

## RESEARCH ARTICLE

## Development of the Global Width Database for Large Rivers

10.1002/2013WR014664

Dai Yamazaki<sup>1,2</sup>, Fiachra O'Loughlin<sup>2</sup>, Mark A. Trigg<sup>2</sup>, Zachary F. Miller<sup>3</sup>,  
Tamlin M. Pavelsky<sup>3</sup>, and Paul D. Bates<sup>2</sup>

### Key Points:

- Global Width Database for Large Rivers (GWD-LR) based on satellite water mask
- River width of GWD-LR is calculated along the HydroSHEDS flow direction map
- GWD-LR shows good consistency to existing basin-scale river width data sets

### Supporting Information:

- ReadMe
- Algorithm Description

### Correspondence to:

D. Yamazaki,  
d-yamazaki@jamstec.go.jp

### Citation:

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

Received 10 SEP 2013

Accepted 7 APR 2014

Accepted article online 11 APR 2014

Published online 28 APR 2014

<sup>1</sup>Department of Integrated Climate Change Projection Research, Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan, <sup>2</sup>School of Geographical Sciences, University of Bristol, Bristol, UK, <sup>3</sup>Department of Geological Sciences, University of North Carolina, Chapel Hill, North Carolina, USA

**Abstract** River width is a fundamental parameter of river hydrodynamic simulations, but no global-scale river width database based on observed water bodies has yet been developed. Here we present a new algorithm that automatically calculates river width from satellite-based water masks and flow direction maps. The Global Width Database for Large Rivers (GWD-LR) is developed by applying the algorithm to the SRTM Water Body Database and the HydroSHEDS flow direction map. Both bank-to-bank river width and effective river width excluding islands are calculated for river channels between 60S and 60N. The effective river width of GWD-LR is compared with existing river width databases for the Congo and Mississippi Rivers. The effective river width of the GWD-LR is slightly narrower compared to the existing databases, but the relative difference is within  $\pm 20\%$  for most river channels. As the river width of the GWD-LR is calculated along the river channels of the HydroSHEDS flow direction map, it is relatively straightforward to apply the GWD-LR to global and continental-scale river modeling.

## 1. Introduction

River width is a fundamental parameter for river hydrodynamics. The river discharge is calculated as the product of river width, mean flow depth, and mean flow velocity. Flow conveyance capacity of river channels is mainly decided by river width and channel depth, topographic slope and friction, thus inundation often occurs at narrow segments of river channels where conveyance capacity is lower. Most physically based hydrodynamic flood models, which simulate movement of surface waters along prescribed topography based on mass and momentum conservation equations, use river width as a basic topographic parameter for the calculation of river flow [e.g., Bates and De Roo, 2000; Yamazaki et al., 2012]. The development of a river width database is an essential process for applying hydrodynamic flood models to any river.

In many studies for regional-scale river modeling, the river width parameter is generated by field observations [e.g., Wilson et al., 2007] or manually measuring river cross sections on air photos or satellite images [e.g., Biancamaria et al., 2009]. The manual processing of the river width parameter is acceptable when a target domain is not large and/or when river width does not vary much along the river segment of interest. Some algorithms for automatically calculating river width from water masks have been developed [e.g., Pavelsky et al., 2008; Trigg et al., 2012; O'Loughlin et al., 2013]. These algorithms are very helpful for generating river width parameters for river basins where manual processing of river width parameters is difficult [e.g., Durand et al., 2010a; Neal et al., 2012]. Some existing algorithms for river width calculation (e.g., Riv-Width by Pavelsky et al. [2008]) are close to being fully automated, but even with these a small amount of manual processing, such as correction of the input water mask, is unavoidable. For example, gaps in the water mask due to some obstacles such as bridges need to be manually filled in order to ensure upstream-downstream relationships along river networks. The application of previous algorithms to the global river width calculation is not straightforward because of this.

In the case of global and continental-scale river modeling, river width parameters are commonly generated from empirical equations. For example, it has long been recognized that river width has a power law relationship with river discharge or drainage area [e.g., Leopold and Maddock, 1953; Maidment, 1993; Dingman and Sharma, 1997]. Most global and continental-scale river models have estimated river width by empirical functions of river discharge [e.g., Arora and Boer, 1999; Oki et al., 2001; Decharme et al., 2008; Yamazaki et al., 2011, 2012; Getirana et al., 2012, 2013; Andreadis et al., 2013] or drainage area [e.g., Coe et al., 2008; Paiva

*et al.*, 2013]. However, empirical equations cannot capture the local variability of river channels because geologic and hydraulic controls differ in different regions. Therefore, the river width parameters estimated by empirical equations are not realistic for many river reaches. Furthermore, it is not clear whether an identical empirical equation can be applied at global or continental scales, under varied climate and geomorphological conditions.

The purpose of this study is to generate a Global Width Database for Large Rivers (GWD-LR) that is based on satellite water masks, which will replace the river width parameters of global and continental-scale river models that have previously been estimated by empirical equations. Here we propose a new algorithm that calculates river width at global scale from satellite-based water masks and flow direction maps. In order to create a GWD-LR which is applicable for global-scale hydrodynamic simulations, the proposed algorithm satisfies the following requirements: (1) only globally available databases are used as input data; (2) the entire procedure of river width calculation is automated; (3) consistency between the river width data set and the flow direction map is ensured; and (4) effective river width excluding islands is generated as well as bank-to-bank river width.

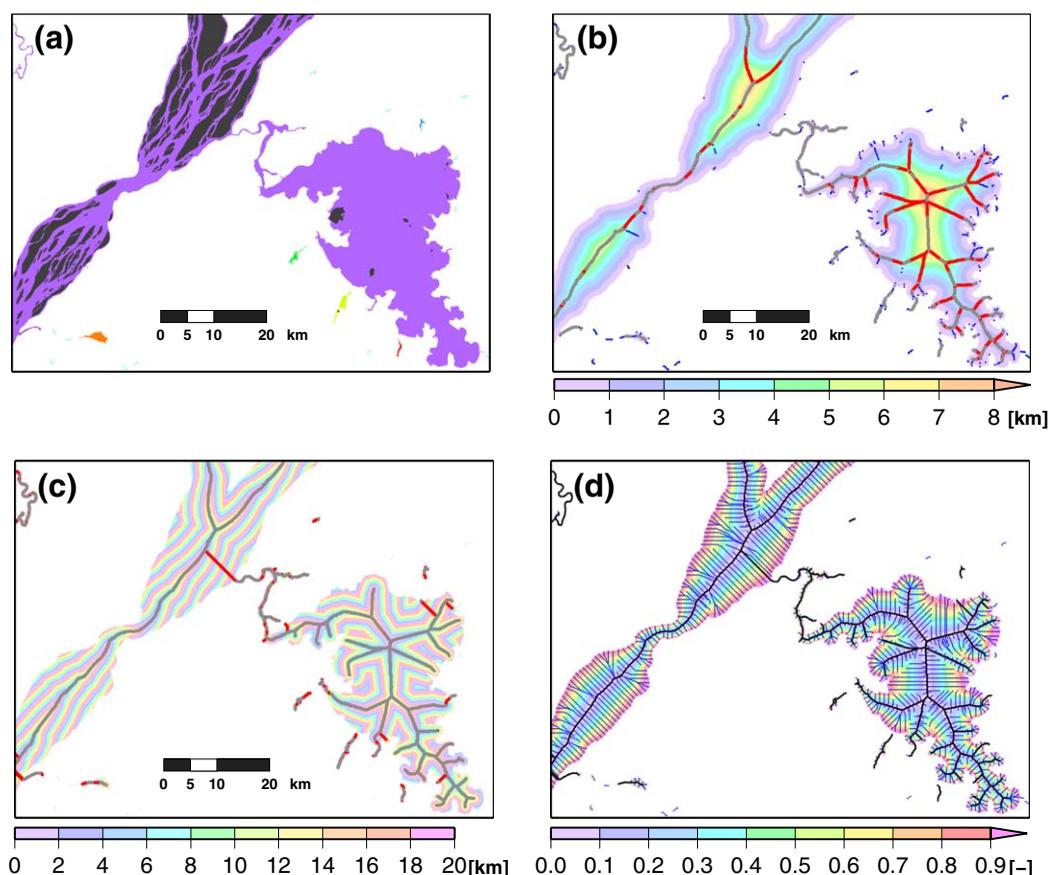
## 2. Baseline Data

We used the SRTM Water Body Data (SWBD) (NASA/NGA, SRTM Water Body Data Product Specific Guidance, Version 2.0, 2003, [http://dds.cr.usgs.gov/srtm/version2\\_1/SWBD/SWBD\\_Documentation/](http://dds.cr.usgs.gov/srtm/version2_1/SWBD/SWBD_Documentation/)) as a baseline water mask in this study. The SWBD can be freely downloaded from the US Geological Survey webpage ([http://dds.cr.usgs.gov/srtm/version2\\_1/SWBD/](http://dds.cr.usgs.gov/srtm/version2_1/SWBD/)). The SWBD is a by-product of the SRTM (Shuttle Radar Topography Mission) digital elevation model, in which ocean, lake, and river shorelines were identified and delineated. River channels whose width is greater than 183 m for a length of 600 m or more are represented in the SWBD water mask. Islands with a medial axis length greater than 300 m are also depicted in SWBD. River channels whose effective river width is narrower than 300 m are not well represented in the SWBD water mask, due to these minimum criteria of water body size, especially in braided sections with many small islands. The original SWBD data were prepared in ArcGIS shape file format at 1 arc sec resolution, and we converted them to raster format at 3 arc sec resolution using the Geospatial Data Abstraction Library (GDAL). We also filled some water bodies missing in the SWBD using the Global Land Cover Facility (GLCF) MODIS Water Mask database [Carroll *et al.*, 2009]. The missing water bodies in the SWBD can occur when there are midstream islands and/or where cloud cover was persistent [Carroll *et al.*, 2009].

