

A new dataset of river flood hazard maps for Europe and the Mediterranean Basin

Francesco Dottori¹, Lorenzo Alfieri², Alessandra Bianchi³, Jon Skoien¹, Peter Salamon¹

1: European Commission, Joint Research Centre, Via E. Fermi 2749, 21027 Ispra, Italy.

2: CIMA Research Foundation, Savona, Italy

3: FINCONS SPA, Italy

Correspondence to: francesco.dottori@ec.europa.eu

Keywords: river flooding, flood hazard mapping, Europe, EFAS, Mediterranean Basin.

Abstract

In recent years, the importance of continental-scale hazard maps for riverine floods has grown. Nowadays, such maps are used for a variety of research and commercial activities, such as evaluating present and future risk scenarios and adaptation strategies, as well as supporting national and local flood risk management plans. In this paper we present a new set of high-resolution (100 metres) hazard maps for river flooding that covers most European countries, as well as all of the river basins entering the Mediterranean and Black Seas in the Caucasus, Middle East and Northern Africa countries. The new river flood hazard maps represent inundation along 329,000 km of the river network, for six different flood return periods, expanding on the datasets previously available for the region. The input river flow data for the new maps are produced by means of the hydrological model LISFLOOD using new calibration and meteorological data, while inundation simulations are performed with the hydrodynamic model LISFLOOD-FP. In addition, we present here a detailed validation exercise using official hazard maps for Hungary, Italy, Norway, Spain and the UK, which provides a more detailed evaluation of the new dataset compared with previous works in the region. We find that the modelled maps can identify on average two-thirds of reference flood extent, but they also overestimate flood-prone areas for flood probabilities below 1-in-100 years, while for return periods equal to or above 500 years the

29 maps can correctly identify more than half of flooded areas. Further verification is required in
30 North African and Eastern Mediterranean regions, in order to understand better the performance
31 of the flood maps in arid areas outside Europe. We attribute the observed skill to a number of
32 shortcomings of the modelling framework, such as the absence of flood protections and rivers
33 with upstream area below 500 km², and the limitations in representing river channels and
34 topography of lowland areas. In addition, the different designs of reference maps (e.g. extent of
35 areas included) affect the correct identification of the areas for the validation, thus penalizing the
36 scores. However, modelled maps achieve comparable results to existing large-scale flood models
37 when using similar parameters for the validation. We conclude that recently released high-
38 resolution elevation datasets, combined with reliable data of river channel geometry, may greatly
39 contribute to improving future versions of continental-scale river flood hazard maps. The new
40 high-resolution database of river flood hazard maps is available for download at
41 <http://data.europa.eu/89h/1d128b6c-a4ee-4858-9e34-6210707f3c81> (Dottori et al., 2020a).

42

43 *1) Introduction*

44 Nowadays, flood hazard maps are a basic requirement of any flood risk management strategy (EC
45 2007). Such maps provide spatial information about a number of variables (e.g. flood extent,
46 water depth, flow velocity) that are crucial to quantify flood impacts and therefore to evaluate
47 flood risk. Moreover, they can be used as a powerful communication tool, enabling the quick
48 visualization of the potential spatial impact of a river flood over an area.

49 In recent years, continental- and global-scale flood maps have grown in importance, and these
50 maps are now used for a variety of research, humanitarian and commercial activities, and as a
51 support of national and local flood management (Ward et al., 2015; Trigg et al., 2016). Global
52 flood maps are used to provide flood risk information and to support decision-making in spatial
53 and infrastructure planning, in countries where national level assessments are not available (Ward
54 et al., 2015). Moreover, continental and global hazard maps are vital for consistent quantification
55 of flood risk and for projecting the impacts of climate change (Alfieri et al., 2015; Trigg et al.,
56 2016; Dottori et al., 2018), thereby allowing for comparisons between different regions, countries
57 and river basins (Alfieri et al., 2016). Quantitative and comparable flood risk assessments are also
58 necessary to derive measurable indicators of the targets set by international agreements such as
59 the Sendai Framework for Disaster Risk Reduction (UNISDR, 2015).

60 In Europe, continental-scale flood hazard maps have been produced by Barredo et al. (2007),
61 Feyen et al. (2012), Alfieri et al. (2014), Dottori et al. (2016a) and Paprotny et al. (2017). These
62 maps have been used for a variety of studies, such as the evaluation of river flood risk under future
63 socio-economic and climate scenarios (Barredo et al., 2007; Feyen et al., 2012; Alfieri et al.,
64 2015), the evaluation of flood adaptation measures (Alfieri et al., 2016) and near real-time rapid
65 risk assessment (Dottori et al., 2017).

66 The quality of continental-scale flood maps is constantly improving, thanks to the increasing
67 accuracy of datasets and modelling tools. Wing et al., (2017) developed a dataset of flood hazard
68 maps for the conterminous United States using detailed national datasets and high-resolution
69 hydrodynamic modelling, and demonstrated that continental-scale maps can achieve an accuracy
70 similar to official national hazard maps, including maps based on accurate local-scale studies.
71 Moreover, Wing et al. used the same official hazard maps to evaluate the performance of the
72 global flood hazard model developed by Sampson et al. (2015). While the global model was less
73 accurate than the continental version, it was able to identify correctly over two-thirds of flood

74 extent. Conversely, European-scale maps have undergone limited testing against the official
75 hazard maps, due to limitations in accessing official data (Alfieri et al., 2014).
76 Here, we present a new set of flood hazard maps at 100 metres resolution (Dottori et al., 2020a),
77 developed as a component of the Copernicus European Flood Awareness System (EFAS,
78 www.efas.eu). The new dataset builds upon the map catalogue developed by Dottori et al (2016a),
79 and features several improvements. The geographical extent of the new maps has been expanded
80 to include all geographical Europe (with the exclusion of the Volga river basin), the rivers entering
81 the Mediterranean Sea and the Black Sea (with the partial inclusion of the Nile river basin), plus
82 Turkey, Syria and the Caucasus region. To the best of our knowledge, these are the first flood
83 hazard maps available at 100 metres resolution for the whole region of the Mediterranean Basin.
84 The hydrological input data are calculated using the LISFLOOD hydrological model (van der
85 Knijff et al., 2010; Burek et al, 2013; <https://ec-jrc.github.io/lisflood/>), based on updated routines
86 and input data in respect to the previous dataset by Dottori et al. (2016a). Flood simulations are
87 performed with the hydrodynamic model LISFLOOD-FP (Bates et al., 2010; Shaw et al., 2021),
88 following the approach developed by Alfieri et al., (2014; 2015).
89 To provide a comprehensive overview of the skill of the new hazard maps, we perform a
90 validation exercise using official hazard maps for a number of countries, regions and large river
91 basins in Europe. The number and extent of the validation sites allows for a more detailed
92 evaluation with respect to previous efforts by Alfieri et al. (2014) and Paprotny et al. (2017), even
93 though none of the validation sites is located outside Europe (due the unavailability of national
94 flood maps). Finally, we discuss the results of the validation in light of previous literature studies,
95 we compare the performance of the present and previous versions of the flood hazard map dataset,
96 and we discuss a number of tests with alternative datasets and methods.

97 *2) Data and methods*

98 In this Section we describe the procedure adopted to produce and validate the flood hazard maps.
99 The hydrological input data consist of daily river flow for the years 1990-2016, produced with
100 the hydrological model LISFLOOD (see Section 2.1), based on interpolated daily meteorological
101 observations. River flow data are analysed to derive frequency distributions, peak discharges and
102 flood hydrographs, as described in Section 2.2. Flood hydrographs are then used to simulate
103 flooding processes at local scale with the LISFLOOD-FP hydrodynamic model (Section 2.3).

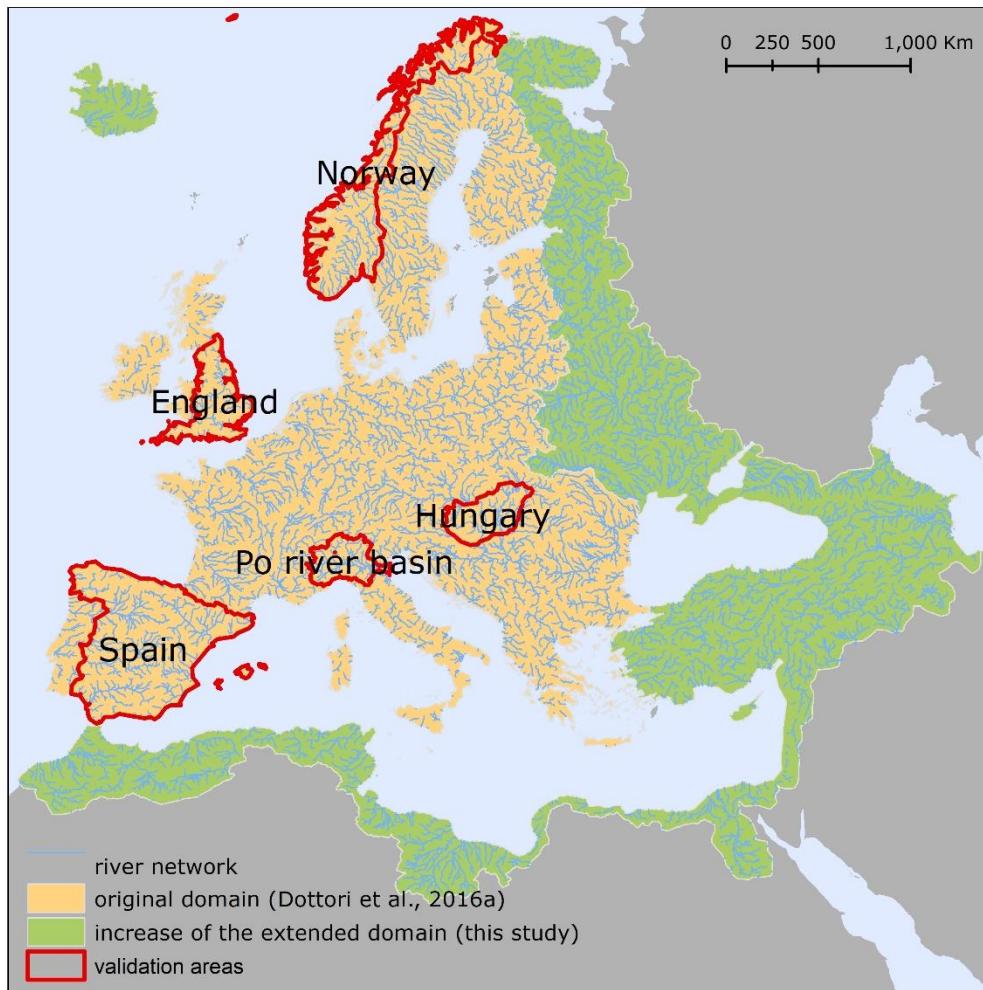
104 Finally, Section 2.4 describes the validation exercise and the comparison of different approaches
105 and input datasets.

106 *2.1 The LISFLOOD model*

107 LISFLOOD (Burek et al, 2013; van der Knijff et al., 2010) is a distributed, physically-based
108 rainfall-runoff model combined with a routing module for river channels. For this work we used
109 an updated version of LISFLOOD, released as open-source software and available at [https://ec-](https://ec-jrc.github.io/lisflood/)
110 [jrc.github.io/lisflood/](https://ec-jrc.github.io/lisflood/). The new version features an improved routine to calculate water
111 infiltration, the possibility of simulating open water evaporation and minor adjustments that
112 correct previous code inconsistencies (Arnal et al., 2019). The model is applied to run a long-term
113 hydrological simulation for the period 1990-2016 at 5 km grid spacing and at daily resolution,
114 which provides the hydrological input data for the flood simulations. Note that the same
115 simulation also provides initial conditions for daily flood forecast issued by EFAS.

116 The long-term run of LISFLOOD is driven by gridded meteorological maps, derived by
117 interpolating meteorological observations from stations and precipitation datasets (see Appendix
118 A for details). The meteorological dataset has been updated with respect to the dataset used by
119 Dottori et al. (2016a), to include new stations and gridded datasets across the new EFAS domain
120 (Arnal et al. 2019). In addition, LISFLOOD simulations require a number of static input maps
121 such as land cover, digital elevation model (DEM), drainage network, soil parameters and
122 parameterization of reservoirs. All the static maps have been updated to cover the whole EFAS
123 domain depicted in Figure 1. Further details on the static maps are provided by Arnal et al. (2019).
124 The current LISFLOOD version also benefits from an updated calibration at European scale,
125 based on the Evolutionary Algorithm approach (Hirpa et al., 2018) with the modified Kling-Gupta
126 efficiency criteria (KGE; Gupta et al., 2009) as objective function, and streamflow data for 1990-
127 2016 from more than 700 gauge stations. The same stations have been used to validate model
128 results, considering different periods of the time-series. The calibration and validation procedure
129 and the resulting hydrological skill are described by Arnal et al (2019), and summarized in
130 Appendix B. While we did not carry out a formal comparison with the previous LISFLOOD
131 calibration, which used a different algorithm and performance indicators (Zajac et al., 2013), the
132 larger dataset of streamflow observations and the improvement of the calibration routines should
133 provide a better performance.

134 The geographical extent used in the present study to produce the flood maps follows the recent
135 enlargement of EFAS (Arnal et al., 2019), and is shown in Figure 1. The new domain is
136 approximately 8,930,000 km² wide (an increase of 76% compared with the previous extent). The
137 new extent covers the entire area of geographical Europe (with the exclusion of the Volga river
138 basin and a number of river basins of the Arctic Sea in Russia), all the rivers entering the
139 Mediterranean and Black Seas (with a partial inclusion of the Nile river basin), plus the entire
140 territories of Armenia, Georgia, Turkey, and most of Syria and Azerbaijan.



