological attributes.

Climatological attributes			
Attribute	**Notation**	**Units**	**Description**
Mean Annual Precipitation	MAP MAP_cv	mm	Spatial mean and coefficient of variation of the total mean annual precipitation (Braca et al., 2021).
B1, B2, C1, C2	B1_rain, B2_rain, C1_rain, C2_rain	-	Mean values of the coefficients of the Fourier series representative of the precipitation regime computed from the mean monthly precipitation (see Appendix A).
Coefficient of variation of rainfall regimes	cv_rain	-	Temporal coefficient of variation calculated from monthly mean precipitation depths derived from Braca et al. (2021).
Time step between maximum and minimum of mean monthly rainfall	seas_prec	-	Number of months between the occurrence of the absolute annual maximum precipitation and the subsequent absolute minimum precipitation.
Mean Annual Temperature	MAT MAT_cv	°C	Spatial mean and coefficient of variation of the mean annual temperature computed using Braca et al. (2021).
B1, B2, C1, C2	B1_temp, B2_temp, C1_temp, C2_temp	-	Mean values of the coefficients of the Fourier series representation of temperature regimes (see Appendix A).

360

Once representing the results, one can notice in Figure 6a that the mean annual precipitation (MAP) is higher in mountainous regions, with largest values in the Alps. The same areas also show low mean annual temperature (MAT; Figure 6b).

(a)   (b)

365 **Figure 6. Spatial representation of the mean annual precipitation (a) and mean annual temperature (b). Administrative boundaries: GADM v3.6.**

4.3 Extreme rainfall catchment attributes

With respect to other national databases, like the LamaH-CE dataset (Klingler et al., 2021), one of the main novelties of this work is the introduction of information related to parameters of the sub-daily precipitation extremes. This inclusion was made

possible thanks to the availability of a rich collection of in-situ data, characterized by a much greater accuracy in capturing extremes compared to reanalysis data like ERA5 and ERA5-Land datasets (Muñoz Sabater, J., 2019), widely used in the creation of other datasets. Annual maximum rainfall depths used to derive spatial extreme rainfall statistics are obtained from the Improved Italian – Rainfall Extreme Dataset (I^2-RED; Mazzoglio et al., 2020). The dataset consists of official and quality-controlled short-duration (1, 3, 6, 12 and 24 hours) annual maximum rainfall depths recorded by more than 5200 rain gauges over Italy between 1916 and 2019. The stations were subjected to a quality-control procedure to correct errors in the plano-altimetric positions, duplicates and incorrect rainfall measurements (Mazzoglio et al., 2020). This FOCA dataset represents the first national-scale collection of mean extreme rainfall catchment attributes. Thanks to FOCA, it is now possible to perform simple regional and national hydrological studies without the need of retrieving information from 21 different agencies, as mentioned in the Introduction. Due to the complex data policy that regulates the data collected by the hydrological agencies (most of them provides the data free of charge only for research purposes, while people interested in using them for commercial purposes are requested to pay) we were not allowed no include the annual maxima time series in the FOCA dataset. However, nationwide maps of the parameters allowing the computation of the IDF (Intensity – Duration – Frequency) curves have been prepared and made freely available within the FOCA dataset.

Rainfall data and related statistics are processed at 250 m resolution with the *autokrige* R function (Hiemstra and Skoien, 2023), that applies an automatic ordinary kriging using the variogram that better fits the data, automatically generated by the *autofitVariogram* R function. Rainfall statistics obtained with this procedure are:

- the scale factor *a* and the scaling exponent *n* of the average Intensity-Duration (ID) curves, obtained by linear regression of the logarithm of the average of rainfall depth h_d over the 1- to 24-hour durations with the logarithm of the duration, where

$$h_d = a \cdot d^n; \tag{5}$$

- coefficient of L-variation (*L-CV*) of the 1- to 24-hours durations, evaluated with Eq. 6 of Laio et al. (2011);
- coefficient of L-skewness (*L-CA*) of the 1- to 24-hours durations, evaluated with Eq. 7 of Laio et al. (2011).

Time series with at least 10 years of data were used to evaluate the *a* and *n* parameters, while series with 20 and 30 years were used for *L-CV* and *L-CA*, respectively. The different record lengths were selected because higher order statistics can not be evaluated from short time series (Koutsoyiannis, 2019). These maps represent the first attempt to reconstruct and represent updated extreme rainfall statistics over the entire Italy, following what has been released in Switzerland (i.e., the Hydrological Atlas of Switzerland, available on https://hydrologicalatlas.ch/, last access: 23 October 2023), Austria (i.e., the Hydrological Atlas of Austria, Fürst et al. (2009)), Germany (i.e., the KOSTRA-DWD, or "Coordinated heavy precipitation regionalization and evaluation of the DWD", available on https://www.dwd.de/DE/leistungen/kostra_dwd_rasterwerte/kostra_dwd_rasterwerte.html, last access: 23 October 2023), the United States (i.e., NOAA Atlas 14, available on https://www.weather.gov/owp/hdsc_currentpf and on https://hdsc.nws.noaa.gov/hdsc/pfds/, last access: 23 October 2023), the Hawaii (i.e., the Rainfall Atlas of Hawai'i, available on http://rainfall.geography.hawaii.edu/, last access: 23 October 2023) and Canada (CSAGroup, 2019).

Table 5 summarizes the characteristics of the attributes computed and averaged over the 631 catchments.

Table 5. List of rainfall attributes.

Extreme rainfall attributes			
Attribute	**Notation**	**Units**	**Description**
a	a a_cv	mm/h	Scale factor of the IDF curve. Mean value and (spatial) coefficient of variation were computed.
n	n n_cv	-	Scaling exponent of the IDF curve. Mean value and (spatial) coefficient of variation were computed.
L-CV *d*h	LCV_*d*h LCV_*d*h_cv	-	Coefficient of L-variation for *d* = 1, 3, 6, 12 and 24-hour duration. Mean value and (spatial) coefficient of variation were computed.
L-CA *d*h	LCA_*d*h LCA_*d*h_cv	-	Coefficient of L-skewness for *d* = 1, 3, 6, 12 and 24-hours duration. Mean value and (spatial) coefficient of variation were computed.