We used the HydroSHEDS (Hydrological data and maps based on SHuttle Elevation Derivatives at multiple Scales) [Lehner *et al.*, 2008; Lehner and Grill, 2013] flow direction map, which is a global hydrographic database at 3 arc sec resolution, as a baseline river network database. HydroSHEDS was chosen as the baseline data because it is the only high-resolution flow direction database with a global coverage. The HydroSHEDS flow direction map represents river networks by describing the downstream direction of each pixel toward one of its eight neighboring pixels. The use of a flow direction map in addition to a water mask is the key characteristic of the width calculation algorithm proposed in this study. Upstream-downstream connectivity along river networks is prescribed by a flow direction map, therefore manual correction of the input water mask to ensure connectivity, as required in previous algorithms is not needed. Given that the HydroSHEDS flow direction map was created based on the SRTM DEM and the SWBD, the topographies in the two input data sets are considered to be consistent. The coverage of both the SWBD and HydroSHEDS is between 60N and 60S.

## 3. Method

We developed a new algorithm, which automatically calculates river width at a global scale. The algorithm consists of four steps: (1) filling island gaps in water body mask; (2) determination of river centerlines; (3) generation of flow directions toward centerlines; and (4) calculation of river width. As an example of the global-scale calculation, intermediate outputs of the algorithm for the Congo River basin (1.0S–0.4S, 17.4E–18.2E) are shown in Figure 1. A more detailed description of the algorithm is available online (<http://hydro.iis.u-tokyo.ac.jp/~yamada/GWD-LR/>). Note that trial-and-error calibration of parameters/thresholds is performed globally by checking the calculation results for visibly gross errors in river width.



**Figure 1.** Procedure for river width calculation. (a) Gap filling: gaps in the water mask (dark gray) are filled. Each color represents one “water body unit” which shares the same outlet. (b) Centerline determination: distance from bank pixels (background colors), original centerlines (gray), extended centerlines (red), and spurious centerlines (blue). (c) Centerline connection: distance from outlet pixels (periodic colors), original centerlines (gray), and connected centerlines (red). (d) Determination of flows to centerlines: normalized distance from centerline pixels (background colors), centerlines (black), and flows to centerlines (blue).

### 3.1. Filling Islands Within Water Body Mask

Gaps in the water bodies, including islands and sandbars, are filled in order to determine the outer banks of river channels. Gaps within water masks whose area is smaller than a threshold value (assumed to be 1000 km<sup>2</sup> in this study) were filled (gray pixels in Figure 1a). A relatively large area (1000 km<sup>2</sup>) was selected as the threshold to fill all island gaps except for very large ones (e.g., Ilha do Bananal in the Amazon River), because the main target of GWD-LR development is application to large-scale hydrodynamic models that cannot yet represent channel bifurcation. Water body pixels adjacent to unfilled land pixels are considered as river bank pixels. Flow directions within a gap-filled water body mask are modified in the subsequent two steps (sections 3.2 and 3.3). Note that the modification of flow directions is applied to each “water body unit” (colors in Figure 1a), which is defined as the aggregation of water body pixels sharing the same water body outlet (a water body pixel with flow direction toward a land pixel). Therefore, the modification only alters flow directions locally within each water body unit and not the entire river network structure. The algorithm is applied for all water bodies represented by the SWBD, including lakes and floodplains.

### 3.2. Determination of Centerlines

Distance to the nearest bank pixel (hereafter “bank distance”) is calculated for each pixel (background colors in Figure 1b). A pixel is judged to be a centerline pixel when the following two conditions are satisfied: (1) the bank distance of the considered pixel is longer than that of at least six neighboring pixels; and (2) the maximum gradient of the bank distances between the considered pixel and its neighboring pixel is not larger than the threshold gradient (set to be 0.26 in this study). The gradient of bank distances is calculated by:

$$\Delta D_b / \Delta x = \frac{D_{bj} - D_{bi}}{L}, \quad (1)$$

where  $\Delta D_b / \Delta x$  is the gradient of bank distances,  $D_{bi}$  is the bank distance of the considered pixel  $i$ ,  $D_{bj}$  is the bank distance of the neighboring pixel  $j$ , and  $L$  is the distance between the pixels  $i$  and  $j$ . The first condition was introduced because only the upstream and downstream pixels on a centerline may have a bank distance longer than that of the considered centerline pixel. The second condition was introduced to remove spurious centerlines detected by the first condition. These spurious centerlines are caused by the curvatures of river banks, and tend to extend from a river bank toward a true centerline. Thus, the gradient of bank distances along a spurious centerline tends to be larger than that along a true centerline. The threshold gradient to distinguish true and spurious centerlines was chosen to be 0.26 by trial and error. We found that smaller threshold gradient produces less spurious centerlines, but some true centerlines are not detected when the threshold gradient is too small. The true centerlines determined by these conditions are shown in gray, and rejected spurious centerline pixels are shown in blue in Figure 1b.

The centerlines detected by the above procedure can have large gaps between them where river width is increasing due to the second condition. In order to improve the connectivity along centerlines, centerlines are extended by the following method: (1) for each centerline pixel, the pixel with maximum bank distance among eight neighboring pixels is selected; (2) if the selected neighboring pixel has a larger bank distance than the considered centerline pixel, the selected neighboring pixel is marked as a new centerline pixel. The extended centerline pixels are shown in red in Figure 1b. Remaining upstream-to-downstream discontinuities in the centerline were filled based on the riverline distance between centerline pixels and the outlet of the water body unit (background periodic colors in Figure 1c), see supporting information for additional details.

### 3.3. Generation of Flow Directions Toward Centerlines

The flow directions of noncenterline pixels are decided based on the gradient of the distance from each noncenterline pixel to its nearest centerline pixel (hereafter "centerline distance"). The centerline distance is normalized by the bank distance of the nearest centerline pixel in order to avoid unrealistic accumulation of flow from an area outside of a tributary's width, within the zone where a tributary merges into its main channel. The normalized centerline distance is given by equation (2):

$$D_c = \frac{D_g}{D_{bc}}, \quad (2)$$

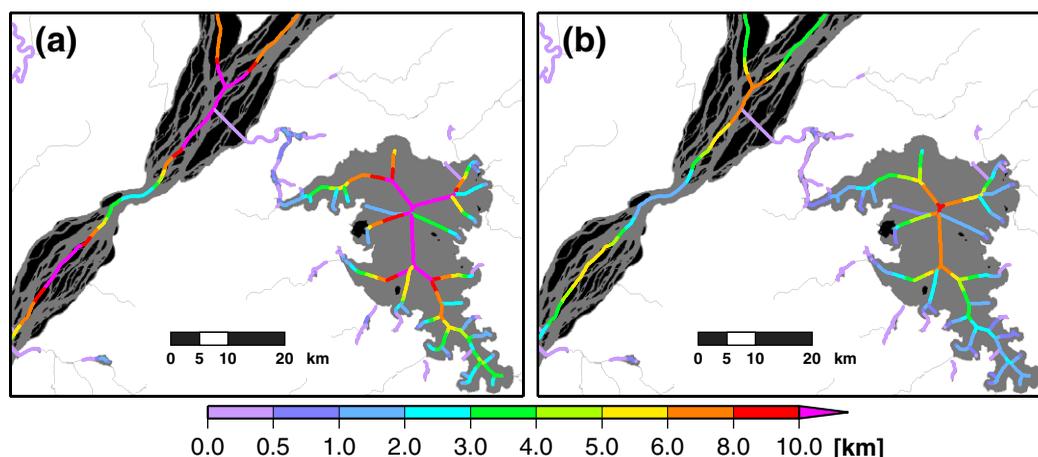
where  $D_c$  is the normalized centerline distance,  $D_g$  is the geometric centerline distance, and  $D_{bc}$  is the bank distance of the nearest centerline pixel. The normalized centerline distance becomes smaller for centerline pixels whose bank distance is larger and vice versa. The normalized centerline distance is shown by background colors in Figure 1d. Then, the modified downstream direction is determined for each noncenterline pixel by choosing the maximum gradient of the normalized centerline distances from the considered pixel toward the eight neighboring pixels (blue lines in Figure 1d). The gradient of normalized centerline distance is calculated by equation (3):

$$\Delta D_c / \Delta x = \frac{D_{cj} - D_{ci}}{L}, \quad (3)$$

where  $\Delta D_c / \Delta x$  is the gradient of normalized centerline distances,  $D_{ci}$  is the normalized centerline distance of the considered pixel  $i$ ,  $D_{cj}$  is the normalized centerline distance of the neighboring pixel  $j$ , and  $L$  is the distance between the two adjacent pixels  $i$  and  $j$ .

