SCIENTIFIC DATA

OPEN

Data Descriptor: GFPLAIN250m, a global high-resolution dataset of Earth's floodplains

F. Nardi[1], A. Annis[1], G. Di Baldassarre[2,3,4], E. R. Vivoni[5,6] & S. Grimaldi[7,8]

Received: 27 June 2018
Accepted: 21 November 2018
Published: 15 January 2019

Identifying floodplain boundaries is of paramount importance for earth, environmental and socioeconomic studies addressing riverine risk and resource management. However, to date, a global floodplain delineation using a homogeneous procedure has not been constructed. In this paper, we present the first, comprehensive, high-resolution, gridded dataset of Earth's floodplains at 250-m resolution (GFPLAIN250m). We use the Shuttle Radar Topography Mission (SRTM) digital terrain model and set of terrain analysis procedures for geomorphic floodplain delineations. The elevation data are processed by a fast geospatial tool for floodplain mapping available for download at https://github.com/fnardi/GFPLAIN. The GFPLAIN250m dataset can support many applications, including flood hazard mapping, habitat restoration, development studies, and the analysis of human-flood interactions. To test the GFPLAIN250m dataset, we perform a consistency analysis with floodplain delineations derived by flood hazard modelling studies in Europe.

Design Type(s)	modeling and simulation objective • process-based data transformation objective
Measurement Type(s)	flood plain
Technology Type(s)	computational modeling technique
Factor Type(s)	geographic location
Sample Characteristic(s)	Earth (Planet)

[1]Water Resources Research and Documentation Centre (WARREDOC), University for Foreigners of Perugia, Perugia, Italy. [2]Department for Earth Sciences, Uppsala University, Uppsala, Sweden. [3]Centre of Natural Hazards and Disaster Science, CNDS, Uppsala, Sweden. [4]IHE Delft Institute for Water Education, Delft, The Netherlands. [5]School of Earth and Space Exploration, Arizona State University, Tempe, AZ, USA. [6]School of Sustainable Engineering and the Built Environment, Arizona State University, Tempe, AZ, USA. [7]Department for Innovation in Biological, Agro-food and Forest systems (DIBAF), University of Tuscia, Viterbo, Italy. [8]New York University Tandon School of Engineering, Department of Mechanical and Aerospace Engineering, New York City, USA. Correspondence and requests for materials should be addressed to F.N. (email: fernando.nardi@unistrapg.it)

Continent	f_p [km^2]	f_{pr} [−]
Europe	806525	0.08
Africa	3853197	0.13
North America	1482914	0.06
South America	2931955	0.16
Asia	3452714	0.08
Oceania	866834	0.10

Table 1. The GFPLAIN250m dataset of Earth's floodplains. The floodplain areas are estimated at the continental scale. Estimated variables are: Total floodplain area or f_p [km^2]; and floodplain area divided by total continental area, the floodplain ratio or f_{pr} [−].

Background & Summary

Floodplains are clearly recognizable from aerial photography by their distinguishable shapes and colors[1]. Riverine areas are not only clearly visible, but are spatially organized following well-known hydrologic and geomorphic properties[2]. Nevertheless, significant uncertainty is associated with existing floodplain delineation methods[3,4]. While floodplain thematic maps are often available, they typically only reflect the context for which they were derived, limiting their broad, multi-sectorial use. For instance, a hydrologic investigation and an aquatic ecology study would likely identify different floodplain extents for the same river corridor depending on the spatiotemporal scale, event or process of interest. To date, a scale-invariant and consistent morphometric zoning of river corridors to identify floodplain landscapes on Earth is still lacking[4].

The aim of this paper is to present the first global floodplain dataset at 8.33 arcsecond resolution that is equivalent at the equator to a 250-m grid cell size. The GFPLAIN250m dataset is derived implementing a unifying framework for fluvial valley zoning. This framework captures the spatial extent of floodplains by implementing geomorphic algorithms able to identify the alluvium extent as a morphometric descriptor of digital terrain models[5–10].