A discussion on the overall spatial and temporal variability of the rainfall extremes is available in Mazzoglio et al. (2020; 2022a; 2022b; 2023). What is useful to comment on here is the result of the spatial averaging of the parameters *a* and *n* at the

410 catchment scale (Figure 7a-b). The parameter *a* can be combined with *n* to obtain the mean rainfall for different durations (1- to 24-hours). The first one represents the mean 1-hour duration: the higher is *a*, the higher is the intensity of short-duration rainfall extremes over the catchments. This parameter is particularly relevant for small catchments whose characteristic time lag are of the order of 1 hour. Figure 7, compared to Figure 2, shows a modest correlation between *a* and elevation (reverse orographic effect; Avanzi et al., 2015; Mazzoglio et al., 2022a; Mazzoglio et al., 2023). One can also notice that the regions

415 where the most severe short-duration rainfall events occur (Figure 7a) are different from those that present higher MAP (Figure 6a), which is a characteristic of the Mediterranean climate (Mazzoglio et al., 2022a).

(a)   (b)

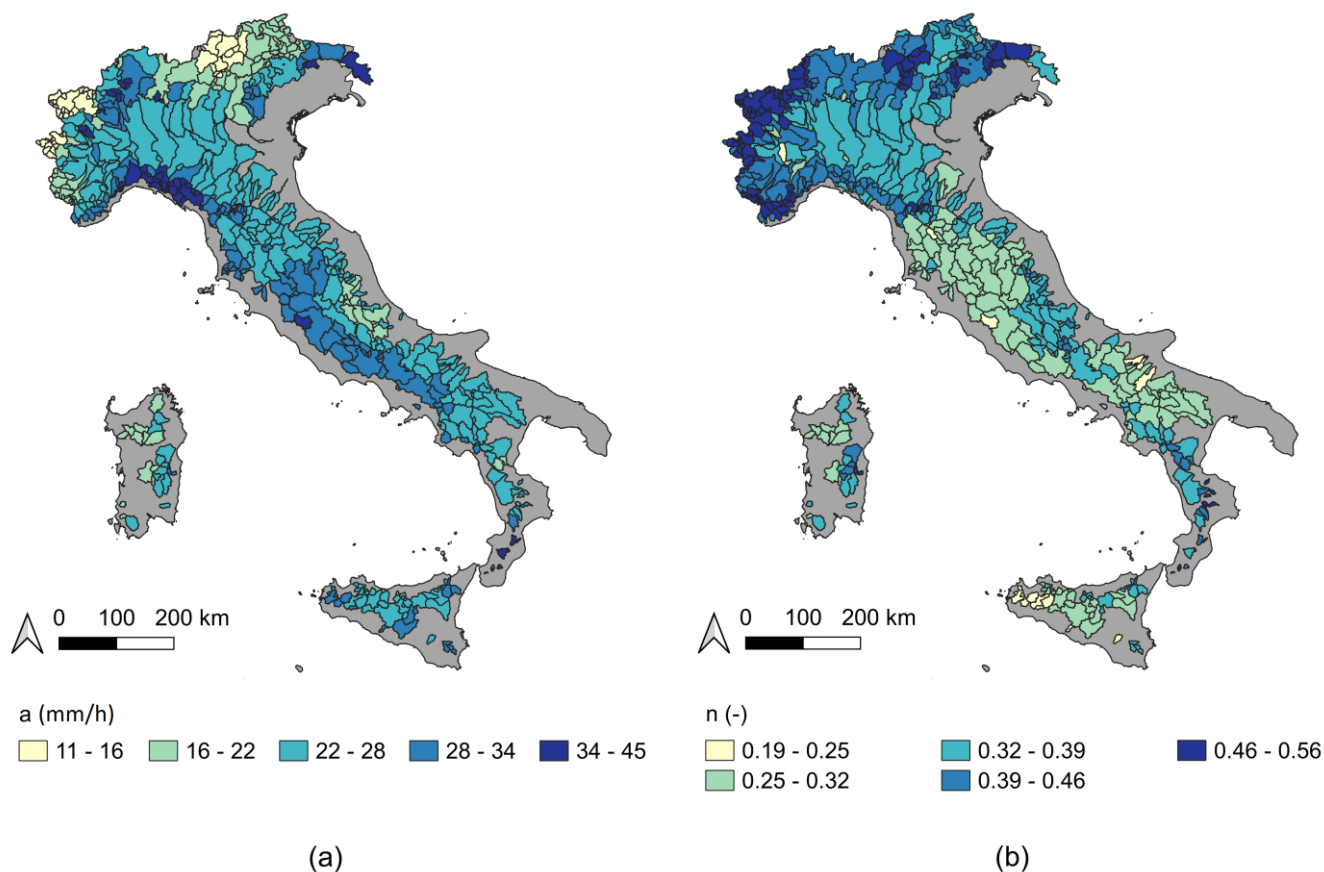



Figure 7. Spatial representation of the spatial average of the parameter *a* (a) and parameter *n* (b) of the ID curves. Administrative
420 **boundaries: GADM v3.6.**

In terms of the assessment of the uncertainty associated to the spatial average values, Figure 8 provides interesting insights, as it shows the standard deviation of the kriging estimates of the *a* parameter in each individual pixel. The figure shows an almost uniform spatial distribution of the interpolation uncertainty, with higher values near the coastline (due to the lower station network density) and lower values where the station density is high. Thus, this map not only serves to show the interpolation
425 uncertainty but also provides information on the rain gauge density throughout the country.

