The Global Width Database for Large Rivers (GWD-LR) version 1.2

Dai Yamazaki
JAMSTEC – Japan Agency for Marine Earth Science and Technology
d-yamazaki@jamstec.go.jp

1. Introduction

The Global Width Database for Large Rivers (GWD-LR) is a satellite-based global-scale database of river width [Yamazaki et al., 2014a]. GWD-LR is calculated from satellite water mask and flow direction map using fully-automated width calculation algorithm. The GWD-LR is freely distributed for research purposes. Please visit the GWD-LR webpage (http://hydro.iis.u-tokyo.ac.jp/~yamadai/GWD-LR/index.html). In order to acquire the password to download the GWD-LR, please e-mail to the developer (Dai Yamazaki).

Note that both the input dataset and the calculation algorithm has been updated from the previous versions. A brief description for river width calculation is available online

.

2. Product Description

The GWD-LR is prepared as a raster data. The data format is “plain binary” (i.e. Fortran direct access, GrADS binary, ArcGIS BIL/BSQ/FLT format).

Effective river width (channel width excluding islands) is represented in the GWD-LR. A land area <500 km^2 surrounded by water mask is treated as an island. River width is calculated for each river centerline pixel, and the upstream-downstream relationships of centerline pixels are given by flow direction maps. Therefore, river width data is available sequentially along river networks.

The effective width data, flow directions, drainage area (and additional data such as elevation) are prepared at three different resolutions.

2.1 Original 3-sec resolution database

River width is calculated at 3-sec resolution (about 90 m at the equator), and the GWD-LR is available at the original 3-sec resolution. The data is divided into 5 degree by 5 degree tile (6000 pixels by 6000 pixels). Filename indicates the west and south corner of the domain, e.g. **n30w140_wth.flt** is the width file for the domain of N30-N35, W140-W135. Effective river width, flow direction map, drainage data is available in the original 3-sec resolution database.

[Effective River Width] ${LAT}${LON}_wth.flt

Effective river width (in [m], >0) is calculated for each centerline pixel. Land pixel, non-centerline water pixel, and island pixel are represented by 0, -1, -2, respectively. Ocean pixels are filled with undefined value (-9999). Data is prepared in 4 byte real format (in little endian).

[Flow Directions] ${LAT}${LON}_dir.bil

Flow direction of each pixel is represented by (1:N, 2:NE, 3:E, 4:SE, 5:S, 6:SW, 7:W, 8:NW). River mouth is represented by 0, while inland river terminating point is represented by -1. Ocean pixels are filled by undefined value (-9). Flow direction is also a raster data, thus all land pixels have a downstream direction. Note that the flow directions in GWD-LR is modified from the original HydroSHEDS and GDBD flow directions, in order to represent the centerline of each river segment. Flow direction map is prepared in 1 byte integer format.

[Drainage Area] ${LAT}${LON}_upa.flt

Drainage area of each pixel is saved in [m^2]. Ocean pixels are filled by undefined value (-9999). Drainage area is saved in 4 byte real format in little endian.

