

Reply to Reviewers

RC1: 'Comment on essd-2020-313', Anonymous Referee #1, 09 Apr 2021

This paper evaluates new European flood maps provided by the Copernicus European Flood Awareness System. The methods behind the mapping have been developed and published elsewhere and are only briefly documented here (too briefly in a few places – see comments below). The advantage of the presentation approach is that the paper should be accessible to non-experts in this type of modelling.

The paper focuses on the performance of the hazard layers against several national flood hazard maps and international studies using a similar class of regional flood models. The paper is well presented and easy to follow. It makes a useful contribution to the literature and more validation studies of this type are needed. Generally the conclusions are well supported by the analysis, although the early focus on the Mediterranean basin region is lost later in the manuscript. I agree fully with the premise of the paper and would recommend publication subject to the revisions outlined below.

We thank the Referee for his/her positive comments on our manuscript.

Line 81: Add reference for LISFLOOD-FP for consistency with LISFLOOD. There is also now source code published for LISFLOOD-FP and you could cite this later for consistency with the presentation of LISFLOOD <https://doi.org/10.5194/gmd-2020-340>

We thank the Reviewer for providing this reference, which is now included in the main text as well as in the Data Availability section

Line 140: I appreciate that the article is trying to avoid repeating technical details published elsewhere. However, I would like a little more detail on the statistical analysis of the extreme flows to be presented. Specifically what data were used (AMAX)? What distributions were fitted? Alfieri et al 2014 describe the overall method but I also think so direct citation to the extreme value analysis method followed would be useful.

We applied the statistical analysis of extreme flows over the long-term hydrological simulation generated with LISFLOOD. For each pixel of the river network, we extracted annual maxima for the period 1990-2016 and we used the L-Moments approach to fit a Gumbel distribution and calculate extreme flow return periods. This explanation has been placed at the beginning of the new section 2.2 “Hydrological input of flood simulations”.

Line 146: What is the source of the high-resolution river network data?

The high-resolution river network data is taken from the CCM River and Catchment Database for Europe, the same database of the DEM (Vogt et al., 2007). We have added this reference at the beginning of Section 2.3.

Line 157: Does the DEM include building and vegetation or are these removed to approximate a DTM?

Thanks for raising this point. The elevation values of CCM DEM are influenced by buildings and vegetation cover, as already mentioned in Section 2.4.2. . This motivated our quick correction of elevation values using land use information. In the revised text we have rewritten Section 2.4.2 to clarify the importance of these source of uncertainty on DEM values.

Line 161: Given that a 2D model is used without the river channels how are river flows accounted for? For example, is a design flow subtracted from the volume entering the model (this approach would approximate the method JBA used for the original 1 in 1000 year extreme hazard map for the UK), is zero in channel flow assumed (this would potentially put more water onto the floodplain than in reality and results in a somewhat precautionary model), are the channels represented as 1D components (for example the approach taken in LFPtools <https://doi.org/10.1016/j.envsoft.2019.104561>), or do you burn a channel into the DEM (tricky at 100 m resolution on smaller rivers and usually only used for high-resolution simulations). I don't believe it matters which approach was taken from a publication perspective, but a short note is needed here to acknowledge and justify the choice made, especially as there are examples of LISFLOOD-FP being applied in all of these ways.

In each flood simulation we modify the input flow hydrograph by subtracting the 1-in-2-year flood discharge value to all daily time step values, therefore reducing the overall flood volume entering the model. Conversely, the original DEM is not modified, following the approach proposed by Alfieri et al. (2014,2015). We have added this piece of information in the expanded description of input hydrographs (see end of updated Section 2.2).

Line 243: How are the reference maps treated in the comparison. Do they maintain some native resolution or are the polygons rasterised to the same 100 m resolution as the modelled maps. Assuming 100 m resolution what impact does this have, I assume any loss of resolution usually makes the reference maps easier to fit?

