The Global River Width Algorithm

Dai Yamazaki
School of Geographical Sciences, University of Bristol
Dai.Yamazaki@bristol.ac.uk

Note:

This document describes the Global River Width Algorithm (GRW Algorithm), which was used to develop the Global Width Database for Large Rivers (GWD-LR) [Yamazaki et al., in 2014]. The Fortran90 code of the GRWA with sample dataset is available on request to the developer (Dai Yamazaki).

1. Input and Output Datasets

1.1 Input Datasets

The GRW algorithm requires three input datasets: (1) a water body mask; (2) a flow direction map; and (3) a drainage area map. These three input datasets should be prepared in a consistent grid coordination system and at the same spatial resolution. The SRTM Water Body Data (SWBD) [NASA/NGA, 2003] and HydroSHEDS flow direction map [Lerner et al., 2013] are used as the input datasets to develop GWD-LR.

The water body mask describes whether each pixel is a land pixel (value: 0) or a water body pixel (value: 1). Types of water body (i.e. river, lake, ocean) do not have to be distinguished, and all water bodies should be indicated by the value 1 (Figure 1.1).

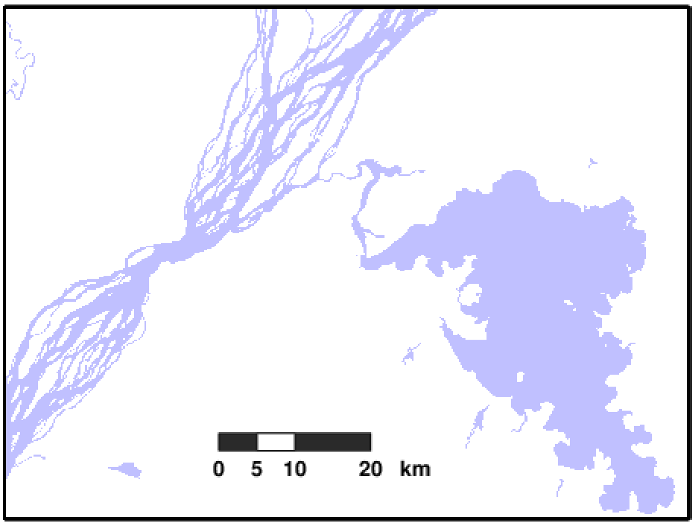

Figure 1.1 Input water body mask. Blue: water body pixels, white: land pixels. A part of the Congo River (17.4-18.2E, 0.4-1.0S) is shown as an example.

The flow direction map describes the downstream direction of surface water flow at each pixel toward one of the eight neighboring pixels (1: north, 2: northeast, 3: east, 4: southeast, 5: south, 6: southwest, 7: west, 8: northwest). The river mouth is indicated by the value 0, while ocean pixels are represented by the value -9.

The drainage area map (or flow accumulation map) describes the accumulative drainage area of each pixel. The drainage are map can be calculated from the flow direction map.

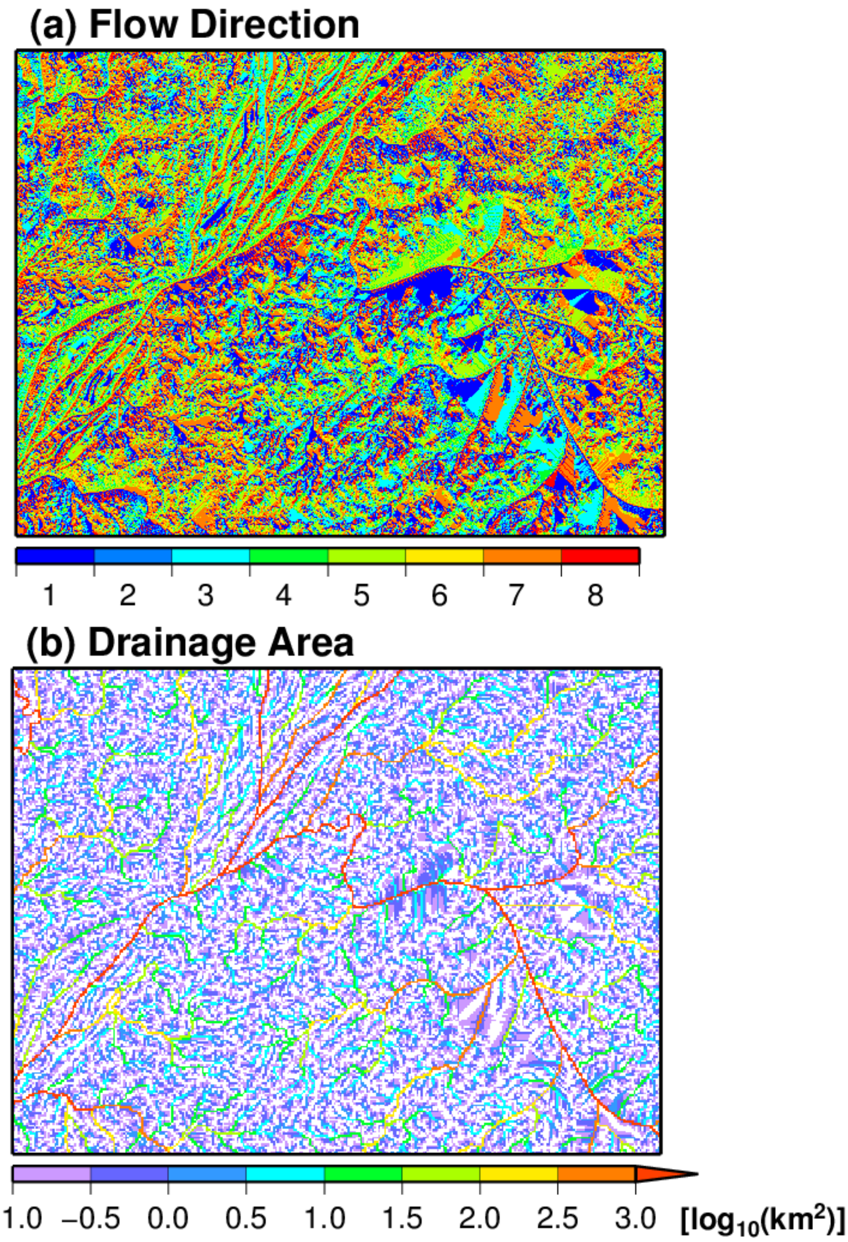

Figure 1.2: (a) Input flow direction map (HydroSHEDS). Colors represent flow directions (1: north, 2: northeast, 3: east, 4: southeast, 5: south, 6: southwest, 7: west, 8: northwest). (b) Input drainage area map.

1.2 Output Data

The algorithm automatically calculates bank-to-bank river width (Figure 1.3a) and effective river width excluding islands (Figure 1.3b) for all water bodies in the calculation domain. The procedures of the algorithm are explained in Section 2. The modified flow direction map is also outputted as a by-product (Figure 1.4b). The river width is calculated along the modified flow direction map, thus it is straightforward to use the river width database as a topographic parameter of large-scale hydrodynamic models.

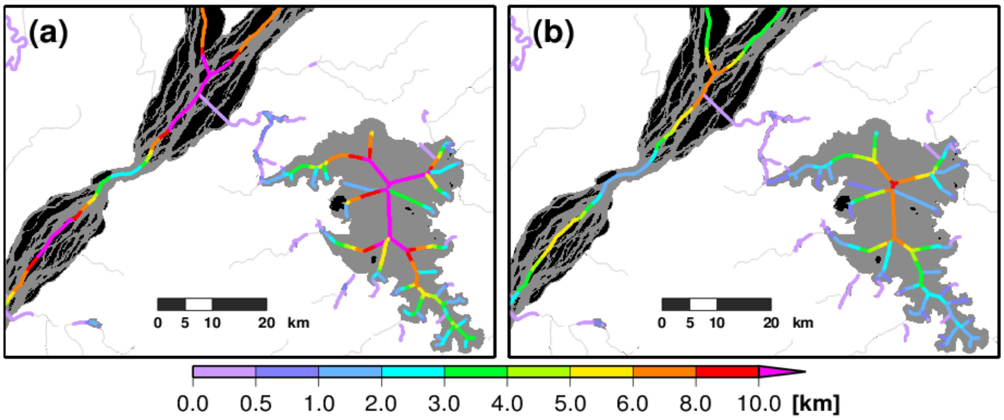

Figure 1.3: (a) Bank-to-bank river width. (b) Effective river width excluding islands. Water bodies are shown gray, while islands are represented by black.

