

Article

I²-RED: A Massive Update and Quality Control of the Italian Annual Extreme Rainfall Dataset

Paola Mazzoglio , Ilaria Butera  and Pierluigi Claps

Politecnico di Torino, Department of Environment, Land and Infrastructure Engineering, 10129 Torino, Italy; ilaria.butera@polito.it (I.B.); pierluigi.claps@polito.it (P.C.)

* Correspondence: paola.mazzoglio@polito.it

Received: 6 October 2020; Accepted: 23 November 2020; Published: 25 November 2020



Abstract: The collection and management of hydrological data in Italy has been dealt with at national level, initially, by the National Hydrological Service (SIMN), and at regional level in the last 40 years. This change has determined problems in the availability of complete and homogeneous data for the whole country. As of 2020, an updated and quality-controlled dataset of the historical annual maxima rainfall in Italy is still lacking. The Italian Rainfall Extreme Dataset (I-RED) has recently been created to allow studies to be performed with a homogeneous dataset at a national level. In this paper, the methodological approach adopted to build an improved and quality-controlled version of I-RED (in terms of both the rainfall depth values and the position of the rain gauges) is presented. The new database can be used as a more reliable research support for the frequency analysis of the rainfall extremes. This new I²-RED database contains rainfall annual maxima rainfall of 1, 3, 6, 12 and 24 h from 1916 until 2019, counts 5265 rain gauges and has been corroborated by a re-positioning and elevation-checking of 15% of the stations. A descriptive analysis of the maximum values of the stations, which provides an additional quality check and reveals different intriguing spatial features of Super-Extreme rainfall events, is also presented.

Keywords: annual maxima; dataset; rainfall; rainfall frequency analysis; Super-Extreme

1. Introduction

The frequent recurrence of intense rainfall events that trigger catastrophic flash floods is a climatic signature of Italy, and is one of the causes of the high vulnerability of the Italian territory, where a high population density is associated with a complex topography, extended coastlines and high climate variability. According to the Istituto Superiore per la Protezione e la Ricerca Ambientale (ISPRA), 4.1% of the Italian territory is characterized by a high probability of inundation scenario (return period of 20–50 years), 8.4% by a medium-probability scenario (return period of 100–200 years) and 10.9% by a low-probability scenario [1]. By merging these data with the population density, it is possible to estimate that more than 2 million inhabitants (3.5% of the entire Italian population) are currently living in a location with a high probability of such an event, while 6.2 and 9.3 million inhabitants live in medium- and low-probability event locations, respectively [1]. An accurate estimation of extreme rainfall (in terms of magnitude, duration and frequency) is urgently needed for such a complex and vulnerable territory for planning, design and management purposes. The need to build a uniform dataset of rainfall extremes for Italy is evident, due to its position in the middle of the Mediterranean Sea, in order to integrate meteorological and climate data from the past, and to better discriminate the hazard of rainfall deriving from Synoptic Systems, Convective Systems or Mediterranean Hurricanes (“Medicanes”) [2–5].

Rainfall frequency and large-scale trend analyses require a complete, updated and quality-controlled dataset of extreme rainfall depths, which are usually available in the form of annual maxima records.

Unfortunately, this type of information still requires a uniformity re-assessment for the Italian territory: such information is fragmented in Italy (each Italian Region collects a subset of this information) and processed/stored in different formats [2]. The most recent effort made to complete a uniform Italian database of rainfall extremes is the Italian–Rainfall Extreme Dataset (I-RED) [2], which is an upgrade of the first database, created as part of the Characterisation of Ungauged Basins by Integrated uSe of hydrological Techniques (CUBIST) project [6]. The latter was mainly based on the digitalization of the Hydrological Yearbooks of the National Hydrological and Mareographic Service (SIMN) [7] and on some additions resulting from the Valutazione delle Piene (VAPI) project [8]. Among other databases, it is worth mentioning the one available for extreme daily rainfall [5] and the one created by the Italian National Research Council (CNR) Research Institute of the Geo-Hydrological Protection (IRPI), which made reference to approximately 800 rain gauges in north-eastern Italy (Trentino-Alto Adige, Veneto and Friuli Venezia Giulia) [9]. As far as daily observational precipitation datasets are concerned, the ARchivio Climatologico per l'Italia centro-Settentrionale (ARCIS) database [10] is also interesting, as it was created using 1762 stations in north and central Italy, and interpolating them to produce a high-resolution (5×5 km) daily gridded dataset from 1961 to 2015. Other datasets that cover the Italian domain are reanalysis datasets, such as ERA5-Land (1981–present, 0.1° spatial resolution) [11], NOAA 20th-century reanalysis V3 (1836–2015, 1° spatial resolution) [12] and MERIDA (1990–present, 7 km spatial resolution) [13]. The advantages of working with a rain-gauges-based dataset, like I-RED, emerge both as regards the greater spatial resolution and in consideration that its annual maxima reflect the actual highest rainfall in moving windows with length for 1 to 24 h. This is not the case for the reanalysis datasets.

The original I-RED database has been already used in a few investigations (see, e.g., [14]), but its coverage is limited to 2010, which does not allow, e.g., to answer technical questions about recent and very severe weather events in Italy (e.g., the 2018 “Vaia Storm” [15]). We have therefore decided to update the available sources of data and to proceed with an advanced data quality control, in terms of consistency, redundancy, gap-filling of records and systematic geographical checks, to ensure that the corrected position and elevation of the stations could be of help in investigations on the spatial distribution of extreme statistics. The elevation of the Italian territory is quite variable, and it is therefore very important to deal with correct data about the geographic positions of the rain gauges in view of, for example, a geostatistical analysis, where the elevation provides additional information as a secondary variable [16,17].

Such an effort has led to the production of a newly compiled Improved I-RED (I^2 -RED) database, which includes the annual maximum rainfall depths for 1, 3, 6, 12 and 24 consecutive hours from 5265 weather stations across Italy, recorded between 1916 and 2019. Details about the quality-control procedures applied to correct inconsistencies and errors are presented in the following sections, as well as details about the automatic and non-automatic quality checks applied by the agencies involved in the data collection before the data can be released.

Findable, Accessible, Interoperable and Re-usable (“FAIR”) data principles [18,19] are addressed in I^2 -RED. The database shows a high level of uniformity: all the rainfall measurements are expressed in mm, while the positions of all the rain gauges are converted into a single, uniform reference system. Metadata are included in the database in pdf file format. The final version of the I^2 -RED database is available on Zenodo, albeit with restricted access, due to the data policies of the regional hydrological agencies. Information about the data policies is reported at the end of this manuscript. A Digital Object Identifier (<https://doi.org/10.5281/zenodo.4269509>) and relevant keywords have also been assigned. Data are stored in both mat and in csv format (to make it accessible to all open-source programs). No specific code is required to access or work with the data. The volume of the final database is less than 10 MB, thus making the database easily transferable and storable.

This paper is structured as follows: Section 2 deals with the details of the data collection; the data updating and gap filling are described in Section 3, while the applied quality control procedures, e.g., to recover station elevations, are described in Section 4. A description of the quality-controlled

database is given in Section 5, together with climatological information. An extreme value analysis and an additional error detection procedure are reported in Section 6. The discussion and conclusions are reported in Section 7, while the data policy is summarized in Section 8.

2. Data Collection

The historical data (i.e., data acquired before the dismantlement of SIMN) included in I-RED are freely available in the SIMN Hydrological Yearbooks [7]. The dismantlement of SIMN, which was planned with Legislation Decree 112/1998, resulted in the data collection and management tasks being transferred to a regional level, that is, to 19 Regions and 2 Autonomous Provinces. Once the regional services had been set-up, the local authorities adopted different data distribution policies, even though, according to the Italian law, an Open Data policy should be applied to public data for research and no-profit purposes. The most important laws regarding Open Data in Italy are Legislative Decree 82/2005, Legislative Decree 36/2006, Law 221/2012, Legislative Decree 179/2012 and Law 114/2014 [2]. As of 2020, some agencies have not published the data (raw or validated) on their website, and only make them available upon request, in order to track data usage and prevent them from being used for commercial purpose. Access to the data is granted after either sending a request or filling a request form that users can find on the agency's website. Those who have received authorization are then granted access to a database, the delivery of digitized tables or, on occasion, the delivery of pdf files. Information about the validation criteria applied by the regional agencies and measurement accuracy are, unfortunately, not provided together with the data. In early 1900, most of the instruments were mechanical ones. Starting from about 1980, all old mechanical stations were replaced with mechanical and digital ones. This led to data with different accuracies within the same dataset.

