Hydrography90m: A new high-resolution global hydrographic dataset

Giuseppe Amatulli[1,2], Jaime Garcia Marquez[2], Tushar Sethi[3], Jens Kiesel[2,5], Afroditi Grigoropoulou[2,5], Maria M. Üblacker[2,6], Longzhu Q. Shen[2,3,7], and Sami Domisch[2]

[1]Yale University, School of the Environment, 195 Prospect Street, New Haven, CT, 06511, USA

[2]Leibniz Institute of Freshwater Ecology and Inland Fisheries, Department of Community and Ecosystem Ecology, Müggelseedamm 310, 12587 Berlin, Germany

[3]Spatial Ecology, 35A, Hazlemere Road, Penn, Buckinghamshire, HP10 8AD, United Kingdom.

[4]MARGOSA Environmental Solutions Ltd, 35A, Hazlemere Road, Penn, Buckinghamshire, HP10 8AD, United Kingdom.

[5]Christian-Albrechts-University Kiel, Institute for Natural Resource Conservation, Department of Hydrology and Water Resources Management, Olshausenstr. 75, 24118 Kiel, Germany

[6]Freie Universität Berlin, Department of Biology, Chemistry, Pharmacy, Institute of Biology, Königin-Luise-Str. 1-3, Berlin, 14195 Germany

[7]Carnegie Mellon University, Center for Green Science, Pittsburgh, PA 15213, USA

Correspondence: Giuseppe Amatulli (giuseppe.amatulli@gmail.com), Sami Domisch (sami.domisch@igb-berlin.de)

Abstract. The geographic distribution of streams and rivers drives a multitude of patterns and processes in hydrology, geomorphology, geography and ecology. Therefore, a hydrographic network that accurately delineates both small streams and large rivers with equal precision, along with their topographic and topological properties, would be indispensable in the earth sciences. Currently, available global hydrographies do not feature small headwater streams in great detail. However, these headwaters are vital because they are estimated to contribute to more than 70% of overall stream length. We aimed to fill this gap by using the MERIT Hydro Digital Elevation Model at 3 arc-sec ($\sim$90 m at the equator) to derive a globally seamless, standardised hydrographic network, the "Hydrography90m", with corresponding stream topographic and topological information. A central feature of the network is the minimal upstream contributing area, i.e. flow accumulation, of 0.05 km^2 (or 5 ha) to initiate a stream channel, which allowed us to extract headwater stream channels in great detail. By employing a suite of GRASS GIS hydrological modules, we calculated the range-wide upstream flow accumulation and flow direction to delineate a total of 1.6 million drainage basins, and extracted globally a total of 726 million unique stream segments with their corresponding sub-catchments. In addition, we computed stream topographic variables comprising stream slope, gradient, length, and curvature attributes, as well as stream topological variables to allow for network routing and various stream order classifications. We validated the spatial accuracy and flow accumulation of Hydrography90m against NHDPlus HR, an independent, national high-resolution hydrographic network dataset of the United States. Our validation shows that the newly developed Hydrography90m has the highest spatial precision, and contains more headwater stream channels compared to three other global hydrographic datasets. This comprehensive approach provides a vital, and long-overdue baseline for assessing actual streamflow in headwaters, and opens new research avenues for high-resolution studies of surface water worldwide. Hydrography90m thus offers significant potential to facilitate the assessment of freshwater quantity and quality, inundation risk, biodiversity and conservation, as well as resource management objectives in a globally comprehensive and standardised manner. The Hy-

drography90m layers are available at https://doi.org/10.18728/igb-fred-762.1 and while they can be used directly in standard GIS applications, we recommend the seamless integration with hydrological modules in open-source QGIS and GRASS GIS software to further customise the data and derive optimal utility from it.

Plain language summary:

25   Streams and rivers drives several processes in hydrology, geomorphology, geography and ecology. A hydrographic network that accurately delineates streams and rivers, along with their topographic and topological properties, is needed for environmental applications. Using the MERIT Hydro Digital Elevation Model at 90 m resolution we derived a globally seamless, standardised hydrographic network: "Hydrography90m". The validation demonstrates improved accuracy compared to other datasets.

1   Introduction

Global information on spatial hydrographic attributes, including stream topographic and topological properties, is fundamental to numerous disciplines, such as hydrological and hydraulic studies, flood and drought impact investigations, agricultural and land management, freshwater ecosystem and biodiversity assessments, conservation, element cycling, as well as for investigating the effects of climate change on the earth's freshwater resources (Lowe and Likens, 2005; Thoms et al., 2018; Maasri et al.,

35   2021a). These hydrographic attributes contain the geographic location and distribution of the world's streams and rivers, along with their network topologies and catchments.

The delineation of a hydrographic network across a wide geographic range is based on remotely-sensed Digital Elevation Models (DEM). From such datasets it is possible to derive potential water flow channels, given that water follows the steepest downstream slope (Seibert and McGlynn, 2007). Defining the upstream contributing area, i.e. flow accumulation, which initi-

40   ates a stream channel is central to delineating the streams within a hydrographic network. The smaller the threshold applied to the flow accumulation, the more detailed the resulting network and its headwaters.

Various DEMs have been used for global stream channelisation, beginning with the GTOPO30 DEM at 30 arc-sec ($\sim$1 km at equator) (USGS, 1996) from which the US Geological Survey (USGS) created the HYDRO1k dataset (USGS EROS Archive, 2018), using a 1000 km^2 threshold of upstream contributing areas. Then, in the year 2000, based on the Shuttle

45   Radar Topography Mission (SRTM) DEM (USGS, 2015) with a near-global, sub-60°latitude coverage at 3 arc-sec ($\sim$90 m at the equator) spatial resolution, Lehner et al. (2008) delineated the first near-global HydroSHEDS river network at 7.5x7.5 arc-sec ($\sim$500 m at the equator) spatial resolution, with a minimum of 10 km^2 upstream contributing area. This hydrographic dataset was later revised as the global HydroRIVERS product that used HYDRO1k for the northern latitudes. Subsequently, Yamazaki et al. (2019) computed stream channels and river widths together with flow accumulation and direction, with a

50   stream channelisation threshold of 5 km^2, based on the 3x3 arc-sec Multi-Error-Removed Improved-Terrain (MERIT) Hydro DEM. Recently, Lin et al. (2021) computed the MERIT Hydro–Vector hydrography dataset, which features global variable drainage density, and was derived from the Multi-Error-Removed Improved-Terrain (MERIT) Hydro DEM (Yamazaki et al., 2019). Despite employing MERIT Hydro at 3x3 arc-sec, the channelisation of the MERIT Hydro–Vector dataset was initialised

using a 1 km^2 threshold for the upstream contributing area, followed by a machine learning procedure to trim network density.

55 While MERIT Hydro–Vector features an up-to-date hydrographic network (i.e., stream channels) in the highest available global spatial resolution at 3x3 arc-sec, its coarse channelisation threshold does not yield a worldwide distribution of small headwater stream channels in substantial detail.

High spatial resolution of hydrographic data is key for informed water management, as it offers a detailed distribution of stream and river channels, and thus enables accurate discharge and biogeochemical cycling simulations, (Marzadri et al.,
60 2021; Liu et al., 2021; Hosen et al., 2021), nutrient concentration estimations (Shen et al., 2020), as well as biodiversity and environmental protection (Benstead and Leigh, 2012; Domisch et al., 2015a; Jackson et al., 2016). In addition, it allows for the delineation of small streams and their headwaters, i.e. the 1st and 2nd Strahler order streams (Strahler, 1957). These streams are estimated to comprise >70% of the overall length of a hydrographic network (Lowe and Likens, 2005; Leopold et al., 1964; Benstead and Leigh, 2012), and contribute significantly to flow and nutrient dynamics (Shumilova et al., 2019; Shanafield et al.,
65 2021) that are essential for maintaining biodiversity-rich habitats (Finn et al., 2011; Meyer et al., 2007). Delineating stream channels at high spatial resolution also allows for the assessment of the distribution of ephemeral streams, i.e. intermittent streams that run dry in certain seasons (Datry et al., 2014). While Messager et al. (2021) has mapped the global distribution of such non-perennial streams based on HydroRIVERS, small headwater streams are nevertheless missing from this dataset because of a significantly larger stream channelisation threshold, as mentioned in the preceding paragraph.

70 With Hydrography90m, we address the all-important issue of headwaters and present a globally seamless and standardised hydrographic network at 3x3 arc-sec (equal $0.0008\overline{3}$ °), together with their corresponding stream topographic and topological attributes. We use the Multi-Error-Removed Improved-Terrain (MERIT) Hydro DEM (Yamazaki et al., 2019) and employ a standard worldwide channelisation threshold of 0.05 km^2 (or six 3x3 arc-sec grid cells as the equator). This results in a dense network, which depicts small headwater stream channels in fine detail. This choice of small channelisation threshold (in the
75 low hectare range) is fundamental in computing headwater stream variables, which are essential for intermittent and flood flow modelling (Ågren et al., 2015). Thus, thresholding with a higher value (e.g. 1 km^2) would fail to include vital headwater stream hydrographic features.

The Hydrography90m dataset consists of a global rendition of stream channels and drainage basins, the sub-catchment of each stream segment, in-stream and among-stream distance metrics, and various stream slope and stream order metrics.
80 Additionally, the dataset provides a full topology for flow routing, owing to a unique stream segment identifiers, each of which contains the attributes of the related upstream and downstream segments. The Hydrography90m stream network and flow accumulation have been validated against the National Hydrography Dataset Plus High Resolution (NHDPlus HR) (Moore et al., 2019) product, revealing high precision on the spatial accuracy and flow accumulation computation. We note that we are in the process of providing monthly discharge estimates for each stream segment, which will be used to reduce the channel
85 density so as to retain only those channels that have potentially held water during a given time frame within a 30-year period. The hydrographic dataset (https://doi.org/10.18728/igb-fred-762.1) is available for download in raster and vector formats. We recommend downloading the data via http://hydrography.org, where users can click on a map interface to choose the

required tile to directly download the corresponding layers, or follow a scripting procedure for batch download and subsequent processing (likewise available at http://hydrography.org.

2 Methodology

2.1 Terminology used

To facilitate the understanding of the various layers we provide a description of terms used in the manuscript and Hydrography90m dataset (Figure 1) below.

- *Flow direction*: the direction of water flow in a grid cell, given that water follows the steepest downstream slope.
- *Flow accumulation*: the upstream contributing area, i.e. the drainage of water into a given downstream cell. It is expressed in area units (in our case in km^2)
- *Flow accumulation threshold*: the upstream contributing area that initiates a stream channel. In Hydrography90m it has been set to 0.05 km^2 ($\sim$six $3''$ cells at the equator)).
- *Stream cell*: the grid cell that marks a stream channel's presence. It is the smallest spatial unit in Hydrography90m with a size of 3x3 arc-sec (equal to $\sim$90x90 m at the equator).
- *Stream channel*: part of the hydrographic network as extracted from the DEM. A stream channel consists of many stream segments. In Hydrography90m the stream channel network does not assume the presence of water but indicates its potential as a flow path.
- *Stream segment*: the stream channel between two segment nodes (or from initialisation to the first confluence) of the network where the stream order is unchanged. Each stream segment worldwide is labelled with a unique ID.
- *Drainage basin*: any area of land where precipitation collects and drains into a common outlet. The outlet can be into the sea or an inland depression. If the drainage basin can be included completely in one tile, it is labelled *entire drainage basin* but if it intersects a tile border, it is termed *truncated drainage basin*. Each drainage basin worldwide is assigned a unique ID. Adjacent basins share a border that corresponds to the *basin drainage divide* (i.e., ridgeline between basins).
- *Sub-catchment*: land area between two segment nodes that contributes to the local flow accumulation of a given stream segment. Sub-catchments and stream segments have a common unique ID worldwide. Adjacent sub-catchments share a border that corresponds to the *sub-catchment drainage divide*, i.e. the ridgeline between sub-catchments.
- *Base layers*: comprise raster flow accumulation and flow direction, which are the primary layers for extracting the hydrographic network and basins.
- *Network layers*: raster and vector layers that are derived from flow accumulation and flow direction. The network layers include drainage basins, sub-catchments, and stream segments.
- *Topographic and topological variables*: additional attributes characterising the topography (e.g. stream slope, stream distance) and topology (e.g. stream order) of the hydrographic network at the stream cell or segment resolution.
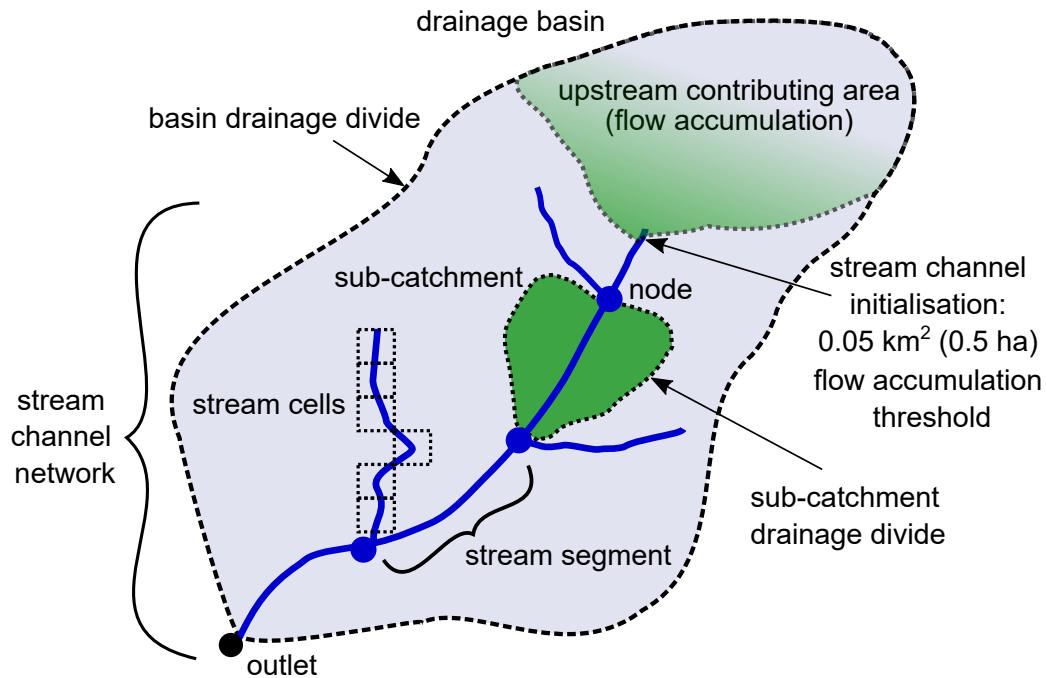
Figure 1. Schematic overview of the terminology used in Hydrography90m. See the main text for detailed descriptions.

- *Tiling System*: two vector layers that consist of the Irregular Tiling System (ITS) and Regular Tiling System (RTS), used to derive the Hydrography90m.
- *Regional unit*: an area that contains only entire drainage basins, masking the truncated ones. Useful for selecting entire drainage basin towards custom study areas.

2.2 Digital Elevation Model (DEM)

As the basis for all calculations, we used the MERIT Hydro DEM that represents the best available globally seamless, high-resolution DEM to date (Yamazaki et al., 2019). The MERIT Hydro DEM is available for download at http://hydro.iis.u-tokyo.ac.jp/~yamadai/MERIT_Hydro/index.html. In general, DEMs represent the elevated land surface in relation to a reference height such as sea level. In addition, DEMs are extensively deployed in geo-computational applications, as land surface plays a fundamental role in modulating earth-dynamic operations such as atmospheric, geomorphological, hydrological, and ecological processes. DEMs built from space-borne observations can achieve global coverage and thus have broad applications. However, the original space-borne DEMs are prone to systematic biases as well as random noise (Rodríguez Ernesto A4 - Morris, Charles S. A4 - Belz, J. Eric, 2006; O'Loughlin et al., 2016). The systematic bias stems from the influence of tree canopies, while random noise can be classed into speckle, stripe noise and absolute biases depending on their wavelengths (Rodríguez Ernesto A4 - Morris, Charles S. A4 - Belz, J. Eric, 2006; Takaku et al., 2016). The Multi-Error-Removed Improved Terrain (MERIT) DEM(Yamazaki et al., 2017), at $3''$ resolution extended from $90°$N to $60°$S, was the first global product with a consistent

135 systematic bias and random noise removal procedure, and is considered the best available, seamless DEM with global coverage (Hirt, 2018; Moudrỳ et al., 2018). MERIT DEM is a fusion of the National Aeronautics and Space Administration (NASA) SRTM3 version 2.1 (Farr et al., 2007), the Japan Aerospace Exploration Agency (JAXA) AW3D global high resolution 3D map (version 1) (Tadono et al., 2015) and the Viewfinder Panorama's DEM (available at http://www.viewfinderpanoramas.org/dem3.html). The quality of MERIT DEM is unique because it eliminates stripe noise using a 2-D Fourier filtering technique

140 that is able to detect unrealistic regular terrain undulations. Absolute bias has been corrected by calculating the difference between the DEM and the ICESat elevations (Harding and Carabajal, 2005). Tree-height bias is addressed by combining tree density (Hansen et al., 2013) and tree height (Simard et al., 2011), and by comparing the obtained MERIT DEM to ICESat.