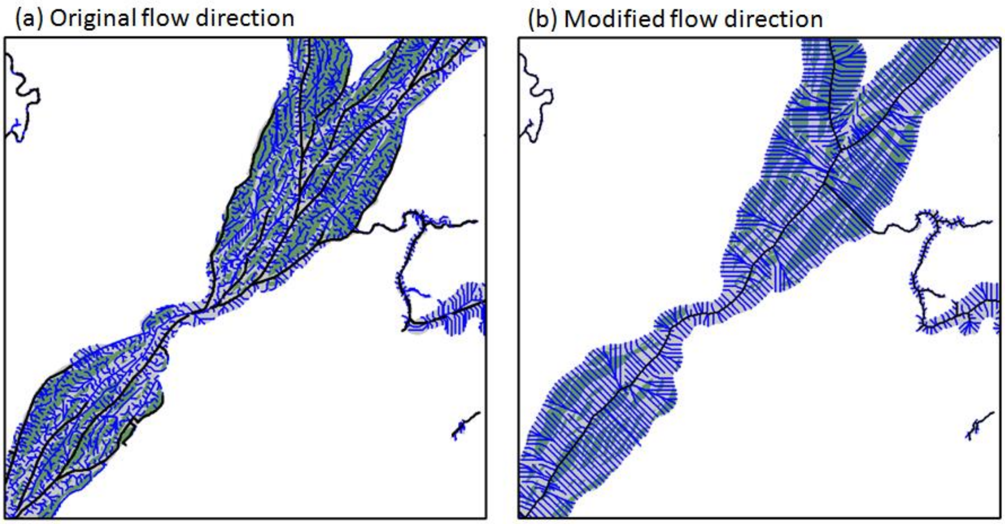

Figure 1.4: (a) Original flow direction map. Major streams are shown by black lines, while associated flows are shown by blue lines. (b) Modified flow direction map. Centerline pixels (black lines) and perpendicular flows to centerlines (blue lines) are illustrated. Islands are represented by dark green. Note that flow directions of limited pixels are shown to represent the difference between two panels.

2 Calculation Steps

2.1 Modification of Water Body Mask

Some modifications of the input water body mask are needed due to discrepancies between the input water body mask and the input flow direction map.

[Step 1.1] Pixels whose drainage area is larger than a threshold value (default: 100 km^2) are changed to water body pixels. This modification is needed because some rivers in the input flow direction map may run outside of water body areas of the input water body mask (see Figure 2.1).

The water body pixels added by this procedure are used for the calculation of bank-to-bank river width, but they are excluded in the calculation of effective river width.

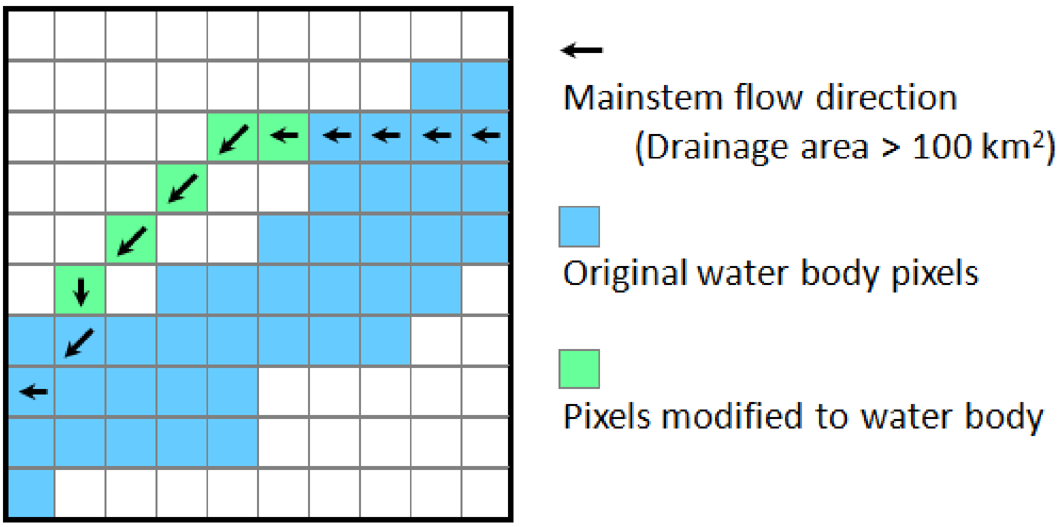

Figure 2.1: Water mask modification using drainage area.

[Step 1.2] Gaps between water masks whose area is smaller than the threshold value (default: 1000 km^2) are filled as island (Figure 2.2). Water body pixels and island pixels are termed as “in-bank pixels” which are used for the calculation of bank-to-bank river width.

Bank-to-bank width is calculated for all in-bank area (water body and island pixels), while island pixels are excluded when effective river width is calculated (see Section 2.10).

The default threshold for island filling is set to a relatively large value (1000 km^2). All island gaps except for very large ones (e.g. Ilha do Bananal in the Amazon River) are filled by this threshold. This large threshold is used to amalgamate bifurcated channels into one merged channel, because the GWD-LR is mainly developed for application to large-scale river models which cannot represent channel bifurcation.

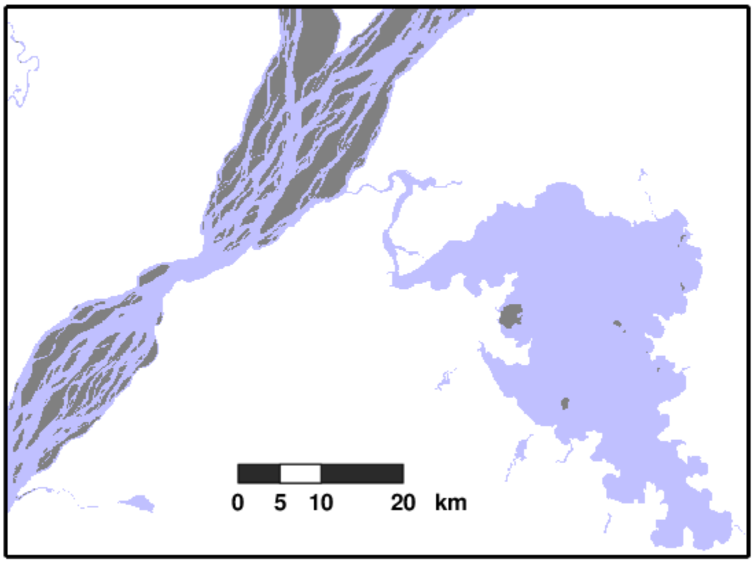

Figure 2.2: Island gap filling. Gaps in water body mask whose area is smaller than 1000 km^2 are filled as island pixels (gray).

[Step 1.3] Water body pixels which represent very narrow channels (with 1-pixel width) are changed to land pixels (Figure 2.3). This modification is applied to improve the computational efficiency.

Most of these very nrrow channels are not represented in the original SWBD water mask, given that the minimum channel width of the SWBD water mask is 183 m. Thus, most of them are introduced by the modification in Step 1.1.

