Reply to the Editor

Dear authors,

many thanks for submitting the revised version of the manuscript on flood hazard mapping in Europe. The paper has now been seen by two referees. While referee #1 has now further comments, the new referee #3 did raise a number of important points which have to considered. In particular the questions on model validation, the reliability of extrapolations to 500 and 1000 year events and the choice of the Gumble over the GEV distribution are highly relevant since they affect the overall dependability of the data product.

Please find below a detailed reply to all the points raised by Referee #3.

Reply to Referee #3

General remarks

The paper is clearly structured and has generally a good reading flow. The dataset presented improves the spatial coverage of existing flood hazard maps. However, the paper lacks methodological detail not evident to readers unfamiliar with previous work published by the authors. In particular, important information on hydrological model evaluation with respect to high flows and the statistical models (i.e. Gumbel distribution and hydrograph construction procedure) is missing.

We thank the Referee for his/her positive view of our work and for the useful comments. Please find below a detailed reply to all the points raised

1) What kind of new features that the latest version of the LISFLOOD model have that make it more suitable to derive flood hazard maps than previous versions (l. 83-86)?

The new LISFLOOD version benefits from updated in the model components, in the input dataset and in the calibration routine. Specifically, we expanded the description of the new LISFLOOD version in Section 2.1 as follows (see lines 120-128): "The new version features an improved routine to calculate water infiltration, the possibility of simulating open water evaporation and several minor adjustments that correct previous code inconsistencies (Arnal et al., 2019)". In addition, most of the input datasets (e.g. meteorological data, digital elevation model etc) have been updated, and a new calibration algorithm has been developed and applied with more calibration and validation data". All these updates are likely to improve the estimation of river flow regimes and hence also the estimation of peak flows.

2) I guess that the official hazard maps were derived using locally-calibrated hydrological models and are therefore considered to be more reliable than maps

derived using global models (and are therefore chosen as a reference). This is not evident though and should be mentioned somewhere (l. 89-91).

We thank the Reviewer for pointing out this issue. In the revised manuscript, we have modified Section 2.3.1 to motivate the use of official flood hazard maps in the validation (lines 245-255): "In Europe, all member states of the European Union as well as the United Kingdom have developed national datasets of flood hazard maps for a range of flood probabilities (usually expressed with the flood return period), following the guidelines of the EU Floods Directive (EC 2007). These maps are usually derived using multiple hydrodynamic models of varying complexity (AdB Po 2012) based on high-resolution topographic and hydrological datasets, such as DEMs of at least 5 m resolution in England (Sampson et al., 2015), LIDAR elevation data in Spain (MITECO 2011), and river sections based on LIDAR surveys in the Po River basin (AdB Po, 2012). Even though official maps might be prone to errors or be incomplete (Wing et al 2017), they are likely to provide a higher accuracy than the modelled maps presented here, and therefore have been selected as reference maps for the validation."

3) The authors briefly describe how the LISFLOOD model has been calibrated (l. 124-126), however, I could not find any information about how the model was validated and what the outcome of the validation step was. It would be important to provide some validation results with respect to high-flow simulation performance in order to establish trust into the streamflow simulations used for the hazard assessment.

In the revised text (Appendix B) we have included a more detailed description of the joint calibration/validation done by Arnal et al. (2019). In particular, we added an overview of the calibration/validation results using the Nash – Sutcliffe efficiency index (NSE), and a new table summarizing the model skill. The NSE index is more suited to evaluate the model accuracy in simulating discharge peaks and complements the Kling-Gupta efficiency index.

4) The authors estimate 500- and 1000-year floods using a sample consisting of 26 annual maxima only (l. 151). Such extrapolations are extremely dangerous because of large sampling uncertainty. I would therefore limit the analysis to extremes with 100-year return period.

We agree with the Referee on that using short time series bring substantial uncertainty in extrapolating peak flow extremes. However, the comparison with the previous version of the flood maps described in Section 3.4 shown that much of the uncertainty in peak flow estimation is actually smoothed out by the low sensitivity of flood extent and depth to return period, in particular for return periods above 100 years. Such low sensitivity was observed by Dottori et al. (2016) and by Trigg et al (2016) for a global-scale application of the same flood hazard mapping procedure, and derives from the relatively low accuracy of the available topographic information, in particular the absence of river channels and structures *such as river embankments. We have updated Section 3.4 to include this explanation, which is also mentioned in the Conclusions (lines 671-674).*

5) The Gumbel distribution is chosen for extreme value analysis. I have my doubts that this 2-parameter distribution is a good fit for the data. Goodness-of-fit testing is required here to show the suitability of the Gumbel distribution to model annual maxima. If it is rejected in many cases, I would rather use the more flexible GEV distribution.

