MERIT Hydro: A high-resolution global hydrography map based on latest topography datasets

Dai Yamazaki[1], Daiki Ikeshima[2], Jeison Sosa[3], Paul D. Bates[3], George Allen[4], Tamlin Pavelsky[5]

[1]Institute of Industrial Science, The University of Tokyo,
4-6-1 Komaba, Meguro-ku, Tokyo, 153-8505, Japan

[2]Department of Civil and Environmental Engineering, Tokyo Institute of Technology
2-12-1-M1-6, Ookayama, Meguro-ku, Tokyo 152-8552, Japan

[3]School of Geographical Sciences, University of Bristol,
University Road, Clifton, Bristol BS8 1SS, United Kingdom

[4]Deparment of Geography, Texas A&M University,
3147 TAMU, College Station, Texas 77843

[5]Department of Geological Sciences, University of North Carolina,
104 South Rd CB#3315, Chapel Hill, NC 27599

Corresponding author: Dai Yamazaki (yamadai@rainbow.iis.u-tokyo.ac.jp)

Key Points:

- A global hydrography map was generated using latest topography datasets
- Near-automatic algorithm applicable for global hydrography delineation
- Adjusted elevation and river width layers consistent with flow direction map

Abstract

High-resolution raster hydrography maps are a fundamental data source for many geoscience applications. Here we introduce MERIT Hydro, a new global flow direction map at 3 arc-second resolution (~90 m at the equator) derived from the latest elevation data (MERIT DEM) and water body datasets (G1WBM, GSWO, and OpenStreetMap). We developed a new algorithm to extract river networks near-automatically by separating actual inland basins from dummy depressions caused by the errors in input elevation data. After a minimum amount of hand-editing, the constructed hydrography map shows good agreement with existing quality-controlled river network datasets in terms of flow accumulation area and river basin shape. The location of river streamlines was realistically aligned with existing satellite-based global river channel data. Relative error in the drainage area was <0.05 for 90% of GRDC gauges, confirming the accuracy of the delineated global river networks. Discrepancies in flow accumulation area were found mostly in arid river basins containing depressions that are occasionally connected at high water levels and thus resulting in uncertain watershed boundaries. MERIT Hydro improves on existing global hydrography datasets in terms of spatial coverage (between N90 and S60) and representation of small streams, mainly due to increased availability of high-quality baseline geospatial datasets. The new flow direction and flow accumulation maps, along with accompanying supplementary layers on hydrologically adjusted elevation and channel width, will advance geoscience studies related to river hydrology at both global and local scales.

Plain Language Summary

Rivers play important roles in global hydrological and biogeochemical cycles, and many socio-economic activities also depend on water resources in river basins. Global-scale frontier studies of river networks and surface waters require that all rivers on the Earth are precisely mapped at high resolution, but until now, no such map has been produced. Here we present "MERIT Hydro", the first high-resolution, global map of river networks developed by combining the latest global map of land surface elevation with the latest maps of water bodies that were built using satellites and open databases. Surface flow direction of each 3-arcsecond pixel (~90m size at the equator) is mapped across the entire globe except Antarctica, and many supplemental maps (such as flow accumulation area, river width, and a vectorized river network) are generated. MERIT Hydro thus represents a major advance in our ability to represent the global river network and is a dataset that is anticipated to enhance a wide range of geoscience applications including flood risk assessment, aquatic carbon emissions, and climate modelling.

1 Introduction

A hydrography map is important baseline data source for many geoscience studies, such as land hydrology and flood inundation modeling (Miguez-Macho et al. 2007; Yamazaki et al. 2014a), analysis of ecosystem and biodiversity (Turner et al., 2012), global carbon budget estimation (Raymond et al., 2013), and terrain type classification (Hengle and Evans, 2009; Nobre et al., 2011). Typically, a hydrography map is provided as a high-resolution raster grid of surface flow directions (Lehner et al. 2008), with river networks represented by pixels with large flow accumulation areas. By analyzing surface flow directions, many hydrological parameters can be delineated, such as catchment boundaries, flow distance, height above nearest drainage (Noble et al., 2008), and river channel width (Yamazaki et al., 2014b). Thus, the accuracy of the hydrography map is critically important for many applications to reduce uncertainties.

The flow direction of each high-resolution pixel can be precisely determined if very accurate topography data are available, however, construction of a high quality hydrography map is still difficult for much of the globe because of the errors and limitations in available topography datasets. Digital Elevation Models (DEMs) are the primary topography data for the development of a hydrography map, but they usually contain non-negligible vertical errors which distort the terrain slope that is used to estimate flow directions (Yamazaki et al. 2017). Small streams whose width is smaller than the pixel size of the DEM are not represented in many cases (Turcotte et al., 2001). Even wide rivers and large lakes may not be well captured because DEMs usually represent mean water surface elevations rather than the bed elevation of these features.

At regional scales where the focus is on one or a few river basins, several methods for extracting a high-accuracy hydrography map from DEMs have been proposed (e.g. Tarboton, 1997). In many cases, supplementary information on river streamlines is used to modify the DEM to generate realistic river networks. However, at a continental or global scale, automatic or near-automatic river network delineation has not yet been realized in a practical manner because it is difficult to separate actual inland endorheic basins from dummy depressions caused by DEM errors. Up to now, HydroSHEDS, which was developed based on the SRTM3 DEM (Lehner et al., 2008), has been the only available global-scale high-resolution (3 arc-second, about 90m at the equator) hydrography map, but the development of the HydroSHEDS data set required a substantial amount of manual editing to ensure the reality of the represented river networks (Lehner et al., 2006). Because of this manual editing, reproducing this process so it can be repeated with more recent high-quality terrain data sets has not been feasible.

In recent years, a number of highly accurate topography datasets that are potentially helpful in producing more accurate hydrography maps have been released. For example, high-resolution DEMs such as TanDEM-X (Krieger et al., 2007) and AW3D-30m (Tadono et al., 2016) have become available. To enhance the applicability of spaceborne DEMs, the MERIT DEM (Yamazaki et al., 2017) was developed by applying a global-scale error-removal algorithm to existing spaceborne DEMs. The availability of global water layer data, another input required by hydrography delineation, has also increased rapidly. Global-scale analysis of Landsat images is now possible (e.g. Gorelick et al., 2017), and water body maps considering the frequency of water existence have been produced (GSWO by Pekel et al, 2016; G3WBM by Yamazaki et al. 2015). Furthermore, the availability of local geospatial information is increasing rapidly (such as OpenStreetMap, Haklay et al., 2008), following recent trends towards "Open Data" policies by local and national governments and the crowd-sourcing of vector maps.

In the decade that has passed since the development of HydroSHEDS, the appearance of new global topography datasets and enhanced computing capacity means there is now an urgent need to produce methods to near-automatically delineate global hydrography maps. Furthermore, as the accuracy and spatial-coverage of the baseline high-resolution topography datasets have increased in recent years, more precise representation of river networks should also be possible by overcoming the limitations of HydroSHEDS. For example, the locations of small rivers were not well represented in HydroSHEDS, especially in forested areas (Figure S1), because the elevations in satellite DEMs were biased due to tree canopy artefacts. Representation of the flow directions over large water bodies in HydroSHEDS was not also adequate due to the limitations in the water body data and GIS algorithms that were available at the time (Figure S2), which degrades the consistency between HydroSHEDS and other river-relevant datasets such as GRWL river width data (Allen and Pavelsky, 2018). Perhaps most importantly, the coverage of HydroSHEDS was limited to below N60 because of the availability of the SRTM3 DEM, and thus only coarser-resolution hydrography maps such as Hydro1K (US Geological Survey, 2000) could be used for studies at high latitudes (Figure S3f). These limitations can be addressed if a new hydrography map is developed based on recent topography datasets.

Here, we developed a new algorithm for delineating hydrography data at a global scale, with only a minimum amount of hand-editing. We applied the new algorithm to the MERIT DEM enhanced by supplementary water body layers, and generated the new global hydrography map at 3 arc-second resolution. In this paper, we describe the methodology used to create the new hydrography map, named MERIT Hydro, and undertake a number of validation tests of this new dataset.

2 Method

The schematic diagram for generating the new hydrography map and its supplementary data layers is shown in Figure 1. Detailed descriptions of input data, algorithm, and additional data layers are found below.

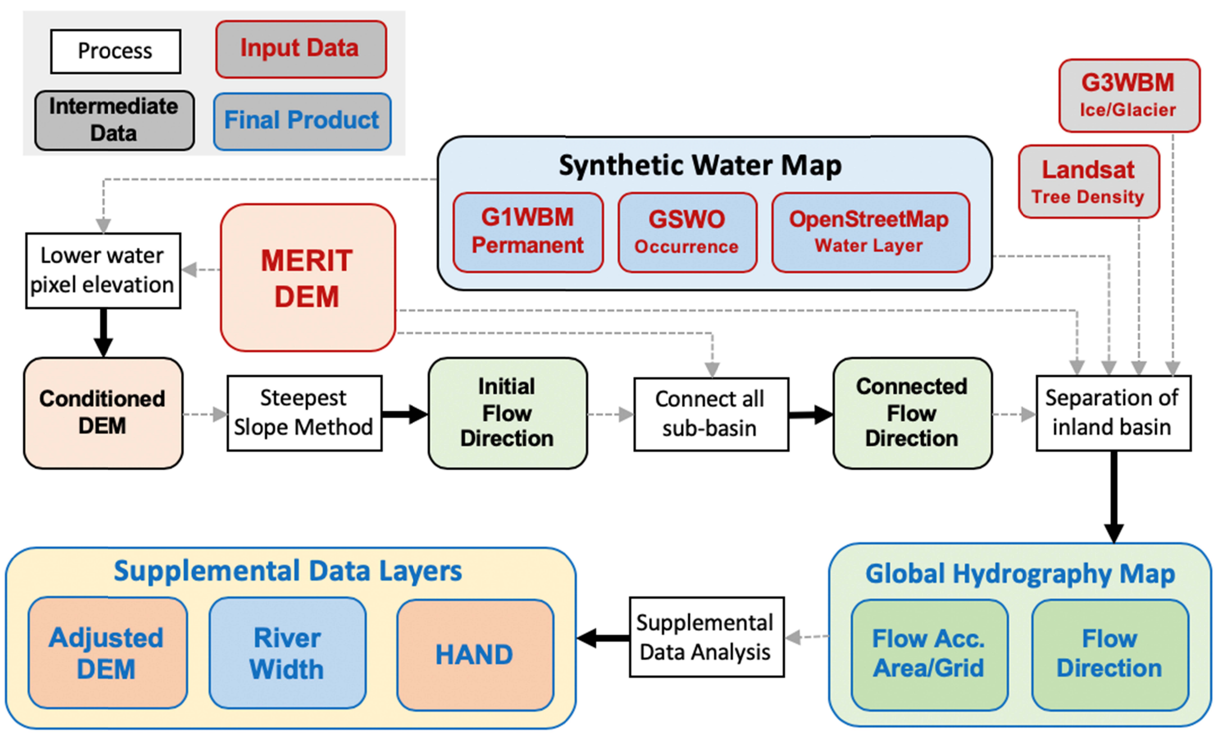


Figure 1. Procedures of hydrography delineation.

2.1 Input data sources

We used the MERIT DEM (Multi-Error-Removed Improved-Terrain Digital Elevation Model; Yamazaki et al. 2017, available at: http://hydro.iis.u-tokyo.ac.jp/~yamadai/MERIT_DEM) as the baseline elevation data for the hydrography delineation. The MERIT DEM was developed by removing multiple error components from the SRTM3 (Farr et al, 2007) and AW3D-30m DEMs (Tadono et al., 2016). As the original DEMs were affected by non-negligible height errors and tree canopy biases that distort river network structures, the use of the error-removed DEM was essential for the hydrography analysis. The spatial resolution of the MERIT DEM is 3 arc-second (~90m at the equator), and it covers the entire globe except for Antarctica (between 90N and 60S).

A water layer dataset is also needed to improve hydrography delineation as a complement to the elevation data. As the accuracy of the MERIT DEM is limited by the remaining height errors and its spatial resolution, the water layer data is used to mitigate the impact of remaining errors and to represent streams smaller than the DEM pixel size. We used multiple water layer datasets to reflect the different characteristic of each product. The synthetic water layer map was generated by combining the G1WBM (Global 1-second Water Body Map; Yamazaki et al. 2015), GSWO (Global Surface Water Occurrence; Pekel et al. 2016) and water-related layers from OpenStreetMap (Figure 2a). The synthetic water layer map represents the "likelihood" of water existence at each pixel using a value ranging between 0 and 100 (Figure 2b), and the elevation data was modified following this likelihood value (see section 2.2). The synthetic water layer map was generated at 1-arcsecond resolution, and then upscaled to 3-arcsecond resolution by taking the maximum value within 3x3 pixels. The procedure to generate the synthetic water layer map is described below.

As the baseline data for the synthetic water layer map, we used the G1WBM permanent water layer at 1 arc-second spatial resolution (Figure 2a, red color). G1WBM is the new global water body map we generated for MERIT Hydro development. It is the new version of the G3WBM water layer (Yamazaki et al. 2015) with the increased spatial resolution (enhanced from 3 arc-second to 1 arc-second), though the same input data and algorithm were used (see Yamazaki et al. 2015 for details). G1WBM is available at: http://hydro.iis.u-tokyo.ac.jp/~yamadai/G3WBM/. For the hydrography map development, it is better to use a water layer dataset that corresponds to the baseline MERIT DEM, because any temporal changes in water layers, for example those caused by channel migration, could have a negative impact on the hydrography delineation. The G1WBM dataset was created by merging water layers from 4-epochs in the Landsat GLS collection (Gutman et al., 2013), and the continuity of river channels was ensured by integrating the SWBD (SRTM Water Body Data, acquired simultaneously with SRTM DEM). Thus, G1WBM is considered to have better consistency to the MERIT DEM compared to other global water layer datasets (e.g. GSWO), because SRTM-related products were used both in MERIT DEM and G1WBM. In the synthetic water layer map, the likelihood value of 100 is given to the G1WBM permanent water layer pixels.

We also integrate GSWO (Pekel et al. 2016) into the synthetic water layer map (Figure 2a, cyan and blue colors), in addition to G1WBM. GSWO represents the water occurrence frequency based on the entire global Landsat archive (~3 million images) at 1-arcsecond resolution. It has the potential to correct the remaining error in the MERIT DEM, because pixels with higher water occurrence value are expected to be lower elevation than adjacent pixels with lower water occurrence frequency. In the synthetic water layer map, the GSWO occurrence value (originally between 0 and 100) was rescaled to the range 0-70, and overlaid onto the G1WBM permanent water layer. The rescaling was adopted in order to enhance the contrast between the permanent water bodies (such as river channels) and seasonally inundated water bodies (such as floodplains).

To represent small streams that are not visible in 1 arc-second (~30m at the equator) resolution Landsat data, we also used water layers from OpenStreetMap. First, we extracted all water related components from the OpenStreetMap datasets (i.e. "planet.osm" file, downloaded from https://planet.openstreetmap.org/ on 16 January 2018). The water-related features were extracted by using the OSM tags: ***"natural=water"***, ***"waterway=*"***, ***"landuse=reservoir"***. Then, the extracted water-related features were classified to three different types: [1] "large rivers and lakes" represented as closed vector polygon data; [2] "middle-sized river channels" represented as line data with the OSM tag ***"waterway=riverbank, river"***; and [3] "small streams" with the tag ***"waterway=canal, drain, ditch, stream, brook, wadi, drystream"***. The "large rivers and lakes", "middle-sized river channels", and "small stream" classes extracted from OpenStreetMap are represented by orange, green, and gray colors in Figure 2a. Then, extracted water-related vector data were converted to raster format at 1-arcsecond resolution, and these OpenStreetMap water layers were integrated with the synthetic water layer map. When integrating, the water occurrence likelihood value for "large rivers and lakes", "middle-size rivers", "small streams" were set to 25, 20 and 5 respectively, with these values selected by trial and error. Relatively small likelihood values were used for OpenStreetMap water layers especially for "small streams" as its mapping accuracy is considered to be lower than Landsat-based water maps. The extracted OpenStreetMap water-related layer data are made available online at: http://hydro.iis.u-tokyo.ac.jp/~yamadai/OSM_Water/.

In addition to the above elevation and water layer datasets, the Landsat tree density map (Hansen et al. 2013) was used as a quality flag for the MERIT DEM elevation. Even though the tree canopy bias was removed in the MERIT DEM, the elevation value in forested pixels has higher uncertainty compared to non-forested pixels. When elevation data has a problem with hydrological consistency (e.g. catchment upstream elevations are lower than downstream elevations), the elevations in areas covered by higher tree density are likely to be the cause of the problem. A more detailed description of how tree density data are used to reduce these errors is given in the following algorithm section.

2.2 Hydrography generation algorithm

The schematic diagram for the hydrography delineation procedure is shown in Figure 1. First, a "conditioned DEM" was generated by lowering the elevation of water pixels in the MERIT DEM. Similar to other spaceborne DEMs, the MERIT DEM represents the elevation of the water surface for water bodies, not the bathymetric elevation of the river or lake bed. Furthermore, streams smaller than the pixel size cannot be well represented in MERIT. In order to better represent river networks, the elevation of DEM pixels overlain by water should be lowered before the flow direction is calculated (a similar approach was taken when HydroSHEDS was generated, i.e. the elevation of the SRTM3 DEM was lowered using the SWBD water mask, see Lehner et al. 2006). We lowered the original MERIT DEM elevation based on the water likelihood value of the synthetic water layer data. The conditioned elevation Z_{con} for a water body pixel ($L_{wat} > 0$) is given by the equation (1):

$$Z_{con} = Z_{ori} - (3.0 + 0.17L_{wat}) \quad (1)$$

where the original elevation Z_{ori} is lowered by the range between 3m and 20m following the water likelihood value L_{wat} (range between 0 and 100). The minimum range (3m) was used in order to make sure that a water pixel drains surface water flow from its surrounding land pixels, and the lowering amount increased up to 20m along with water occurrence likelihood assuming the bathymetry is deeper for higher water occurrence pixels. After this lowering process, the elevation is further conditioned to satisfy the rule that a pixel with a higher water likelihood should be lower than any adjacent pixel with a lower water likelihood value. The original and conditioned elevation and the water likelihood in the Tone River basin in Japan is shown as an example in Figure 2. Note that the "conditioned DEM" is used only for the calculation of the initial flow direction in the next step. It is different from the "hydrologically adjusted DEM" created at a later stage in the processing chain as a supplementary data layer of the hydrography dataset.

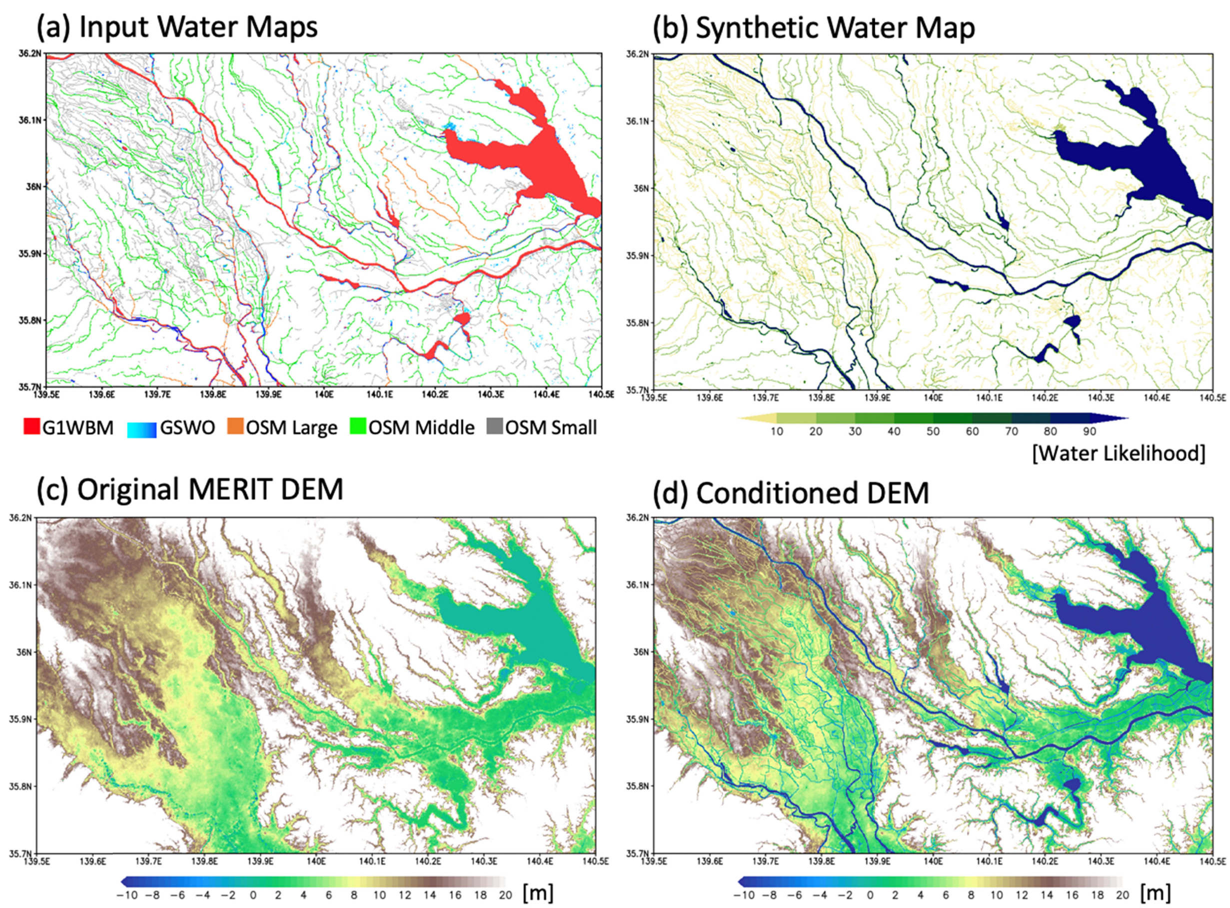


Figure 2. Elevation conditioning for initial flow direction calculation for the Tone River in Japan. (a) Input water maps. (b) Synthetic water layer data. (c) Original MERIT DEM elevation. (d) Conditioned DEM elevation.