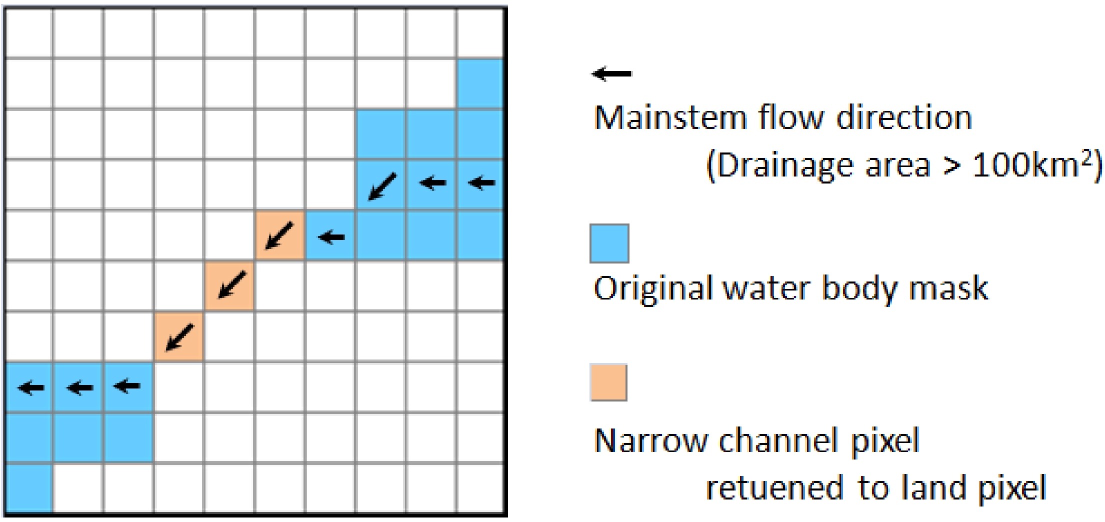

Figure 2.3: Narrow channels with 1-pixel width are changed to land pixels.

[Step 1.4] Pixels located within 1 km downstream from water bodies are changed to water body pixels. This modification is required because it is assumed in the proposed algorithm that river width is calculated for each “water body unit” which shares one outlet pixel. If there are multiple outlet from one water body, the water body is treated as multiple water body units, and river width is calculated separately for each water body unit. The modification in downstream of a water body generates a shared outlet pixel for each water body unit (Figure 2.4).

The water body pixels added by this procedure are used for the calculation of bank-to-bank river width, but they are excluded when effective river width is calculated.

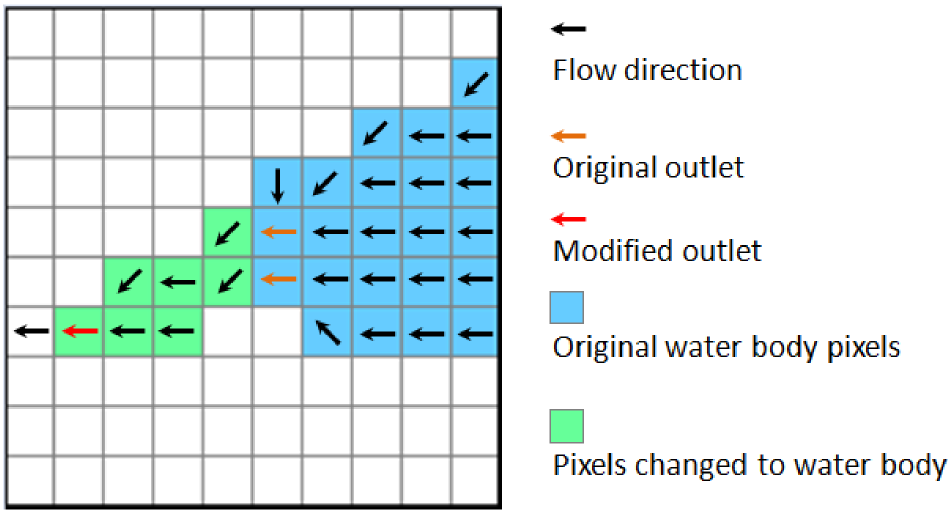

Figure 2.4: Modification on water body downstream.

2.2 Calculation of Water Body unit ID

[Step 2] Water body and island pixels which share the same water body outlet (a water body pixel with flow direction toward a land pixel) are treated as one water body unit. Identical ID number is given to each water body unit.

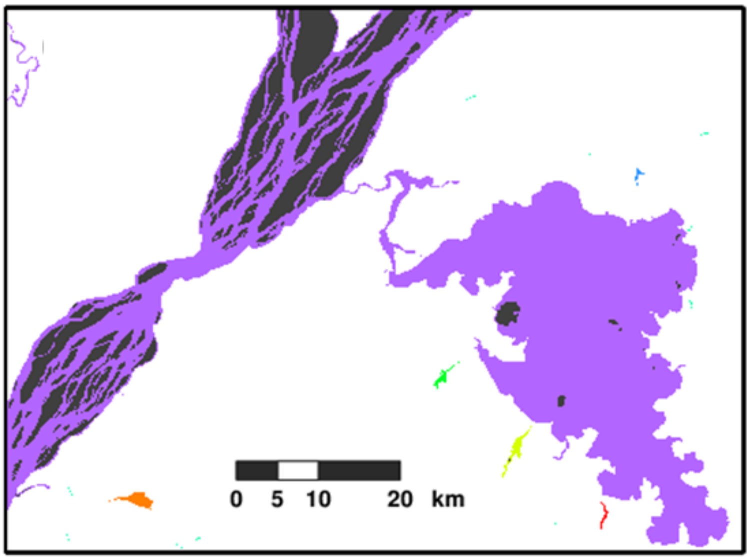

Figure 2.5: Water body unit ID. Each color represents one water body unit. Island pixels are represented by gray.

2.3 Calculation of Bank Distance

[Step 3] Bank pixels are identified by searching eight neighboring pixels of each water body pixel. If any of neighboring pixels is a land pixel, the water body pixels is identified as a bank pixel. Then, distance to the nearest bank pixel (hereafter “bank distance”) is calculated for each in-bank pixel (Figure 2.6).

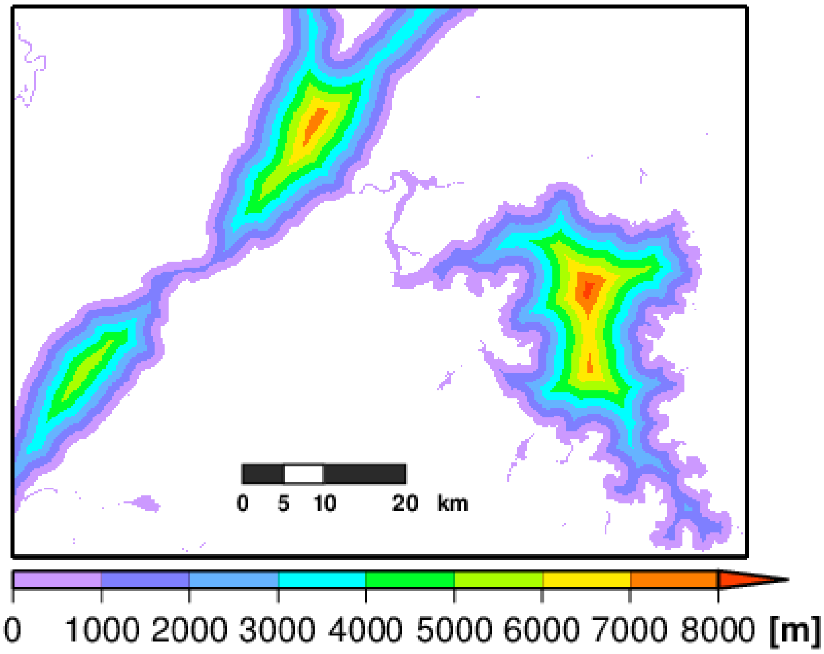

Figure 2.6: Bank distance. Distance to the nearest bank pixel is calculated for each in-bank pixel.

Figure 2.7 is a schematic illustration of the definitions of bank distance, centerline distance, and local distance between adjacent pixels. The bank distance is given as the distance to the nearest river bank pixel. The bank distance of the blue squared pixel is given by the length of blue vector. The centerline distance is given as the normalized distance to the nearest centerline pixel (explained in Section 2.8). The centerline distance of the red squared pixel is given by “the length of the red vector divided by the length of the black vector” (Equation 2.2). The black vector represents the bank distance of the black squared pixel, which is the nearest centerline pixel of the red squared pixel. The local distance between the two adjacent pixels (red and green squared pixels) is given by the length of the green line.