A list of the hydrological agencies that conduct the data collection and management after the dismantlement of SIMN are reported in Appendix A. The Provinces of Trento and Bolzano are located in the same Region, but are handled separately, because they have different data providers. In order to include them in the list, we generically refer to Italian "regions". Figure 1 summarizes the data policies of the original data, as reported in Appendix A. Information about the data policies and how raw data can be accessed is also included. In the new database, I²-RED, each Italian "region" has a specific progressive number that ranges from 1 to 21, and this number is reported in the rainfall station ID assigned to I²-RED. The last three digits of each rainfall station ID refer to a progressive number within each "region", while the first number (1 to 21) identifies the "region" and refers to the progressive numbers reported in Appendix A.

Because of the complexity of the data-merging operations, most of the work requested to build the new I²-RED database was performed in manual mode, by visually inspecting the data. Data merging was of fundamental importance, particularly as some "regions" did not have an editable, easily accessible and complete version of the database (most of them store the data in pdf files, on a yearly basis). Only some "regions" have carried out activities in order to merge historical and new data in an easily accessible format. Different geographic coordinates systems are used by the regional agencies. In some cases, historical and new data are even geo-referenced within the same "region" using different systems. In the I²-RED database, all the station coordinates were made available in the Universal Transverse Mercator coordinate system (that is, WGS 84 UTM 32 N, EPSG 32632) to guarantee uniformity at the national level.

Like the I-RED database, which was created following four different procedures, depending on the data availability [2], the creation of I²-RED required the application of different methods that would be able to allow a rigorous quality check to be made, as described in detail in Section 3. Other quality checks are also described in Sections 4 and 6. The entire workflow is summarized in Figure 2.



Figure 1. Data policy of the data reported in I²-RED for the 21 Italian “regions”. The provinces of Trento and Bolzano, despite being located in the same Region, are shown separately in the map as the data providers and the policies are different.

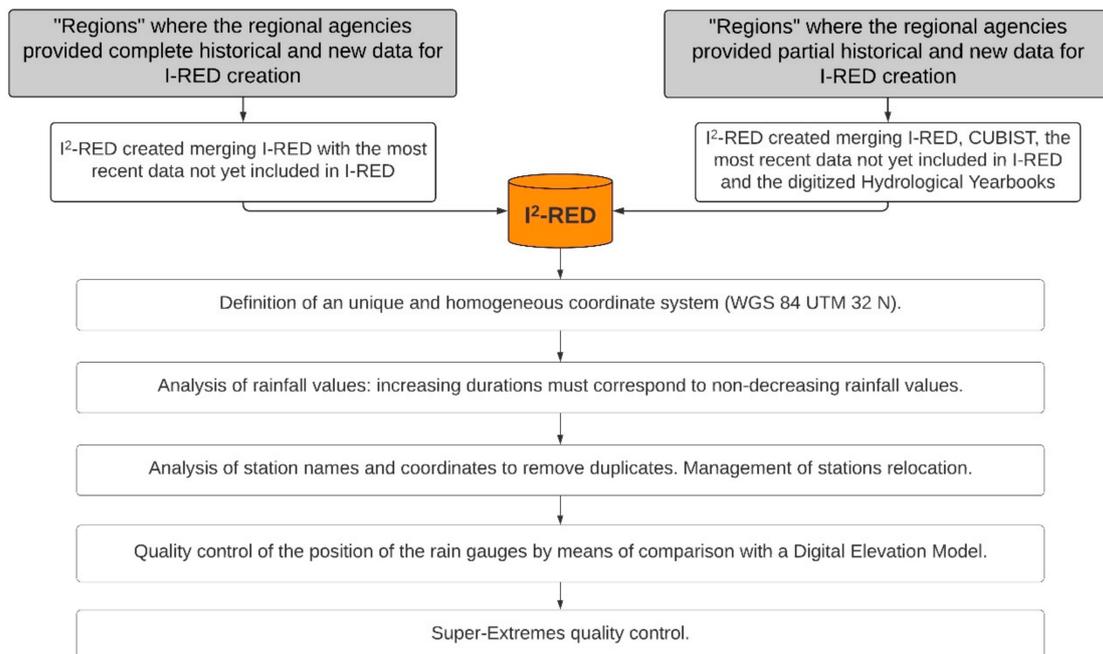


Figure 2. Quality control workflow applied during the creation of I²-RED.

The newly compiled I²-RED database consists of 5265 stations, and the data covers the 1916–2019 period (the number of available stations can be seen in Figure 3). The lack of a considerable

number of measurements between 1940 and 1950 is due to the Second World War, while the apparent reduction in the number of rain gauges between 1996 and 2000 was mainly induced by the effect of the dismissing of SIMN: in those years, some “regions” decided not to upgrade or repair the old rainfall stations, and most of the old mechanical tipping-bucket rain gauges were progressively substituted by automatic ones. Finally, the decreasing number of stations that emerges in recent years is due to the fact that, for the most recent years, data were not yet available in some “regions” (e.g., the most recent data available for the Trento province date back to 2009, the most recent data for Piemonte date back to 2017).

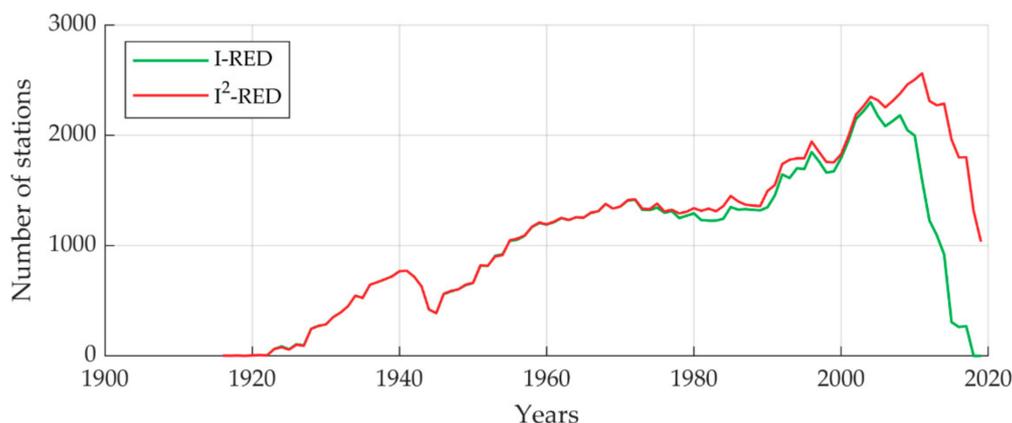


Figure 3. Number of stations included in the two I-RED databases year by year. The green line refers to I-RED, while the red line refers to I²-RED.

3. Data Update and First Quality Control

As reported in Section 2, the data updating was accompanied by a robust and ad-hoc quality check that has taken into account how I-RED was originally created. In particular, in “regions” where the data provider had already digitized their part of the SIMN database, the CUBIST database was not used. In this case, the data updating performed in I²-RED involved updating I-RED with the most recent data. In the other cases (i.e., in “regions” where the agencies only had partial coverage of historical data), we extracted the remaining historical data from the CUBIST database, and the data updating performed in I²-RED was carried out manually in three steps, to maximize the available data and avoid duplicates. First, the I²-RED database was updated to include the data for the most recent years that was not included in I-RED. Secondly, any data that were missing were extracted from CUBIST and then included. Thirdly, because of the low number of digitized data for some years, a manual data input from the Hydrological Yearbooks proved to be necessary to fill the gaps in both the CUBIST database and in the digitized versions of the historical data received from the local authorities. In order to do so, the Hydrological Yearbooks (available as a scanned copy in pdf format) were digitized using optical character recognition (OCR) software and compared with the already available data to detect values that had not been included.

