An Improved Upscaling Method to Construct a Global River Map

Dai YAMAZAKI[1], Yuji MASUTOMI[2], Taikan OKI[1] & Shinjiro KANAE[1]
[1] Institute of Industrial Science, the University of Tokyo, Japan
[2] National Institute for Environmental Studies, Japan
yamadai@rainbow.iis.u-tokyo.ac.jp

Abstract

A global river model is useful for (1) the climate model simulation to close water circulation, (2) validation of the land surface model by converting soil runoff to river discharge and (3) the water resource assessment by estimating renewable freshwater amount. A river model consists of a river map which indicates the downstream of each grid to express major rivers and river routine which calculates water transportation along wit the river map. Construction of the river map requires two processes. The first process is what is called upscaling, which convert super fine resolution river datasets into a coarse resolution river map. Global fine resolution (about 1km grid size) river datasets are already available, but lowering resolution is required according to computational limitation. However, this upscaling method cannot perfectly produce a realistic river map, thus we need manual error correction of the upscaled river map as the second process. To reduce the burden of error correction, a new upscaling method is proposed in this study. Previous upscaling methods tend to cause many errors where more then two rivers run very closely and they are located in a single grid, but the new method can reduce the number of errors by distinguishing multiple rivers in a single grid. The efficiency of some upscaling methods is discussed with Nash-Sutcliffe coefficient which compares drainage area of the original fine resolution river data and that of the upscaled river map. The Nash-Sutcliffe coefficient of the new method reached up to 0.99, which shows much increase from previous methods. This result indicates that the upscaled river map with the new method expresses realistic river channel networks and it requires much less error correction than any previous upscaling methods.

Key Words: global river model, upscaling, DEM

1. Introduction

Global river models were originally developed for use in climate models to close water circulation by representing water transport from soil runoff to oceans [Millar et al., 1994]. After that, global river models are applied to validate the land surface processes of climate models [Oki et al., 1999; Hirabayashi et al., 2005]. Because river discharge is an integration of soil runoff within a river basin wide, amount and timing of the runoff generation can be evaluated by comparing model simulated and observed river discharge. Global river models are also helpful for the water resources assessment, because simulated river discharge can be translated into the renewable freshwater amount which human well-beings can use for agriculture, industry or domestic purpose [Hanasaki et al., 2008].

A global river model often consists of a river routing and a global river map [Millar et al, 1994; Oki et al, 1999], and it simulates river discharge by dividing the entire globe into longitude-latitude based grids. The river routing estimates outflow discharge from each grid mainly based on linear reservoir model, and the global river map indicates the downstream grid into which calculated outflow from each grid enters. Thus, generated soil runoff is transported from upstream to downstream along with the global river map, and the global river model is able to simulate continental scale river flows.

To simulate river discharge properly, the global river map should accurately represent realistic river channel networks. Various algorithms to construct a realistic global river map have been studied in recent decades. The most simple and traditional one is so called the Steepest Slope Method; it determines the downstream of each grid by choosing the grid which produces the steepest slope among eight neighboring grids. Gradient between grids are estimated with grid averaged elevation, however in global scale, grid averaged topography does not always agree with the micro-scale topography which decides the ways of river streams, and as a result manual correction of an estimated global river map is need to reconstruct the global river map with acceptable accuracy. To reduce the burden of manual correction, Fekete [2001] has developed the method which uses a global super fine resolution river datasets instead of digital elevation model (DEM) to construct a global river map. His method is now classified as “upscaling method”, since it converts a super fine resolution global river datasets into a coarse resolution river map which is acceptable for computer calculation. Upscaling method produces a more realistic river map than the Steepest Slope Method, because super fine resolution river datasets have more useful information about river channel networks than DEM. Later, Doell [2002] and Olivera [2002] improved the efficiency of upscaling methods, and constructed river maps by their method became more realistic. However, their upscaling methods were not perfect that manual correction of the upscaled river map was still required.

In this study, previous upscaling methods are reviewed to discuss why manual correction is required for each method. Then an improved upscaling method is newly introduced, and results of previous and new upscaling methods are compared. Description about the target global river map and input datasets are written in section 2, and previous and the newly improved upscaling methods are explained in section 3. Results of each upscaling method are compared in section 4, and summary and discussions are written in section 5.