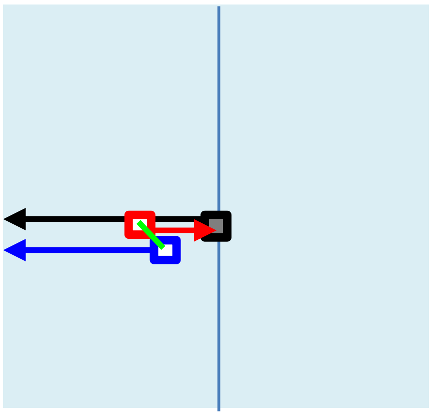

Figure 2.7: Schematic illustration of the definitions of bank distance D_b, centerline distance D_c, and the local distance between adjacent pixels L. The bank distance of the blue squared pixel is given by the length of blue vector. The centerline distance of the red squared pixel is given by “the length of the red vector (geometric centerline distance D_g) divided by the length of the black vector”. The local distance between the two adjacent pixels (red and green squared pixels) is given by the length of the green line.

2.4 Definition of Centerline Pixels

[Step 4.1] Centerline pixels are defined by searching convex points in the bank distance field. 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 the bank distance of six or more neighboring pixels; and (2) the maximum gradient of bank distance between the considered pixel and its neighboring pixel is not larger than the threshold gradient (set to be 0.26, or ~tan(15^{deg})). The gradient of bank distance is calculated by Equation (2.1):

$$\Delta D_b / \Delta x = \frac{D_{bj} - D_{bi}}{L} \quad (2.1),$$

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 local distance between the centers of the pixels i and j. Note that the distance between two point is calculated as a function of longitude and latitude assuming the earth ellipsoid in this study, so that the difference of actual distance in a degree coordination at different latitude is considered.

The first condition was introduced because only the upstream and downstream pixels on the 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 curvatures of river banks (see blue lines in Figure 2.8), and tend to extend from a river bank toward a true centerline. Thus, the gradient of bank distance tends to be larger on a spurious centerline than on a true centerline. The threshold gradient to distinguish true and spurious centerlines was set to be 0.26 (~tan(15^{deg})) 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 centerline pixels determined by these conditions are shown in gray lines in Figure 2.8, while spurious centerline pixels are shown in blue dots.

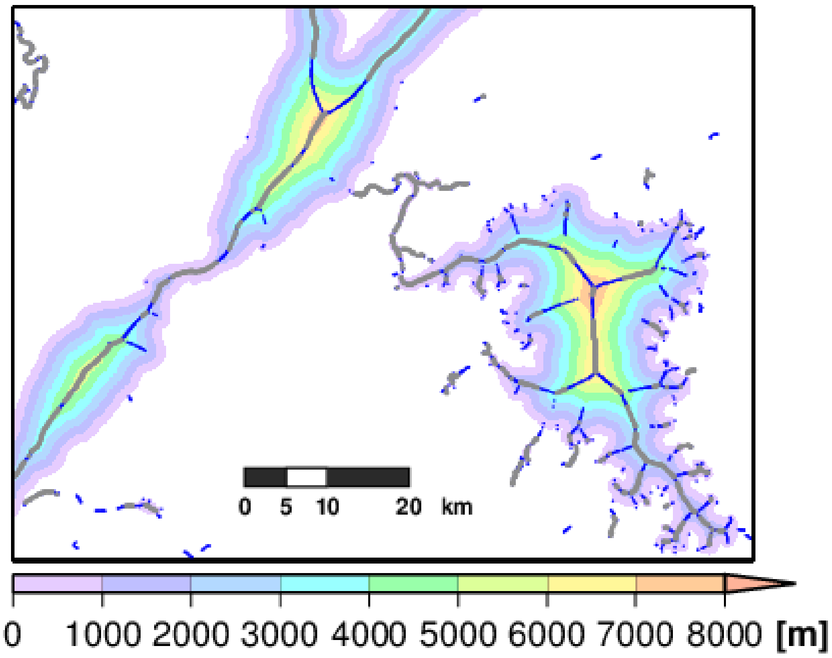

Figure 2.8: Defining centerline pixels by searching convex points in bank distance field. Gray: defined centerline pixels, blue: spurious centerline pixels. The bank distance field is shown by the background colors.

[Step 4.2] The centerline pixels detected by Step 4.1 have large gaps between them where river width is increasing because of the second condition. In order to improve the connectivity of centerlines, centerline pixels are extended by the following procedures: (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 converted to a new centerline pixel; (3) the extension procedure is repeated during the criteria (2) is true. The extended centerline pixels are shown by red lines in Figure 2.9.

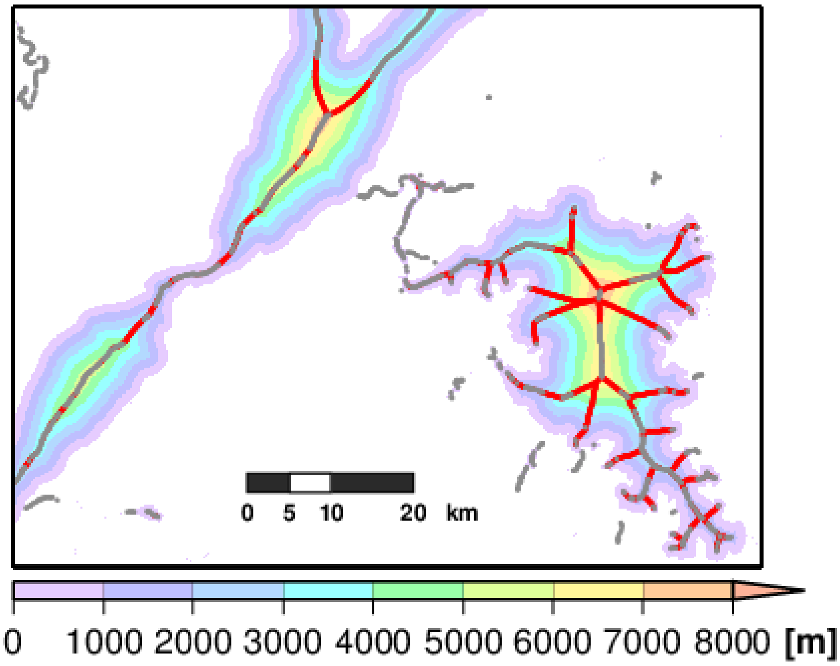

Figure 2.9: Centerline extension. Gray: original centerline pixels, red: extended centerline pixels. The bank distance field is shown by the background colors.

2.5 Calculation of Outlet Distance

[Step 5] The riverline distance from the outlet pixel of each water body unit (hereafter “outlet distance”) is calculated. The outlet distance is calculated by accumulating the local distance between pixels from the outlet pixel toward upstream (Figure 2.10).

In order to connect centerline pixels in a sequential downstream direction, the local distance between adjacent pixels is weighted by 10 when the pixel in the upstream side is not a centerline pixel. (blue vectors in Figure 2.10). Sensitivity to this weighing parameter is discussed in Section 3.

The calculated outlet distance is shown in Figure 2.11.

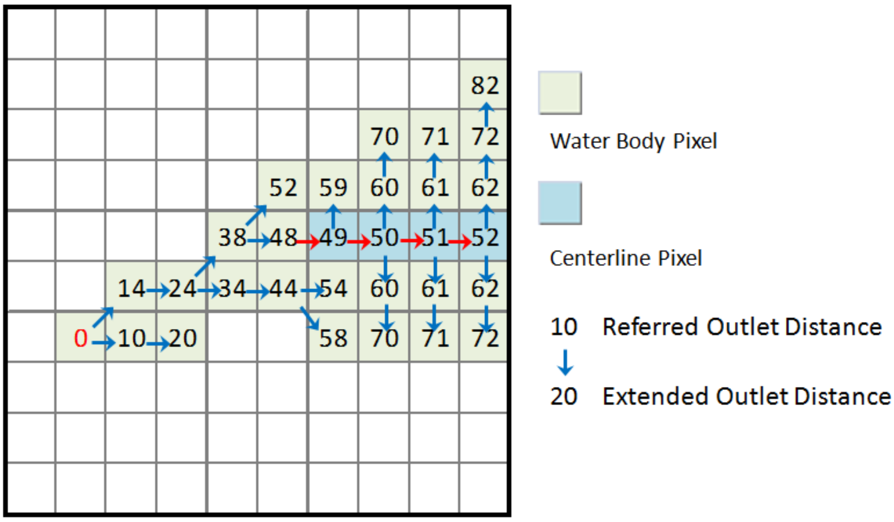

Figure 2.10: Schematic illustration of the outlet distance calculation. The pixel with number 0 denotes the outlet pixel of the water body unit. For easy explanation, the local distance between adjacent pixels in an orthogonal position is set to 1, while the local distance between adjacent pixels in a diagonal position is set to 1.4. The weighted local distances are set to 10 for orthogonal direction and 14 for diagonal direction (blue vectors). The outlet distance is calculated cumulatively from the outlet pixel of a water body unit toward upstream.