Thanks to the previously mentioned gap-filling procedures in I²-RED, it was possible to reduce the percentage of missing transcriptions. However, some inconsistencies in the series still persisted, e.g., cases in which a rainfall measurement was not available for a specific duration but was available for others. This could be due to transcription mistakes in the Hydrological Yearbooks. Table 1 shows the percentage of these missing transcription within both databases, and also highlights the greater completeness reached with I²-RED.

Table 1. Missing transcription in I-RED and I²-RED for the different durations. The percentages were evaluated by comparing the total number of records related to each duration (i.e., 105,940 for I-RED and 122,217 for I²-RED).

Database	1 h	3 h	6 h	12 h	24 h
I-RED	0.98%	1.23%	1.12%	1.19%	0.88%
I ² -RED	0.87%	1.06%	1.01%	1.06%	0.78%

A first algorithm-based analysis was performed to verify that rainfall values corresponding to increasing durations were consistent (i.e., the annual maxima related to a 1-h interval had to be lower or at least equal to the 3-h one). Several errors emerged within the database, thus highlighting the necessity of checking the value of the raw data in order to substitute it with the correct one or with a “Not a Number”.

For a complete description of the first raw data quality improvement achieved in I²-RED, it is important to highlight that almost every hydrological agency also has rainfall stations in nearby “regions”, as a result of the joint management of interregional basins. The values reported in the I²-RED database were therefore checked by comparing the time series of each rain gauge with all the other ones: this check allowed us to identify some stations located along the regional boundaries that had been included twice (sometimes with the same station name, sometimes not) with similar positions and the same equal rainfall values. As some “regions” shared their data with other agencies before applying a validation procedure, the visual data check was of fundamental importance to avoid data duplication. Stations with the same name, similar coordinates and with a time overlap were merged if the rainfall measurements were the same, otherwise they were considered as different stations in the presence of geographic or record length discrepancies. A few cases of station-overlapping were also found in the case of stations with almost the same latitude/longitude (i.e., at a distance of less than 200 m).

A brief mention should also be made of the thorny problem of station relocation. Unfortunately, some “regions” did not track the relocation of some stations by changing the name, the position, or the instrument code in their databases. When it was possible to recognize that new and historical data had been collected by different weather stations (e.g., thanks to some years of missing data or to the information collected by regional authorities, reports), the data were handled separately, and two different stations were reported in I²-RED. When it was not possible to obtain such information, the time series were merged and only one rain gauge was considered.

4. Quality Control on the Station Elevation

As mentioned in the Introduction, one of the main innovations of the I²-RED database, with respect to I-RED, is the extensive application of quality control, both through manual and automated methods. We used the European Digital Elevation Model (EU—DEM), version 1.1 [20], a Digital Surface Model (DSM) with European coverage, a 25 m resolution and a vertical accuracy equal to ± 7 m to check the reliability of each station elevation.

The planimetric position of each rain gauge was used to extract the corresponding elevation values from EU-DEM. After a comparison between the elevations obtained from EU-DEM and those reported in I-RED, several errors emerged in the original database. The relative difference between these two measurements was evaluated, and a threshold value equal to 150 m was used to identify rain gauges that needed to be checked and corrected before using the data for further analyses. As reported in Figure 4a, 790 stations out of a total of 5265 were affected by significant elevation errors, which varied from 150 to more than 2000 m.

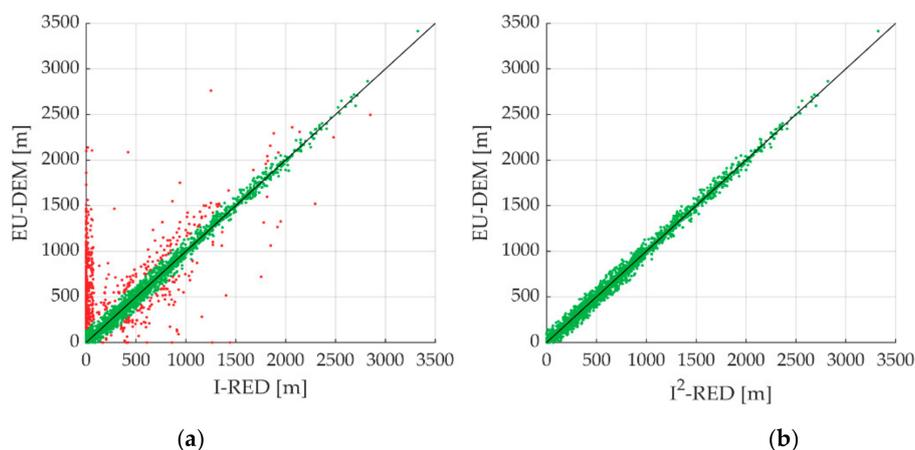


Figure 4. Comparison between the elevation values before (a) and after (b) the quality control. The red dots in (a) identify the rain gauges included in I-RED that are characterized by an elevation discrepancy greater than 150 m, while the green ones identify rain gauges with a discrepancy lower than 150 m. (b) refers to the same information for the I²-RED database.

During this geographic analysis, it emerged that there were three types of elevation errors: (i) only an incorrect elevation, (ii) an incorrect position but correct elevation, (iii) both an incorrect position and an incorrect elevation. In order to understand the reasons for this variety of errors, it is important to note that the accuracy of the coordinates is not uniform in either the I-RED database or in the I²-RED one. The coordinates of the rain gauges installed in recent years are characterized by metric or sub-metric precision, while the old ones have kilometric precision (i.e., in the Hydrological Yearbooks, they are expressed in degrees and in minutes, and no information is reported about seconds). Such a low positional accuracy explains why considerable errors could arise in the position and, more importantly, in the elevation, especially in mountainous areas.

A manual correction was applied to each of the points highlighted with a red dot in Figure 4a. In particular, information about the correct position was collected from other sources available online (e.g., reports, scientific publications). In some cases, the agencies involved in the data collection were contacted directly to discuss the issue. In the case of the Autonomous Province of Bolzano-Alto Adige, it was possible to partially substitute the positions derived from the Hydrological Yearbooks with more accurate ones, thanks to detailed information received from the competent regional agency. When possible (i.e., for the most recent stations), Google Earth was used to visually check whether a rain gauge was present in the position reported in the database, taking advantage of the Street View mode. In a few cases, when it was not possible to obtain further information about the position (e.g., for stations that were active in the early 1900s), a rain gauge was not relocated, and the elevation extracted for EU-DEM was confirmed. The results of the elevation quality control are shown in Figure 4b, which shows that all the rain gauges previously affected by an elevation error greater than 150 m have now been corrected in the I²-RED database.

The distribution of the rain gauges over the different elevations in Italy is reported in Table 2. As can be seen, 34% of the rain gauges are located in sites with a lower elevation than 200 m a.s.l., while 62% of the rain gauges are located in sites with a lower elevation than 500 m a.s.l. Only 5% of the rain gauges are located in areas with a high elevation (i.e., >1500 m a.s.l.).

Table 2. Distribution of the rain gauges included in I²-RED over different elevations.

Elevation (m)	Number of Stations	Percentage
≤200 m	1782	34%
200 m < x ≤ 500 m	1499	28%
500 m < x ≤ 1000 m	1314	25%
1000 m < x ≤ 1500 m	401	8%
>1500 m	269	5%

5. Time and Space Uniformity of the I²-RED Database

After the operations described in the previous Sections, a new and reliable data source was compiled for the extreme rainfall in Italy. The main features of the data collection and of the update are presented in this section.

I-RED [2] was made up of 105,940 different records for every duration. The new I²-RED database presents 122,217 records. By dividing the number of records by the number of stations (5265), it is possible to obtain the average length of the time series: 22.5 years for I-RED and 23.2 years for I²-RED.

Some features of the rainfall stations are mapped in Figure 5. Figure 5a, in particular, maps the length of the time series of each rain gauge included in I²-RED. Out of a total of 5265 rain gauges:

- 3782 rain gauges have a time series equal to or longer than 10 years;
- 2283 rain gauges have a time series equal to or longer than 20 years;
- 1436 rain gauges have a time series equal to or longer than 30 years;
- 693 rain gauges have a time series equal to or longer than 50 years;
- 52 rain gauges have a time series equal to or longer than 80 years.