Second, an initial flow direction is calculated from the conditioned DEM based on the topographic slope. Among the possible eight flow directions for each pixel, the direction that generates the steepest topographic gradient is selected as the initial flow direction. At the boundary of the land and ocean, a pixel is determined to be a "river mouth pixel" when any of its adjacent pixels is treated as ocean in the DEM. If the elevations of all adjacent pixels are higher than the target pixel, the target pixel is treated as "inland basin termination".

However, one river basin could theoretically be separated into multiple sub-basins in the initial flow direction data, due to artificial depressions caused by the height errors in the DEM (Figure 3a). In the third step of the hydrography delineation algorithm, the initial flow directions were modified to connect all sub-basins to a river mouth (Figure 3b). For this purpose, the lowest elevation pixel in each sub-basin (dark blue dot in Figure 3) and the lowest pixel along sub-basin boundaries (orange square in Figure 3) were detected. If the lowest elevation in one sub-basin is lower than that of its adjacent sub-basin, the flow directions between their lowest sub-basin boundary and the lowest pixel were reversed (red arrows in Figure 3b), and these sub-basins were merged. This procedure (finding lowest pixels and reversing flow directions) was repeated until all sub-basins were connected to a river mouth, and thus "connected flow direction" data was generated.

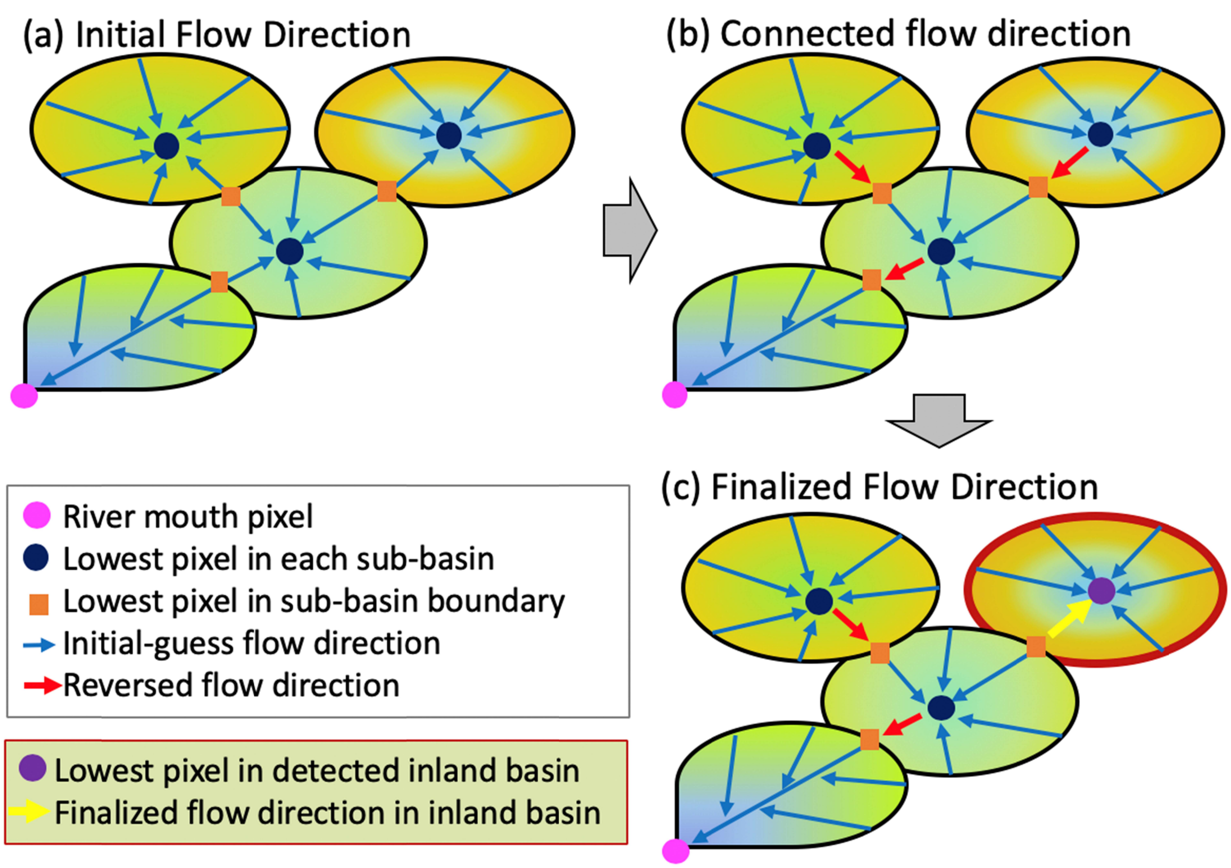


Figure 3. Procedures of flow direction calculation. (a) Initial flow direction. (b) Connected flow direction. (c) Finalized flow direction by separating inland basins.

In the last step, actual inland basins were detected and separated, because artificial depressions caused by DEM errors and actual inland basins were treated together in the previous step. Inland basins were identified by calculating how much volume of topography needed to be modified to reverse the flow directions in the previous step (i.e. reducing downstream elevations and/or lifting upstream elevations). We assumed that the dummy depression could be connected to its adjacent sub-basin by slightly modifying the elevations around the sub-basin boundary, while actual inland basins should remain independent unless the topography was modified significantly. We calculated the minimum amount of topography modification by combining the downstream reduction and upstream lifting, following the method developed by Yamazaki et al. (2012).

The schematic illustration of the method for inland basin detection is shown in Figure 4. First, depressions are defined as an area where the downstream elevation is higher than the upstream elevation (Figure 4ab). Then, the highest elevation on the depression downstream ridge (Z_{max}) and the lowest elevation in the depression area (Z_{min}) were detected. The area consisting of the "upstream depression" lower than Z_{max} and the "downstream uplift" higher than Z_{min} is considered for inland basin detection. By assuming flat topography after modification for reducing computational complexity, the modified elevation (Z_{mod}) after the depression removal can take a value ranging between Z_{min} and Z_{max}. For simplification, in case of Figure 3, it is assumed that $Z_{max} = Z_{min} + 2$ and the elevation increment is 1. Thus, three values (Z_{min}, Z_{min}+1, and Z_{max}) should be considered as a potential modified elevation Z_{mod}. Then, the required volume of topography modification (V) is calculated for each possible modification. The modification pattern that requires the minimum modification volume is selected as the final modified elevation. In case of Figure 3, the required volume becomes minimum (V=2) when the modified elevation is Z_{min}+1, thus it is decided that $Z_{mod} = Z_{min} + 1$. Note that the original algorithm by Yamazaki et al. (2012) was developed for the SRTM3 DEM, which is in integer format (1m increment), but the MERIT DEM is provided as real-numbers (32 bit float). In order to reduce computational cost, we converted the original MERIT DEM elevations from continuous real numbers to discrete values with 10cm increments.

Then, the depression area is determined to be an actual inland basin if the required modification volume is larger than a threshold value. After several trial and error tests, we decided to adopt 2,500,000 m^3 as the threshold modification volume to separate actual inland basins from dummy depressions (equivalent to 2.5 m constant depth depression with 1 km^2 area). Here, we considered some uncertainties in the DEM to avoid confusion between the actual inland basin and dummy depressions due to height errors. First, DEM elevations over a water body are not reliable as they are usually estimated by interpolating surrounding terra firma elevations, thus 1 m was removed from the modified height value when calculating the modified volume over a water body. Second, the DEM height uncertainty is larger in forested areas (Yamazaki et al. 2017), thus the calculated modified volume of a forest pixel was reduced by 50%. We assumed a pixel is treated as forest when its Landsat tree density (Hansen et al. 2013) is >50%. Third, the elevation over glaciers has large errors, and thus we did not modify depressions over glacier pixels. Glaciated pixels were identified by the "ice" tag in the G3WBM data (Yamazaki et al., 2015).

Finally, the flow directions in the detected inland basins were returned to the original direction (Figure 3c, yellow arrows), thereby separating inland basins from dummy depressions. By applying the above algorithm, the automatic calculation of the hydrography data is realized, including the separation of inland basins, which was previously difficult at the global scale. However, due to the remaining errors in the DEM and water layer data (e.g. very narrow valleys that cannot be resolved at the DEM resolution, discontinuities in the water layer data etc.), it was impossible to perfectly delineate a river network map fully automatically. Therefore, we visually checked the calculated result, and manually modified the original elevation or synthetic water layer map in ~400 locations). This visual quality check was unavoidable to ensure the accuracy of the hydrography data, but this methodology required significantly less manual editing/human effort (about 1-2 expert-person days) compared to previous methodologies.

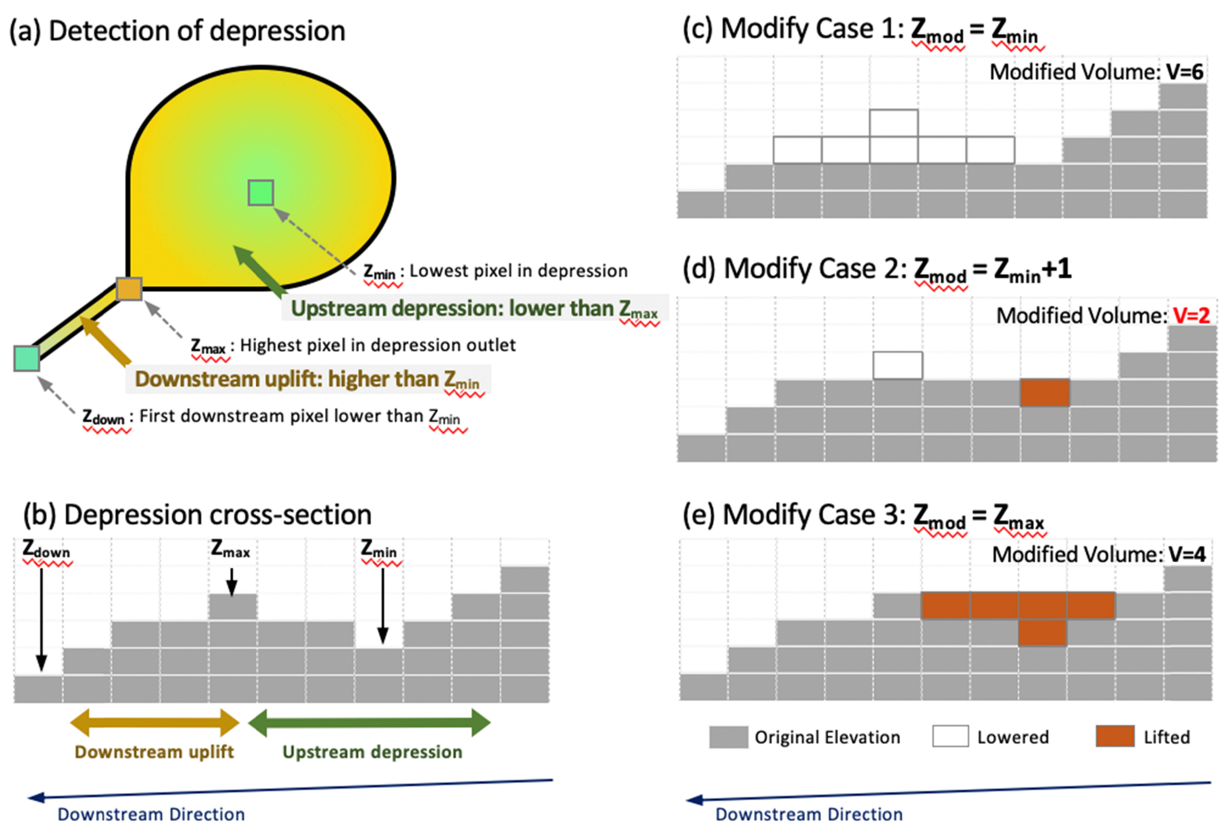


Figure 4. Schematic illustration of the method for inland basin detection. (a) A depression is detected as an area where downstream elevations are higher than upstream. The upstream depression and downstream uplift needed to connect the apparently enclosed basin are defined for each depression. (b) Cross-sectional illustration of the depression. Z_{max} represents the highest elevation of the depression ridge, while Z_{min} represents the lowest elevation in the depression. Z_{down} is the first downstream pixel lower than Z_{min}. (c-e) The elevation after the depression removal (Z_{mod}) can take a value ranging between Z_{min} and Z_{max}. For example, in Figure 3, Z_{max} is assumed to be Z_{min}+2, and increment of height modification is set to 1. The amount of required modification volume (V) can be calculated for different Z_{mod} values.

2.3 Supplementary Data Layer

In addition to the basic hydrography parameters (e.g. flow direction, flow accumulation in pixels, flow accumulation area), other potentially useful supplementary data layers can also be produced from the hydrography processing chain (Figure 1, bottom row). In particular, the HAND parameter (Height Above Nearest Drainage; Nobre et al. 2011), hydrologically adjusted elevation, and river width were determined.

The hydrologically adjusted elevation represents the DEM in which elevations were modified to satisfy the condition that "downstream is not higher than upstream". In order to minimize the amount of modification from the original DEM, for this we used the same algorithm employed for inland sink detection (Figure 4; Yamazaki et al. 2012).

The HAND parameter represents the relative height of each pixel above the elevation of its nearest downstream drainage pixel. This topography index is useful for many types of hydrologically-relevant terrain analysis (Nobre et al. 2011). We calculated the HAND value using a 0.5km^2 threshold to define drainage. The threshold for defining drainage could differ by region or by climate, but we used a globally-uniform value to prepare the HAND data. Users are recommended to re-calculate HAND using their own region-specific threshold if their application is sensitive to the thresholding value. Note that the hydrologically adjusted elevation was used to calculate the HAND parameter.

The river width is an important parameter for many applications such as flood inundation modelling. We calculated river width using the developed flow direction data and the G1WBM permanent water body layer by applying an existing algorithm for river width calculation (Yamazaki et al., 2014b). As the original river width algorithm was only applicable to the binary water mask (land/water classification), we modified the code to handle sub-pixel water fraction data (percentage of 1 arc-second permanent water pixels within a 3 arc-second flow direction pixel).

3 Results and Validation

3.1 Delineated River Network

The delineated river network in MERIT Hydro is illustrated in Figure 5. As an example, three regions were selected: (a) the Pearl River (Zhujiang) basin in Southern China, which contains both mountainous areas and alluvial floodplains; (b) the Ob River mouth region in Western Siberia in Russia as a representative of high latitude regions not covered in HydroSHEDS, and (c) the Danakil Desert in Ethiopia, which contains many inland basins. Pixels with flow accumulation area >5km^2 are represented as streams. The dark blue lines represent the streams that overlap with the Landsat water layer data, while black lines represent streams not coincident with Landsat water observations. The thickness of stream lines corresponds to the flow accumulation area. Similar figures for all other regions on the globe are accessible from the data product webpage.

The Pearl River basin (Figure 5a) contains an alluvial delta region (around N22.8, E113.2), and also high mountain areas in the basin headwaters. In high mountain regions, many dummy depressions exist because height errors are larger for high relief topography. The river network of the Pearl River was reasonably delineated by connecting these dummy depressions using the newly developed algorithm. The green dot in the southwest part (N21.7, E110.9) represents an open mine which is treated as an inland basin. The algorithm also succeeded in generating river networks in very flat topography in the river mouth delta region, even though it has been noted previously that calculation of acceptable flow directions is difficult in flat regions because of both real and artificial DEM depressions (Pan et al., 2012). However, the bifurcating channels in the delta region cannot be fully represented because only one downstream direction was assumed at each pixel. Because of this limitation, many distributary channels have relatively small flow accumulation areas, whereas in reality those channels could have large river discharge bifurcated from the main channel. Some countermeasures for handling bifurcating channels are needed to further enhance the developed hydrography datasets for certain applications (e.g. analysis of bifurcation in flood inundation models, Yamazaki et al., 2014a) and will be developed in subsequent work.

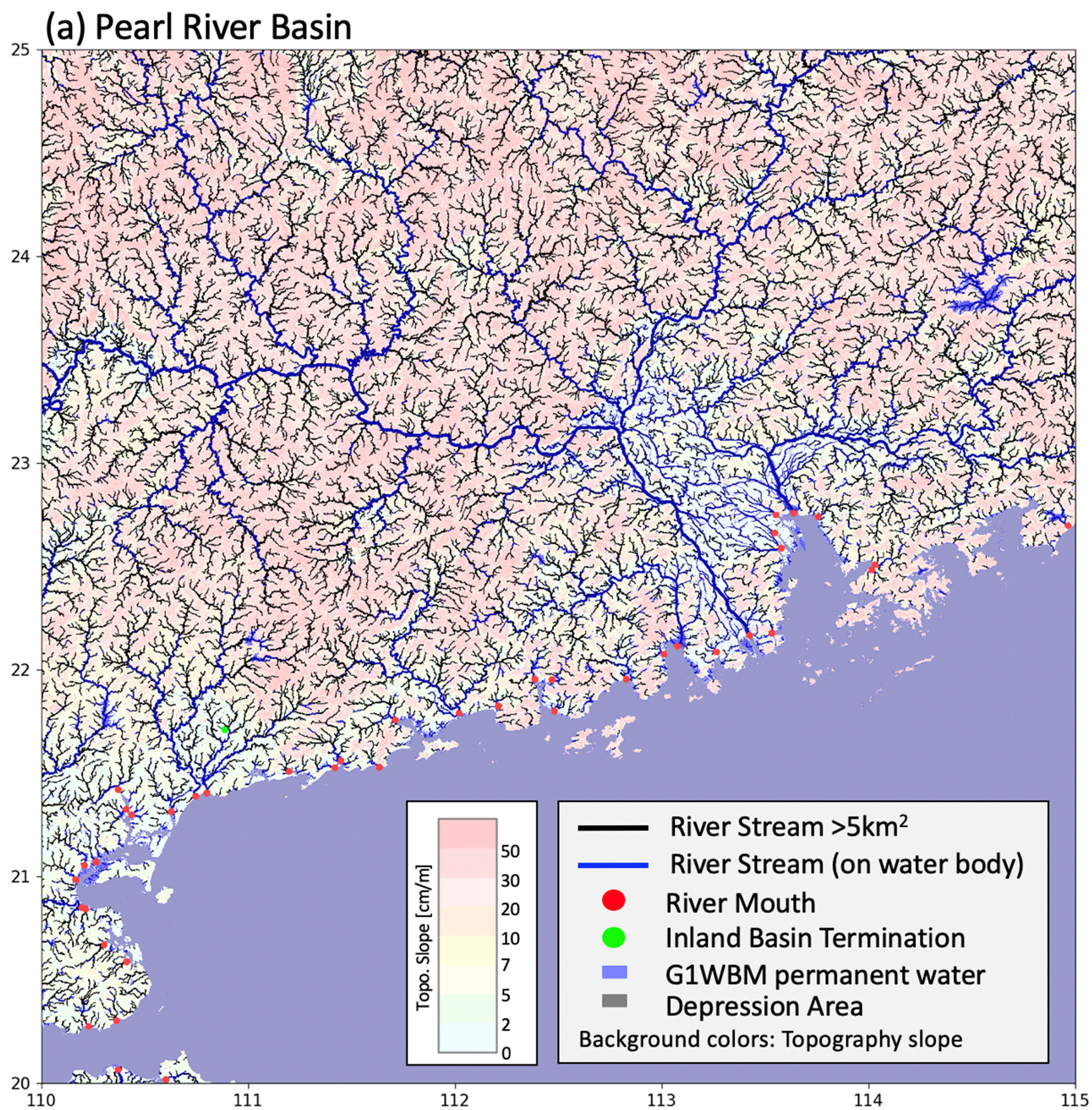


Figure 5. Delineated river networks in MERIT Hydro. The streams with >5km^2 flow accumulation area are shown. The dark blue line represents streams that overlap with Landsat water observations, while black lines represent streams without coincident Landsat water observation. The red dots correspond to river mouth pixels whose drainage area is >100 km^2. The green dots represent the terminating pixel of inland basins (the size of green dot corresponds to the drainage area). The background map represents the topographic slope. (a) the Pearl River basin in Southern China.