### 3.4. Calculation of River Width

The bank-to-bank river width is calculated for each centerline pixel as twice its bank distance. The effective river width excluding islands is calculated by:



**Figure 2.** (a) Bank-to-bank river width and (b) effective river width. Water mask is shown with dark gray, while islands are represented by black. Channels not covered by the SWBD are shown by light gray lines.

$$W_e = W_b \times \left( \frac{A_w}{A_t} \right), \quad (4)$$

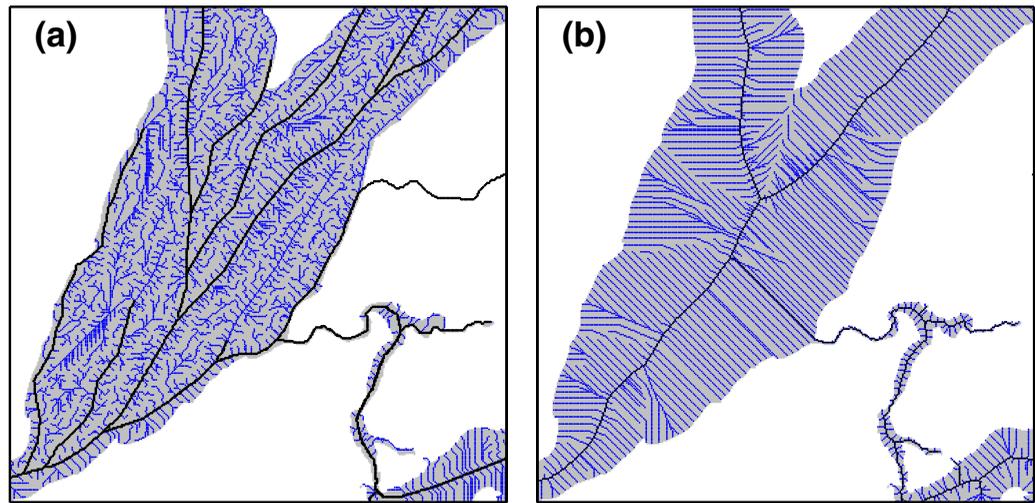
where  $W_e$  is the effective river width,  $W_b$  is the bank-to-bank river width,  $A_t$  is the total in-water area of the effective river segment (i.e., water body and island),  $A_w$  is the water body area of the effective river segment. The effective river segment for a given centerline pixel is defined as; the longitudinal reach within the bank distance length either side of the pixel, and includes the noncenterline pixels draining to that segment of centerline. The water body area is defined by excluding the area of islands from the effective river segment.

#### 4. Results

The Global Width Database for Large Rivers (GWD-LR) was constructed by applying the proposed algorithm to the SWBD water mask and the HydroSHEDS flow direction map. The calculated bank-to-bank river width and effective river width for a domain in the Congo River basin (1.0S–0.4S, 17.4E–18.2E) are shown in Figure 2. The algorithm succeeded in calculating centerlines for both rivers and lakes covered by the SWBD water mask. The continuity of river width data is ensured because all the centerline pixels are connected to their downstream centerline pixel. River width is not calculated for channels narrower than 183 m which are not covered by the SWBD (thin light gray lines in Figure 2), and coverage for rivers narrower than 300 m is spotty. These gaps in river width data have to be filled by another method (such as an empirical equation) when the river width database is applied to river models.

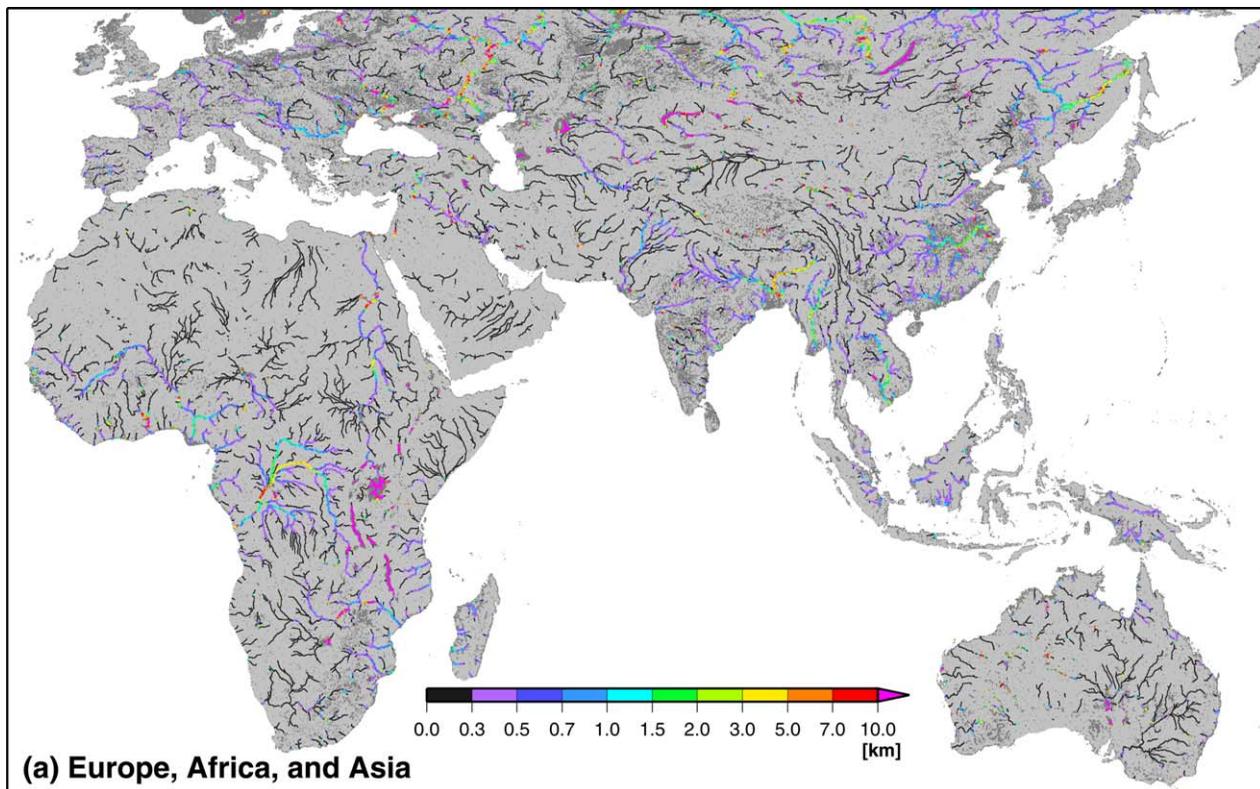
The algorithm reasonably captured local variations of bank-to-bank river width, including the narrow segments around (–0.1S, 17.7E). The calculated bank-to-bank river widths generally agreed with the widths visually estimated from the figure (see the scale bar in Figure 2a). Detailed analysis of calculated river widths is performed in section 5. The effective river widths also show general agreement to the values visually estimated from the water masks (islands are shown by black color in Figure 2b). The effective river widths are smaller than 60% of the bank-to-bank river widths in some segments where there are many islands.

The algorithm also outputs modified flow directions in addition to the river widths. The original flow directions and the modified flow directions are shown in Figure 3. Given that only one downstream direction is given to each pixel in the flow direction map, braided river segments in the original flow direction map are represented by one major channel and its “branches” (black lines in Figure 3a). Therefore, it is difficult to appropriately represent in-bank flow for braided river segments in the original map. In the modified flow direction map, multiple channels are merged into a single “effective” channel for braided river segments (black lines in Figure 3b), which is exactly the input required by a number of recently developed large-scale hydraulic models [e.g., Neal *et al.*, 2012].