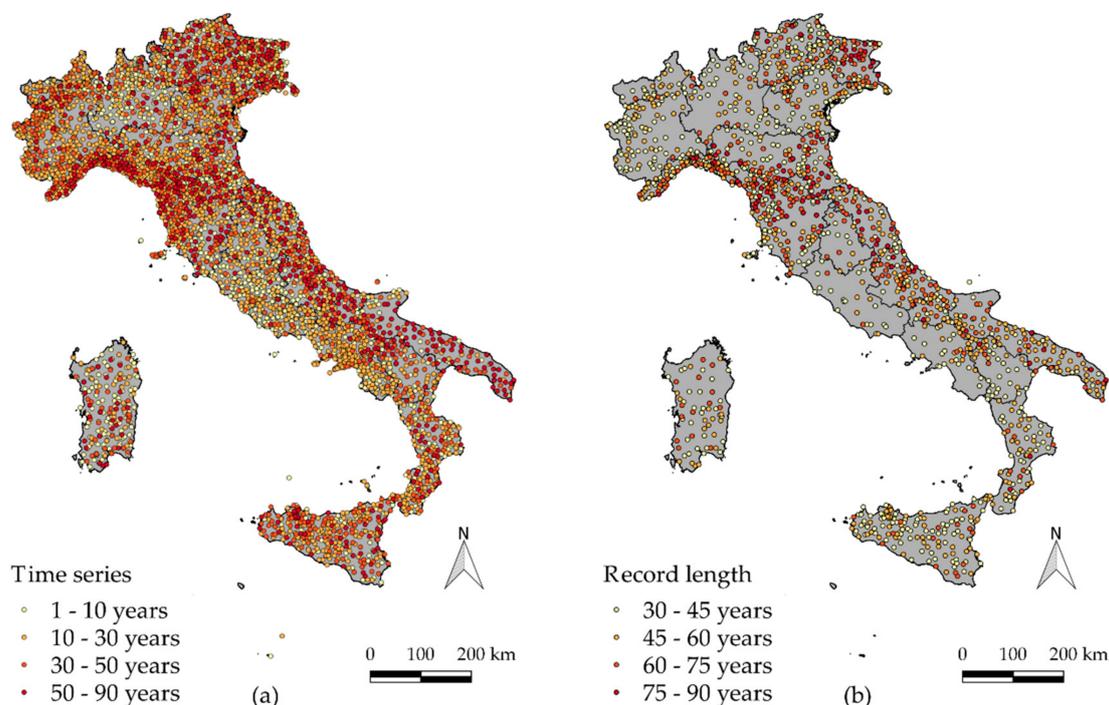


Figure 5. (a) Position of the rain gauges in I²-RED. The color refers to the length of the time series. (b) Spatial distribution of the rain gauges with a time series longer than 30 years.

The evolution in the management of the rain gauge network is one of the main factors that has led to a short and fragmented dataset in Italy. It is important to highlight that, although the database consists of 5265 rain gauges, only 1436 of these are characterized by a time series of at least 30 years, which is

the standard length for climatological analyses [21,22] (Figure 5b). Although the most recent stations were installed to locally increase the network density, they unfortunately still provide little suitable information for these types of analysis. The inhomogeneity of the record length is also influenced by the different degrees of accuracy adopted by each regional agency. In fact, some agencies recorded small relocations of the station position, which do not break the series, while others did not (as already pointed out in Section 3). In order to become aware of the spatial and temporal evolution of the Italian rain gauge network, it may be useful to watch a stop-motion video that shows a map of the active rainfall stations for each year of the whole observation period (Supplementary Material, Figure S1).

The geographical uniformity and density of the stations throughout Italy are also worth considering. Observing the map in Figure 5b, it is possible to observe that Umbria, Lazio, Sardegna, Campania and Lombardia are characterized by the lowest densities (for named “regions” see Figure 1). In particular, the coastal area of Campania shows a very low number of rain gauges, even though it is known to be an area that is affected by a high number of extreme events. A general picture of the network density in the different “regions” in Italy is reported in Table 3.

Table 3. Number of rain gauges with at least 30 years of measurements. The areas were derived from ISTAT data [23].

ID	“Region”	Number of Rain Gauges	Area (km ²)	Average Contributing Area (km ² /rain Gauge)
1	Abruzzo	69	10,832	157
2	Basilicata	37	10,073	272
3	Calabria	69	15,222	221
4	Campania	36	13,671	380
5	Emilia-Romagna	125	22,453	180
6	Friuli Venezia Giulia	64	7862	123
7	Lazio	28	17,232	615
8	Liguria	94	5416	58
9	Lombardia	64	23,864	373
10	Marche	47	9401	200
11	Molise	26	4461	172
12	Piemonte	106	25,387	240
13	Puglia	98	19,541	199
14	Sardegna	58	24,100	416
15	Sicilia	143	25,832	181
16	Toscana	152	22,987	151
17	Autonomous Province of Trento	39	6207	159
18	Autonomous Province of Bolzano-Alto Adige	46	7398	161
19	Umbria	13	8464	651
20	Valle d’Aosta	14	3261	233
21	Veneto	108	18,407	170

However, more detailed information about the station density is still needed, due to the particular form of the Italian peninsula, and to an apparent proliferation of stations in highly urbanized areas. A clear variation in station density can, in fact, affect the spatial analyses of extreme rainfall quantiles, or evaluations related to the areal persistence of trends, as pointed out by [14]. Figure 6 shows two density maps built with the I²-RED database, using a spatial grid of 2 × 2 km. The number of rain gauges available within a 10 km radius, centered in the cell, was computed for each pixel of the grid (Figure 6a). Dense spots may be observed in correspondence to the cities of Rome, Turin, Bologna, etc. The sum of all the observations pertaining to the stations contained in the circle was systematically computed to take into account the density of observations (Figure 6b), thereby allowing any evident non-uniformity in the data density to emerge. High-density values can be observed for Northern Italy, especially over Liguria, southern areas of Piemonte, Toscana and Friuli Venezia Giulia.

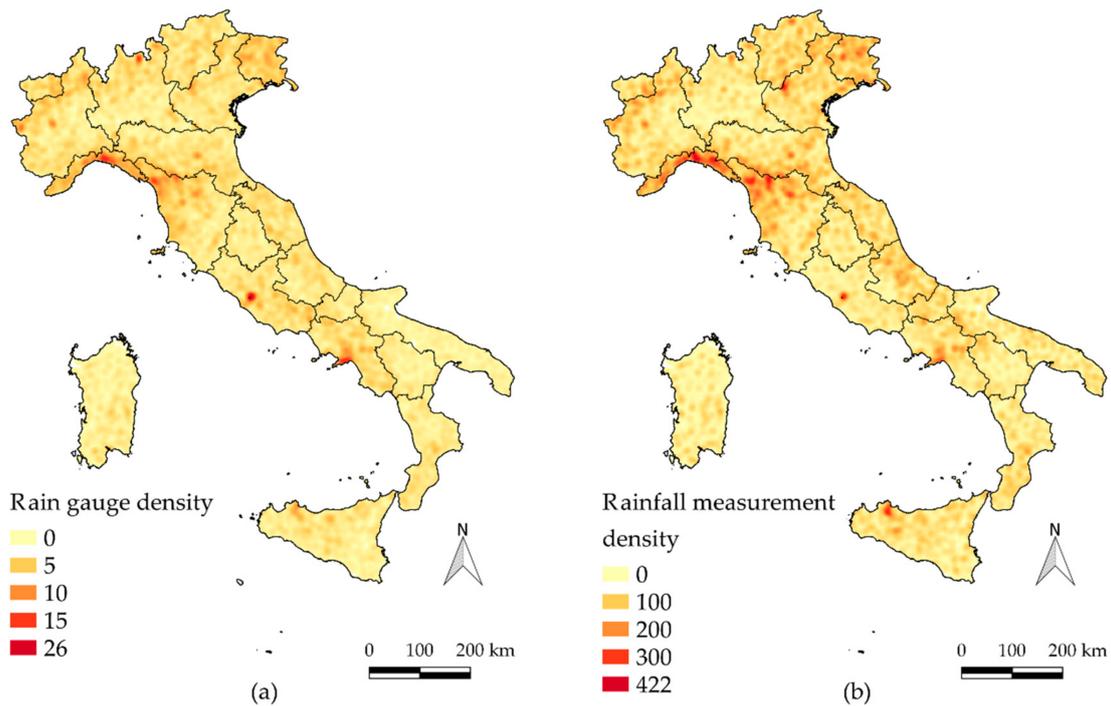


Figure 6. (a) Kernel density map (obtained using QGIS Heatmap Plugin) with a spatial resolution of 2×2 km. The colors refer to the number of rain gauges available within a 10 km radius centered in each cell. (b) Kernel density map obtained with a distance equal to 10 km, spatial resolution of 2×2 km, and weighted with the length of the time series. The colors refer to the number of measurements (i.e., the sum of the lengths of all the time series available within a 10 km radius centered in each cell).