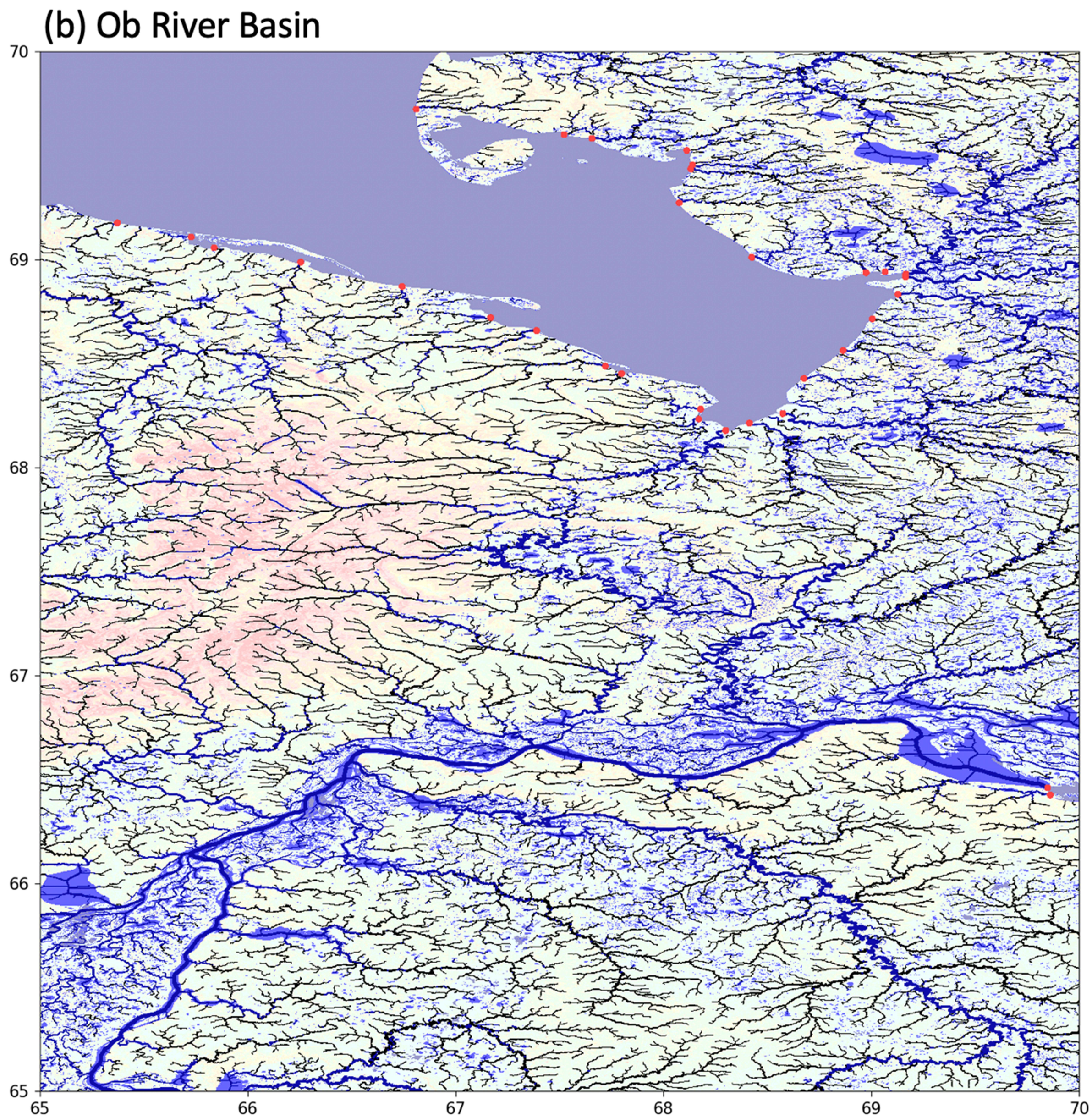


Figure 5. Delineated river network in MERIT Hydro (continued). (b) the Ob River mouth region in Western Siberia in Russia. The stream density looks smaller compared to other two regions because of the projection, as the size of 3-sec pixels is smaller at high latitudes.

Figure 5b illustrates the detected river networks around the river mouth of the Ob River. Previously, high-resolution hydrography data were not available in high latitudes because HydroSHEDS was developed based on the SRTM3 DEM, which only covers between N60 and S56. As we used the new MERIT DEM, which covers between N90 and S60 as the input data, the new high-resolution hydrography can be produced above 60N. Here, the detailed meandering structure of small streams can be observed, which were not resolved in previous maps at coarser resolution (see comparison to GDBD in Figure S3f).

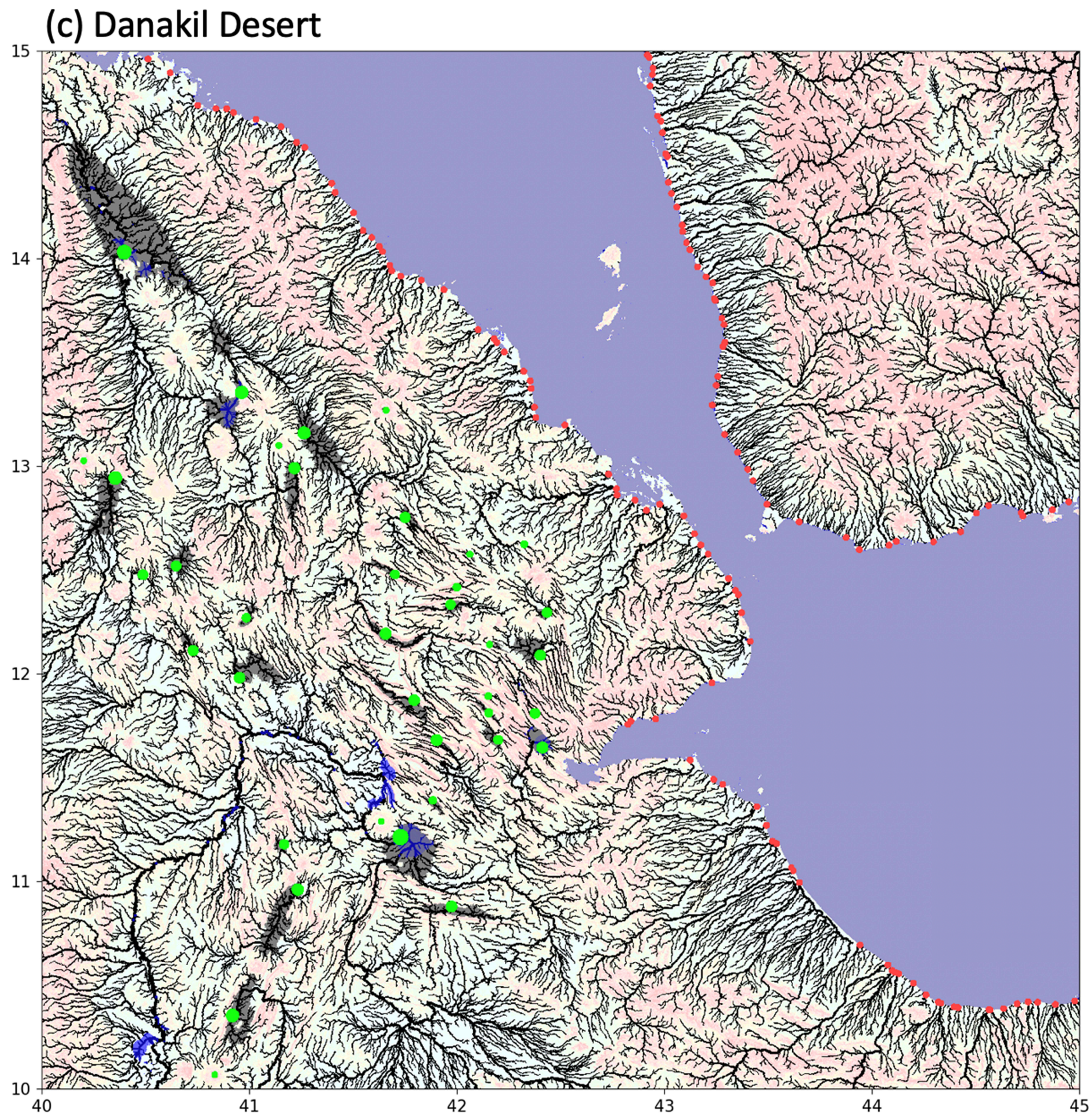


Figure 5. Delineated river network in MERIT Hydro (continued). (c) Danakil Desert in Ethiopia. The green dots represent the flow termination points of inland basins, while dot size corresponds to the size of inland basin. Gray shaded background represents the location of enclosed depressions.

Figure 5c illustrates the delineated river network around the Danakil Desert in Ethiopia. This region includes many inland river basins and depressions such as volcanic craters. The separated inland basins identified by the developed algorithm are shown as green dots in Figure 5c. Major inland basins such as Lake Abbe (N11.2, E41.8), Lake Asal in Djibouti (N11.6, E41.2) , Lake Afrera (N13.2, E40.9), and Lake Asal in the Danakil Depression in Ethiopia (N14.0, E40.4) are well represented, as are some volcanic craters (e.g. Aruku Volcano, N13.27, E41.66). The separation of actual inland basins and dummy depressions is one of the most difficult issues in hydrography delineation (Pan et al. 2012), and previously the separation had to be done using a time consuming manual process (Lehner et al. 2008). The inland depressions detected near-automatically using the method developed in this paper are almost identical to the manually-edited HydroSHEDS data, except for minor depressions (Figure S4). Thus, the newly developed algorithm is considered to be effective for inland basin separation.

The spatial distribution of the detected inland basins is shown in Figure 6. Most inland basins are in arid climate regions, but some are observed in humid karst regions (e.g. Indochina Peninsula, Balkan Peninsula). Though these depression in karst regions are in many cases connected to another basin by underground water pathways, it is difficult to detect and represent these underground pathways within the current framework of global hydrography maps. Also, some open mines which are represented as large topographical depressions, are treated as inland basins (e.g. inland basins in Germany). In total, 9703 inland basins were detected in the new hydrography map, which is significantly less than the number identified in HydroSHEDS (16604 inland basins). It was found that HydroSHEDS has a larger number of smaller-sized inland basins (<100 km^2) compared to the new hydrography map, but the spatial distribution of inland basins was similar between the two products (Figure S5). This difference is probably caused by the methodology (i.e. manual separation or thresholding). Note that we comprehensively checked whether the detected inland depressions are located reasonably by visually comparing the hydrography data to existing airborne/satellite images.

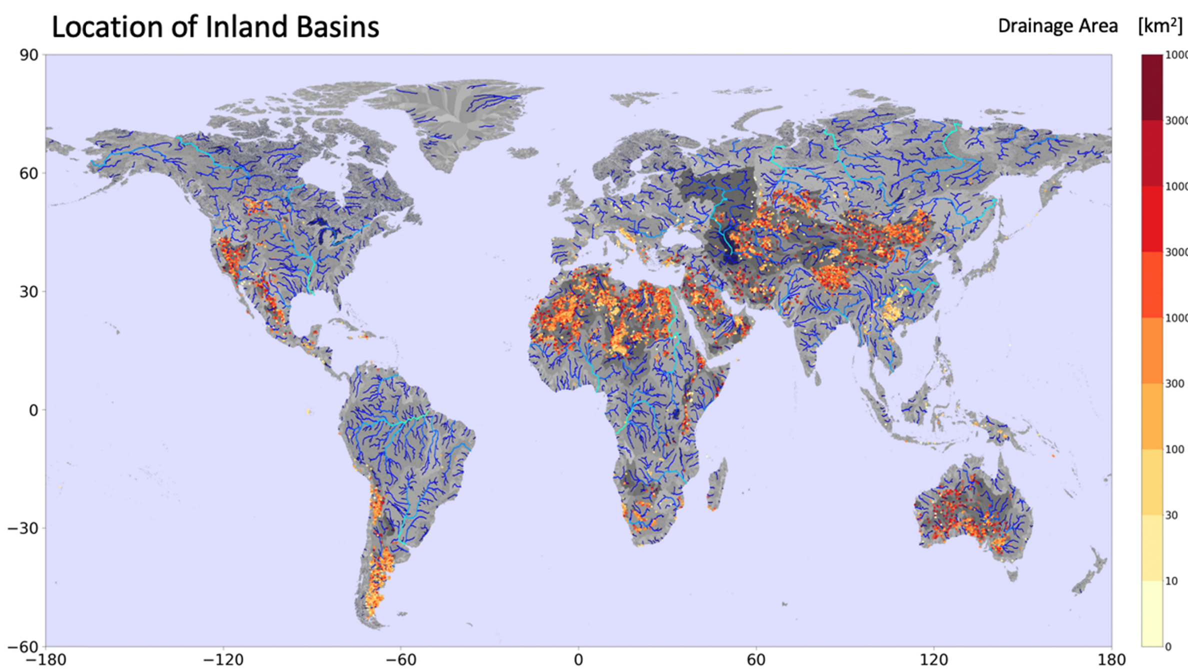


Figure 6. Spatial distribution of the detected inland basins in MERIT Hydro. The dot represents the locations of inland basin terminations, with colors representing their drainage area.

3.2 Supplementary Data Layers

The calculated river channel width of the Pearl River basin is illustrated in Figure 7a. By using the Landsat-based G1WBM water mask at 1 arc-second resolution, narrow streams whose width is around 100m are well represented. The algorithm to calculate river width is designed to keep consistency between the flow direction, water body location, and channel width (for the detailed explanation of the method, see Yamazaki et al., 2014b). The width value is therefore given to the pixels which represent the flow path of the high-resolution hydrography data. This consistency between different layers (i.e. river width, water body, flow direction and flow accumulation) is a significant advantage of the constructed global hydrography datasets when used for hydrology/hydrodynamic models, especially given that consistency between the hydrography map and river width map was not fully considered in previous datasets (see Figure S2 as an example). Note that in the developed river width layer, rivers and lakes are not distinguished (e.g. width is given within the reservoir around N23.8 E114.5). Development in the future of a river-lake classification mask will be helpful for a more detailed analysis of river morphology for advanced land hydrology modelling.

The Height Above Nearest Drainage (HAND) was also developed as a supplemental data layer (Figure 7a). As discussed by Nobre et al. (2011), HAND is a good indicator of hydrology-relevant topography such as floodplains. For example, flat basins at high elevation (e.g. the flat region around N23.5 E110.5) can be recognized in the HAND layer, while this is difficult to visualize when absolute elevation values are used. In addition to the Pearl River basin in Figure 7, the tiled figures of HAND and river channel width for the entire globe are accessible from the product webpage.

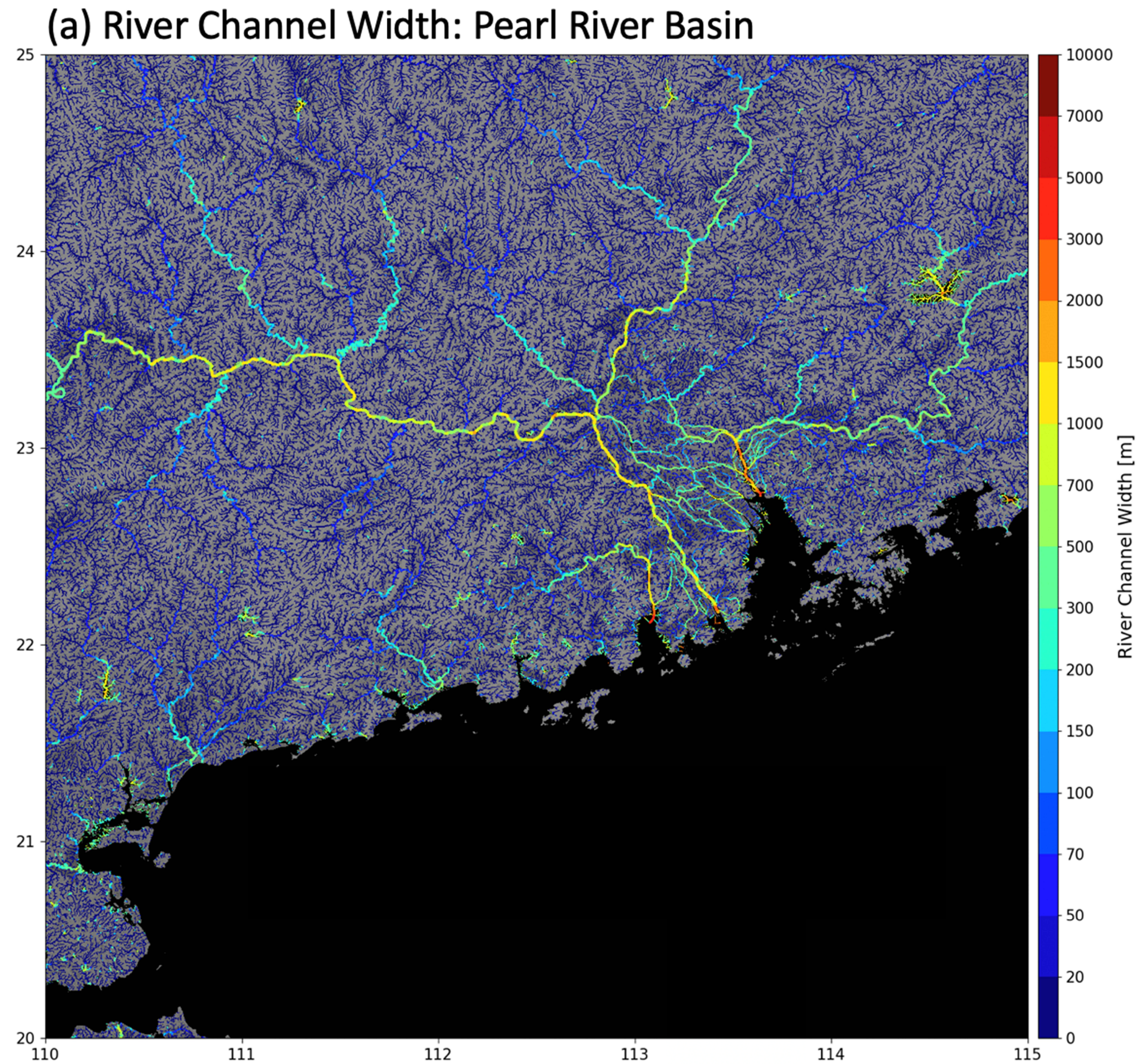


Figure 7. Supplementary data layers of the developed hydrography map. The area around the river mouth of the Pearl River is illustrated. (a) River channel width. Dark blue lines represent streams without river channel width value (e.g. not Landsat-visible stream). (b) Height Above River Channel (HAND).

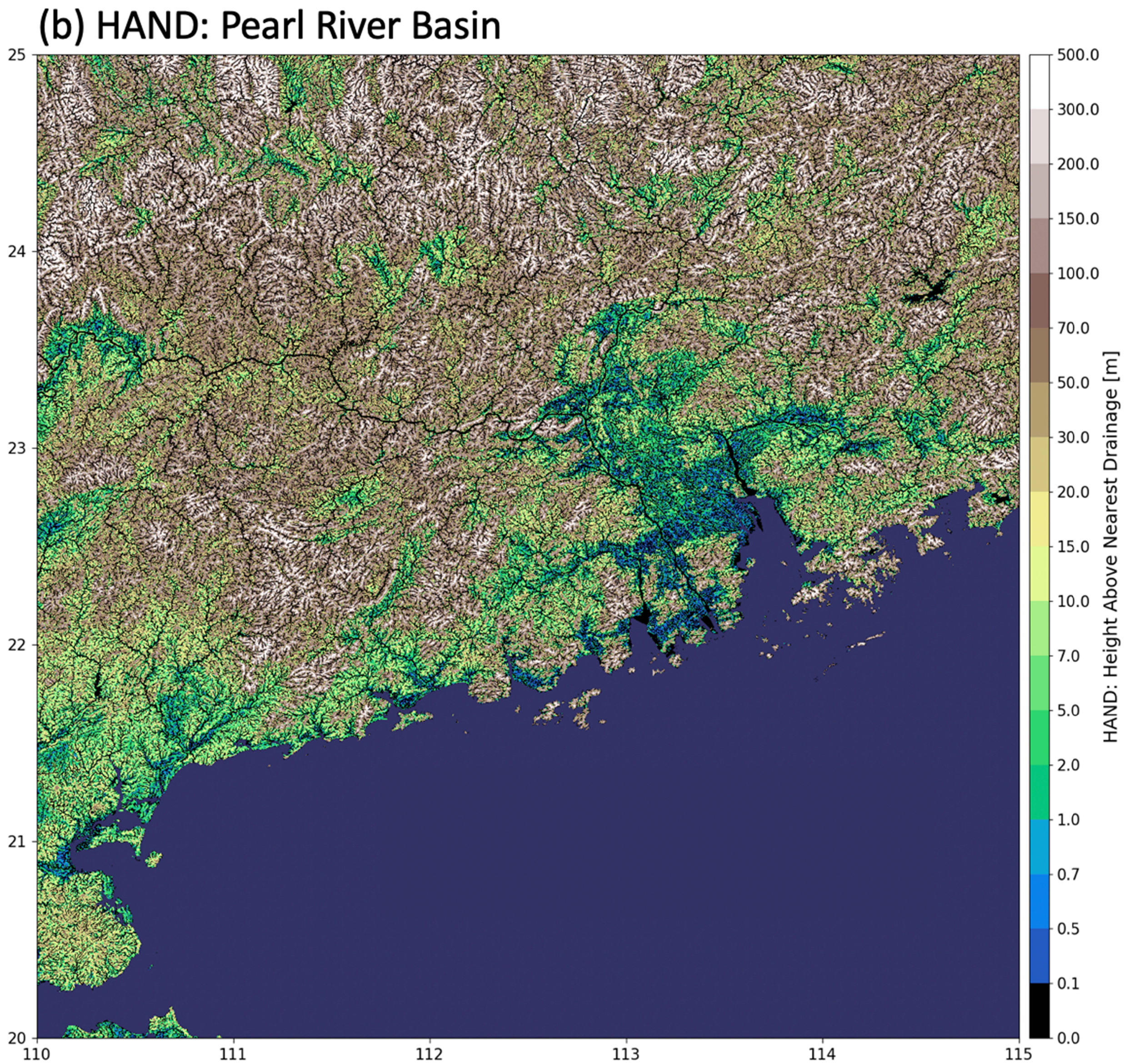


Figure 7. Supplementary data layers of the developed hydrography map. The area around the river mouth of the Pearl River is illustrated. (a) River channel width. Dark blue lines represent streams without river channel value. (b) Height Above River Channel (HAND).