Even though the tree canopy bias was removed in MERIT DEM, the elevation information in grid cells with substantial tree coverage has a higher uncertainty compared to those without tree coverage (Yamazaki et al., 2017, 2019). Hence, the tree

145 density map and G3WBM glacier map were used to enforce the separation of actual inland basins and dummy depressions by means of a correction of a predefined topographic volume that ascertains whether a depression is present or not (Yamazaki et al., 2017). Finally, speckle noise was removed using an adaptive-scale smoothing filter (Gallant and Wilson, 2000). Yamazaki et al. (2017) reported that after the error removal, areas mapped with ±2 m or better vertical accuracy increased by 19% and slope distortions were reduced.

150 In 2019, Yamazaki et al. (2019) released the MERIT Hydro – a new global hydrologically-adjusted DEM, which included depression, flow direction, flow accumulation, river width and height above the nearest drainage (HAND) layers. The hydrologically-adjusted elevation incorporates various surface water datasets (G1WBM (Yamazaki et al., 2015), GSWO (Pekel et al., 2016) and OpenStreetMap (OpenStreetMap contributors, 2017)), as well as a Landsat-derived tree density map (Hansen et al., 2013) and G3WBM glacier map to allow for an additional round of hydrological corrections (Yamazaki et al., 2015).

155 The water bodies serve as a carving template to modify the elevation of the MERIT DEM, satisfying the condition that "downstream is not higher than upstream" and to include valleys that are not depicted because they are smaller than the grid cells of the DEM (Yamazaki et al., 2019). The G1WBM and GSWO are Landsat-derived (30 m resolution) and therefore of limited use to depict tributaries smaller than 30 m river width or rivers with a width >30 m that are covered by tree canopy (Amatulli et al., 2020). On the other hand, OpenStreetMap (OpenStreetMap contributors, 2017) does depict small tributaries, depending

160 on the region and the extent of survey efforts on the concerned water bodies. To date, not all countries in OpenStreetMap provide high spatial accuracy for headwater streams, and thus, headwater streams are not yet carved consistently into MERIT Hydro DEM (Amatulli, 2020; Amatulli et al., 2018a).

2.3 Flow routing algorithms

The flow accumulation operation performs a cumulative count of the number of grid cells (or other surface area unit) that drain

165 into outlets given the terrain surface. Calculating flow accumulation involves three sequential algorithms: determining flow direction, addressing depressions and flat areas, and finally, calculating flow accumulation.

Several flow-routing algorithms exist for identifying streams channels at various spatial resolutions (Yang et al., 2010; Orlandini et al., 2003; Tarboton, 1997; Zhang et al., 2007b; O'Callaghan and Mark, 1984). They are built upon the observation

that water follows the steepest route along a relief, and accumulates in valleys, lowlands, flat areas and depressions (Heine et al., 2004).

The most widely used algorithm is the single-flow (D8) (O'Callaghan and Mark, 1984) algorithm that assigns flow from a focal grid cell to only one of the eight neighbouring grid cells with the steepest slope. This algorithm accumulates, or pools, the entire flow from one cell to another, producing often distinct, artificially straight stream channels (Erskine et al., 2006), where the steepest gradient might lie between two of the eight directions (Seibert and McGlynn, 2007). D8 has been used to develop the HydroRIVERS (Lehner et al., 2008), the MERIT Hydro hydrography map (Yamazaki et al., 2019), and the MERIT Hydro–Vector dataset (Lin et al., 2021). Nonetheless, it is acknowledged that D8 algorithm is not adequate to represent flow contributing area in headwater regions (Yamazaki et al., 2019).

To overcome the limitation of artificially straight stream channels (Erskine et al., 2006) of the D8, the multi-flow direction algorithm (MD8) has been proposed as an improvement. It splits the flow into multiple directions as a function of the slope in each of the neighbouring grid cells (Quinn et al., 1991, 1995) and produces stream channel patterns closer to reality than D8 does. In case of small elevation differences between two or more neighbouring grid cells, both cells receive about the same proportion of the accumulated area. The disadvantage of MD8, as presented by Quinn et al. (1991), is that the area from one grid cell is routed to all downslope grid cells without considering a divergence or convergence hillslope factor, which would increase or decrease the dispersion rate as a function of the tangential curvature. To minimise this problem, Holmgren (1994) suggested partitioning the flow according to a convergence factor ranging from 1 to 10 (suggested value 5).

This MD8 is defined as

$$fi = \frac{(tan\beta_i)^x}{\sum(tan\beta_i)^x}$$

where β is the slope gradient and x is a weighting factor. The MD8 algorithm, as opposed to D8, is better able to handle a wider range of terrain, including flat areas where flow routing is challenging (Liang and MaCkay, 2000). It therefore allows the extraction of stream channels of headwaters and small, non-perennial streams in greater detail than D8 does.

Alternatively, Tarboton (1997) has suggested the use of triangular facets to overtake the eight possible directions of the D8. Tarboton (1997) named this method D∞ that describes the infinite behaviour of dispersion of the single-direction flow pathways. Nevertheless, we opted for MD8 using the traditional Holmgren method which has been implemented in GRASS GIS (named "FD8") (Neteler et al., 2012; Neteler and Mitasova, 2013) within several hydrological modules, and which shows similar performance than the D∞ (Seibert and McGlynn, 2007) in hydrologically and algorithmically challenging terrain, such as flat areas (Jasiewicz and Metz, 2011).

2.4 Depression and endorheic basins

We overlaid the depression layer from MERIT Hydro (Yamazaki et al., 2019) with the HydroLAKES dataset (Messager et al., 2016) and identified 1400 interior lakes that coincide with depression points that mark inland depressions. We then rasterised

these lakes to the 3x3 arc-sec grid cell resolution and together with the MERIT Hydro depression layer, we assigned these areas as NoData in MERIT Hydro (i.e., no outflow from these lakes). Such a procedure was needed, for instance, to treat the Caspian Sea an inland depression to route the flow accumulation correctly until the coast line. This procedure also created many small, surrounding drainage basins only connected to lakes. In addition, we assigned the geographic locations (i.e. the single 3x3 arc-sec grid cells) of all remaining depression points in MERIT Hydro as NoData.

2.5 Computational stages

The overall computation of the Hydrography90m consisted of four stages:

- Splitting the global DEM into smaller spatial units (tiles) to achieve computational scalability.
- Computing flow accumulation and direction, and the subsequent extraction of stream channels and basins.
- Validating the spatial distribution of stream channels and basins using independent data sources.
- Computing geophysical, morphological and topological properties of the stream channels and basins.

Several procedures within the entire workflow were repeated at different stages, including the use of intermediate layers from preceding stages in the creation of final layers. The entire work flow was automated with Bash scripts to integrate a hybrid operation of multiple open-source software. These procedures were run at the High Performance Computing (HPC) facility of the Center for Research Computing, Yale University:

- Geospatial Data Abstraction Library (GDAL), version number 3.1.0 (GDAL Development Team, 2020): for tiling, cropping, mosaicking, merging, image compression.
- Processing Kernel for geospatial data (Pktools), version number 2.6.7.6 (Kempeneers, 2018; McInerney and Kempeneers, 2015): for masking, histogram, re-classification.
- Geographic Resources Analysis Support System - Geographic Information System (GRASS GIS) software , version number 7.8.0 (GRASS Development Team, 2019): for computing the hydrography.

These tools provided fast, flexible and scalable features and functions for raster-based analysis with Python APIs and Bash command access (Amatulli et al., 2014). They also enabled multi-core parallel processing of very large datasets owing to efficient algorithms and optimised memory management. The entire workflow consisted of eight main tasks (labelled with letters in Figure 4) for which we used a total of three GRASS GIS modules and several GIS commands (i.e. cropping, reclass, merging, etc.), and yielded both intermediate and final outputs within the given tiling system.

3 Tiling the DEM

3.1 Irregular tiling system (ITS)

To address the high computational demand for calculating flow accumulation at 3x3 arc-sec resolution globally, we split the entire MERIT Hydro DEM into 59 tiles of varied sizes to yield an Irregular Tiling System (ITS, red squares in Figure 2).

These irregular tiles were large enough to contain one or more entire drainage basins such that their lateral and longitudinal connectivity is maintained within each tile. For an initial approximation of the location and size of the basins, and hence the position of the tiles, we aggregated MERIT Hydro from 3x3 arc-sec to a 30x30 arc-sec resolution, while preserving the minimal value of elevation in each cell. The resulting global DEM matrix of 751,680,000 grid cells at 30x30 arc-sec resolution, allowed us to compute drainage basins on a global extent using the GRASS GIS module *r.stream.basins*. We then manually created rectangular tiles considering (i) the maximum possible number of ∼2 billion grid cells (2^31-1, that requires ∼67 GB of RAM) within a tile, (ii) that large drainage basins were completely within a given tile, while (iii) keeping a buffer of 1°longitude and latitude. The tiles naturally overlap with each other because of the irregular shapes of adjacent drainage basins. By merging all entire drainage basins within the 59 irregular tiles we obtained a representation of the world's drainage basins at 3x3 arc-sec resolution (Figure 2).
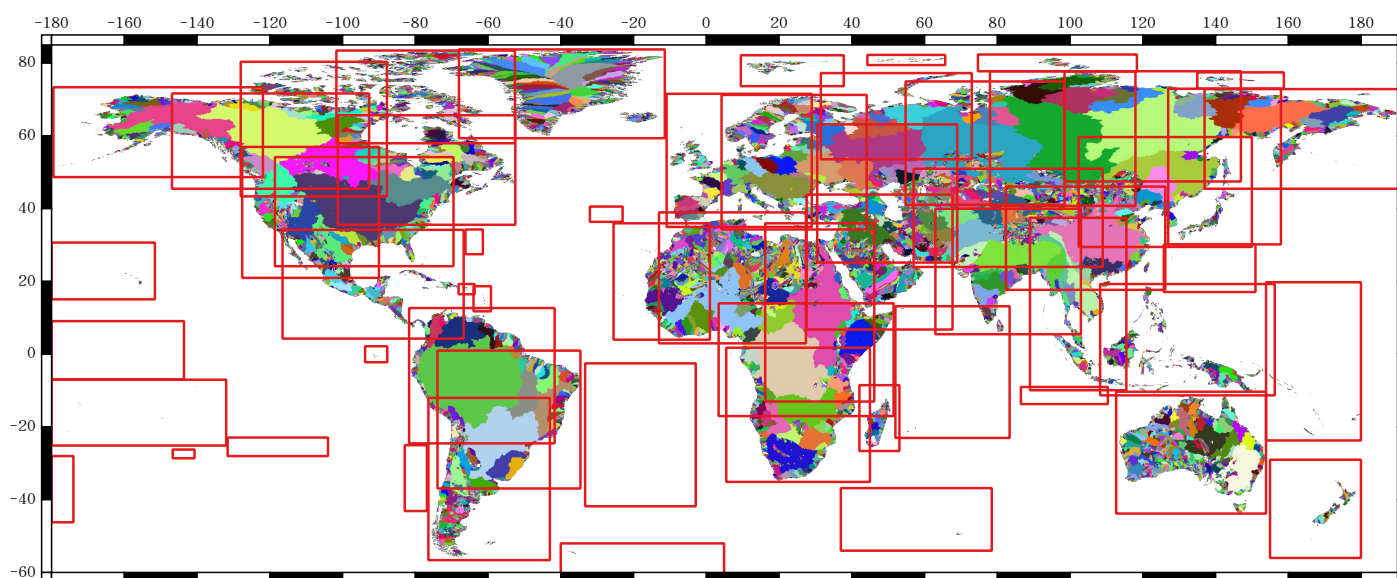
Figure 2. The irregular tiling system (ITS, in red) overlaid with the global drainage basins at 3x3 arc-sec resolution (random colours for illustrative purposes). Within a tile we retain only the areas that belong to entire drainage basins, so as to preserve lateral and longitudinal connectivity within the basin.

3.2   Regular tiling system (RTS)

In addition to the ITS, we also built a Regular Tiling System (RTS) consisting of 116 tiles with a fixed dimension of 20°longitude x 20°latitude (ranging from -180°to +191°longitude and from +85°to -60°latitude). This RTS was implemented to avoid the handling and distribution of a single, and computationally heavy, global file. We modified the size of two eastern tiles, since the traditional map view of MERIT Hydro splits drainage basins at -180°and +180°. We set the Bering Strait as the border

of the map in the north-east, where the tiles reached 31°longitude x 20°latitude (i.e., the eastern boundary was expanded to 191°to include the Chukotka region in Russia). We repeated this for southern tiles, reaching 20°longitude x 25°latitude (i.e., the southern boundary was expanded to -60°latitude to include southern islands). All distributed Hydrography90m raster and vector files in are available for download in tiles. The tile labels are reported in Figure 3. In case the area of interest crosses several tiles, the data needs to be merged to combine the cross-border drainage basins. We provide easy-to-use efficient code and instructions to merge the tiles for vector or raster files at http://hydrography.org. No border effects or artefacts remain after the merge.
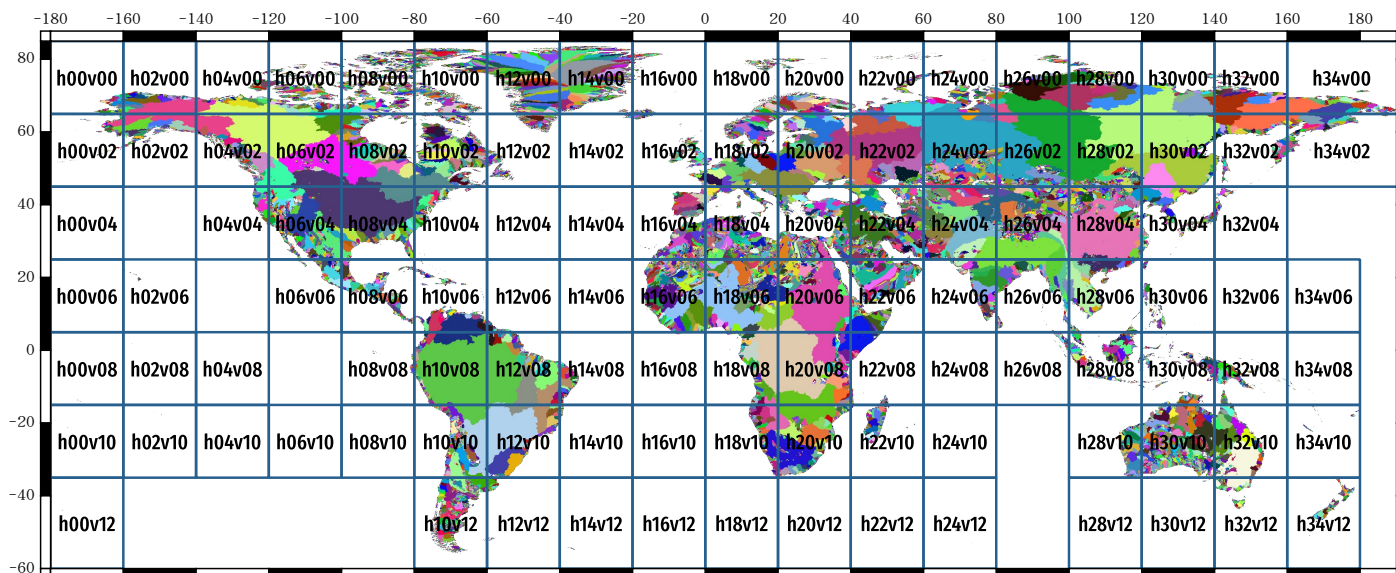
Figure 3. The regular tiling system (RTS, in blue) overlaid with global drainage basins at 3x3 arc-sec resolution (random colours for illustrative purposes). Tile labels correspond to the names of raster and vector data available for download, which are listed in Tables 1, 3, 4, 5, 6, 7. We provide an interactive map at http://hydrography.org that allows clicking on a given tile to directly download the data.