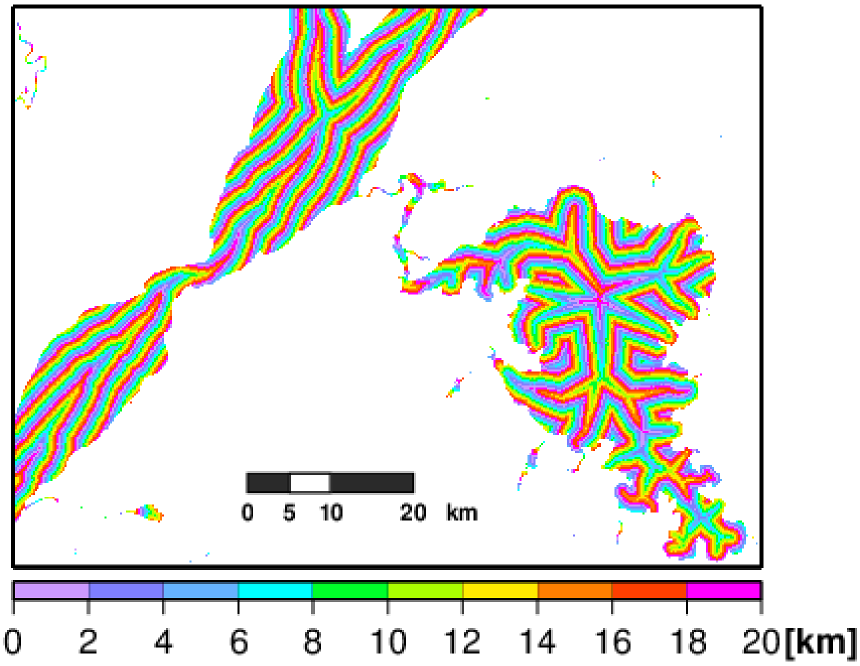

Figure 2.11: Outlet distance. The cumulative riverline distance with weight for non-centerline pixels from the outlet of each water body unit is shown by periodic colors.

2.6 Setting Centerline Flow Directions

[Step 6.1] The downstream direction of each centerline pixel is determined by choosing the pixel with minimum outlet distance among its eight neighboring pixels (Figure 2.12). If the selected downstream pixel is not a centerline pixel, the selected downstream pixel is changed to a new centerline pixel. The centerline extension is repeated until the downstream centerline pixel or the outlet pixel is connected (connected centerline pixels are shown by orange boxes in Figure 2.12 and red lines in Figure 2.13). Thus, the connectivity between all centerline pixels is ensured within a water body unit (Figure 2.13).

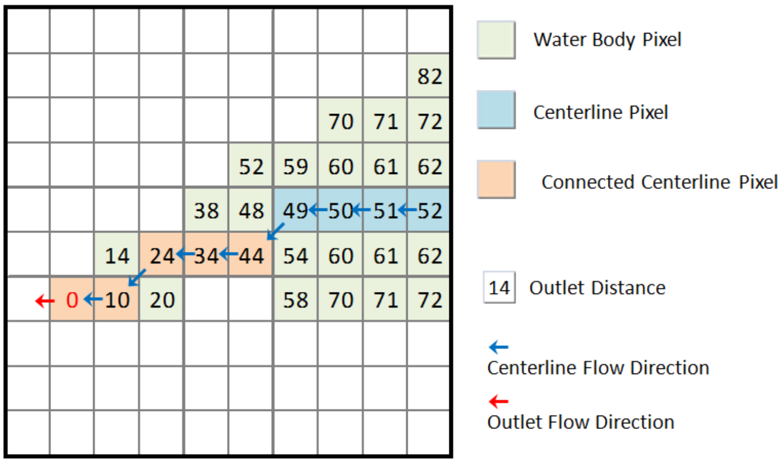

Figure 2.12: Determination of centerline flow directions.

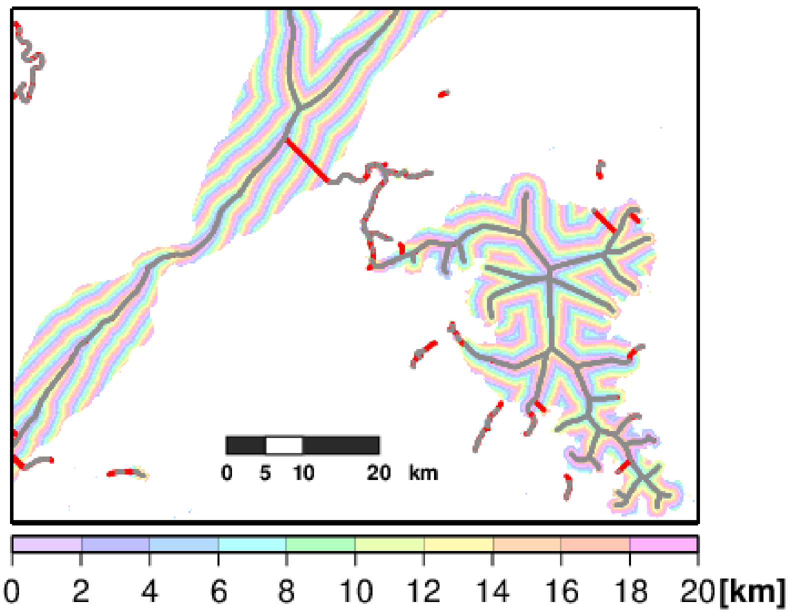

Figure 2.13: Centerline connection. Gray: original centerline pixels, red: connected centerline pixels. The outlet distance field is shown by the background colors.

[Step 6.2] The topmost centerline pixels are modified to non-centerline pixels. Most of the topmost centerline pixels which do not have an upstream centerline pixel are caused by the determination of flow directions (Figure 2.14) and they don't have to be treated as centerline pixels any more.

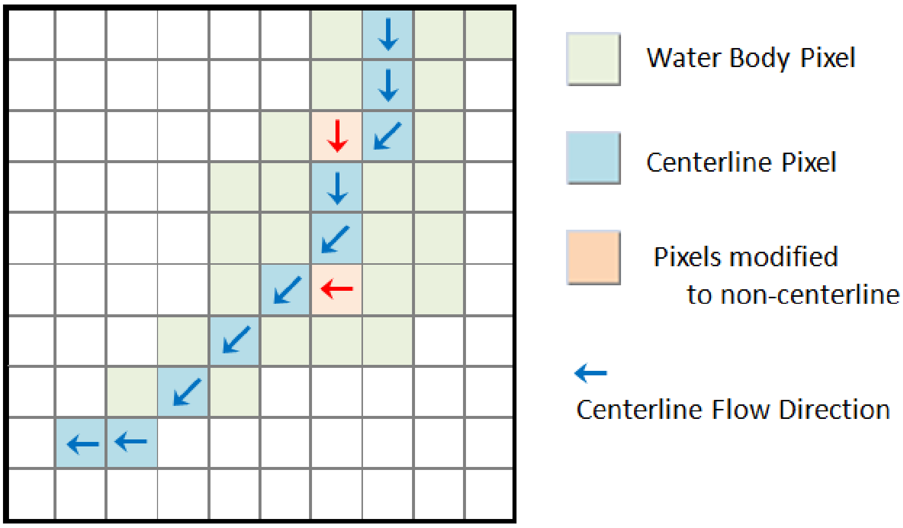

Figure 2.14: Modification on topmost centerline pixels. Topmost centerline pixels (orange) are converted to non-centerline.