In the revised version Section 2.2. we have added some paragraphs to motivate the choice of the Gumbel distribution: "We used the Gumbel distribution to keep a parsimonious parameterization (2 parameters instead of 3 of the generalized extreme value (GEV), log-normal and other distributions) and thus avoid over-parameterization when extracting high return period maps from a relatively short time series. The same distribution was also adopted for the extreme value analysis in previous studies regarding flood frequency and hazard (Alfieri et al., 2014, 2015; Dottori et al., 2016)." In addition, we mentioned the uncertainty arising from the extreme value analysis (distribution used for extreme value fitting, length of time series etc) in the discussion of results in Section 3 (lines 387-395)

6) I do not fully understand how the design flood estimates were derived (l.156-165). The description of how event duration is included needs more/clearer explanation because the FDC itself is only a CDF of daily flow and does not really say anything about event duration:How is event duration derived? How is event volume derived? How is the event shape derived?

We have carefully rewritten the last part of Section 2.1 to better explain how we derive synthetic flood hydrographs **: "***The synthetic flood hydrographs are derived using daily time steps. The peak value of the hydrograph is given by the peak discharge for the selected T-year return period QT, while the other values Qi are derived multplying QT by the ratio εi. The hydrograph peak QT is placed in the centre of the hydrograph, while the other values Qi are sorted alternatively to produce a triangular hydrograph shape, as shown in Figure 2. The total duration of the synthetic hydrograph is given by the local value of the time of concentration Tc, therefore all the durations > Tc are discarded from the final hydrograph (Figure 2). Because river channels are usually not represented in continental scale topography, flood hydrograph values are reduced by subtracting the 2-years discharge peak QT(2), which is commonly considered representative of river bank-full conditions (note that the original DEM is not modified with this procedure). Hence, the overall volume of the flood hydrograph is given by the sum of all daily flow values with duration < Tc."*

7) I find it a inconsistent to compare existing flood risk maps which have been derived taking flood protection measures into account to estimated flood risk maps derived ignoring these measures (l.365-369). This seems as if you were comparing apples

with pears instead of apples with apples. Wouldn't a flood risk map not considering protection measures depict a wrong (and overestimated) picture of flood risk?

We fully agree on that the modelled flood hazard maps should be compared with reference maps not accounting for flood protections. To better explain our approach we rewrote part of Section 3.1.2 (lines 269-277): "The modelled maps does not include the effect of protections, as mentioned in Section 2.3. Wherever possible, for the comparison exercise we selected either reference flood maps that do not account for protections (e.g. Hungary) or maps for flood return periods exceeding local protection standards, assuming that the resulting flood extent is little conditioned by flood defences. For instance, the main stem of the Po river is protected against the 1-in-200-year flood events (Wing et al., 2019), whereas protection standards in England and Norway are usually above 20 years (Scussolini et al., 2016). Reference maps where the extent and design level of protection is not known (e.g. Spain) have been also included in the comparison to increase the number of validation areas." Note that in this latter case, we explicitly mention in Section 3.2.3 that the influence of protections can condition the outcomes of the comparison.

8) Overall, a more nuanced discussion of different uncertainty sources not limited to uncertainties related to the hydraulic modeling step would be required. Such uncertainty sources include hydrological model performance, statistical modeling, design hydrograph estimation, …

In the revised version, Section 3 we now mention different uncertainty sources related to the elaboration of the hydrological input (lines 387-395): "Differences between simulated and reference hydrological input are likely to influence the skill of modelled flood maps and may depend on several factors such as the hydrological model performance for peak flows, extreme value analysis (distribution used for extreme value fitting, length of available time series) and design hydrograph estimation. However, further analyses are difficult because we have no specific information on the hydrological input used for the reference flood maps (e.g. peak flows, statistical modelling of extremes, hydrograph shape). In the following sections, we use the skill of the LISFLOOD long-term simulation to evaluate the agreement between modelled and observed hydrological regime, but this does not necessarily translate to extreme values."