141
142 *Figure 1. Geographical extent of the EFAS extended domain covered by the present dataset of*
143 *flood hazard maps. The extent of the map dataset produced by Dottori et al. (2016a) is depicted*
144 *in beige, while the regions added with the extended domain are in green. The Figure also displays*
145 *the river network considered by the flood maps and the areas used for the validation exercise (see*
146 *Sections 2.3 and 3).*

147 The river network included in the new flood hazard maps has a total length of 329,000 km, with
148 an 80% increase compared with the previous flood maps (Alfieri et al., 2015; Dottori et al.,
149 2016a).

150 *2.2 Hydrological input of flood simulations*

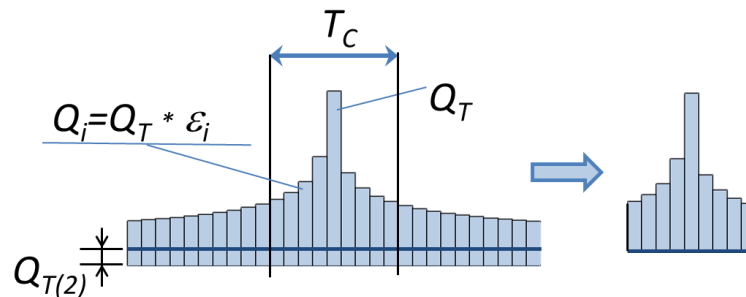
151 The hydrological input data required for the flood simulations are provided using synthetic flood
152 hydrographs, following the approach proposed by Alfieri et al. (2014).

153 We use the streamflow dataset derived from the long-term run of LISFLOOD described in Section
154 2.1, considering the rivers with upstream drainage areas larger than 500 km². This threshold was
155 selected because the meteorological input data cannot accurately capture the short and intense
156 rainfall storms that induce extreme floods in small river basins, and therefore the streamflow
157 dataset does not represent accurately the flood statistics of smaller catchments (Alfieri et al.,
158 2014).

159 For each pixel of the river network we selected annual maxima over the period 1990-2016 and
160 we used the L-moments approach to fit a Gumbel distribution and calculate peak flow values for
161 reference return periods of 10, 20, 50, 100, 200 and 500 years. We also calculated the 30- and
162 1,000-year return periods in limited parts of the model domain to allow validation against official
163 hazard maps, see Section 2.3. The resulting goodness-of-fit is presented and discussed in
164 Appendix B. We used the Gumbel distribution to keep a parsimonious parameterization (two
165 parameters instead of three for the generalized extreme value (GEV), log-normal and other
166 distributions), thus avoiding over-parameterization when extracting high return period maps from
167 a relatively short time-series. The same distribution was also adopted for the extreme value
168 analysis in previous studies regarding flood frequency and hazard (Alfieri et al., 2014, 2015;
169 Dottori et al., 2016).

170 Subsequently, we calculate a Flow Duration Curve (FDC) from the streamflow dataset. The FDC
171 is obtained by sorting in decreasing order all the daily discharges, thus providing annual
172 maximum values Q_D for any duration i between 1 and 365 days. Annual maximum values are
173 then averaged over the entire period of data, and used to calculate the ratios ε_i between each
174 average maximum discharge for i -th duration $Q_{D(i)}$ and the average annual peak flow (i.e. $Q_D = 1$
175 day). Such a procedure was carried out for all the pixels of the river network.

176 The synthetic flood hydrographs are derived using daily time-steps, following the procedure
 177 proposed by Maione et al. (2003). The peak value of the hydrograph is given by the peak discharge
 178 for the selected T-year return period Q_T , while the other values for Q_i are derived by multiplying
 179 Q_T by the ratio ε_i . The hydrograph peak Q_T is placed in the centre of the hydrograph, while the
 180 other values for Q_i are sorted alternatively as shown in Figure 2. The resulting hydrograph shape
 181 is therefore fully consistent with the empirical values of the flow duration curve. The total
 182 duration of the synthetic hydrograph is given by the local value of the time of concentration T_c ,
 183 such that all of the durations $> T_c$ are discarded from the final hydrograph (Figure 2).



184

185 *Figure 2. General scheme of flood hydrographs (adapted from Alfieri et al., 2014).*

186

187 Because river channels are usually not represented in continental-scale topography, flood
 188 hydrograph values are reduced by subtracting the 2-year discharge peak $Q_{T(2)}$, which is commonly
 189 considered representative of river bank-full conditions. (Note that the original DEM is not
 190 modified with this procedure). Hence, the overall volume of the flood hydrograph is given by the
 191 sum of all daily flow values with duration $< T_c$.

192

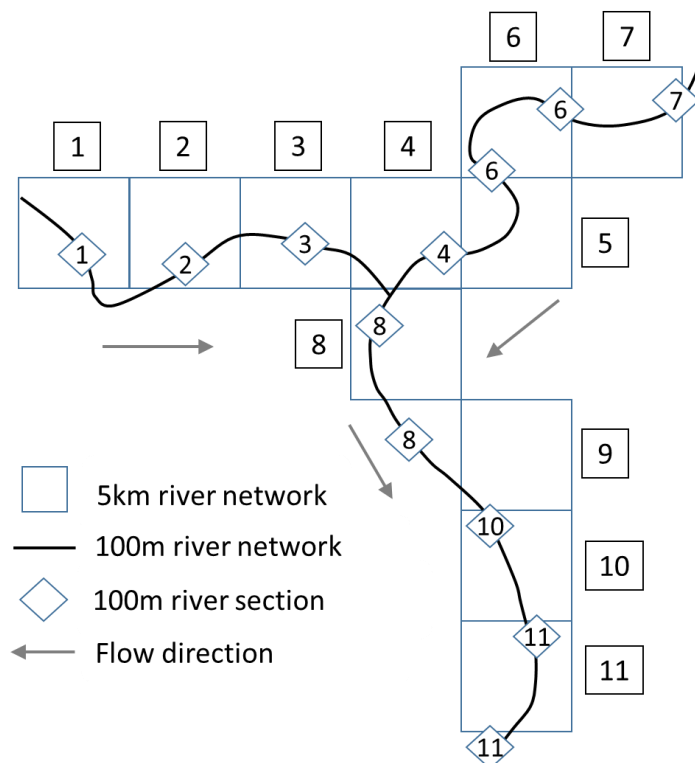
193 **2.3 Flood hazard mapping**

194 The continental-scale flood hazard maps are derived from local flood simulations run along the
 195 entire river network, as in Alfieri et al. (2014). We use the DEM at 100 metres resolution
 196 developed for the Catchment Characterization and Modelling Database (CCM; Vogt et al., 2007)
 197 to derive a high-resolution river network at the same resolution. Along this river network we
 198 identify reference sections every 5 km along the stream-wise direction, and we link each section
 199 to the closest upstream section (pixel) of the EFAS 5 km river network, using a partially
 200 automated procedure to ensure a correct linkage near confluences. In this way, the hydrological

201 variables necessary to build the flood hydrographs can be transferred from the 5 km to the 100
 202 metres river network. Figure 3 describes how the 5 km and 100 metres river sections are linked
 203 using a conceptual scheme.

204 Then, for every 100 metres river section we run flood simulations using the two-dimensional
 205 hydrodynamic model LISFLOOD-FP (Shaw et al., 2021), to produce a local flood map for each
 206 of the six reference return periods. Simulations are based on the local inertia solver of
 207 LISFLOOD-FP developed by Bates et al. (2010), which is now available as open-source software
 208 (<https://www.seamlesswave.com/LISFLOOD8.0>). We use the CCM DEM as elevation data, the
 209 synthetic hydrographs described in Section 2.2 as hydrological input data, and a mosaic of
 210 CORINE Land Cover for the year 2016 (Copernicus LMS, 2017) and Copernicus GlobCover
 211 (Global Land Cover Map) for the year 2009 (Bontemps et al., 2009) to estimate the friction
 212 coefficient based on land use.

213



214

215 *Figure 3. Conceptual scheme of the EFAS river network (5 km, squares) with the high-resolution*
 216 *network (100 metres) and river sections (diamonds) where flood simulations are derived. The*
 217 *related sections of the two networks are indicated by the same number. Source: Dottori et al.*
 218 *(2017).*

219

220 Finally, the flood maps with the same return period are merged together to obtain the continental-
221 scale flood hazard maps. The 100 metres river network is included as a separate map in the dataset,
222 to delineate those water courses that were considered in creating the flood hazard maps.

223 It is important to note that the flood maps developed do not account for the influence of local
224 flood defences, in particular dyke systems. Such limitation has been dictated mainly by the
225 absence of consistent data at European scale. None of the available DEMs for Europe has the
226 required accuracy and resolution to embed artificial embankments into elevation data.
227 Furthermore, there are no publicly available continental or national datasets describing the
228 location and characteristics (e.g. dyke height, distance from river channel) for flood protections.
229 Currently available datasets are based on the design return period of flood protection, e.g. the
230 maximum return period of flood events that protections can withstand before being overrun,
231 (Jongman et al., 2014; Scussolini et al., 2016). Most of the protection standards reported by these
232 datasets for Europe are based on empirical regressions derived using proxy variables (e.g. GDP,
233 land use), with few data based on actual design standards. While these datasets have been applied
234 to calculate flood risk scenarios (Alfieri et al., 2015) and flood impacts (Dottori et al., 2017), they
235 have important limitations when used for mapping flood extent. Wing et al. (2017) linked the
236 flood return period of protection standards with flood frequency analysis to adjust the bank height
237 of the river channels, however with impaired performance of the model. Moreover, recent studies
238 for United States suggest that empirical regressions based on gross domestic product and land use
239 may not be reliable (Wing et al., 2019).

240 Despite these limitations, maps not accounting for physical flood defences may be applied to
241 estimate the flood hazard in case of failure of the protection structures, and for flood events
242 exceeding protection levels.

243 *2.3 Validation of flood hazard maps*

244 *2.3.1 Selection of validation areas and maps*

245 The validation of large-scale flood hazard maps requires the use of benchmarks with one or more
246 datasets with extension and accuracy commensurate to the modelled maps. For instance Wing et
247 al. (2017) used the official hazard maps developed for the conterminous United States to evaluate

248 the performance of two flood hazard models, respectively designed to produce global- and
249 continental-scale flood maps (see Section 1). In Europe, all EU Member States as well as the UK
250 have developed national datasets of flood hazard maps for a range of flood probabilities (usually
251 expressed with the flood return period), following the guidelines of the EU Floods Directive (EC
252 2007). These maps are usually derived using multiple hydrodynamic models of varying
253 complexity (AdB Po, 2012) based on high-resolution topographic and hydrological datasets, such
254 as DEMs of at least 5 metres resolution in England (Sampson et al., 2015), LIDAR elevation data
255 in Spain (MITECO 2011), and river sections based on LIDAR surveys in the Po River basin (AdB
256 Po, 2012). Although official maps might be either prone to errors or incomplete (Wing et al 2017),
257 these are likely to provide higher accuracy than the modelled maps presented here, and therefore
258 they have been selected as reference maps for the validation. While official flood maps are
259 generally available online for consultation on Web-GIS services, only a few countries and river
260 basin authorities make the maps available for download in a format that allows comparison with
261 geospatial data. Table 1 presents the list of flood hazard maps that could be retrieved and used for
262 the validation exercise, while their geographical distribution is shown in Figure 1. Note that the
263 relevant links to access these maps are provided in the Data Availability section.

264 While more of such official maps are likely to become available in the near future, the maps
265 considered here offer an acceptable overview of the different climatic zones and floodplain
266 characteristics of the European continent. Conversely, we could not retrieve national or regional
267 flood hazard maps outside Europe, meaning the skill of the modelled maps could not be tested in
268 the arid regions in Northern Africa and Eastern Mediterranean. In Norway, Spain, the UK and the
269 Po River Basin the official maps take flood defences into account, which are not represented in
270 the modelling framework. Official maps for England also include areas prone to coastal flooding
271 events (such as tidal and storm surges). None of the official maps include areas prone to pluvial
272 flooding, which are therefore not considered in this analysis.

273 As mentioned in Section 2.3, the modelled maps do not include the effect of flood protections.
274 Wherever possible, for the comparison exercise we selected either reference flood maps that do
275 not account for protections (e.g. Hungary) or maps for flood return periods exceeding local
276 protection standards, assuming that the resulting flood extent is relatively unaffected by flood
277 defences. For example, the main stem of the Po river is protected against 1-in-200-year flood
278 events (Wing et al., 2019), whereas protection standards in England and Norway are usually

279 above 20 years (Scussolini et al., 2016). Reference maps where the extent and design level of
 280 protection are not known (e.g. Spain) have been also included in the comparison to increase the
 281 number of validation areas.