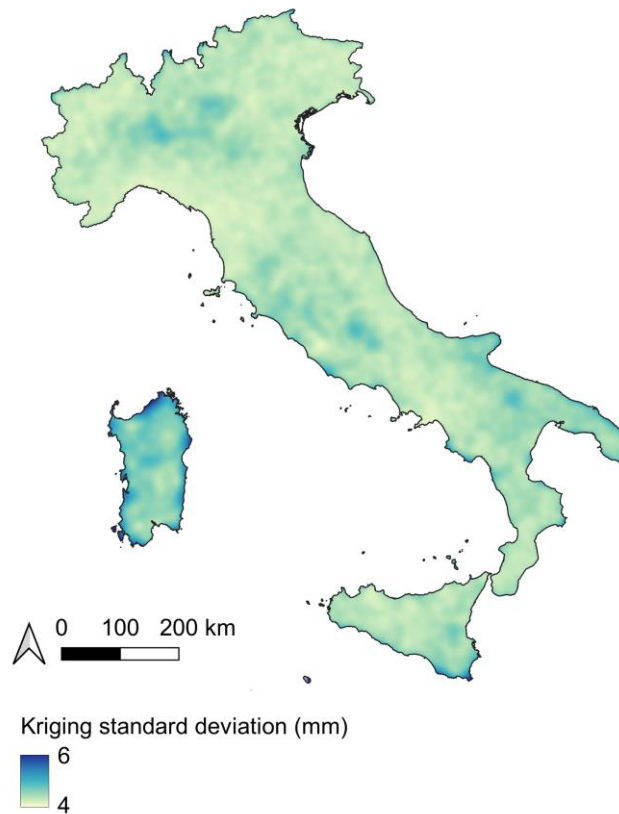



Figure 8. Kriging standard deviation of the *a* parameter. Administrative boundaries: GADM v3.6.

4.4 Peak and daily discharge collections

430   As outlined in Section 2, peak and daily discharges of each gauging station come from Claps et al. (2020a, 2020b, 2020c), which is the most updated systematic flood collection in Italy. Until the 1970s, the data reflect most of the content of the "Pubblicazione n°17" of the SIMN (Servizio Idrografico, 1980). Then, a consistent integration was carried out by merging data from different sources (such as Bencivenga et al., 2011; Barbero et al., 2012; Hall et al., 2015; Brath et al., 2017; Settore Idrologico e Geologico Regionale, 2022; ARPA Lombardia Sistema Informativo Idrologico, 2023), using the same database

435   set up for the CUBIST project (Claps et al., 2008).

   The provision of this information in this paper (in terms of both the time series and of their statistics, like the mean values of the flood peaks, Q_p) offers the opportunity to fully characterize the climatology of the extremes of the catchments upstream of the gauging stations. Uncertainty of observed discharge values cannot be reported, since it is a piece of information that is not available to the public. In essence, before their publication on the Hydrological Yearbooks, these data were validated

440   through application of the corresponding rating curves for every year of measurement. However, it is not possible to rate the uncertainty of the level-to-discharge conversion, which were made decades ago. They would involve very site-specific technical issues that can be dealt with only in a strict cooperation with the offices where every peak discharge value was validated. And this should be done for each and every measurement section.

   The spatial variability of the mean annual flood is shown in Figure 9a, in which one can recognize the presence of nested sub-

445   catchments, resulting from the presence of different measurement stations located along the river path. This situation is particularly evident in the North-West of Italy, where different small mountainous sub-catchments of the Po basin are highlighted, while the Po basin is the partially visible catchment depicted in black. To have an idea of the length of the discharge time series available one can refer to Figure 9b: the different dot colors show that most of the longer time series (> 50 years) are generally available in North and Central Italy. Moreover, while the reference period covered by the dataset is from 1911

450   up to 2016, the time series are fragmented and characterized by different time coverage, and thus do not characterize the entire interval (Figure 10). The decreasing number of time series in recent decades depends on the lack of quality-controlled discharge

data published by the regional hydrological agencies, mostly due to the unavailability of updated rating curves. In most cases, the regional agencies are now publishing only water levels, while information about discharges are often missing.

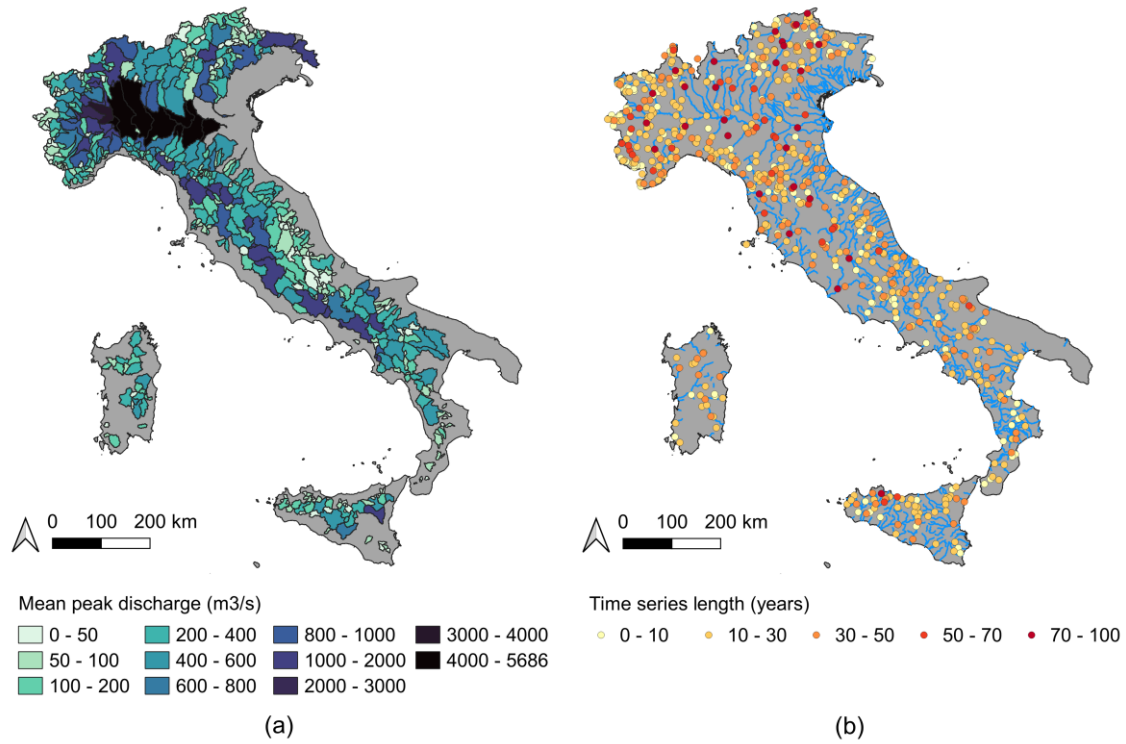
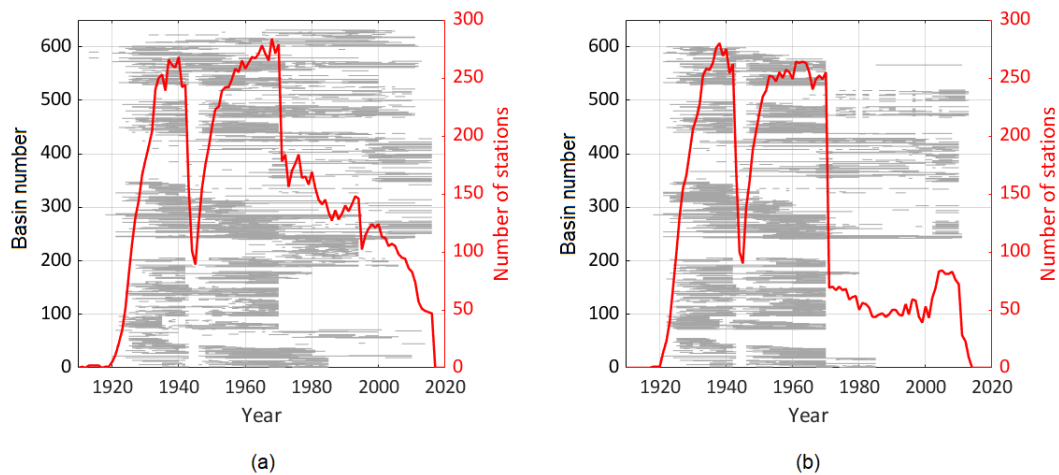