A hydrologically adjusted DEM which satisfies the condition that "downstream is not higher than upstream" is another essential input for many hydrological applications, and thus it is developed as the supplementary data layer of the new hydrography map. In the original MERIT DEM (Figure 8a) the elevations over water bodies have large uncertainty because they are estimated mainly by interpolating surrounding terra firma topography in the baseline SRTM and AW3D DEMs. Thus, oscillations in the implied water surface elevations are observed along river channels and within some large lakes. However, in the hydrological adjusted DEM (Figure 8b) these oscillations are removed successfully, and an intuitively correct smooth variation of water surface elevations is represented. The elevation profile along the Ob River mainstem (the pink colored streamline in Figure 8d is illustrated in Figure 8e. The elevation oscillations in the original MERIT DEM are removed and smoothed in the adjusted DEM, which is likely to be important for using the data in many applications such as floodplain inundation modelling. Note that even in the hydrologically adjusted DEM, the elevation over water bodies is expected to represent the elevation of the water surface and not the channel or lake bathymetry. Further studies on estimating under-water bathymetry at global scales are needed to fully understand the geomorphology of rivers, lakes and floodplains.

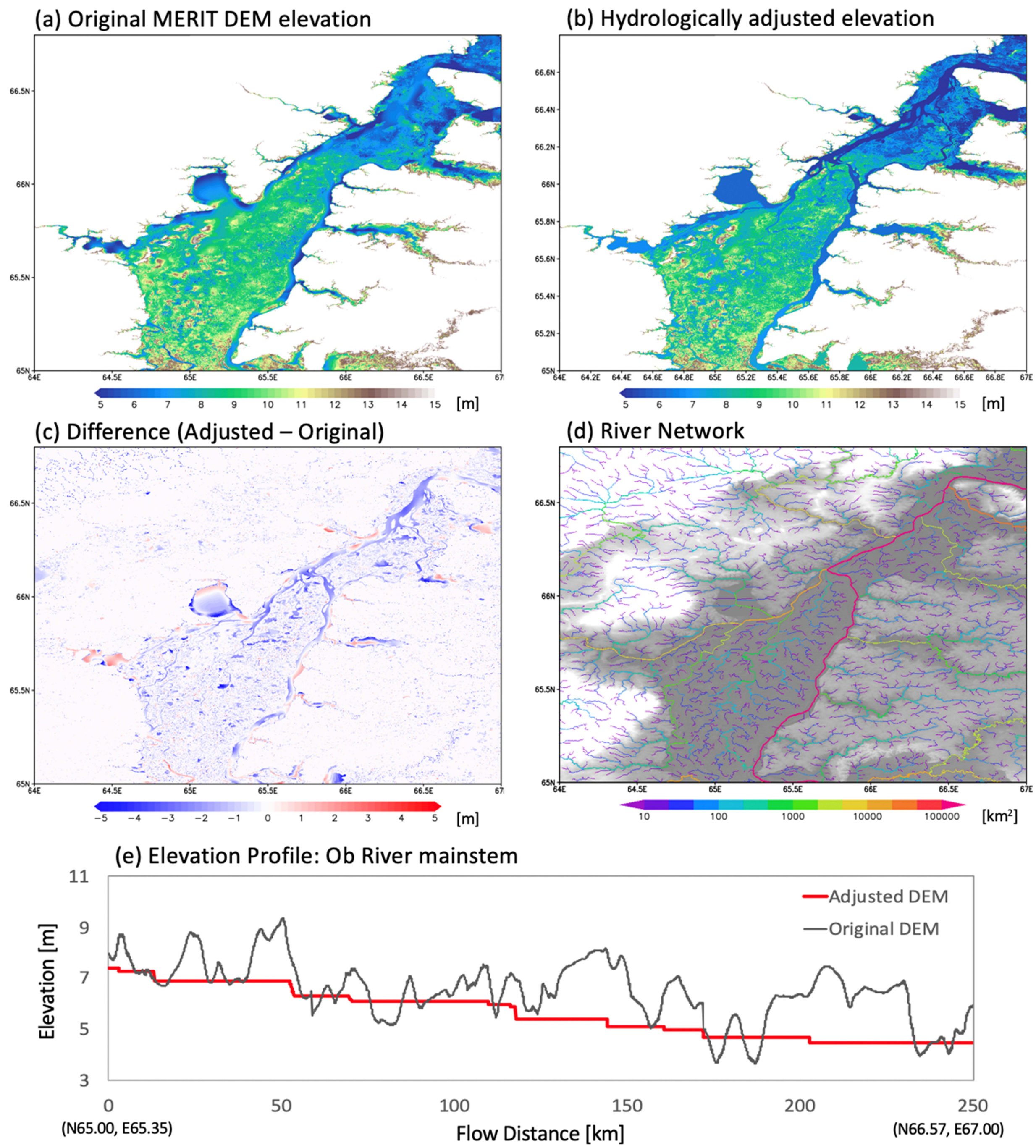


Figure 8. Adjusted elevation in the lower Ob River basin. (a) Original MERIT DEM elevation. (b) Hydrologically adjusted elevation. (c) Difference between the adjusted and original elevations. (d) Delineated river networks. (e) Elevation profile along the Ob River mainstem (pink streamline in Figure 8d).

3.3 Evaluation Against Existing Global Products

In addition to extensive quality control by visual inspection, we validated the accuracy of the developed hydrography maps by comparison to previous products. First, the flow accumulation area of the new hydrography map was compared against HydroSHEDS (Lehner and Grill, 2008) whose resolution is also 3 arc-second resolution (~90m at the equator). HydroSHEDS was developed based on the SRTM3 DEM, and is currently the most widely used global-scale hydrography map. Even though the location accuracy of streams may be limited in HydroSHEDS due to the errors in the SRTM3 DEM and limitations with available water-related data at the time of its development, the large-scale river network structure (i.e. upstream-downstream relationship) should be reliable because of the extensive quality control (Lehner and Grill, 2008). Thus, we assumed that the flow accumulation area of HydroSHEDS could be used to evaluate the river network structure of the new hydrography map. Note that the flow accumulation data are not included in the original HydroSHEDS datasets, thus we calculated them from the flow direction data using our own algorithm used for MERIT Hydro.

In order to compare the flow accumulation areas between the two datasets, the following method was used to consider the difference in stream locations. 1) Flow accumulation area is upscaled to 1-arcmin resolution to reduce computational cost. 2) For each pixel with >1000 km^2 flow accumulation area in the new hydrography map, the flow accumulation area of HydroSHEDS pixels within 2-arcmin distance (i.e. 5x5 pixels) was checked to find the minimum relative error. Note that for northern region above N60, we used GDBD (Global Drainage Basin Database, Masutomi et al. 2009) as a reference hydrography map instead of HydroSHEDS.

Figure 9a illustrates the relative error of flow accumulation area between the new hydrography map and previous datasets (HydroSHEDS and GDBD above N60). It is found that for most rivers in the world, the relative error is smaller than 5% (white lines in Figure 9a), suggesting that the upstream-downstream relationship of the river network was well reproduced by the new algorithm. A significant difference was found mainly in arid regions, because inland basins are sometimes treated differently in the different hydrography maps. For example, the flow accumulation area of the Niger River was underestimated up to 15% in the new hydrography map compared to HydroSHEDS. This difference was caused because some depressions that were not delineated separately in HydroSHEDS were treated as inland basins in the new hydrography (green colored areas). These depressions are connected to the Niger River mainstem in high water years (Pekel et al. 2016), but are isolated in more typical years. There is a difficulty in treating this seasonally variable connectivity behaviour in different hydrography products. Discrepancies between the hydrography maps due to treatment of inland basins was also found in the Amur River basin, Arabian Peninsula, and southern coast of Australia.

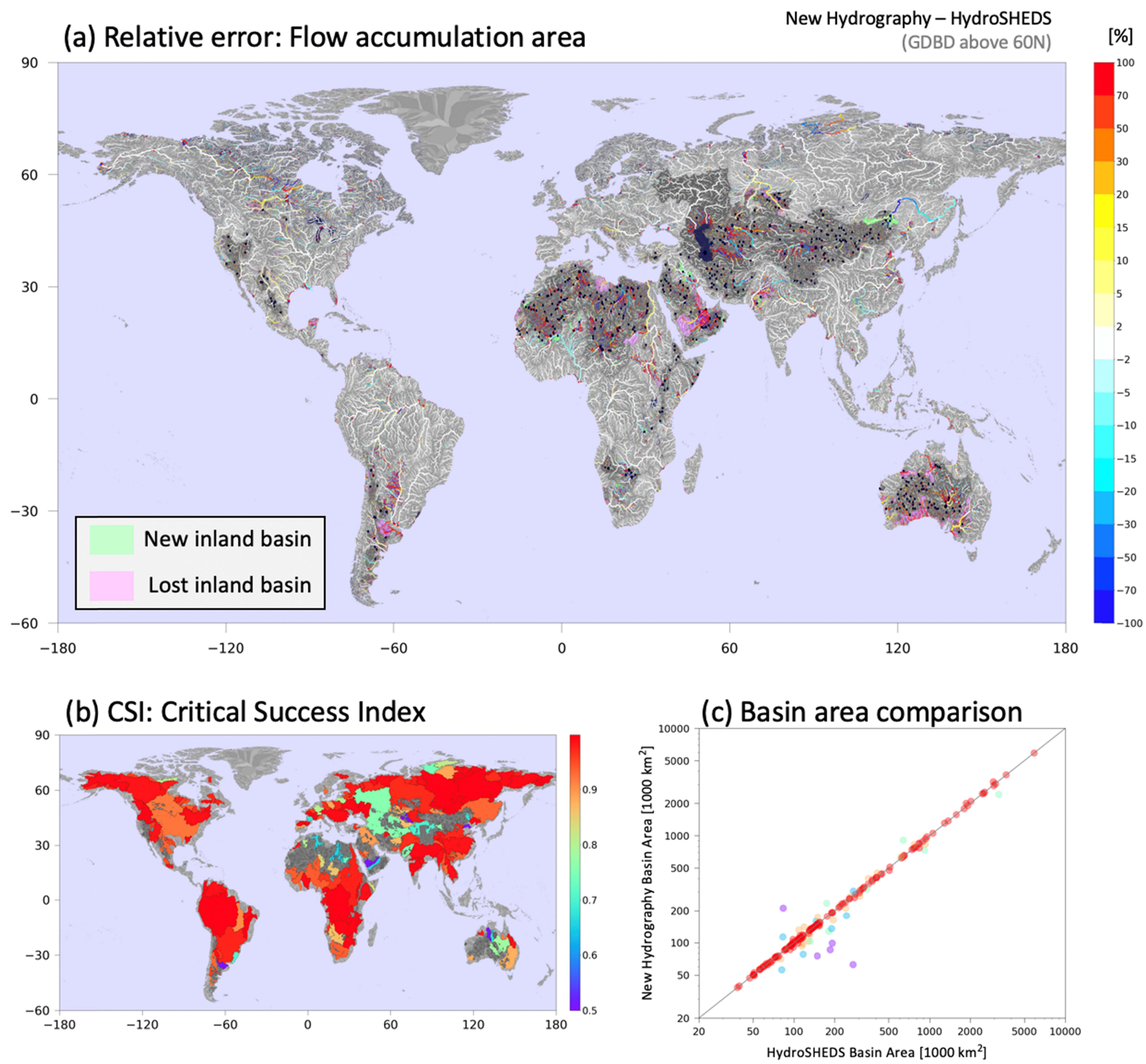


Figure 9. Comparison of the river networks between MERIT Hydro (new hydrography map) and HydroSHEDS (GDBD above N60). (a) Relative error in flow accumulation area for rivers larger than 1000km^2. Yellow and red colors represent rivers whose flow accumulation areas are larger in the new hydrography data, while blue colors represent where flow accumulation area is smaller in the new hydrography. The green color represents areas which are treated as inland basins only in the new hydrography data, while the pink color represents the "lost" inland basins which only existed in HydroSHEDS. (b) Critical Success Index (CSI) for the 200 largest basins. (c) Scatter plot for drainage area for the 200 largest basins. The colors in of dots represent the CSI in panel (b).

Varying treatment of bifurcating channels is another reason for the discrepancy. As only one downstream direction is assumed at each pixel, the current framework of global hydrography maps cannot fully represent bifurcating channels. If there is a distributary with multiple downstream channels, only one should be selected as the downstream flow direction. This problem can be recognized in some of the river segments with overestimation and underestimation of flow accumulation areas (e.g. Ob, Makenzie, Mississippi-Atchafalaya). The discrepancy is similarly present for lakes with multiple outlets, such as Southern Indian Lake (N57.3, W98.4) and Wollaston Lake (N58.2, W103.3) in Canada.

The reasons for other discrepancies between the hydrography maps were basin-specific. The flow accumulation area was larger in the Pyasina River and smaller in the Khatanga River and Taymyr River (around N71, E93) in the new hydrography data. We found that some parts of the Pyasina River basin were wrongly treated as upstream areas of the Khatanga River and Taymyr River in the GDBD hydrography, probably because it is difficult to calculate river networks appropriately in these regions with very flat topography. The Caspian Sea was treated as “sea” (no data) in HydroSHEDS, so no flow directions or flow accumulation areas were assigned. The difference in the Parana River basin (around S25, W60) was probably caused by the lack of water layer information. It is difficult to decide the location of streams in this region because topography is relatively flat and the river width is too narrow to be observed by Landsat. The stream location data from OpenStreetMap was also limited. Availability of higher-resolution data on elevation and water bodies must be a key to improve the accuracy of hydrography maps in the future.

In addition to flow accumulation areas, we also compared the shape of river basins between the new hydrography map and HydroSHEDS using the Critical Success Index (CSI). CSI for one river basin is calculated as in Equation (2):

$$CSI = \frac{N \cap H}{N \cup H} \qquad (2),$$

where N and H are the group of pixels in the considering river basin in the new hydrography map and HydroSHEDS, respectively. The CSI is 1 when the shape of the considering river basin is exactly same in the two datasets, while CSI is zero when there is no overlap between the two datasets. The CSI values for the world’s 200 largest river basins are shown in Figure 9b. In addition, the drainage area of the 200 largest river basins is compared between the new hydrography map and HydroSHEDS in Figure 9c.

It is found that the CSI is very close to 1 for most river basins, suggesting the shapes of river basins are similar between the two datasets. River basins which contain arid and semi-arid regions tend to have lower CSI index (e.g. the Mississippi and Nelson Rivers, which contain arid inland basins) because the connectivity of some inland depressions is treated differently in the two datasets. The CSI of inland river basins located in desert regions is relatively low (CSI<0.8). The CSI was also low in cases where the location of a river mouth is different in the two datasets (e.g. a river basin in one dataset could be represented in two separate river basins if the boundary of land and ocean is different in the other dataset). For example, the Caspian Sea is treated as “land” and all rivers flowing to the Caspian Sea are treated as one large basin in the new hydrography, while HydroSHEDS treats the Caspian Sea as “ocean (not land)”. As the shapes of river basins are highly affected by the treatment of inland depressions and by the specific land/ocean mask used, it is difficult to compare the accuracy of the river basin shape only using the Critical Success Index.

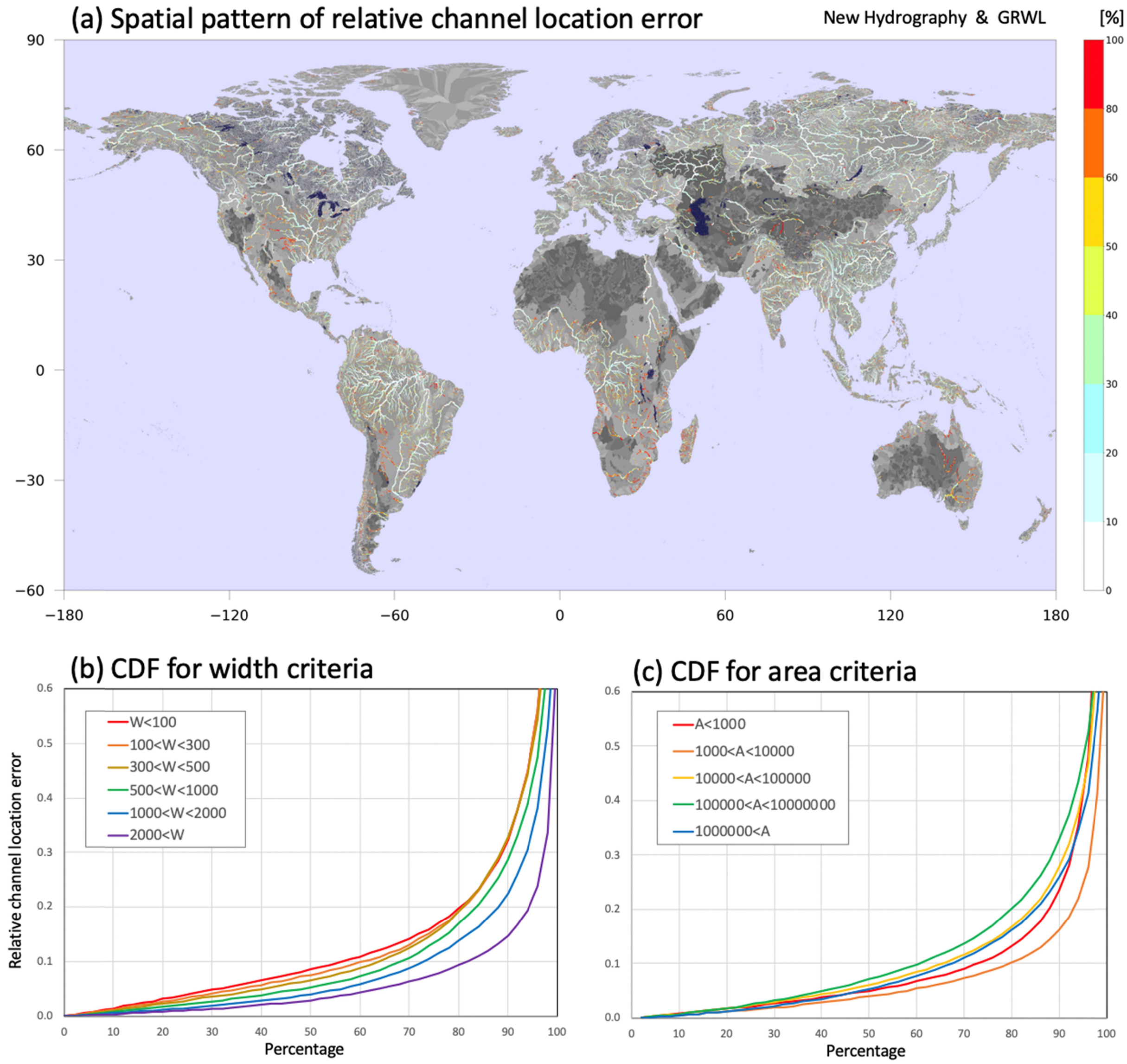


Figure 10. Relative distance error of stream centerline locations between the new hydrography and GRWL. (a) Spatial pattern. (b,c) Cumulative distribution function of relative centerline location error for channel width criteria and flow accumulation area criteria.

In order to validate the accuracy of stream location, we use river vector data from the Global River Widths from Landsat dataset (GRWL, Allen and Pavelsky, 2018). GRWL river vector data contains river centerline location and width, calculated from a Landsat-based water mask. We calculated the distance between river centerlines in the new hydrography and the GRWL vector data, assuming the centerline location of GRWL vector data is accurate because it is well quality-controlled by extensive visual editing. Given that the impact of absolute distance error in centerline location depends on the size of a river channel (e.g. 100m distance error is critical for 100m width rivers, but relatively insignificant for a 1km wide river), we used a "relative channel location error" to assess the accuracy of streamline location. The relative channel location error was calculated as "absolute distance of centerline locations divided by river channel width". A relative distance error larger than 50% means the centerline in one product is located outside of the river channel mask of another product.

The calculated relative channel location error is shown in Figure 10a. For most large rivers, the relative error was smaller than 20%, and thus stream locations are nearly identical in the GRWL product and the new hydrography. However, the relative error is sometimes larger than 50% in small rivers, probably because a 1-pixel shift in stream location is more critical here. The cumulative distribution function (CDF) of the relative channel location error for each river width bin is shown in Figure 10b. The relative location error was smaller than 50% for 95% of river segments for any channel width bin, suggesting the channel centerline locations are very similar between the new hydrography map and GRWL. This is reasonable given that channel location and river width in both datasets were based on Landsat water body data. From Figure 10b, we can observe that the relative channel location error tends to be smaller for wider rivers. On the other hand, the relationship between channel location error and flow accumulation area was not clear (Figure 10c). This is probably because larger rivers tend to have more bifurcating or braided sections where determination of river centerlines and channel width is difficult.