282

Country	Geographical extent	Return periods used	Defences included
Hungary	Country scale	30 - 100 – 1,000 years	No
Italy	Po River Basin	500 years	Yes
Norway	Country scale	100 years	Yes
Spain	Country scale	10 - 100 - 500 years	Yes
UK	England	100 – 1,000 years	Yes

283 *Table 1. Characteristics of the flood hazard maps used in the validation exercise. The links for*
 284 *downloading the maps are provided in the Data Availability section.*

285 **2.3.2 Performance metrics and validation procedure**

286 The national flood hazard maps listed in Table 1 are provided as polygons of flood extent, with
 287 no information on water depth or on original resolution of data. According to Sampson et al.
 288 (2015), the official flood hazard maps for England are constructed using DEMs of at least 5 metres
 289 resolution, therefore flood extent maps should be of comparable resolution. Reference flood maps
 290 for the Po basin and Spain are likely to have a similar resolution since they are based on LIDAR
 291 elevation data (AdB Po, 2012; MITECO 2011). For the comparison, official reference maps have
 292 been converted to raster format with the same resolution as the modelled maps (i.e. 100 metres),
 293 while the latter have been converted to binary flood extent maps. To improve the comparison
 294 between modelled and reference maps we applied a number of corrections. Firstly, we used the
 295 CORINE Land Cover map to exclude permanent water bodies (river beds of large rivers or
 296 estuaries, lakes, reservoirs, coastal lagoons) from the comparison. Secondly, we restricted the
 297 comparison area around modelled maps to exclude the elements of river network (e.g. minor
 298 tributaries) included in the reference maps but not in the modelled maps. We used a different

299 buffer extent according to each study area, considering the floodplain morphology and the
 300 variable extent and density of mapped river network. For example, in Hungary we applied a 10-
 301 km buffer around modelled maps to include the large flooded areas reported in reference maps
 302 and to avoid overfitting. In England, we used a 5 km buffer due to the high density of the river
 303 network mapped in the official maps. The buffer is also applied to mask out coastal areas far from
 304 rivers estuaries, because official maps include flood-prone areas due to 1-in-200-year coastal
 305 flood events. We calculated that flood-prone areas inside the 5 km buffer correspond to 73% of
 306 the total extent for the 1-in-100-year flood. For the Po river basin, we excluded from the
 307 comparison the areas belonging to the Adige river basin and the lowland drainage network, which
 308 are not included in the official hazard maps. In Spain and Norway official flood hazard maps have
 309 only been produced where relevant assets are at risk, according to available documentation
 310 [MITECO 2011; NVE 2020]. We therefore restricted the comparison only to areas where official
 311 flood hazard maps have been produced. Table 2 provide the list of parameters used to determine
 312 the areas used for the comparison.

Test area	Buffer value (reference maps)	Buffer value (modelled maps)
Hungary	NA	10 km
Po River Basin	NA	See main text
Norway	5 km	5 km
Spain	5 km	5 km
England	NA	5 km

313 *Table 2. List of parameters used to determine the extent of areas used for comparing reference*
 314 *and modelled maps (NA: buffer not applied).*

315
 316 We evaluate the performance of simulated flood maps against reference maps using a number of
 317 indices proposed in literature (Bates and De Roo, 2000; Alfieri et al., 2014; Dottori et al., 2016b;
 318 Wing et al., 2017). The hit ratio (HR) evaluates the agreement of simulated maps with
 319 observations and it is defined as:

320
$$HR = (Fm \cap Fo)/(Fo) \times 100 \quad (1)$$

321 where $Fm \cap Fo$ is the area correctly predicted as flooded by the model, and Fo indicates the total
322 observed flooded area. HR scores range from 0 to 1, with a score of 1 indicating that all wet cells
323 in the benchmark data are wet in the model data. The formulation of the HR does not penalize
324 over-prediction, which can be instead quantified using the false alarm ratio FAR:

325
$$FAR = (Fm/Fo)/(Fm) \times 100 \quad (2)$$

326 where Fm/Fo is the area wrongly predicted as flooded by the model. FAR scores range from 0
327 (no false alarms) to 1 (all false alarms). Finally, a more comprehensive measure of the agreement
328 between simulations and observations is given by the critical success index (CSI), defined as:

329
$$CSI = (Fm \cap Fo)/(Fm \cup Fo) \times 100 \quad (3)$$

330 where $Fm \cup Fo$ is the union of observed and simulated flooded areas. CSI scores range from 0
331 (no match between model and benchmark) to 1 (perfect match between benchmark and model).

332 **2.4 Additional tests**

333 To choose the best possible methodologies and datasets to construct the flood hazard maps, we
334 performed a number of tests using recent input datasets, as well as alternative strategies to account
335 for vegetation effects on elevation data.

336 **2.4.1 Elevation data**

337 It is well recognized that the quality of flood hazard maps strongly depend on the accuracy of
338 elevation data used for modelling (Yamazaki et al., 2017). This is especially crucial for
339 continental-scale maps, since the quality of available elevation datasets is rarely commensurate
340 to the accuracy required for modelling flood processes [Wing et al., 2017]. Moreover, high-
341 resolution and accurate elevation data such as LIDAR-based DEMs cannot be used for reasons of
342 consistency, since these data are only available for few areas and countries.

343 The recent release of new global elevation models have the potential to improve the accuracy of
344 large-scale flood simulations, and hence the quality of flood hazard maps. Here, we test the use
345 of the MERIT DEM (Yamazaki et al., 2017) within the proposed modelling approach and we
346 compare the results with those obtained with CCM DEM. The MERIT DEM is based on the
347 SRTM data, similarly to CCM DEM, but has been extensively corrected and improved through
348 comparisons with other large-scale datasets, to eliminate error bias, improve data accuracy at high

349 latitudes (areas above 60° are not covered by SRTM), and compensate for factors like vegetation
350 cover. Note that areas above 60° in CCM DEM were derived from national datasets, and therefore
351 these areas are where the two datasets are likely to differ most.

352 *2.4.2 Correction of elevation data with land use*

353 The CCM DEM elevation dataset is mostly based on SRTM data, and so the elevation values can
354 be spuriously increased by the effect of vegetation canopy in densely vegetated areas, and by
355 buildings in urban areas. Recent research work has proposed advanced techniques to remove
356 surface artefacts, based on artificial neural networks (Wendi et al., 2016, Kulp and Strauss, 2018)
357 or other machine learning methods (Liu et al., 2018; Meadows and Wilson, 2021). Most
358 approaches correct DEM elevation with higher-accuracy datasets, using auxiliary data such as
359 tree density and height for correcting vegetation bias (as done for the MERIT-DEM by Yamazaki
360 et al., 2017), whereas elevation bias in urban areas can be corrected using night light, population
361 density, or OpenStreetMap elevation data (Liu et al., 2018). Given that improving elevation data
362 is not the main scope of this work, we opted for applying a simpler method for quickly correcting
363 the CCM DEM elevation data. Specifically, we use the land cover map derived from CORINE
364 Land Cover and Copernicus GlobCover to identify densely vegetated areas and urban areas, and
365 we applied a correction factor as a function of local land use to reduce elevation locally. The
366 correction factor varies from 8 metres for densely forested areas, to 2 metres for urban areas. Note
367 that these values are based on the findings of previous literature studies such as Baugh et al.
368 (2013) and Dottori et al. (2016b), while a formal calibration was not undertaken.

369

370 *3) Results and discussion*

371 We present the outcomes of the validation exercise by first describing the general results at
372 country and regional scale (Section 3.1). Then, we discuss the outcomes for England, Hungary
373 and Spain (Section 3.2), while the Norway and Po river basin case studies are presented in the
374 Appendix C. We also complement the analysis with additional validation over major river basins
375 in England and Spain. In Section 3.3 we compare our results with the validation exercise carried
376 out by Wing et al. (2017) and with the findings of other literature studies. Finally, in Section 3.4
377 and 3.5 (and Appendix B) we compare the performance of the present and previous versions of

378 the flood hazard map dataset, and we discuss the results of the tests with different elevation data
379 and strategies to account for vegetation.

380 *3.1 Validation of modelled maps at national and regional scale*

381 Table 3 presents the validation results for each testing area and return period. The performance
382 metrics are calculated using the total extent of the reference and modelled maps with the same
383 return period. The first visible outcome is the low scores for the comparisons with reference maps
384 with high probability of flooding, i.e. low flood return periods (< 30 years). Performances improve
385 markedly with the increase of return periods due to the decrease of false alarm rate (FAR), while
386 the hit rate (HR) does not vary significantly. In particular, critical success index (CSI) values
387 approach 0.5 for the low probability flood maps, i.e., for return periods equal or above 500 years.
388 Considering that most of the reference flood maps include the effect of flood defences (unlike the
389 modelled maps), these results suggest that the majority of rivers in the study areas may be
390 protected for flood return periods of around 100 years or less, as indeed reported by available
391 flood defence databases (Scussolini et al., 2016). Differences between simulated and reference
392 hydrological input are likely to influence the skill of modelled flood maps and may depend on
393 several factors such as the hydrological model performance for peak flows, the extreme value
394 analysis (distribution used for extreme value fitting, length of available time-series) and the design
395 hydrograph estimation. In the following Sections, we evaluate the modelled hydrological regime
396 considering the skill of the LISFLOOD long-term simulation and the uncertainty of the extreme
397 value analysis (see Appendix B2). However, further analysis is difficult as we have no specific
398 information on the hydrological input used for the reference flood maps (e.g. peak flows,
399 statistical modelling of extremes, hydrograph shape). High-probability floods are also sensitive
400 to the method used to reproduce river channels, and the simplified approach used in this study
401 might underestimate the conveyance capacity of channels (see Section 3.2.2 for an example).
402 Finally, the better performance for low-probability floods may also depend on floodplain
403 morphology, where valley sides create a morphological limit to flood extent.

404

405

406

407

408 *Table 3. Results of the validation against official flood hazard maps: value of the performance*
 409 *indices at country and regional scale. RP=Return Period, HR=Hit Ratio, FAR= False Alarm*
 410 *Ratio, CSI=Critical Success Index.*

REGION	RP (years)	HR	FAR	CSI
Spain	10	0.58	0.65	0.28
Hungary	30	0.77	0.88	0.11
Spain	100	0.63	0.44	0.42
Hungary	100	0.76	0.74	0.24
Norway	100	0.70	0.72	0.25
England	100	0.53	0.31	0.43
Po River Basin	500	0.60	0.13	0.56
Spain	500	0.61	0.36	0.45
Hungary	1000	0.76	0.45	0.47
England	1000	0.52	0.12	0.48

411 **3.2 Discussion of results at national and regional scale**

412 The results in Table 3 highlight considerable differences in the skill of the flood maps across
 413 countries and regions. While some differences may arise from the variability of floodplain
 414 morphology and model input data, others are attributable to the different methods applied to
 415 produce the reference maps (MITECO 2011; NVE 2020). In the following sections we examine
 416 in more detail the outcomes for each study area.

417 **3.2.1 England**

418 According to Table 3, modelled flood maps tend to underestimate flood extent in England, as
 419 visible by the HR values around 0.5 (e.g. out of every two flooded cells, only one is correctly
 420 identified as flooded by the model). Such result is confirmed when focusing the analysis on the
 421 major river basins of England, as reported in Table 4. Notably, HR has generally marginal or no
 422 increases with the increase of return period considered, while FAR values have a marked
 423 decrease. Results of reported by Arnal et al. (2019) and summarized in Figure B1 suggest a fair
 424 hydrological skill of the LISFLOOD calibration in England, with KGE values generally above
 425 0.5. The difference between estimated and reference discharge annual maxima is also acceptable,

426 generally below 25%. However, there is not a clear correlation between hydrological and flood
 427 map skill, with some basins (e.g. Thames) showing high KGE values but relatively low CSI
 428 values.

429 For the Thames basin, the low CSI value is likely influenced by the tidal flooding component
 430 from London eastwards. According to Sampson et al. (2015), the official flood hazard map
 431 assumes a 1 in 200 year coastal flood along with failure of the Thames tidal barrier, whereas our
 432 river flood simulations use the mean sea level as boundary condition and do include storm surge
 433 and tidal flooding. Concurrent fluvial-tidal flooding processes occur in other river estuaries, so
 434 this might reduce the skill of the modelled maps. Furthermore, the Thames catchment is heavily
 435 urbanized and has extensive flood defence and alleviation schemes compared to the other
 436 catchments (Sampson et al., 2015). Both aspects might increase the elevation bias of CCM DEM
 437 and complicate the correct simulation of extreme flood events.

438

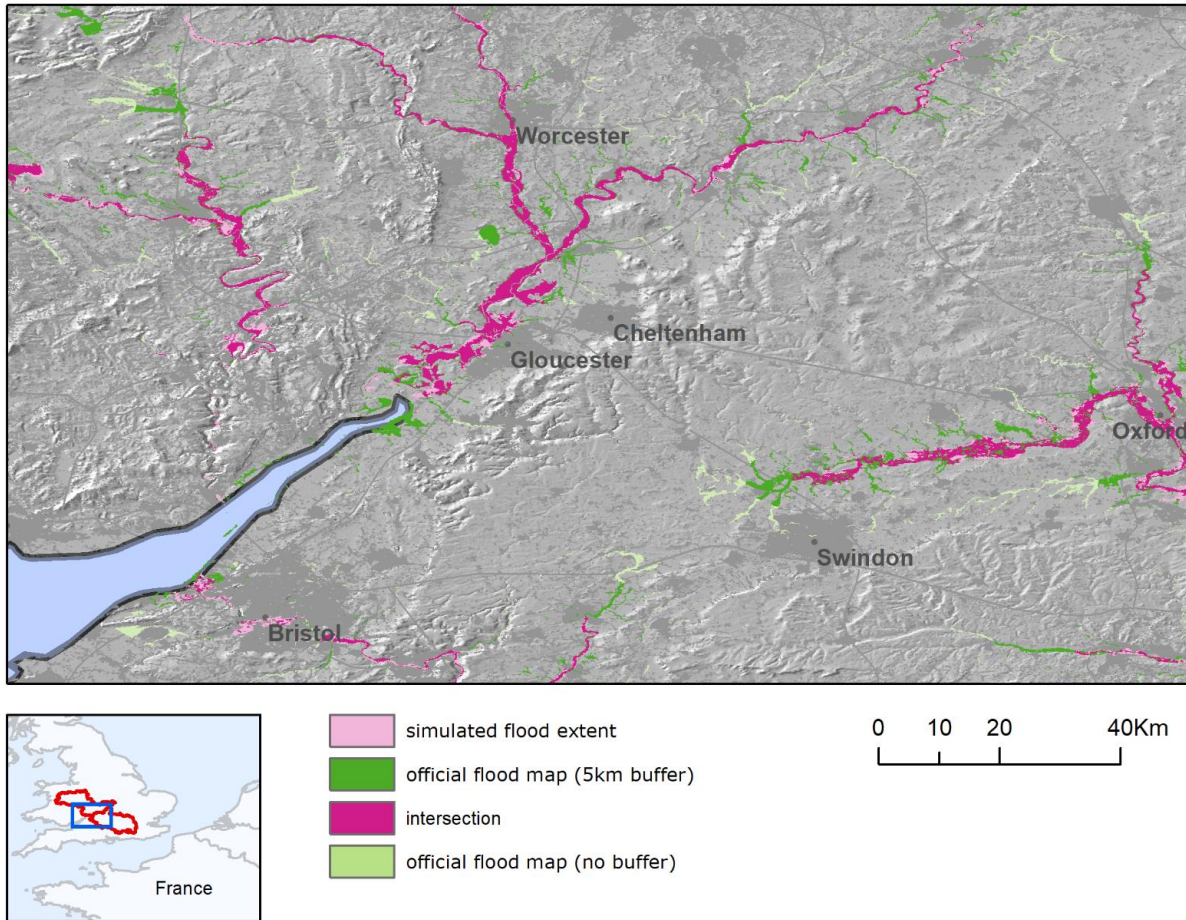
439 *Table 4. Validation indices in England and in major river basins.*