2.7 Calculation of bank-to-bank river width

[Step 7.1] Centerline pixels are classified into actual centerlines and virtual connecting centerlines which connect a small tributary to a wide main channel. Centerline pixels are considered to be actual centerline when the gradient of bank distances toward downstream centerline pixel (calculated by Equation 2.1) is not larger than 0.57 (~tan(30^{deg})), while it's considered to be virtual connecting centerlines when the gradient is larger than 0.57 (Figure 2.15). Virtual connecting centerlines have rapidly increasing bank distance because they run from bank toward centerline in a wide main channel (blue lines in Figure 2.15). The threshold value of 0. 57 [~tan(30^{deg})] is decided by trial and error.

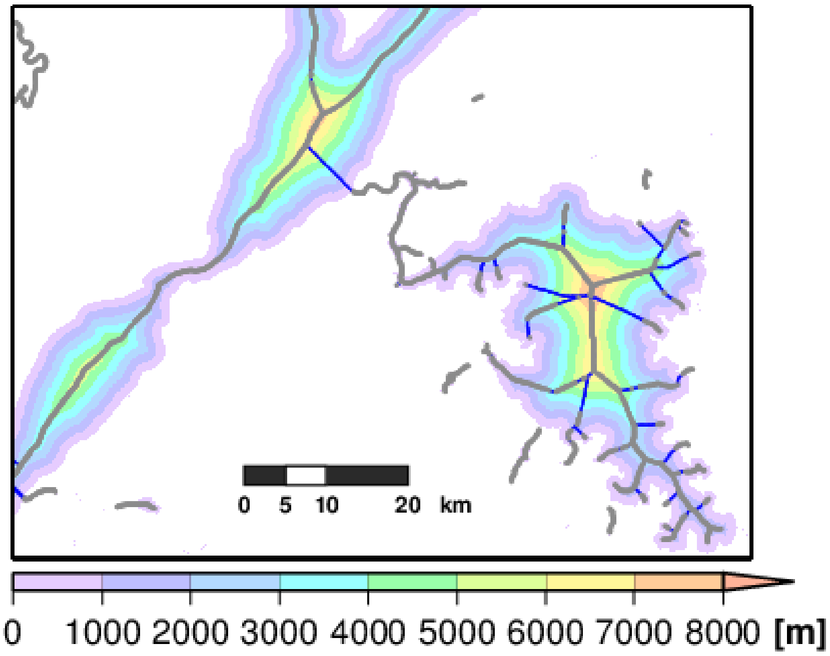

Figure 2.15: Centerline classification. Gray: actual centerline. Blue: virtual connecting centerline. Background color represents bank distance.

[Step 7.2] The bank-to-bank river width of actual centerline pixels is set to twice of its bank distance. The bank-to-bank distance of the virtual connecting centerlines (blue lines in Figure 2.15) is set to the same value as the bank-to-bank river width of its nearest upstream actual centerline pixel. The calculated bank-to-bank river width is shown in Figure 2.16.

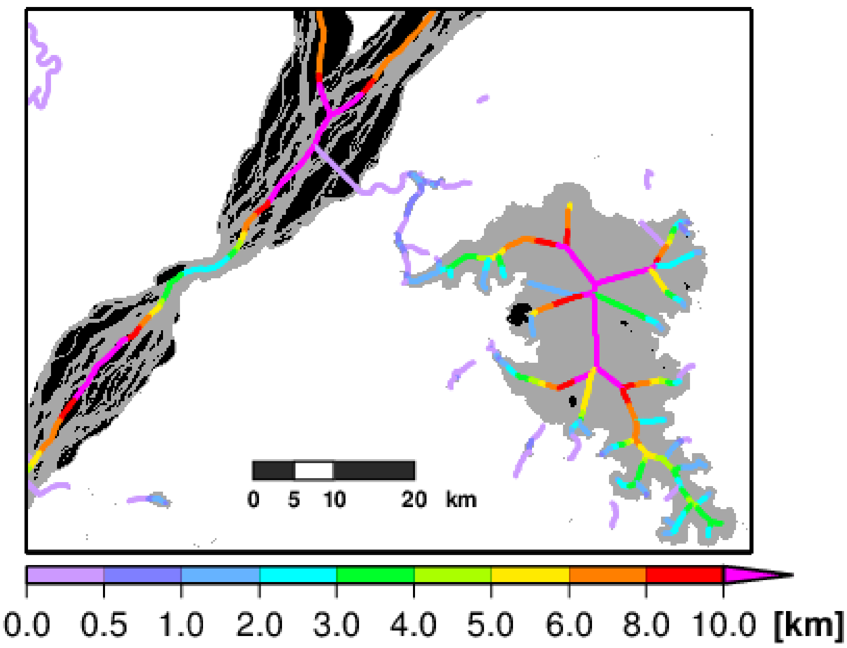

Figure 2.16 Bank-to-bank river width. Gray: non-centerline water body. Black: island.

2.8 Calculation of Centerline Distance

[Step 8] For every in-bank pixels, the distance to its nearest centerline pixel (hereafter "centerline distance") is calculated. 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 the tributary merges into its main channel (Explained in Section 3.2).

The scaled centerline distance is given by Equation (2.2):

$$D_c = \frac{D_g}{D_{bc}} \quad (2.2),$$

where D_c is the normalized centerline distance, D_g is the geometric distance to the centerline pixel, and D_{bc} is the bank distance of the nearest centerline pixel. The normalized centerline distance of a considered pixel becomes smaller when its nearest centerline pixel has larger bank distance and vice versa. The weighted centerline distance is shown by colors in Figure 2.17.

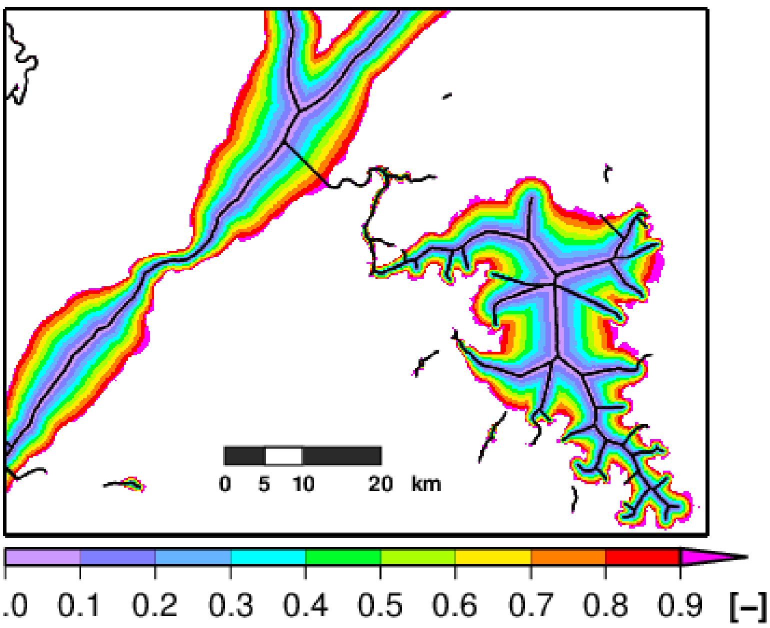

Figure 2.17: Centerline distance. The background colors represent centerline distance (i.e. normalized distance to its nearest centerline). Centerlines are illustrated by black lines.

2.9 Determination of flow direction

[Step 9] The flow direction of each non-centerline pixel is decided based on the gradient of centerline distance given by Equation 2.3.

$$\Delta D_c / \Delta x = \frac{D_{cj} - D_{ci}}{L} \quad (2.3),$$

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

Then, the flow direction is determined for each pixel by choosing the maximum gradient among the eight neighboring pixels. The modified flow directions are shown in Figure 2.18.

