

Interactive comment on “Technical note: Space-time analysis of rainfall extremes in Italy: clues from a reconciled dataset” by Andrea Libertino et al.

Andrea Libertino et al.

andrea.libertino@polito.it

Received and published: 4 February 2018

We greatly appreciate the insightful comments from the reviewer. The comments from the reviewer have been reproduced in italic below, interspersed with our responses.

The manuscript presents the Italian Rainfall Extreme Dataset (I-RED) of annual maxima values for 1-3-6-12 and 24 hours duration together with an exploratory statistical analysis of Italian annual maxima. The authors report in detail: (i) the efforts needed to collect the datasets from the complex universe of Italian institutions that are responsible for managing the measurement network; (ii) the techniques used and the choices made to merge the datasets into I-RED. Standard statistical techniques have been

C1

used for the statistical analysis and, as far as I can judge, without major flaws. The presentation of the manuscript is clear and concise. The authors credit the sources of the data (even for those 12 Institutions out of 21 which have not yet provided data, see Table 1, “Under request” under the column “Digitized data availability”).

We need to clarify a misunderstanding on this point. In Table 1, the “Under request” category is related to those agencies that do not provide data directly on their websites. Most of them have already provided their datasets (or a part of them) upon request. As it is not clear, the Table will be edited in the revised manuscript under the category “available upon request”.

The potential of I-RED for both research and risk management/reduction is clearly evident and it is a pity that it cannot be made freely available because of the data policies of some of the institutions involved (as written in the paragraph “Data availability”). There are some questions I was left with after having gone through the manuscript about I-RED data policy and availability. In my opinion, this is the main limitation of your technical note. A little more detail on this major concern is below. My advice to the editor is to accept the manuscript for publication with only minor revisions.

Major Comments:

1. Data policy 1. The I-RED data policy is not clearly specified while it must be clearly presented to the reader both in the main text and briefly mentioned in the abstract too. The authors’ work is valuable for the research community and the society whether I-RED could be made totally public or not. However, I believe most readers would be interested in knowing if I-RED is accessible in total or only in part. For instance, could you add a column to Table 1 to make it clear if you are allowed to re-distribute the data through I-RED? Is there any website where the reader could access the public part of I-RED?

We thank the reviewer for the suggestion. Data availability is one of our main concerns; that is why we have reported in the caption of Table 1 all the websites where the reader

C2

can find the original datasets. In the revised version of the manuscript we plan to add information on the procedure to access the data that we have merged and harmonized in the *I-RED*. In essence, as the data owners want to supervise their use, we only have permission to use them for purposes connected to our project. Consequently, we can provide the full database access only to research individuals or groups who join our project. Nevertheless, some individual regional databases can be provided upon evidence of permission received by the regional agencies, in particular, those that release data upon request.

2. Data policy 2. You made clear in the manuscript that I-RED is related to CUBIST, which can be accessed on the polito website and as you write it is "first important attempt of making the large Italian hydrological heritage freely available in computer-readable format" (page 2, line 19). My question is: why CUBIST data is freely available and I-RED data has restricted access?

This is a drawback of the change in the owner of the network that we have described in the manuscript. The data included in the *CUBIST* database are the digitized version of the data included in the Hydrological Yearbooks of the *SIN-SIMN*, that are freely available in a PDF format under the Hydrological Yearbooks Project of the Italian National Institute for Environmental Protection and Research (*ISPRA*) (<http://www.isprambiente.gov.it/progetti/acque-interne-e-marino-costiere-1/progetto-annali>). As a digital version of public data, they are public too. After the *SIMN* has been dismantled, every local authority has adopted a different policy for data distribution. This further clarifies the difficulties we have in making the *I-RED* dataset fully open. Moreover, the data are often provided only upon request, because the regional authorities want to supervise the spreading and the correct use of them, as stated in answer 1.