**Figure 3.** Flow direction networks. (a) Original flow directions: major streams (black lines) and associated flows (blue lines). (b) Modified flow directions: centerline pixels (black lines) and perpendicular flows to centerlines (blue lines). Islands are shown by green.

The global maps of GWD-LR effective river width are shown in Figure 4. The GWD-LR covers most major rivers in tropical and temperate regions. The SWBD covers river channels wider than 183 m, thus river widths were not calculated for small branches or certain rivers in arid regions (black lines in Figure 4).



**Figure 4.** (a) Effective river width (Europe, Africa, and Asia). Large rivers whose drainage area is larger than  $10,000 \text{ km}^2$  are shown. Dark gray color represents noncenterline water bodies or rivers whose drainage area is smaller than  $10,000 \text{ km}^2$ . (b) Effective river width (North and South America).

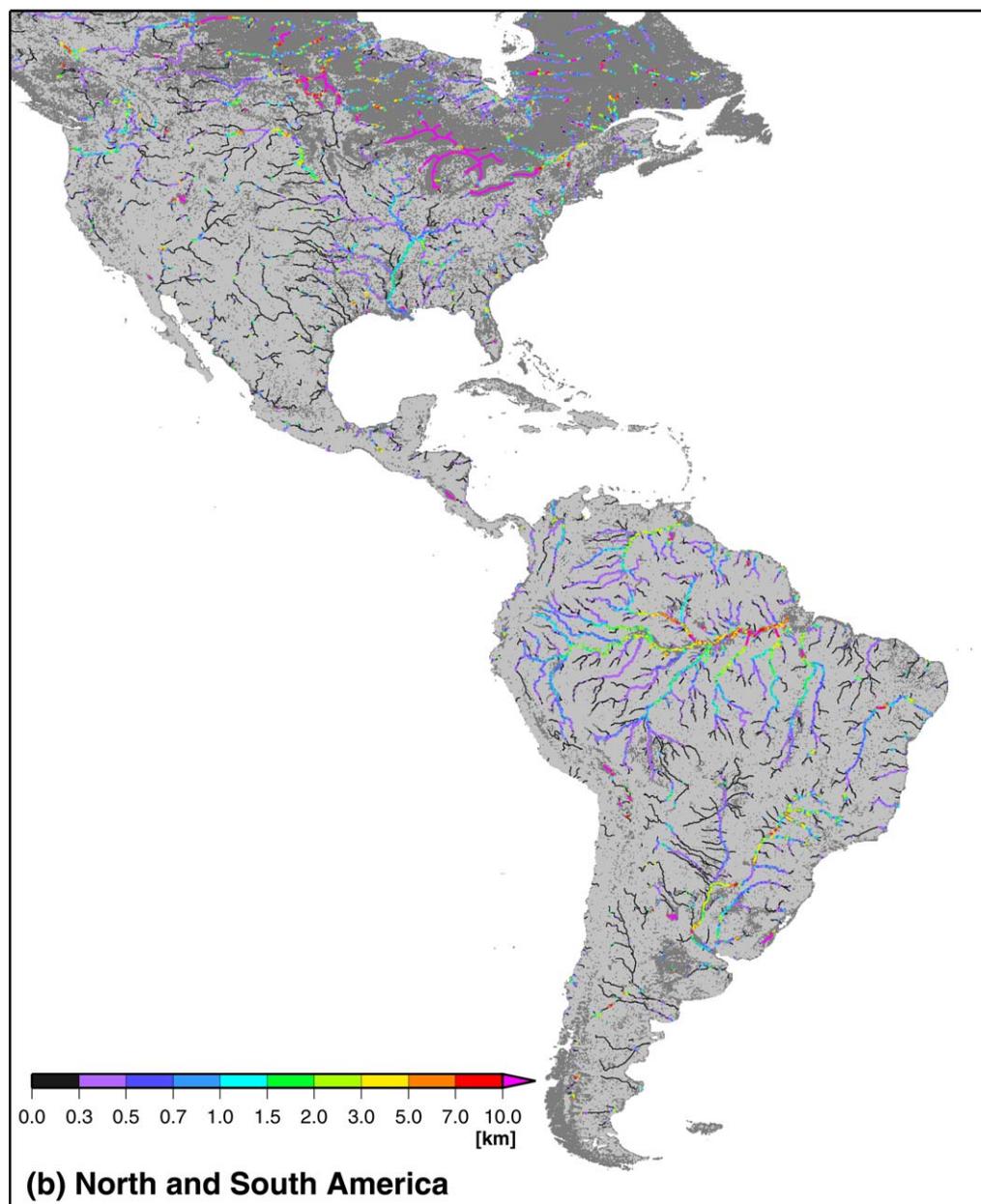


Figure 4. (continued)

Stacked bar graphs of effective river widths for drainage area bins and for annual mean discharge bins are shown in Figure 5. The annual averaged discharge was calculated using the runoff output from the land surface model MATSIRO forced by JRA-25 reanalysis climate forcing [Kim *et al.*, 2009]. Note that river basins which have a drainage area located above 60N are excluded from the data used for Figure 5 because drainage area and annual discharge cannot be calculated for these river basins. The total channel length with GWD-LR coverage for drainage area bins and for annual mean discharge bins are summarized in Table 1(a and b), respectively. Note that channel length is calculated by tracking the HydroSHEDS flow direction map. Coverage of the developed river width database increases for larger drainage areas and larger annual discharges. Only 18% of the total length of river channels with drainage areas between 10,000 and 100,000 km<sup>2</sup> is covered by the GRWD, while GRWD coverage increases to 56% of channel length for river segments with drainage area between 100,000 and 1,000,000 km<sup>2</sup>. The coverage increased to 92% for river channels whose drainage area is larger than 1,000,000 km<sup>2</sup>. A similar increasing trend of GWD-LR coverage appears

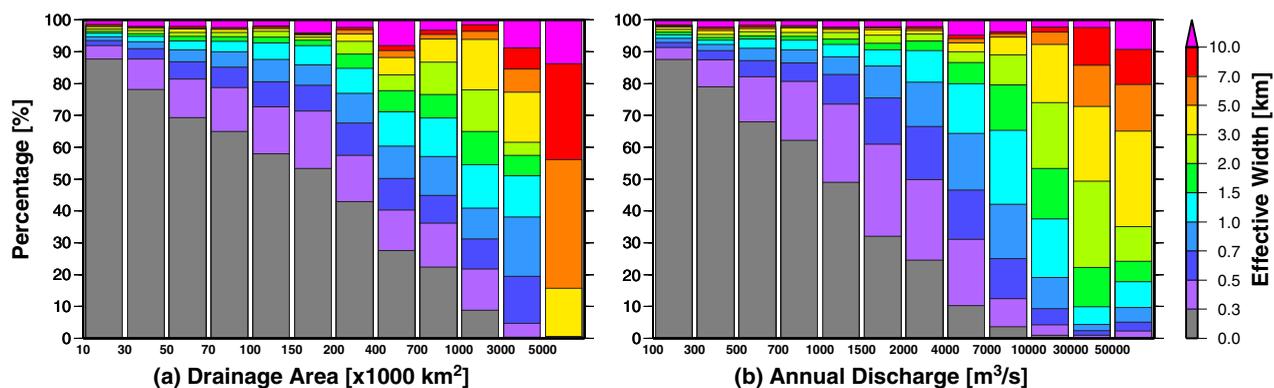


Figure 5. Stacked bar graphs of effective river width: (a) for drainage area bins; (b) for annual mean discharge bins.

for annual mean discharge bins as listed in Table 1(b). Effective river widths become generally wider when drainage area is larger or when annual averaged discharge is larger. However, variability of river widths in each drainage area bin or annual discharge bin is very large. This fact suggests that it is difficult to realistically represent river width by an empirical and monotonically increasing function of drainage area or annual discharge, and is consistent with the large spread in data around logarithmic relationships typically applied [e.g., *Andreadis et al., 2013; Leopold and Maddock, 1953*].