Figure 9. Spatial distribution of the mean peak discharge (a) and discharge time series length (b). Administrative boundaries: GADM v3.6.

 Figure 10. Temporal availability of the peak (a) and daily (b) discharge. Each grey line represents a time series: gray identify available records, white indicates missing data. The red bold line identifies the total number of yearly measurements available in each year.

The catchments cover a wide variety of morphological features. Figure 11 shows the empirical cumulative distribution function (ECDF) of the specific mean annual flood (i.e., the ratio between the mean peak discharge Q_p and the catchment area; Figure 11b). The relation between the area, the specific mean annual flood and the mean catchment elevation is represented in Figure 11, which shows how the Italian catchments are characterized by different hydrological regimes, and a modest dependence of the mean elevation on the specific peak discharge. Catchments with high mean elevation (2000 up to more than 3000 m a.s.l.)

cover a wide range of possible specific mean annual floods, but the interval is smaller than those that characterizes catchments
470 located at lower elevation.

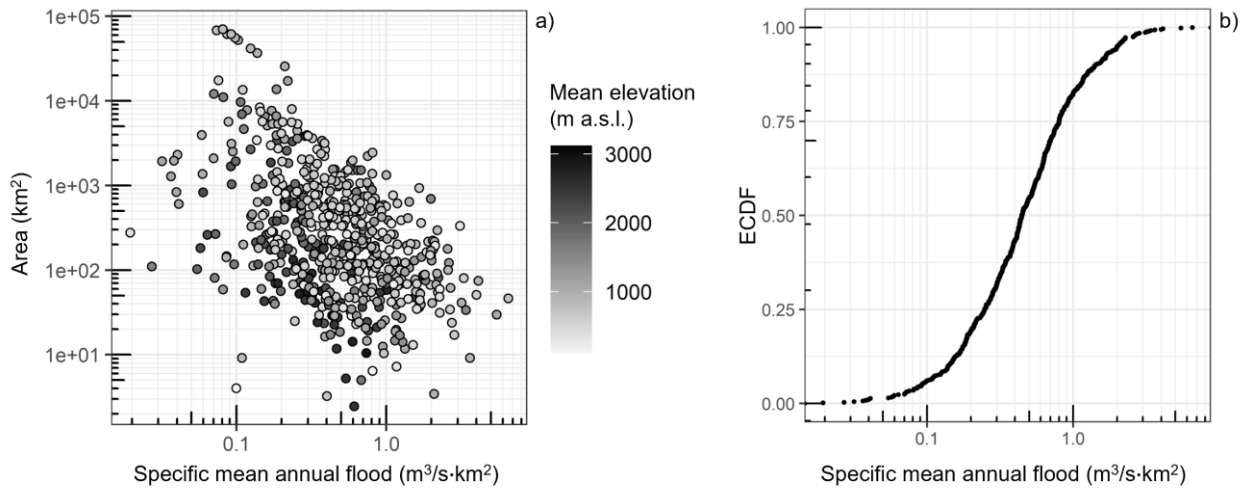



Figure 11. Relation among area, specific peak discharge and mean elevation of the catchments (a) and empirical cumulative distribution function (ECDF) of the specific mean annual flood (b).

5 Conclusions

In this work we present FOCA (Italian FlOod and Catchment Atlas), a collection of attributes of 631 Italian river catchment derived from data sources that comply with the following criteria: (a) nationwide coverages; (b) consistency in data quality (i.e., no regional or local biases); (c) adequacy of the original resolution, in relation to the type of information.

FOCA represents the most updated collection of catchment boundaries and related attributes at the Italian national scale,
480 summarized in features related to climate (discharge, rainfall and temperature), geomorphology, soil and land use. The dataset covers an overall area representative of the most landscapes of Italy, including the mountainous ones (elevation ranges from 0 to 4800 m). One of the main novelties with respect to other national-scale datasets is the inclusion of a rich set of geomorphological and extreme rainfall attributes. The second notable novelty is the inclusion, in the same dataset, of the most updated information regarding extreme rainfall and discharge.

485 As mentioned before, Italy has a very wide variability of climate, land and morphological features and this peculiarity emerges from this new dataset. Figure 2b shows that the FOCA dataset includes catchments with significantly different areas, with prevailing small-to-medium ones: 105 out of 631 catchments have an area of less than 50 km^2 and 300 out of 631 catchments have an area of less than 200 km^2. In addition, 279 out of 631 catchments have a mean elevation higher than 1,000 m a.s.l., highlighting that, despite the existence of some large plains, Italy is a mountainous country. On the other hand, on the large
490 catchments in the North of Italy, the spatial averaging operation over the catchment area becomes less significant, being evaluated over an area within which significantly different features coexist.