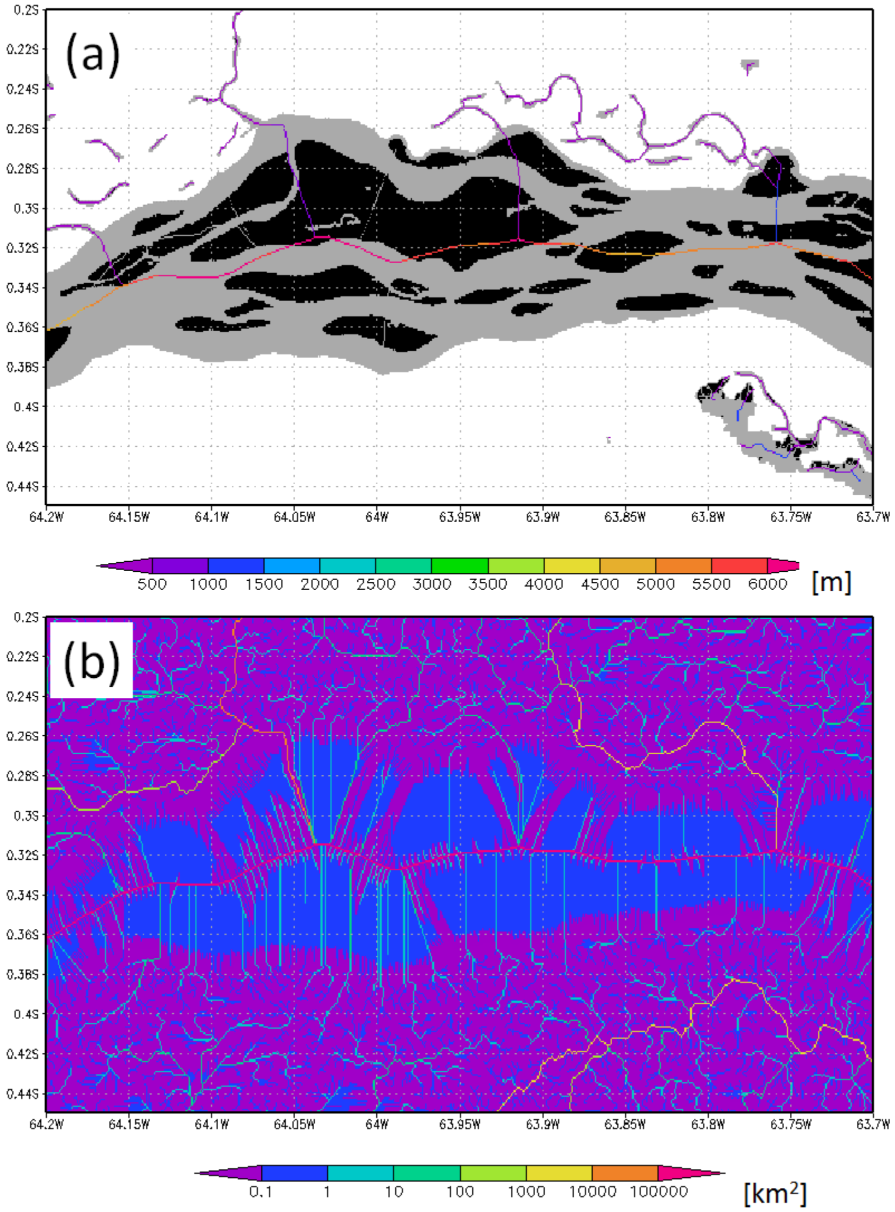

Figure 1: (a) Effective river width. Non-centerline water bodies are represented by gray, while islands are represented by black. (b) Drainage Area of modified flow direction map.

2.2 Regional 0.005 degree resolution database

The original 3-sec resolution database was upscaled to 0.005 degree resolution by the FLOW algorithm [Yamazaki et al., 2009] in order to reduce the file size for easy data handling. The global domain is divided into 14 areas tagged by 3 characters.

Table 1: List of areas in 0.005 degree resolution database

Area Tag	sa1	ca1	na1	af1	eu1	eu2	as1
ID	1	2	3	4	5	6	7
West	−85	−120	−130	5	−20	5	55
East	−30	−60	−50	60	20	70	100
North	15	40	60	35	60	60	60
South	−60	5	25	−35	0	20	5
nx	11000	12000	16000	11000	8000	13000	9000
ny	15000	7000	7000	14000	12000	8000	11000
cell size	0.005	0.005	0.005	0.005	0.005	0.005	0.005

Area Tag	as2	as3	oc1	na2	eu3	si1	si2
ID	8	9	10	11	12	13	14
West	90	90	110	−170	0	55	100
East	150	155	180	−55	70	115	195
North	60	35	−10	75	80	80	75
South	20	−15	−50	50	45	45	50
nx	12000	13000	14000	23000	14000	12000	19000
ny	8000	10000	8000	5000	7000	7000	5000
cell size	0.005	0.005	0.005	0.005	0.005	0.005	0.005

ID, domain boundary, number of cells, cell size are listed.