We also compared the flow accumulation area of the new hydrography map against the reported area at gauging stations registered in the Global Runoff Data Center (GRDC) archive. The flow accumulation area was compared at 5795 gauging stations whose area was >1000 km^2 (Figure 11). For 90% of gauging stations, the relative error was <0.05, suggesting the modelled flow accumulation area agreed with reported areas. We found some gauging stations with large errors in flow accumulation area. The large differences in reported and modelled flow accumulation area can have various causes. First, some lakes and reservoirs have multiple outlets, and the downstream accumulation area changed significantly depending on which outlet was chosen in the hydrography map. For example, the Churchill River in Canada is diverted to the Nelson River at South Indian Lake for a hydropower project, but MERIT-Hydro treated the diverted route as the trunk stream. This resulted in the overestimation of flow accumulation in the Nelson River, while underestimating the Churchill River accumulation area compared to GRDC reported values. Second, determining the connectivity of inland depressions in arid rivers is difficult and thus both the modelled and the reported flow accumulation values have uncertainties (e.g. Lake Chad basin, Rio Salado in Argentina). Third, the metadata of GRDC gauges (i.e. longitude, latitude and accumulation area) sometimes contains significant errors, and it is probably incorrect to assume all reported values are precise. Despite these limitations, the general agreement between the modelled and reported flow accumulation area suggested that the new hydrography map has an adequate accuracy for global-scale hydrology studies; we expect it to also be of sufficiently high quality for analysis in smaller-scale rivers (>1000 km^2) .

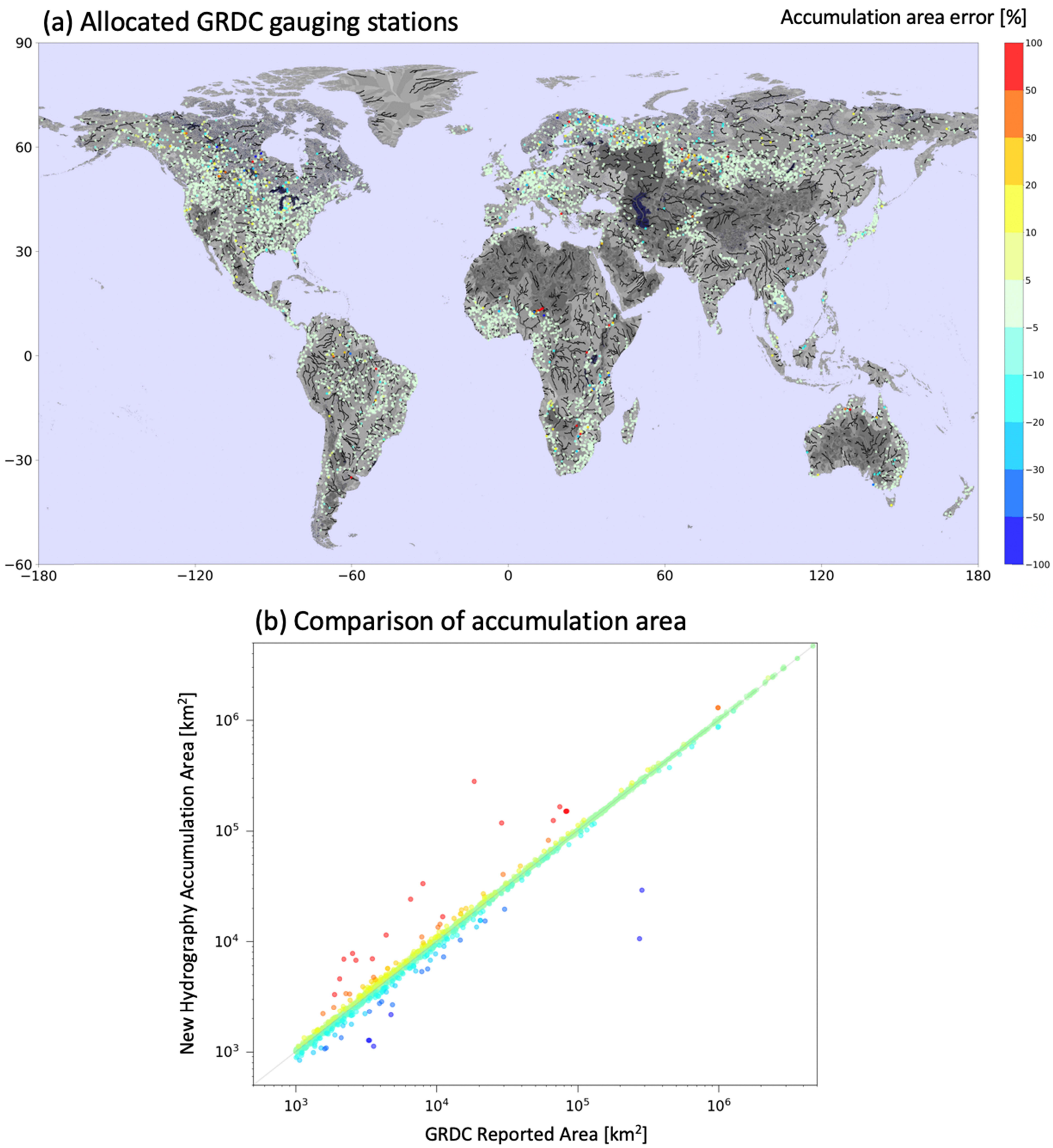


Figure 11. Comparison of flow accumulation area at GRDC gauging stations. (a) Spatial distribution of analyzed gauging stations. The colors represent the relative error in accumulation area between the new hydrography and the GRDC reported value. (b) Scatter plot of accumulation areas. The vertical axis represents the modelled flow accumulation area in the new hydrography map, and the horizontal axis represents the original reported area in the GRDC data. The color of the dots corresponds to panel (a).

4 Discussion

4.1 Importance of input data

As described in the methods section, the new hydrography map was constructed using multiple input data sources. In order to check the importance of each input data, we constructed hydrography maps with different configurations of input data usage. Figure 12 illustrates the constructed river network maps of the Tone River basin in Japan for cases: (a) using all input data (MERIT DEM, G1WBM, GSWO, and OpenStreetMap); (b) using only MERIT DEM (no water-related input); (c) using MERIT DEM, G1WBM and GSWO (i.e. without OpenStreetMap). Note that the availability of the water-related input data for the same domain is shown in Figure 2a. As a reference, the river network of HydroSHEDS is illustrated in Figure 12d. The new hydrography map captures more detailed river networks compared to HydroSHEDS, mainly because of the increased availability of input datasets.

It is found that the river networks cannot be constructed precisely if water-related input data are not used (Figure 12b). For example, the two rivers flowing parallelly from north to south (Kinu River and Kokai River, marked with "A" in Figure 12e) were wrongly merged around [N36.0, E140.0]. If elevation data are not enhanced using water-related input layers it is difficult to resolve narrow river segments, especially when a river is running through a narrow valley (see Figure 2c-d). It is also confirmed that the information from OpenStreetMap water layers is essential to represent smaller streams whose channel width is smaller than the DEM pixel size. Especially in flat topography (e.g. around N36.1, E139.6, marked with "E" in Figure 12e), it is virtually impossible to generate actual stream networks that mostly consist of narrow, man-made narrow irrigation canals without OpenStreetMap data. This result suggest that the Landsat-based water input is essential to represent the major river networks, and it is desirable to use the information from OpenStreetMap to realistically represent streams narrower than the pixel size.

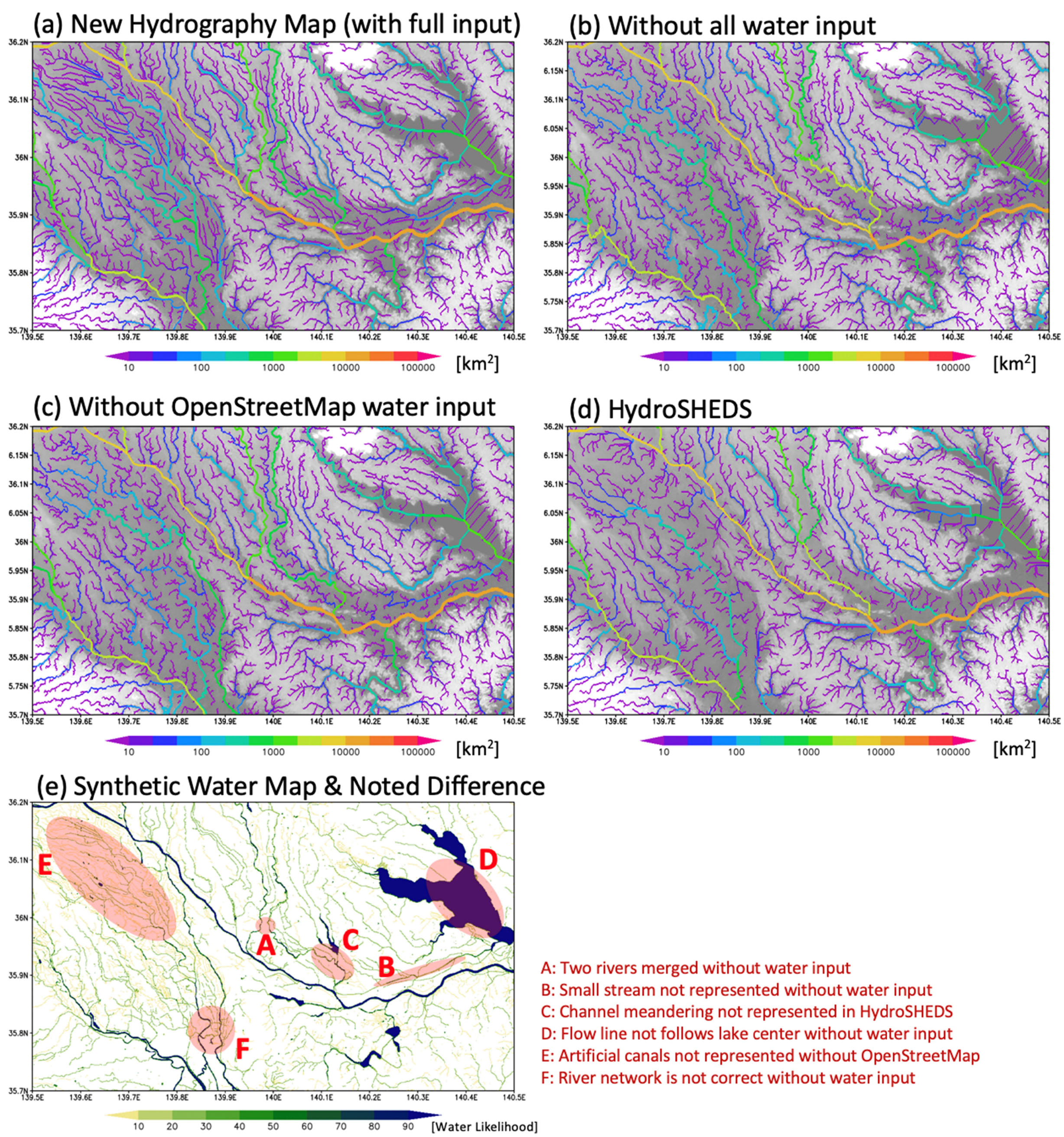


Figure 12. Comparison of river networks between the new hydrography map and HydroSHEDS in The Tone River in Japan (same region as Figure 2). In order to discuss the importance of the input data, the hydrography map was generated by different settings. (a) The new hydrography map using all input data as described in the method section. (b) Hydrography map generated without al water-layer input. (c) Hydrography map generated without OpenStreetMap water layers. (d) The river networks in HydroSHEDS. (e) Noted differences on the synthetic water map.

4.2 Limitations & Future Work

Although the newly developed hydrography map has improved accuracy compared to previous products, there still exist several limitations that should be addressed in the future. First, channel bifurcations are not represented in the current hydrography map framework, as only one downstream direction is assumed at each pixel. As delta regions, where such bifurcations are common, are important for climate change (Chmura et al. 2003) and flood risk (Ikeuchi et al. 2017), representation of channel bifurcations by allowing multiple downstream directions is required. There is a method to analyze bifurcation channels in the field of flood inundation modelling (Yamazaki et al. 2014a), and applying a similar method may be useful even for a global high-resolution hydrography map.

Similarly, multi-level crossing of flow pathways cannot be represented in the current framework. For example, there are many underground channels in karst topography, and representing underground channels is essential to estimate large-scale water balance beyond watershed boundaries at the terrain surface. Man-made canals sometimes have under-ground and over-ground crossings, so representation of a multi-level stream network could also be important for local and regional water resources management purposes.

Representation of human-made structures is a remaining challenge. MERIT Hydro contains some artificial channels which are visible in Landsat imagery or included in OpenStreetMap, but most small canals are likely not represented in current datasets. Furthermore, human-made channel networks usually have a complex upstream-downstream relationship, especially in urbanized flat terrain, so estimated flow directions in artificial channels may contain errors. Also, representation of artificial reservoirs in MERIT Hydro needs to be enhanced for more advanced applications. Currently, reservoirs are represented simply as "pixels with water bodies", without any separation from natural lakes and rivers. Aggregation and classification of water body pixels as rivers, lakes, or reservoirs will be a future task.

It is known that D8 flow direction methods are not adequate to represent flow contributing area in headwater regions, and for these zones more flexible D16 or D-Infinity flow direction representations have been proposed (Tarboton et al. 1997). In this study, a D8 approach was adopted for achieving the calculation of flow directions at a global scale, but probably more flexible and precise flow direction methods, such as D-Infinity, could be applied as post-processing.

Careful inspection is recommended when the MERIT Hydro is used in coastal areas. Even though the flow directions were calculated based on the latest topography datasets, coastlines are sometimes not well represented due to discrepancies in input datasets or temporal change in shorelines. The definition of the boundary between river and sea is usually ambiguous, thus it is recommended that users are recommended to check the river networks and river mouth locations of MERIT Hydro, especially for coastal hydrology applications.

Even though the quality of input datasets (MERIT DEM, G1WBM, GSWO, OpenStreetMap) was improved compared those available at the time of HydroSHEDS development, the currently-used input data still have some uncertainties. It is therefore recommended to re-generate the hydrography map regularly when new and higher-quality input data becomes available. In particular, the availability and quality of OpenStreetMap varies greatly from region to region, and any improvements to the OpenStreetMap water layer could have a significant impact. Updating of the global hydrography map is now achievable, given that a nearly-automated algorithm for flow direction calculation was developed in this study.

5 Conclusions

In this study, we constructed MERIT Hydro, a new global hydrography map (raster flow direction map) based on the latest topography and water layer data (MERIT DEM, G1WBM, GSWO, and OpenStreetMap). The MERIT Hydro more precisely represents river networks compared to previous hydrography maps such as HydroSHEDS, mainly because of improved data availability and quality that has been achieved over the past 10 years, especially for small streams and rivers in high latitude. Comparison to the GRDC, HydroSHEDS and GRWL datasets suggested that the new hydrography map does not contain significant errors in upstream-downstream relationships and channel locations in continental-scale rivers, which is very important for many application studies.

In addition to the flow direction and river networks, we also prepared supplementary data layers such as river width and hydrologically adjusted elevation. These supplementary products are carefully developed to ensure consistency among the different hydrography data layers. For example, the streamlines of the new hydrography map follow the channel centerline calculated based on the high-resolution water body data, and the channel width value is given to the river streamline pixels. This consistency among different layers will be helpful when utilizing the hydrography data for hydrology modelling, especially flood inundation models which require precise and coherent river networks, floodplain elevations, and channel cross-section parameters.

Even though some limitations remain, as discussed in section 4, we anticipate that the new hydrography map will be useful to studies relevant to river hydrology and hydrodynamics. The new hydrography data "MERIT Hydro" (MERIT-DEM based Hydrography map) will be freely available for academic research and education purpose.

Acknowledgments

All baseline input datasets are available online. The new hydrography data is freely available for research and education purposes from the developer's webpage (http://hydro.iis.u-tokyo.ac.jp/~yamadai/MERIT_Hydro/).

This research was supported by "MEXT: TOUGOU Program" and "JSPS: KAKENHI 16H06291". Paul Bates was supported by a Research Fellowship from the Leverhulme Trust and a Royal Society Wolfson Research Merit Award.

References

Allen, G. H., T. M. Pavelsky (2018), Global extent of rivers and streams, Science, 361(6402), 585-588, doi: 10.1126/science.aat0636

Chmura, G. L., S. C. Anisfeld, D. R. Cahoon, and J. C. Lynch (2003), Global carbon sequestration in tidal, saline wetland soils, Global Biogeochem. Cycles, 17, 1111, doi:10.1029/2002GB001917

Farr, T. G., et al. (2007), The Shuttle Radar Topography Mission, Rev. Geophys., 45, RG2004, doi:10.1029/2005RG000183.

Gorelick, N., M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore (2017), Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sensing of Environment, 202, 18-27, doi: 10.1016/j.rse.2017.06.031

Gutman, G., C. Huang, G. Chander, P. Noojipady, and J. G. Masek (2013), Assessment of the NASA-USGS Global Land Survey (GLS) datasets. Remote Sensing of Environment, 134, 249–265, doi: 10.1016/j.rse.2013.02.026.

Haklay, M., A. Singleton, and C. Parker (2008), Web mapping 2.0: The neogeography of the GeoWeb. Geography Compass, 2(6), 2011-2039, doi: 10.1111/j.1749-8198.2008.00167.x

Hansen, M. C., et al. (2013), High-resolution global maps of 21st-century forest cover change, Science, 342, 850–853, doi:10.1126/science.1244693.

Hengl, T., and I. S. Evans (2009), Mathematical and digital models of the land surface. Developments in Soil Science, 33, 31-63, doi: 10.1016/S0166-2481(08)00002-0

Ikeuchi, H., Y. Hirabayashi, D. Yamazaki, S. Muis, P. J. Ward, H. C. Winsemius, M. Verlaan, and S. Kanae (2017), Compound simulation of fluvial floods and storm surges in a global coupled river-coast flood model: Model development and its application to 2007 Cyclone Sidr in Bangladesh, J. Adv. Model. Earth Syst., 9, 1847–1862, doi:10.1002/2017MS000943.

Krieger, G., A. Moreira, H. Fiedler, I. Hajnsek, M. Werner, M. Younis, and M. Zink (2007), TanDEM-X: A satellite formation for high-resolution SAR interferometry. IEEE Transactions on Geoscience and Remote Sensing, 45(11), 3317-3341, doi: 10.1109/TGRS.2007.900693

Lehner, B., K. Verdin, and A. Jarvis (2006), HydroSHEDS technical documentation, World Wildlife Fund, Washington, D. C. (Available at http://hydrosheds.cr.usgs.gov).

Lehner, B., K. Verdin, and A. Jarvis (2008), New global hydrography derived from spaceborne elevation data. Eos, Transactions American Geophysical Union, 89(10), 93-94, 10.1029/2008EO100001

Miguez-Macho, G., Y. Fan, C. P. Weaver, R. Walko, and A. Robock (2007), Incorporating water table dynamics in climate modeling: 2. Formulation, validation, and soil moisture simulation, J. Geophys. Res., 112, D13108, doi:10.1029/2006JD008112.

Masutomi, Y., Y. Inui, K. Takahashi, and Y. Matsuoka (2009), Development of highly accurate global polygonal drainage basin data, Hydrological Processes, 23, 572-584, doi: 10.1002/hyp.7186

Nobre, A. D., L. A. Cuartas, M. Hodnett, C. D. Rennó, G. Rodrigues, A. Silveira, W. Waterloo, and S. Saleska (2011), Height Above the Nearest Drainage–a hydrologically relevant new terrain model. Journal of Hydrology, 404(1-2), 13-29, doi: 10.1016/j.jhydrol.2011.03.051

Pan, F., M. Stieglitz, and R. B. McKane (2012), An algorithm for treating flat areas and depressions in digital elevation models using linear interpolation, Water Resour. Res., 48, W00L10, doi:10.1029/2011WR010735.

Pekel, J. F., A. Cottam, N. Gorelick, and A. S. Belward (2016), High-resolution mapping of global surface water and its long-term changes. Nature, 540(7633), 418, doi: 10.1038/nature20584

Raymond, P. A., J. Hartmann, R. Lauerwald, S. Sobek, C. McDonald, M. Hoover, D. Butman, R. Striegl, E. Mayorga, C. Humborg, P. Kortelainen, H. Dürr, M. Maybeck, P. Ciais, P. Kortelainen (2013), Global carbon dioxide emissions from inland waters, Nature, 503(7476), 355, doi: 10.1038/nature12760

Tarboton, D. G. (1997), A new method for the determination of flow directions and upslope areas in grid digital elevation models, Water Resour. Res., 33(2), 309–319, doi:10.1029/96WR03137.