The GFPLAIN250m dataset depicts floodplains as unique and identifiable morphological entities that have been primarily shaped by the accumulated effects of geomorphic and hydrologic processes and secondarily by diffusive biotic processes[6,11]. In such a manner, river basins are dissected into domains of low-lying riparian corridors separated from their surrounding landscapes. This scale-invariant, theoretically-consistent representation of the Earth's floodplains is thus applicable in regions where water-driven erosion and depositional processes govern the morphology of floodplain landscape features. This excludes areas on Earth classified as deserts with low water availability and ice-covered regions with insignificant river flows[12].

Methods

General procedure

The global floodplain map is developed with the GFPLAIN algorithm[6]. Terrain analysis techniques are implemented in GFPLAIN to extract the stream network from a digital terrain model (DTM) of the Earth[13,14]. Each drainage network cell is assigned the maximum potential channel flow depth (h) adopting the power law of equation (1) using the contributing area (A) as a scaling parameter[15,16]. Equation 1 constitutes an adapted version of the Leopold scaling law[15] to represent the proportionality, expressed by the $\propto$ term, between the potential energy associated with floodplain flow shaping process and the river basin morphometric parameter A.

$$h \propto A^b \tag{1}$$

The GFPLAIN algorithm[6,17] produces a gridded floodplain layer by flagging low-lying cells along river corridors. The algorithm recognizes the floodplain extent as formed by those cells, draining to the selected channel location, that are characterized by elevations that are lower than the corresponding maximum channel flow level $H = z + h$, where z is the channel cell elevation obtained from the DTM expressed as absolute elevation in meters above sea level. Figure 1 depicts the three main processing steps of the floodplain identification procedure.

The variation of floodplain flow levels across spatial scales is evaluated by means of the dimensionless b exponent[17] to produce a consistent floodplain zoning analysis (see Technical Validation). The Shuttle Radar Topography Mission (SRTM)[18,19] DTM, provided by the Consortium for Spatial Information (CGIAR-CSI) at 8.33 arcsecond resolution, covering all regions of the world between −60° and 60° of latitude, is used for floodplain delineations of river basins with a contributing area (A) greater than 1000 km^2. This resolution, equivalent to 250 meters at the equator, is consistent with the spatial scale of other global datasets derived in earth, environmental, social and behavioural science applications for depicting fluvial corridor processes and features[20–22].

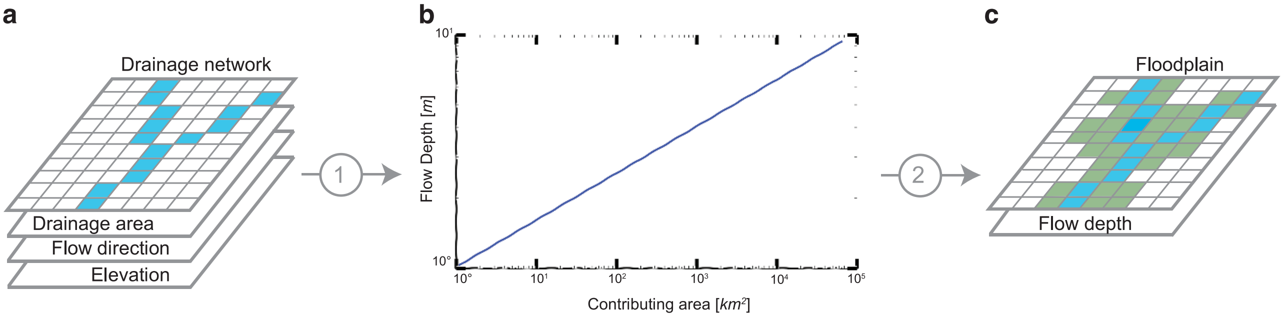

Figure 1. Flow chart describing the DTM analysis and geomorphic scaling law processing for floodplain delineation. Three main steps of the procedure are depicted. (**a**) DTM analysis for flow direction, drainage area and network identification from elevation data. (**b**) Scaling laws implemented for associating a floodplain flow depth to the contributing area of each drainage network grid cell. (**c**) GFPLAIN250m gridded layer is derived by flagging as a floodplain those cells whose elevations are lower than corresponding drainage network flow levels.