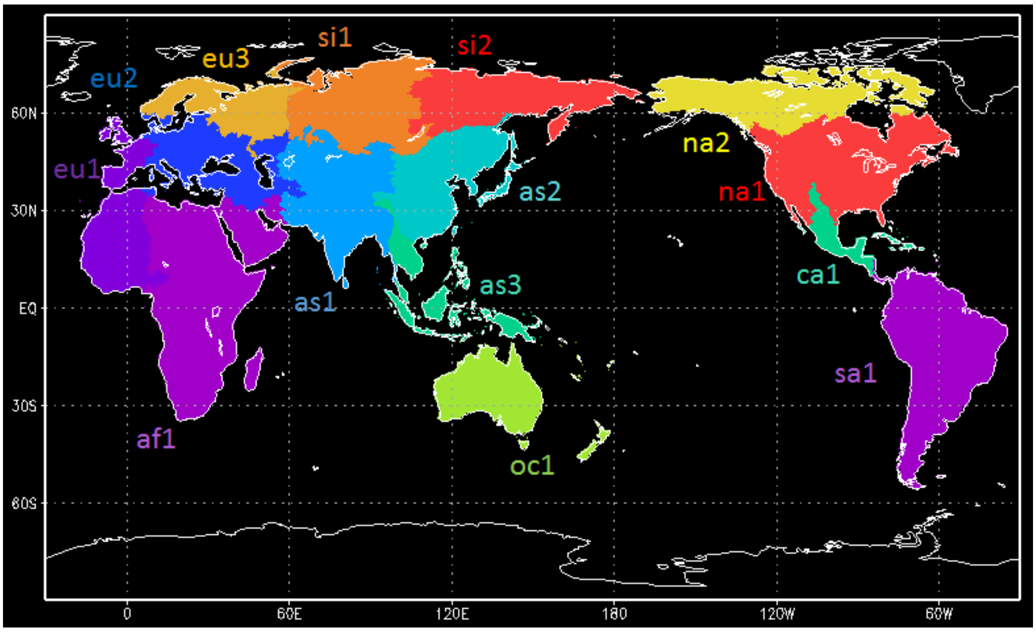

Figure 2: Areas for 0.005 degree resolution database.

[Effective River Width] ${AREA}.rivwth.flt

Effective river width (in [m], >0) is calculated for each centerline cell. Land cell, non-centerline water cell, and island cell are represented by 0, -1, -2, respectively. Ocean cells are filled with undefined value (-9999). Data is prepared in 4 byte real format (in little endian).

[Downstream XY] ${AREA}.nextxy.bsq

Downstream cell (jx,jy) of each cell (ix,iy) is saved. Record=1 represents downstream jx, while record=2 represents downstream jy. River mouth is represented by -9. Ocean cells are filled with undefined value (-9999). Data is prepared in 2 byte integer format.

The “downstream XY format” is used because upscaling of river network is mathematically impossible when “D8 flow direction format” is used. For detail, see [Yamazaki et al., 2009].

[Drainage Area] ${AREA}.uparea.flt

Drainage area of each cell is saved in [km^2]. Ocean cells are filled by undefined value (-9999). Drainage area is saved in 4 byte real format in little endian.

[Adjusted Elevation] ${AREA}.elevtn.flt

Elevation of each cell is saved. Ocean cells are filled by undefined value (-9999). Elevation is saved in 4 byte real format in little endian.

Elevation is based on SRTM3 DEM (GTOPO30 DEM above 60N), and is adjusted (smoothed) along river networks to remove depressions [Yamazaki et al., 2012].