3. Data policy 3. How could you reconcile these two statement: (1) page 2 line 27 "...the Italian law adopted an Open Source policy for the public data..." and (2) your statement on "data availability" where you write that you have signed agreements with

C3

the data provider (which are all public institutions) that restrict the use of the data to the aims of your project? In my view, if you get the dataset only after signing an agreement that limit the use of a dataset then the dataset is not openly available (by definition).

The reviewer is right: the two sentences seem to be incompatible, but the situation is the result of the evolving framework of the national and regional laws concerning the "open data policy". The Italian law concerning the open data refers mainly to the D.Lgs.82/2005, D.Lgs. 36/2006, Law 221/2012, D.Lgs. 179/2012, Law 114/2014. Substantially it requires that all the data of public interest collected from public authorities should be made freely available on the internet for non-commercial use. The complex framework of the regional hydrological agencies actually comply to the law in a varied fashion. Some of the regional agencies started immediately to provide the data on the internet as they become available. Other agencies, to make sure of the non-commercial use of the data, provide them only upon request, after a certification of the purposes for which the data would be used. Other regions are still developing the platform for the digitalization and the distribution of the data, and their local archive are still not officially ready for the publication. Moreover, in some regions agencies became full operational after the dismissal of the *SIMN* only years later: in this time lag data may have been collected in a non-systematic way from authorities different from the one which finally came in charge of the service. This involves further complexity in the definition of who and how should work on patching and publishing the data. Considering the global complexity of the topic, supplementary material including the official data policies of the different authorities will be added.

4. Conclusions and Future plans. Page 10, line 14. "The final aim is to make the update of the database systematic and unsupervised." A few words on how the authors plan to achieve the goal would do here. In particular, if the author could present any link with national and/or international project/activities this would strengthen their statement. For example, I'm aware of initiatives aimed at collecting Italian datasets such as ArCIS (<https://www.arcis.it/wp/en/home-2/>) or ISPRA-SCIA (<http://www.scia.isprambiente.it>).

C4

Do the authors have any contact with them? In particular, it seems to me that ISPRA-SCIA is doing a work on the establishment of a national database that is very similar to I-RED, though mostly for aggregated values. Do you plan to join your efforts with other institutions? Do you have any contact with international institutions? There are several ongoing projects at an international level to collect and organize in-situ observations (See for example COPERNICUS: <https://climate.copernicus.eu/global-land-and-marine-observations-database>)

The reviewer is right. There are other significant projects in the Italian framework that will be mentioned in the revised version of the manuscript. In particular, the *ISPRA-SCIA* dataset pursues the same aims of this work, but just relative to the daily precipitation. The *ISPRA-SCIA* dataset is a significant source of data, considered also in the development of the *CUBIST* database and a mention to the project will be added in the revised version of the manuscript. In general, the projects the reviewer refers to (as most of the project related to hydrology in Italy) have been developed from the same agencies that collaborated with us for the development of the *I-RED*, as partners of *ISPRA*. We plan to contact again all the agencies for a feedback on the work that we have done, in order to strength the collaborations and integrate our work in a national framework. We are also looking forward to join to international institutions in starting EU-wide projects aimed at maximizing the outcomes of this work.

Minor Comments:

4. Figure 2. Panels b and c shows exactly the same histogram. Please, check it.

The distribution of the length classes is similar across the different durations. As it is not possible nor useful to show such behavior, figure 2 will be edited in the revised manuscript.

5. Figure 4. Excellent Figure. Page 8, lines 7-8 (“When short...appears.”). I agree with your statement, nonetheless this could also be an artifact due to the color scale chosen. Is there any particular reason for your choices of min/max values for color

C5

scale of the different durations? Does the min/max values correspond to any percentile of the distribution of values, for example?

The reviewer is right, the apparent distribution of the median values could be also an artifact of the chosen color scale (the min/max values correspond to the min/max values of the distributions). Nonetheless, the coefficient of variation of the nationwide distribution of the medians grows when larger durations are considered, and this seems to confirm our hypothesis on more “peaky” distribution of larger durations values. These considerations are included in the revised manuscript.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2017-752>, 2018.

C6