4 Computing the base and network hydrographic layers

The stream channel extraction and the drainage basin delineation was performed in GRASS GIS software using the *r.watershed* (Metz et al., 2011) module, followed by *r.stream.extract* (Jasiewicz and Metz, 2011), as well as the *r.stream.basins* (Jasiewicz and Metz, 2011) module, respectively. The work was split into two phases (Figure 4): the first phase produced a globally seamless representation of the flow accumulation, whereas the second phase generated the stream channels and associated data.
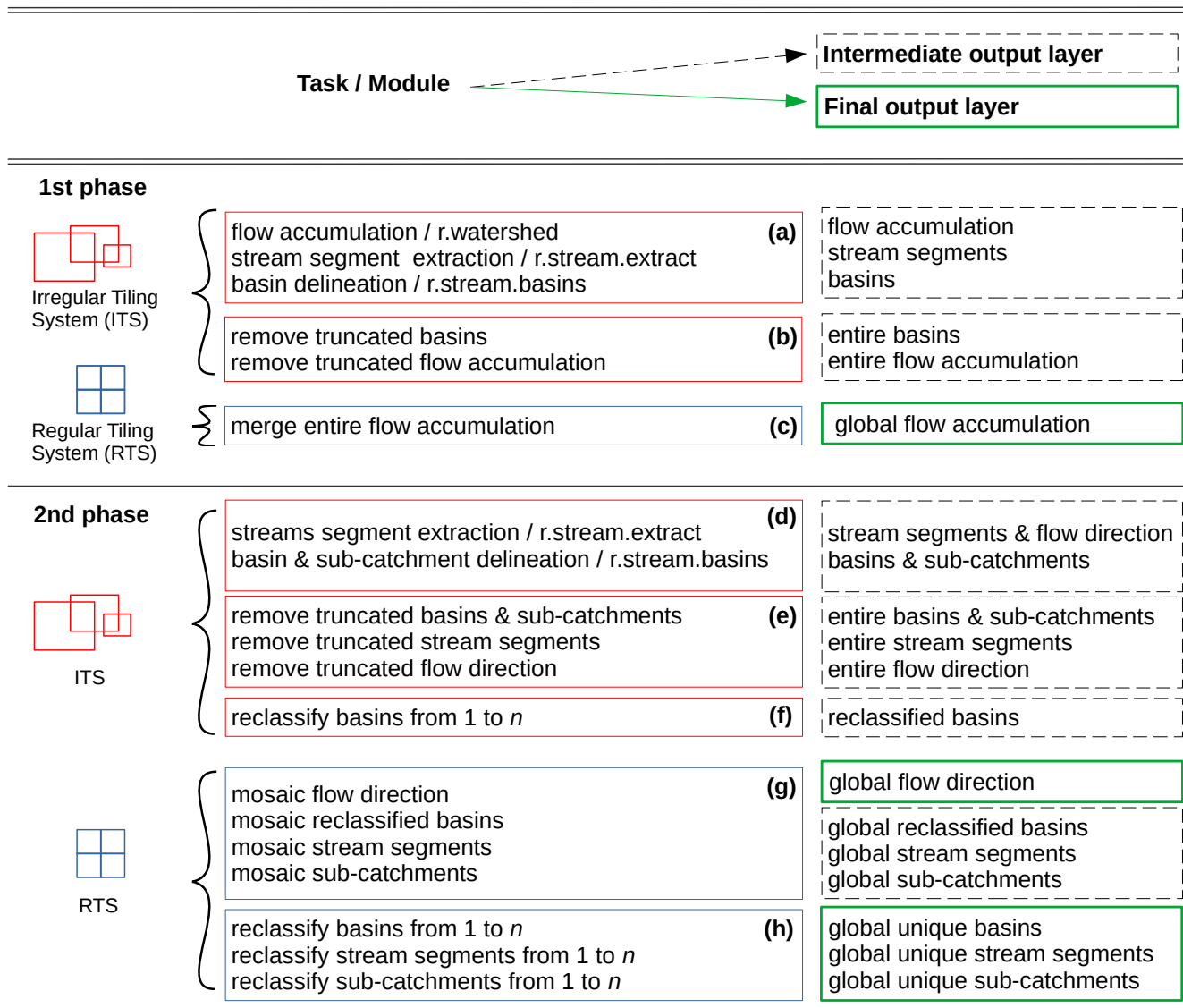
Figure 4. Overview of the Hydrography90m dataset computation workflow. Task labels correspond to the performed computation, and command labels refer to the GRASS GIS module used. The parenthesised letters listed in the figure correspond to steps detailed in the methodology section. A schematic scripting procedure of the workflow for three South American tiles is reported at http://hydrography.org. The scripting procedure is related to the 1st and 2nd phases but without the final reclassification step to have a simplified workflow example.

4.1 Flow accumulation within the irregular tiling system (ITS)

260   For each irregular tile we ran the *r.watershed* module to produce a flow accumulation map (one for each tile) based on the MD8 multi-flow direction algorithm (Holmgren, 1994). In *r.watershed* we used the MERIT Hydro elevation and depression

layers as inputs in addition to a "surface area in km^2" layer at 3x3 arc-sec resolution for downstream area accumulation. The latter provides the surface area expressed in km^2 within each 3x3 arc-sec grid cell. This was necessary, as we computed the hydrography in the unprojected WGS84 coordinate reference system (Figure 4a).

4.2 Checking for truncated drainage basins (ITS)

Each irregular tile included the flow accumulation, as well as the drainage basins and stream channels within the circumscribed drainage basins. Some drainage basins were entirely hemmed within a given tile, while others were partially included because they spread across multiple tiles (and where the tile could not be enlarged beyond the maximum number of grid cells). These partially-included or truncated basins were identified (and removed) by querying for those that intersected a tile border. This left only entire drainage basins and the associated flow accumulation (Figure 4b).

4.3 Merging the global flow accumulation

We merged the 59 irregular flow accumulation tiles, which yielded the 116 smaller tiles of the RTS (ranging in size from a few MB to 2 GB, stored as Float32 data type), to a globally seamless flow accumulation layer at 3x3 arc-sec spatial resolution (Figure 4c). This was done in the interest of producing manageable file sizes.

The creation of this 3x3 arc-sec resolution globally seamless flow accumulation layer computed with the MD8 algorithm can be considered a pioneering computational achievement. It serves as the foundation for Hydrography90m, and in our understanding will allow for significant expediency in the future computation of any derivative hydrological products.

4.4 Stream channel and basin delineation computation (ITS)

We used the seamless global flow accumulation to re-compute the drainage basins, flow direction and stream network. This resolved any errors that could have occurred at the tile borders and truncated drainage basins, given possible rounding errors in the grid cell alignment when cropping drainage basins at 3x3 arc-sec resolution at a global extent. Again, we ran *r.stream.extract* followed by *r.stream.basins*, masking all previously identified truncated drainage basins (Figure 4d,e).

4.5 Mosaic drainage basins, sub-catchments, streams segments and flow direction

In each tile in the ITS, several drainage basins were computed having a unique identifier (ID) ranging from 1 to n ID. Prior to creating the global mosaic, we reclassified all drainage basin IDs from 1 to n in order to consecutively number the basins across the globe. 4f). We repeated this re-classification after the global merging to ensure that the ID series from 1 to n was continuous and thereby avoided any gaps in the ID sequence. Ultimately, the re-classification yielded a total of 1,560,490 globally unique drainage basin IDs. A similar reclassification procedure was performed on the sub-catchment and stream segment IDs 4g,h).

A global representation of the network hydrography layers produced with such methodology is shown in Figure 5. A more detailed representation is depicted in Figure 6 and the corresponding Table 1. The table lists the file description and GRASS

GIS commands for locating these layers in the data repository and reproducing the calculations (Table 1). Additional features of the layers are available in the relevant GRASS GIS module manual pages.
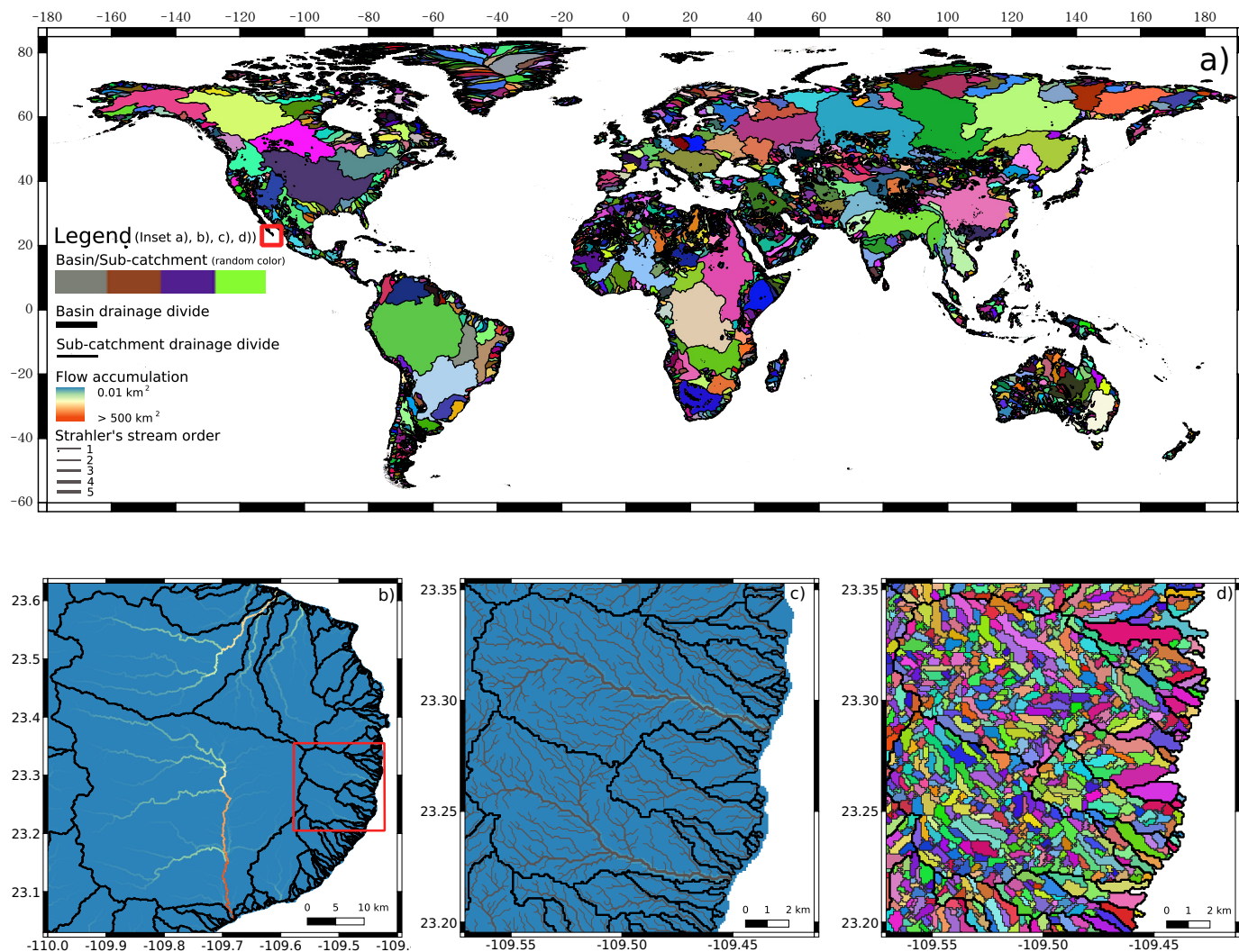
Figure 5. Map (a) shows the global distribution of the newly-delineated 1.6 million drainage basins. The red box in Baja California Sur, Mexico, represents the location of inset (b). This inset shows flow accumulation and the basin drainage divide. Inset (c) shows flow accumulation and the stream channel network, where line width corresponds to the Strahler stream order. Map (d) illustrates the corresponding sub-catchments, sharing an identifier ID with stream channels. Drainage basin and sub-catchment colour assignment is random and for illustrative purposes only.

Table 1. Base and network layers of Hydrography90m: flow accumulation, flow direction, drainage basins, outlets, stream segments, sub-catchments, regional units and depression; Map reference corresponding to Figure 6 for raster visualisation and Figure 5 for vector visualisation; Unit; Commands for computation and output file names. The asterisk stands for the regular tile ID for downloading the data, available in 20°x 20°tiles (Figure 3).

Output map description	Map reference	Unit	Command	Output tif file name
Flow accumulation (raster)	Figure 6d	km^2	r.watershed -b accumulation=acc	accumulation_*.tif
Flow direction (raster)	Figure 6f	NE-N-NW-W-SW-S-SE-E correspond to 1 - 2 - 3 - 4 - 5 - 6 - 7 - 8	r.stream.extract direction=dir threshold=0.05	direction_*.tif
Drainage basin (raster)	Figure 6g	IDs from 1 to 1,676,628	r.stream.basins -l basins=basin	basin_*.tif
Drainage basin (vector)	Figure 5b	IDs from 1 to 1,676,628	gdal_polygonize.py	basin_*.gpkg
Outlets (raster)	Figure 6h	ID=1	r.stream.extract stream_vector=stream threshold=0.05; v.to.rast input=stream	outlet_*.tif
Outlets (vector)	Figure 6h	ID=1	gdal_polygonize.py	outlet_*.gpkg
Depression (raster)	-	ID=1	pksetmask	depression_*.tif
Stream segment (raster)	Figure 6h	IDs from 1 to 726,723,221	r.stream.extract stream_raster=stream threshold=0.05	segment_*.tif
Sub-catchment (raster)	Figure 6i	IDs from 1 to 726,723,221	r.stream.basins basins=sub_catchment	sub_catchment_*.tif
Sub-catchment (vector)	Figure 5d	IDs from 1 to 726,723,221	gdal_polygonize.py	sub_catchment_*.gpkg
Regional unit (raster)	Figure 7	IDs from 1 to 116 IDs from 150 to 200	pkreclass pksetmask	regional_unit_*.tif
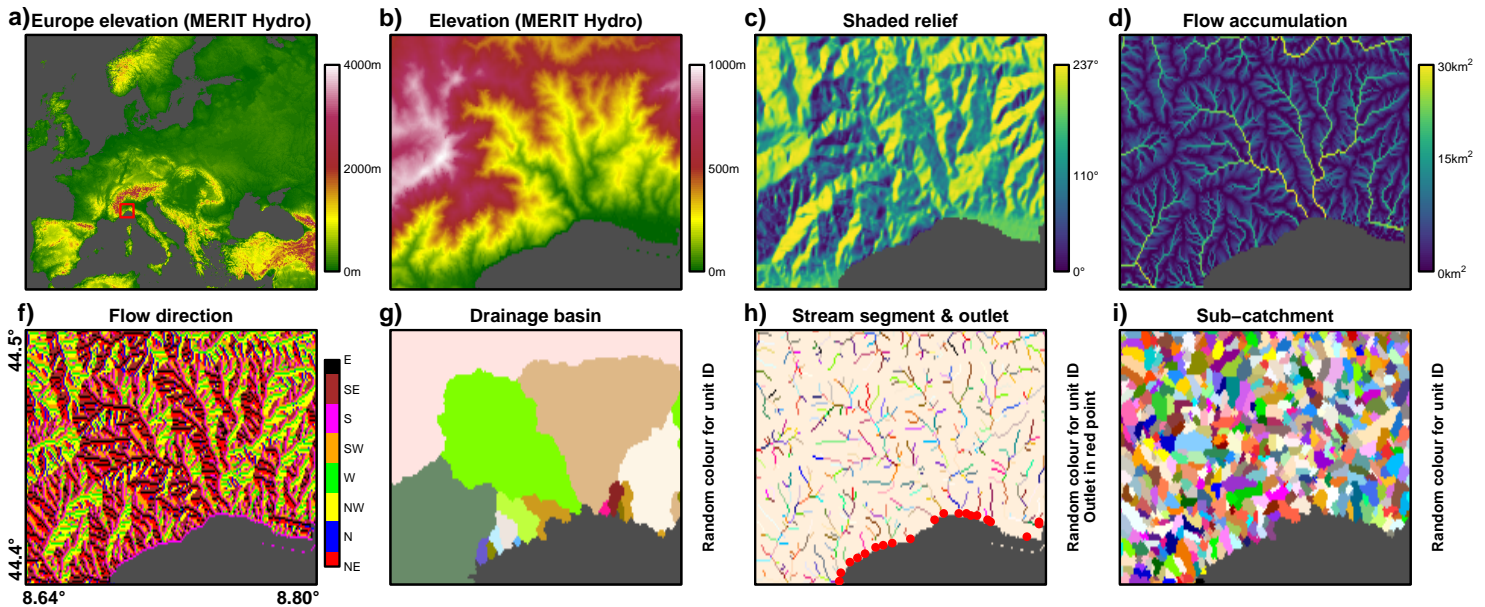
Figure 6. Map (a) shows the MERIT Hydro DEM for an area of 13x11 km in north-west Italy, and (b–i) the base and network layers of Hydrography90m. The sea is depicted in dark grey. The outlets are shown in panel (h) as red points.