2. Framework of the target global river map and input datasets

2.1 Framework of the target global river map

To compare the efficiency of each upscaling method, framework of the target global river map should be same. Resolution of the target global river map is fixed to T213. T213 is the way of divide the entire globe into grids based on spectrum method, and it is often adopted in atmospheric global circulation models. The grid size is about 0.56 arc-degrees (around 50km in the equator) and entire globe is divided into 640x320 grids.

In global river maps, each grid point is assumed to have only one outflow direction to one of eight neighboring grids, so flow direction of each grid is indicated with the direction of downstream neighboring grid: North, Northeast, East, Southeast, South, Southwest, West and Northwest. This method which expresses river channel networks with eight directions is classified as “D8” method [Costa-Cabral and Burges, 1994]. D8 method has limitation to represent complex river flows in river mouth deltas or artificial water withdrawal to artificial canals where water tends to flow into multiple directions. However in global scale, the continental scale river discharge is able to be simulated with river maps expressed by D8 method.

2.2 Input datasets

In this study, GTOPO30 by United States Geological Survey (USGS) is used as input DEM for the Steepest Slope Method. GTOPO30 express elevation in 30 arc-seconds resolution for the entire glove, and it is aggregated into T213 grid averaged elevation. Area of 30 arc-seconds cells are calculated by assuming the earth ellipsoid. The area between latitude ϕ_1 and ϕ_2, and longitude λ_1 and λ_2 is given by equation (1)

$$S(\phi_1,\phi_2,\lambda_1,\lambda_2) = (\lambda_1 - \lambda_2)\left[\frac{\pi a^2(1-e^2)}{180}\left(\frac{e\sin\phi}{2(1-e^2\sin^2\phi)}\right)+\frac{1}{4}\ln\left|\frac{1+e\sin\phi}{1-e\sin\phi}\right|\right]_{\phi_2}^{\phi_1} \tag{1}$$

Here, the radius of earth a=6378.136 km, and the ellipticity of the earth e^2=0.006699447.

As an input super fine resolution river datasets for upscaling methods, flow direction map of Global Drainage Basin databases (GDBD) [Masutomi, 2007] is used in this study. GDBD is 1km resolution global river datasets, and the flow direction map of GDBD expresses the river channel networks by D8 methods.

3. Description about upscaling methods

3.1 Steepest Slope Method (SSM)

The Steepest Slope Method is the most simple and traditional method to construct a river map. It determines the flow direction by choosing the steepest slope among slopes toward neighborhood eight grids (Fig.1-a). Gradient between grids is calculated by grid averaged elevation and distance between girds. Grid averaged elevation is derived from GTOPO30 (umbers in Fig.1-a) and distance between grids is calculated by assuming the earth ellipsoid. The point of longitude ϕ and latitude λ is converted into orthogonal coordinate *(X, Y, Z)* by equation (2)

$$X = \frac{a\cos\phi\cos\lambda}{\sqrt{1-e^2\sin^2\phi}}, Y = \frac{a\cos\phi\sin\lambda}{\sqrt{1-e^2\sin^2\phi}}, Z = \frac{a\sin\phi}{\sqrt{1-e^2\sin^2\phi}} \tag{2}$$

Here, the radius of the earth a, and the ellipticity of the earth e^2 is set to the same values as equation (1). Distance between two points is calculated geometrically based on the orthogonal coordinate derived by equation (2).

3.2 Network Scaling Algorism (NSA)

The Network Scaling Algorithm is developed by Fekete [2001], and it uses super fine resolution river datasets instead of DEM (Fig.1-b). The cell of GDBD with maximum drainage basin area within each grid is chosen as a representative cell. Representative cells are marked with green colored small squares, and drainage area at each representative cell is indicated by numbers in Fig.1-b. Then, flow direction of each grid is decided toward the grid whose representative cell has the maximum drainage area among the eight neighboring grids.

3.3 Improved NSA by Doell (Doell)

Doell [2003] improved the Network Scaling Algorithm by constructing a medium resolution river map. In this method, every grid at the target resolution is divided into 9 sub-grids, and then Network Scaling Algorithm is applied to construct the medium resolution river map from GDBD

(Black solid line in Fig.1-c). Upper drainage area of medium resolution river map is recalculated using equation (1), and the sub-grid which has maximum drainage area within each grid is defined as representative sub-grid (green colored squares in Fig.1-c). Finally, the river channel of the medium resolution river map is traced towards downstream from each representative sub-grid, and downstream of each representative sub-grid is chosen as the downstream of each grid.