Catchments	100-year RP			1,000-year RP		
	HR	FAR	CSI	HR	FAR	CSI
England	0.53	0.31	0.43	0.52	0.12	0.48
Ouse	0.57	0.39	0.42	0.56	0.19	0.49
Severn	0.64	0.24	0.53	0.63	0.20	0.54
Thames, above Lea	0.56	0.46	0.38	0.55	0.23	0.47
Trent	0.63	0.28	0.50	0.59	0.06	0.57
Tyne	0.51	0.43	0.37	0.52	0.28	0.43

440

441 Besides these results, the visual inspection of reference maps suggest that the underestimation is
 442 partly caused by the high density of mapped river network in the reference maps, in respect to
 443 modelled maps. Indeed, the modelling framework excludes river basins with an upstream basin
 444 area below 500 km², meaning that EFAS maps only cover main river stems but miss out several
 445 smaller tributaries. This is clearly visible over the Severn and in the upper Thames basins (Figure
 446 4), and might also explain the lower skill in the lowlands of Ouse and Trent rivers, where the
 447 contributions of main river stems and tributaries to the flood extent are difficult to separate.
 448 Including minor tributaries in the flood maps would require either to increase the resolution of

449 the climatological forcing to reproduce intense local rainfall, or to add a pluvial flooding
 450 component as done by Wing et al. (2017). Finally, areas prone to storm surge and tidal flooding
 451 around river estuaries might further reduce the overall skill of modelled maps, despite the 5 km
 452 buffer applied.

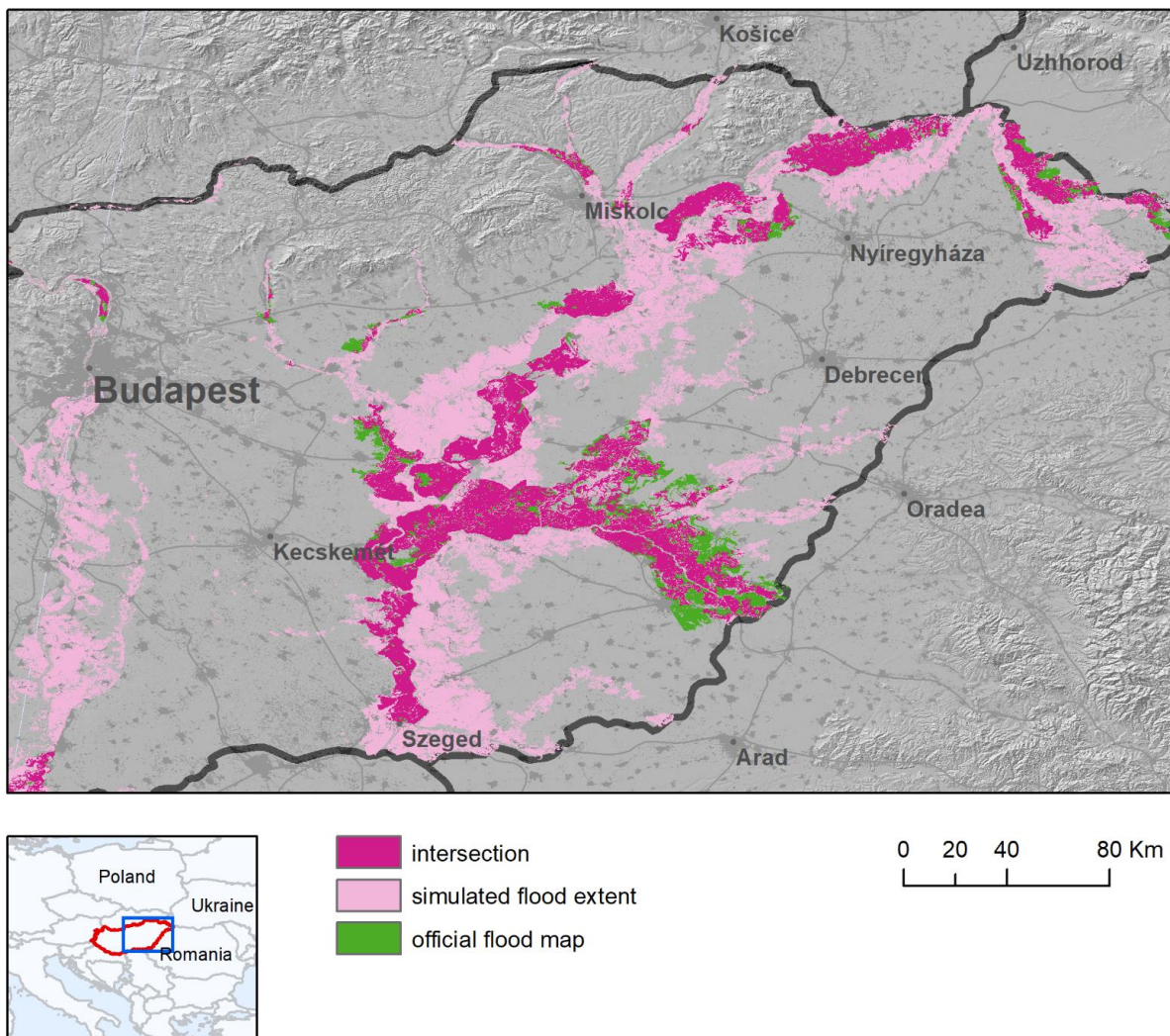


453
 454 *Figure 4. Comparison of modelled and reference flood hazard maps (1-in-100-year) over the*
 455 *Severn (centre) and the upper Thames (right) river basins in England. Purple areas denotes*
 456 *intersection (agreement) between the modelled and reference set of maps. The original*
 457 *reference maps (i.e. with no masking around modelled maps) are shown in light green.*
 458

459 3.2.2 Hungary

460 The results in Table 3 for Hungary show a general tendency to overestimate flood extent for all
 461 return periods. HR values are consistently high and do not change much with the return period.
 462 Conversely, FAR is very high for the 1-in-30 year flood map and still considerable even for the

463 1-in-1000 year flood map. Arnal et al (2019) reported a fair hydrological skill of LISFLOOD
 464 (KGE values >0.5) for the calibration period, even though KGE validation values were
 465 considerably low for the Tisza River. The uncertainty on estimated discharge annual maxima is
 466 also comparable to the average values reported in Appendix B2.
 467 Given that flood defences are not modelled in reference maps, the observed results may be
 468 explained by assuming a large conveyance capacity of river channels. For instance, the 1-in-100
 469 year reference map shows relatively few flooded areas for the Danube main stem (Figure 5), thus
 470 suggesting that the main channels can convey the 1-in-100-year discharge without overflowing.



471
 472 *Figure 5. Comparison of modelled and reference flood hazard maps (1-in-100-year) over the*
 473 *Danube (left) and Tisza (right) rivers in Hungary. Purple areas denotes the intersection between*
 474 *the modelled and reference set of maps.*

475 Conversely, river channels in the modelling framework are assumed to convey only the 1-in-2-
476 year discharge. Obviously, the same considerations can be made for 1-in-30-year discharge for
477 the majority of river network, which explains the very low scores. Furthermore, artificial
478 structures such as road embankments and drainage network may further reduce flood extent in
479 lowland areas, leading to further overestimation given the fact that these features are not
480 represented in the DEM. These findings highlight the need for high-resolution DEM fed with
481 local-scale information to achieve adequate performance in lowland areas, as observed also by
482 Wing et al (2019b).

483 *3.2.3 Spain*

484 The performance of the modelled maps in Spain show a fairly stable HR value and decreasing
485 FAR values with increasing return periods, similarly to what was observed for England and
486 Hungary. The analysis of the results for the major river basins of the Iberian Peninsula, reported
487 in Table 5, provide further insight on the skill of flood maps. A number of basins exhibit both
488 large HR and FAR such as the Duero, Tajo and Guadalquivir basins. Rivers in South-East Spain
489 (Segura, Jucar) have relatively low HR values, while the modelled maps perform better in the
490 Ebro river basin. The interpretation of results requires the consideration of different aspects. The
491 poor results for the 1-in-10-year maps are likely due to the effect of flood protection structures,
492 such as dykes and flood regulation systems, which are probably relevant also for the 1-in-100-
493 year map. Indeed, most Iberian rivers are regulated by multiple reservoirs, which are often used
494 to reduce flood peaks according to specific operating rules. While dykes are not represented in
495 the inundation model, reservoirs are included in the LISFLOOD model through a simplified
496 approach, given that operating rules are not known. Therefore, the real and modelled hydrological
497 regimes might differ significantly, including flow peaks of low-probability flood events. This is
498 also reflected by the low hydrological skill of LISFLOOD, with KGE values generally below 0.5
499 with few exceptions (Figure B1).

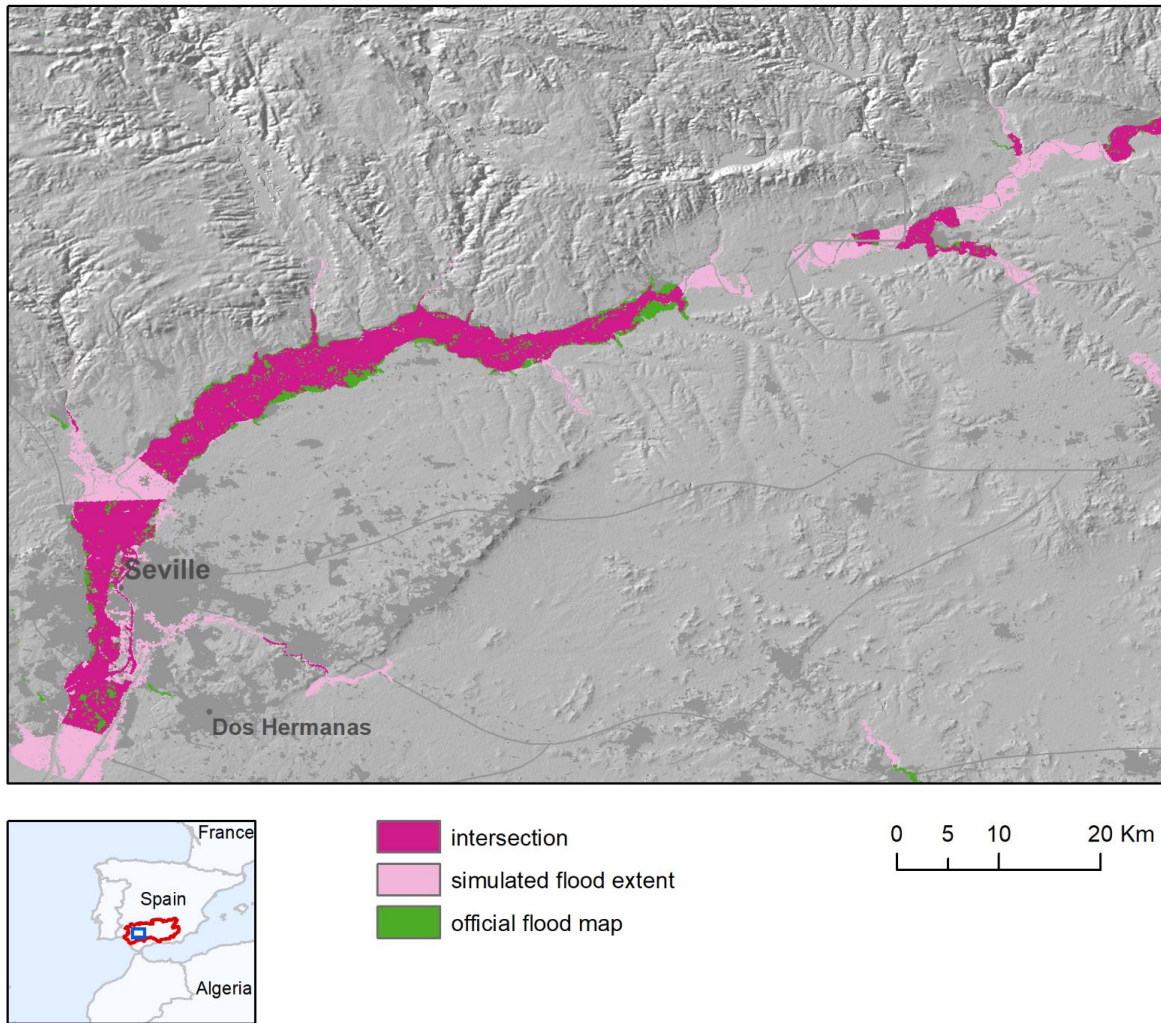
500 In addition, the comparison of modelled and reference maps is affected by the partial coverage of
501 the reference inundation maps in several river basins. According to the information available in
502 the official website (MITECO 2011) large sections of the river network in the basins of the Duero,
503 Tajo, Guadiana and Guadalquivir rivers have not been analysed, due to the absence of relevant
504 assets or inhabited places at risk. Even though this has been accounted for by restricting the area

505 of comparison around reference maps, a visual inspection of the maps being compared shows
 506 spurious overestimation around the edges of reference map polygons (Figure 6). Finally, the low
 507 HR values scored in rivers in South-East Spain (Segura, Jucar) are partially explained by the
 508 presence of several tributaries not included in EFAS maps.