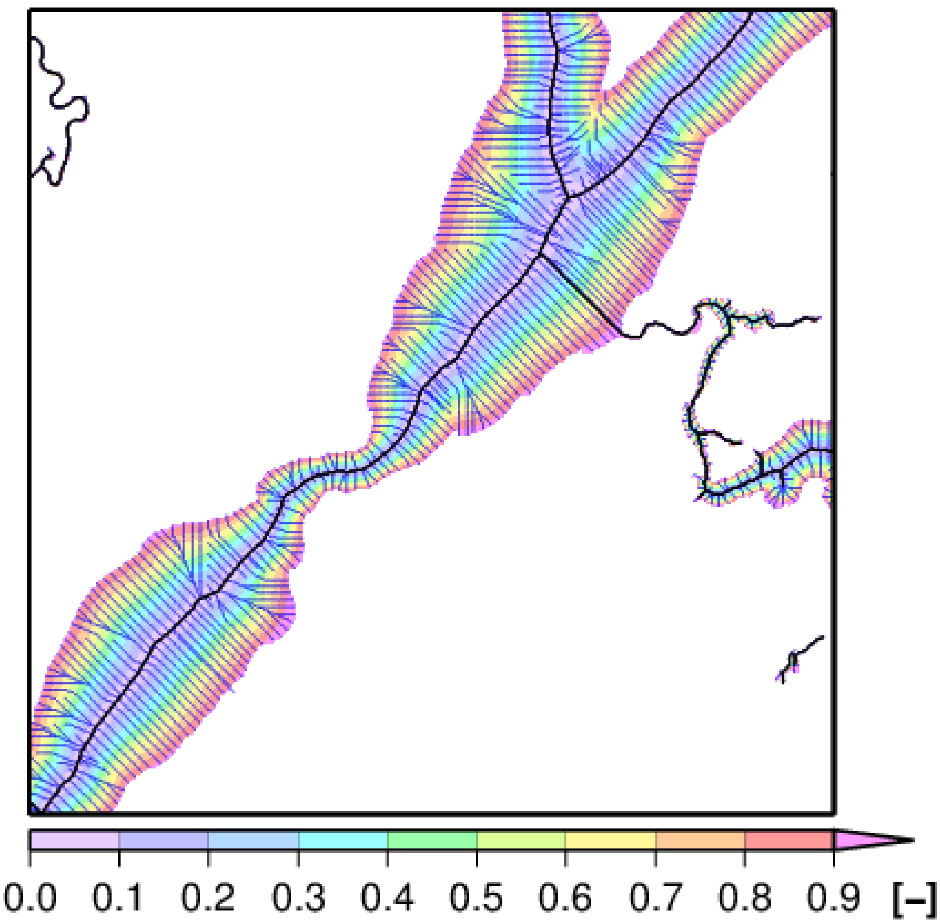

Figure 2.18 Calculation of Flow Directions. Flow directions are represented by blue lines while centerlines are indicated by black lines. Background colors represent normalized centerline distance. Note that flow directions of limited pixels are shown.

2.10 Calculation of effective river width

[Step 10.1] The effective river segment for each centerline pixel is defined for the calculation of effective river width excluding islands. 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 (the red dashed line in Figure 2.19), and includes the none-centerline pixels draining to that segment of centerline. (the area shaded with green in Figure 2.19).

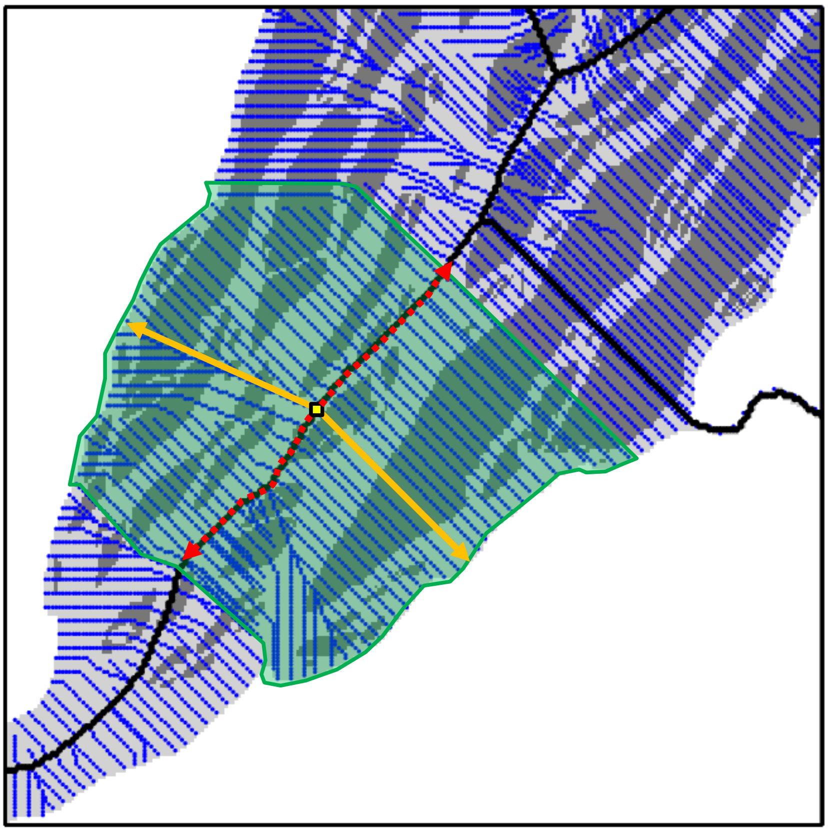

Figure 2.19: Calculation of effective channel segment. The green shaded area represents the effective river segment of the centerline pixel marked by the yellow square. The dashed red vector represents the effective centerline of the considered centerline pixel, which is determined by the bank distance of the considered pixels (orange vectors).

[Step 10.2] The total in-water area (water body and island) and the total water body area within the effective river segment are calculated. Then, effective river width is calculated by Equation (2.4).

$$W_e = W_b \times \left(\frac{A_w}{A_t} \right) \quad (2.4),$$

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 (i.e. water body and islands), A_w is the total water body area.

The effective river width of the virtual connecting centerlines (blue lines in Figure 2.15) is modified to the same value as the effective river width of its nearest upstream actual centerline pixel.

The calculated effective river width is shown in Figure 2.20.

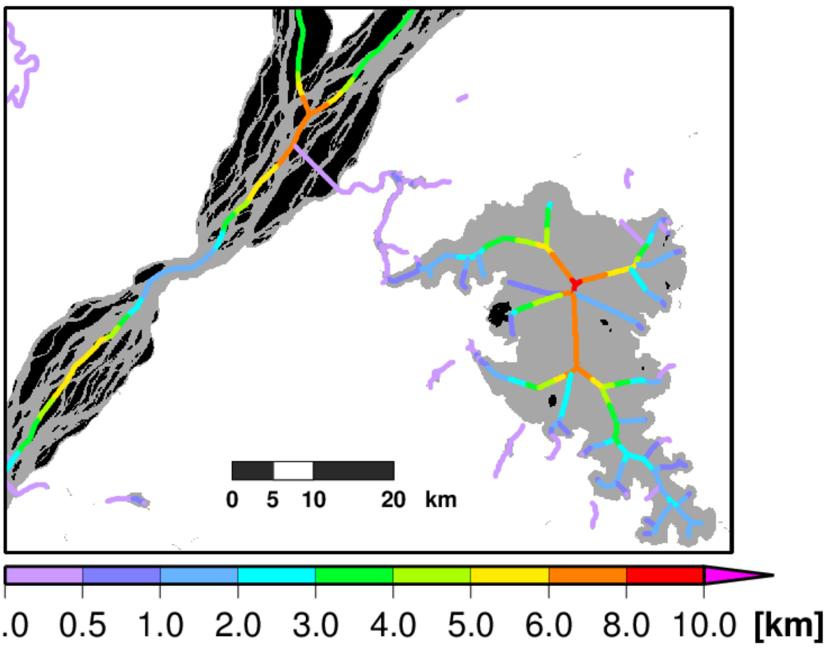

Figure 2.20. Effective river width. Non-centerline water bodies are represented by gray, while islands are represented by black.

3. Sensitivity to Parameters

Sensitivity of the river width calculation to parameters used in the global river width algorithm is discussed in this section.