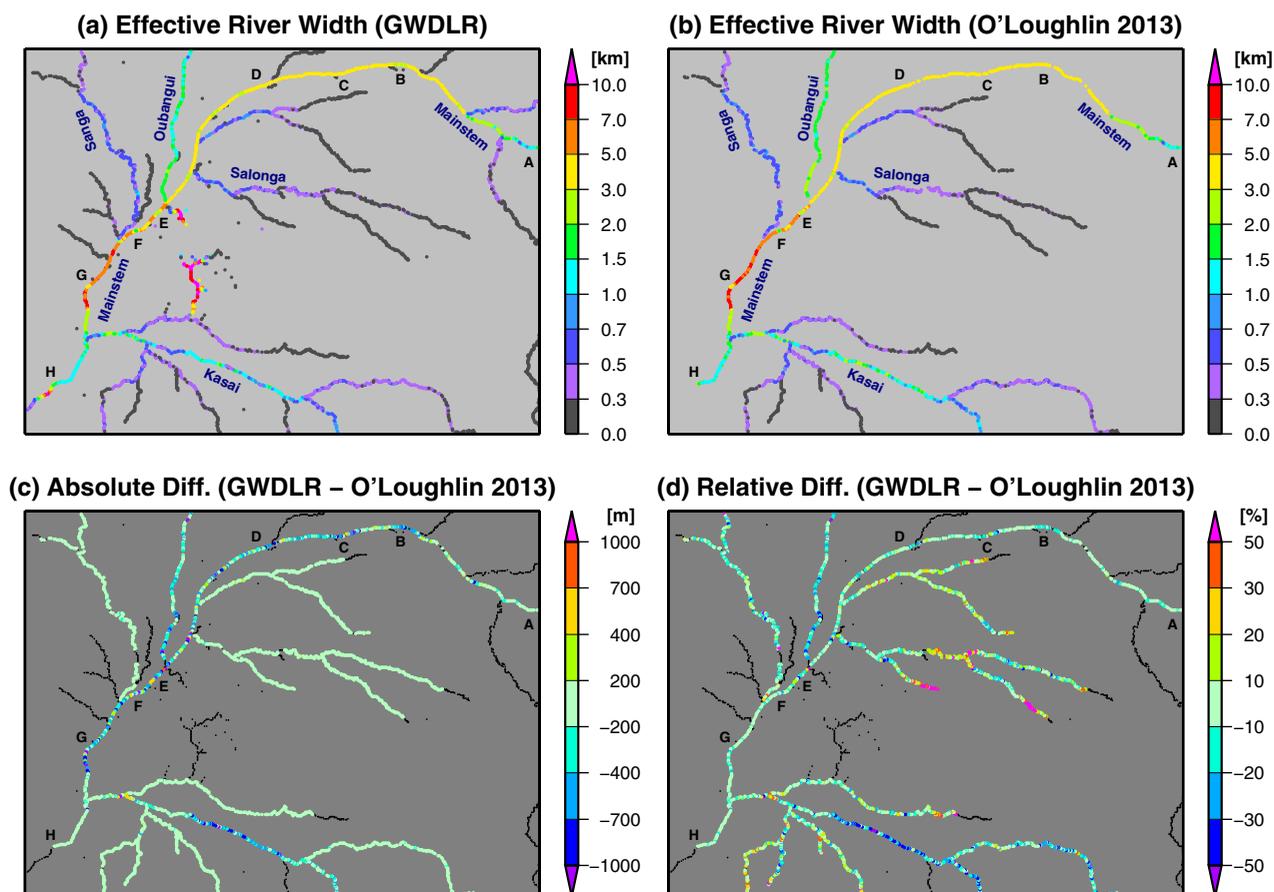
### 5. Comparison to Other Databases

The effective river width of GWD-LR was compared against other river width data sets. We used the river width data sets of the Congo River [*O’Loughlin et al., 2013*] and the Mississippi River [*Miller et al., 2014*]. The river width data for the Mississippi River were calculated by applying the RivWidth algorithm [*Pavelsky et al., 2008*] to the US National Land Cover Database 2001 (NLCD 2001) water masks [*Homer et al., 2007*]. The Riv-Width algorithm calculates bank-to-bank river width by finding a centerline of rivers and then measuring the length of cross sections perpendicular to the centerline. Then, the effective river width is calculated by excluding the length of islands on the cross sections from the bank-to-bank river width [for details, see *Pavelsky et al., 2008*]. The river width of the Congo River [hereafter *O’Loughlin et al., 2013*] was calculated from a LANDSAT water mask using a perpendicular cross-section algorithm similar to RivWidth [for details, see *O’Loughlin et al., 2013; Trigg et al., 2012*]. For the comparison, river widths are aggregated at 0.02° resolution. The aggregated width was given by averaging the values from the 25th to 75th percentiles within each 0.02° grid box.

The effective river widths of GWD-LR and *O’Loughlin et al. [2013]* are shown in Figures 6a and 6b, respectively. The absolute and relative differences between the two data sets are shown in Figures 6c and 6d, respectively. The effective river widths of the two data sets are found to be similar, though the effective river widths of the GWD-LR tend to be smaller. The smaller widths in the GWD-LR can be explained as follows. First, the GWD-LR algorithm tends to underestimate river width because the bank-to-bank river width

**Table 1.** Total Channel Length and Channel Length Width GWD-LR Coverage: (a) for Drainage Area Bins; and (b) for Annual Mean Discharge Bins

(a) Drainage area ( $\times 1000 \text{ km}^2$ )	10–100	100–1000	>1000
Total length (km)	564,590	139,875	18,119
GWD-LR coverage (km)	100,139	77,641	16,696
Percentage	17.7%	55.5%	92.2%
(b) Annual discharge ( $\text{m}^3/\text{s}$ )	100–1000	1000–10,000	>10,000
Total length (km)	432,190	125,681	20,292
GWD-LR coverage (km)	80,483	88,927	20,125
Percentage	18.6%	70.8%	99.2%

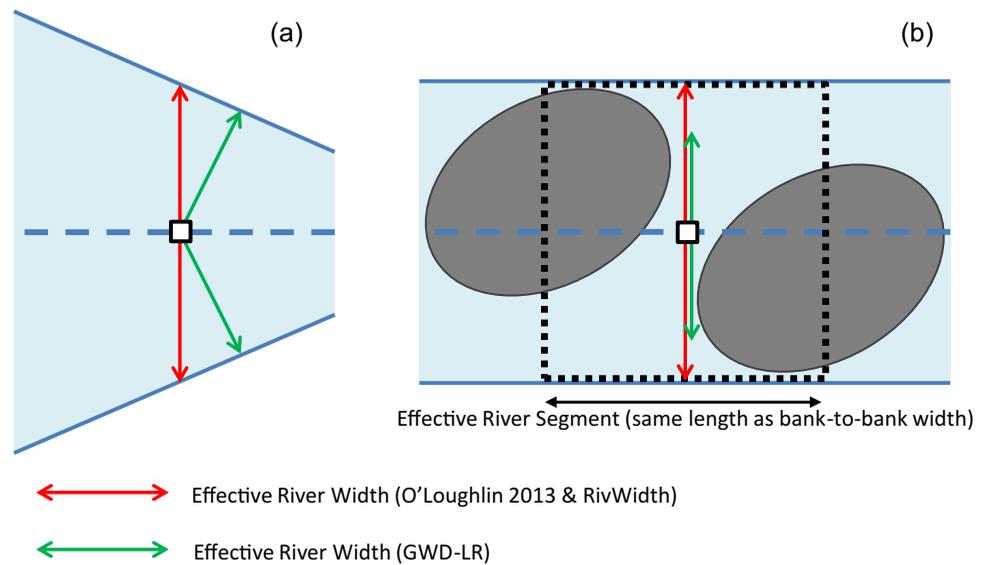


**Figure 6.** Effective river widths of the Congo River basin (15E–25E, 2.5N–5S). (a) Effective river width of GWD-LR, (b) effective river width by *O'Loughlin et al.* [2013], (c) difference, and (d) relative difference between the two databases. The letters A–H along the main stem corresponds to the points in Figure 8.

is calculated as twice the distance to the nearest bank, and the line connecting a centerline pixel and its nearest bank pixel is not necessarily perpendicular to the river centerline, such as in the case where bank-to-bank river width varies longitudinally (as illustrated in Figure 7a). Second, the algorithm of *O'Loughlin et al.* [2013] tends to overestimate effective river width when braided channels are not parallel to the calculated centerline (Figure 7b). The effective river widths of the GWD-LR algorithm are calculated using equation (2), considering the ratio of water area to the total area in the effective river segment (dotted square in Figure 7b). Thus, the effective river widths of GWD-LR tend to be smaller in bifurcated sections when compared to *O'Loughlin et al.* [2013].