All the national flood hazard maps used in this work are available as polygons of flood extent, and the original resolution is not known. As such, we could not quantify the impact of coarsening reference flood maps, however it is likely to be negligible compared to the overall error in modelled flood maps. Reference maps were then converted to raster format with the same resolution as the modelled maps, while the latter have been converted to binary flood extent maps (as explained in the original manuscript).

Line 272: Did you consider using a DEM derived from TanDEM-X data?

We are aware of the recently released 90m DEM derived from TanDEM-X data (<https://geoservice.dlr.de/web/dataguide/tdm90/>), but it was not considered in this work due to time constraints. We have mentioned the possibility of using this dataset for future research in the Conclusions, along with Copernicus DEM and MERIT-Hydro datasets.

Line 287: Some acknowledgement that the approach taken to the vegetation and urban correction is far from the state of the art is needed here. Some of the newer machine learning based approaches are likely to do a significantly better job of removing surface

artifacts than this approach and we know that flood simulation is very sensitive to the quality of vegetation and building removal in the global DEM's. I do not think this detracts from the value of the study, but it should be clear that there are known routes to potentially better modelling here.

We take the point. Indeed, working on the improvement of elevation data was not the main scope of this work (and the MERIT DEM applied in the tests already includes a correction of vegetation effects), so we modified Section 2.4.2 to make clear that we proposed a rather basic and quick approach. We agree on that mentioning more advanced research on this topic would benefit the manuscript. To this end, we have included the following text to Section 2.4.2 :“ Recent research works proposed advanced techniques to remove surface artifacts, based on artificial neural networks (Wendi et al., 2016, Kulp and Strauss, 2018) or other machine learning methods (Liu et al., 2018; Meadows and Wilson, 2021). Most approaches compare the DEM to be corrected with higher-accuracy datasets, using auxiliary data such as tree density and height for correcting vegetation bias (Yamazaki et al., 2017), whereas elevation bias in urban areas can be corrected using night light, population density, or Open Street Map elevation data (Liu et al., 2018)”.

Line 309: depending on the approach taken to represent the river channel network, see comment above, this might also be significant for the simulation of higher probability floods. Channels are very important flow pathways and especially so for smaller floods.

We fully agree with the Reviewer on this point. We have included his/her remark in the revision of Section 3.1 as follows: “Considering that most of the reference flood maps include the effect of flood defences (contrary to the modelled maps), these results suggest that the majority of rivers in the study areas may be protected for flood return periods around 100 years or lower, as indeed reported by available flood defence databases (Scussolini et al., 2016). High-probability floods are also sensitive to the method used to reproduce river channels, and the simplified approach used in this study might underestimate the conveyance capacity of channels (see Section 3.2.2 for an example). Finally, the better performance for low-probability floods may depend on floodplain morphology, where valley sides create a morphological limit to flood extent.”

General: How did you deal with coastal areas in the England flood maps? These are 1 in 200 year return periods in the flood map but also this source is not included in your modelling.

We used a 5km buffer around modelled flood maps to mask out coastal areas far from rivers estuaries, because the geo datasets available did not allow separating areas prone to coastal or river flooding. Moreover, the estuary area of several rivers show concurrent fluvial-tidal flooding, as mentioned for the Thames by Sampson et al. (see following comments), and this interaction is again not simulated in our modelling framework. We have mentioned the fact that reference maps for England include areas prone to 1-in-200-year coastal flood events (revised Section 2.3.2), and we further clarify this point in the description of results for England (section 3.2.1)

Line 330: Thames will have significant tidal flooding from London eastwards.

Thanks for this valuable insight. We used it to improve the comment of results for England (section 3.2.1) as follows: “there’s not a clear correlation between hydrological and flood map skill, with some basins (e.g. Thames) showing high KGE values but relatively low CSI values. For the Thames basin, the low CSI value is likely influenced by tidal flooding component from London eastwards. According to Sampson et al. (2015), the official flood hazard map assumes a 1 in 200 year coastal flood along with failure of the Thames tidal barrier, whereas river flood simulations use the mean sea level as boundary condition and do include storm surge and tidal flooding. Given that other river estuaries are characterized by concurrent fluvial-tidal flooding processes, this might reduce the skill of the modelled maps.”