A key decision that we had to take while developing FOCA was whether to use local or global/quasi-global datasets, or both. The use of global datasets facilitates the comparison of the results obtained in large-scale hydrology works, even when continental analyses are performed. However, local datasets are, without doubts, characterized by higher-quality information.
495 In this work we decided to opt for the second approach, selecting for each variable the best possible data, prioritizing local information, and resorting to global data only in few cases. The availability of the digital information of the catchment boundaries allows the users to evaluate mean areal values of other variables, as the evapotranspiration and the snow cover, using their own models or other large-scale reanalysis datasets, like ERA5 and ERA5-Land.

The FOCA dataset provides new opportunities to perform both regional and national-scale studies using catchments shapes
500 and attributes extracted using a common framework, subjected to the same quality-control procedure. Information about the

algorithms used in this work are also reported in Section 3 to ensure replicability and the calculation of the attribute characteristics in any ungauged basin. Moreover, the database provides the opportunity to investigate how catchment attributes control river flows and allows to improve data-intensive investigations as the prediction in ungauged catchments.

Data availability

505   The dataset detailed in this paper is available at https://doi.org/10.5281/zenodo.10446258 (Claps et al., 2023). It contains all the catchment boundaries and related catchment attributes described before. To access the latest version of the database, in case of future updates, the readers can refer to (https://doi.org/10.5281/zenodo.8060736) to download the most recent version.

Competing interests

The contact author has declared that none of the authors has any competing interests.

510   **Acknowledgements**

This study was carried out within the RETURN Extended Partnership and received funding from the European Union Next-GenerationEU (National Recovery and Resilience Plan – NRRP, Mission 4, Component 2, Investment 1.3 – D.D. 1243 2/8/2022, PE0000005) – SPOKE TS2. This study has been partially conducted within the framework of the technical-scientific collaboration agreement signed between the Southern Apennine District Basin Authority and the Consorzio Interuniversitario

515   per l'Idrologia (CINID).

Author contribution

PC: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – review & editing. GE: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft preparation, Writing – review & editing. DG: Conceptualization, Methodology, Writing

520   – review & editing. PM: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft preparation, Writing – review & editing. IM: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft preparation, Writing – review & editing.

Appendix A

525   The rainfall/temperature/NDVI regimes are described using Fourier series, that allows to reconstruct the shape of the regime using a reduced number of parameters, compared to the 12 different monthly values (or 36 10-day values in the case of the NDVI) used to describe it. According to Fourier's theorem, a wave of period τ and pulse $\omega = 2\pi/\tau$ can be described as

$$f(t) = A_0 + \sum_{i=1}^{N}(A_i \cos(i\omega t + \phi_i)) \tag{6}$$

where t is the time, A_0 is the mean of $f(t)$ over t, N is the number of harmonics, A_i is the harmonic amplitude and ϕ_i is the

530   phase. In the case of two harmonics, Eq. 6 can be written as

$$f(t) = A_0 + A_1 \cos\left(\frac{2\pi}{\tau} + \phi_1\right) + A_2 \cos\left(\frac{4\pi}{\tau} + \phi_2\right) = A_0 + A_1 \cos\left(\frac{2\pi}{\tau}t\right) \cdot \cos(\phi_1) - A_1 \sin\left(\frac{2\pi}{\tau}t\right) \cdot \sin(\phi_1) +$$

$$A_2 \cos\left(\frac{4\pi}{\tau}t\right) \cdot \cos(\phi_2) - A_2 \sin\left(\frac{4\pi}{\tau}t\right) \cdot \sin(\phi_2) \tag{7}$$

By separating the variables that depend on the time from those that are not time-dependent we obtain

$$f(t) = A_0 + B_1 \cdot X_1(t) + C_1 \cdot Y_1(t) + B_2 \cdot X_2(t) + C_2 \cdot Y_2(t) \qquad (8)$$

535   In this analysis we evaluated *B1*, *B2*, *C1*, *C2* with the ordinary least squared method (see Eq. 9 to 12).

$$B_1 = A_1 \cos(\phi_1) \qquad (9)$$

$$B_2 = A_2 \cos(\phi_2) \qquad (10)$$

$$C_1 = -A_1 \sin(\phi_1) \qquad (11)$$

$$C_2 = -A_1 \sin(\phi_2) \qquad (12)$$

540   The equation can be further simplified by assuming as null the mean A_0 and solving it for $\tau = 2\pi$.

References

Addor, N., Newman, A. J., Mizukami, N., and Clark, M. P.: The CAMELS data set: catchment attributes and meteorology for large-sample studies, Hydrol. Earth Syst. Sci., 21, 5293–5313, https://doi.org/10.5194/hess-21-5293-2017, 2017.

Alvarez-Garreton, C., Mendoza, P. A., Boisier, J. P., Addor, N., Galleguillos, M., Zambrano-Bigiarini, M., Lara, A., Puelma, 545 C., Cortes, G., Garreaud, R., McPhee, J., and Ayala, A.: The CAMELS-CL dataset: catchment attributes and meteorology for large sample studies – Chile dataset, Hydrol. Earth Syst. Sci., 22, 5817–5846, https://doi.org/10.5194/hess-22-5817-2018, 2018.

Andréassian, V., Delaigue, O., Perrin, C., Janet, B., and Addor, N.: CAMELS-FR: A large sample, hydroclimatic dataset for France, to support model testing and evaluation, EGU General Assembly 2021, online, 19–30 Apr 2021, EGU21-13349, 550 https://doi.org/10.5194/egusphere-egu21-13349, 2021.

ARPA Lombardia Sistema Informativo Idrologico: https://idro.arpalombardia.it/it/map/sidro/, last access: 29 May 2023.

Arsenault, R., Bazile, R., Ouellet Dallaire, C., and Brissette, F.: CANOPEX: A Canadian hydrometeorological watershed database, Hydrol. Process., 30, 2734– 2736, https://doi.org/10.1002/hyp.10880, 2016.

Avanzi, F., De Michele, C., Gabriele, S., Ghezzi, A., and Rosso, R.: Orographic signature on extreme precipitation of short 555 durations, J. Hydrometeorol., 16(1), 278-294, https://doi.org/10.1175/JHM-D-14-0063.1, 2015.