4.6 Regional units

In addition to the base hydrography layers we provide a "regional unit" raster map that holds 166 regional unit IDs, which
295   contains only entire drainage basins. Such units are useful for splitting whole global hydrography layers into single drainage basins or other units of manageable size. Each regional unit together with the correspondent Hydrography90m layers can be loaded into GRASS GIS for additional computation, accounting for less than ∼2 billion grid cells. Such regional units are meant only to address computational requirements and are not for the consideration of any eco-region context or hydrological similarity. Here, the 50 largest drainage basins, such as the Nile, Amazon, or Mississippi drainages, correspond to 50 single
300   regional units, and the remaining 116 regional units include two or more smaller entire drainage basins. A global representation of the 166 regional units is shown in Figure 7. The details for regional unit IDs are provided in the last row of Table 1. In case users want to perform a hydrological analysis within each single "regional unit", they may first need to merge the data across tiles, and then identify and mask the specific "regional unit" of interest. We provide easy-to-use efficient code and instructions at http://hydrography.org.
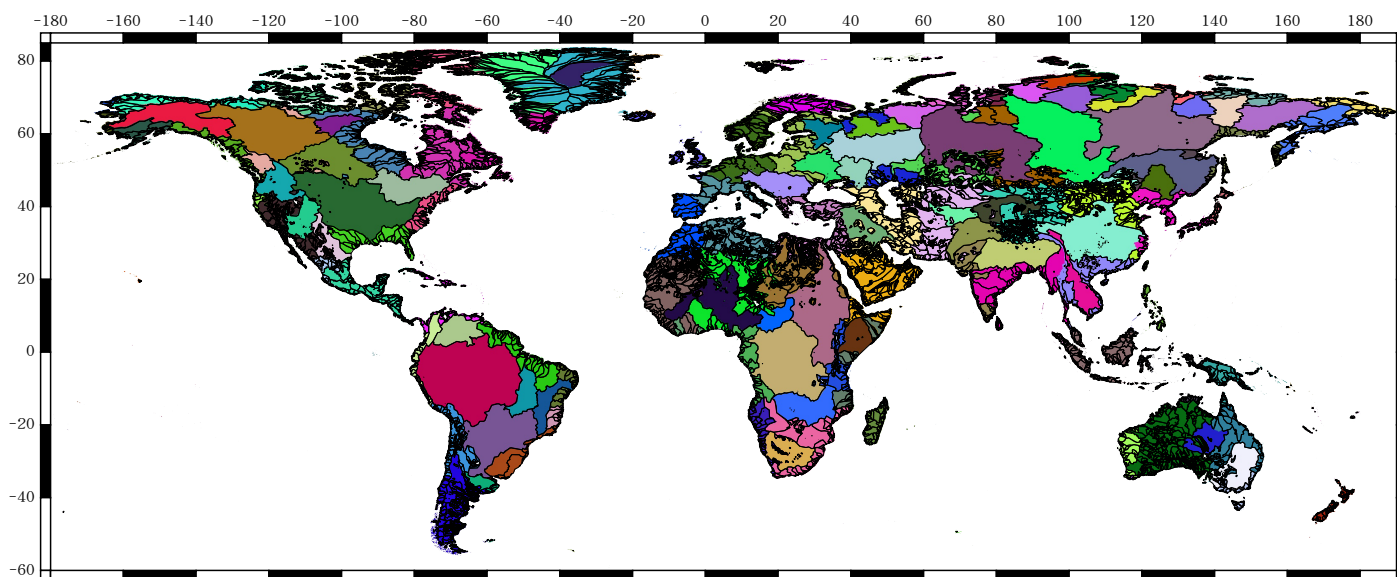
Figure 7. The 166 regional units that facilitate the splitting of all the Hydrography90m layers into customisable zones, including those areas that are delimited by an entire basin. The first 116 regional units (IDs 1 to 116) include two or more entire drainage basins, e.g. the green regional unit in the north of Europe, which includes several such basins. The remaining 50 units (IDs 150 to 200) each contain one of the 50 largest drainage basins in the world, e.g. Nile basin, Amazon basin, etc. The colour assignment is random and for illustrative purposes only. Each coloured unit holds an ID at 3x3 arc-sec grid cell resolution and serves to mask any neighbouring drainage basins that may not be within a given area of interest (see usage notes for more details).

5 Validation of the Hydrography90m

We validated the accuracy of the stream channels spatial position and the flow accumulation against the independent and observed, NHDPlus HR vector dataset (Moore et al., 2019) of the United States.

5.1 Spatial accuracy of the streams

We then compared the newly-delineated Hydrography90m against three other global datasets: the HydroRIVERS dataset (Lehner et al., 2008), the Global River Widths from Landsat (GRWL, Simplified Vector Product V01.01) (Allen et al., 2018), and the MERIT Hydro–Vector hydrography dataset (Lin et al., 2019). These vector-based datasets were brought to a 3x3 arc-sec grid cell resolution using *gdal_rasterize* to allow a direct comparison with the newly-developed Hydrography90m stream network dataset.

Since none of these previous products had delineated headwater streams at high spatial precision, we used the NHDPlus HR vector dataset as a reference to compare the accuracy of our newly-delineated stream channels. The NHDPlus HR was

at 1:24,000 scale, 3D Elevation Program data at 10 m spatial resolution, and the nationally complete Watershed Boundary Dataset (Buto and Anderson, 2020). We likewise rasterised the NHDPlus HR vector lines to a 3x3 arc-sec grid cell resolution to allow for a direct comparison. We then buffered the NHDPlus HR gridded stream lines in 5 categories, the first category (buffer-0) being the grid cells where the NHDplus HR network overlaps, and the other four categories (from buffer-1 to buffer-4) representing 100 m interval buffers (Figure 8). We overlaid the four datasets with each of the buffered ranges and calculated the number of overlapping grid cells in each dataset within the given buffer distance. This procedure accounted for the lateral accuracy of each hydrographic dataset when compared to the NHDPlus HR reference dataset. In addition, in order to estimate stream length, we quantified the length of each stream channel as the number of stream grid cells, relative to the length of the NHDPlus HR streams.
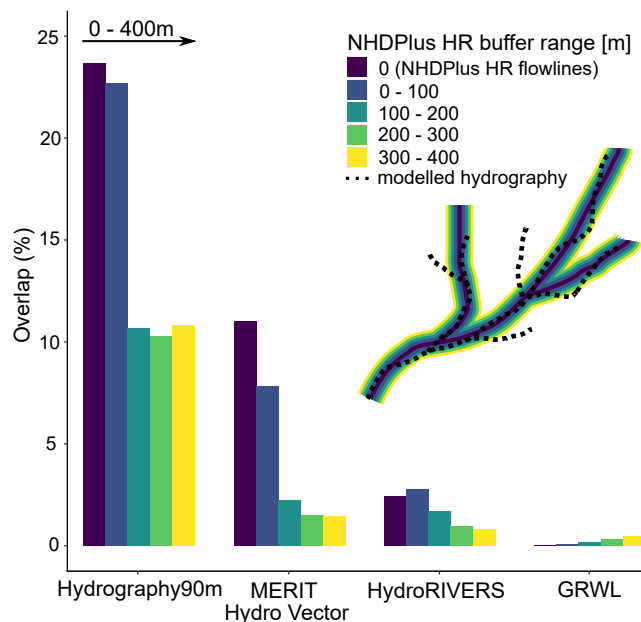
Figure 8. Comparison of the spatial accuracy of the newly-developed Hydrography90m and the MERIT Hydro-Vector, HydroRIVERS and GRWL datasets against NHDPlus HR. The inset shows the validation method: we first buffered NHDPlus HR by 100, 200, 300 and 400 m (see corresponding colours), and then calculated the fraction of overlapping stream grid cells between NHDPlus HR and each dataset. The latter are illustrated here as dotted lines.

The validation showed that the Hydrography90m has the highest lateral accuracy within the smallest buffer ranges among each of the compared datasets pairs (Figs. 8, 9). Within a buffer distance of 100 m, Hydrography90m overlaps with NHDPlus HR by 46%, achieving the highest overlap among all available global hydrographic datasets (Figure 8). Assessing the proportion of stream channels in each hydrography dataset versus NHDPlus HR showed that Hydrography90m underestimated the total river length with 28% less stream grid cells than NHDPlus HR. In comparison, the MERIT Hydro-Vector contained 77%, the HydroRIVERS 92% and the GRWL 99% less stream grid cells than the NHDPlus HR (Figure 9). This discrepancy

in overall stream length and coverage among the different hydrographies is also shown in Figure 9. Hydrography90m hence provided an all-inclusive approach by also considering potential stream channels contingent on water availability, delineating small headwater streams for the first time globally, and providing an important baseline for the future assessment of stream flow within these channels.
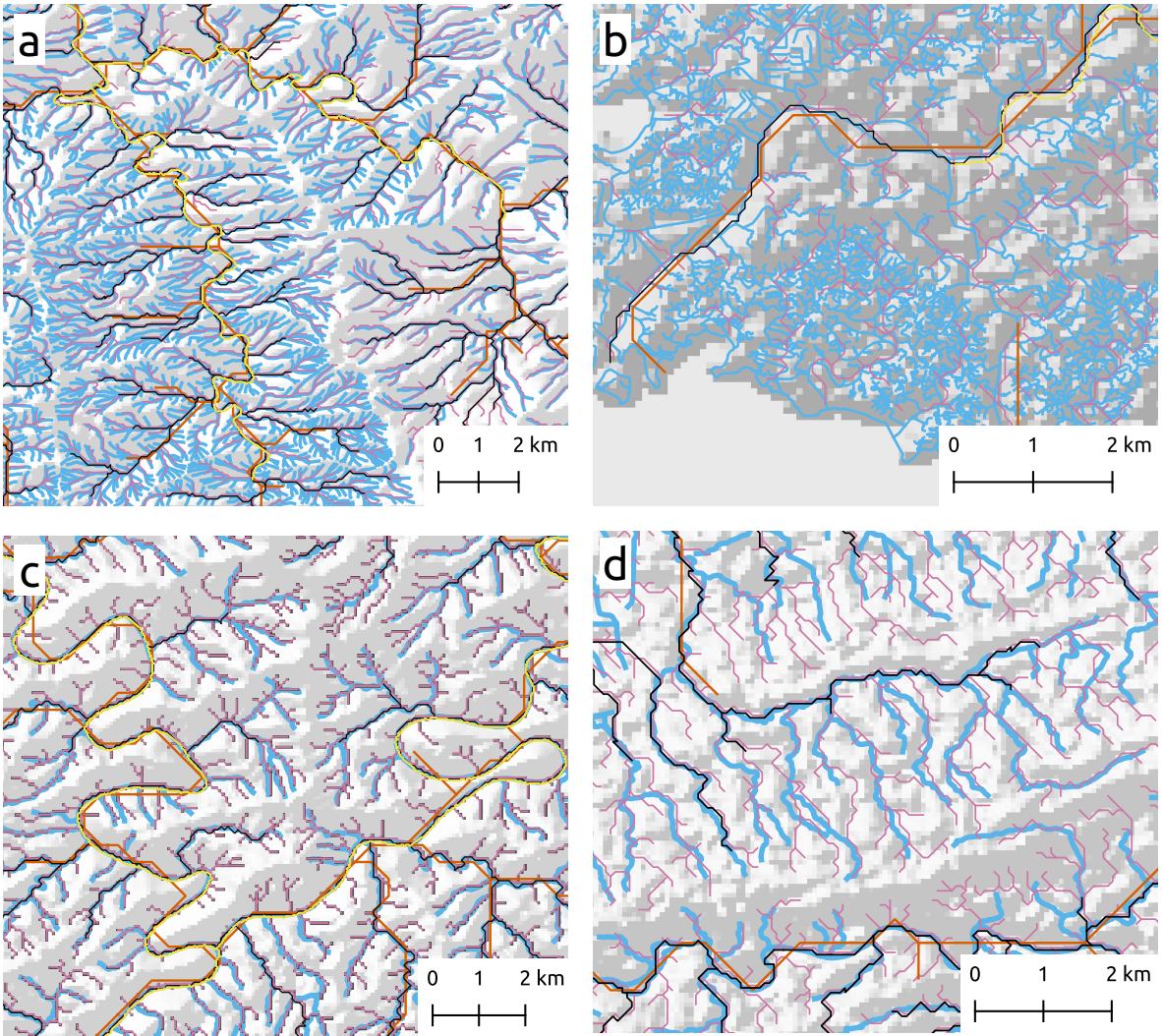
Figure 9. Visualisation of the different global hydrographic datasets (Hydrography90m, MERIT Hydro-Vector, HydroRIVERS, GRWL) against the high-resolution NHDPlus HR reference dataset of the United States (Buto and Anderson, 2020). The geographic location of the four panels a–d corresponds to the labels in the map.

5.2 Flow accumulation accuracy

The NHDPlus HR vector dataset attribute table reports the flow accumulation for each stream segment. The flow accumulation was computed using a 10x10 m 3D Elevation Program (3DEP) (Sugarbaker et al., 2014) DEM. The high resolution of 3DEP allows for precise flow routing and flow accumulation, which can then be used to validate the Hydrography90m flow accumulation. The flow accumulation values are reported in the NHDPlus HR vector attributes, labelled as TotDASqKm. We rasterised these stream segment TotDASqKm attributes to a 3x3 arc-sec grid cell resolution using *gdal_rasterize*. The rasterisation process was performed using the maximum value of the TotDASqKm flow accumulation as standard, in case more than one stream segment fell within the same grid cell. We did not include all the streams that appear inside lakes, due to the emergence of the "fishbone" structure (Domisch et al., 2015a), which depicted stream channels as artificial straight lines due to zero slope. We masked such features using the HydroLAKES dataset (Messager et al., 2016).
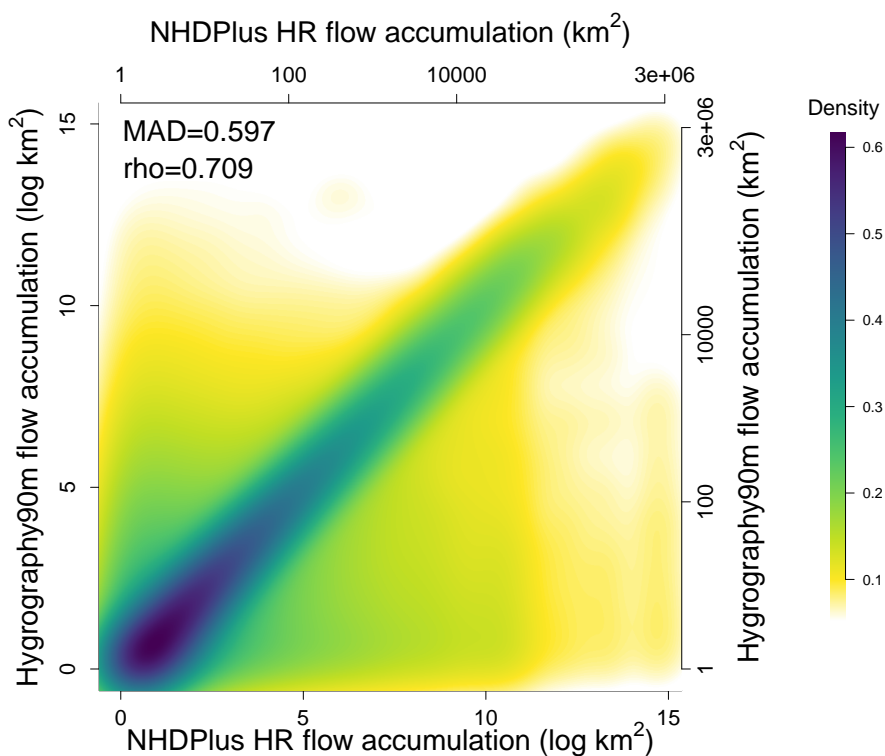
Figure 10. Density plot of Hydrography90m vs NHDPlus HR flow accumulation (log-scale) of those 28 million stream channel grid cells with a flow accumulation larger than 1 km^2. Median Absolute Deviation (MAD) and Spearman coefficient (rho) were computed using all 61 million stream channel grid cells. The colour bar indicates a 2D kernel density estimate.

345   For statistical accuracy, we then selected those rasterised NHDPlus HR streams that overlapped with the Hydrography90m stream channels, and hence extracted the Hydrography90m flow accumulation. 61 million grid cell values were used in this procedure to compute the Median Absolute Deviation (MAD) and Spearman coefficient (rho). These values are depicted in the 2D kernel density scatter-plot in Figure 10.

    In addition to the NHDPlus HR flow accumulation comparison, we compared the surface areas of the ten largest drainage

350 basins worldwide (Table 2) among the HydroBASINS level 3 (Lehner and Grill, 2013) and HYDRO1k (USGS EROS Archive, 2018) datasets. These ten basins drain a substantial amount of 31.5% of the world's land surface (Table 2), and the comparison showed in general a high agreement in the basin surface area among the datasets.

Table 2. The surface area of the ten largest drainage basins worldwide in km^2, compared among the Hydrogryphy90m, HydroBASINS (Lehner and Grill, 2013) and Hydro1k (USGS EROS Archive, 2018) datasets.

River basin	Hydrography90m	HydroBASINS	HYDRO1k
Amazon	5,831,589	5,912,923	5,880,705
Congo	3,675,948	3,705,302	3,699,151
Mississippi	3,210,426	3,240,617	3,221,209
Nile	3,043,683	3,057,772	3,078,270
Ob	2,575,465	3,088,705	2,950,753
Yenisey	2,554,228	2,505,668	2,564,574
Paraná	2,512,913	2,626,303	2,713,595
Lena	2,462,216	2,453,648	2,397,583
Yangtze	2,652,142	1,924,625	1,919,832
Niger	2,022,256	2,122,996	2,117,687

6   Stream topographic and topological variables

    In addition to the base hydrography layers (flow direction and flow accumulation) and network layers (drainage basins, stream

355 channels and sub-catchments), we produced layers that characterise the topographic and topological properties of the hydrography. These variables were computed along stream channels, e.g. stream slope, or across continuous land surfaces, e.g. distance to the stream.

6.1   Stream slope

    We calculated various stream channel slope properties at 3x3 arc-sec stream grid cell resolution (as opposed to segment-level),

360 (Figure 1), including minimum and maximum curvatures, gradient (slope), and elevation differences across the hydrography using the *r.stream.slope* (Jasiewicz and Metz, 2011) GRASS GIS module. Stream slope metrics were calculated between the current cell and the adjacent downstream and upstream cell. Stream channel properties, such as curvature, can be important for

estimating channel bank shear stress and channel evolution (Buraas et al., 2014), whereas stream slope can be used in ecological studies as an indirect indicator of flow velocity and gas transfer velocities (Raymond et al., 2012; Kuemmerlen et al., 2014).

365 All stream slope variables play a role in in-stream sediment transport (Yang, 1977) and the calculation of hydraulic flow and stream power (Hankin et al., 2019).