Line 339: The smaller tributaries, and coastal flooding issues is discussed for the Thames and Severn in Sampson et al 2015. I think that would be a better comparison/citation specifically in this section than Wing et al 2017. Their CSI values from the Sampson global flood model might also be useful to report for these basins and compare with your values to complement

Line 345: see comments above.

We thank the Reviewer for the useful suggestion. We included the considerations by Sampson et al. about the influence of smaller tributaries, coastal flooding and urban areas to expand the discussion of Section 3.2.1. Moreover, we added a comparison of performance metrics between our maps and Sampson et al. in Section 3.3.

Figure 4: Could you include floodplains outside of the 5 km buffer in another colour? TBH this map doesn’t really reflect how much flooding in the UK is not being simulated by this modelling setup – which is absolutely fine but the paper should be upfront about it.

We have modified Figure 4 to include flood-prone areas outside the 5km buffer. Moreover, we calculated that flood-prone areas inside the 5km buffer correspond to 73% of the total extent for the 1-in-100-year flood map (as specified now in Section 2.3.2 of the revised manuscript).

The EA flood map doesn’t include surface water flooding from pluvial flooding, that would be an even more detailed layer, so the flooding missed is fluvial and coastal.

We now mention in Section 2.3.1 that areas prone to pluvial flooding are missing from our analysis in all study areas.

Line 410: this is an unfair comment given the publication date, but there is an updated US validation in Bates et al 2021 WRR. I don’t think this would have any significant impact on the discussion here but it might be worth citing.

We thank the Referee for the suggestion. We have included this recent work in the discussion of Section 3.3.

I’ve no experience with the flood maps outside of the UK but the comparisons undertaken look robust.

We thank the Referee for the appreciation

The conclusions are well supported by the analysis, however little validation has been undertaken around Mediterranean basins, particularly those areas into which the new maps have extended. Flood simulation in arid areas are often more challenging and the performance from Europe might not translate well to North Africa and the Eastern Mediterranean. I think some discussion of this issue is needed given the focus on the Mediterranean basin region in the title and introduction... Or perhaps less focus on the Mediterranean basin region and more on Europe earlier in the manuscript if the discussion is going to be too vague in this regard.

In the revised text we now highlight the issue of incomplete validation of the dataset. For instance, in the Abstract we state that "Further verification in North African and Eastern Mediterranean regions is needed to better understand the performance of the flood maps in arid areas outside Europe" while the end of the Introduction we state that "The number and extent of the validation sites allows for a more detailed evaluation in respect to previous efforts by Alfieri et al. (2014) and Paprotny et al. (2017), even though none of the validation sites is located outside Europe due the unavailability of national flood maps". In the conclusions, we say that "(...) the unavailability of reference flood maps outside Europe did not allow any validation in the arid regions in North Africa and Eastern Mediterranean. In these areas, further research will be needed to better understand the performance of the flood mapping procedure here proposed."

However, we believe that mentioning the Mediterranean basin region in the title and introduction is justified and constitutes a major point of the work, because the dataset is the first example of high-resolution (100m) and freely available flood hazard maps available for the whole Mediterranean basin region, including North Africa and the Eastern Mediterranean and.

Additional References

Bates, P. D., Quinn, N., Sampson, C., Smith, A., Wing, O., Sosa, J., et al. (2021). Combined modeling of US fluvial, pluvial, and coastal flood hazard under current and future climates. *Water Resources Research*, 57, e2020WR028673. <https://doi.org/10.1029/2020WR028673>

Kulp, S.A.; Strauss, B.H. CoastalDEM: A global coastal digital elevation model improved from SRTM using a neural network. *Remote Sens. Environ.* 2018, 206, 231–239.

Liu, Y., Bates, P.D., Neal, J.C. and Yamazaki, D., 2019, December. Bare-earth DEM Generation in Urban Areas Based on a Machine Learning Method. In *AGU Fall Meeting Abstracts (Vol. 2019, pp. H41N-1899)*.