The longitudinal variation of the effective river widths along the Congo main stem (between symbols A and H in Figure 6) is shown in Figure 8. The Congo River is known to have some very narrow sections (indicated by symbols E, F, and G) [*O'Loughlin et al.*, 2013]. It can be said from Figure 8 that both data sets successfully capture the narrow segments of the main stem. The large-scale bifurcation of the channel is another characteristic of the Congo River. The Congo main stem has some large islands, which cause channel bifurcation for long reaches (e.g., the symbols B, C, and D in Figures 6a, 6b, and 8). The effective widths of these sections with large islands are generally well captured but slightly underestimated. The reason for the underestimation within the bifurcated sections are: (1) the islands between the bifurcated channel are not recognized as a gap because small channels in a bifurcated section are not connected in the water mask data (Figure 9); and (2) *O'Loughlin et al.* [2013] created a water mask based on high flows while SWBD is a snap shot at a particular observation period (February 2000).

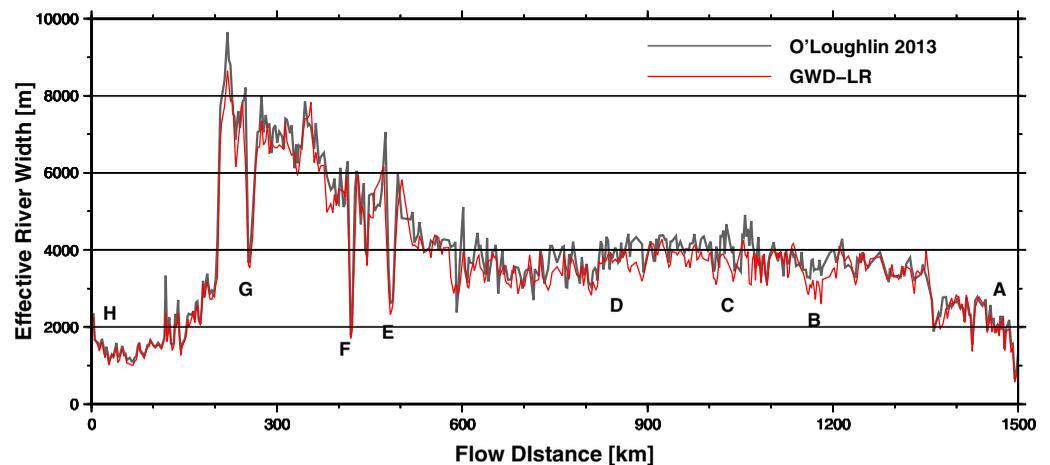
The scatter plot between the effective river width and the relative difference is shown in Figure 10a. It was found that the relative difference between the two data sets is less than 20% for 82% of the river channels,



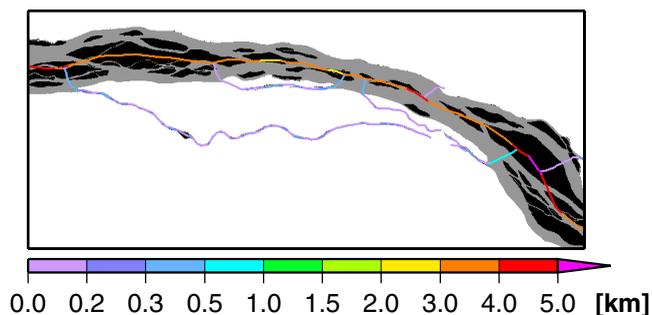
**Figure 7.** Schematic illustrations of the possible difference in width calculation. (a) A pattern when the GWD-LR algorithm (green vector) underestimates river width compared to the algorithms of *O'Loughlin et al.* [2013] and RivWidth (red vector). (b) A pattern when the algorithms of *O'Loughlin et al.* [2013] and RivWidth overestimate effective river width (red vector). The effective river width of the GWD-LR algorithm (green vector) is estimated by equation (4) considering the ratio of water body area to the area of effective river segment (dotted square). Note that gray circles in Figure 7b represent islands.

and it is less than 30% for 93% of the river channels. The large negative relative differences for river widths between 1000 and 2000 m in Figure 10a correspond to the underestimated sections in one of the main Congo tributaries: the Kasai River (Figure 6). The river channel of the Kasai River is braided and meandering, thus it is difficult to determine river widths with both algorithms used in GWD-LR and *O'Loughlin et al.* [2013] as explained in Figure 7. Both positive and negative relative differences become large where the effective river width is smaller than 700 m. These large differences are mainly caused by the limitation of pixel size. Given that the GWD-LR algorithm is based on the 90 m resolution pixels, it is difficult to capture the width variations in these small channels. However, even though the spread of the relative difference becomes larger for effective river width smaller than 700 m, about 80% of these channels still have a relative difference smaller than 20%.

In order to distinguish the differences between the algorithms and water mask data sets, the algorithm proposed in this study was applied to the same LANDSAT water mask used in *O'Loughlin et al.* [2013]. The



**Figure 8.** Comparison of effective river width along the Congo main stem. The letters A–H corresponds to the locations in Figure 6.



**Figure 9.** Example of underestimation of river width. The effective river width is shown by the color range. The region around Symbol B in Figure 6 (20.0–20.7E, 1.95–2.25N) is illustrated. The SWBD water mask is represented by gray, while island gaps are shown by black.

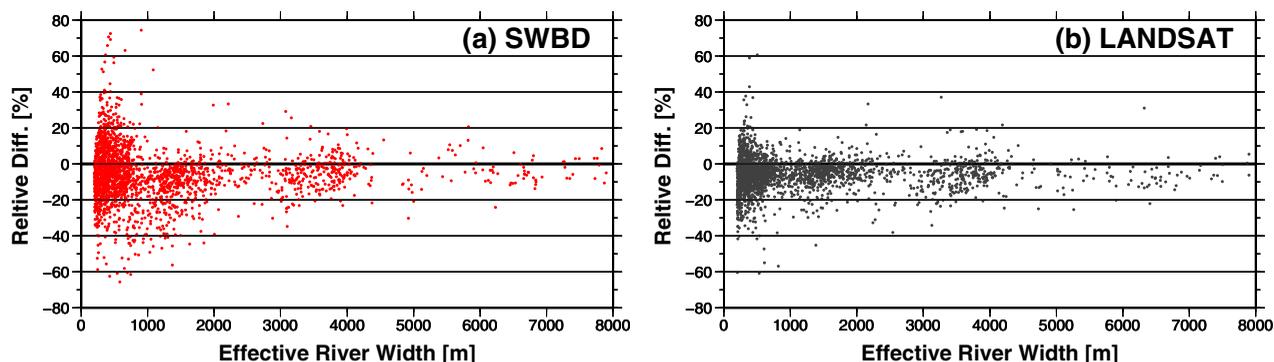
relative difference in effective river width between the GWD-LR algorithm using the LANDSAT water mask and *O’Loughlin et al.* [2013] is shown in Figure 10b. Effective river widths by the GWD-LR algorithm are generally smaller than those estimated by the algorithm of *O’Loughlin et al.* [2013], even when the same water mask is used. However, the relative differences are smaller than 20% for 93% of the river channels when the GWD-LR algorithm is applied to the LANDSAT water mask. Relative differences larger than 40% are found only in

0.5% of the river channels. Thus, the differences in the water masks between SWBD and LANDSAT are considered to be the main cause of the large relative differences (>40%) in Figure 10a.