Tadono, T., H. Nagai, H. Ishida, F. Oda, S. Naito, K. Minakawa, and H. Iwamoto (2016), Generation of the 30 M-MESH global digital surface model by ALOS PRISM. International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences, 41.157-162, doi: 10.5194/isprsarchives-XLI-B4-157-2016

Turcotte, R., J. P. Fortin, A. N. Rousseau, S. Massicotte, and J. P. Villeneuve (2001), Determination of the drainage structure of a watershed using a digital elevation model and a digital river and lake network. Journal of Hydrology, 240(3-4), 225-242, doi: 10.1016/S0022-1694(00)00342-5

Turner, W. R., K. Brandon, T. M. Brooks, C. Gascon, H. K. Gibbs, K. S. Lawrence, R. A. Mittermeier, and E. R. Selig (2012), Global biodiversity conservation and the alleviation of poverty, BioScience, 62(1), 85-92, doi: 10.1525/bio.2012.62.1.13

U.S. Geological Survey (2000), HYDRO1k Elevation Derivative Database, Cent. for Earth Resour. Obs. and Sci., Sioux Falls, S. D. (Available at http://edc.usgs.gov/products/elevation/gtopo30/hydro/).

Yamazaki, D., C. Baugh, P. D. Bates, S. Kanae, D. E. Alsdorf, and T. Oki (2012), Adjustment of a spaceborne DEM for use in floodplain hydrodynamic modeling, J. Hydrol., 436-437, 81–91, doi: 10.1016/j.jhydrol.2012.02.045.

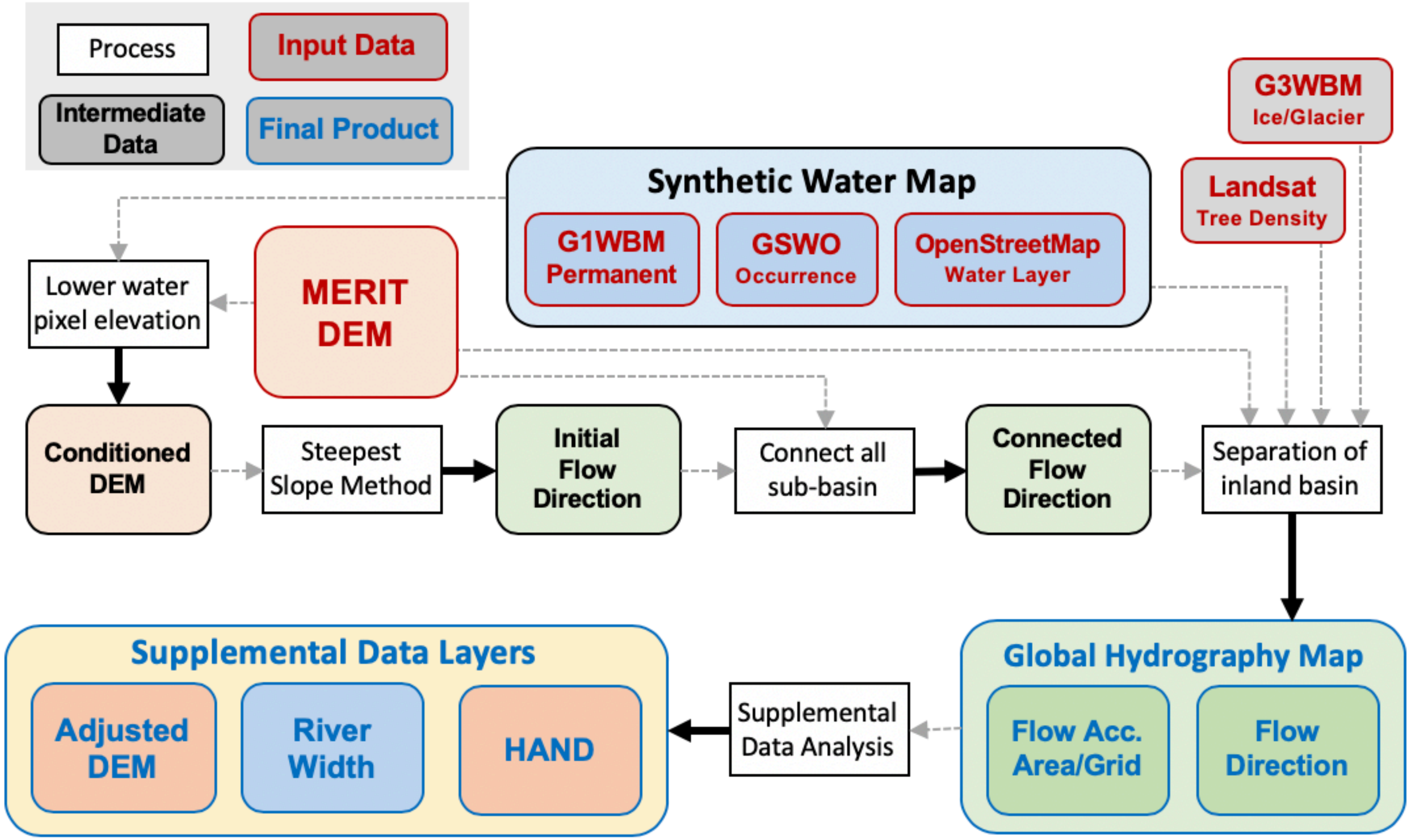
Yamazaki, D., T. Sato, S. Kanae, Y. Hirabayashi, and P. D. Bates (2014a), Regional flood dynamics in a bifurcating mega delta simulated in a global river model, Geophys. Res. Lett., 41, 3127–3135, doi:10.1002/2014GL059744.

Yamazaki, D., F. O'Loughlin, M. A. Trigg, Z. F. Miller, T. M. Pavelsky, and P. D. Bates (2014b), Development of the Global Width Database for Large Rivers, Water Resour. Res., 50, 3467-3480, doi:10.1002/2013WR014664

Yamazaki, D., M. A. Trigg, and D. Ikeshima (2015), Development of a global~ 90 m water body map using multi-temporal Landsat images. Remote Sensing of Environment, 171, 337-351, doi: 10.1016/j.rse.2015.10.014

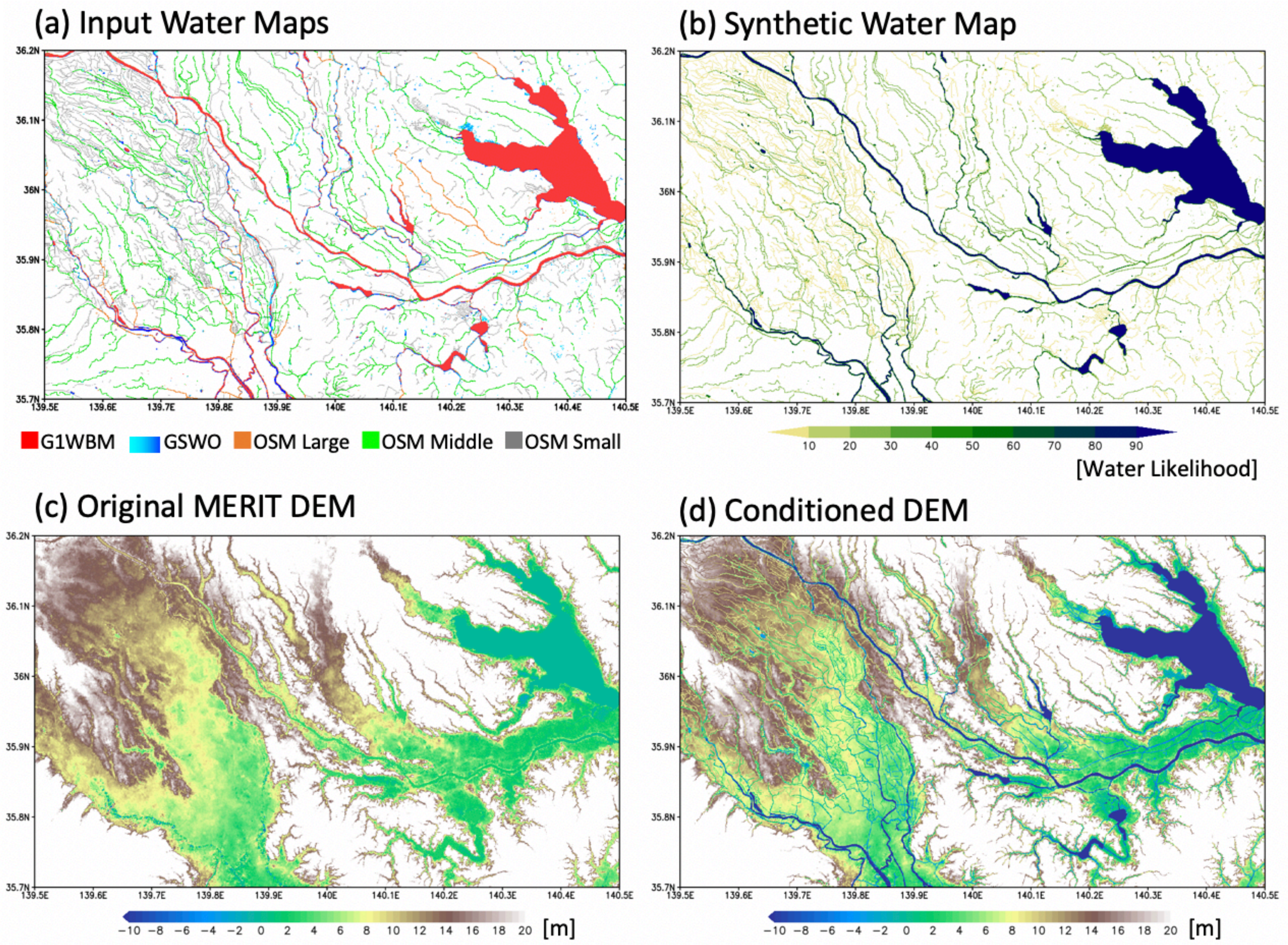
Yamazaki, D., D. Ikeshima, R. Tawatari, T. Yamaguchi, F. O'Loughlin, J. C. Neal, C. C. Sampson, S. Kanae, and P. D. Bates (2017), A high-accuracy map of global terrain elevations, Geophys. Res. Lett., 44, 5844–5853, doi: 10.1002/2017GL072874.

Figure 1.



Process
Input Data
Intermediate Data
Final Product
G3WBM
Ice/Glacier
Synthetic Water Map
Landsat
Tree Density
G1WBM
Permanent
GSWO
Occurrence
OpenStreetMap
Water Layer
Lower water pixel elevation
MERIT DEM
Conditioned DEM
Steepest Slope Method
Initial Flow Direction
Connect all sub-basin
Connected Flow Direction
Separation of inland basin
Supplemental Data Layers
Adjusted DEM
River Width
HAND
Supplemental Data Analysis
Global Hydrography Map
Flow Acc. Area/Grid
Flow Direction

Figure 2.



(a) Input Water Maps
(b) Synthetic Water Map
(c) Original MERIT DEM
(d) Conditioned DEM
G1WBM
GSWO
OSM Large
OSM Middle
OSM Small
[Water Likelihood]
10 20 30 40 50 60 70 80 90
−10 −8 −6 −4 −2 0 2 4 6 8 10 12 14 16 18 20 [m]
36.2N 36.1N 36N 35.9N 35.8N 35.7N
139.5E 139.6E 139.7E 139.8E 139.9E 140E 140.1E 140.2E 140.3E 140.4E 140.5E

Figure 3.

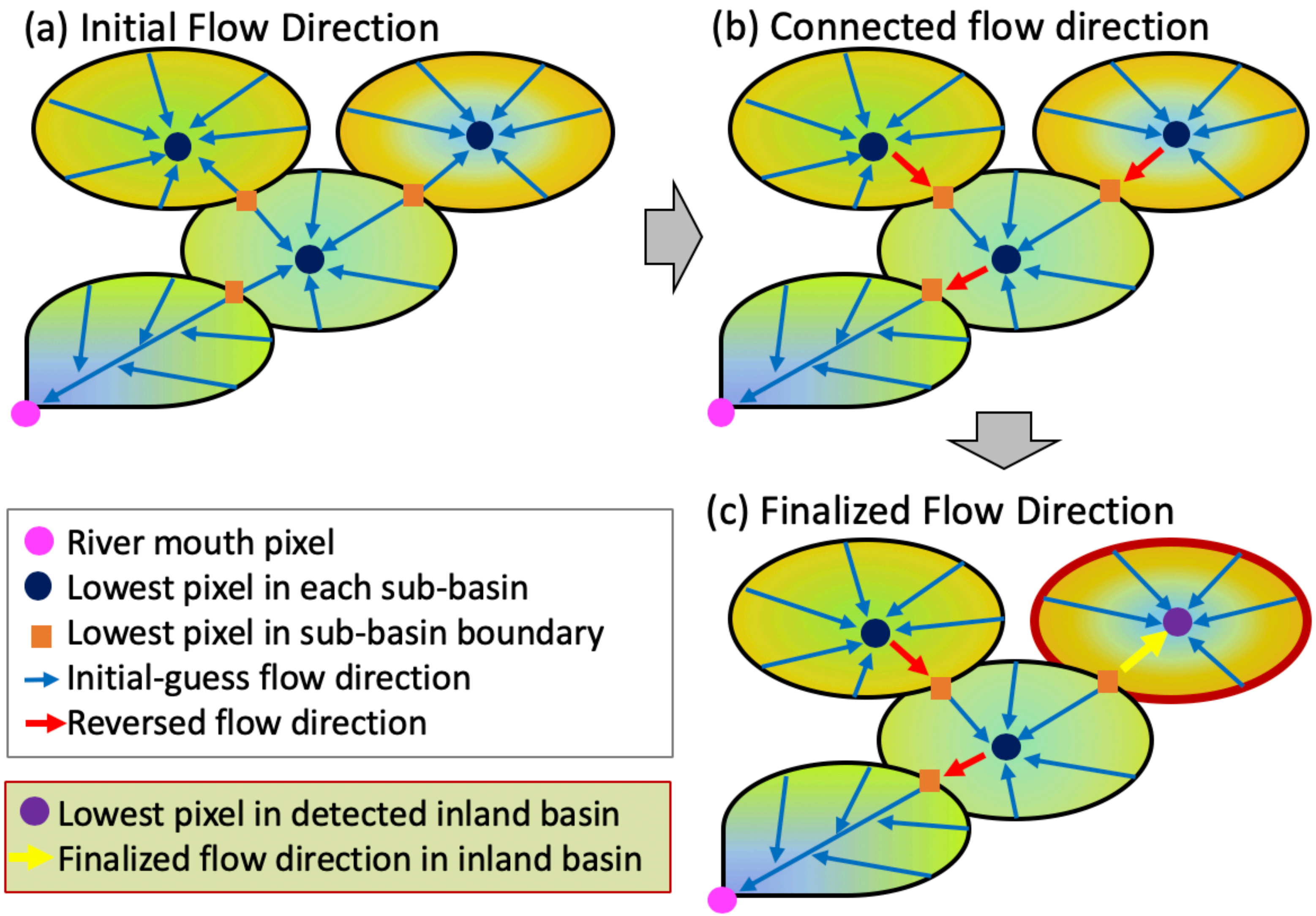
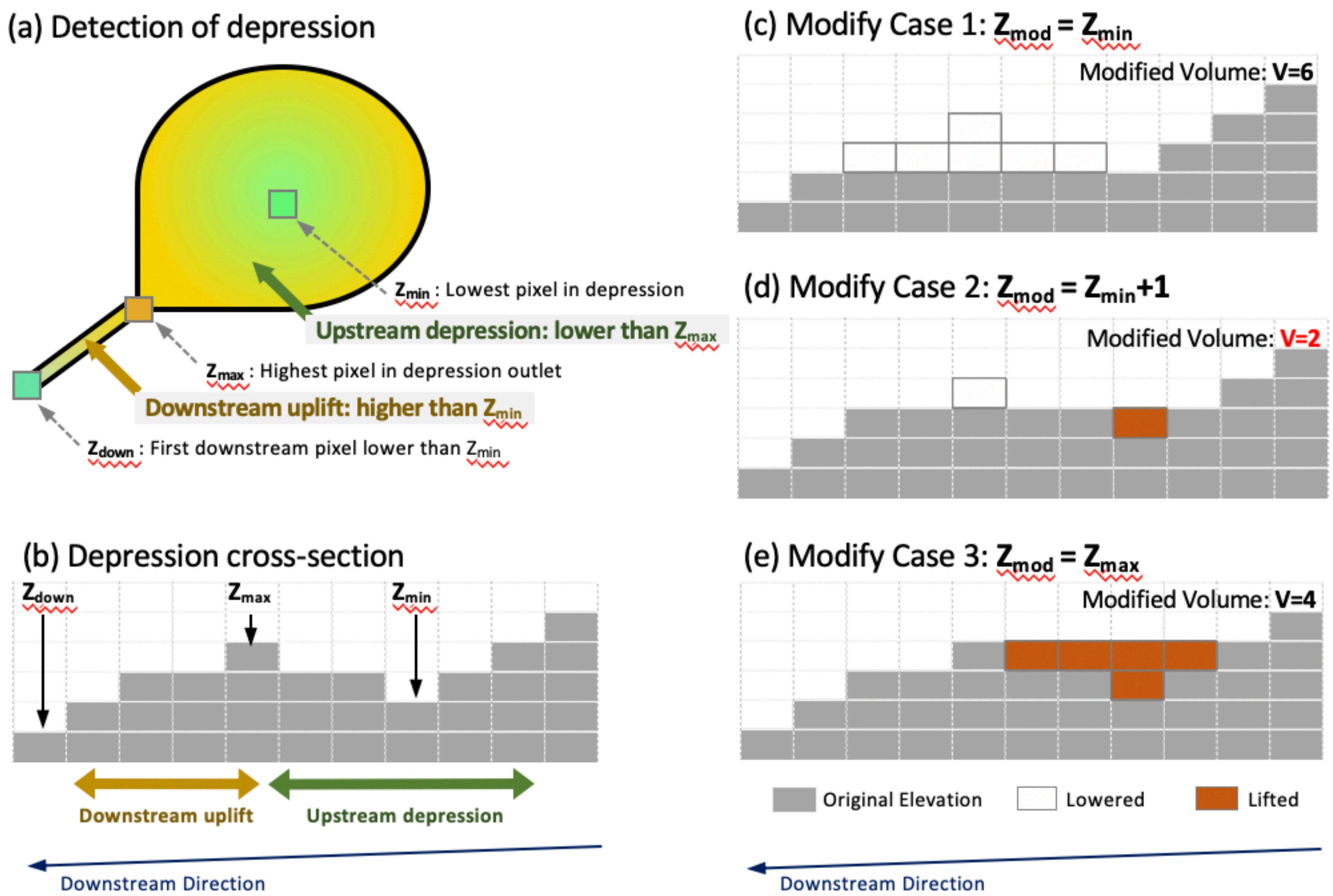

(a) Initial Flow Direction
(b) Connected flow direction
(c) Finalized Flow Direction
River mouth pixel
Lowest pixel in each sub-basin
Lowest pixel in sub-basin boundary
Initial-guess flow direction
Reversed flow direction
Lowest pixel in detected inland basin
Finalized flow direction in inland basin

Figure 4.



(a) Detection of depression
Zmin : Lowest pixel in depression
Upstream depression: lower than Zmax
Zmax : Highest pixel in depression outlet
Downstream uplift: higher than Zmin
Zdown : First downstream pixel lower than Zmin
(b) Depression cross-section
Zdown
Zmax
Zmin
Downstream uplift
Upstream depression
Downstream Direction
(c) Modify Case 1: Zmod = Zmin
Modified Volume: V=6
(d) Modify Case 2: Zmod = Zmin+1
Modified Volume: V=2
(e) Modify Case 3: Zmod = Zmax
Modified Volume: V=4
Original Elevation
Lowered
Lifted
Downstream Direction

Figure 5a.

(a) Pearl River Basin

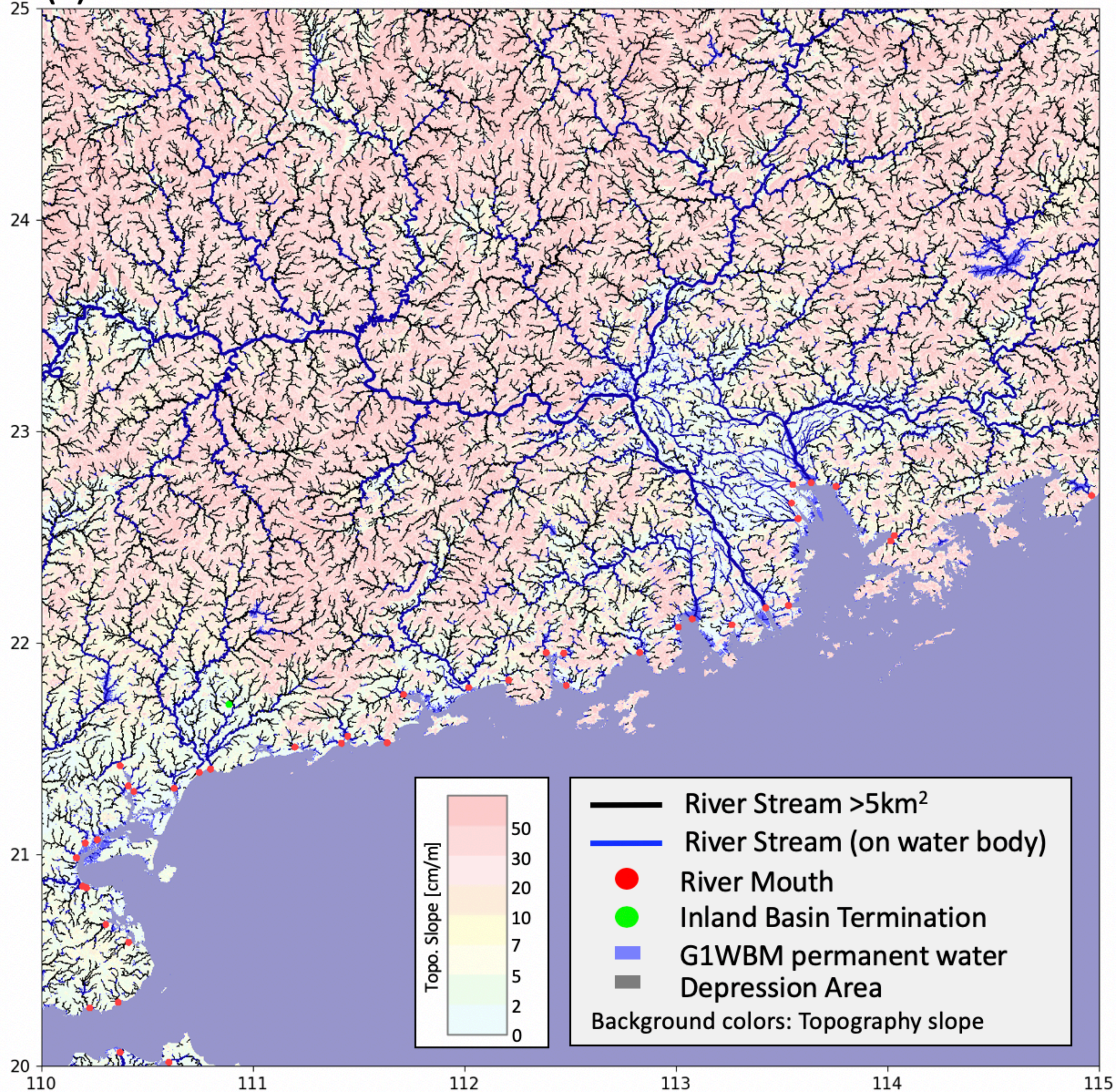


Figure 5b.