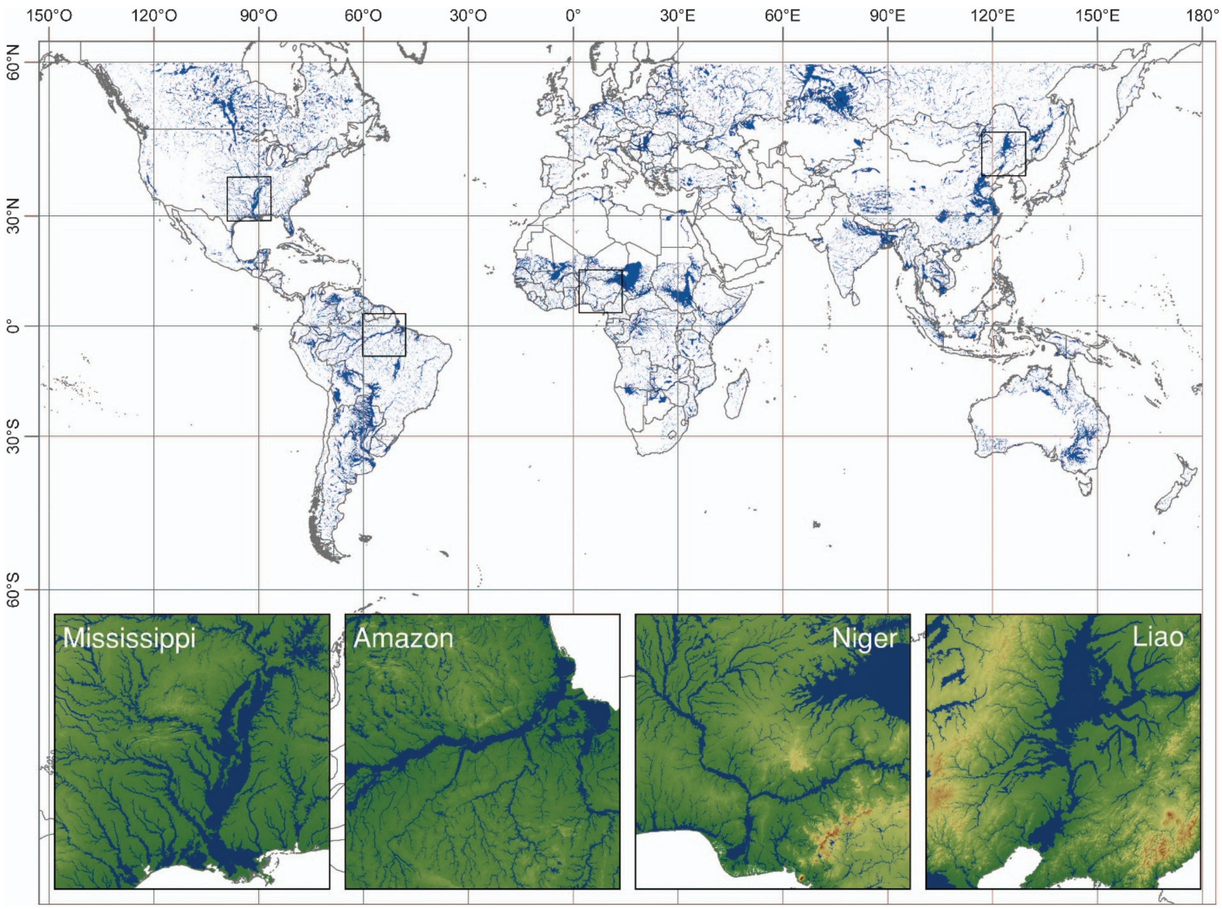

Figure 2. The GFPLAIN250m global floodplain dataset. The GFPLAIN250m is presented in blue color. Insets show floodplains of four major global rivers superimposed on the SRTM dataset.

GFPLAIN algorithm

The GFPLAIN algorithm is organized as a set of Python routines implementing the two main steps of the procedure: (1) Terrain analysis of a DTM for watershed drainage extraction (Fig. 1a), and (2) floodplain delineations (Fig. 1b and c).