In order to explore the potential of the database to map the climatology of extreme rainfall at the national scale, a first preliminary analysis was conducted over the entire domain. We started with the representation of the parameters of the depth–duration–frequency curves (DDF). The mean depth–duration curve is often expressed in the form of a power–law relationship [24]

$$h_d = ad^n \quad (1)$$

where h_d is the average of the annual maximum precipitation depth for the duration d , and a and n are site parameters estimated by means of linear regression between d and the available values of h_d . If a rainfall quantile h_d is required for a given return period T , the above expression can be multiplied with a growth factor $K(T)$, according to the index-value approach [25].

Rain gauges with at least 10 years of data were used to estimate both a and n for all the gauge sites. Kriged maps were then produced by applying an ordinary kriging with a spherical variogram and variable windows (a maximum value of 30 km was set to find at least five rain gauges). In order to avoid the smoothing effect of kriging, induced by the fact that a station grid may not have coincided with the interpolation grid, pixels that contained one or more rain gauges were forced to have the DDF parameters equal to the maximum ones obtained within the pixel. The mapped values for a and n are shown in Figure 7a,b.

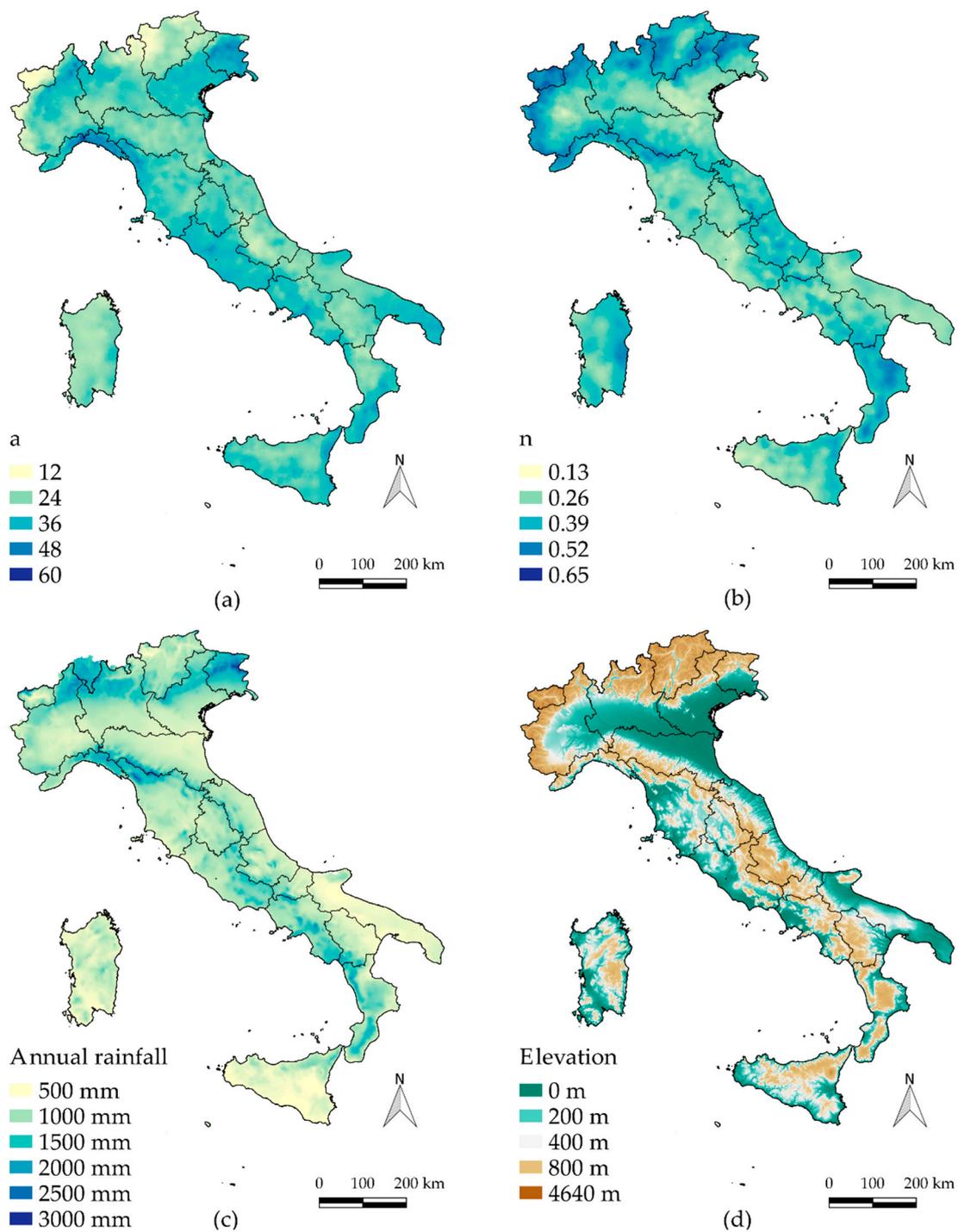


Figure 7. Parameters a (a) and n (b) of the DDF curves evaluated on the basis of the data contained in I^2 -RED. The maps are characterized by a 1×1 km spatial resolution. (c) Mean annual precipitation as mapped by [26]. (d) Elevation data, Shuttle Radar Topography Mission (SRTM) digital elevation model.

The parameter a is approximately the average maximum rainfall recorded for a 1-h interval, and it shows higher values in Liguria, Friuli Venezia Giulia and in the northern part of Piemonte (Figure 7a). This parameter shows some similarities with the spatial distribution of the mean annual rainfall (Figure 7c) that is used here for climatic comparisons. This last map was compiled by [26] at a 30-arc-second resolution using rainfall measurements related to the 1961–1990 period. The parameter n (Figure 7b) shows higher values for high-altitude areas, such as in the alpine region. To show

to which extent the spatial distribution of these parameters can agree with known phenomena, we reported a Digital Elevation Model of Italy in Figure 7d. Local factors, such as the steepness of the terrain, slope orientation, aerosol content and stability of the atmosphere, can influence the orographic enhancement of precipitation, especially for long aggregation intervals [27]. This can be recognized in the map of the exponent n (Figure 7b) as the higher n , the higher the increase in h_{24} with respect to the corresponding value h_1 (parameter a). On the other hand, short aggregation intervals (e.g., 1 h) show a different behavior, which leads to decreasing rainfall amounts as the altitude increases, especially over Northern Italy, as already shown [28,29].

Similar results, in terms of DDF parameters, were obtained in regional-scale studies over Piemonte [30], the Trento province [31] and Sicilia [32], thus confirming the importance of working with a complete and updated dataset in order to avoid problems along the regional boundaries. The maps presented here can help show the influence of the terrain morphology on the spatial variability of the averages of the rainfall extremes and point out the importance of having reliable information for the plano-altimetric position of stations.

6. Super-Extreme Quality Control and Descriptive Statistics

The highest values measured in each station, which were defined as “Super-Extremes” [33], have a particular importance in relation to the occurrence of the highest rainstorms in the Mediterranean area, and are worth emphasizing. We analyzed these “station maxima”, computed for all the stations and all the durations, to possibly identify further errors in the data, which had not been detected in the data digitization and in the record upgrade process. By examining these values, it was indeed possible to detect unrealistically high annual extremes, after dividing them by the local mean annual precipitation depth (shown in Figure 7c).