509

510 *Table 5. Validation indices in Spain and in some test river basins.*

Catchments	10-year RP			100-year RP			500-year RP		
	HR	FAR	CSI	HR	FAR	CSI	HR	FAR	CSI
Spain	0.58	0.65	0.28	0.63	0.44	0.42	0.61	0.36	0.45
Duero	0.60	0.74	0.22	0.65	0.55	0.36	0.65	0.46	0.42
Ebro	0.71	0.46	0.45	0.75	0.27	0.59	0.74	0.23	0.61
Guadalquivir	0.67	0.66	0.29	0.69	0.49	0.42	0.66	0.46	0.42
Guadiana	0.52	0.63	0.28	0.60	0.42	0.42	0.61	0.31	0.48
Jucar	0.32	0.89	0.09	0.53	0.46	0.36	0.51	0.39	0.39
Tajo	0.60	0.85	0.14	0.70	0.63	0.32	0.69	0.49	0.41
Segura	0.18	0.89	0.07	0.38	0.52	0.27	0.41	0.24	0.36



511
 512 *Figure 6. Comparison of modelled and reference flood hazard maps (1-in-100-year) over a*
 513 *stretch of the Guadalquivir river basin, Spain. Purple areas denote the intersection between the*
 514 *two set of maps.*

515 **3.3 Comparison with previous continental-scale validation studies**

516 To put the previously described results in context, we compare them with the validation exercises
 517 performed by Sampson et al. (2015) over the Thames and Severn rivers in England, and by Wing
 518 et al. (2017) over the United States. The study by Wing et al. is, to our knowledge, the first study
 519 that carried out a consistent validation of modelled flood hazard maps at the continental scale.
 520 Bates et al. (2021) have recently updated the work by Wing et al. by including pluvial and coastal
 521 flooding components in the modelling framework, but their work is not considered here. A

522 comparison of validation metrics of the three studies are shown in Table 6 and 7. For our
 523 framework, we calculated each index in Table 6 using the overall modelled and reference flood
 524 extent available for each return period (e.g. the value for the 100-year maps includes reference
 525 and modelled maps for England, Spain and Norway). As such, each area is weighted according
 526 to the extent of the corresponding flood map.

527 As can be seen in Table 6, the continental-scale model by Wing et al. achieved the highest scores
 528 both for 100-year and 500-year return periods. However, this model is based on national datasets
 529 with higher accuracy and resolution than those available for the European continent (e.g. a 10
 530 metres resolution DEM and a detailed catalogue of flood defences). The global and European
 531 models have comparable hit rates for the 100-year flood maps (0.68 and 0.65 respectively), but
 532 the former exhibits a much lower FAR value (0.34 compared to 0.61 for the European model),
 533 and a higher HR value for the 500-year maps.

534

535 *Table 6. Comparison of the performance metrics for the European model described in the present*
 536 *study and the two models evaluated in the study by Wing et al. (2017).*

	RP (years)	HR	FAR	CSI
US model (Wing et al.)	100	0.82	0.37	0.55
Global model (Wing et al.)	100	0.69	0.34	0.50
European model (this study)	100	0.66	0.61	0.32
US model (Wing et al.)	500	0.86	na	na
Global model (Wing et al.)	500	0.74	na	na
European model (this study)	500	0.61	0.24	0.51
European model (this study)	1000	0.68	0.39	0.47

537

538 The higher HR values scored by the US and global models might depend on the higher density of
 539 the modelled river network, which includes river reaches up to 50 km² by simulating both pluvial
 540 and fluvial flooding processes. The lower FAR values of the US and global models might be
 541 explained by the inclusion of flood defences. In the US model, defences are explicitly modelled
 542 using the US dataset of flood defences, while the global model parameterizes flood defences
 543 through the adjustment of channel conveyance using socioeconomic factors and degree of
 544 urbanization (Wing et al., 2017). However, Wing et al observed that the latter methodology had

545 a negligible effect on HR values in defended areas, when compared with an undefended version
546 of the model.

547 Another possible reason for the low FAR values is the different approach used in the validation
548 method. Wing et al. applied a narrow 1 km buffer around official maps to constrain the area of
549 comparison and avoid spurious over-prediction in areas not considered by official maps.
550 However, this might result in a reduction of true false alarms, because part of overestimated flood
551 areas can go undetected. To verify this hypothesis, we recalculated the performance indices
552 against the 100-year reference map in Spain using a 1 km buffer instead of the 5 km previously
553 applied to constrain the validation area. As a result the FAR dropped from 0.44 to 0.34, similar
554 to the performance of the global model. However, we observed a reduction of true false alarms,
555 especially in river basins with continuous map coverage such as the Ebro, Jucar and Segura.
556 The comparison of HR, FAR and CSI values show better scores for the global maps by Sampson
557 et al. (2015) in respect to our modelled maps (Table 7).

558

559 *Table 7. Comparison of the performance metrics for the maps described in the present study and*
560 *the global maps by Sampson et al. (2015). Metrics for the latter study are calculated removing*
561 *all channels with upstream areas of less than 500 km².*

	HR	FAR	CSI
Thames (this study)	0.56	0.46	0.38
Thames (Sampson et al. 2015)	0.73	0.3	0.56
Severn (this study)	0.64	0.24	0.53
Severn (Sampson et al. 2015)	0.83	0.23	0.67

562

563 The different masking applied to reference flood maps may explain some of the differences:
564 Sampson et al. removed all channels with upstream areas of less than 500 km², whereas here we
565 use a simpler 5 km buffer around modelled maps. The exclusion of permanent river channels in
566 our comparison may further penalize the overall score especially for the Thames, which as a rather
567 large channel estuary. Besides these differences in the validation, the better metrics of the maps
568 by Sampson et al. may depend on a more accurate hydrological input (based on regionalization
569 of gauge station data) and a better correction of urban elevation bias (based on a moving window
570 filter instead of the constant correction values applied here).

571 To provide further context, the US model by Wing et al. (2017) attained average CSI values of
 572 ~0.75 against a number of detailed local models, whereas flood models built and calibrated for
 573 local applications may achieve CSI scores up to 0.9 when benchmarked against very high quality
 574 data (see Wing et al., 2019a). Fleischmann et al. (2019) recently proposed that regional-scale
 575 models can provide locally relevant estimates of flood extent when $CSI > 0.65$. Although the
 576 overall values shown in Table 3 are consistently below this threshold, better results are observed
 577 for a number of river basins, as shown in Tables 4 and 5.

578 *3.4 Comparison with the previous flood map dataset*

579 Table 8 compares the performances of the flood hazard maps described in the present study
 580 (version 2) with the previous version developed by Dottori et al. (2016a; version 1). The
 581 comparison is shown for England and Hungary. Results for all other areas are comprised within
 582 the range of results shown in Table 3. As can be seen, differences are generally reduced across
 583 the different areas and return periods. Version 1 of the flood maps produced slightly better results
 584 in Hungary for the 100- and 1000-year return period (increased CSI and HR, lower FAR), while
 585 version 2 has somewhat improved performances in England, mainly driven by higher HR.

586
 587 *Table 8. Comparison of performances of the flood hazard maps described in the present study*
 588 *and developed by Dottori et al. (2016a). Table reports the ratio between flood extents (F2/F1)*
 589 *and the difference between Version 2 and 1 of the HR, FAR and CSI values.*

	RP (years)	F2/F1	ΔHR	ΔFAR	ΔCSI
Hungary	30	0.97	-0.5%	-0.4%	2.9%
Hungary	100	1.00	-2.1%	0.7%	-2.4%
Hungary	1000	1.01	-3.6%	5.7%	-6.3%
England	100	1.05	9.4%	1.7%	7.3%
England	1000	1.04	8.2%	-1.1%	7.7%

590
 591 These outcomes may be interpreted considering the changes in input data between the two
 592 versions, and the structure of the modelling approach and of input data, which in turn has not
 593 changed substantially. The main difference between the two map versions is given by the
 594 hydrological input, with Version 2 using the latest calibrated version of the LISFLOOD model.

595 For the 100-year return period, peak flow values of Version 2 are on average 35% lower than
 596 Version 1 in Hungary, and 16% lower in England. However, similar decreases are also observed
 597 for the 1-in-2-year peak discharge that determines full-bank discharge. The resulting reduction in
 598 channel hydraulic conveyance in respect to Version 1 is likely to offset the decrease of peak flood
 599 volumes, which explain the small difference in overall flood extent given by the F2/F1 parameter
 600 in Table 8. Such results confirm the low sensitivity of the modelling framework to the
 601 hydrological input observed by Dottori et al. (2016) and by Trigg et al (2016) for a global-scale
 602 application. This low sensitivity is likely to offset the uncertainty related to the estimation of peak
 603 flow values reported in Appendix B. The results also confirm that the knowledge of river channel
 604 geometry is crucial to correctly model the actual channel conveyance and thus improve inundation
 605 modelling. Other differences in input data are given by minor changes in Manning’s parameters
 606 and in the EFAS river network, which might contribute to the observed differences.

607 *3.5 Influence of elevation data*

608 Table 9 compares the metrics calculated with CCM DEM elevation data against the same metrics
 609 for the modelled flood maps based on MERIT-DEM. The comparison is carried out for England,
 610 Hungary and the Po river basin. Performance is slightly improved by the use of MERIT-DEM
 611 data for all areas and return periods, in particular through the reduction of FAR, even though the
 612 overall increase of CSI values is limited to few percentage points.

613

614 *Table 9. Comparison of performances of the flood hazard maps described in the present study*
 615 *and developed by Dottori et al. (2016a) based on the MERIT-DEM (a) and CCM-DEM (b). Table*
 616 *reports the ratio between flood extents F and the differences for HR, FAR, and CSI (e.g. $(HRa-$*
 617 *$HRb)/HRa$).*

	RP (years)	ΔF	ΔHR	ΔFAR	ΔCSI
Hungary	100	-5.3%	0.0%	-2.0%	5.1%
Hungary	1000	-5.9%	-0.1%	-7.6%	5.2%
England	100	0.0%	2.6%	-5.7%	3.8%
England	1000	1.7%	2.8%	-7.8%	3.2%
Po	500	0.2%	0.9%	-4.3%	3.4%

618

619 Because of this limited improvement and the considerable amount of time required to re-run the
620 complete set of flood hazard maps (several days for each return period) it was decided not to
621 update the flood maps using the MERIT-DEM as elevation data. Moreover, new high-resolution
622 datasets such as the Copernicus DEM (ESA-Airbus 2019), the 90m version of TanDEM-X dataset
623 (<https://geoservice.dlr.de/web/dataguide/tdm90>), and MERIT-HYDRO (Yamazaki et al., 2019)
624 have recently become available, and therefore future research could focus on performing
625 additional comparisons to identify which dataset is most suitable for inundation modelling in
626 Europe.

627

628 *4) Conclusions and ongoing work*

629 We presented here a new dataset of flood hazard maps covering the geographical Europe and
630 including large parts of the Middle East and river basins entering the Mediterranean Sea. This
631 dataset significantly expands the previous available flood maps datasets at continental scale
632 (Alfieri et al., 2014; Dottori et al., 2016a), and therefore constitutes a valuable source of
633 information for future research studies and flood management, especially for countries where no
634 official flood hazard maps are available. The new maps also benefit from updated models and
635 new calibration and meteorological data. The maps are being used for a range of applications at
636 continental scale, from evaluating present and future river flood risk scenarios, to the cost-benefit
637 assessment of different adaptation strategies to reduce flood impacts, and for comparisons
638 between different regions, countries and river basins (Dottori et al, 2020b). Moreover, the flood
639 hazard maps are designed to be integrated with the Copernicus European Flood Awareness
640 System (EFAS), and will be used to perform operational flood impact forecasting in EFAS
641 (Dottori et al., 2017).

642 We performed a detailed validation of the modelled flood maps in several European countries
643 against official flood hazard maps. The resulting validation exercise is the most complete
644 undertaken so far for Europe to our best knowledge, and provided a comprehensive overview of
645 the strengths and limitations of the new maps. Nevertheless, the unavailability of reference flood
646 maps outside Europe did not allow any validation in the arid regions in North Africa and Eastern
647 Mediterranean. In these areas, further research will be needed to better understand the
648 performance of the flood mapping procedure here proposed. Modelled maps generally achieve

649 low scores for high and medium probability of flooding. For the 1-in-100-year return period, the
650 modelled maps can identify on average two-thirds of reference flood extent, however they also
651 largely overestimate flood-prone areas in many regions, thus hampering the overall performance.
652 Performances improves markedly with the increase of return period, mostly due to the decrease
653 of the false alarm rates. In particular, critical success index (CSI) values approach and in some
654 cases exceed 0.5 for return periods equal or above 500 years, meaning that the maps can correctly
655 identify more than half of flooded areas in the main river stems and tributaries of different river
656 basins.

657 It is important to note that the validation was affected by problems in identifying the correct areas
658 for a fair comparison, because of the different density of the mapped river network in reference
659 and modelled maps. In our study we used large buffers to constrain comparison areas, which
660 possibly penalized the model performance by generating spurious false alarms in areas not
661 considered by official maps. However, we observed that the proposed maps achieve comparable
662 results to other large-scale flood models when using similar parameters for the validation.