Barbero, S., Claps, P., Graziadei, M., Zaccagnino, M., Saladin, A., Ganora, D., Laio, F., and Radice, R.: Catalogo delle massime portate annuali al colmo del bacino occidentale del Po, Arpa Piemonte, Torino, Italy, 301 pp., ISBN 978-88-7479-112-5, 2012.

Batjes, N. H.: Harmonized soil profile data for applications at global and continental scales: Updates to the 857 WISE database, 560 Soil Use and Management, 25, 124–127, https://doi.org/10.1111/j.1475-2743.2009.00202.x, 2009.

Beighley, R.E., and Gummadi, V.: Developing channel and floodplain dimensions with limited data: a case study in the Amazon Basin, Earth Surf. Process. Landforms, 36, 1059-1071, https://doi.org/10.1002/esp.2132, 2011.

Bencivenga, M., Calenda, G., and Mancini, C.P.: Ricostruzione storica delle scale di deflusso delle principali stazioni di misura del bacino del fiume Tevere: il secolo XX, Istituto Poligrafico e Zecca dello Stato, Roma, Italy, 355 pp., 2011.

565 Braca, G., Bussettini, M., Lastoria, B., Mariani, S., and Piva, F.: Elaborazioni modello BIGBANG versione 4.0, Istituto Superiore per la Protezione e la Ricerca Ambientale – ISPRA. http://groupware.sinanet.isprambiente.it/bigbang-data/library/bigbang40, 2021.

Brath A., Castellarin A., Domeneghetti A., and Persiano, S.: Rapporto sintetico riguardante l'esame critico degli studi idrologici sui bacini delle alpi orientali e dell'appennino settentrionale, DICAM - Università degli studi di Bologna, Bologna, 570 Italy, 2017.

Chagas, V. B. P., Chaffe, P. L. B., Addor, N., Fan, F. M., Fleischmann, A. S., Paiva, R. C. D., and Siqueira, V. A.: CAMELS-BR: hydrometeorological time series and landscape attributes for 897 catchments in Brazil, Earth Syst. Sci. Data, 12, 2075–2096, https://doi.org/10.5194/essd-12-2075-2020, 2020.

Carriero, D.: Analisi della distribuzione delle caratteristiche idrologiche dei suoli per applicazioni di modelli di simulazione

575   afflussi-deflussi, PhD thesis, Università degli studi della Basilicata, 2004.

Cho, H.: A recursive algorithm for calculating the longest flow path and its iterative implementation, Environ. Model Software, 131, 104774, https://doi.org/10.1016/j.envsoft.2020.104774, 2020.

Chow, V.T.: Hydrologic determination of waterway areas for the design of drainage structures in small drainage basins, University of Illinois Engineering Experiment Station Bulletin No. 462, Urbana, IL, 104 pp.,

580   https://core.ac.uk/download/pdf/4814575.pdf, 1962.

Claps, P., Barberis, C., De Agostino, M., Gallo, E., Laguardia, G., Laio, F., Miotto, F., Plebani, F., Vezzù, G., Viglione, A., M. Zanetta, M.: Development of an Information System of the Italian basins for the CUBIST project, Geophysical Research Abstracts, 10, 2008.

Claps, P., Evangelista, G., Ganora, D., Mazzoglio, P., and Monforte, I: FOCA (Italian FlOod and Catchment Atlas), Zenodo

585   [data set], https://doi.org/10.5281/zenodo.10446258, 2023.

Claps, P., Ganora, D., Apostolo, A., Brignolo, I., and Monforte, I.: Catalogo delle Piene dei Corsi d'acqua Italiani Vol. 1, Ed. CINID., 499 pp., ISBN 978-88-945568-0-3, 2020.

Claps, P., Ganora, D., Apostolo, A., Brignolo, I., and Monforte, I.: Catalogo delle Piene dei Corsi d'acqua Italiani Vol. 2, Ed. CINID, 537 pp., ISBN 978-88-945568-2-7, 2020.

590   Claps, P., Ganora, D., Apostolo, A., Brignolo, I., and Monforte, I.: Catalogo delle Piene dei Corsi d'acqua Italiani Vol. 3, Ed. CINID, 401 pp., ISBN 978-88-945568-4-1, 2020.

Coxon, G., Addor, N., Bloomfield, J. P., Freer, J., Fry, M., Hannaford, J., Howden, N. J. K., Lane, R., Lewis, M., Robinson, E. L., Wagener, T., and Woods, R.: CAMELS-GB: hydrometeorological time series and landscape attributes for 671 catchments in Great Britain, Earth Syst. Sci. Data, 12, 2459–2483, https://doi.org/10.5194/essd-12-2459-2020, 2020.

595   Crespi, A., Brunetti, M., Lentini, G., and Maugeri, M.: 1961–1990 high-resolution monthly precipitation climatologies for Italy, Int. J. Climatol., 3, 878–895, https://doi.org/10.1002/joc.5217, 2018.

Crespi, A., Matiu, M., Bertoldi, G., Petitta, M., and Zebisch, M.: A high-resolution gridded dataset of daily temperature and precipitation records (1980–2018) for Trentino-South Tyrol (north-eastern Italian Alps), Earth Syst. Sci. Data, 13, 2801–2818, https://doi.org/10.5194/essd-13-2801-2021, 2021.

600   CSAGroup: Technical guide. Development, interpretation, and use of rainfall intensity-duration-frequency (IDF) information: guideline for Canadian water resources practitioners, CSA PLUS 4013:19, 2019.

Delaigue, O., Brigode, P., Andréassian, V., Perrin, C., Etchevers, P., Soubeyroux, J.-M., Janet, B., and Addor, N.: CAMELS-FR: A large sample hydroclimatic dataset for France to explore hydrological diversity and support model benchmarking, IAHS-AISH Scientific Assembly 2022, Montpellier, France, 29 May–3 Jun 2022, IAHS2022-521,

605   https://doi.org/10.5194/iahs2022-521, 2022.

Di Leo, M., and Di Stefano, M.: An open-source approach for catchment's physiographic characterization, Abstract H52E-06 presented at 2013 Fall Meeting, AGU, San Francisco, CA, USA, 2013.