(b) Ob River Basin

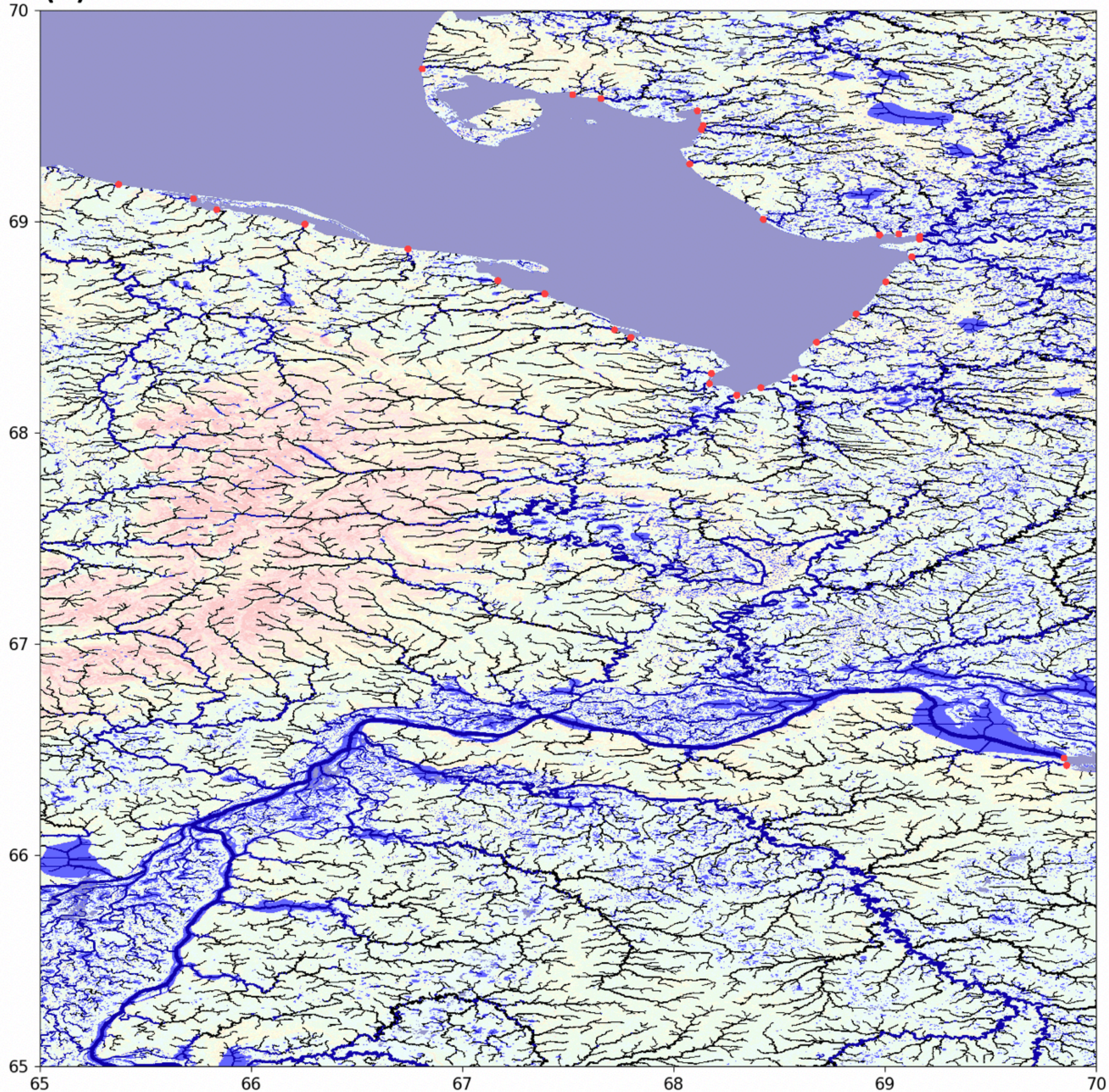


Figure 5c.

(c) Danakil Desert

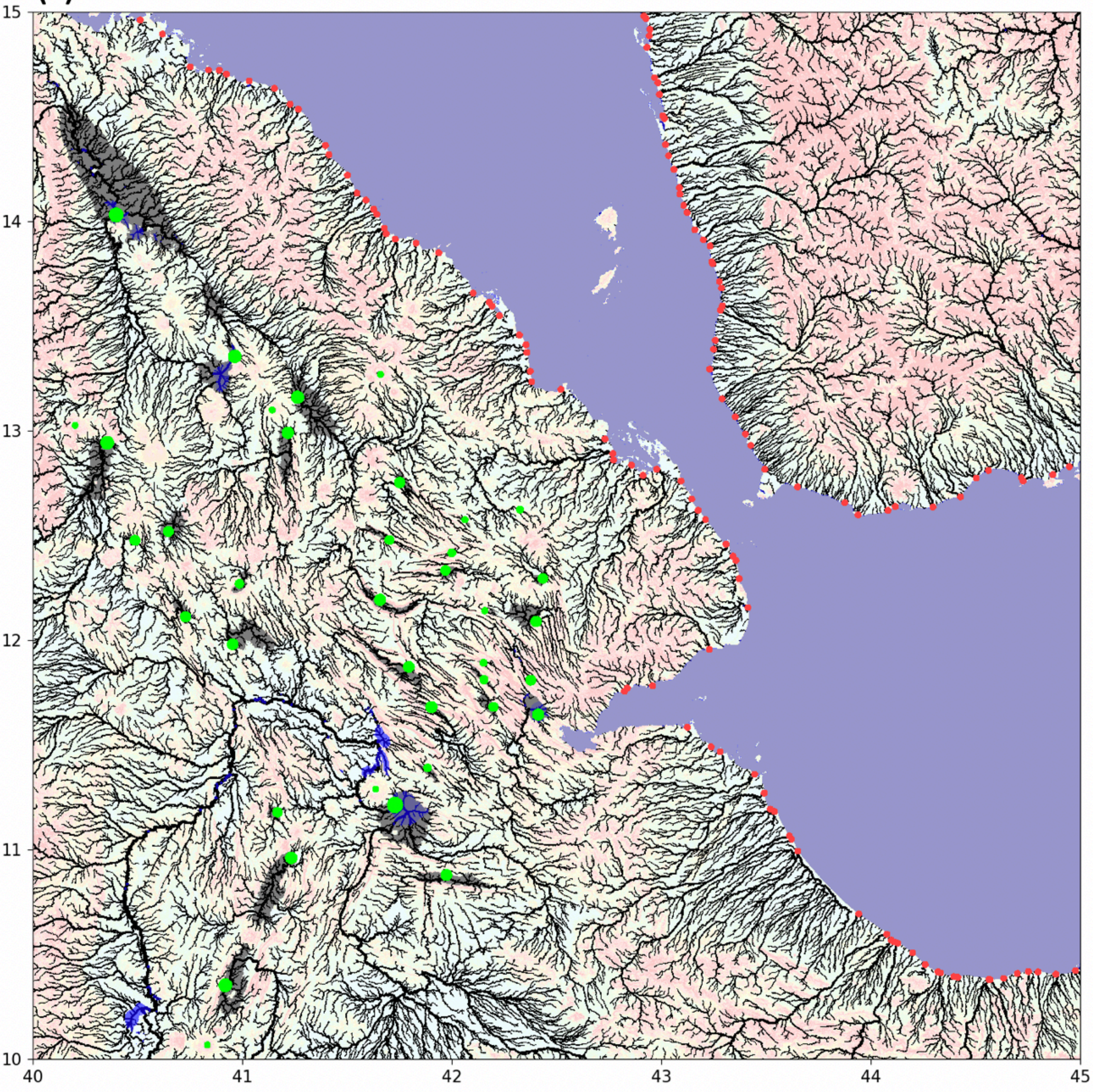


Figure 6.

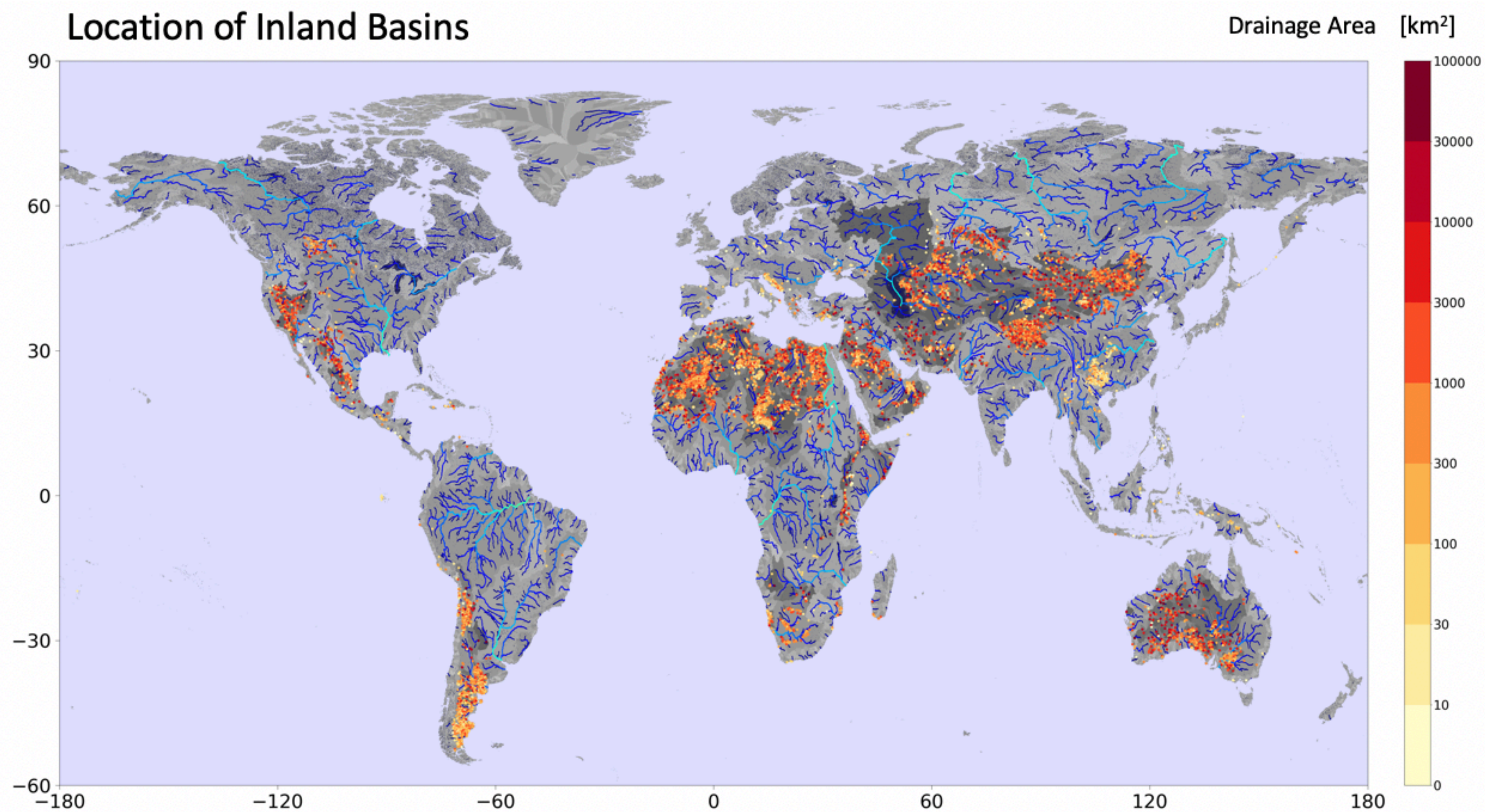

Location of Inland Basins
Drainage Area [km²]
90
60
30
0
−30
−60
−180
−120
−60
0
60
120
180
100000
30000
10000
3000
1000
300
100
30
10
0

Figure 7a.

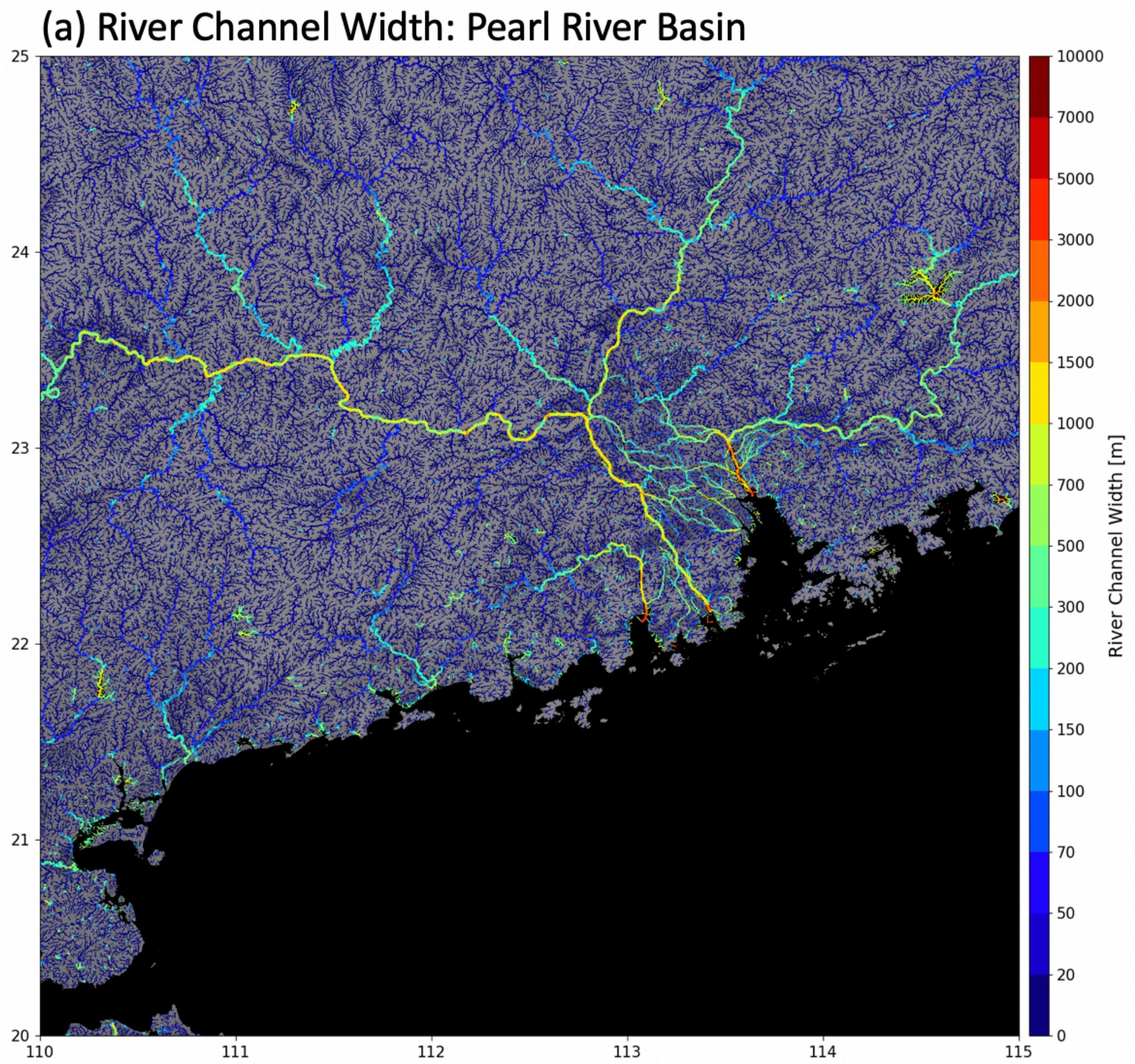

(a) River Channel Width: Pearl River Basin
25
24
23
22
21
20
110
111
112
113
114
115
10000
7000
5000
3000
2000
1500
1000
700
500
300
200
150
100
70
50
20
0
River Channel Width [m]

Figure 7b.

(b) HAND: Pearl River Basin

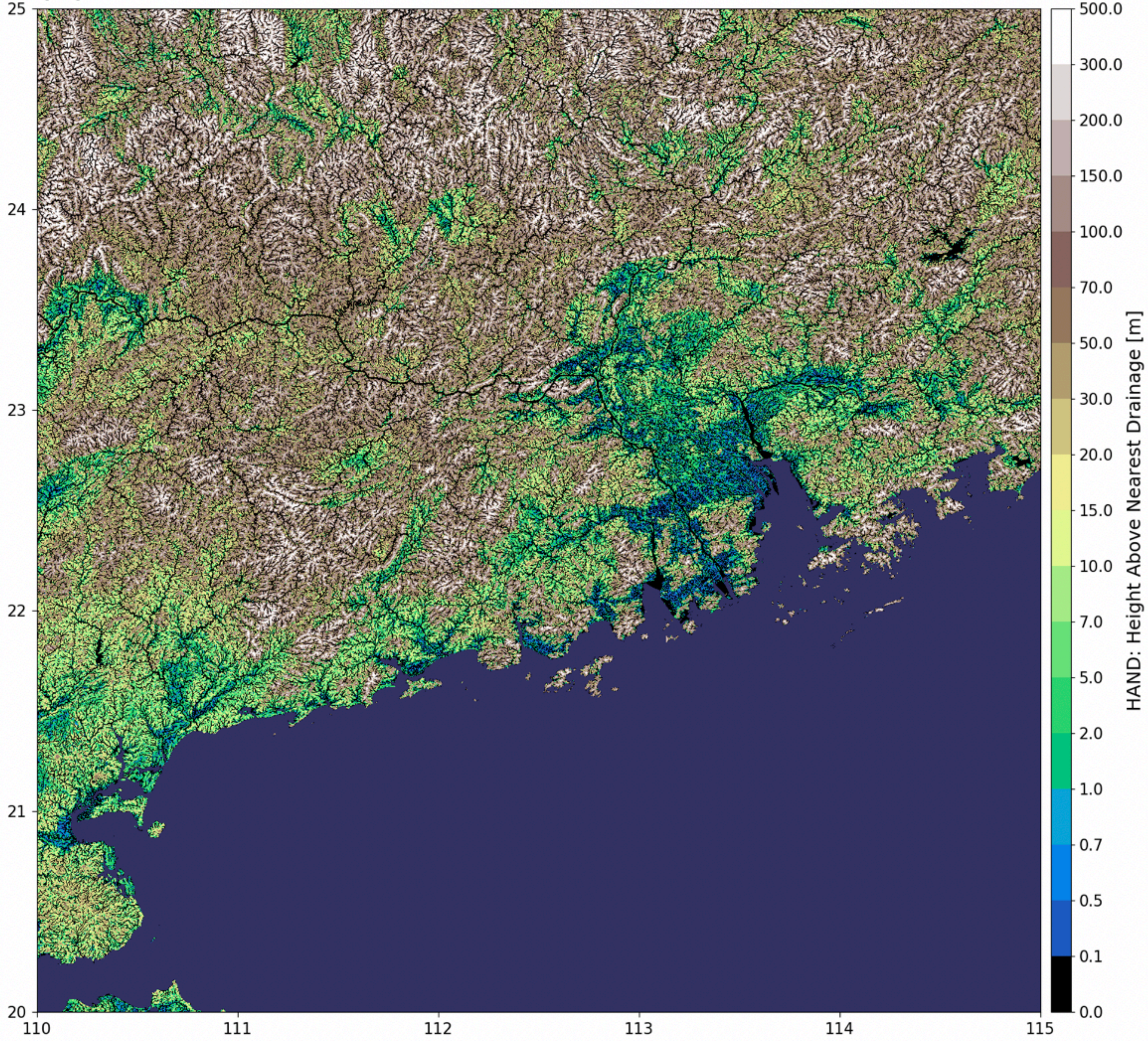
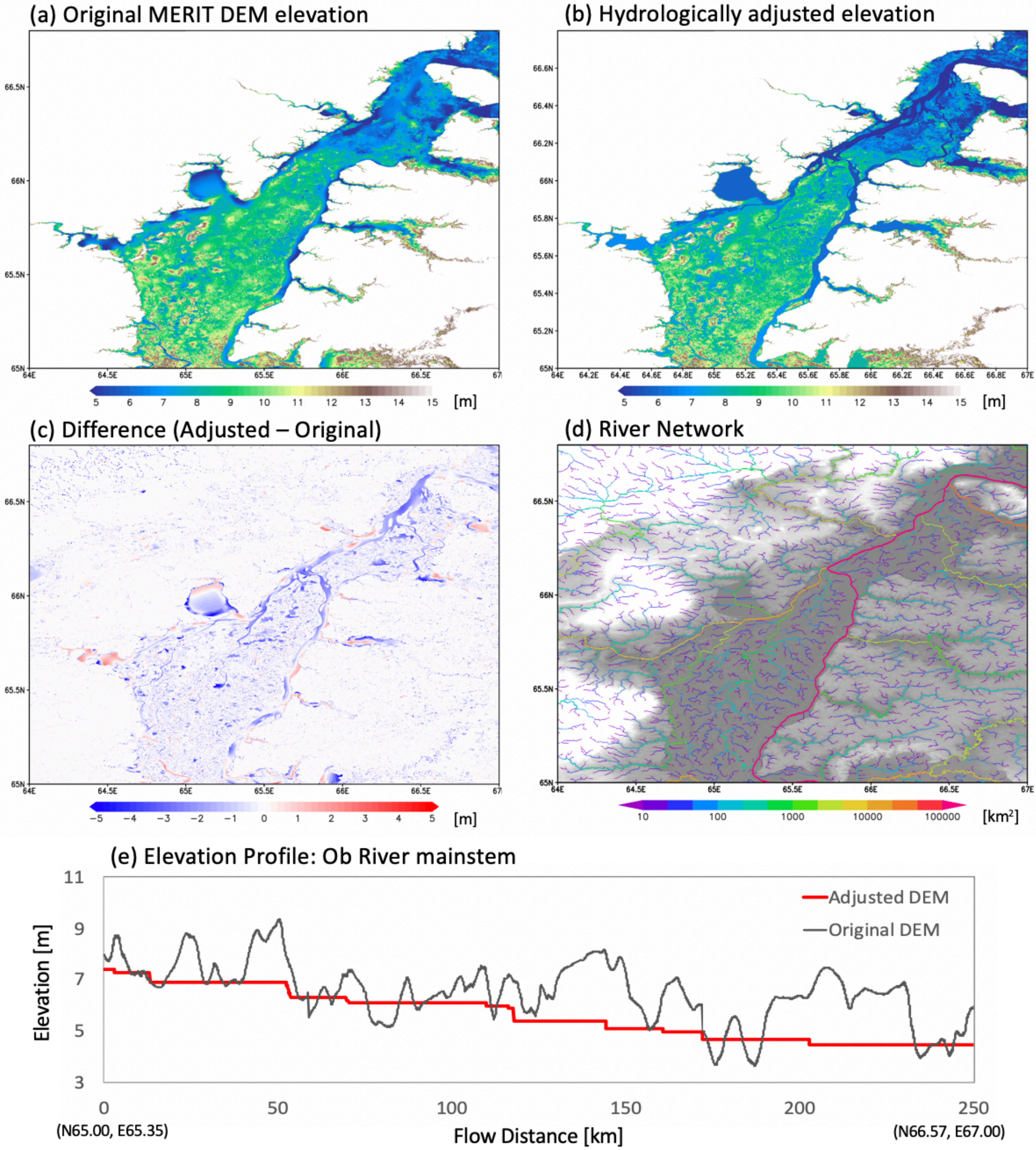