663 The low skill of modelled maps for high and medium probability of flooding, with large
664 overestimations observed in different lowland areas, is likely motivated by the non-inclusion of
665 flood defences in the modelling framework and the simplified representation of channel hydraulic
666 conveyance, due to the absence of datasets at European scale describing river channels and
667 defence structures (i.e. design standards and location of dyke systems). Such information
668 combined with high-resolution DEM fed with local-scale information (artificial and defence
669 structures) is crucial to improve the performance of large-scale flood models and apply more
670 realistic flood modelling tools, as observed also by Wing et al (2017, 2019b). Uncertainty in peak
671 flow estimation can also influence the skill of the modelled maps; however, we found that the
672 limited sensitivity of the modelling approach to changes in the hydrological input smooths out
673 this uncertainty source, because channel conveyance is linked to streamflow characteristics. Such
674 finding highlight the need for independent data of river channel width, shape and depth to better
675 reproduce streamflow and flooding processes. Moreover, the improved results offered by the use
676 of the MERIT-DEM elevation data suggest that recent high-resolution datasets such as the
677 Copernicus DEM (ESA-Airbus 2019), TanDEM-X
678 (<https://geoservice.dlr.de/web/dataguide/tm90>), and MERIT-HYDRO (Yamazaki et al., 2019)

679 may offer a viable solution to improve future versions of continental-scale flood hazard maps in
680 Europe.

681 Increasing map coverage by including the minor river network is likely to improve the skill of
682 modelled maps. However, this might require the use of a different modelling approach to account
683 for pluvial flooding (Wing et al., 2017; Bates et al., 2021), along with reliable model climatology
684 to represent small-scale precipitation processes. Improving the simulation of reservoirs may also
685 reduce the difference between the real and modelled hydrological regimes in regions such as the
686 Iberian Peninsula and the Alps.

687 *Data availability*

688 The dataset described in this manuscript is accessible as part of the data collection “Flood Hazard
689 Maps at European and Global Scale” at the JRC Data Catalogue
690 (<https://data.jrc.ec.europa.eu/collection/id-0054>).

691 Please refer to the dataset as follows: Dottori F., Bianchi A., Alfieri, L., Skoien, J., Salamon P.:
692 River flood hazard maps for Europe and the Mediterranean Basin. JRC Data Catalogue,
693 <http://data.europa.eu/89h/1d128b6c-a4ee-4858-9e34-6210707f3c81> , doi: 10.2905/1D128B6C-
694 A4EE-4858-9E34-6210707F3C81, 2020.

695 The dataset comprises the following maps (eight in total), each one available as a raster (Geotiff)
696 file:

- 697 • Map of permanent water bodies for Europe and the Mediterranean Basin
- 698 • River network in Europe and the Mediterranean Basin
- 699 • River flood hazard maps for Europe and the Mediterranean Basin (return periods of 10, 20,
700 50, 100, 200 and 500 years)

701 The official flood hazard maps used for the validation exercise are freely accessible at the
702 following web-sites:

- 703 • Spain: <https://www.miteco.gob.es/es/cartografia-y-sig/ide/descargas/agua/zi-lamina.aspx> (in
704 Spanish);
- 705 • Po River Basin: <https://pianoalluvioni.adbpo.it/mappe-del-rischio-2/download-mappe/> (in
706 Italian);
- 707 • Norway: <https://www.nve.no/flaum-og-skred/kartlegging/flaum/> (in Norwegian);

- 708 • England: [https://data.gov.uk/dataset/bed63fc1-dd26-4685-b143-2941088923b3/flood-map-](https://data.gov.uk/dataset/bed63fc1-dd26-4685-b143-2941088923b3/flood-map-for-planning-rivers-and-sea-flood-zone-3)
709 [for-planning-rivers-and-sea-flood-zone-3](https://data.gov.uk/dataset/bed63fc1-dd26-4685-b143-2941088923b3/flood-map-for-planning-rivers-and-sea-flood-zone-3) ; [https://data.gov.uk/dataset/cf494c44-05cd-4060-](https://data.gov.uk/dataset/cf494c44-05cd-4060-a029-35937970c9c6/flood-map-for-planning-rivers-and-sea-flood-zone-2)
710 [a029-35937970c9c6/flood-map-for-planning-rivers-and-sea-flood-zone-2](https://data.gov.uk/dataset/cf494c44-05cd-4060-a029-35937970c9c6/flood-map-for-planning-rivers-and-sea-flood-zone-2) (in English)
711 • Hungary: <https://www.vizugy.hu/index.php?module=content&programelemid=62> (in
712 Hungarian)
713 The LISFLOOD hydrological model used in this research is released as open-source software and
714 available at <https://ec-jrc.github.io/lisflood/> (last access on 14 February 2022).
715 The streamflow dataset derived from the long-term run of the LISFLOOD model is available at
716 <https://cds.climate.copernicus.eu/cdsapp#!/dataset/efas-historical> .
717 The LISFLOOD-FP hydrodynamic model used in this research is available as open-source
718 software at <https://www.seamlesswave.com/LISFLOOD8.0> for research and non-commercial
719 purposes.
720 All links have been accessed on 14 February 2022.

721 *Appendix A: Meteorological observations used for LISFLOOD simulations*

722 The long-term run of the hydrological model LISFLOOD is based on observed data from
723 meteorological stations and precipitation datasets, which are collected and continuously expanded
724 as part of the development work for EFAS. The meteorological variables considered are:
725 precipitation, minimum and maximum temperature, wind speed, solar radiation and vapour
726 pressure. The number of stations with available meteorological observations depends on the
727 period and variable considered, with an increasing availability towards the end of the historical
728 simulation period. As an example, for the year 2016 the number of daily observations available
729 ranged from ~8.800 for temperature to ~5.500 for precipitation and ~3.700 for vapour pressure.
730 The input from meteorological stations is completed by a number of precipitation datasets
731 (EURO4M-APG, INCA-Analysis Austria, ERA-Interim GPCP corrected and Carpat-Clim; for
732 details see Arnal et al., 2019). Note that the same datasets are used to drive the LISFLOOD
733 calibration and to calculate the initial conditions for the EFAS forecasts. The data from
734 meteorological stations and gridded datasets were then interpolated using the interpolation
735 scheme SPHEREMAP to produce meteorological grids with a daily time step. The reader is
736 referred to Arnal et al. (2019) for further details.
737

738 *Appendix B: Calibration and validation of hydrological components*

739 *B1: LISFLOOD calibration and validation results*

740 We report here an overview of the results of the LISFLOOD calibration and validation presented
 741 by Arnal et al. (2019). The skill of LISFLOOD in reproducing observed flow regimes
 742 (hydrological skill) is expressed using two indices, the Kling-Gupta Efficiency (KGE; Gupta et
 743 al., 2009) and the Nash-Sutcliffe Efficiency (NSE; Nash and Sutcliffe, 1970). The NSE index is
 744 widely applied in literature and is useful to measure the hydrological skill under high-flow
 745 conditions, given its sensitivity to flow extremes (Krause et al., 2005). The KGE index provides
 746 a more complete evaluation of the model skill under variable flow conditions, and is therefore
 747 useful for calibration purposes (Gupta et al., 2009; Knoben et al., 2019)

748 Table B1 summarizes the results of KGE and NSE indices, and Figure B1 shows the spatial
 749 distribution of the KGE index values across the EFAS domain. The spatial distribution of NSE is
 750 roughly similar. For a detailed list of scores for all stations, please refer to Arnal et al. (2019).

751

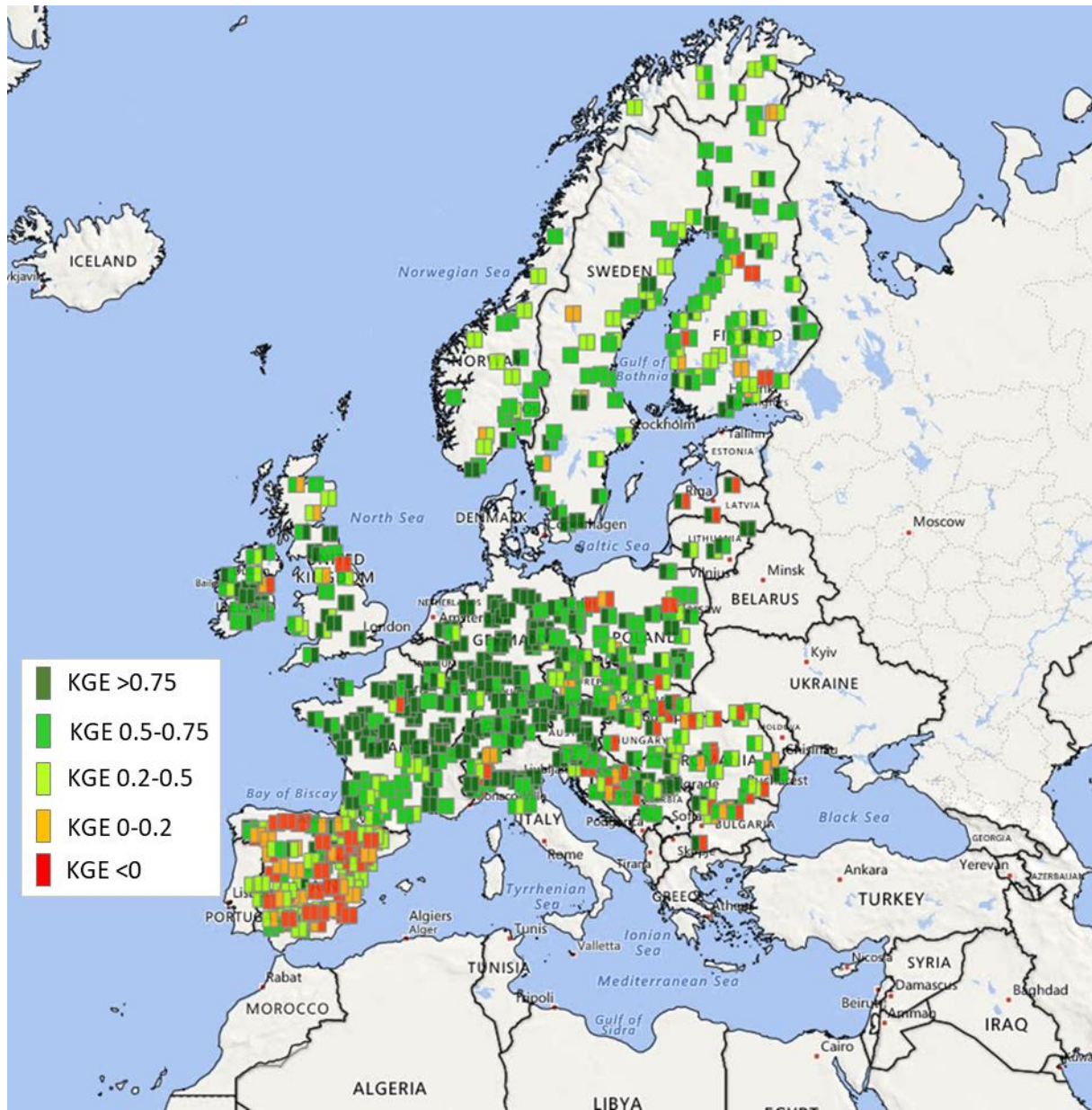
752 *Table B1. Overview of the hydrological skill of LISFLOOD for the calibration and validation*
 753 *stations.*

NSE	calibration		validation		KGE	calibration		validation	
	no. of stations	[%]	no. of stations	[%]		no. of stations	[%]	no. of stations	[%]
> 0.75	147	21%	101	14%	> 0.75	303	42%	174	24%
> 0.5–0.75	277	39%	207	30%	> 0.5–0.75	240	33%	235	33%
> 0.2–0.5	165	23%	171	25%	> 0.2–0.5	91	13%	172	24%
> 0–0.2	35	5%	65	9%	> 0–0.2	36	5%	44	6%
≤0	93	13%	153	22%	≤0	47	7%	73	10%
	Σ 717		Σ 698			Σ 717		Σ 698	

754

755 As can be seen from Table B1, 75 % of all stations scored a KGE higher than 0.5 during
 756 calibration, and 57 % during validation. NSE index values above 0.5 are scored for 60% and 44%
 757 of stations, respectively for the calibration and validation periods.

758 It is clearly noticeable that the skill is not homogeneously distributed across Europe, with higher
759 skills in large parts of Central Europe, and lower skill mostly in Spain caused by the strong
760 influence of reservoirs and flow control structures. The other study areas considered in the
761 validation exercise (England, Hungary, Norway, Po river basin) exhibit KGE and NSE values
762 generally above 0.5.
763



764

765 *Figure B1. Hydrological skill of EFAS at the calibration locations. Colour coding denotes the*
 766 *quality of the KGE during calibration (left half of square) and validation (right half of the square).*
 767 *Adapted from Arnal et al. (2019).*

768

769

770

771 **B2: performance of the extreme value analysis**

772 Here we evaluate the performance of the Gumbel distribution in fitting the available reference
 773 discharge values (26 annual maxima calculated for all the grid points of the LISFLOOD long-
 774 term run). Specifically, we compare the empirical and fitted distributions of streamflow annual
 775 maxima using the Cramer-von Mises test (Anderson, 1962), and we calculate the average
 776 differences between reference and fitted discharge values. Table B2 summarizes the resulting p-
 777 values over the study area. Figure B2 compares empirical and fitted distributions in two locations
 778 of the rivers Rhine and Danube.