This ratio, called “Normalized” Maximum Value (NMV), has helped to detect very high rainfall amounts relative to the climatology of the site. The NMV index was built as

$$NMV = \frac{X_n}{MAP} \quad (2)$$

where X_n is the station maximum value, while the mean annual precipitation (MAP) comes from an existing digital map [26]. This index was introduced to highlight truly “extreme” events, that is, to distinguish them from other events, albeit very intense, which occurred in locations characterized by high annual rainfall amounts. Using various thresholds for NMV and considering the related frequency of occurrence, our aim was to provide an empirical, yet objective, representation of the spatial distribution of the highest relative extreme rainfall events in Italy, and to obtain clues about any local anomalies. In this regard, Figure 8 shows, as an example, the positions where NMV exceeds 0.5 for the 24-h aggregation interval. In comparison to [5], and to some other well-known hydrometeorological disasters that had happened in Italy, the presence of some clusters was predictable, e.g., in Sicilia, in Sardegna and in the Calabria Regions.

However, some anomalies emerged for some very isolated values (circled dots in Figure 8). These values were checked by verifying the hourly or daily rainfall measurements acquired by the same and by nearby stations (when available) and by looking for evidence of a relevant heavy rainfall event in the news, in scientific articles and/or in disaster reports. After all these controls, all of these exceptional values proved to be the results of errors, and in all cases were attributable to a misplaced decimal separator in the original printed figures; the corresponding data were therefore adjusted in the dataset.

Figure 9a shows the highest NMV values of the corrected dataset for the 24-h aggregation interval, after lowering the threshold to 0.35. The map shows that most of the values are concentrated in areas with low mean annual precipitation (see Figure 7c as a reference). This may imply that the flooding impact in those areas may be higher, as infrastructures and the populations are not accustomed to frequent and intensive rainfall events in a dry climate.

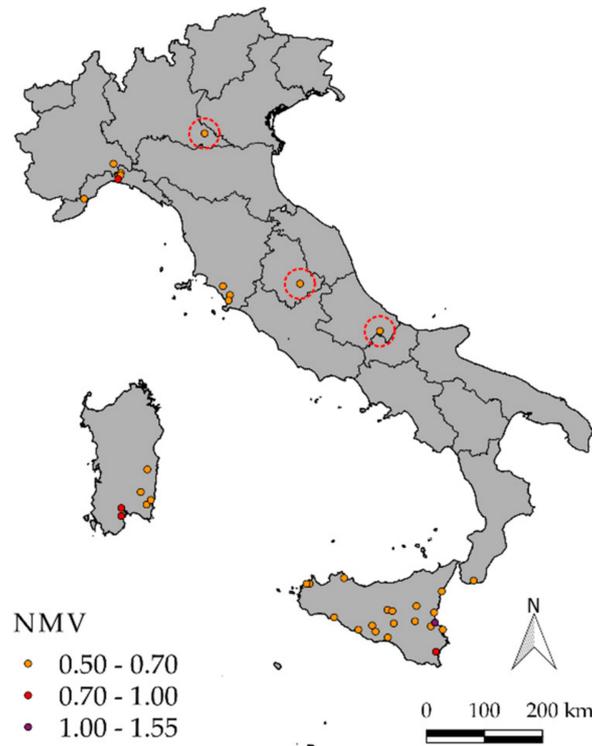


Figure 8. Rain gauges characterized by greater “Normalized” Maximum Values (NMV) than 0.5 for the 24-h interval. The red circles indicate anomalous values.

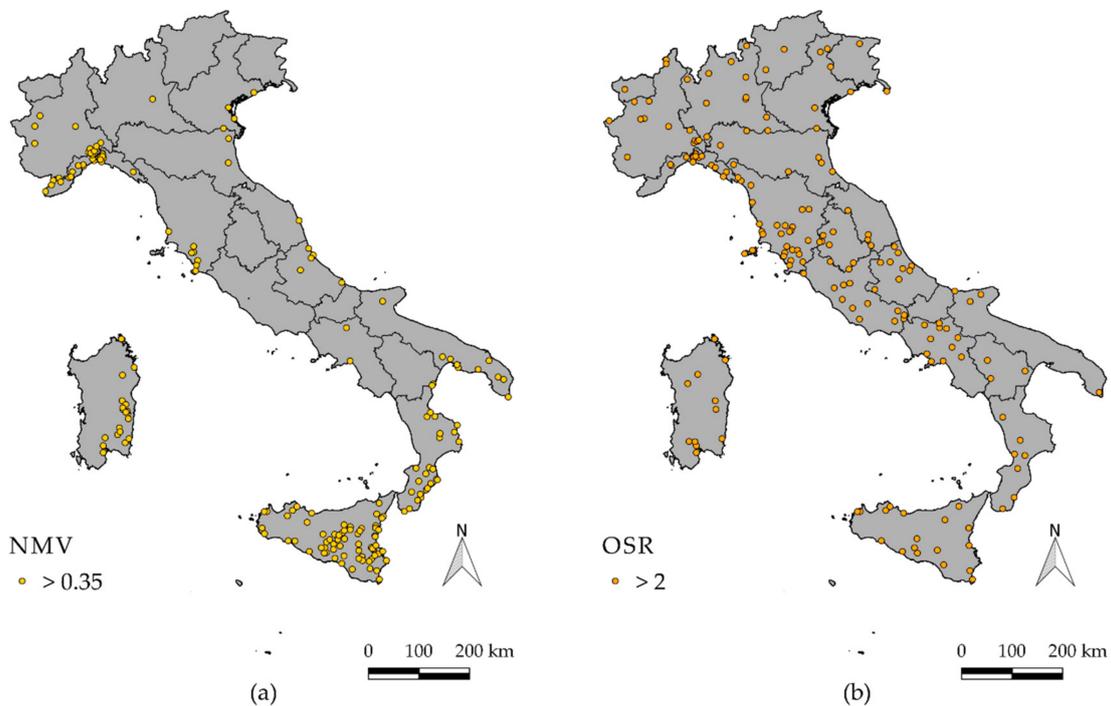


Figure 9. (a) NMV of the corrected 24-h dataset. Values lower than 0.35 were removed from the figure in order to highlight the localities where the most extreme rainfall events occurred. (b) Rain gauges characterized by greater OSR than 2 for the 24-h interval.

A different empirical index was applied to station maxima considering the highest k th sample values. In a record of size n , the n th order statistic is the maximum value of the record (X_n), while the

$n-1$ th order statistic is the second-highest value (X_{n-1}) and so on [34]. By extracting the n th and the $n-1$ th order statistics of the annual maxima for each rain gauge, we were able to compute a non-dimensional index, named the “Order Statistic Ratio”, as

$$OSR = \frac{X_n}{X_{n-1}} \quad (3)$$

In [35], one can find some results on the distribution of the ratio of two smallest order statistics from an exponential variate. Even though no available results were found in the literature about finite-sample distribution of the ratio of two highest-order statistics from extreme-value distributions, we assume that can be of interest to observe its spatial variability. This empirical ratio can be considered as an “outlier index” and has the advantage of being independent of preliminary assumptions about the probability distribution of the various samples.

The spatial distribution of the higher *OSR* values, computed for the 24-h maxima, can be observed in Figure 9b, where only rain gauges characterized by a greater *OSR* than 2 are reported. Counter-intuitively to the outcomes of previous analyses, the spatial distribution of these points is rather uniform, thus showing that high *OSR* values occur (and may occur in the future) in almost every site: even the areas with the lowest data density show these events. However, the Liguria Region is an exception, as it presents a significant concentration of high *OSR* values.

An attempt was made to find *NMV* and *OSR* threshold values that would allow a few but significant extreme events to be identified. In this regard, applying greater *NMV* thresholds than 0.35 and *OSR* ones greater than 2 allowed us to identify almost the same number of “Super-Extremes” (176 in the first case, 172 in the second one). These maps show that, with almost the same number of points, different selection criteria produce significantly different spatial distributions of these “severity” indexes. In both cases, the applied normalizations make the relevance of the selected events somewhat “relative”, while absolute criteria (i.e., a threshold value of 250 mm/day) are applied in other cases [5].

7. Conclusions

Despite the high rain gauge network density throughout the Italian territory, an updated and quality-controlled dataset of annual maxima rainfall events over the entire Italian territory was not available before this work. The availability of only short and fragmented series of rainfall measurements, combined with the reduced spatial domains covered by each regional agency, had led to a somewhat limited representativeness of the results used to represent the climatology of extreme events in Italy, as shown, e.g., by [14]. Regional frequency analyses and geostatistical analyses applied to regional-level datasets are also known to lead to artefacts along the regional boundaries (i.e., the borders), when measurements in the nearby “regions” are not included [36,37].