3.4 Double Maximum Method (DMM)

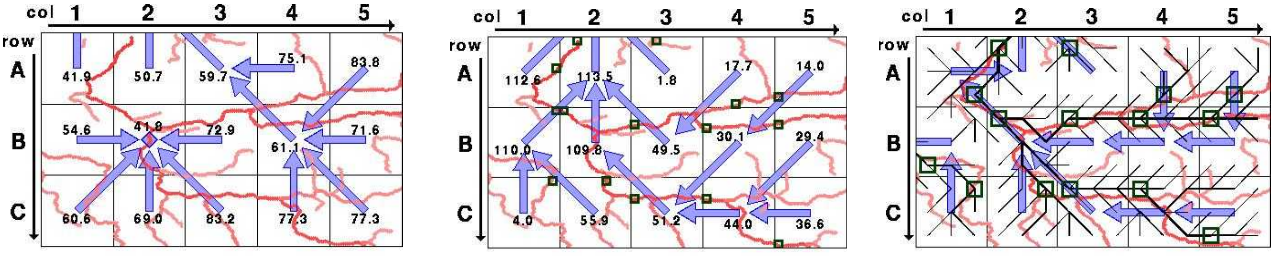
The Double Maximum Method [Olivera, 2002] defines the representative cell for each grid by the same way as the Network Scaling Algorithm (marked with green colored small squares in Fig.1-d). Then, Double Maximum Method defines the buffered area for each grid, whose width is a half of the target resolution. In case of the grid A4 in Fig.1-d, the area colored with grey is defined as buffered area. The river channel expressed by GDBD is traced towards downstream from each representative cell of the target grid, and flow direction of the target grid is decided when the traced river channel of GDBD get out of the buffered area.

3.5 Effective Area Method (DMM)

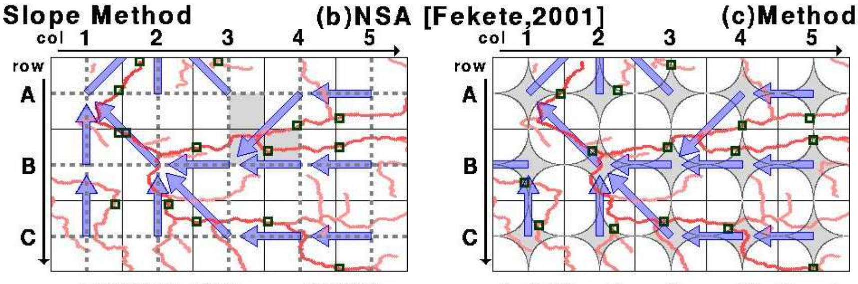
The Effective Area Method is a newly proposed upscaling method in this paper. It is the improved version of the Double Maximum Method, but the way of choosing representative cells is different. Firstly, effective area for each grid is defined by equation (3), (grey colored area in Fig.1-e).

$$\left\{(\varphi,\lambda)\,|\,(\varphi-\varphi_0)^{0.5}+(\lambda-\lambda_0)^{0.5}<R^{0.5}\right\} \tag{3}$$

Here, ϕ and λ are the longitude and latitude, ϕ_0 and λ_0 are the longitude and latitude of the center of each grid respectively, and R is the half of the grid size. Then the GDBD cell which has the maximum drainage area in effective area is chosen as the representative cell for each grid, and the river channel of GDBD is traced toward downstream from each representative cell. Flow direction of each grid is decided when traced river channel of GDBD enters to the effective area of neighboring eight grids.



(a)Steepest Slope Method (b)NSA [Fekete,2001] (c)Method by Döll [2002]



(d)DMM [Olivera,2002] (e)Effective Area Method

Fig.1 Flow direction maps derived by (a)SSM, (b)NSA, (c)Method by Doell, (d)DMM, (e)EAM.
Blue arrows are upscaled flow directions. Red colored line is the river channel in GDBD. Green squares are representative cells for each method. Numbers in (a) and (b) represent the grid average elevation and upper drainage area of representative cells respectively.

4. Validation of constructed river maps by each method

4.1 Characteristics of constructed river maps

A part of the upscaled global river map by each method (Odra River basin in Poland) is shown in Fig.1. Blue arrows represent flow directions of the constructed river map, and red lines represent the river channel derived from GDBD.