[Longitude / Latitude of outlet pixel] ${AREA}.lonlat.flt

Longitude and latitude of the outlet of each pixel are saved. Record=1 is for longitude, record=2 is for latitude. Longitude and latitude is saved in 4 byte real format in little endian.

Each 0.005 degree cell has 36 3-sec pixels. The outlet pixels of each cell represent the position of river channel within the 0.005 degree cell.

2.3 Global database (1/40, 1/10, 1/4 degree resolution)

The 0.005 degree resolution database is again upscaled by the FLOW algorithm to construct global coverage database. Global database is prepared at various resolutions. This global database can be used in the global hydrodynamic model CaMa-Flood [Yamazaki et al., 2011, 2013, 2014b].

[Effective River Width] width.flt

Effective river width (in [m], >0) is calculated for each centerline grid. Land grid, non-centerline water grid, and island grid are represented by 0, -1, -2, respectively. Ocean grids are filled with undefined value (-9999). Data is prepared in 4 byte real format (in little endian).

[Downstream XY] nextxy.bsq

Downstream grid (jx,jy) of each grid (ix,iy) is saved. Record=1 represents downstream jx, while record=2 represents downstream jy. River mouth is represented by -9. Ocean grids are filled with undefined value (-9999). Data is prepared in 4 byte integer format.

[Associated Maps]

Global database contain files listed in the below table.

Table 2: List of files in global domain database

File	Description		Unit	Format
nextxy.bsq	Downstream X	(rec=1)		integer
	Downstream Y	(rec=2)		
width.flt	GWD-LR Effective Width		[m]	real
elevtn.flt	Base Elevation		[m]	real
uparea.flt	Upstream Drainage Area		[m2]	real
rivlen.flt	Channel Length		[m]	real
basin.bsq	Bain ID		–	integer
bsncol.bsq	Basin Color Patern for Visualization		–	integer
lonlat.flt	Outlet Longitude	(rec=1)	deg	real
	Outlet Latitude	(rec=2)	deg	real

Note: data format is 4 byte real or 4 byte integer, both in little endian.

3. Input Data & Method

3.1 Flow Direction Map

HydroSHEDS 3 arc-sec flow direction map [Lehner and Grill, 2013] is used as baseline data between 60N and 60S. GDBD 1km flow direction [Masutomi et al., 2009] is used for region above 60N. GDBD 1km flow direction map was downscaled to 3 arc-sec resolution, and fused to HydroSHEDS flow direction map [Downscaling and fusing method: not published yet]

3.2 Water Mask

SRTM Water Body Data (SWBD) [NASA/NGA, 2003] is used as baseline data between 60N and 60S. GLCF's MODIS Water Mask [Carrol et al., 2009] is used for region above 60N. MODIS Water Mask is downscaled to 3 arc-sec resolution for calculating river width.

3.3 Elevation Data

For adding elevation to river network map, SRTM3 DEM (CSI version 4.1) [Reuter et al., 2007; Farr et al., 2007] is used between 60N and 60S. The area above 60N is covered by GTOPO30 [USGS]. Elevations are adjusted to remove depressions along river networks [Yamazaki et al., 2012]. Then, integer-value elevations are converted to real-value in order to smooth stepwise elevation increase due to integer representation [Smoothing method not published yet].

3.4 River Width Calculation

Brief description on the width calculation algorithm is explained in Yamazaki et al., 2014a. Detailed description document and sample source code are available online.

3.5 Upscaling Method

Flow direction and river width at 3 arc-sec resolution are upscaled to 0.005 degree database by the FLOW algorithm [Yamazai et al., 2009]. Then, the 0.005 degree database is again upscaled by the FLOW algorithm to create global database.

4. Update in version 1.2

4.1 Database above 60N

GWD-LR is extended to regions above 60N, using GDBD flow direction map and MODIS water mask. Because low-resolution baseline data is used, the accuracy of river width is lower in regions above 60N, compared to the accuracy between 60N and 60S.