The GFPLAIN is a computationally efficient algorithm. Module 2 runs on the order of minutes. Using a standard workstation and the 250-m resolution river network as input, it takes 15 min for delineating

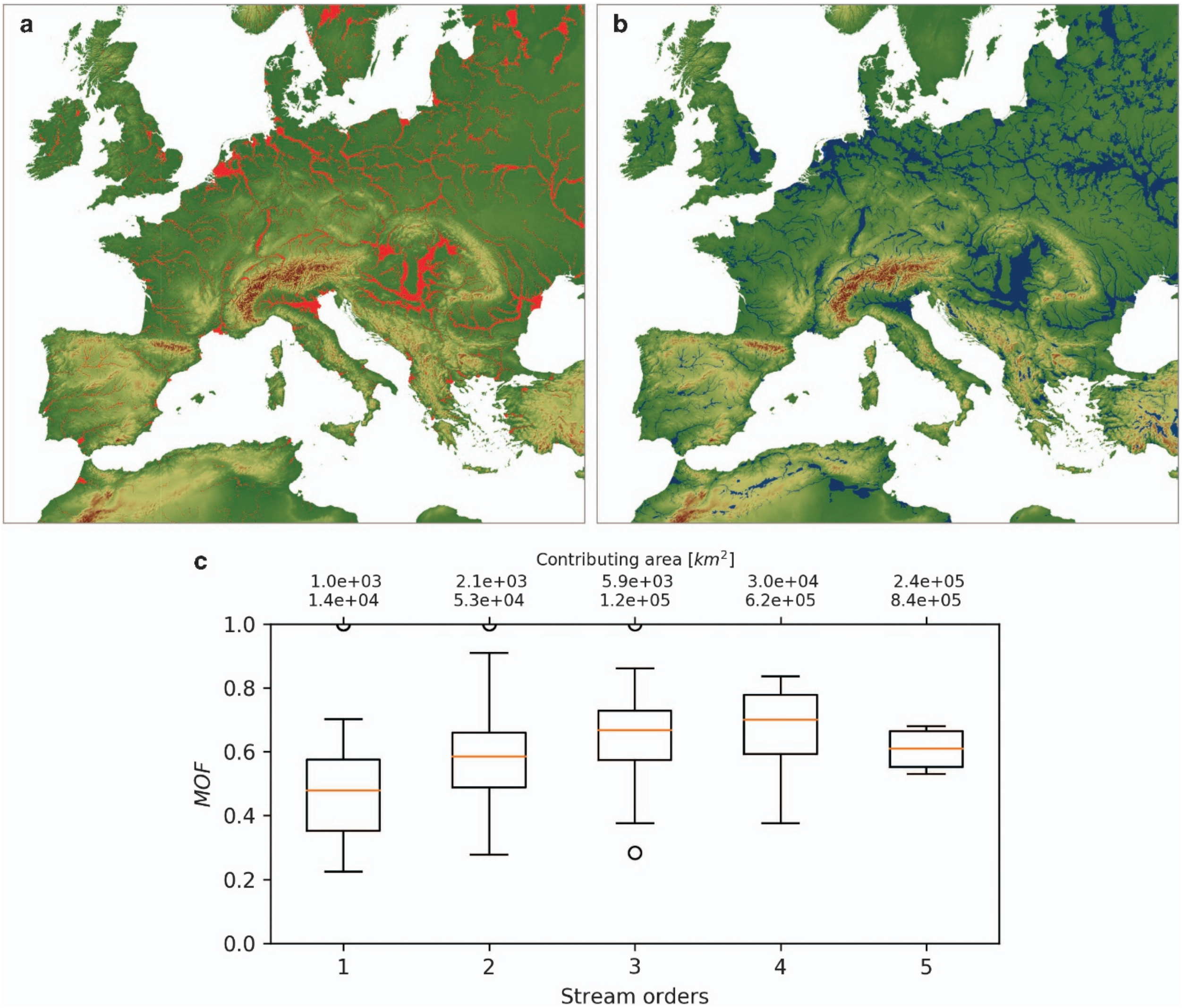

Figure 3. Evaluating the consistency of the geomorphic floodplain model with respect to a flood hazard map. Example of floodplain mapping in Europe using two paradigms. (**a**) Flood hazard event-based mapping using hydraulic simulations of the 200-year synthetic flood design (red color by European Commission, Joint Research Centre). (**b**) Geomorphic floodplain map (blue color). (**c**) Evaluation of the GFPLAIN250m dataset is performed by varying the *b* parameter of the geomorphic scaling law and performing a quantitative comparison with the reference dataset using a measure-of-fit (MOF) index[17]. Box plots represent the statistics of the MOF index obtained by comparing the GFPLAIN250m floodplain zoning with respect to flood hazard zones for European basins of different stream orders.

the entire floodplains of North and South America. This implies that the largest river basins of the world can be analysed in less than 10 min.