In case of the Steepest Slope Method, grid averaged elevation of grid B2 is lower than any neighboring grids that it is recognized as an inland depression. Such kind of unexpected inland depressions often appear because grid averaged elevation does not consider sub-grid scale topography like valleys or small hills. On the other hand, the Network Scaling Algorithm never produces such unexpected inland depressions due to use of super fine resolution river datasets instead of DEM as an input data. However, flow directions tend to gather towards the grid which has a large river inside, so boundaries of basins and sub-basins are not represented precisely.

The medium resolution river map in method by Doell contributes to create the proper basin boundaries, because large rivers are less likely to be allocated in neighboring grids in medium resolution. However, medium resolution is not fine enough to separate all river basins that basin merging occurs in some regions (See grid A4 and A5 in Fig.1-c). Instead of construct a medium resolution river map, the 1km resolution river channel of GDBD is directly traced in the Double Maximum Method, and it succeeded to avoid the unexpected basin merging. Upscaling efficiency of the Double Maximum Method is higher than other methods in this point but there still exist a significant error. If more than two rivers are allocated in a single grid, smaller rivers are merged into the largest river (see grid B1 in Fig.1-d).

Number of unexpected river merging due to existence of multiple rivers within one grid is reduced in the Effective Area Method, by introducing effective area into each grid. Because rivers just enters into the edge of the grid are neglected when choosing representative cells, multiple rivers in a single grid can be more often recognized separately.

4.2 Drainage area of upscaled river maps

Fig.2 illustrates the upper drainage area at representative cells calculated from GDBD and the drainage area calculated from the upscaled global river map for each grid. Grid area of the upscaled river map is derived by equation (1). Those two values are highly correlated if river network is upscaled correctly, but they differ if there are errors in upscaling processes. Underestimation of drainage area seen in both methods by Doell and the Double Maximum Method is result of unexpected river merging, where upstream of one river is deprived by another river. Overestimation in small drainage area in method by Doell is caused by merging of small basins due to errors in creating the medium resolution river map. Compared to those two methods, results by the Effective Area Method are clustered around one to one line, and this implicates errors in upscaling is less than previous methods.

Accuracy of upscaling can be statistically validated by modeling efficiency (ME) [Janssen, 1995], which is described as equation (4)

$$ME = \frac{\sum_{i=1}^{N}\left(A_{GDBD,i} - \overline{A}_{GDBD}\right)^2 - \sum_{i=1}^{N}\left(A_{T213,i} - A_{GDBD,i}\right)^2}{\sum_{i=1}^{N}\left(A_{GDBD,i} - \overline{A}_{GDBD}\right)^2} \tag{4}$$

Here, $A_{GDBD,i}$ is drainage area at each representative cell, $\overline{A}_{GDBD}$ is average of them, and $A_{T213,i}$ is drainage area calculated with the upscaled global river map at each grid. ME equals to one if upscaling is perfect, and it decreases if there are errors in the upscaled river maps. ME for method by Doell, the Double Maximum Method, and the Effective Area Method are 0.90, 0.69, and 0.99, respectively, and this also shows upscaling efficiency of the effective Area Method is the best among these three algorithms. The difference of ME between the Effective Area Method and others is result of the way of choosing representative cells, that is, the Effective Area Method has higher ability to recognize and separate two independent rivers located in a single grid by introducing effective area.