Meadows, M.; Wilson, M. A Comparison of Machine Learning Approaches to Improve Free Topography Data for Flood Modelling. *Remote Sens.* 2021, 13, 275. <https://doi.org/10.3390/rs13020275>

Wendi, D.; Liang, S.-Y.; Sun, Y.; Doan, C.D. An innovative approach to improve SRTM DEM using multispectral imagery and artificial neural network. *J. Adv. Model. Earth Syst.* 2016, 8, 691–702.

RC1: 'Comment on essd-2020-313', Anonymous Referee #1, 09 Apr 2021

The authors present a dataset of flood hazard maps, updated with respect to the ones presented in Alfieri et al. (2014) and Dottori et al. (2016). Both in the abstract and in the introduction, it is not clear what are the improvements and the differences between the new and the old version of the dataset. Later in the manuscript some information are provided but this point is not clearly discussed. Please add the needed details.

Following the Referee's comment, we have highlighted in the Introduction the main advances of the new dataset in respect of the previous versions (namely, the extended domain, the use of updated hydrological and hydrodynamic models, and new calibration and meteorological input data). Furthermore, we pointed out the relevance of the validation exercise, that allows for a more detailed evaluation in respect to previous efforts in literature

Many times the authors refer to previous own works. Even if I understand the reason for which some details are not given in the manuscript, some additional information e.g., about the methodology applied to obtain synthetic flood hydrograph or the meteorological data used as input to the LISFLOOD model, could be useful for the reader.

We have expanded the description of the data and methods used for this study, following similar comments from both Referees. The revised text features new sections describing the meteorological dataset driving the LISFLOOD model (Appendix A) and the synthetic flood hydrographs used in flood simulations (Section 2.2 in the main text). Moreover, we provide now more details about the extreme value analysis and the validation procedure. More detailed explanations are given in the following replies

The results discussed throughout the manuscript should be better explained. In particular, it is not clear how performance scores reported in the Tables are obtained for each study areas and it is not clear how the comparison shown in Table 6 has been carried out. For major details, please refer to specific comments.

The description of methodology applied to obtain the results has been revised and clarified. Please refer to the following point-by-point reply for more details.

The authors attribute the differences between modelled and reference flood maps to a number of shortcomings of the modelling framework specifically related to the hydrodynamic simulation. No description (magnitude of peak, duration of the hydrograph) is given about the hydrographs used as input to the hydrodynamic LISFLOOD-FP model that are (could be) different from the ones used to obtain reference flood maps. Please, add details on the hydrological inputs used and comment how they impact on the definition of the flood extension.

We have added a new section (2.2) with a detailed description of the methodology used to derive the flood hydrographs, reporting the following text: “The input hydrographs necessary for the flood simulations are derived from the LISFLOOD streamflow dataset described in Section 2.1. Streamflow data is available for the EFAS river network at 5 km grid spacing for rivers with upstream drainage areas larger than 500 km². For each pixel of the river network we selected annual maxima over the period 1990-2016 and we used the L-moments approach to fit a Gumbel distribution and calculate peak flow values for reference return periods of 10, 20, 50, 100, 200 and 500 years. Note that we also calculated the 30- and 1000-year return periods in limited parts of the model domain to allow validation against official hazard maps, see Section 2.3. Subsequently, we calculate a Flow Duration Curve (FDC) from the long-term simulation. The FDC is obtained by sorting in decreasing order all the daily discharges, thus providing annual maximum values Q_D for any duration i between 1 and 365 days. Annual maximum values are then averaged over the entire period of data, and used to calculate the ratios ϵ_i between each average maximum discharge for i -th duration $Q_{D(i)}$ and the average annual peak flow (i.e. $Q_D = 1$ day). Design flood hydrographs are derived using daily time steps. The peak value is given by the peak discharge for the selected T -year return period Q_T , while the other values Q_i are derived multiplying Q_T by the ratio ϵ_i . The hydrograph peak Q_T is always placed in the centre of the hydrograph, while the other values Q_i are sorted alternatively to produce a triangular hydrograph shape, as shown in Figure xx. Because river channels are usually not represented in the CCM DEM, flood hydrograph values are reduced by subtracting the 2-years discharge peak, which is commonly considered representative of river bank-full conditions. Note that the original DEM is not modified. The total duration of the hydrograph is given by the local value of the time of concentration T_c , therefore all the durations $> T_c$ are discarded from the final hydrograph.”