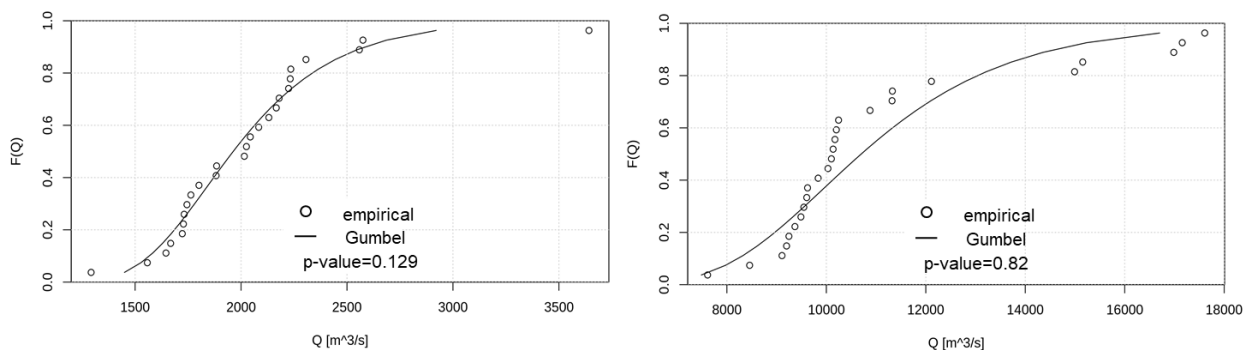
779

780 *Table B2. Overview of the performance of the Gumbel distribution calculated with the Cramer-
 781 Vin Mises criterion.*

P value	% LISFLOOD points
<0.1	5%
0.1–0.25	6%
0.25–0.5	14%
0.5–0.75	23%
>0.75	52%

782

783 *Figure B2. Comparison of the empirical and fitted distributions of annual discharge maxima at
 784 selected locations of the rivers Rhine (left) and Danube (right).*



785

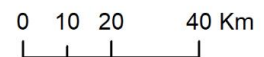
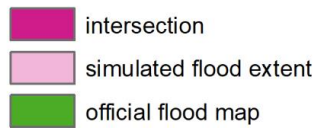
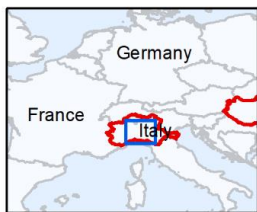
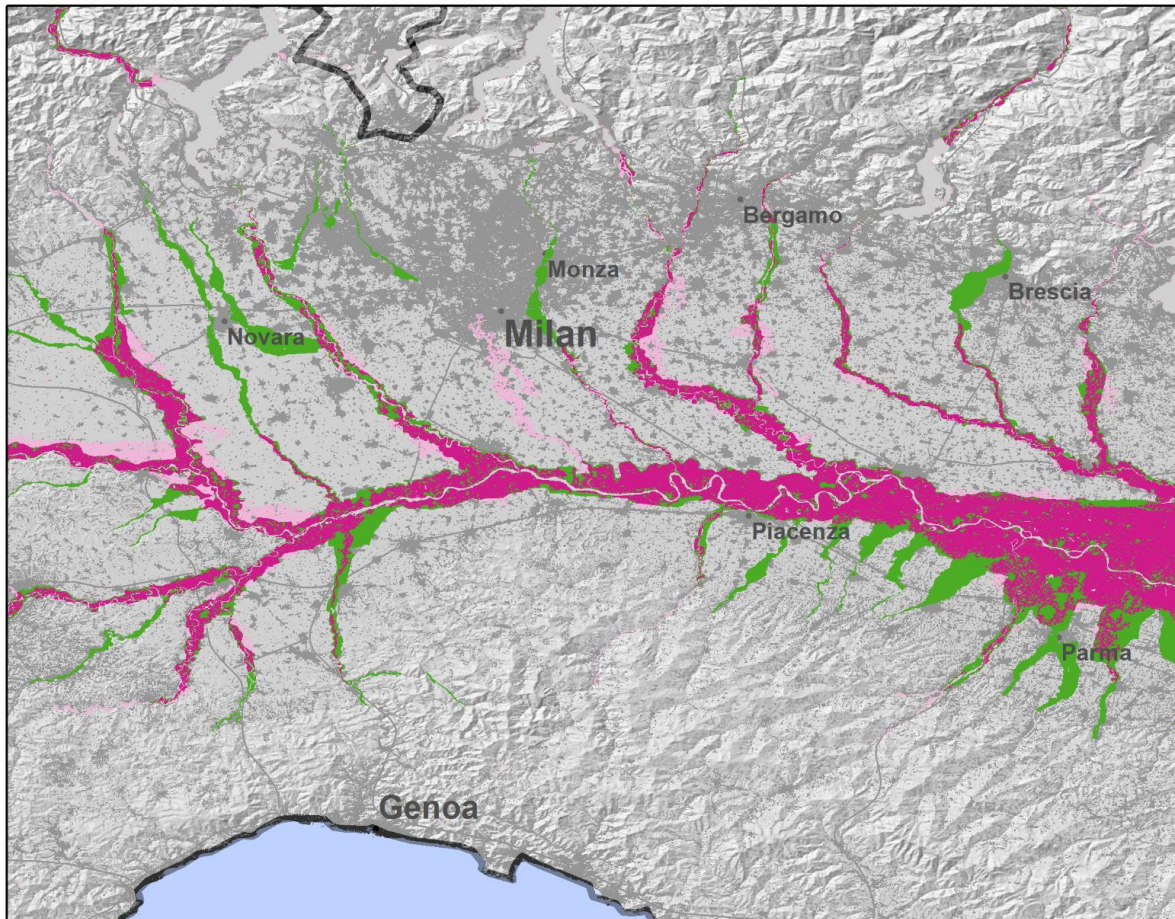
786 P-values in Table B2 suggest a low skill of the fitted Gumbel distributions; however the resulting
 787 uncertainty in the estimates of discharge maxima is generally below 25%, as in the examples

788 shown in Figure B2. This is considered acceptable because the reference discharge maxima are
789 modelled and not observed values. Due to limited sample size, it is not possible to evaluate the
790 extrapolation error for peak flows beyond the available sample; however, previous studies
791 suggested the suitability of the Gumbel distribution. Cunnane (1989) stated that the Gumbel
792 distribution is effective for small sample sizes, whereas the Generalized Extreme Value (GEV)
793 distribution shows a better overall performance if the size is greater than 50. More recently,
794 Papalexiou and Koutsoyiannis (2013) found similar results for extreme precipitation values. In
795 particular, they demonstrated that short record lengths affects the estimation the GEV shape
796 parameter, and thus the choice between a two-parameter (Gumbel) and a three-parameter GEV.
797 Di Baldassarre et al. (2009) observed that the Gumbel distribution might estimate flood extremes
798 with high return periods (e.g. 100-year) with smaller errors than other distributions, if the
799 available sample size is small. Further research could use longer observed streamflow series to
800 compare different extreme value distributions across European regions, similarly to what done by
801 Villarini and Smith (2010) for the eastern United States and Rahman et al. (2013) for Australia.
802

803 *Appendix C: Additional results*

804 *C1: validation of the hazard maps for the Po River Basin*

805 According to Table 3, the modelled flood maps provide a better reproduction of reference maps
806 for the Po River, compared to other study areas. False alarms are low, while hit ratio (HR) values
807 indicate that two out of every three pixels in the reference map are correctly identified as flooded.
808 The analysis of reference and modelled maps (Figure C1), suggests that the underestimation is
809 partly caused by flooded areas along some tributaries which are not included in modelled maps.
810 Other areas with omission errors are located near confluences of the Po main stem and the major
811 tributaries in Emilia-Romagna, which may depend on the underestimation of peak flow on
812 tributaries. In fact, the results of the LISFLOOD calibration in Figure B1 show better hydrological
813 skill along the Po main stem, compared to some tributaries. Finally, it is likely that the inclusion
814 of smaller tributaries of the river network in the modelled maps would improve the overall
815 performance.
816



817

818 *Figure C1. Comparison of modelled (blue) and reference (green) flood hazard maps (1-in-500-*
 819 *year) over the Po river basin, Italy. Purple areas denotes the intersection (agreement) between*
 820 *the two set of maps.*

821

822 ***C2: validation of the hazard maps for Norway***

823 The results of the modelled flood maps in Norway show a general tendency to overestimate flood
 824 extent for the 1-in-100-year events, with high values for both hit ratio (HR) and false alarm ratio
 825 (FAR). Such a result is in fact largely influenced by the relatively small extent and discontinuous

826 coverage of reference maps. Flood-prone areas for the 1-in-100-year official maps only cover 215
827 km², possibly due to the low density of populated places in Norway, while they cover between
828 4700 and 5700 km² for England, Spain and Hungary. As for Spain, we applied a 5km buffer to
829 restrict the area of comparison around reference maps, yet this leads to spurious overestimation
830 around the edges of reference map polygons. Notably, the performance improves markedly with
831 the use of a 1km buffer as in Wing et al., (2017), which results in increased critical success index
832 (CSI) scores up to nearly 0.50.

833 The results of reported by Arnal et al. (2019) and summarized in Figure B1 suggest an acceptable
834 hydrological skill of the LISFLOOD calibration in Norway, with a majority of gauge stations
835 scoring KGE values above 0.5. In the areas with lower scores, the model performance for low-
836 probability flood events might be influenced by an incorrect estimation of peak discharges driven
837 by snow melt, which plays a relevant role in determining low-probability flood events.

838

839 *C3: Influence of correcting elevation data with land use*

840 We tested the results of correcting CCM DEM elevation data with vegetation cover in
841 Scandinavia, where the percentage of land covered by forests is more relevant than in the other
842 regions included in the modelled flood maps. For the 1-in100-year flood maps, the overall
843 difference in flood extent between the corrected and uncorrected maps is less than 4%, and similar
844 values were found for the other return periods. Moreover, the HR, FAR and CSI values of two
845 set of maps differ by less than 2% when calculated against the 1-in100-year official map in
846 Norway, probably because forested areas have not been considered as relevant flood-prone areas.
847 These results suggest that the simulation of densely vegetated areas have a limited importance in
848 determining the overall performance of modelled flood maps in Europe.

849 *Author contribution*

850

851 FD: conceptualization, formal analysis, investigation, data curation writing (original draft, review
852 and editing); LA: methodology, investigation, writing (review and editing); AB: data curation,
853 validation, visualization; JS: investigation, writing (review and editing); PS: conceptualization,
854 project administration, writing (original draft, review and editing)

855

856 *Competing interests*

857 The authors declare that they have no conflict of interest.

858

859 *Acknowledgements*

860 This study has been partially funded by the COPERNICUS programme and by an administrative
861 arrangement with Directorate General 'European Civil Protection and Humanitarian Aid
862 Operations (DG ECHO) of the European Commission. EFAS is operated and financed as part of
863 the Copernicus Emergency Management Service. The authors would like to thank Niall
864 McCormick for his valuable suggestions on the early versions of the manuscript

865

866

867 *References*

- 868 Alfieri, L., Salamon, P., Bianchi, A., Neal, J., Bates, P.D., and Feyen, L.: Advances in pan-
869 European flood hazard mapping, *Hydrol. Process.* 28 (18), 4928-4937, doi:10.1002/hyp.9947,
870 2014.
- 871 Alfieri L., Feyen L., Dottori F., and Bianchi A.: Ensemble flood risk assessment in Europe
872 under high end climate scenarios. *Glob. Environmental Change* 35, 199–212, 2015.
- 873 Alfieri, L., L. Feyen, and Di Baldassarre, G.: Increasing flood risk under climate change: a
874 pan-European assessment of the benefits of four adaptation strategies, *Clim. Change*, 136(3),
875 507–521, doi:10.1007/s10584-016-1641-1, 2016.
- 876 Anderson, T. W.: On the Distribution of the Two-Sample Cramer–von Mises Criterion.
877 *Annals of Mathematical Statistics.* Institute of Mathematical Statistics, 33 (3), 1148–1159.
878 doi:10.1214/aoms/1177704477, 1962.
- 879 Arnal, L., Asp, S.-S., Baugh, C., de Roo, A., Disperati, J., Dottori, F., Garcia, R., Garcia
880 Padilla, M., Gelati, E., Gomes, G., Kalas, M., Krzeminski, B., Latini, M., Lorini, V., Mazzetti,
881 C., Mikulickova, M., Muraro, D., Prudhomme, C., Rauthe-Schöch, A., Rehfeldt, K., Salamon,
882 P., Schweim, C., Skoien, J.O., Smith, P., Sprokkereef, E., Thiemig, V., Wetterhall, F., and Ziese,
883 M.: EFAS upgrade for the extended model domain – technical documentation, EUR 29323 EN,
884 Publications Office of the European Union, Luxembourg, 2019, ISBN 978-92-79-92881-9,
885 doi:10.2760/806324, JRC111610, 2019.
- 886 Autorita` di bacino del fiume Po (AdB Po): Progetto di Variante al PAI: mappe della
887 pericolosita` e del rischio di alluvione (in Italian), [https://pianoalluvioni.adbpo.it/progetto-](https://pianoalluvioni.adbpo.it/progetto-esecutivodelleattivita/)
888 [esecutivodelleattivita/](https://pianoalluvioni.adbpo.it/progetto-esecutivodelleattivita/), last access 3 April 2020, 2012.
- 889 Bates, P. D., and De Roo, A. P. J.: A simple raster-based model for flood inundation
890 simulation, *J. Hydrol.*, 236 (1–2), 54–77, 2000.
- 891 Bates, P.D., Horritt, M.S., and Fewtrell, T.J.: A simple inertial formulation of the shallow
892 water equations for efficient two-dimensional flood inundation modelling. *J. Hydrol.*, 387, 33-45,
893 2010.
- 894 Bates, P. D., Quinn, N., Sampson, C., Smith, A., Wing, O., Sosa, J., et al. (2021). Combined
895 modeling of US fluvial, pluvial, and coastal flood hazard under current and future climates. *Water*
896 *Resour. Res.*, 57, e2020WR028673, <https://doi.org/10.1029/2020WR028673>

897 Baugh, C. A., Bates, P. D., Schumann G., and Trigg, M.A.: SRTM vegetation removal and
898 hydrodynamic modeling accuracy, *Water Resour. Res.* 49, 5276–5289, doi:10.1002/wrcr.20412,
899 2013.