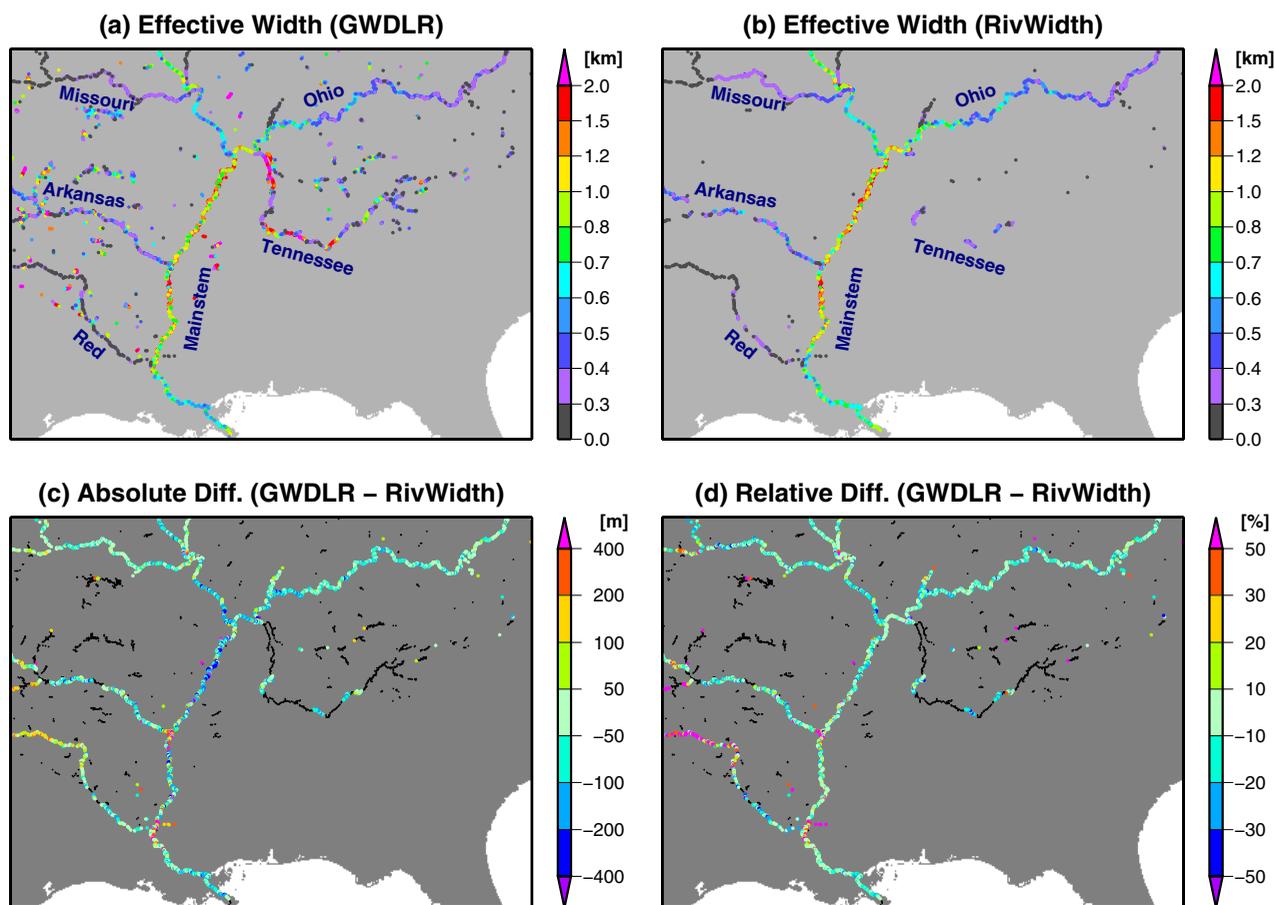
For the Mississippi River, the effective river widths of GWD-LR are compared to the effective river widths estimated by RivWidth in Figure 11. Note that the RivWidth product does not include in-channel lakes and reservoirs, so the effective river width is not calculated in some parts of the Tennessee River (Figure 11b). The effective river width of GWD-LR generally agrees with that of RivWidth, even though different water masks were used for the two databases. The absolute and relative differences of the two databases are smaller than 100 m and 20%, respectively, for most of river segments. The effective river width is estimated to be smaller in GWD-LR than in RivWidth for the middle reach of the Mississippi main stem. The river channels of the underestimated sections are braided and meandering with frequently changing width, thus the underestimation by GWD-LR is probably due to the differences explained in Figures 7a and 7b. The river width is overestimated in the GWD-LR by more than 50% in the upstream reach of the Red River. This overestimation is caused because islands in channels narrower than 300 m are represented in the NLCD but not in the SWBD. For the 470 km reach of the Red River, the SWBD data include only 87 bifurcated sections with more than one channel, while the NLCD data include 1163 multichannel measurements. Thus, the differences between the SWBD and NLCD water masks are considered to be the major cause of this large difference in effective river width.

## 6. Discussion and Conclusions

We developed a new algorithm to calculate river width from water masks and flow direction maps. The algorithm is fully automated, so that it can be applied at global-scale without manual intervention. The



**Figure 10.** Scatter plot between effective river width and relative difference. (a) Relative difference between GWD-LR and *O’Loughlin et al.* [2013]. (b) Relative difference between the LANDSAT-based GWD-LR algorithm and *O’Loughlin et al.* [2013].



**Figure 11.** Effective river widths of the Mississippi River basin (96W–80W, 40N–29N). (a) Effective river width of GWD-LR, (b) effective river width by RivWidth, (c) absolute difference, and (d) relative difference between two databases.

Global Width Database for Large Rivers (GWD-LR) was constructed by applying the developed algorithm to the SRTM Water Body Database (SWBD) and the HydroSHEDS flow direction map. Both bank-to-bank river width and effective river width excluding islands are calculated. The GWD-LR represents the longitudinal variability of river width, which cannot be represented by the monotonic empirical equations currently used in global- and continental-scale river modeling.

The river width of the GWD-LR was compared against existing river width data sets for the Congo and Mississippi Rivers. The GWD-LR river width generally agreed with the two river width databases, but the GWD-LR algorithm was found to have slightly smaller river widths. The relative difference between the GWD-LR and other two data sets were smaller than 20% for the majority of river segments, except for sections where channels are braided and meandering and where channel width is smaller than 300 m. It was also found that the proposed algorithm cannot capture the effective river width correctly when two channels in a bifurcated river sections are treated separately due to the lack of connectivity in the water mask data. This limitation in bifurcated river sections is caused because only one downstream direction is given to each pixel in the HydroSHEDS flow direction map.

The GWD-LR provides only one effective river width in a bifurcated section with multiple channels by taking a relatively large threshold value for island filling (i.e., 1000 km<sup>2</sup>, see section 3.1). The relatively large threshold value is adopted because the GWD-LR was mainly developed for application to large-scale river modeling. Since many large-scale river models do not represent channel bifurcation [e.g., Yamazaki et al., 2009; Coe et al., 2008; Paiva et al., 2013], widths of bifurcated channels should be

amalgamated into one effective river width in order to execute realistic hydrodynamic simulations. However, given that the proposed algorithm is fully automated, it is easy to generate a global width database using a smaller threshold value for island filling when river width should be calculated separately for each channel in a bifurcated section [e.g., analysis on channel geomorphology, *Kleinhaus and van den Berg*, 2011].

The coverage of data sets is currently limited between 60N and 60S, following the availability of the SWBD and HydroSHEDS. However, the flow direction maps of HydroSHEDS will be extended to boreal river basins in the near future [*Lehner and Grill*, 2013] and water mask databases other than the SWBD are available (e.g., products from LANDSAT and MODIS). Hence, the calculation of river widths in boreal rivers will be possible in near future. In addition, narrow river channels (<300 m) are not well represented in the SWBD, thus river width is not calculated for small rivers in the GWD-LR. The use of higher resolution data (e.g., LANDSAT or InSAR-based water masks) would be useful in extending the coverage to small rivers. The future satellite mission SWOT (Surface Water and Ocean Topography) [*Durand et al.*, 2010b] will also provide high-resolution water extent as well as water level data.

Temporal variability of river width is not represented because only one water mask (i.e., the SWBD) was used in the calculation of river width in this study. The SRTM measurement was performed in February 2000, so that GWD-LR should be considered as a snapshot database. Both bank-to-bank river widths and effective river widths may vary in time in some rivers, thus it may not be appropriate to assume that river width is constant for river hydrodynamic modeling. Thus, calculation of temporal variation of river widths is a potential future research topic. The future satellite mission SWOT will provide frequent observations (at least twice in 21 day orbit cycle) with water level and water extent data, which will be useful for calculating the temporal variability of river width.

It would be interesting to see the differences between hydrodynamic simulations using the river width parameters from the GWD-LR and from empirical equations. Hydrodynamic flood models simulate discharge and water level using a nonlinear function of the river width parameter (i.e., mass and momentum conservation equations), thus it is obviously better to have a channel width parameter with realistic longitudinal variation. However, the comparison is not straightforward because river hydrodynamics is also affected by other channel parameters (e.g., channel depth and roughness). Given that the channel depth and roughness parameters have been estimated empirically in many global and continental-scale river models [e.g., *Yamazaki et al.*, 2011; *Coe et al.*, 2008; *Paiva et al.*, 2013], it may be difficult to attribute simulated errors to a specific channel parameter because different combinations of channel parameters may produce similar river discharges. On the other hand, the GWD-LR may be useful for the estimation of channel depth parameters. Once the river width is known, the channel depth may be estimated by comparing the simulated and observed river discharges, water levels, and wave propagation speeds [e.g., *Durand et al.*, 2010a].

The consistency between the river width data and the flow direction map is an advantage of the GWD-LR. The application of the GWD-LR to large-scale river modeling is relatively straightforward because most large-scale river modeling utilizes the HydroSHEDS flow direction map to derive the river network map and the associated topographic parameters [e.g., *Yamazaki et al.*, 2011; *Paiva et al.*, 2013; *Lehner and Grill*, 2013]. The representation of realistic river width is expected to increase the accuracy of simulated flow depth and flow velocity, which is critical for the simulation of sediment transportation as well as studies on aquatic environments and ecosystems.