Differences between simulated and reference hydrological inputs are likely to influence the skill of modelled flood maps. 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, hydrograph shape). Along the text, we use the skill of the long-term simulation of LISFLOOD to evaluate the agreement between modelled and observed hydrological regime, but this does not necessarily translate to extreme values (See also the following comments for more details). These considerations have been included at the beginning of Section 3.1.

Moreover, it is expected that higher differences are found for basins with properties and characteristics not well described by the approximations used in the procedure. For instance, if a leveed river is simulated without considering flood defence structures, the identified maps will be different from those that can be obtained by using a detailed morphology description. Actually, the considered simplifications can modify significantly the flooding dynamics. A comment of the authors is required.

We fully agree with the Referee on this point. Indeed, the influence of not including defence structures in the simulations is already discussed in the manuscript at different points (see L 175-195, 304-308, 383-385). Also, the limitations given by the simplified representation of

river channels are discussed in Section 3.2.2. We have carefully revised the manuscript to make sure that all the limitations of the modelling framework are clearly stated.

Specific comments:

Line 19 and 164: “six different flood return 20 periods...”. Reading Line 141, seems that the analysed return periods are seven. Please, modify the manuscript where needed.

To clarify this, we have modified the paragraph and moved it in the new section 2.2. The text now reads as follows: “For each pixel of the river network we selected annual maxima over the period 1990-2016 and we used the L-moments approach to fit a Gumbel distribution and calculate peak flow values for reference return periods of 10, 20, 50, 100, 200 and 500 years. Note that we also calculated the 30- and 1000-year return periods in limited parts of the model domain to allow validation against official hazard maps, see Section 2.3.” In other words, The 30- and 1000-year return period flood maps were produced only for Hungary and England for comparison exercise while the hazard maps for the other six return periods cover the whole domain.

Lines 30-31. What does the authors mean with “large variability”?

The original text in the abstract (“In addition, the large variability of reference maps affects the correct identification of the areas for the validation, thus penalizing scores”) will be replaced by the following: “In addition, the different design of reference maps (e.g. extent of areas included) affects the correct identification of the areas for the validation, thus penalizing scores”

Lines 104-108. Please, add some details on the meteorological observations used to force the LISFLOOD model. Moreover, what does the authors mean with “the static input maps have been updated and expanded”? Please, specify.

We rephrased this part in Section 2.1 as follows: “The long-term run of LISFLOOD is driven by gridded meteorological maps, derived by interpolated meteorological observations from stations and precipitation datasets (see Appendix A for details). The meteorological dataset has been updated in respect to the dataset used by Dottori et al. (2016a) to include new stations and gridded datasets across the new EFAS domain (Arnal et al. 2019). In addition, LISFLOOD simulations require a number of static input maps such as land cover, digital elevation model, drainage network, soil parameters and parameterization of reservoirs. All the static maps have been updated to cover the whole EFAS domain depicted in Figure 1.”