900 Barredo, J.I., de Roo, A., and Lavallo, C.: Flood risk mapping at European scale. *Water*
901 *Science and Technology* 56, 11–17, 2007.

902 Bontemps, S., et al.: GLOBCOVER -Products description and validation report, Univ.
903 Catholique de Louvain, 2009.

904 Burek, P., van der Knijff, J., and de Roo, A.: LISFLOOD, Distributed Water Balance and
905 Flood Simulation Model Revised User Manual 2013. Publications Office, Luxembourg, 2013

906 Copernicus Land Monitoring Service. Corine Land Cover. [http://land.copernicus.eu/pan-](http://land.copernicus.eu/pan-european/corine-land-cover)
907 [european/corine-land-cover](http://land.copernicus.eu/pan-european/corine-land-cover), last access on 12 February 2020.

908 Cunnane, C.: Statistical Distributions For Flood Frequency Analysis. Operational Hydrology
909 Report no. 33, World Meteorological Organization, 1989.

910 Di Baldassarre, G., Laio, F., and Montanari, A.: Design flood estimation using model
911 selection criteria. *Physics and Chemistry of the Earth* 34, 606–611, 2008.

912 Dottori, F., Alfieri, L., Salamon, P., Bianchi, A., Feyen, L., and Lorini, V.: Flood hazard map
913 for Europe - 100-year return period. European Commission, Joint Research Centre (JRC)
914 [Dataset] PID: http://data.europa.eu/89h/jrc-floods-floodmapeu_rp100y-tif, 2016a.

915 Dottori, F., Salamon, P., Bianchi, A., Alfieri, L., Hirpa, F.A., and Feyen, L.: Development
916 and evaluation of a framework for global flood hazard mapping. *Adv. Water Resour.* 94, 87–102,
917 2016b

918 Dottori, F., Kalas, M., Salamon, P., Bianchi, A., Alfieri, L., and Feyen, L.: An operational
919 procedure for rapid flood risk assessment in Europe. *Nat. Hazards Earth Syst. Sci.*, 17, 1111-
920 1126, <https://doi.org/10.5194/nhess-17-1111-2017>, 2017.

921 Dottori, F., Szewczyk, W., Ciscar, J.C., Zhao, F., Alfieri, L., Hirabayashi, Y., Bianchi, A.,
922 Frieler, K., Betts, R.A., Feyen, L.: Increased human and economic losses from river floods with
923 anthropogenic warming. *Nat. Clim. Change* 8(9), 781-786, [https://doi.org/10.1038/s41558-018-](https://doi.org/10.1038/s41558-018-0257-z)
924 [0257-z](https://doi.org/10.1038/s41558-018-0257-z), 2018.

925 Dottori, F., Bianchi, A., Alfieri, L., Skoien, J., and Salamon P.: River flood hazard maps for
926 Europe and the Mediterranean Basin region, doi:10.2905/1D128B6C-A4EE-4858-9E34-
927 6210707F3C81, 2020.

928 Dottori, F., Mentaschi, L., Bianchi, A., Alfieri, L., and Feyen, L.: Adapting to rising river
929 flood risk in the EU under climate change, EUR 29955 EN, Publications Office of the European
930 Union, Luxembourg, 2020, ISBN 978-92-76-12946-2 , doi:10.2760/14505, JRC118425, 2020b.

931 European Commission (EC): Directive 2007/60/EC of the European Parliament and
932 of the Council on the assessment and management of flood risks, Official Journal of
933 the European Communities, Brussels, [http://eur-lex.europa.eu/legal-](http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32007L0060)
934 [content/EN/TXT/?uri=CELEX%3A32007L0060](http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32007L0060), last access on 13 May 2020, 2007.

935 ESA-Airbus: Copernicus Digital Elevation Model Validation Report,
936 [https://spacedata.copernicus.eu/documents/12833/20611/GEO1988-CopernicusDEM-RP-](https://spacedata.copernicus.eu/documents/12833/20611/GEO1988-CopernicusDEM-RP-001_ValidationReport_V1.0/9bc5d392-c5f2-4118-bd60-db9a6ea4a587)
937 [001_ValidationReport_V1.0/9bc5d392-c5f2-4118-bd60-db9a6ea4a587](https://spacedata.copernicus.eu/documents/12833/20611/GEO1988-CopernicusDEM-RP-001_ValidationReport_V1.0/9bc5d392-c5f2-4118-bd60-db9a6ea4a587), last access on 14 May
938 2020, 2019.

939 Feyen, L., Dankers, R., Bódis, K., Salamon, P., and Barredo, J.I.: Fluvial flood risk in Europe
940 in present and future climates. *Climatic Change* 112(1), 47–62, doi:10.1007/s10584-011-0339-7,
941 2012.

942 Fleischmann, A., Paiva, R. and Collischonn, W.: Can regional to continental river
943 hydrodynamic models be locally relevant? A cross-scale comparison, *J. Hydrol. X*,
944 <https://doi.org/10.1016/j.hydroa.2019.100027>, 2019.

945 Gupta, H.V., Kling, H., Yilmaz, K.K. and Martinez, G.F.: Decomposition of the mean squared
946 error and NSE performance criteria: implications for improving hydrological modelling. *J.*
947 *Hydrol.*, 377, 80-91, 2009.

948 Hirpa, F.A., Salamon, P., Beck, H.E., Lorini, V., Alfieri, L., Zsoter, E., and Dadson, S.J. :
949 Calibration of the Global Flood Awareness System (GloFAS) using daily streamflow data. *J.*
950 *Hydrol.*, 566, 595–606, 2018.

951 Jongman, B., Hochrainer-Stigler, S., Feyen, L., Aerts, J.C.J.H., Mechler, R., Botzen, W.J.W.,
952 Bouwer, L.M., Pflug, G., Rojas, R., and Ward, P.J.: Increasing stress on disaster-risk finance due
953 to large floods. *Nat. Clim. Change* 4, 264-268, doi:http://dx.doi.org/10.1038/nclimate2124, 2014.

954 Knoben, W. J. M., Freer, J. E., and Woods, R. A.: Technical note: Inherent benchmark or not?
955 Comparing Nash–Sutcliffe and Kling–Gupta efficiency scores, *Hydrol. Earth Syst. Sci.*, 23,
956 4323–4331, <https://doi.org/10.5194/hess-23-4323-2019>, 2019.

957 Krause, P., Boyle, D. P., and Bäse, F.: Comparison of different efficiency criteria for
958 hydrological model assessment, *Advances in Geosciences*, European Geosciences Union, 5, 89-
959 97, 2005.

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

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

965 Maione, U., Mignosa, P., and Tomirotti, M.: Regional estimation of synthetic design
966 hydrographs. *International Journal of River Basin Management*, 1(2), 151-163, 2003.

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

970 Ministerio de Medio Ambiente y Medio Rural y Marino (MITECO):. Guía Metodologica para
971 el desarrollo del sistema nacional de cartografía de zonas inundables (in Spanish).
972 [https://www.miteco.gob.es/es/agua/temas/gestion-de-los-riesgos-de-inundacion/snczi/Guia-](https://www.miteco.gob.es/es/agua/temas/gestion-de-los-riesgos-de-inundacion/snczi/Guia-metodologica-determinacion-zonas-inundables/default.aspx)
973 [metodologica-determinacion-zonas-inundables/default.aspx](https://www.miteco.gob.es/es/agua/temas/gestion-de-los-riesgos-de-inundacion/snczi/Guia-metodologica-determinacion-zonas-inundables/default.aspx), last access on 18 May 2020, 2011.

974 Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I – A
975 discussion of principles, *J. Hydrol.*, 10, 282–290, [https://doi.org/10.1016/0022-1694\(70\)90255-](https://doi.org/10.1016/0022-1694(70)90255-6)
976 6, 1970.

977 The Norwegian Water Resources and Energy Directorate,: Flood Zone Maps (in Norwegian),
978 <https://www.nve.no/flaum-og-skred/kartlegging/flaum/> , last access on 24 April 2020.

979 Papalexiou, S. M., and Koutsoyiannis, D.: Battle of extreme value distributions: A global
980 survey on extreme daily rainfall, *Water Resour. Res.* 49, doi:10.1029/2012WR012557,
981 2013.Paprotny, D., Morales-Nápoles, O., and Jonkman, S. N.: Efficient pan-European river flood
982 hazard modelling through a combination of statistical and physical models, *Nat. Hazards Earth*
983 *Syst. Sci.*, 17, 1267-1283, <https://doi.org/10.5194/nhess-17-1267-2017>, 2017.

984 Rahman, A. S., Rahman, A., Zaman, M.A., Haddad, K., Ahsan, A., Imteaz, M.: A study on
985 selection of probability distributions for at-site flood frequency analysis in Australia. *Nat.*
986 *Hazards* 69, 1803–1813 , doi:10.1007/s11069-013-0775-y, 2013.

987 Sampson, C. C., Smith, A. M., Bates, P. D., Neal, J. C., Alfieri, L., and Freer, J. E.: A high-
988 resolution global flood hazard model. *Water Resour. Res.* 51(9), 7358-7381, 2015.

989 Scussolini, P., Aerts, J. C. J. H., Jongman, B., Bouwer, L. M., Winsemius, H. C., de Moel, H.,
990 and Ward, P. J.: FLOPROS: an evolving global database of flood protection standards, *Nat.*
991 *Hazards Earth Syst. Sci.*, 16, 1049–1061, <https://doi.org/10.5194/nhess-16-1049-2016>, 2016.

992 Shaw, J., Kesserwani, G., Neal, J., Bates, P., and Sharifian, M. K.: LISFLOOD-FP 8.0: the
993 new discontinuous Galerkin shallow-water solver for multi-core CPUs and GPUs, *Geosci. Model*
994 *Dev.*, 14, 3577–3602, <https://doi.org/10.5194/gmd-14-3577-2021>, 2021.

995 Thielen J., Bartholmes J., Ramos M.H., and De Roo A.: The European flood alert system -
996 part 1: concept and development. *Hydrol. Earth Syst. Sci.* 13, 125-140, 2009.

997 Trigg, M. et al.: The credibility challenge for global fluvial flood risk analysis. *Environ. Res.*
998 *Lett.* 11 094014, <http://dx.doi.org/10.1088/1748-9326/11/9/094014>, 2016.

999 United Nations Office for Disaster Risk Reduction (UNISDR): Sendai Framework for
1000 Disaster Risk Reduction 2015–2030, www.unisdr.org/we/inform/publications/43291, last access
1001 on 25 May 2020, 2015.

1002 Van der Knijff, J.M., Younis, J., and de Roo, A.P.J.: LISFLOOD: a GIS-based distributed
1003 model for river basin scale water balance and flood simulation. *Int. J. Geogr. Inf. Sci.* 24, 189-
1004 212, 2010.

1005 Villarini, G., and J. A. Smith (2010), Flood peak distributions for the eastern United States,
1006 *Water Resour. Res.*, 46, W06504, doi:10.1029/2009WR008395.

1007 Vogt, J., Soille, P., De Jager, A., Rimaviciute, E., Mehl, W., Foisneau, S., Bodis, K., Dusart,
1008 J., Paracchini, M., Haastруп, P., Bamps, C.: A pan-European River and Catchment Database, EUR
1009 22920 EN. Luxembourg (Luxembourg): OPOCE, JRC40291, doi:10.2788/35907, 2007.

1010 Ward, P.J., Jongman, B., Salamon, P., Simpson, A., Bates, P., De Groeve, T., Muis, S.,
1011 Coughlan de Perez, E., Rudari, R., Trigg, M.A., and Winsemius, H.C.: Usefulness and limitations
1012 of global flood risk models. *Nat. Clim. Change* 5, 712–715, 2015.

1013 Wendi, D., Liong, S.-Y., Sun, Y., and Doan, C.D.: An innovative approach to improve SRTM
1014 DEM using multispectral imagery and artificial neural network. *J. Adv. Model. Earth Syst.* 8,
1015 691–702, 2016.

1016 Wing, O. E., Bates, P. D., Sampson, C. C., Smith, A. M., Johnson, K. A., and Erickson, T. A.:
1017 Validation of a 30 m resolution flood hazard model of the conterminous United States. *Water*
1018 *Resour. Res.* 53(9), 7968-7986, doi:10.1002/2017WR020917, 2017.

1019 Wing, O. E. J., Sampson, C. Bates, P. D., Quinn, N., Smith, A. M., and Neal, J. C.: A flood
1020 inundation forecast of Hurricane Harvey using a continental-scale 2D hydrodynamic model. *J.*
1021 *Hydrol. X* 4, 100039, 2019a.

1022 Wing, O. E. J., Bates, P. D., Neal, J. C., Sampson, C. C., Smith, A. M., Quinn, N., Shustikova,
1023 I., Domeneghetti, A., Gilles, D.W., Goska, R., Krajewskiet W.F.: A New Automated Method for
1024 Improved Flood Defense Representation in Large-Scale Hydraulic Models. *Water Resour. Res.*
1025 55, 11007-11034, <https://doi.org/2019WR025957>, 2019b.

1026 Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J., Sampson,
1027 C., Kanae, S., and Bates, P. D.: A high accuracy map of global terrain elevations. *Geophysical*
1028 *Research Letters*, 2017.

1029 Yamazaki, D., Ikeshima, D., Sosa, J., Bates, P. D., Allen, G. H., & Pavelsky, T. M.: MERIT
1030 Hydro: a high-resolution global hydrography map based on latest topography dataset. *Water*
1031 *Resour. Res.*, 55, 5053–5073. <https://doi.org/10.1029/2019WR024873>, 2019.

1032 Zajac, Z., Zambrano-Bigiarini, M., Salamon, P., Burek, P., Gentile, A., and Bianchi, A.:
1033 Calibration of the LISFLOOD hydrological model for Europe. JRC technical report JRC87717,
1034 2013.