Figure 8.



(a) Original MERIT DEM elevation
(b) Hydrologically adjusted elevation
[m]
(c) Difference (Adjusted – Original)
[m]
(d) River Network
[km2]
(e) Elevation Profile: Ob River mainstem
Adjusted DEM
Original DEM
Elevation [m]
Flow Distance [km]
(N65.00, E65.35)
(N66.57, E67.00)

Figure 9.

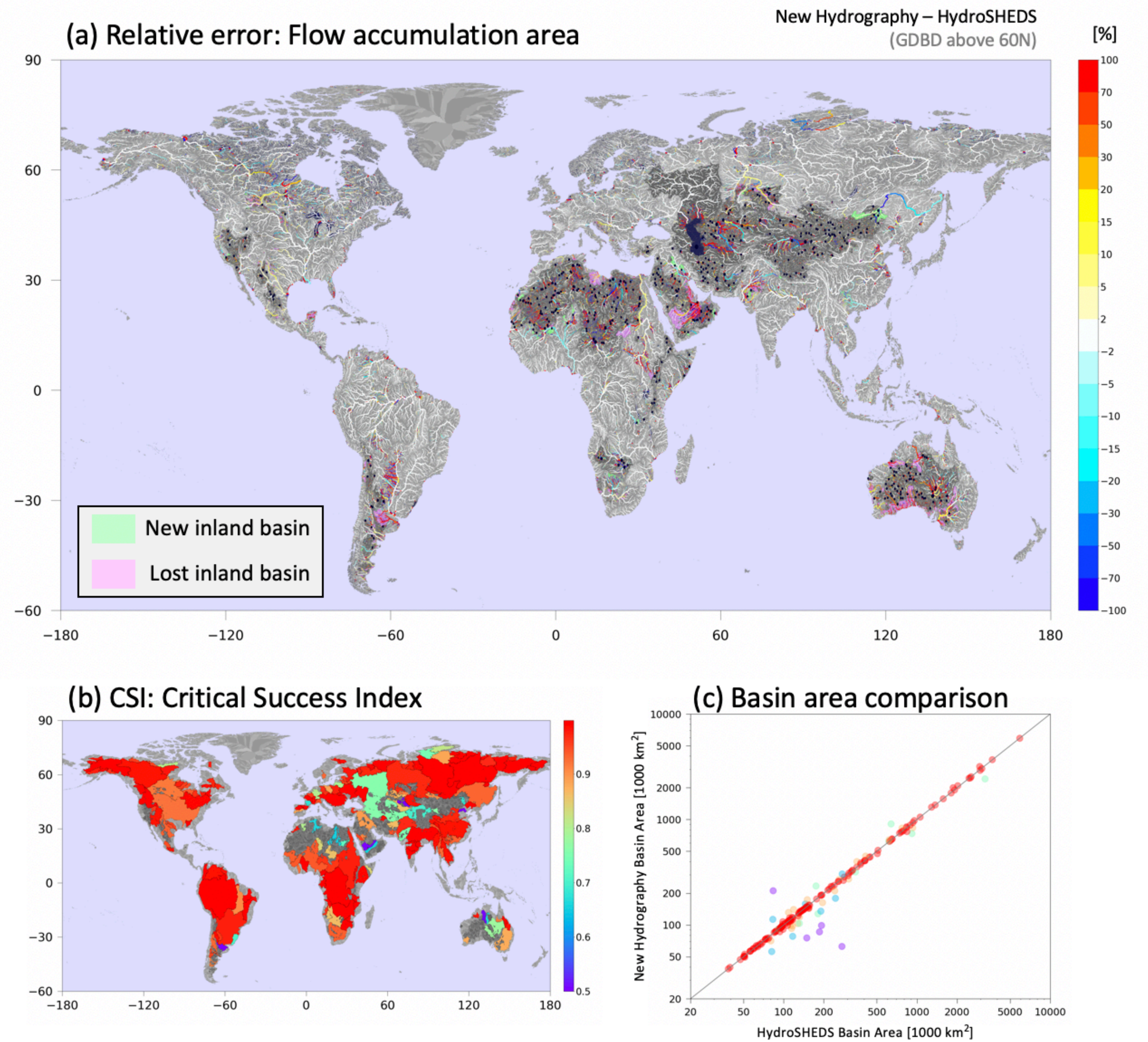
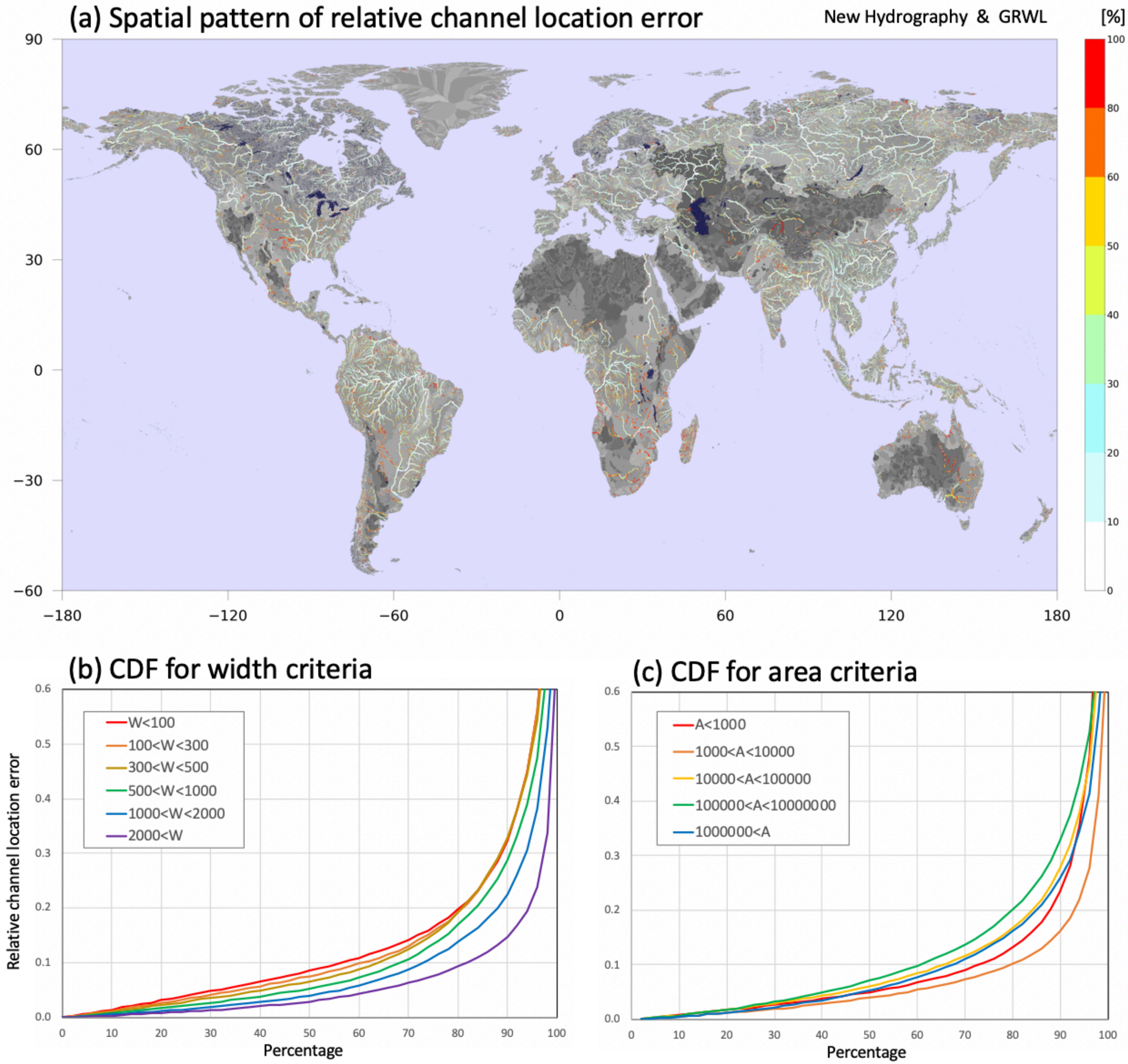
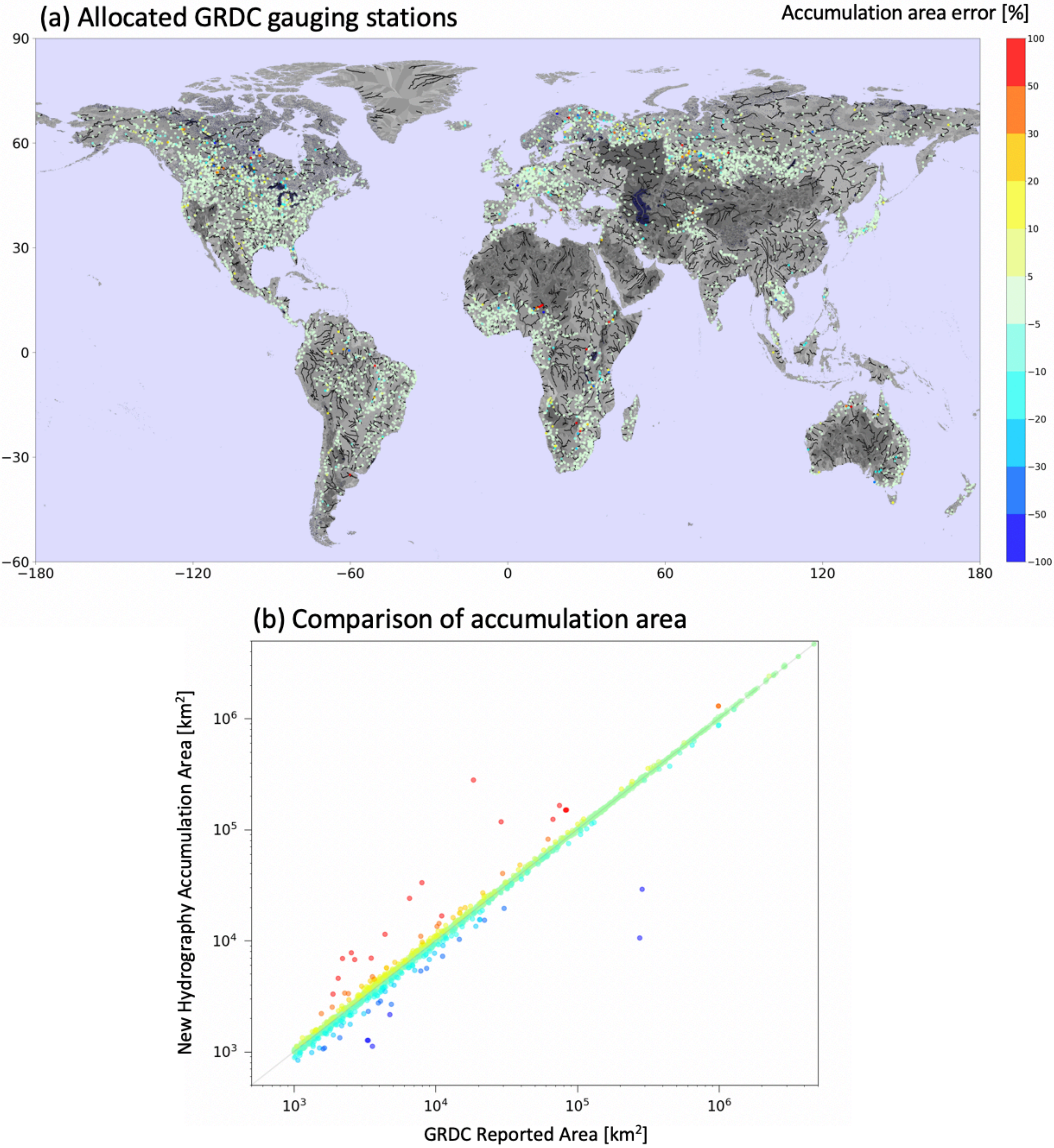

(a) Relative error: Flow accumulation area
New Hydrography – HydroSHEDS
(GDBD above 60N)
[%]
100
70
50
30
20
15
10
5
2
−2
−5
−10
−15
−20
−30
−50
−70
−100
New inland basin
Lost inland basin
90
60
30
0
−30
−60
−180
−120
−60
0
60
120
180
(b) CSI: Critical Success Index
0.9
0.8
0.7
0.6
0.5
(c) Basin area comparison
New Hydrography Basin Area [1000 km²]
HydroSHEDS Basin Area [1000 km²]
20
50
100
200
500
1000
2000
5000
10000

Figure 10.



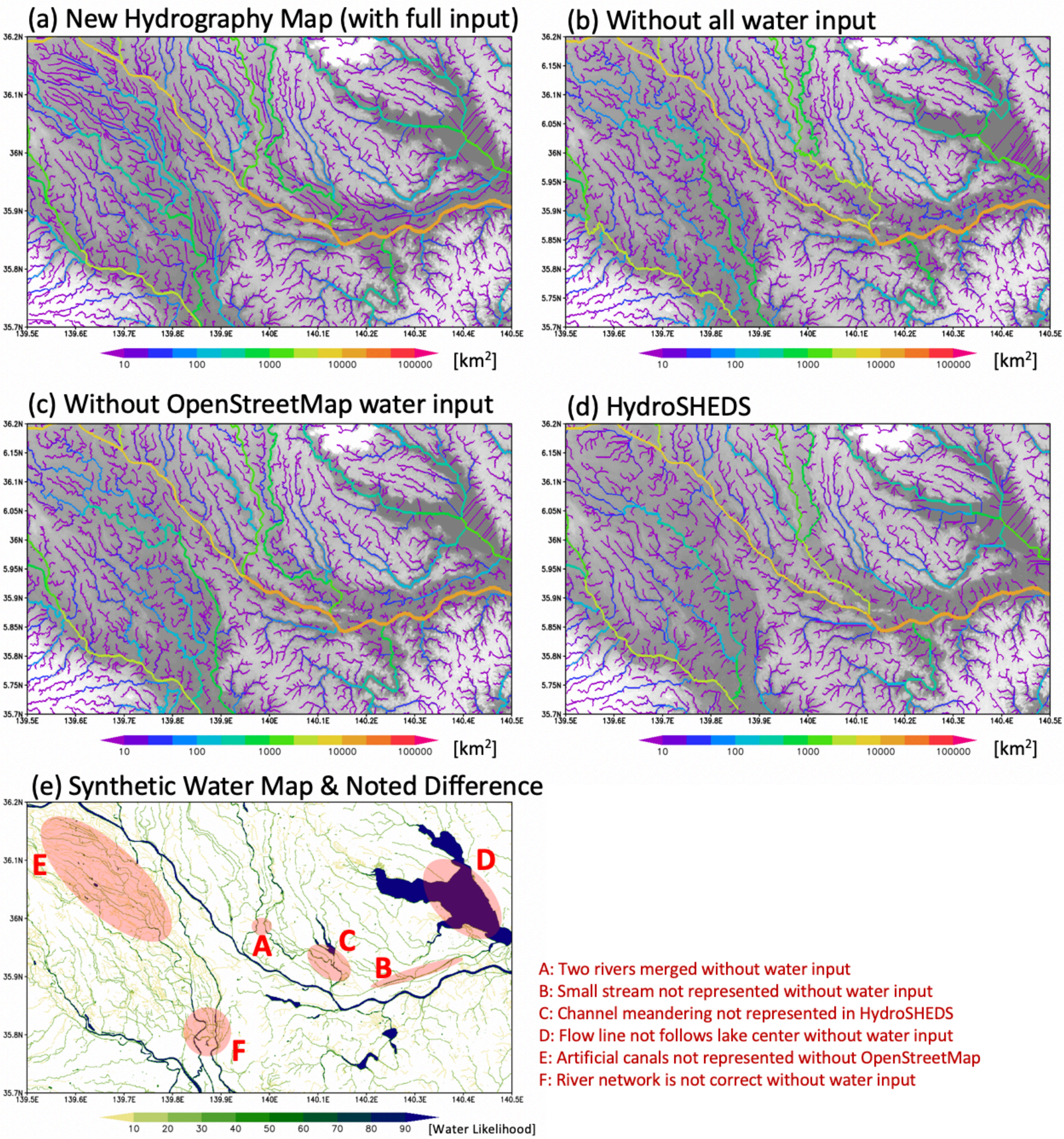
(a) Spatial pattern of relative channel location error
New Hydrography & GRWL
[%]
100
80
60
50
40
30
20
10
0
90
60
30
0
−30
−60
−180
−120
−60
0
60
120
180
(b) CDF for width criteria
W<100
100<W<300
300<W<500
500<W<1000
1000<W<2000
2000<W
Relative channel location error
Percentage
(c) CDF for area criteria
A<1000
1000<A<10000
10000<A<100000
100000<A<1000000
1000000<A
Percentage

Figure 11.



(a) Allocated GRDC gauging stations
Accumulation area error [%]
90
60
30
0
−30
−60
−180
−120
−60
0
60
120
180
100
50
30
20
10
5
−5
−10
−20
−30
−50
−100
(b) Comparison of accumulation area
New Hydrography Accumulation Area [km^2]
GRDC Reported Area [km^2]
10^3
10^4
10^5
10^6

Figure 12.


(a) New Hydrography Map (with full input)
(b) Without all water input
(c) Without OpenStreetMap water input
(d) HydroSHEDS
10 100 1000 10000 100000 [km2]
(e) Synthetic Water Map & Noted Difference
10 20 30 40 50 60 70 80 90 [Water Likelihood]
A: Two rivers merged without water input
B: Small stream not represented without water input
C: Channel meandering not represented in HydroSHEDS
D: Flow line not follows lake center without water input
E: Artificial canals not represented without OpenStreetMap
F: River network is not correct without water input