Moreover, we have included a new section Appendix A describing the meteorological forcing of the LISFLOOD model, taken from the report by Arnal et al. (2019). “The long-term run of the hydrological model LISFLOOD is based on observed data from meteorological stations and precipitation datasets, which are collected and continuously expanded as part of the development work for EFAS. The meteorological variables considered are: precipitation, minimum and maximum temperature, wind speed, solar radiation and vapour pressure. The number of stations with available meteorological observations depends on the period and variable considered, with an increasing availability towards the end of the historical simulation period. As an example, for the year 2016 the number of daily observations

available ranged from ~ 8.800 for temperature to ~ 5.500 for precipitation and ~ 3.700 for vapour pressure. The input from meteorological stations is completed by a number of precipitation datasets (EURO4M-APG, INCA-Analysis Austria, ERA-Interim GPCP corrected and Carpat-Clim; for details see Arnal et al., 2019). Note that the same datasets are also used to drive the LISFLOOD calibration and to calculate the initial conditions for the EFAS forecasts. The data from meteorological stations and gridded datasets were then interpolated using the interpolation scheme SPHEREMAP to produce meteorological grids with a daily time step. The reader is referred to Arnal et al. (2019) for further details.”

Lines 140-145: Please, add details on this part to allow the reader understanding the procedure. How the statistical analysis has been carried out? how the synthetic flood hydrographs have been defined?

Following previous comments from both Referees, we have added a new section (2.2) providing more details on the statistical analysis of extremes and the derivation of flood hydrographs.

Lines 146-160: this part should be better explained and modified as it is very similar to paragraph 2.21 in Dottori et al. (2017). Please rephrase.

We rewrote this part (now in Section 2.3) as follows: “The continental-scale flood hazard maps are derived from local flood simulations run all along the river network as in Alfieri et al. (2014). We use the DEM at 100 m resolution developed for the Catchment Characterization and Modelling Database (CCM; Vogt et al., 2007) to derive a high-resolution river network at the same resolution. Along this river network we identify reference sections every 5 km along stream-wise direction, and we link each section to the closest upstream section (pixel) of the EFAS 5km river network, using an partially automated procedure to ensure a correct linkage near confluences. In this way, the hydrological variables necessary to build the flood hydrographs can be transferred from the 5km to the 100m river network. Figure 3 describes how the 5km and 100m river sections are linked using a conceptual scheme.

Then, for every 100 m river section we run flood simulations using the 2D hydrodynamic model LISFLOOD-FP (Shaw et al., 2021), to produce a local flood map for each of the six reference return periods. Simulations are based on the local inertia solver of LISFLOOD-FP developed by Bates et al. (2010), which is now available as open-source software (<https://www.seamlesswave.com/LISFLOOD8.0>). We use the CCM DEM as elevation data, the synthetic hydrographs described in Section 2.2 as hydrological input, and a mosaic of Corine Land Cover for the year 2016 (Copernicus LMS, 2017) and GlobCover for the year 2009 (Bontemps et al., 2009) to estimate the friction coefficient based on land use. Finally, the flood maps with the same return period are merged together to obtain the continental-scale flood hazard maps. The 100m river network is included as a separate map in the dataset, to delineate which water courses have been considered in the creation of the flood hazard maps.”

Lines 215-216: this sentence seems to be in contrast e.g., with the results shown in Table 3 for Spain for return period of 10 years.

This was indeed not consistent in the previous version. We modified this paragraph in Section 2.3.1, which now reads as follows: "For the comparison exercise, we selected available maps for return periods for which flood extent is likely to be less 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). Conversely, we consider the 1-in-30-year map for Hungary and the 1-in-10-year map for Spain because flood defences are either not accounted for (Hungary) or their extent and design level is not known (Spain)".

Line 223: which is the native resolution of the official reference maps? After the conversion to raster format, which is the adopted resolution to make the comparison with simulated flood maps?

The official reference maps are provided as polygons with no indication of the original resolution. According to Sampson et al. (2015), the official flood hazard maps for England are constructed using DEMs of at least 5 m resolution, therefore the resolution of reference maps should be similar. Reference flood maps for the Po basin and Spain are likely to have a similar resolution since they are based on LIDAR elevation data. All the reference maps have been converted to 100m resolution for the comparison with modelled maps. All these details are now included in the revised version (section 2.3.2).

Lines 247, 251 and 255: Please, correct the numbers of the equations.

Thanks for spotting this inconsistency, we amended the numbers of the equations

Lines 243-256: Please, add the variability range of HR, FAR and CSI and define the perfect score for each of them.