As we state in the revised version, a more in-depth discussion of the hydrological uncertainty is hampered by the lack of information about reference flood maps. This means that we cannot quantify the differences between modelled and reference maps on crucial aspects such as the statistical modelling of extremes, input peak flows, the shape of hydrographs etc. However, we do provide an overview of LISFLOOD performance in all the subsections dedicated to the study areas, and where possible we discuss the possible influence of hydrological input on the skill scores (for instance, in Section 3.2.3 we mention that that modelled flow peaks for low-probability flood events are more uncertain).

9) Some additional language editing would further improve the reading flow.

The paper has been carefully edited to improve the language.

Minor points

L. 109: when was LISFLOOD last updated?

The LISFLOOD version used to run the hydrological simulations and documented in Arnal et al. (2019) was finalized in the second half of 2018. The code of the model version presently available as source code in GitHub (see Data Availability) is basically unchanged besides bug fixing and the possibility to use 6-hourly time step. To avoid confusion, we change all references to the LISFLOOD version from "the latest version" to "updated version" (in respect to the version used by Dottori et al., 2016).

L. 118-120: can you please provide the data sources for all these datasets?

We modified the text to state that all *the data sets are described in Arnal et al (2019)*

L. 156: what do you mean by 'long-term' simulation?

We refer here to the LISFLOOD long-term simulation described in Section 2.1. In the revised paper we replaced it with the term "streamflow dataset" and we specify at the beginning of Section 2.1 that it is derived from LISFLOOD long-term simulation.

Figure B1: color legend is missing.

The colour legend has been added to the Figure.

Suggestion for slight title adjustment: 'A new dataset of river flood hazard maps for Europe and the Mediterranean Basin' or 'River flood hazard maps for Europe and the Mediterranean Basin: a new dataset derived using LISFLOOD'

We thank the Referee for the suggestion and we modified the title to "A new dataset of river flood hazard maps for Europe and the Mediterranean Basin". We prefer not to include the LISFLOOD model in the title, because it is not the only model applied (flood simulations are run using the hydrodynamic model LISFLOOD-FP) and because the overall procedure include several steps not directly related to LISFLOOD.

References

Alfieri, L., Salamon, P., Bianchi, A., Neal, J., Bates, P.D., Feyen, L., 2014. Advances in pan-European flood hazard mapping, Hydrol. Process.,28 (18), 4928-4937, doi:10.1002/hyp.9947.

Alfieri, L., Burek, P., Feyen, L., and Forzieri, G.: Global warming increases the frequency of river floods in Europe, Hydrol. Earth Syst. Sci., 19, 2247–2260, https://doi.org/10.5194/hess-19-2247-2015, 2015.

Arnal, L. et al., 2019. EFAS upgrade for the extended model domain – technical documentation, EUR 29323 EN, Publications Office of the European Union, Luxembourg, 2019, doi: 10.2760/806324, JRC111610

Autorita` di bacino del fiume Po (AdB Po): Progetto di Variante al PAI: mappe della pericolosita` e del rischio di alluvione (in Italian), https://pianoalluvioni.adbpo.it/progetto-esecutivodelleattivita/, accessed on 2020-04-03, 2012.

Dottori, F., Salamon, P., Bianchi, A., Alfieri, L., Hirpa, F.A., Feyen, L., 2016. Development and evaluation of a framework for global flood hazard mapping. Advances in Water Resources 94, 87–102.

Knoben, W. J. M., Freer, J. E., and Woods, R. A.: Technical note: Inherent benchmark or not? Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores, Hydrol. Earth Syst. Sci., 23, 4323–4331, https://doi.org/10.5194/hess-23-4323-2019, 2019.

Krause, P., Boyle, D. P., Bäse, F. Comparison of different efficiency criteria for hydrological model assessment. Advances in Geosciences, European Geosciences Union, 2005, 5, pp.89-97.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I – A discussion of principles, J. Hydrol., 10, 282–290, https://doi.org/10.1016/0022-1694(70)90255-6, 1970.

Trigg, M. et al., 2016. The credibility challenge for global fluvial flood risk analysis. Environ. Res. Lett.11 094014