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., Seal, D., Shaffer, S., Shimada, J., Umland, J., Werner, M., Oskin, M., Burbank, D., and Alsdorf, D.: The Shuttle Radar Topography

610   Mission, Rev. Geophys., 45, RG2004, https://doi.org/10.1029/2005RG000183, 2007.

Fowler, K. J. A., Acharya, S. C., Addor, N., Chou, C., and Peel, M. C.: CAMELS-AUS: hydrometeorological time series and landscape attributes for 222 catchments in Australia, Earth Syst. Sci. Data, 13, 3847–3867, https://doi.org/10.5194/essd-13-3847-2021, 2021.

Fürst, J., Godina, R., Peter Nachtnebel, H., and Nobilis, F.: The hydrological atlas of Austria - comprehensive transfer of

615   hydrological knowledge and data to engineers, water resources managers and the public, IAHS-AISH Publication 327, 36-44, 2009.

Gallo, E., Ganora, D., Laio, F., and Claps, P.: Atlante dei Bacini Imbriferi Piemontesi, Final report of the RENERFOR-ALCOTRA project, ISBN 978-88-96046-06-7, 2013.

Ganora, D., Laio, F., Claps, P.: Valutazione probabilistica delle piene in Piemonte e Valle d'Aosta: Metodologia Regionale
620 Spatially Smooth, http://www.idrologia.polito.it/web2/open-data/Dati_Reports_Piemonte/Piene_Prog_Flora_2014.pdf, 2014.

Ganora, D., Evangelista, G., Cordero, S., and Claps. P.: Design flood hydrographs: a regional analysis based on flood reduction functions, Hydrol. Sci. J., 68(2), 325-340, https://doi.org/10.1080/02626667.2022.2153051, 2023.

Hall, J., Arheimer, B., Aronica, G. T., Bilibashi, A., Boháč, M., Bonacci, O., Borga, M., Burlando, P., Castellarin, A., Chirico, G. B., Claps, P., Fiala, K., Gaál, L., Gorbachova, L., Gül, A., Hannaford, J., Kiss, A., Kjeldsen, T., Kohnová, S., Koskela, J.
625 J., Macdonald, N., Mavrova-Guirguinova, M., Ledvinka, O., Mediero, L., Merz, B., Merz, R., Molnar, P., Montanari, A., Osuch, M., Parajka, J., Perdigão, R. A. P., Radevski, I., Renard, B., Rogger, M., Salinas, J. L., Sauquet, E., Šraj, M., Szolgay, J., Viglione, A., Volpi, E., Wilson, D., Zaimi, K., and Blöschl, G.: A European Flood Database: facilitating comprehensive flood research beyond administrative boundaries, Proc. IAHS, 370, 89–95, https://doi.org/10.5194/piahs-370-89-2015, 2015.

Hack, J.: Studies of longitudinal stream profiles in Virginia and Maryland, US Geological Survey Professional Paper, 294-B,
630 1, 1957.

Hao, Z., Jin, J., Xia, R., Tian, S., Yang, W., Liu, Q., Zhu, M., Ma, T., Jing, C., and Zhang, Y.: CCAM: China Catchment Attributes and Meteorology dataset, Earth Syst. Sci. Data, 13, 5591–5616, https://doi.org/10.5194/essd-13-5591-2021, 2021.

Hengl, T., Mendes de Jesus, J., Heuvelink, G.B.M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotic, A., Shangguan, W., Wright, M.N., Geng, X., Bauer-Marschallinger, B.,Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars,
635 J.G.B., Ribeiro, E.,Wheeler, I., Mantel, S., and Kempen, B.: SoilGrids250m: global gridded soil informa-tion based on machine learning, PLoS ONE, 12 (2), e0169748, https://doi.org/10.1371/journal.pone.0169748, 2017.

Hiemstra, P., and Skoien, J.O.: Package 'automap', https://cran.r-project.org/web/packages/automap/automap.pdf, 2023.

Höge, M., Kauzlaric, M., Siber, R., Schönenberger, U., Horton, P., Schwanbeck, J., Floriancic, M. G., Viviroli, D., Wilhelm, S., Sikorska-Senoner, A. E., Addor, N., Brunner, M., Pool, S., Zappa, M., and Fenicia, F.: CAMELS-CH: hydro-meteorological
640 time series and landscape attributes for 331 catchments in hydrologic Switzerland, Earth Syst. Sci. Data, 15, 5755–5784, https://doi.org/10.5194/essd-15-5755-2023, 2023.

ISPRA: Bacini idrografici principali [data set], http://geoportale.isprambiente.it/dettagli/?uuid=ispra_rm%3A20101021%3A123000, 2005a.

ISPRA: Bacini idrografici secondari [data set],
645 http://geoportale.isprambiente.it/dettagli/?uuid=ispra_rm%3A20101021%3A120000, 2005b.

Jasiewicz, J.: r.stream.stats, https://grass.osgeo.org/grass82/manuals/addons/r.stream.stats.html, 2021.

Kirpich, Z.: Time of concentration of small agricultural watersheds, Civil Engineering 10(6), 1940, 362.

Klingler, C., Schulz, K., and Herrnegger, M.: LamaH-CE: LArge-SaMple DAta for Hydrology and Environmental Sciences for Central Europe, Earth Syst. Sci. Data, 13, 4529–4565, https://doi.org/10.5194/essd-13-4529-2021, 2021.

650 Koutsoyiannis, D.: Knowable moments for high-order stochastic characterization and modelling of hydrological processes, Hydrol. Sci. J., 64(1):19–33, https://doi.org/10.1080/02626667.2018.1556794, 2019.

Kratzert, F., Nearing, G., Addor, N., Erickson, T., Gauch, M., Gilon, O., Gudmundsson, L., Hassidim, A., Klotz, D., Nevo, S., Shalev, G., and Matias, Y.: Caravan - A global community dataset for large-sample hydrology, Sci. Data, 10, 61, https://doi.org/10.1038/s41597-023-01975-w, 2023.

655 Laio, F., Ganora, D., Claps, P., and Galeati, G.: Spatially smooth regional estimation of the flood frequency curve (with uncertainty), J. Hydrol., 408(1–2), 67-77, https://doi.org/10.1016/j.jhydrol.2011.07.022, 2011.

Lindsay, J.B.: The practice of DEM stream burning revisited, Earth Surf. Process. Landforms, 41, 658–668, https://doi.org/10.1002/esp.3888, 2016.