In the present study, we propose a new, updated and quality-controlled dataset of the annual maxima rainfall events and a representation of descriptive statistics that may be used to highlight the areas that are more vulnerable to extreme rainfall. More than 2000 different errors (in terms of incorrect planimetric/altimetric positions of the rain gauge or rainfall values) were detected in the original data and subsequently adjusted to improve the accuracy of a previous version of the database (I-RED), and this in turn led the release of I²-RED.

The study shows simple but effective methods for the detection of errors in both the rainfall records and in the station positions. These methods are mainly based on the use of a digital elevation model to identify any position/elevation errors, and on the analysis of the severity of the most relevant extreme rainfall events.

Finally, analyses of the 24-h maxima of the stations were performed to preliminarily exploit the potential of the proposed database and to fine-tune the selection of the possible data errors. The maps of these Super-Extremes indices suggested that a high spatial persistence of particularly extreme rainfall events is concentrated in the coastal areas of the Italian territory, possibly due to the effects induced by a steep and complex orography near the sea [38]. Further studies, based on the new database

presented here, can suggest methods that can be used to objectively identify a coherent selection of Super-Extreme events, with the aim of contributing to the knowledge on the spatial distribution of this particular hazard.

8. Data Policy

As mentioned in Section 2, although the Italian law requires an Open Source policy for all public data, this right has not yet been implemented by all the Italian agencies. The agreements we signed with some of the agencies involved in the data collection, aimed at monitoring the correct use of the data, restricted their use to the aims of the authors' project. As a result of these legal restrictions, a complete version of I²-RED can only be provided to two groups of people:

- People in the authors' research group (which is already fully authorized to use the data);
- People who can prove they have received clearance from the regional authorities.

Currently, the entire quality-controlled database is available on Zenodo, albeit with restricted access. The data can be used by third parties, for an indefinite timeframe, upon having completed an agreement either with the authors or the regional agencies involved in the data collection.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2073-4441/12/12/3308/s1>, Figure S1: temporal evolution of the length of the time series of every rain gauge included in I²-RED.

Author Contributions: Conceptualization, P.C., I.B., P.M.; methodology, P.C., I.B., P.M.; software, P.M.; validation, P.M.; formal analysis, P.M.; investigation, P.M.; resources, P.C., I.B.; data curation, P.M.; writing—original draft preparation, P.M.; writing—review and editing, P.C., I.B., P.M.; visualization, P.C., I.B., P.M.; supervision, P.C., I.B.; project administration, P.C., I.B., P.M.; funding acquisition, P.C., I.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: The authors would like to thank the regional agencies involved in the data collection and management for having provided the data. Credits are reported in Appendix A.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

This section summarizes the information about the regional agencies involved in the data collection and management after the dismantling of SIMN. The number assigned to each "region" mentioned in this list is equal to the station ID reported in I²-RED.

1. Abruzzo, Ufficio Idrografico e Mareografico Regione Abruzzo: data acquired until 2008 are available free of charge on <https://www.regione.abruzzo.it/content/annali-idrologici>, while data acquired from 2009 onward are available free of charge for research and no-profit purposes, upon request.
2. Basilicata, Dipartimento Protezione Civile Regione Basilicata: data are available free of charge on <http://www.centrofunzionalebasilicata.it/it/annali1.php>.
3. Calabria, ARPACAL–Centro Funzionale Mulirischi: data are available on <http://www.cfd.calabria.it> (password-protected website, access granted upon request).
4. Campania, Centro Funzionale Regione Campania: data are available free of charge for research and no-profit purposes, upon request.
5. Emilia-Romagna, ARPA Emilia Romagna: data are available free of charge on https://www.arpae.it/documenti.asp?parolachiave=sim_annali&cerca=si&idlivello=64.
6. Friuli Venezia Giulia, ARPA Friuli Venezia Giulia: data are available free of charge for research and no-profit purposes, upon request.

7. Lazio, Centro Funzionale Regione Lazio: data acquired until 2004 are available free of charge on http://www.regione.lazio.it/rl_protezione_civile/?vw=contenutidettaglio&id=153, while data acquired from 2005 onward are available free of charge for research and no-profit purposes, upon request.
8. Liguria, ARPA Liguria: data are available free of charge on <https://www.regione.liguria.it/articoli/26-servizi-online/14279-banca-dati-meteoclimatica.html>.
9. Lombardia, ARPA Lombardia: data are available free of charge on <http://idro.arpalombardia.it/pmapper-4.0/map.phtml>.
10. Marche, Dipartimento di Protezione Civile Regione Marche: data are available on <http://console.protezionecivile.marche.it> (password-protected website; access granted upon request).
11. Molise, Centro Funzionale Regione Molise: data are available free of charge for research and no-profit purposes, upon request.
12. Piemonte, ARPA Piemonte: data are available free of charge on https://www.arpa.piemonte.it/rischinaturali/accesso-ai-dati/annali_meteoidrologici/annali-meteo-idro/banca-dati-meteorologica.html.
13. Puglia, Dipartimento di Protezione Civile Regione Puglia: data are available free of charge on <https://protezionecivile.puglia.it/centro-funzionale-decentrato/rete-di-monitoraggio/annali-e-dati-idrologici-elaborati/annali-idrologici-parte-i-download/>.
14. Sardegna, ARPA Sardegna: data acquired until 2011 are available free of charge on <https://www.regione.sardegna.it/j/v/25?s=205270&v=2&c=5650&t=1>, while data acquired from 2012 onward are available free of charge on <http://www.sardegnaambiente.it/index.php?xsl=611&s=21&v=9&c=93749&na=1&n=10>. Other data acquired from ARPAS are available upon payment.
15. Sicilia, Osservatorio delle Acque Regione Siciliana: data are available free of charge for research and no-profit purposes, upon request.
16. Toscana, Servizio Idrografico Regione Toscana: data are available free of charge on <http://www.sir.toscana.it/consistenza-rete>.
17. Autonomous Province of Trento, Centro Funzionale Provincia Autonoma di Trento: data are available free of charge on <https://www.meteotrentino.it/#!/content?menuItemDesktop=143>.
18. Autonomous Province of Bolzano – Alto Adige, Ufficio Idrografico Provincia Autonoma di Bolzano–Alto Adige: data are available free of charge for research and no-profit purposes, upon request.
19. Umbria, Regione Umbria–Servizio Idrografico Regionale: data are available free of charge on <https://annali.regione.umbria.it/#>.
20. Valle d’Aosta, Centro Funzionale Regione Autonoma Valle d’Aosta: data are available free of charge on http://presidi2.regione.vda.it/str_dataview.
21. Veneto, ARPA Veneto: data freely available on <https://www.arpa.veneto.it/bollettini/storico/precmax/>.

References

1. Trigila, A.; Iadanza, C.; Bussetini, M.; Lastoria, B. Dissesto idrogeologico in Italia: Pericolosità e indicatori di rischio—Edizione 2018. *ISPRA Rapp.* **2018**, *287*, 31–32.
2. Libertino, A.; Ganora, D.; Claps, P. Technical note: Space-time analysis of rainfall extremes in Italy: Clues from a reconciled dataset. *Hydrol. Earth Syst. Sci.* **2018**, *22*, 2705–2715. [[CrossRef](#)]
3. Cavicchia, L.; von Storch, H.; Gualdi, S. A long-term climatology of medicanes. *Clim. Dyn.* **2014**, *43*, 1183–1195. [[CrossRef](#)]
4. Greco, A.; De Luca, D.L.; Avolio, E. Heavy precipitation systems in Calabria Region (Southern Italy): High-resolution observed rainfall and large-scale atmospheric pattern analysis. *Water* **2020**, *12*, 1468. [[CrossRef](#)]