Code availability

The Python script and user manual of the GFPLAIN algorithm used for generating the GFPLAIN250m dataset are accessible at https://github.com/fnardi/GFPLAIN with instructions for applications and code reuse.

Data Records

The original SRTM dataset used in this study can be accessed at http://srtm.csi.cgiar.org/ and includes the 250-m SRTM version 4.1 DTM. Figure 2 provides an overview of the dataset, while Table 1 reports a summary of the floodplain mapping for the continents on Earth, except Antarctica.

The GFPLAIN250m dataset can be accessed via figshare (Data Citation 1). Files are stored using both the Esri ASCII raster and the GeoTIFF formats and provided as a seamless dataset using the World Geodetic System 1984 (WGS84) datum and geographic coordinate system. Floodplain raster layers are compressed into a single file zipped for each continent, including the corresponding ASCII or GeoTIFF file. The coding used for each continent and additional information are detailed in the metadata included in the GFPLAIN250m data repository.

Technical Validation

Evaluation of the quality of the GFPLAIN250m dataset is linked to two main factors: (1) the sources of error and potential uncertainties of the DTM processing for drainage network extraction, and (2) the validation of the geomorphic algorithm for floodplain identification.

The first issue refers to sources of error that impact digital terrain data and known assumptions of DTM analysis techniques for earth science applications. Although it is known that DTM resolution and production method may have a direct impact on the outcomes of the stream network extraction[23–25], this uncertainty does not propagate to the geomorphic floodplain zoning considering the simulated channel always flows within the fluvial valley[26]. DTM corrections and the use of updated terrain and hydrologic datasets can mitigate this uncertainty[27,28]. Moreover, the potential sources of error of the river network location and profile do not impact the validity of the GFPLAIN250m dataset considering that it is a topographic data descriptor consistent with other morphometric parameters in river basins[20,21].

For the latter, validation of the geomorphic floodplain algorithm is performed by evaluating the outcomes of the GFPLAIN model to varying parameterization of the scaling law. In particular, the sensitivity of results to varying the b parameter is investigated. The b parameter is varied within a physically feasible range (floodplain flow energy levels within the 10^0–10^2 order of magnitude). The optimal b is associated to floodplain modelling results that maximize the performances of the geomorphic zoning with respect to a reference floodplain dataset. This consistency analysis is developed by quantifying the effect of b value variations on the floodplain zoning behaviour expressed by means of a measure-of-fit index (MOF) based on overlapping, underprediction and overprediction of the floodplain zones[17]. Global fluvial landscape feature zoning is available to depict river channel surface water domains[29,30]. To date, large scale studies delineating floodplain extents using geologic, morphologic or ecologic criteria are not available to benchmark the GFPLAIN250m dataset. Therefore, the 200 years flood prone zoning[31], based on hydrodynamic models, is used as the only available homogeneous floodplain reference dataset at the global scale[32]. The consistency analysis confirms the validity of the GFPLAIN algorithm in capturing the geomorphic signature of fluvial flooding dynamics. MOF value statistics depict consistent floodplain identification behaviour across the geomorphic, climatic and ecologic diversity of European river basins (Fig. 3). Tests confirm that reasonable ranges of MOF values are obtained, with varying b parameters, supporting the use of a constant parameterization at the global scale with $b = 0.30$. As such, the GFPLAIN250m dataset can be used in combination with global datasets of human settlements, to support large-scale studies of human-flood interactions[32–34], human pressure on rivers[35], and changes over time of floodplain and wetland habitats at risk[36–38]. Regional values for the scaling law parametrization can be further refined to capture local variations of geologic, climatic and ecological properties.