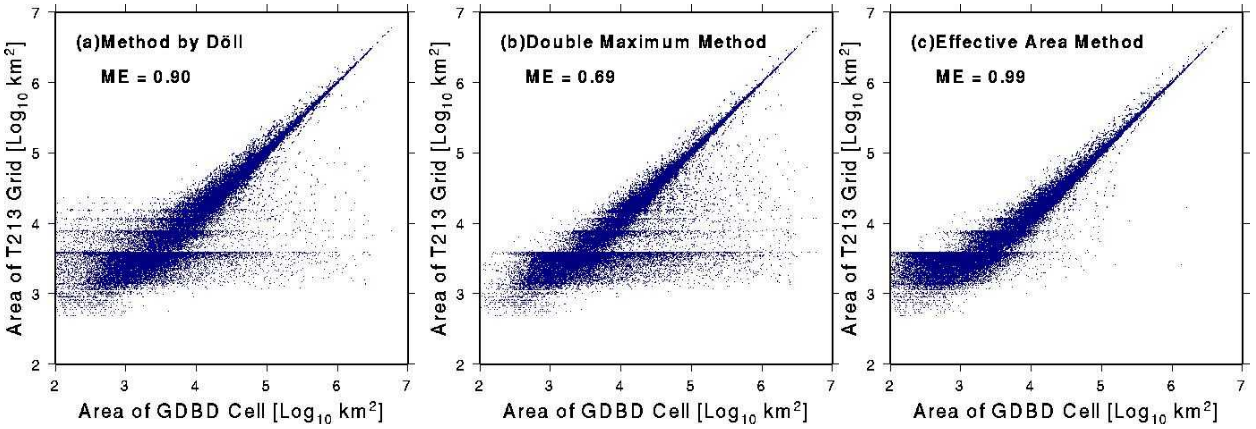


Fig.2 Drainage area of GDBD (horizontal) and that of the upscaled river map (vertical)
If there are no error in upscaling, drainage area of GDGD and that of the upscaled river map are highly correlated and plots are clustered around one to one line.

4.3 River channel length of upscaled river maps

Fig.3 illustrates cumulative river channel length (CRCL) of each flow direction calculated for the upscaled global river map derived by the Steepest Slope Method, the method by Doell, the Double Maximum Method, and the Effective Area Method. Channel length is calculated by the equation (2). CRCL should be evenly distributed for each direction if the upscaling method does not have bias to prefer certain flow directions. The Steepest Slope Method shows the most evenly distributed CRCL because the steepest slope is used as one absolute index to examine eight directions. Differences of CRCL between some directions seen in the Steepest Slope Method are considered to be the result of uneven terrain distribution such as location of mountains in continents. CRCL calculated for the method by Doell has longer bias for orthogonal directions, on the other hand CRCL for the Double Maximum Method and the Effective Area Method are relatively even. This implicates that the Double Maximum Method and the Effective Area Method can produce less biased flow directions compared to the method by Doell, thus impact of defining buffered area or effective area is large for less biased upscaling. Even tough both the Double Maximum Method and the Effective Area Method have tendency to choose orthogonal directions than diagonal directions compared to the Steepest Slope Method, the Effective Area Method shows the best fit to the Steepest Slope Method.

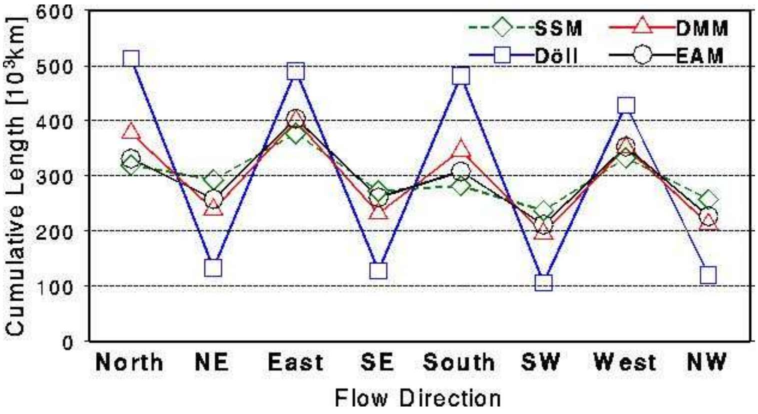


Fig.3 Cumulative river channel length for every flow direction

5. Summary and discussion

The Effective Area Method is validated to have better upscaling efficiency than previous methods, but it still has limitations to construct an acceptable global river map. Even though the ability to recognize and separate independent rivers located in a single grid is improved by introducing effective area, unexpected river merging occurs where two rivers are located in the effective area of one single grid. Example of regions with significant error is shown in Fig.4-a, where three rivers run parallel within quite close distance. The Effective Area Method fails to separate Mekong River from Salween River and Yangtze River, and as a result upstream of Mekong River is merged with other two rivers. In such regions, flow directions should be manually corrected by allowing one grid shift of river streams and the global river map have to be reconstructed like Fig.4-b. In the future, global river maps in much higher resolution will be required for more complex climate simulation or more accurate water resources assessment, so there is a need to develop an upscaling algorithm which does not require manual corrections.