Table 3. Curvature, gradient (elevation difference divided by distance), and elevation difference raster maps computed with the *r.stream.slope* GRASS GIS module; map reference corresponding to Figure 11; specific GRASS GIS command; and output layer name.

Output raster map description	Map reference	Unit	Command	Output tif file name
Maximum curvature between highest upstream cell, focal cell and downstream cell	Figure 11a	m^{-1} (scale factor 10^6)	r.stream.slope maxcurv=slope_curv_max_dw_cel	slope_curv_max_dw_cel_*.tif
Minimum curvature between lowest upstream cell, focal cell and downstream cell	Figure 11b	m^{-1} (scale factor 10^6)	r.stream.slope mincurv=slope_curv_min_dw_cel	slope_curv_min_dw_cel_*.tif
Elevation difference between focal cell and downstream cell	Figure 11c	m	r.stream.slope difference=slope_elv_dw_cel	slope_elv_dw_cel_*.tif
Focal cell gradient	Figure 11d	unit-less (scale factor 10^6)	r.stream.slope gradient=slope_grad_dw_cel	slope_grad_dw_cel_*.tif
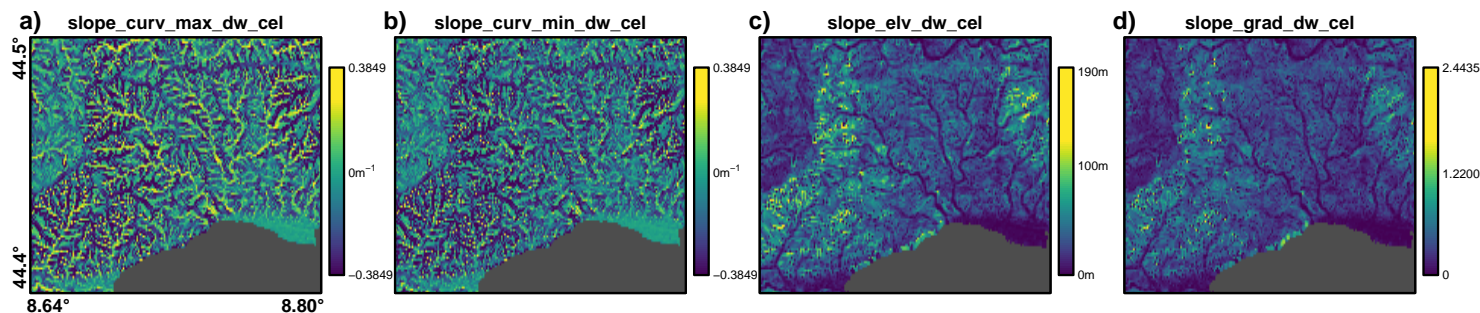
Figure 11. Maps (a–b) show curvature and gradient (elevation difference divided by distance) attributes of each land grid cell to the closest cell along the stream channel computed using the *r.stream.slope* GRASS GIS module. The panel letters correspond to those in Table 3.

6.2 Stream distance

We calculated various distance metrics by setting the (i) stream channels, (ii) outlets or (iii) stream nodes as starting points, using the *r.stream.distance* (Jasiewicz and Metz, 2011) GRASS GIS module. These metrics were based on the distance to (or elevation difference between) the shortest (nearest) or longest (farthest) paths calculated along the upstream and downstream directions. In the case of upstream direction, the shortest or longest paths are given by the MD8 algorithm, which distributes flow accumulation to multiple grid cells. Therefore, the nearest path is considered the shortest trajectory that the largest quantity of water follows from a given stream and/or focal cell to the drainage divide, while the farthest one represents the longest possible path, and is the one that receives less water. Instead, for the downstream direction, water always follows the shortest path going from divide to stream.

These metrics are important for estimating the peak-to-valley time-lag effects of water flow, and can aid in the prediction of travel time. The stream/outlet distance raster files are listed in Table 4. Both euclidean ("as-the-crow-flies") and dendritic ("as-the-fish-swims") stream distance metrics along the network are widely implemented in spatial species distribution modelling, with the latter metric being more effective in modelling the dispersal of aquatic organisms (Mozzaquattro et al., 2020; Altermatt, 2013; Grant et al., 2007). Moreover, stream distance metrics are essential for calculating sediment delivery ratios (Walling, 1983).

Table 4. Stream/outlet distance and elevation difference raster maps computed with the *r.stream.distance* GRASS GIS module; map reference corresponding to Figure 12; unit; GRASS GIS command; and output layer name.

Output raster map description	Map reference	Unit	GRASS GIS command	Output tif file name
Shortest upstream distance between focal grid cell and the nearest sub-catchment drainage divide	Figure 12b	m	r.stream.distance -n method=upstream distance=distance_stream_upstream	stream_dist_up_near_*.tif
Longest upstream distance between focal grid cell and the nearest sub-catchment drainage divide	Figure 12c	m	r.stream.distance method=upstream distance=distance_stream_upstream	stream_dist_up_farth_*.tif
Distance between focal grid cell and its nearest downstream stream grid cell	Figure 12d	m	r.stream.distance method=downstream distance=distance_stream_downstream	stream_dist_dw_near_*.tif
Distance between focal grid cell and the outlet grid cell in the network	Figure 12f	m	r.stream.distance -o method=downstream distance=distance_stream_upstream	outlet_dist_dw_basin_*.tif
Distance between focal grid cell and the downstream stream node grid cell	Figure 12g	m	r.stream.distance -o -s method=downstream distance=distance_stream_upstream	outlet_dist_dw_scatch_*.tif
Euclidean distance between focal grid cell and the stream network	Figure 12h	m	r.grow.distance -m metric=geodesic distance=streams_proximity	stream_dist_proximity_*.tif
Elevation difference of the shortest path from focal grid cell to the sub-catchment drainage divide	Figure 12i	m	r.stream.distance -n method=upstream difference=difference_stream_upstream	stream_diff_up_near_*.tif
Elevation difference of the longest path from focal grid cell to the sub-catchment drainage divide	Figure 12j	m	r.stream.distance method=upstream difference=difference_stream_upstream	stream_diff_up_farth_*.tif
Elevation difference between focal grid cell and its nearest downstream stream pixel	Figure 12k	m	r.stream.distance method=downstream difference=difference_stream_upstream	stream_diff_dw_near_*.tif
Elevation difference between focal grid cell and the outlet grid cell in the network	Figure 12l	m	r.stream.distance -o method=downstream difference=difference_stream_upstream	outlet_diff_dw_basin_*.tif
Elevation difference between focal grid cell and the downstream stream node grid cell	Figure 12m	m	r.stream.distance -o -s method=downstream difference=difference_stream_upstream	outlet_diff_dw_scatch_*.tif
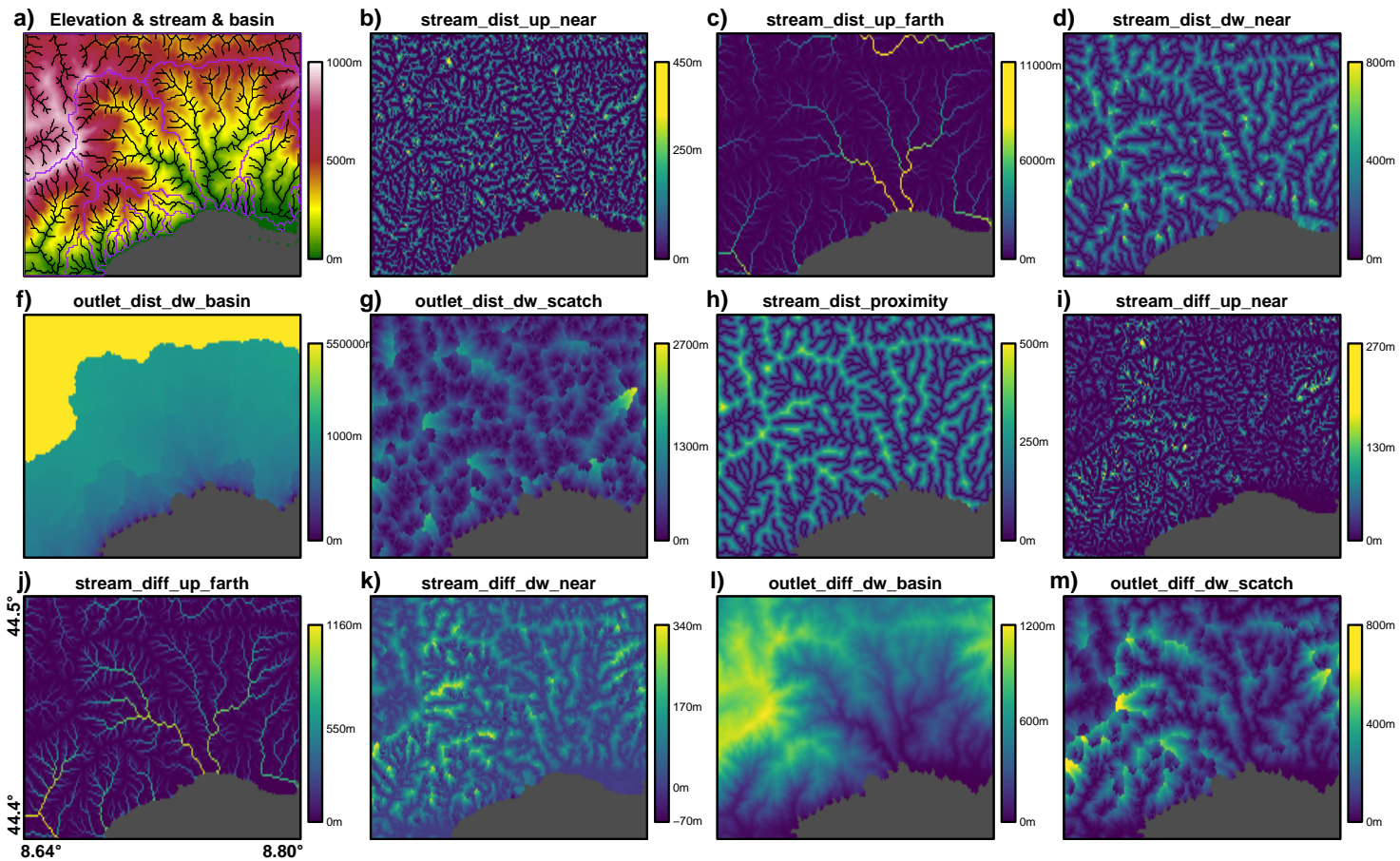
Figure 12. Map (a) shows the stream channels and drainage basins derived from the elevation layer. Maps (b–m) show, for the same area, the distance and elevation difference attributes of each land grid cell to the stream channels, outlets or stream nodes using the *r.stream.distance* GRASS GIS module. The panel letters correspond to those in Table 4.

6.3 Stream segment properties

We calculated segment properties of the stream channels (as opposed to calculating within individual grid cells), (Figure 1) across the hydrography, including the up/down-stream elevation difference, gradient (elevation difference divided by distance),

385 and curvature within each stream segment using the *r.stream.channel* (Jasiewicz and Metz, 2011) GRASS GIS module. The segment properties of the stream channels were calculated downstream for every segment from its initialisation to the outlet/node or from a focal cell to the outlet/node (Figure 1). In contrast, the upstream calculation is done in the opposite direction (from the outlet/node to the initialisation). These stream variables relate only to the stream segments (i.e. across stream channels), as opposed to the stream distance variables that were calculated across the continuous land surface (i.e., sub-catchments). Stream

390 segment properties can be used to classify and distinguish streams, e.g. hydrological delineation of watersheds into similar sub-basins or for in-stream assessments of river structure (Brenden et al., 2008).

Table 5. Curvature, gradient (elevation difference divided by distance), and elevation change raster maps computed with the *r.stream.channel* GRASS GIS module; map reference corresponding to Figure 13; unit; GRASS GIS command; and output layer name.

Output raster map description	Map reference	Unit	GRASS GIS command	Output tif file name
Segment downstream mean gradient between focal cell and the node/outlet	Figure 13b	unit-less (scale factor 10^6)	r.stream.channel -d gradient=channel_grad_dw_seg	channel_grad_dw_seg_*.tif
Segment upstream mean gradient between focal cell and the init/node	Figure 13c	unit-less (scale factor 10^6)	r.stream.channel gradient=channel_grad_up_seg	channel_grad_up_seg_*.tif
Upstream gradient between focal cell and the next cell	Figure 13d	unit-less (scale factor 10^6)	r.stream.channel -l gradient=channel_grad_up_cel	channel_grad_up_cel_*.tif
Cell stream course curvature of the focal cell	Figure 13f	m^{-1} (scale factor 10^6)	r.stream.channel curvature=channel_curv_cel	channel_curv_cel_*.tif
Segment downstream elevation difference between focal cell and the node/outlet	Figure 13g	m	r.stream.channel -d difference=channel_elv_dw_seg	channel_elv_dw_seg_*.tif
Segment upstream elevation difference between focal cell and the init/node	Figure 13h	m	r.stream.channel difference=channel_elv_up_seg	channel_elv_up_seg_*.tif
Upstream elevation difference between focal cell and the next cell	Figure 13i	m (outlet cell value = 99999)	r.stream.channel -l difference=channel_elv_up_cel	channel_elv_up_cel_*.tif
Downstream elevation difference between focal cell and the next cell	Figure 13j	m	r.stream.channel -l -d difference=channel_elv_dw_cel	channel_elv_dw_cel_*.tif
Segment downstream distance between focal cell and the node/outlet	Figure 13k	m	r.stream.channel -d distance=channel_dist_dw_seg	channel_dist_dw_seg_*.tif
Segment upstream distance between focal cell and the init/node	Figure 13l	m	r.stream.channel distance=channel_dist_up_seg	channel_dist_up_seg_*.tif
Upstream distance between focal cell and next cell	Figure 13m	m	r.stream.channel -l distance=channel_dist_up_cel	channel_dist_up_cel_*.tif
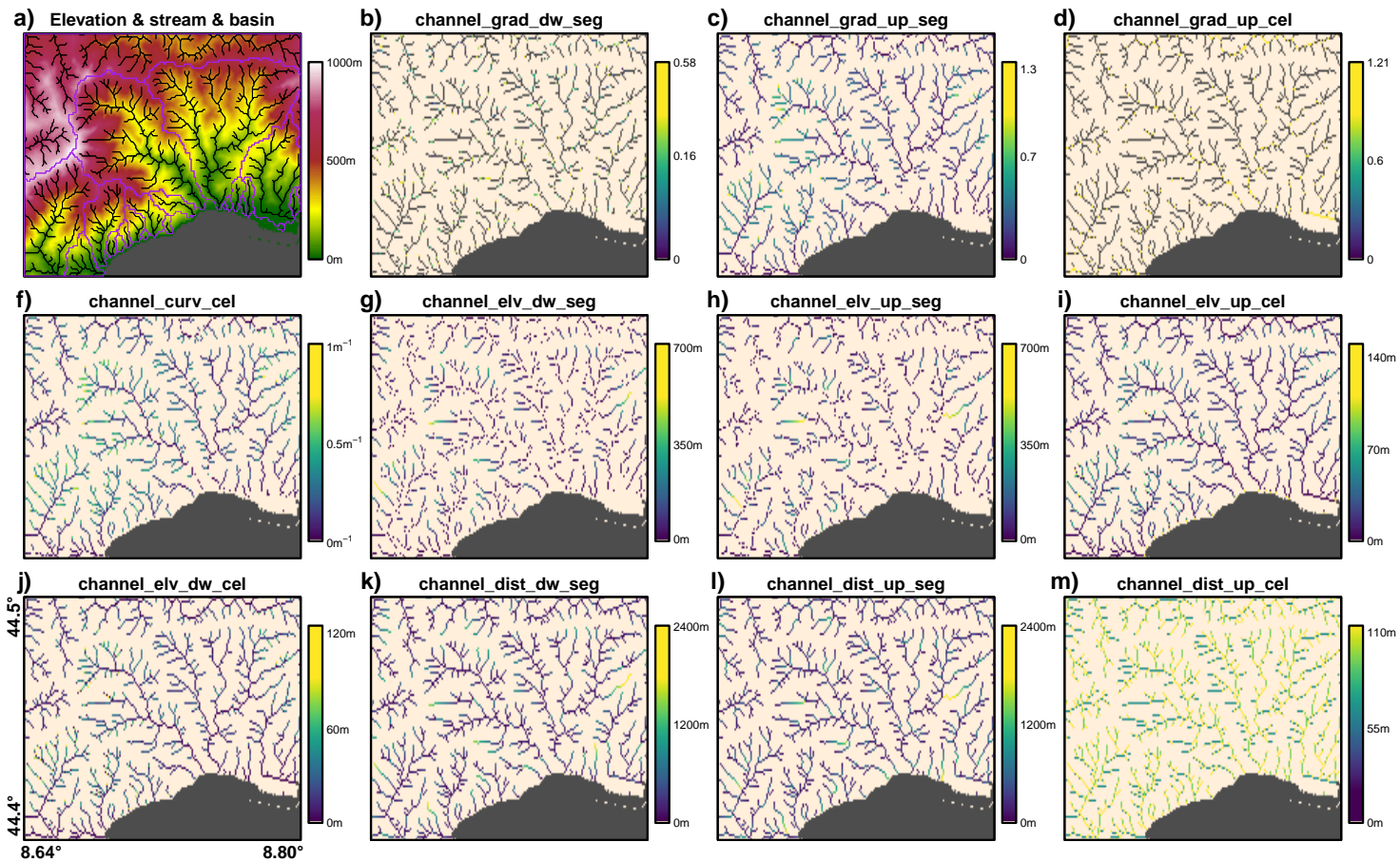
Figure 13. Map (a) shows The stream channels and drainage basins derived from the elevation layer. Maps (b–m) show, for the same area, the curvature, gradient (elevation difference divided by distance), and elevation change raster maps computed with *r.stream.channel* GRASS GIS module. The panel letters correspond to those in Table 5.