References

1. Baynes, E. R. C. *et al.* Erosion during extreme flood events dominates Holocene canyon evolution in northeast Iceland. *PNAS* **112.8,** 2355–2360 (2015).
2. Phillips, C. B. & Jerolmack, D. J. Self-organization of river channels as a critical filter on climate signals. *Science* **352,** 694–697 (2016).
3. Ward, P. J. *et al.* Usefulness and limitations of global flood risk models. *Nat. Clim. Change* **5,** 712–715 (2015).
4. Trigg, M. A. *et al.* The credibility challenge for global fluvial flood risk analysis. *Environ. Res. Lett.* **11,** 094014 (2016).
5. Gallant, J. C. & Dowling, T. I. A multiresolution index of valley bottom flatness for mapping depositional areas. *Water Resour. Res.* **39,** 1347 (2003).
6. Nardi, F., Vivoni, E. R. & Grimaldi, S. Investigating a floodplain scaling relation using a hydrogeomorphic delineation method. *Water Resour. Res.* **42,** W09409 (2006).
7. Dodov, B. A. & Foufoula-Georgiou, E. Floodplain morphometry extraction from a high-resolution digital elevation model: a simple algorithm for regional analysis studies. *IEEE Geosci. Remote Sens. Lett.* **3,** 410–413 (2006).
8. Rennó, C. D. *et al.* HAND, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazonia. *Remote Sens. Environ.* **112,** 3469–3481 (2008).
9. Manfreda, S. *et al.* Investigation on the use of geomorphic approaches for the delineation of flood prone areas. *J. Hydrol.* **517,** 863–876 (2014).
10. Jafarzadegan, K. & Merwade, V. A DEM-based approach for large-scale floodplain mapping in ungauged watersheds. *J. Hydrol.* **550,** 650–662 (2017).
11. Dietrich, W. E. & Perron, J. T. The search for a topographic signature of life. *Nature* **439,** 411–418 (2006).
12. Dinerstein, E. *et al.* An ecoregion-based approach to protecting half the terrestrial realm. *BioScience* **67,** 534–545 (2017).
13. Jenson, S. K. & Domingue, J. O. Software tools to extract topographic structure from digital elevation data for geographic information system analysis. *Photogramm Eng Remote Sensing* **54,** 1593–1600 (1988).
14. Tarboton, D. G., Bras, R. L. & Rodriguez-Iturbe, I. On the extraction of channel networks from digital elevation data. *Hydrol. Process.* **5,** 81–100 (1991).
15. Leopold, L. B. & Maddock, T. The hydraulic geometry of stream channels and some physiographic implications. Report No. 252 (US Government Printing Office, 1953).
16. Bhowmik, N. G. Hydraulic geometry of floodplains. *J. Hydrol.* **68,** 369377–374401 (1984).
17. Nardi, F., Morrison, R. R., Annis, A. & Grantham, T. E. Hydrologic scaling for hydrogeomorphic floodplain mapping: Insights into human-induced floodplain disconnectivity. *River Res. Appl.* **34,** 675–685 (2018).
18. Farr, T. G. *et al.* The Shuttle Radar Topography Mission. *Rev. Geophys.* **45,** RG2004 (2007).
19. Reuter, H. I, Nelson, A. & Jarvis, A. An evaluation of void filling interpolation methods for SRTM data. *Int. J. Geogr. Inf. Sci.* **21,** 983–1008 (2007).