We updated the description of the indices according to the Referee's suggestion, with the following information. HR ranges from 0 to 1, with a score of 1 indicating that all wet cells in the benchmark data are wet in the model data. FAR scores range from 0 (no false alarms) to 1 (all false alarms). CSI scores range from 0 (no match between model and benchmark) to 1 (perfect match between benchmark and model).

Lines 301-301: from Table 3 does not seem that "Performances improve markedly with the increasing of return periods, with a general increase in the hit rate HR". Form Table 3, HR is quite constant with the return period. Please rephrase.

We rephrased these lines in Section 3.1 as follows: "Performances improve with the increasing of return periods due to the decrease of false alarm rate FAR, while the hit rate HR does not vary significantly."

Lines 304-310 and throughout the manuscript: the authors comment on the performances of simulated flood maps to reproduce the reference maps, stressing that differences could be ascribed to floodplain morphology, presence of flood defence structure etc.. According to me the differences between simulated and reference flood maps could be ascribed also to the hydrological input routed through the river channel. How different is the flood

hydrograph used by the authors with respect to the one used by to build the reference flood maps? Do the authors have any information on this point?

Thanks for this comment. We agree on that differences between simulated and reference hydrological inputs could explain some of the observed differences between flood maps. As specified in a previous comment, the revised manuscript now includes a specific section about the hydrological input for flood simulations. Unfortunately, we could not find the necessary data to reconstruct the input flood hydrographs of the reference flood maps (e.g. peak flows, hydrograph shape). For official flood maps in Spain and in the Po river basin, only a description of general methods applied is available online. The evaluation of the skill of the LISFLOOD model can provide some hint on the similarity of modelled and observed hydrological regime, but this does not necessarily translate to extreme values. We added these considerations at the beginning of Section 3.1.

Table 3: How are obtained the performance indices for the study areas? Are these values obtained as average values? Please, add details.

The performance indices are calculated using the total extent of the reference and modelled maps with the same return period. As such each score is a single value and it is not averaged. The corresponding text has been updated accordingly.

Line 316: Should be Table 3.

Lines 323, 354, 355 and throughout the manuscript. Please use the abbreviations for hit rate, false alarm rate and critical success index.

Line 415: should be Table 6.

We amended the manuscript as suggested

Lines 415 – 400 and Table 6: how the values in Table 6 have been obtained? Do they refer to all the five study areas? Please, specify.

We have added the following text at the beginning of Section 3.3 to clarify this point "The performance indices in Table 6 were calculated by first summing up all the reference and modelled maps for the same return period, where available. Then we calculated each index using the overall modelled and reference flood extent (e.g. the value for the 100-year maps includes reference and modelled maps for England, Spain and Norway). As such, each area is weighted according to the extent of the corresponding flood map."

Table 1-3 and Line 364: for Hungary a 30-y return period has been used from the reference maps. How these maps have been compared with the simulated maps? Reading the manuscript seems that only 10, 20, 50, 100, 200, 500 and 1000 return period have been simulated.

This was on oversight of the previous version, a 30-year return period flood maps was produced only for Hungary for comparison exercise. We modified the text (now in Section 2.2) as follows: "For each pixel of the river network we selected annual maxima over the period 1990-2016 and we used the L-moments approach to fit a Gumbel distribution and

calculate peak flow values for reference return periods of 10, 20, 50, 100, 200 and 500 years. Note that we also calculated the 30- and 1000-year return periods in limited parts of the model domain to allow validation against official hazard maps, see Section 2.3.”

Figure 3 could be removed from the manuscript and the limits of the test areas could be added in Figure 1.

We have removed Figure 3 and modified Figure 1 as suggested

Figure 2: Please, modify the number 10 and 11 in the diamond. Specifically, write them in a horizontal line.

The Figure has been amended as requested

Table 1: Please, for each country add the link where the reference flood maps can be downloaded.

We now mention in the revised text and in the caption of Table 1 that the links for downloading the maps are provided in the Data Availability Section.