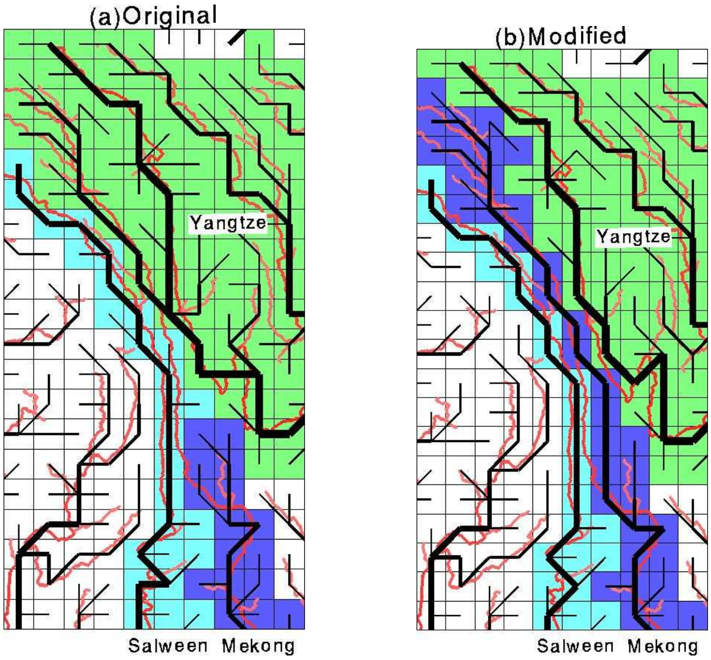


Fig.4. (a) The original and (b) the modified river map for Mekong, Salween, and Yangtze River basins

Acknowledgements

This research is part of the results of the Innovative Program of Climate Change Projection for the 21st Century (*KAKUSHIN Program*) by the Japanese Ministry of Education, Culture, Sports, Science, and Technology, and it is also supported by the Grant in-Aid for Scientific Research (S, Representative: Taikan Oki Ph.D) by Japan Society for the Promotion of Science.

References

Costa-Cabral M. C., and Burges S. J.: Digital elevation model networks (DAEMON): A model of flow over hill slopes for computation of contributing and dispersal areas, *Water Resources research*, **30**, 1681-1692, 1994

Doell P., and Lehner B.: Validation of a new global 20-min drainage direction map. *Journal of Hydrology* **258**: 214-231, 2002

Fekete B. M., Vorostarty C. J., and Lammers R. B.: Scaling gridded river networks for macroscale hydrology: Development, analysis, and control of error. *Water Resource Research* **37**: 1955-1967. 2001

Hanasaki N., Kanae S., Oki T., Masuda K., Motoya K., Shirakawa N., Shen Y., and Tanaka K.: An integrated model for the assessment of global water resources - Part 1: Model description and input meteorological forcing, *Hydrology and Earth System Science*, **12**, 1007-1025, 2008.

Hirabayashi Y., Kanae S., Struthers I., and Oki T.: A 100-year (1901-2000) global retrospective estimation of the terrestrial water cycle, *Journal Geophysical. Research*, **110**, D19101, doi:10.1029/2004JD005492, 2005

Janssen P. H. M., and Heubergar P. S. C.: Calibration of process oriented models. *Ecological Modelling* **83**: 55-66, 1995

Masutomi Y., Inui Y., Takahashi K., and Matsuoka Y.: Development of highly accurate global polygonal drainage basin data. *Accepted to Hydrological Processes*. 2007 (available online at http://www-cger.nies.go.jp/cger-e/db/enterprise/gdbd/gdbd_index_e.html)

Miller J. R., Russell G. L., and Caliri G.: Continental-Scale River Flow in Climate Models, *Journal of Climate*, **7**, pp914-928, 1994

Oki T., Nishimura T., and Dirmeyer P.: Assessment of assessment of annual runoff from land surface models using Total Runoff Integrating Pathways (TRIP). *Journal of meteorological Society of Japan* **77**: 235-255. 1999

Oki T., and Sud Y. C.: Design of Total Runoff Integrating Pathways (TRIP) – A global river channel network. *Earth interactions* **2**: 1-36. 1997

Olivera F., Lear M. S., Famiglietti J. S., and Asante K.: Extracting low-resolution river networks from high-resolution digital elevation models. *Water Resource Research* **38**: 1231, doi:10.1029/2001WR000726. 2002

USGS: United States Geological Survey. GTOPO30 (available online at http://edc.usgs.gov/products/elevation/gtopo30/gtopo30.html)

Vorosmarty C. J., Fekete B. M., Meybeck M. M., Lammers R. B.: Geomorphometric attributes of global system of rivers at 30-minute spatial resolution. *Journal of Hydrology* **237**: 17-39. 2000