20. Schneider, A., Jost, A., Coulon, C., Silvestre, M., Théry, S. & Ducharne, A. Global-scale river network extraction based on high-resolution topography and constrained by lithology, climate, slope, and observed drainage density. *Geophys. Res. Lett.* **44,** 2773–2781 (2017).
21. Giachetta, E. & Willett, S. D. A global dataset of river network geometry. *Sci. Data* **5,** 180127 (2018).
22. Ross, C. W. *et al.* HYSOGs250m, global gridded hydrologic soil groups for curve-number-based runoff modeling. *Sci. Data* **5,** 180091 (2018).
23. Grimaldi, S., Nardi, F., Di Benedetto, F., Istanbulluoglu, E. & Bras, R. L. A physically-based method for removing pits in digital elevation models. *Adv. Water Resour.* **30,** 2151–2158 (2007).
24. Nardi, F., Grimaldi, S., Santini, M., Petroselli, A. & Ubertini, L. Hydrogeomorphic properties of simulated drainage patterns using digital elevation models: the flat area issue. *Hydrol. Sci. J* **53,** 1176–1193 (2008).
25. Li, J. & Wong, D. W. Effects of DEM sources on hydrologic applications. *Comput. Environ. Urban Syst.* **34,** 251–261 (2010).
26. Nardi, F., Biscarini, C., Di Francesco, S., Manciola, P. & Ubertini, L. Comparing a large-scale DEM-based floodplain delineation algorithm with standard flood maps: The Tiber River Basin case study. *Irrig. Drain.* **62,** 11–19 (2013).
27. Lehner, B., Verdin, K. & Jarvis, A. New global hydrography derived from spaceborne elevation data. *Eos* **89,** 93–94 (2008).
28. Shen, X., Anagnostou, E. N., Mei, Y. & Hong, Y. A global distributed basin morphometric dataset. *Sci. Data* **4,** 160124 (2017).
29. Pekel, J. F., Cottam, A., Gorelick, N. & Belward, A. S. High-resolution mapping of global surface water and its long-term changes. *Nature* **540,** 418 (2016).
30. Allen, G. H. & Pavelsky, T. M. Global extent of rivers and streams. *Science* **361,** 585–588 (2018).
31. Dottori, F. *et al.* Flood hazard map for Europe, 200-year return period. *European Commission, Joint Research Centre (JRC)* http://data.europa.eu/89h/jrc-floods-floodmapeu_rp200y-tif (2016).
32. Dottori, F. *et al.* Development and evaluation of a framework for global flood hazard mapping. *Adv. Water Resour.* **94,** 87–102 (2016).
33. Di Baldassarre, G. *et al.* Socio-hydrology: conceptualising human-flood interactions. *Hydrol. Earth Syst. Sci.* **17,** 3295–3303 (2013).
34. Mård, J., Di Baldassarre, G. & Mazzoleni, M. Nighttime light data reveal how flood protection shapes human proximity to rivers. *Sci. Adv.* **4,** eaar5779 (2018).
35. Ceola, S., Laio, F. & Montanari, A. Satellite nighttime lights revealing increased human expo- sure to floods worldwide. *Geophys. Res. Lett.* **41,** 7184–7190 (2014).
36. Alfieri, L., Feyen, L. & Di Baldassarre, G. Increasing flood risk under climate change: a pan-European assessment of the benefits of four adaptation strategies. *G. Climatic Change* **136,** 507–521 (2016).
37. Alfieri, L. *et al.* Global projections of river flood risk in a warmer world. *Earth's Future* **5,** 171–182 (2017).
38. Lehner, B. & Döll, P. Development and validation of a global database of lakes, reservoirs and wetlands. *J. Hydrol* **296,** 1–22 (2004).

Data Citation

1. Nardi, F. *et al.* *Figshare* https://doi.org/10.6084/m9.figshare.6665165.v1 (2018).

Acknowledgements

The authors would like to thank the NASA Shuttle Radar Topographic Mission (SRTM), the United States Geological Survey (USGS) and the Consortium for Spatial Information (CGIAR-CSI) for producing and sharing the near-global 250 m DTM. This work was developed within the activities of the Panta Rhei research initiative of the International Association of Hydrological Sciences (IAHS). GDB was supported by the European Research Council (ERC) within the project "HydroSocialExtremes: Uncovering the Mutual Shaping of Hydrological Extremes and Society", ERC Consolidator Grant no. 761678.

Author Contributions

F.N. and S.G. conceived the theoretical study for geomorphic floodplain mapping. F.N. developed the original geomorphic floodplain algorithm as ESRI AML code and produced Figure 1. A.A. developed and coded the Python scripts of the new GFPLAIN algorithm, applied the model for global floodplain mapping, prepared the GFPLAIN250m dataset and produced Figures 2 and 3. E.V. and G.D.B. contributed to the development of the paper idea, and supported the writing process of the manuscript.

Additional Information

Competing interests: The authors declare no competing interests.

How to cite this article: Nardi, F. *et al.* GFPLAIN250m, a global high-resolution dataset of Earth's floodplains. *Sci. Data*. 6:180309 doi: 10.1038/sdata.2018.309 (2019).