6.4 Stream order

We calculated a suite of topological stream order layers at the segment level. Stream order is depicted as a positive integer for indicating the level of branching in the river network (Zhang et al., 2007a; Scheidegger, 1965). There are various approaches
395 to stream ordering, which either start from the source of the river or from the outlet. We used the *r.stream.order* (Jasiewicz and Metz, 2011) module and calculated stream order using the following methods: Strahler's (Strahler, 1957), Hortons's (Horton, 1945), Shereve's (Shreve, 1967), Hack's (Hack, 1957) and topological stream hierarchy (Marani et al., 1991). We provided each stream order layer as an individual raster file, and all stream orders within the stream vector topology attribute table (Table 6). For all items reported in the stream vector topology attribute table, refer to the *r.stream.order* GRASS GIS manual page
400 (https://grass.osgeo.org/grass78/manuals/addons/r.stream.order.html). From a hydrography point of view, the stream order is used in the River Continuum Concept and therefore provides the basis for distinguishing ecological processes from headwaters to river mouths (Vannote et al., 1980; Thoms et al., 2018).

Table 6. Stream order raster and vector files computed with the *r.stream.order* GRASS GIS module, the map reference corresponding to Figure 14, the specific GRASS GIS command and the layer output name.

Output map description	Map reference	Command	Output tif file name
Strahler's stream order (raster)	Figure 14a	r.stream.order strahler=order	order_strahler_*.tif
Shreve's stream magnitude (raster)	Figure 14b	r.stream.order shreve=order	order_shreve_*.tif
Horton's stream order (raster)	Figure 14c	r.stream.order horton=order	order_horton_*.tif
Hack's stream order (raster)	Figure 14d	r.stream.order hack=order	order_hack_*.tif
Topological dimension of streams (raster)	Figure 14f	r.stream.order topo=order	order_topo_*.tif
All stream nodes attributes (vector)	Figure 14g	r.stream.order order_vect=vect	stream_vect_point_*.gpkg
All stream segments attributes (vector)	Figure 14g	r.stream.order order_vect=vect	stream_vect_segment_*.gpkg
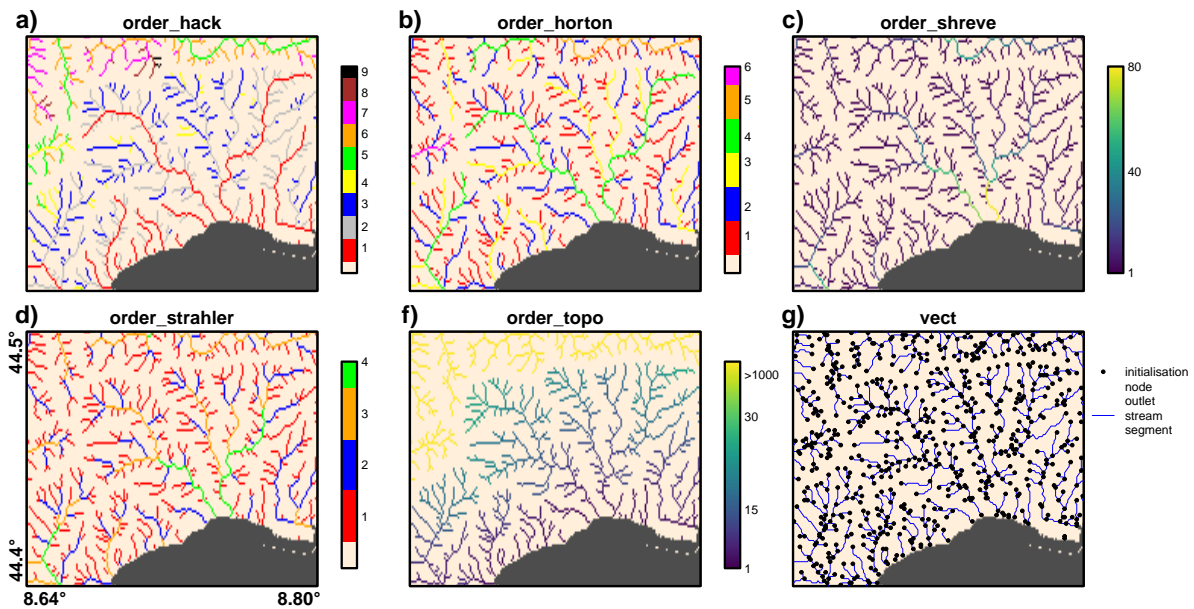
Figure 14. Maps (a–f) show different stream order types computed with the *r.stream.order* GRASS GIS module. All stream order layers are also available as vector data together with their attribute table. Map (g) shows the blue stream segments in vector format with the initialisation, node and outlet vertices labelled as black points in Table 6.

6.5 Flow index

Using flow accumulation and terrain slope we calculated three flow indices at the grid cell resolution: the compound topographic index (cti, or topographic wetness index), the stream power index (spi) and the stream transportation index (sti, Table 7).

The stream power index (spi) (Moore et al., 1991) is computed as the product of the upstream catchment area and the tangent of the terrain slope angle. The stream power index represents the erosive power associated with flow and the gravitational forces that move water downstream (Moore et al., 1991). It is commonly used in soil erosion models (Thalacker, 2014), landslide susceptibility (Pourghasemi et al., 2012) and groundwater estimations (Ozdemir, 2011).

The sediment transport index (sti) (Moore and Burch, 1986) is derived from unit stream-power theory and is equivalent to the length–slope factor in the Revised Universal Soil Loss Equation (RUSLE) (Moore et al., 1991). It is often used to represent the erosive power of surface flow for landslides (Pourghasemi et al., 2012; Hong et al., 2017) or debris-flow modelling (Lay et al., 2019).

The compound topographic index (cti) (Beven and Kirkby, 1979), also known as topographic wetness index, is a steady state wetness index, and is computed as the logarithm of the cumulative upstream catchment area divided by the tangent of the terrain slope angle. This index is a proxy for long-term soil moisture availability (Raduła et al., 2018). It has been often used in species distribution modelling, species richness and composition analyses, as well as landslide susceptibility and soil carbon assessments (Román-Sánchez et al., 2018; Raduła et al., 2018).

Table 7. The compound topographic index (cti), stream power index (spi) and stream transportation index (sti) derived from flow accumulation (α) and terrain slope (β); map reference corresponding to Figure 15; unit; specific GDAL command; and output layer name.

Output raster map description	Map reference	Unit	Command	Output tif file name
Stream power index (spi)	Figure 15a	unit-less (scale factor 10^3)	gdal_calc.py - -calc="α * $\tan(\beta)$"	spi_*.tif
Stream transportation index (sti)	Figure 15b	unit-less (scale factor 10^3)	gdal_calc.py - -calc="1.4 * (α / 22.13)$^{0.4}$ * ($\sin(\beta)$ / 0.0896)$^{1.3}$"	sti_*.tif
Compound topographic index (cti)	Figure 15c	unit-less (scale factor 10^8)	gdal_calc.py - -calc="$\ln(\alpha / \tan(\beta))$"	cti_*.tif
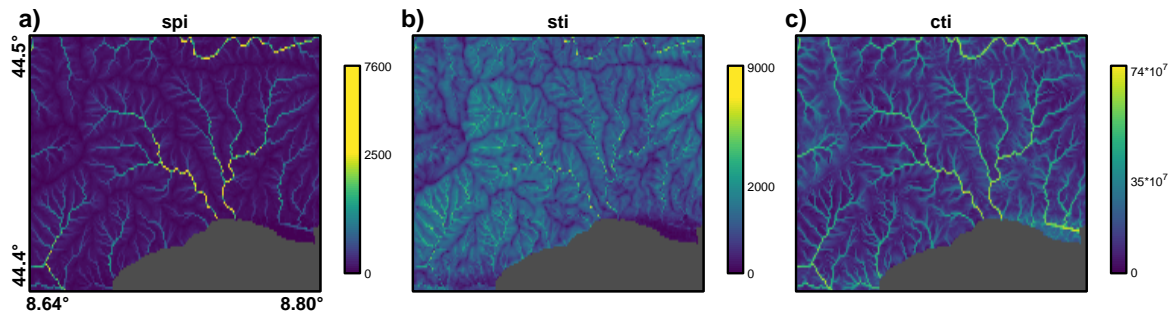
Figure 15. Different flow index layers computed using flow accumulation and terrain slope. Map (a) shows the stream power index (spi), map (b) sediment transport index (sti), and map (c) compound topographic index (cti). The panel letters correspond to those in Table 7.

7 Computational processing framework

In order to produce the standardised Hydrography90m products, we developed Bash scripts that launched each other in a cascading manner as a series of single batch jobs or as job arrays that submit and manage collections of similar jobs. The overall computation, starting from the calculation of flow accumulation to the topographic and topological variable creation, ($\sim$2TB of layer products), which accounted for a total of 52 scripts containing over 4000 code lines, took 12,418 core-hours at the High Performance Computing (HPC) facility of the Center for Research Computing, Yale University.

The entire procedure was run several times during computational development to check for consistency and potential mismatches among the different hydrographic layers. The scripts also employed several bench-marking strategies to check for potential errors in the data flow. The bench-marking strategies focused on:

- Tile geographic extent at integer degree level;
- Predefined ($0.0008\overline{3}$ °) and constant cell resolution during the entire data processing;
- Unique IDs for basins, sub-catchments, and stream segments worldwide;
- Computation of histogram raster values to spot potential outliers;
- Uniform tiling system for all layers;
- Tile resampling at 30x30 arc-sec ($0.008\overline{3}$ °) cells for fast global visualisation;
- Cross-over procedures to obtain consistent results (e.g. outlet number = drainage basins number; stream segments number = sub-catchments number)

8 Discussion

8.1 Methodological considerations

High-resolution information regarding the delineation of drainage basins and stream channels is vital for a wide array of earth system sciences, hydrology, chemistry, freshwater biodiversity research and for informed management applications (Lowe and Likens, 2005; Reichl et al., 2009; Oudin et al., 2008; Maasri et al., 2021b; Amatulli et al., 2018b). Hydrography90m presents such information within a globally standardised and seamless hydrographic dataset that delineates headwaters in unparalleled detail. Hydrography90m is the first-ever data product that allows for global and comparative area-of-interest studies on small headwater stream channels. The high precision of the Hydrography90m has been demonstrated against NHDPlus HR, and achieves high levels of accuracy for stream spatial and flow accumulation values. The increased density of headwaters in Hydrography90m, compared to the bench-mark NHDPlus HR and other reference hydrographies, is a distinctly valuable feature. These headwaters are a crucial component in hydrology and its associated applications (Lowe and Likens, 2005; Finn et al., 2011; Meyer et al., 2007) but have not been depicted globally until this publication. We thus opted for a comprehensive approach that enables headwater mapping at high resolution. To achieve this objective, we delineated the potential headwater stream channels, which were derived entirely from DEM and topographic features.

Through the course of our ongoing research agenda, we will continue to identify unmapped perennial and non-perennial headwater channels by incorporating emerging higher-resolution DEMs (30 m) with benchmarking accuracy procedures (Strobl et al., 2021), and climatic and meteorological data. In arid regions, the delineated headwater stream channels receive water and produce floods (Farquharson et al., 1992), albeit with lower frequency compared to streams in humid and wet climates. However, these flow pulses are important for in-stream aquatic organisms (Bunn et al., 2006) and should not be neglected on a global stream network dataset. To address the frequency and duration of flows in these headwater streams, we are currently modelling discharge using a 30-year monthly climate time-series. This allows for the estimation of mean monthly discharge for each of the 726,723,221 identified stream segments, and provides additional stream flow attributes to Hydrography90m. We shall thus be able to use this information to assign the probability of water occurrence to each stream channel, i.e. a stream channel will appear present if discharge > 0 m^3/sec). The overall outcome of this separate project will be a dynamic hydrological assessment, with stream channel length changing as a function of discharge. The final output will be monthly stream networks that are dynamic, i.e. longer during the rainy season, and shorter and intermittent during dry months. Such monthly temporal assessment widens the scope for an improved and more complete representation of the network than assigning, for example, a channelisation threshold contingent on the geographic region alone, as suggested by Vogt et al. (2003). In summary, we employed a low threshold of 0.05 km^2 in this paper, and in subsequent research, will shorten (prune) the network dynamically given the modelled discharge in the stream segments. Most importantly, the implementation of a low threshold allows the computation of topographic and topological variables for small headwaters streams that can be used to assess flow regimes and stream properties, such as sediment transportation. Pruning the network with a wider threshold in the first place would not allow the computation of such stream variables.

While Hydrography90m offers improved spatial accuracy compared to previous global hydrographic products, it can still benefit from several enhancements in the future. Currently, stream channel bifurcations are not represented in the Hydrography90m, and despite the MD8 algorithm distributing the flow accumulation to multiple grid cells, the stream channel follows only one downstream direction. We note that stream width was not considered in our approach, and due to the 3x3 arc-sec spatial resolution, small headwater streams were located within grid cells. Standing water bodies were not yet integrated in Hydrography90m, and are also part of our ongoing research. In the meantime, we invite users interested in integrating standing water bodies into Hydrography90m to contact the authors for a preliminary product.

Improvements to the state of the art are possible with even higher-resolution digital terrain models than the employed MERIT Hydro. Nonetheless, increased resolution may also introduce the challenge of accounting for man-made canals and other engineered structures. For instance, we found in Hydrography90m that the Tongji Canal in China, part of the Grand Canal that connects the Yellow and Yangtze rivers, routes sections of the Yellow's flow accumulation into the Yangtze, leading to discrepancies among the compared global hydrographic datasets (Table 2). Similarly, we identified missing bifurcations within the Niger river, owing to the lack of bifurcation options in the implemented flow routing algorithm. While also a challenge in other global hydrographic datasets (Lehner et al., 2008; Yamazaki et al., 2019), missing bifurcation routines created difficulties with flow routing in Hydrography90m. Such bifurcations occurred mainly in very flat areas around the globe. Thus, any

discharge computations derived from this dataset would need to account not only for missing bifurcations but also apportion flow among two (or more) channels, according to each tributary's flow capacity.

8.2 Applications

Hydrography90m has been developed with a wide range of natural science applications in mind. In the broadest sense, its appeal lies in the potential for combining and extending the scope of both remote and field sensor technologies. The computational approach behind Hydrography90m not only overcomes spectral limitations but also the spatial and accessibility constraints of conventional resource monitoring instruments. While the dataset's uptake is obviously relevant within hydrography and hydrology, its value in numerous other geoinformatics applications is also well-recognised.

The high-resolution base layers (flow accumulation and flow direction) and network hydrography layers (drainage basins, stream segments and sub-catchments) can inform studies on flow estimation, sediment transport and ecology. For example, flow accumulation has been utilised in flood susceptibility mapping models, and often serves as a proxy for discharge in ecological modelling (Shafizadeh-Moghadam et al., 2018; Kuemmerlen et al., 2014). Flow direction has been used in metacommunity structure studies (Mozzaquattro et al., 2020), whereas drainage basins and sub-catchments can be used as the spatial units in species distribution modelling (Altermatt et al., 2013; Read et al., 2015). The Hydrography90m products can provide fundamental information for modelling the high-resolution supply and demand of biogeochemically relevant constituents that, so far, have been modelled using hydraulic information derived from low-resolution datasets (Raymond et al., 2016; Wollheim et al., 2018). These layers are a vital input in modelling species distribution (Domisch et al., 2019) and for monitoring invasive species (Haubrock et al., 2022), whether for biodiversity monitoring or for public health measures to combat vector-borne diseases (Bishop et al., 2021; Pless et al., 2021; Saarman et al., 2019, 2018). Specifically, sub-catchments have been used to derive zonal statistics from topographic and environmental layers for small scale species distribution models (Kuemmerlen et al., 2014). Finally, this novel network has the appreciable potential to guide integrated freshwater conservation efforts given its delineation of headwater stream channels, which are typically species-rich (Abell et al., 2007).

Beyond their direct applications, the layers within Hydrography90m also offer important spatial and statistical information. For instance, the assessment of catchment similarity is useful for the prediction of ungauged basins (Reichl et al., 2009). Concurrently, in machine learning-based approaches, e.g. Long Short-Term Memory (LSTM) models (Kratzert et al., 2019), additional information on catchment attributes is highly sought-after and advantageous to model accuracy.

Besides scientific studies, the aforementioned analyses would serve to address major geopolitical and natural resource challenges involving transboundary rights and water security, water resource management and food production (water quantity, quality and nutrient flows), and catastrophe risk management (flooding, erosion, and drought), to name but a few. Nowadays, such issues notably fall under the ambit of several of the United Nations' Sustainable Development Goals (SDGs; Connor, 2015), such as the ones that concern water resource management, human health, peaceful and equitable societies, and sustainable economic development (Blöschl et al., 2017).