Lindsay, J. B., and Creed, I. F.: Removal of artifact depressions from digital elevation models: towards a minimum impact approach, Hydrol. Process., 19, 3113-3126, https://doi.org/10.1002/hyp.5835, 2005.

Linke, S., Lehner, B., Ouellet Dallaire, C., Ariwi, J, Grill, G., Anand, M., Beames, P., Burchard-Levine, V., Maxwell, S., Moidu, H., Tan, F., and Thieme, M.: Global hydro-environmental sub-basin and river reach characteristics at high spatial resolution, Sci. Data, 6, 283, https://doi.org/10.1038/s41597-019-0300-6, 2019.

Loritz, R., Stölzle, M., Guse, B., Kiesel, J., Haßler, S., Mälicke, M., Tarasova, L., Heidbüchel, I., Ebeling, P., Hauffe, C., Müller-Thomy, H., Jehn, F. U., Brunner, M., Götte, J., and Rohini, K.: CAMELS-DE: Initiative für einen konsistenten, frei verfügbaren Datensatz für hydro-meteorologische Analysen in Einzugsgebieten in Deutschland (1.2), Zenodo. https://doi.org/10.5281/zenodo.6517142, 2022.

Mazzoglio, P., Butera, I., and Claps, P.: I^2-RED: a massive update and quality control of the Italian annual extreme rainfall dataset, Water, 12, 3308, https://doi.org/10.3390/w12123308, 2020.

Mazzoglio, P., Butera, I., Alvioli, M., and Claps, P.: The role of morphology in the spatial distribution of short-duration rainfall extremes in Italy, Hydrol. Earth Syst. Sci., 2022a.

Mazzoglio, P., Ganora, D., and Claps, P.: Long-term spatial and temporal rainfall trends over Italy, Environ. Sci. Proc., 21, 28, https://doi.org/10.3390/environsciproc2022021028, 2022b.

Mazzoglio, P., Butera, I., Claps, P.: A local regression approach to analyze the orographic effect on the spatial variability of sub-daily rainfall annual maxima, Geomatics, Nat. Hazards Risk, 14(1), 2205000, https://doi.org/10.1080/19475705.2023.2205000, 2023.

Montgomery, D. R., and Foufoula-Georgiou, E.: Channel network source representation using digital elevation models, Water Resour. Res., 29(12), 3925– 3934, https://doi.org/10.1029/93WR02463, 1993.

Muñoz Sabater, J.: ERA5-Land monthly averaged data from 1981 to present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS) [data set], https://doi.org/10.24381/cds.e2161bac, 2019.

Nagy, E.D., Szilagyi, J., and Torma, P.: Assessment of dimension-reduction and grouping methods for catchment response time estimation in Hungary, J. Hydrol. Reg. Stud., 38, 100971, https://doi.org/10.1016/j.ejrh.2021.100971, 2021.

Pavan, V., Antolini, G., Barbiero, R., Berni, N., Brunier, F., Cacciamani, C., Cagnati, A., Cazzuli, O., Cicogna, A., De Luigi, C., Di Carlo, E., Francioni, M., Maraldo, L., Marigo, G., Micheletti, S., Onorato, L., Panettieri, E., Pellegrini, U., Pelosini, R., Piccinini, D., Ratto, S., Ronchi, C., Rusca, L., Sofia, S., Stelluti, M., Tomozeiu, R., Torrigiani Malaspina, T.: High resolution climate precipitation analysis for north-central Italy, 1961–2015, Clim. Dyn., 52:3435-3453, https://doi.org/10.1007/s00382-018-4337-6, 2019.

Ravazzani, G., Boscarello, L., Cislaghi, A., and Mancini, M.: Review of Time-of-Concentration equations and a new proposal in Italy, J. Hydrol. Eng., 24(10), https://doi.org/10.1061/(ASCE)HE.1943-5584.0001818, 2019.

Rossi, G., and Caporali, E.: Regional analysis of low flow in Tuscany (Italy), IAHS Publ. 340, 135-141, https://iahs.info/uploads/dms/15207.22-135-141-340-17-T1_Rossix.pdf, 2010.

Saxton, K.E., Rawls, W.J., Romberger, J.S., and Papendick, R.I.: Estimating generalized soil-water characteristics from texture, Soil Sci. Soc. Am. J., 50, 1031-1036, https://doi.org/10.2136/sssaj1986.03615995005000040039x, 1986.

Servizio Idrografico: Dati caratteristici dei corsi d'acqua italiani, Pubbl. N. 17, Istituto Poligrafico dello Stato, Roma, Italy, http://www.idrologia.polito.it/web2/open-data/Pubblicazione_17/1980%20new.pdf, 1980.

Settore Idrologico e Geologico Regionale: http://www.sir.toscana.it/index.php, 2022.

Sheridan, J. M.: Hydrograph time parameters for flatland watersheds. Transactions of the American Society of Agricultural Engineers, 37(1), 103–113, https://doi.org/10.13031/2013.28059, 1994.

Singh, P.K., Mishra, S.K., and Jain, M.K.: A review of the Synthetic Unit Hydrograph: from the empirical UH to advanced geomorphological methods, Hydrol. Sci. J., 59 (2), 239–261, 2013.

Soil Conservation Service: National Engineering Handbook, Section 4: Hydrology, Department of Agriculture, Washington DC, 762 pp., 1972.

Tarquini, S., Isola, I., Favalli, M., Battistini, A.: TINITALY, a digital elevation model of Italy with a 10 meters cell size (Version 1.0) [data set], Istituto Nazionale di Geofisica e Vulcanologia (INGV), https://doi.org/10.13127/TINITALY/1.0, 2007.

Vogt, J., Soille, P., De Jager, A., Rimaviciute, E., Mehl, W., Foisneau, S., Bodis, K., Dusart, J., Paracchini, M., Haastrup, P., Bamps, C.: A pan-European river and catchment database, OPOCE, Luxembourg (Luxembourg), JRC40291, https://core.ac.uk/download/pdf/38609451.pdf, 2007.

Yang, D., Herath, S. and Musiake, K.: Spatial resolution sensitivity of catchment geomorphologic properties and the effect on hydrological simulation, Hydrol. Process., 15, 2085-2099, https://doi.org/10.1002/hyp.280, 2001.