In addition to application into hydrodynamic modeling, the GWD-LR may be useful to answer some fundamental questions on channel geomorphology. Given that the GWD-LR is the first river width database with a global coverage, a future study will be an analysis of the global distribution of channel widths. By combining it with other geomorphological databases (e.g., discharge, elevation, sediments, lithology), it may be possible to analyze the controlling factors of channel width. A global-scale study on the difference in controlling factors of channel width in different conditions (e.g., continents, climate) will be an interesting topic.

The authors would be pleased to share the GWD-LR for further applications in hydrology, hydrodynamics, and related research fields. The GWD-LR is available online at <http://hydro.iis.u-tokyo.ac.jp/~yamada/GWD-LR/>.

## Acknowledgments

The first author was dispatched to University of Bristol by "JSPS Postdoctoral Fellowship for Research Abroad." The authors would like to thank Japan Society for Promotion of Science (JSPS) and University of Bristol for supporting this collaborative research.

## References

- Andreadis, K., G. Schumann, and T. Pavelsky (2013), A simple global river bankfull width and depth database, *Water Resour. Res.*, *49*, 7164–7168, doi:10.1002/wrcr.20440.
- Arora, V. K., and G. J. Boer (1999), A variable velocity flow routing algorithm for GCMs, *J. Geophys. Res.*, *104*(D24), 30,965–30,979.
- Bates, P. D., and A. P. De Roo (2000), A simple raster-based model for flood inundation simulation, *J. Hydrol.*, *236*, 54–77, doi:10.1016/S0022-1694(00)00278-X.
- Biancamaria, S., P. D. Bates, A. Boone, and N. M. Mognard (2009), Large-scale coupled hydrologic and hydraulic modeling of the Ob river in Siberia, *J. Hydrol.*, *379*, 136–150, doi:10.1016/j.jhydrol.2009.09.054.
- Carroll, M. L., J. R. Townshend, C. M. Di Miceli, P. Noojipady, and R. A. Sohlberg (2009), A new global raster water mask at 250 m resolution, *Int. J. Digital Earth*, *2*(4), 291–308, doi:10.1080/17538940902951401.
- Coe, M. T., M. H. Costa, and E. A. Howard (2008), Simulating the surface waters of the Amazon River basin: Impact of new river geomorphic and flow parameterizations, *Hydrol. Processes*, *22*(14), 2542–2553, doi:10.1002/hyp.6850.
- Dingman, S. L., and K. P. Sharma (1997), Statistical development and validation of discharge equations for natural channels, *J. Hydrol.*, *199*, 13–35.
- Decharme, B., H. Douville, C. Prigent, F. Papa, F. Aires (2008), A new river scheme for global climate applications: Off-line evaluation over South America, *J. Geophys. Res.*, *113*, D11110, doi:10.1029/2007JD009376.
- Durand, M., E. Rodrigues, D. E. Alsdorf, and M. Trigg (2010a), Estimating river depth from remote sensing swath interferometry measurements of river height, slope, and width, *IEEE Geosci. Remote Sens. Lett.*, *3*(1), 20–31, doi:10.1109/JSTARS.2009.2033453.
- Durand, M., L.-L. Fu, D. P. Lettenmaier, D. E. Alsdorf, E. Rodrigues, and D. Esteban-Fernandez (2010b), The Surface Water and Ocean Topography mission: Observing terrestrial surface water and oceanic submesoscale eddies, *Proc. IEEE*, *98*(5), 766–779.
- Getirana, A. C. V., A. Boone, D. Yamazaki, B. Decharme, F. Papa, and N. Mognard (2012), The Hydrological Modeling and Analysis Platform (HyMAP): Evaluation in the Amazon basin, *J. Hydrometeorol.*, *13*, 1641–1665, doi:10.1175/JHM-D-12-021.1.
- Getirana, A. C. V., A. Boone, D. Yamazaki, and N. Mognard (2013), Automatic parameterization of a flow routing scheme driven by radar altimetry data: Evaluation in the Amazon basin, *Water Resour. Res.*, *49*, 614–629, doi:10.1002/wrcr.20077.
- Homer, C., J. Dewitz, J. Fry, M. Coan, N. Hossain, C. Larson, N. Herold, A. McKerrow, J. N. VanDriel, and J. Wickham (2007), Completion of the 2001 National Land Cover Database for the Conterminous United States, *Photogramm. Eng. Remote Sens.*, *73*(4), 337–341.
- Kim, H., P. J.-F. Yeh, T. Oki, and S. Kanae (2009), Role of rivers in the seasonal variations of terrestrial water storage over global basins, *Geophys. Res. Lett.*, *36*, L17402, doi:10.1029/2009GL039006.
- Kleinans, M. G., and J. H. van den Berg (2011), River channel and bar patterns explained and predicted by an empirical and a physics-based method, *Earth Surf. Processes Landforms*, *36*, 721–738, doi:10.1002/esp.2090.
- Lehner, B., and G. Grill (2013), Global river hydrography and network routing: Baseline data and new approaches to study the world's large river systems, *Hydrol. Processes*, *27*, 2171–2186, doi:10.1002/hyp.9740.
- Lehner, B., K. Verdin, and A. Jarvis (2008), New global hydrography derived from spaceborne elevation data, *Eos Trans. AGU*, *89*(10), doi:10.1029/2008EO100001.
- Leopold, L., and T. Maddock (1953), The hydraulic geometry of stream channels and some physiographic implications, *U.S. Geol. Surv. Prof. Pap.* 252., U.S. Gov. Print. Off., Washington, D. C.
- Maidment, D. (1993), *Handbook of Hydrology*, McGraw-Hill, New York.
- Miller, F. M., T. M. Pavelsky, and G. H. Allen (2014), Quantifying river form variations in the Mississippi Basin using remotely sensed imagery, *Hydrol. Earth Syst. Sci. Discuss.*, *11*, 3599–3636.
- Neal, J., G. J.-P. Schumann, and P. D. Bates (2012), A simple model for simulating river hydraulics and floodplain inundation over large and data sparse areas, *Water Resources Research*, *48*, W11506, doi:10.1029/2012WR012514.
- Oki, T., Y. Agata, S. Kanae, T. Saruhashi, D. Yang, and K. Musiaka (2001), Global assessment of current water resources using Total Runoff Integrating Pathways, *Hydrol. Sci. J.*, *46*(6), 983–995.
- O'Loughlin, F., M. A., Trigg, G. J.-P. Schumann, and P. D. Bates (2013), Hydraulic characterization of the middle reach of the Congo River, *Water Resour. Res.*, *49*, 5059–5070, doi:10.1002/wrcr.20398.
- Paiva, R. C. D., W. Collischonn, M.-P. Bonnet, D. C. Buarque, F. Frappart, S. Calmant, and C. B. Mendes (2013), Large scale hydrologic and hydrodynamic modeling of the Amazon River basin, *Water Resour. Res.*, *49*, 1226–1243, doi:10.1002/wrcr.20067.
- Pavelsky, T. M., and L. C. Smith (2008), RivWidth: A software tool for the calculation of river widths from remotely sensed imagery, *IEEE Geosci. Remote Sens. Lett.*, *5*(1), 70–73, doi:10.1109/lgrs.2007.908305.
- Trigg, M. A., P. D. Bates, M. D. Wilson, G. Schumann, and C. Baugh (2012), Floodplain channel morphology and networks of the middle Amazon River, *Water Resour. Res.*, *48*, W10504, doi:10.1029/2012WR011888.
- Wilson, M. D., P. D. Bates, D. Alsdorf, B. Forsberg, M. Horritt, J. Melack, F. Frappart, and J. Famiglietti (2007), Modeling large-scale inundation of Amazonian seasonally flooded wetlands, *Geophys. Res. Lett.*, *34*, L15404, doi:10.1029/2007GL030156.
- Yamazaki, D., T. Oki, and S. Kanae (2009), Deriving a global river network map and its sub-grid topographic characteristics from a fine-resolution flow direction map, *Hydrol. Earth Syst. Sci.*, *13*, 2241–2251, doi:10.5194/hess-13-2241-2009.
- Yamazaki, D., S. Kanae, H. Kim, and T. Oki (2011), A physically-based description of floodplain inundation dynamics in a global river routing model, *Water Resour. Res.*, *47*, W04501, doi:10.1029/2010WR009726.
- Yamazaki, D., H. Lee, D. E. Alsdorf, E. Dutra, H. Kim, S. Kanae, and T. Oki (2012), Analysis of the water level dynamics simulated by a global river model: A case study in the Amazon River, *Water Resour. Res.*, *48*, W09508, doi:10.1029/2012WR011869.