With regard to the methodology and computational process flows employed in Hydrography90m, we note below some key advantages, as well as considerations for improvement. At the very outset of the dataset computation, a suite of topological and

topographic feature layers accompany the base and network layers. While the former set of layers have been generated previously in coarse spatial resolution (Lehner et al., 2008; Linke et al., 2019; Domisch et al., 2015a), or in higher regional to local resolutions (Domisch et al., 2015b), we anticipated that the more comprehensive, high-resolution layers of Hydrography90m will significantly reduce the burden of ad hoc area-limited computations. Such globally available and analysis-ready data is also in line with our previously released Geomorpho90m (Amatulli et al., 2020) dataset, which seamlessly characterised global land surface using a collection of 26 MERIT-DEM-derived geomorphometry variables (Amatulli et al., 2020). The assimilation of this globally standardised data with various environmental, climate, stream flow (GSIM; Do et al., 2018; Gudmundsson et al., 2018) and freshwater quality observations (GRQA; Virro et al., 2021) provides the requisite quantum of inputs to implement a machine learning approach for high-resolution discharge and quality predictions.

9 Conclusions

In this study, we constructed Hydrography90m, as a globally seamless and standardised hydrographic network with associated stream topographic and topological features. These latter supplementary layers were carefully developed to ensure consistency and compatibility among all of the presented hydrography layers.

The data validation procedures confirmed Hydrography90m as a more accurate representation of stream networks compared to HydroRIVERS, GRWL, and MERIT Hydro–Vector. Improved accuracy was achieved principally by employing a higher resolution DEM, the MD8 flow routing algorithm, and a markedly smaller flow accumulation threshold to initiate stream channels. With these characteristics, Hydrography90m provides a valuable basis for supporting a variety of freshwater-related research disciplines.

Moreover, Hydrography90m is currently being further processed to (i) exclude the DEM-derived headwater streams that are not hydrologically relevant, (ii) include a suite of high-resolution (both spatial and temporal) environmental attributes for each of the 726 million stream segments, (iii) and integrate standing water bodies within the hydrographic network.

10 Usage notes

Below we report the web access points for data description and download.

- The Hydrography90m metadata are reported at https://doi.org/10.18728/igb-fred-762.1.

- The tiled Hydrography90m files are stored at https://public.igb-berlin.de/index.php/s/agciopgzXjWswF4 repository.

- The download procedure via a map interface or a scripting procedure is available at http://hydrography.org.

- The online visualisation of the layers via the IGB-GeoNode can be accessed via https://hydrography.org/hydrography90m/hydrography90m_layers/

The layers in Hydrography90m are compatible with any standard GIS application. We encourage, however, to use the open-source QGIS and GRASS GIS tools to further process the data. The benefit lies in the seamless integration with the processing algorithms as well as the identical spatial definition of the regional and grid cell extents.

550   Since we accounted for inland depressions, the stream channel network may terminate in these depression locations. We provide the regional unit layer (useful for extracting entire basins), allowing the seamless integration of those interior drainage basins into their surrounding and larger basins neighbourhoods.

For a given study area, we recommend users to check the tile ID for the area of interest. The basins, stream channels or sub-catchments will be split at the tile border, and a standard merge (raster) or dissolve by ID (vector) operation can mosaic the data

555   together. If any smaller, surrounding drainage basins should be discarded, we then recommend masking the mosaicked tiles with the specific drainage basin IDs in the regional unit raster. This results in keeping only those drainage basins of interest. We provide example and helper functions for downloading, merging, cropping and masking the data at http://hydrography.org.

Data availability. All metadata of the the Hydrography90m dataset can be found at https://doi.org/10.18728/igb-fred-762.1. In the DOI's landing page are reported the links for the IGB-GeoNode visualization and for the data storing repository. The repository includes an

560   "README.txt file" that explains the folders structure and file names. In addition the data can be directly downloaded at http://hydrography. org by clicking on a given tile on the map.

Code availability. At http://hydrography.org we reported a schematic scripting procedure of the workflow together with a procedure example

570   to merge tiles and operate with regional units.

Video supplement. We describe the main features of the Hydrography90m dataset in a 3 minutes video visible at https://doi.org/10.5446/56343.

Author contributions. G.A and S.D designed the study. G.A. developed/implemented/bench-marked the workflow and processing chain in the Yale-HPC to compute the Hydrography90m layers. J.G.M. and S.D. designed and performed the processing required for the NHDplus

HR comparison and validation assessment. A.G. and M.U. performed the geodata verification. L.S. contributed to development of the DEM section and code trouble shooting during the computational implementation. All authors discussed the results. G.A. and S.D. wrote the first manuscript draft, and all authors contributed to the writing of the manuscript. T.S. designed and assembled the video.

Competing interests. The authors declare no competing interests.

Disclaimer. The authors accept no responsibility for any liability arising from the use of this research paper and its associated dataset.

Acknowledgements. We thank the Yale Center for Research Computing for their guidance and the use of research computing infrastructure, specifically Dr. Thomas Langford, Dr. Kaylea Nelson, and Tyler Trafford. We also thank Lt. Col. Richard White (Retd.) for proofreading the article and offering useful critiques from a user/reader perspective. Special thanks go to the GRASS GIS development team. G.A. was supported by Yale University, in particular the School of the Environment & Centre for Research Computing and by the NSF grant 'GR103812 EAGER Watershed rules of life'. We acknowledge funding by the German Federal Ministry of Education and Research (BMBF grant agreement number 033W034A). This work is part of the Global Freshwater Biodiversity, Biogeography and Conservation project (https://glowabio.org), and J.G.M., A.G., M.U. and S.D. acknowledge the Leibniz Competition funding awarded to S.D. (J45/2018).

References

Abell, R., Allan, J. D., and Lehner, B.: Unlocking the potential of protected areas for freshwaters, Biological Conservation, 134, 48–63, 2007.

Allen, G. H., David, C. H., Andreadis, K. M., Hossain, F., and Famiglietti, J. S.: Global estimates of river flow wave travel times and implications for low-latency satellite data, Geophysical Research Letters, 45, 7551–7560, 2018.

Altermatt, F.: Diversity in riverine metacommunities: a network perspective, Aquatic Ecology, 47, 365–377, 2013.

Altermatt, F., Seymour, M., and Martinez, N.: River network properties shape α-diversity and community similarity patterns of aquatic insect communities across major drainage basins, Journal of Biogeography, 40, 2249–2260, 2013.

Amatulli, G.: A new and extendable global watershed and stream network delineation using GRASS-GIS, Geomorphometry 2020, p. 205, 2020.

Amatulli, G., Casalegno, S., D'Annunzio, R., Haapanen, R., Kempeneers, P., Lindquist, E., Pekkarinen, A., M., W. A., and R., Z.-M.: Teaching spatiotemporal analysis and efficient data processing in open source environment, in: Proceedings of the 3rd Open Source Geospatial Research& Education Symposium, p. 13, 2014.

Amatulli, G., Domisch, S., Kiesel, J., Sethi, T., Yamazaki, D., and Raymond, P.: High-resolution stream network delineation using digital elevation models: assessing the spatial accuracy, Tech. rep., PeerJ Preprints, 2018a.

Amatulli, G., Domisch, S., Tuanmu, M.-N., Parmentier, B., Ranipeta, A., Malczyk, J., and Jetz, W.: A suite of global, cross-scale topographic variables for environmental and biodiversity modeling, Scientific Data, 5, 180 040, 2018b.

Amatulli, G., McInerney, D., Sethi, T., Strobl, P., and Domisch, S.: Geomorpho90m, empirical evaluation and accuracy assessment of global high-resolution geomorphometric layers, Scientific Data, 7, 1–18, 2020.

Benstead, J. P. and Leigh, D. S.: An expanded role for river networks, Nature Geoscience, 5, 678–679, 2012.

Beven, K. J. and Kirkby, M. J.: A physically based, variable contributing area model of basin hydrology/Un modèle à base physique de zone d'appel variable de l'hydrologie du bassin versant, Hydrological Sciences Journal, 24, 43–69, 1979.

Bishop, A. P., Amatulli, G., Hyseni, C., Pless, E., Bateta, R., Okeyo, W. A., Mireji, P. O., Okoth, S., Malele, I., Murilla, G., et al.: A machine learning approach to integrating genetic and ecological data in tsetse flies (Glossina pallidipes) for spatially explicit vector control planning, Evolutionary Applications, 14, 1762, 2021.

Blöschl, G., Hall, J., Parajka, J., Perdigão, R. A., Merz, B., Arheimer, B., Aronica, G. T., Bilibashi, A., Bonacci, O., Borga, M., et al.: Changing climate shifts timing of European floods, Science, 357, 588–590, 2017.

Brenden, T., Wang, L., Seelbach, P., Clark, R., Wiley, M., and Sparks-Jackson, B.: A spatially constrained clustering program for river valley segment delineation from GIS digital river networks, Environmental Modelling & Software, 23, 638–649, 2008.

Bunn, S. E., Thoms, M. C., Hamilton, S. K., and Capon, S. J.: Flow variability in dryland rivers: boom, bust and the bits in between, River Research and Applications, 22, 179–186, https://doi.org/https://doi.org/10.1002/rra.904, 2006.

Buraas, E. M., Renshaw, C. E., Magilligan, F. J., and Dade, W. B.: Impact of reach geometry on stream channel sensitivity to extreme floods, Earth Surface Processes and Landforms, 39, 1778–1789, 2014.

Buto, S. G. and Anderson, R. D.: NHDPlus High Resolution (NHDPlus HR)—A hydrography framework for the Nation, Tech. rep., US Geological Survey, 2020.

Connor, R.: The United Nations world water development report 2015: water for a sustainable world, vol. 1, UNESCO publishing, 2015.

Datry, T., Larned, S. T., and Tockner, K.: Intermittent Rivers: A Challenge for Freshwater Ecology, BioScience, 64, 229–235, 2014.

Do, H. X., Gudmundsson, L., Leonard, M., Westra, S., and Grabs, W.: The Global Streamflow Indices and Metadata Archive (GSIM)-Part 1: The production of a daily streamflow archive and metadata., Earth System Science Data, 10, 2018.

625 Domisch, S., Amatulli, G., and Jetz, W.: Near-global freshwater-specific environmental variables for biodiversity analyses in 1 km resolution, Scientific Data, 2, 1–13, 2015a.

Domisch, S., Jaehnig, S. C., Simaika, J. P., Kuemmerlen, M., and Stoll, S.: Application of species distribution models in stream ecosystems: the challenges of spatial and temporal scale, environmental predictors and species occurrence data, Fundamental and Applied Limnology, pp. 45–61, 2015b.

630 Domisch, S., Friedrichs, M., Hein, T., Borgwardt, F., Wetzig, A., Jähnig, S. C., and Langhans, S. D.: Spatially explicit species distribution models: A missed opportunity in conservation planning?, Diversity and Distributions, 25, 758–769, 2019.

Erskine, R. H., Green, T. R., Ramirez, J. A., and MacDonald, L. H.: Comparison of grid-based algorithms for computing upslope contributing area, Water Resources Research, 42, 2006.

Farquharson, F., Meigh, J., and Sutcliffe, J.: Regional flood frequency analysis in arid and semi-arid areas, Journal of Hydrology, 138,
635 487–501, https://doi.org/https://doi.org/10.1016/0022-1694(92)90132-F, 1992.

Farr, T. G., Rosen, P. A., Caro, E., Crippen, R., Duren, R., Hensley, S., Kobrick, M., Paller, M., Rodriguez, E., Roth, L., et al.: The shuttle radar topography mission, Reviews of Geophysics, 45, 2007.

Finn, D. S., Bonada, N., Múrria, C., and Hughes, J. M.: Small but mighty: headwaters are vital to stream network biodiversity at two levels of organization, Journal of the North American Benthological Society, 30, 963–980, 2011.

640 Gallant, J. and Wilson, J.: Terrain analysis: principles and applications, John Wiley & Sons, 2000.

GDAL Development Team: GDAL - Geospatial Data Abstraction Library, Version 3.1.0, Open Source Geospatial Foundation, http://www.gdal.org, 2020.

Grant, E. H. C., Lowe, W. H., and Fagan, W. F.: Living in the branches: population dynamics and ecological processes in dendritic networks., Ecology Letters, 10 2, 165–75, 2007.

645 GRASS Development Team: Geographic Resources Analysis Support System (GRASS GIS) Software, Version 7.8.0, Open Source Geospatial Foundation, http://grass.osgeo.org, 2019.

Gudmundsson, L., Do, H. X., Leonard, M., and Westra, S.: The Global Streamflow Indices and Metadata Archive (GSIM)–Part 2: Quality control, time-series indices and homogeneity assessment, Earth System Science Data, 10, 787–804, 2018.

Hack, J. T.: Studies of longitudinal stream profiles in Virginia and Maryland, vol. 294, US Government Printing Office, 1957.

650 Hankin, B., Metcalfe, P., Beven, K., and Chappell, N. A.: Integration of hillslope hydrology and 2D hydraulic modelling for natural flood management, Hydrology Research, 50, 1535—-1548, 2019.

Hansen, M., Potapov, P., Moore, R., Hancher, M., Turubanova, S., Tyukavina, A., Thau, D., Stehman, S., Goetz, S., Loveland, T., Kommareddy, A., Egorov, A., Chini, L., Justice, C., and Townshend, J.: High-resolution global maps of 21st-century forest cover change, Science, 342, 850–853, 2013.

655 Harding, M. and Carabajal, C.: ICESat waveform measurements of within-footprint topographic relief and vegetation vertical structure, Geophysical Research Letters, 32, 2005.

Haubrock, P. J., Ahmed, D. A., Cuthbert, R. N., Stubbington, R., Domisch, S., Marquez, J. R., Beidas, A., Amatulli, G., Kiesel, J., Shen, L. Q., et al.: Invasion impacts and dynamics of a European-wide introduced species, Global Change Biology, 2022.

Heine, R. A., Lant, C. L., and Sengupta, R. R.: Development and comparison of approaches for automated mapping of stream channel
660 networks, Annals of the Association of American Geographers, 94, 477–490, 2004.

Hirt, C.: Artefact detection in global digital elevation models (DEMs): The Maximum Slope Approach and its application for complete screening of the SRTM v4. 1 and MERIT DEMs, Remote Sensing of Environment, 207, 27–41, 2018.

Holmgren, P.: Multiple flow direction algorithms for runoff modelling in grid based elevation models: an empirical evaluation, Hydrological processes, 8, 327–334, 1994.

665 Hong, H., Tsangaratos, P., Ilia, I., Chen, W., and Xu, C.: Comparing the performance of a logistic regression and a random forest model in landslide susceptibility assessments. The Case of Wuyaun Area, China, in: Workshop on World Landslide Forum, pp. 1043–1050, 2017.

Horton, R. E.: Erosional development of streams and their drainage basins; hydrophysical approach to quantitative morphology, Geological society of America bulletin, 56, 275–370, 1945.

Hosen, J. D., Allen, G. H., Amatulli, G., Breitmeyer, S., Cohen, M. J., Crump, B. C., Lu, Y., Payet, J. P., Poulin, B. A., Stubbins, A., et al.:
670   River network travel time is correlated with dissolved organic matter composition in rivers of the contiguous United States, Hydrological Processes, 35, e14 124, 2021.

Jackson, M. C., Weyl, O., Altermatt, F., Durance, I., Friberg, N., Dumbrell, A., Piggott, J., Tiegs, S., Tockner, K., Krug, C., et al.: Recommendations for the next generation of global freshwater biological monitoring tools, Advances in ecological research, 55, 615–636, 2016.

675 Jasiewicz, J. and Metz, M.: A new GRASS GIS toolkit for Hortonian analysis of drainage networks, Computers & Geosciences, 37, 1162–1173, 2011.

Kempeneers, P.: PKTOOLS - Processing Kernel for geospatial data, Version 2.6.7.6, Open Source Geospatial Foundation, http://pktools. nongnu.org/html/index.html, 2018.

Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A. K., Hochreiter, S., and Nearing, G. S.: Toward Improved Predictions in Ungauged
680   Basins: Exploiting the Power of Machine Learning, Water Resources Research, 55, 11 344–11 354, 2019.

Kuemmerlen, M., Schmalz, B., Guse, B., Cai, Q., Fohrer, N., and Jähnig, S. C.: Integrating catchment properties in small scale species distribution models of stream macroinvertebrates, Ecological Modelling, 277, 77–86, 2014.

Lay, U. S., Pradhan, B., Yusoff, Z. B. M., Abdallah, A. F. B., Aryal, J., and Park, H.-J.: Data mining and statistical approaches in debris-flow susceptibility modelling using airborne LiDAR data, Sensors, 19, 3451, 2019.

685 Lehner, B. and Grill, G.: Global river hydrography and network routing: baseline data and new approaches to study the world's large river systems, Hydrological Processes, 27, 2171–2186, 2013.

Lehner, B., Verdin, K., and Jarvis, A.: New global hydrography derived from spaceborne elevation data, Eos, Transactions American Geophysical Union, 89, 93–94, 2008.

Leopold, L. B., Wolman, M. G., Miller, J. P., and Wohl, E.: Fluvial processes in geomorphology, Courier Dover Publications, 1964.

690 Liang, C. and MaCkay, D. S.: A general model of watershed extraction and representation using globally optimal flow paths and up-slope contributing areas, International Journal of Geographical Information Science, 14, 337–358, 2000.