5. Pelosi, A.; Furcolo, P.; Rossi, F.; Villani, P. The characterization of extraordinary extreme events (EEEs) for the assessment of design rainfall depths with high return periods. *Hydrol. Process.* **2020**, *34*, 2543–2559. [[CrossRef](#)]
6. CUBIST. Available online: <http://www.cubist.polito.it> (accessed on 4 October 2020).
7. Hydrological Yearbooks. Available online: <http://www.acq.isprambiente.it/annalipdf/> (accessed on 4 October 2020).
8. Rossi, F.; Villani, P. *Valutazione Delle Piene in Campania*; CNR-GNDICI: Salerno, Italy, 1994.
9. Data on Maximum Annual Rainfall in North-Eastern Italy. Available online: <http://www.irpi.cnr.it/en/product-service/dati-sulle-piogge-massime-annuali-nellitalia-nord-orientale/> (accessed on 4 October 2020).
10. Pavan, V.; Antolini, G.; Barbiero, R.; Berni, N.; Brunier, F.; Cacciamani, C.; Cagnati, A.; Cazzuli, O.; Cicogna, A.; De Luigi, C.; et al. High resolution climate precipitation analysis for north-central Italy, 1961–2015. *Clim. Dyn.* **2019**, *52*, 3435–3453. [[CrossRef](#)]
11. Muñoz Sabater, J. ERA5-Land Hourly Data from 1981 to Present. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). Available online: <https://doi.org/10.24381/cds.e2161bac> (accessed on 2 November 2020).
12. Slivinski, L.C.; Compo, G.P.; Whitaker, J.S.; Sardeshmukh, P.D.; Giese, B.S.; McColl, C.; Allan, R.; Yin, X.; Vose, R.; Titchner, H.; et al. Towards a more reliable historical reanalysis: Improvements for version 3 of the Twentieth Century Reanalysis system. *Q. J. R. Meteorol. Soc.* **2019**, *145*, 2876–2908. [[CrossRef](#)]
13. Bonanno, R.; Lacavalla, M.; Sperati, S. A new high-resolution Meteorological Reanalysis Italian Dataset: MERIDA. *Q. J. R. Meteorol. Soc.* **2019**, *145*, 1756–1779. [[CrossRef](#)]
14. Libertino, A.; Ganora, D.; Claps, P. Evidence for increasing rainfall extremes remains elusive at large spatial scales: The case of Italy. *Geophys. Res. Lett.* **2019**, *46*, 7437–7446. [[CrossRef](#)]
15. Davolio, S.; Della Fera, S.; Laviola, S.; Miglietta, M.M.; Levizzani, V. Heavy precipitation over Italy from the Mediterranean storm “Vaia” in October 2018: Assessing the role of an atmospheric river. *Mon. Weather Rev.* **2020**, *148*, 3571–3588. [[CrossRef](#)]
16. Goovaerts, P. Using elevation to aid the geostatistical mapping of rainfall erosivity. *CATENA* **1999**, *34*, 227–242. [[CrossRef](#)]
17. Prudhomme, C.; Reed, D.W. Relationships between extreme daily precipitation and topography in a mountainous region: A case study in Scotland. *Int. J. Climatol.* **1998**, *18*, 1439–1453. [[CrossRef](#)]
18. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **2016**, *3*, 160018. [[CrossRef](#)]
19. FORCE 11—The Future of Research Communications and e-Scholarship. Available online: <https://www.force11.org/fairprinciples> (accessed on 4 October 2020).
20. EU-DEM version 1.1. Available online: <https://land.copernicus.eu/imagery-in-situ/eu-dem/eu-dem-v1.1?tab=mapview> (accessed on 4 October 2020).
21. Li, H.; Sun, J.; Zhang, H.; Zhang, J.; Jung, K.; Kim, J.; Xuan, Y.; Wang, X.; Li, F. What large sample size is sufficient for hydrologic frequency analysis?—A rational argument for a 30-year hydrologic sample size in water resources management. *Water* **2018**, *10*, 430. [[CrossRef](#)]
22. Madsen, H.; Lawrence, D.; Lang, M.; Martinkova, M.; Kjeldsen, T.R. A Review of Applied Methods in Europe for Flood-Frequency Analysis in a Changing Environment. NERC/Centre for Ecology & Hydrology, 2013. Available online: <http://nora.nerc.ac.uk/id/eprint/501751/7/N501751CR.pdf> (accessed on 23 November 2020).
23. La Superficie dei Comuni, Delle Province e Delle Regioni Italiane. Available online: <https://www.istat.it/it/files//2013/02/Superfici-dei-comuni.pdf> (accessed on 4 October 2020).
24. Burlando, P.; Rosso, R. Scaling and multiscaling models of depth–duration–frequency curves for storm precipitation. *J. Hydrol.* **1996**, *187*, 45–64. [[CrossRef](#)]
25. Svensson, C.; Jones, D. Review of rainfall frequency estimation methods. *J. Flood Risk Manag.* **2010**, *3*, 296–313. [[CrossRef](#)]
26. Crespi, A.; Brunetti, M.; Lentini, G.; Maugeri, M. 1961–1990 high-resolution monthly precipitation climatologies for Italy. *Int. J. Climatol.* **2018**, *38*, 878–895. [[CrossRef](#)]
27. Napoli, A.; Crespi, A.; Ragone, F.; Maugeri, M.; Pasquero, C. Variability of orographic enhancement of precipitation in the Alpine region. *Sci. Rep.* **2019**, *9*, 13352. [[CrossRef](#)]

28. Allamano, P.; Claps, P.; Laio, F.; Thea, C. A data-based assessment of the dependence of short-duration precipitation on elevation. *Phys. Chem. Earth Parts A/B/C* **2019**, *34*, 635–641. [[CrossRef](#)]
29. Avanzi, F.; De Michele, C.; Gabriele, S.; Ghezzi, A.; Rosso, R. Orographic signature on extreme precipitation of short durations. *J. Hydrometeor.* **2015**, *16*, 278–294. [[CrossRef](#)]
30. Bartolini, E.; Claps, P.; Laio, F. Analisi Della Variabilità Spaziale di Valori Caratteristici Delle Precipitazioni in Piemonte. Working-Paper 2008-01, Dipartimento di Idraulica, Trasporti ed Infrastrutture Civili, Politecnico di Torino, 2008. Available online: <http://www.idrologia.polito.it/~claps/Papers/WP2008-01.pdf> (accessed on 4 October 2020).
31. Borga, M.; Vezzani, C.; Fontana, G.D. Regional rainfall Depth–Duration–Frequency equations for an Alpine region. *Nat. Hazards* **2005**, *36*, 221–235. [[CrossRef](#)]
32. Forestieri, A.; Lo Conti, F.; Blenkinsop, S.; Cannarozzo, M.; Fowler, H.J.; Noto, L.V. Regional frequency analysis of extreme rainfall in Sicily (Italy). *Int. J. Climatol.* **2018**, *38*, e698–e716. [[CrossRef](#)]
33. Libertino, A.; Macchia, S.; Claps, P. Nubifragi eccezionali in Italia: Analisi preliminare di rilevanza sugli eventi 1920–2000. In Proceedings of the XXXV Convegno Nazionale di Idraulica e Costruzioni Idrauliche, Bologna, Italy, 14–16 September 2016.
34. Kottegoda, N.T.; Rosso, R. *Applied Statistics for Civil and Environmental Engineers*, 2nd ed.; Wiley-Blackwell: Oxford, UK, 2008; pp. 406–415.
35. Adler, A. Strong laws for ratios of order statistics from exponentials. *Bull. Inst. Math. Acad. Sin.* **2015**, *10*, 101–111.
36. Hengl, T. A Practical Guide to Geostatistical Mapping of Environmental Variables. Available online: <http://publications.jrc.ec.europa.eu/repository/bitstream/JRC38153/eur22904en.pdf> (accessed on 2 November 2020).
37. Rolo, R.M.; Radtke, R.; Costa, J.F.C.L. Signed distance function implicit geologic modeling. *REM Int. Eng. J.* **2017**, *70*, 221–229. [[CrossRef](#)]
38. Fiori, E.; Comellas, A.; Molini, L.; Rebora, N.; Siccardi, F.; Gochis, D.J.; Tanelli, S.; Parodi, A. Analysis and hindcast simulations of an extreme rainfall event in the Mediterranean area: The Genoa 2011 case. *Atmos. Res.* **2014**, *138*, 13–29. [[CrossRef](#)]

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).