3.1 Weight for outlet distance

The weight on the outlet distance is introduced to achieve sequential downstream connection of centerline pixels (Section 2.6). If the weight is too small, centerline pixels are not appropriately connected (the weight is set to 1 and 3 in Figures 3.1a and 3.1b, respectively). Intermittent centerlines are preferred to be connected to their nearest downstream centerline, but it's difficult to find the nearest downstream centerline when the weight too small. When the weight is large enough, the result is not sensitive to the weight value (the weight is set to 10 and 30 in Figures 3.1c and 3.1d, respectively). Calculation time becomes longer when a larger weight is used, thus we decided to use the weight of 10.

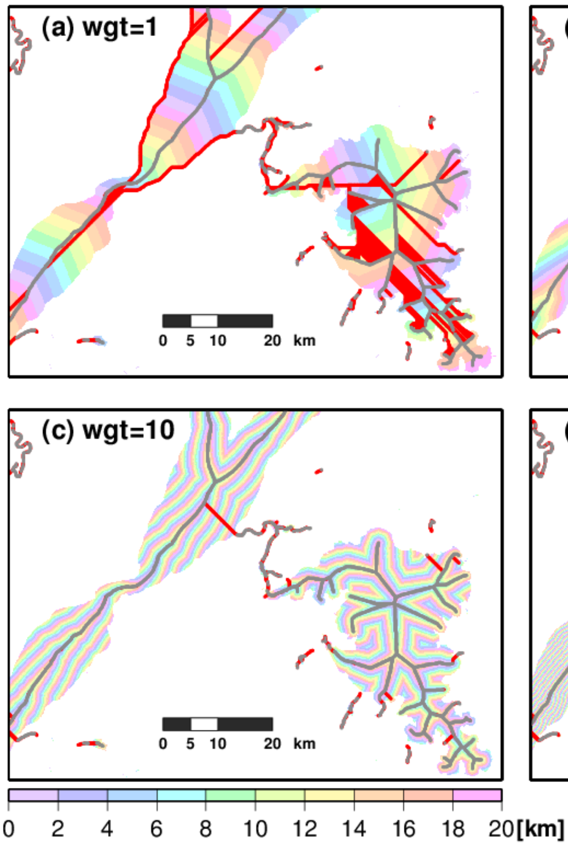

Figure 3.1: Sensitivity of centerline connection to the weight on outlet distance calculation. The original centerlines calculated from bank distance are shown by gray lines, while connected centerlines calculated by outlet distance are shown by red lines. Background colors represent outlet distance.

3.2 Normalized centerline distance

Centerline distance (i.e. distance to the nearest centerline, see Section 2.8) is normalized by the bank distance of the nearest centerline pixel (see, Equation 2.2). The normalization is introduced in order to avoid unrealistic accumulation of flow from an area outside of a tributary's width, within the zone where the tributary merges into its main channel. If centerline distance is not normalized, small tributaries unrealistically gather flows from its main channel (Figure 3.2b).

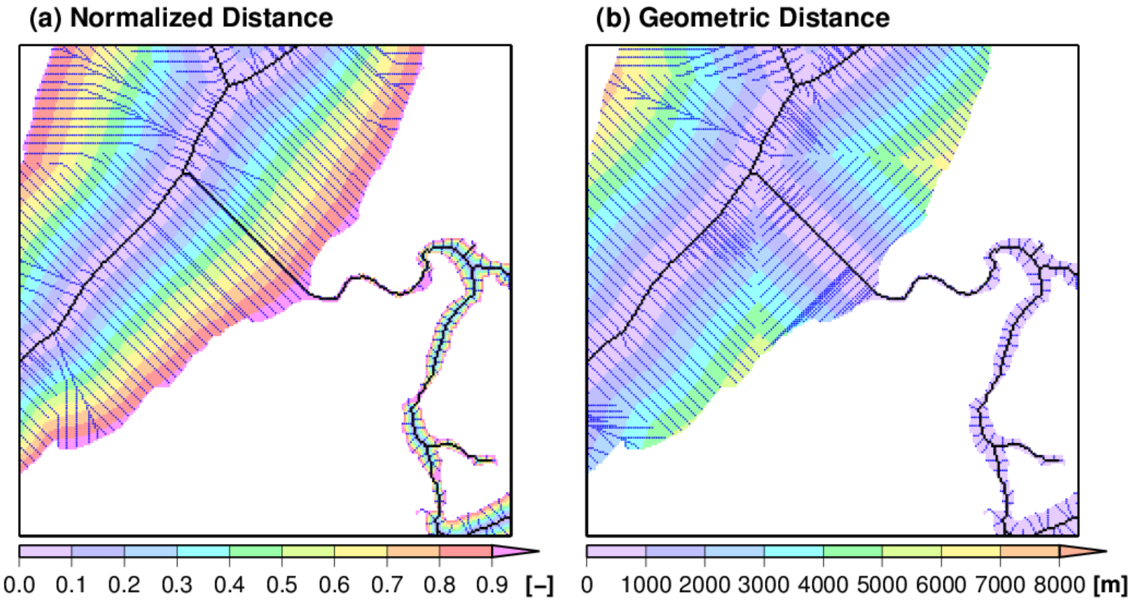

Figure 3.2 Modified flow direction derived from (a) normalized centerline distance, and (b) geometric (non-normalized) centerline distance. Flow directions are shown by blue lines, while centerlines are illustrated by black lines. Background colors represent centerline distance.

3.3 Threshold for island gap filling

Relatively large threshold (1000 km^2) for island gap filling was used in order to amalgamate bifurcated channels into one merged channel (Figure 3.3a), because the main target of GWD-LR is application to large-scale hydrodynamic models which cannot represent channel bifurcations.

In case width values are needed for each bifurcated channels, smaller threshold value for island gap filling may be preferred. By taking the smaller threshold, effective river width can be calculated separately for bifurcated channels (Figure 3.3b). Note that accuracy of effective river width calculation for bifurcated channels depend on the quality of the input water mask and input flow direction map. Manual correction may be needed because small channels are generally not well represented both in the water mask and the flow direction map.

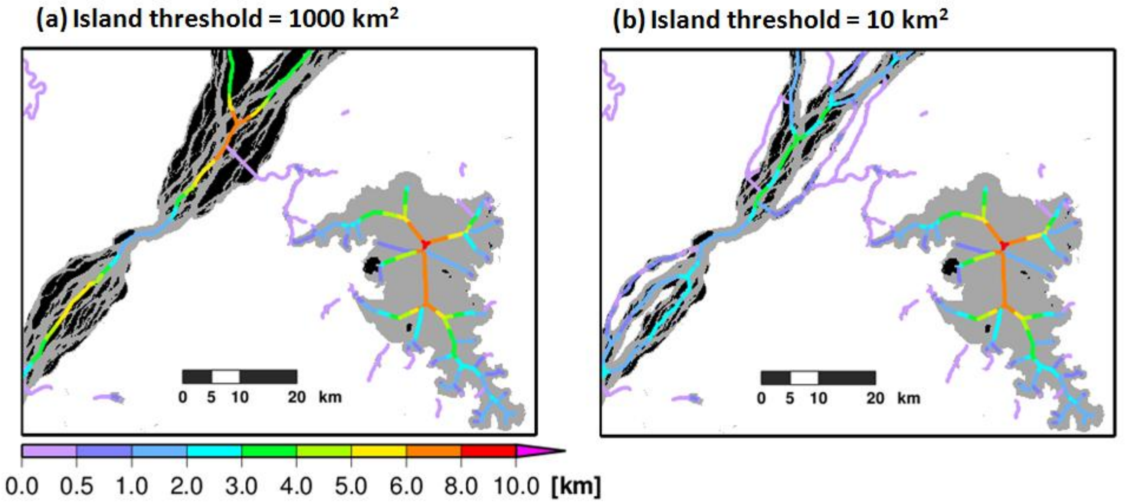

Figure 3.3: Effective river width calculated with a different threshold value for island gap filling. (a) 1000 km^2 threshold, (b) 10 km^2 threshold. Water masks are shown by gray, while filled islands are shown by black.

References

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. Proc., 27, 2171-2186, doi:10.1002/hyp.9740.

NASA/NGA (2003), SRTM Water Body Data Product Specific Guidance, Version 2.0, available online: http://dds.cr.usgs.gov/srtm/version2_1/SWBD/SWBD_Documentation/

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 Resources Research, vol.50, in print, DOI: 10.1002/2013WR014664