Lin, P., Pan, M., Beck, H. E., Yang, Y., Yamazaki, D., Frasson, R., David, C. H., Durand, M., Pavelsky, T. M., Allen, G. H., et al.: Global reconstruction of naturalized river flows at 2.94 million reaches, Water resources research, 55, 6499–6516, 2019.

Lin, P., Pan, M., Wood, E. F., Yamazaki, D., and Allen, G. H.: A new vector-based global river network dataset accounting for variable
695   drainage density, Scientific Data, 8, 1–9, 2021.

Linke, S., Lehner, B., Dallaire, C. O., Ariwi, J., Grill, G., Anand, M., Beames, P., Burchard-Levine, V., Maxwell, S., Moidu, H., et al.: Global hydro-environmental sub-basin and river reach characteristics at high spatial resolution, Scientific Data, 6, 1–15, 2019.

Liu, S., She, D., Gao, C., Amatulli, G., Wang, L., Lu, X., Raymond, P. A., and Xia, X.: Groundwater as a limited carbon dioxide source in a large river (the Yangtze River), Science of The Total Environment, 760, 143 336, 2021.

700 Lowe, W. H. and Likens, G. E.: Moving Headwater Streams to the Head of the Class, BioScience, 55, 196–197, 2005.

Maasri, A., Jähnig, S., Adamescu, M., Adrian, R., Baigun, C., Baird, D., Batista-Morales, A., Bonada, N., Brown, L., Cai, Q., et al.: A Global Agenda for Advancing Freshwater Biodiversity Research, Authorea Preprints, 2021a.

Maasri, A., Thorp, J. H., Kotlinski, N., Kiesel, J., Erdenee, B., and Jähnig, S. C.: Variation in macroinvertebrate community structure of functional process zones along the river continuum: New elements for the interpretation of the river ecosystem synthesis, River Research
705 and Applications, 37, 665–674, 2021b.

Marani, A., Rigon, R., and Rinaldo, A.: A note on fractal channel networks, Water Resources Research, 27, 3041–3049, 1991.

Marzadri, A., Amatulli, G., Tonina, D., Bellin, A., Shen, L. Q., Allen, G. H., and Raymond, P. A.: Global riverine nitrous oxide emissions: The role of small streams and large rivers, Science of The Total Environment, 776, 145 148, 2021.

McInerney, D. and Kempeneers, P.: Open Source Geospatial Tools - Applications in Earth Observation, Springer Verlag, 2015.

710 Messager, M. L., Lehner, B., Grill, G., Nedeva, I., and Schmitt, O.: Estimating the volume and age of water stored in global lakes using a geo-statistical approach, Nature Communications, 7, 1–11, 2016.

Messager, M. L., Lehner, B., Cockburn, C., Lamouroux, N., Pella, H., Snelder, T., Tockner, K., Trautmann, T., Watt, C., and Datry, T.: Global prevalence of non-perennial rivers and streams, Nature, 594, 391–397, 2021.

Metz, M., Mitasova, H., and Harmon, R.: Efficient extraction of drainage networks from massive, radar-based elevation models with least
715 cost path search, Hydrology and Earth System Sciences, 15, 667–678, 2011.

Meyer, J. L., Strayer, D. L., Wallace, J. B., Eggert, S. L., Helfman, G. S., and Leonard, N. E.: The contribution of headwater streams to biodiversity in river networks 1, JAWRA Journal of the American Water Resources Association, 43, 86–103, 2007.

Moore, I. D. and Burch, G. J.: Physical basis of the length-slope factor in the universal soil loss equation, Soil Science Society of America Journal, 50, 1294–1298, 1986.

720 Moore, I. D., Grayson, R., and Ladson, A.: Digital terrain modelling: a review of hydrological, geomorphological, and biological applications, Hydrological processes, 5, 3–30, 1991.

Moore, R. B., McKay, L. D., Rea, A. H., Bondelid, T. R., Price, C. V., Dewald, T. G., Johnston, C. M., et al.: User's guide for the National Hydrography Dataset plus (NHDPlus) High Resolution., Open-File Report-US Geological Survey, 2019.

Moudrỳ, V., Lecours, V., Gdulová, K., Gábor, L., Moudrá, L., Kropáček, J., and Wild, J.: On the use of global DEMs in ecological modelling
725 and the accuracy of new bare-earth DEMs, Ecological Modelling, 383, 3–9, 2018.

Mozzaquattro, L. B., Dala-Corte, R. B., Becker, F. G., and Melo, A. S.: Effects of spatial distance, physical barriers, and habitat on a stream fish metacommunity, Hydrobiologia, 847, 3039–3054, 2020.

Neteler, M. and Mitasova, H.: Open source GIS: a GRASS GIS approach, vol. 689, Springer Science & Business Media, 2013.

Neteler, M., Bowman, H., Landa, M., and Metz, M.: GRASS GIS: A multi-purpose open source GIS, Environmental Modelling & Software,
730 31, 124–130, 2012.

O'Callaghan, J. F. and Mark, D. M.: The extraction of drainage networks from digital elevation data, Computer vision, graphics, and image processing, 28, 323–344, 1984.

O'Loughlin, F., Paiva, R., Durand, M., Alsdorf, D., and Bates, P.: A multi-sensor approach towards a global vegetation corrected SRTM DEM product, Remote Sensing of Environment, 182, 49–59, 2016.

735 OpenStreetMap contributors: Planet dump retrieved from https://planet.osm.org , https://www.openstreetmap.org, 2017.

Orlandini, S., Moretti, G., Franchini, M., Aldighieri, B., and Testa, B.: Path-based methods for the determination of nondispersive drainage directions in grid-based digital elevation models, Water Resources Research, 39, 2003.

Oudin, L., Andréassian, V., Perrin, C., Michel, C., and Le Moine, N.: Spatial proximity, physical similarity, regression and ungaged catchments: A comparison of regionalization approaches based on 913 French catchments, Water Resources Research, 44, 2008.

Ozdemir, A.: Using a binary logistic regression method and GIS for evaluating and mapping the groundwater spring potential in the Sultan Mountains (Aksehir, Turkey), Journal of Hydrology, 405, 123–136, 2011.

Pekel, J.-F., Cottam, A., Gorelick, N., and Belward, A. S.: High-resolution mapping of global surface water and its long-term changes, Nature, 540, 418, 2016.

Pless, E., Saarman, N. P., Powell, J. R., Caccone, A., and Amatulli, G.: A machine-learning approach to map landscape connectivity in Aedes aegypti with genetic and environmental data, Proceedings of the National Academy of Sciences, 118, 2021.

Pourghasemi, H., Pradhan, B., Gokceoglu, C., and Moezzi, K. D.: Landslide susceptibility mapping using a spatial multi criteria evaluation model at Haraz Watershed, Iran, in: Terrigenous mass movements, pp. 23–49, 2012.

Quinn, P., Beven, K., Chevallier, P., and Planchon, O.: The prediction of hillslope flow paths for distributed hydrological modelling using digital terrain models, Hydrological processes, 5, 59–79, 1991.

Quinn, P., Beven, K., and Lamb, R.: The in $(a/\tan/\beta)$ index: How to calculate it and how to use it within the topmodel framework, Hydrological processes, 9, 161–182, 1995.

Raduła, M. W., Szymura, T. H., and Szymura, M.: Topographic wetness index explains soil moisture better than bioindication with Ellenberg's indicator values, Ecological Indicators, 85, 172–179, 2018.

Raymond, P. A., Zappa, C. J., Butman, D., Bott, T. L., Potter, J., Mulholland, P., Laursen, A. E., McDowell, W. H., and Newbold, D.: Scaling the gas transfer velocity and hydraulic geometry in streams and small rivers, Limnology and Oceanography: Fluids and Environments, 2, 41–53, 2012.

Raymond, P. A., Saiers, J. E., and Sobczak, W. V.: Hydrological and biogeochemical controls on watershed dissolved organic matter transport: Pulse-shunt concept, Ecology, 97, 5–16, 2016.

Read, D. S., Gweon, H. S., Bowes, M. J., Newbold, L. K., Field, D., Bailey, M. J., and Griffiths, R. I.: Catchment-scale biogeography of riverine bacterioplankton, The ISME Journal, 9, 516 – 526, 2015.

Reichl, J. P. C., Western, A. W., McIntyre, N. R., and Chiew, F. H. S.: Optimization of a similarity measure for estimating ungauged streamflow, Water Resources Research, 45, 2009.

Rodríguez Ernesto A4 - Morris, Charles S. A4 - Belz, J. Eric, E. A. R.: A Global Assessment of the SRTM Performance, Photogrammetric engineering and remote sensing, v. 72, pp. 249–260–2006 v.72 no.3, 2006.

Román-Sánchez, A., Vanwalleghem, T., Peña, A., Laguna, A., and Giráldez, J.: Controls on soil carbon storage from topography and vegetation in a rocky, semi-arid landscapes, Geoderma, 311, 159–166, 2018.

Saarman, N., Burak, M., Opiro, R., Hyseni, C., Echodu, R., Dion, K., Opiyo, E. A., Dunn, A. W., Amatulli, G., Aksoy, S., et al.: A spatial genetics approach to inform vector control of tsetse flies (Glossina fuscipes fuscipes) in Northern Uganda, Ecology and evolution, 8, 5336–5354, 2018.

Saarman, N., Pless, E., Amatulli, G., and Caccone, A.: Integrating genetic and environmental data to model and forecast movement and habitat use in the major insect vector of sleeping sickness in Uganda (Glossina fuscipes fuscipes), in: Entomology 2019, ESA, 2019.

Scheidegger, A. E.: The algebra of stream-order numbers, United States Geological Survey Professional Paper, 525, 187–189, 1965.

Seibert, J. and McGlynn, B. L.: A new triangular multiple flow direction algorithm for computing upslope areas from gridded digital elevation models, Water resources research, 43, 2007.

775 Shafizadeh-Moghadam, H., Valavi, R., Shahabi, H., Chapi, K., and Shirzadi, A.: Novel forecasting approaches using combination of machine learning and statistical models for flood susceptibility mapping, Journal of environmental management, 217, 1–11, 2018.

Shanafield, M., Bourke, S. A., Zimmer, M. A., and Costigan, K. H.: An overview of the hydrology of non-perennial rivers and streams, Wiley Interdisciplinary Reviews: Water, 8, e1504, 2021.

Shen, L. Q., Amatulli, G., Sethi, T., Raymond, P., and Domisch, S.: Estimating nitrogen and phosphorus concentrations in streams and rivers,
780 within a machine learning framework, Scientific Data, 7, 1–11, 2020.

Shreve, R. L.: Infinite topologically random channel networks, The Journal of Geology, 75, 178–186, 1967.

Shumilova, O., Zak, D., Datry, T., von Schiller, D., Corti, R., Foulquier, A., Obrador, B., Tockner, K., Allan, D. C., Altermatt, F., Arce, M. I., Arnon, S., Banas, D., Banegas-Medina, A., Beller, E., Blanchette, M. L., Blanco-Libreros, J. F., Blessing, J., Boëchat, I. G., Boersma, K., Bogan, M. T., Bonada, N., Bond, N. R., Brintrup, K., Bruder, A., Burrows, R., Cancellario, T., Carlson, S. M., Cauvy-Fraunié, S., Cid,
785 N., Danger, M., de Freitas Terra, B., Girolamo, A. M. D., del Campo, R., Dyer, F., Elosegi, A., Faye, E., Febria, C., Figueroa, R., Four, B., Gessner, M. O., Gnohossou, P., Cerezo, R. G., Gomez-Gener, L., Graça, M. A., Guareschi, S., Gücker, B., Hwan, J. L., Kubheka, S., Langhans, S. D., Leigh, C., Little, C. J., Lorenz, S., Marshall, J., McIntosh, A., Mendoza-Lera, C., Meyer, E. I., Miliša, M., Mlambo, M. C., Moleón, M., Negus, P., Niyogi, D., Papatheodoulou, A., Pardo, I., Paril, P., Pešić, V., Rodriguez-Lozano, P., Rolls, R. J., Sanchez-Montoya, M. M., Savić, A., Steward, A., Stubbington, R., Taleb, A., Vorste, R. V., Waltham, N., Zoppini, A., and Zarfl, C.: Simulating
790 rewetting events in intermittent rivers and ephemeral streams: A global analysis of leached nutrients and organic matter, Global Change Biology, 25, 1591–1611, 2019.

Simard, M., Pinto, N., Fisher, J. B., and Baccini, A.: Mapping forest canopy height globally with spaceborne lidar, Journal of Geophysical Research: Biogeosciences, 116, 2011.

Strahler, A. N.: Quantitative analysis of watershed geomorphology, Eos, Transactions American Geophysical Union, 38, 913–920, 1957.

795 Strobl, P. A., Bielski, C., Guth, P. L., Grohmann, C. H., Muller, J.-P., López-Vázquez, C., Gesch, D. B., Amatulli, G., Riazanoff, S., and Carabajal, C.: The Digital Elevation Model Intercomparison eXperiment DEMIX, a community-based approach at global DEM benchmarking, The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 43, 395–400, 2021.

Sugarbaker, L., Constance, E. W., Heidemann, H. K., Jason, A. L., Lucas, V., Saghy, D., and Stoker, J. M.: The 3D Elevation Program initiative: a call for action, US Geological Survey, 2014.

800 Tadono, T., Takaku, J., Tsutsui, K., Oda, F., and Nagai, H.: Status of "ALOS World 3D (AW3D)" global DSM generation, in: 2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 3822–3825, IEEE, 2015.

Takaku, J., Iwasaki, A., and Tadono, T.: Adaptive filter for improving quality of ALOS PRISM DSM, in: 2016 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), pp. 5370–5373, IEEE, 2016.

Tarboton, D. G.: A new method for the determination of flow directions and upslope areas in grid digital elevation models, Water resources
805 research, 33, 309–319, 1997.

Thalacker, R. J.: Mapping techniques for soil erosion: Modeling stream power index in eastern North Dakota, The University of North Dakota, 2014.

Thoms, M., Scown, M., and Flotemersch, J.: Characterization of river networks: A GIS approach and its applications, JAWRA Journal of the American Water Resources Association, 54, 899–913, 2018.

USGS: Global 30 Arc-Second Elevation (GTOPO30), United States Geological Survey, https://doi.org/https://doi.org/10.5066/F7DF6PQS, 1996.

USGS: Shuttle Radar Topography Mission (SRTM) 1 Arc-Second Global, United States Geological Survey, 2015.

USGS EROS Archive: HYDRO1k Elevation Derivative Database, URL: https://doi.org/10.5066/F77P8WN0, 2018.

Vannote, R. L., Minshall, G. W., Cummins, K. W., Sedell, J. R., and Gushing, C. E.: The River Continuum Concept, Canadian Journal of Fisheries and Aquatic Sciences, 37, 130–137, 1980.

Virro, H., Amatulli, G., Kmoch, A., Shen, L., and Uuemaa, E.: GRQA: Global River Water Quality Archive, Earth System Science Data Discussions, pp. 1–30, 2021.

Vogt, J. V., Colombo, R., and Bertolo, F.: Deriving drainage networks and catchment boundaries: a new methodology combining digital elevation data and environmental characteristics, Geomorphology, 53, 281–298, 2003.

Walling, D. E.: The sediment delivery problem, Journal of Hydrology, 65, 209–237, 1983.

Wollheim, W. M., Bernal, S., Burns, D. A., Czuba, J., Driscoll, C., Hansen, A., Hensley, R., Hosen, J., Inamdar, S., Kaushal, S., et al.: River network saturation concept: factors influencing the balance of biogeochemical supply and demand of river networks, Biogeochemistry, 141, 503–521, 2018.

Yamazaki, D., Trigg, M. A., and Ikeshima, D.: Development of a global~ 90 m water body map using multi-temporal Landsat images, Remote Sensing of Environment, 171, 337–351, 2015.

Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J. C., Sampson, C. C., Kanae, S., and Bates, P. D.: A high-accuracy map of global terrain elevations, Geophysical Research Letters, 44, 5844–5853, 2017.

Yamazaki, D., Ikeshima, D., Sosa, J., Bates, P. D., Allen, G., and Pavelsky, T.: MERIT Hydro: A high-resolution global hydrography map based on latest topography datasets, Water Resources Research, 2019.

Yang, C. T.: The movement of sediment in rivers, Geophysical Surveys, 3, 39–68, 1977.

Yang, W., Hou, K., Yu, F., Liu, Z., and Sun, T.: A novel algorithm with heuristic information for extracting drainage networks from raster DEMs, Hydrology and Earth System Sciences Discussions, 7, 441–459, 2010.

Zhang, L., Wang, G., Dai, B., and Li, T.: Classification and codification methods of stream network in a river basin, a review, Environmental Informatics Archives, 5, 364–372, 2007a.

Zhang, Y., Liu, Y., and Chen, Z.: Multi-flow direction algorithms for extracting drainage network based on digital elevation model, in: Geoinformatics 2007: Geospatial Information Science, vol. 6753, p. 67532B, 2007b.

Ågren, A. M., Lidberg, W., and Ring, E.: Mapping Temporal Dynamics in a Forest Stream Network—Implications for Riparian Forest Management, Forests, 6, 2982–3001, https://doi.org/10.3390/f6092982, 2015.