4.2 Channel/Floodplain Separation

The algorithm is updated in version 1.2, and is now able to separate channels and floodplains more strictly then the previous version. The accuracy of river width calculation is improved.

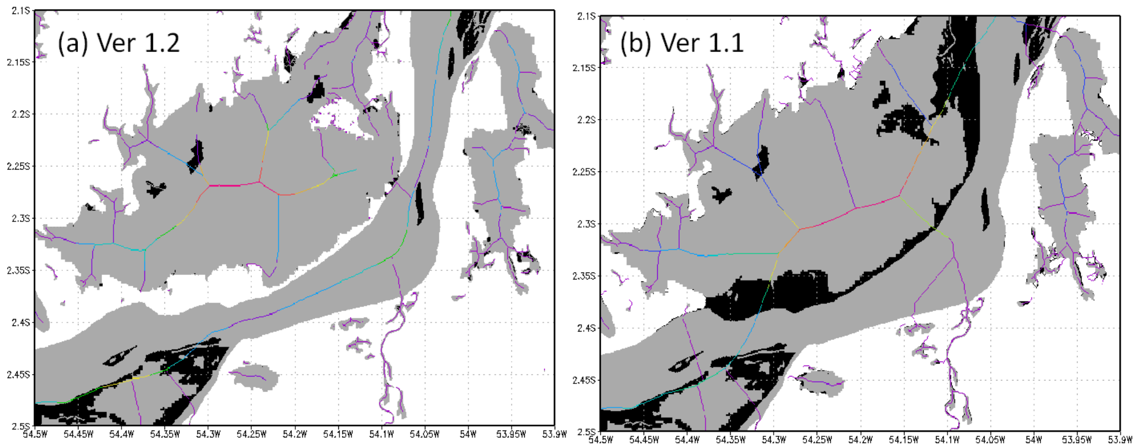

Figure 3: River centerlines calculated by (a) ver 1.2 algorithm and (b) ver 1.1 algorithm. Channels and floodplains was not perfectly separated in version 1.1.

4.3 Seamless 3-sec resolution data

In the previous version (ver. 1.1), the original 3-sec database is divided into 10 areas. River width calculation algorithm is applied to global domain in version 1.2, thus 3-sec resolution database (prepared as 5deg x 5deg tiles) has no area boundary.

4.4 Parallelized Fortran90 code

The width calculation algorithm was parallelized by OpenMP. Now, global calculation at 3-sec resolution can be done within one day. Previously, it took 3-4 days to finish global calculation.

5. Future Tasks

5.1 Lake Mask

In GWD-LR, width is calculated for all water bodies, without separating river channels, floodplains/wetlands, and lakes. This may not be ideal when GWD-LR is applied to hydrodynamic model simulations or any other researches.

In order to distinguish river channels and lakes, a lake mask is currently under construction.

5.2 High-resolution water mask

The accuracy of river width is low in the area above 60N because of low-resolution MODIS water mask is used as baseline data. A new high-resolution water mask is under development using LANDSAT GLS database.

References

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.

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

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.

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

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/

Reuter H.I, A. Nelson, A. Jarvis (2007), An evaluation of void filling interpolation methods for SRTM data, International Journal of Geographic Information Science, 21:9, 983-1008

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., 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., G. A. M. de Almeida, and P. D. Bates (2013), Improving computational efficiency in global river models by implementing the local inertial flow equation and a vector-based river network map, Water Resour. Res., 49, 7221 – 7235, doi:10.1002/wrcr.20552.

Yamazaki, D., F. O'Loughlin, M. A. Trigg, Z. F. Miller, T. M. Pavelsky, and P. D. Bates (2014a), Development of the global width database for large rivers, Water Resour. Res., 50, doi:10.1002/2013WR014664.

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