Extracting low-resolution river networks from high-resolution digital elevation models

Francisco Olivera,1 Mary S. Lear,2 James S. Famiglietti,3 and Kwabena Asante4

Received 14 June 2001; revised 25 March 2002; accepted 3 April 2002; published 13 November 2002.

[1] Including a global river network in the land component of global climate models (GCMs) is necessary in order to provide a more complete representation of the hydrologic cycle. The process of creating these networks is called river network upscaling and consists of lowering the resolution of already available fine networks to make them compatible with GCMs. Fine-resolution river networks have a level of detail appropriate for analysis at the watershed scale but are too intensive for global hydrologic studies. A river network upscaling algorithm, which processes fine-resolution digital elevation models to determine the flow directions that best describe the flow patterns in a coarser user-defined scale, is presented. The objectives of this study were to develop an algorithm that advances the previous work in the field by being applicable at a global scale, allowing for the upscaling to be performed in a projected environment, and generating evenly distributed flow directions. INDEX TERMS: 1848 Hydrology: Networks; 1860 Hydrology: Runoff and streamflow; 9810 General or Miscellaneous: New fields (not classifiable under other headings); KEYWORDS: river network, surface water, global, continental, geographic information systems (GIS), digital elevation models (DEM)

Citation: Olivera, F., M. S. Lear, J. S. Famiglietti, and K. Asante, Extracting low-resolution river networks from high-resolution digital elevation models, Water Resour. Res., 38(11), 1231, doi:10.1029/2001WR000726, 2002.

1. Introduction

[2] Modeling of the hydrologic cycle over large and complex systems, involving land, atmospheric and ocean processes, requires the completion of the horizontal link between the land and ocean components. The consensus of numerous authors who have developed large-scale surface flow routing algorithms [Vörösmarty et al., 1989; Henderson-Sellers et al., 1993; Liston et al. 1994; Miller et al., 1994; Nijssen et al., 1997; Hagemann and Dümenil, 1998; Coe, 1998; Olivera et al., 2000; Arora et al., 2001] supports this concept, although, as pointed out by Miller et al. [1994], most global climate models (GCMs) have neglected the transport of water from the land to the ocean. Among other effects, land-ocean interaction causes changes in the thermohaline circulation because of freshwater influx from rivers to the oceans [Kasse et al., 1994; Kasse & Walsh, 1996; Kassem et al., 1998]. As well, accounting of land-atmosphere interaction in GCMs has allowed, for example, to identify increments in regional precipitation and latent heat flux over northern Africa when considering increased oceanic temperature as compared to less surface water [Coe & Bonan, 1997; Coe, 1997], to better match simulated and observed temperatures when considering a 1% increase in surface temperature (S.T. Graham and J.S. Famiglietti, 1994), and to predict runoff increases for the majority of the rivers in a simulation with increased carbon dioxide levels [Miller and Russell, 1992].

[3] Including a global river network in the land component of GCMs is necessary in order to provide a more complete representation of the hydrologic cycle. The process of creating these networks consists of lowering the resolution of already available (in vector and raster format) fine networks and translating them into coarse networks compatible with the resolution of the GCMs (e.g., approximately 2.8°). This process is called river network upscaling. Fine-resolution river networks have a level of detail appropriate for analysis at the watershed scale, but are too intensive for global scale hydrologic studies; whereas global river networks have the level of detail necessary for analysis at the global scale.

[4] In this paper, a river network upscaling algorithm, which processes fine-resolution digital elevation models (DEMs) to determine the flow directions that best describe the flow patterns in a coarser user-defined scale, is presented. The main objective of this study was to develop an algorithm that advances the previous work in the field by: (1) being applicable at a global scale, (2) allowing for the upscaling to be performed in a projected environment (i.e., a flat representation of the curved surface of the Earth), and (3) generating evenly distributed flow directions. The algorithm was evaluated by comparing the resulting river networks to the O’Donnell et al. [1999] upscaling method results and the Arora and Boer [1999] global river network.

2. Background

[5] In the last decade, many researchers have addressed the problem of determining flow directions at a coarse-
3. Methodology

3.1. Double Maximum Method (DMM)

The algorithm’s name of double maximum method (DMM) was given because the river network is determined after identifying the locations with maximum drainage area within the boxes of two coarse-resolution grids.

Although the physical processes involved in large-scale hydrologic systems, such as those at a continental scale, take place on the curved surface of the Earth, for modeling purposes this curved surface is mapped as a flat surface by using a mathematical algorithm. Different algorithms result in different ways to map a specific area, and correspond to different map projections. In the conversion from curved to flat, the terrain features represented in the map undergo distortion that affects, for example, the length of lines, and/or the area and shape of polygons. Before starting the modeling of a hydrologic process, a projection of the study area into a flat surface is necessary. Selection of the map projection usually depends on the size, shape and location of the study area.

GCMs have traditionally used control volumes that can be defined in terms of longitude and latitude. However, although longitude and latitude coordinates, also known as geographic coordinates, can be used to map the curved surface of the Earth over a flat surface (at the expense of a significant distortion close to the poles), it is not considered a map projection. Thus it is said that GCMs are implemented on a nonprojected environment. Surface water modeling, on the other hand, requires implementing a map projection that, depending on the specific physical process to be represented, preserves lengths, areas or shapes. Surface water modeling is done on a projected environment. Interaction between these two modeling environments requires the capability to convert the results of one environment into the other.

In this paper, in addition to upscaling the resolution, the river networks are converted from projected to non-projected coordinates for global climate modelers to use. The upscaling process occurs in the projected environment, the environment of the original fine-resolution river network, and the resulting river networks are then projected to geographic coordinates to match the coarse-resolution grids of the GCMs.

The DMM requires two inputs: a fine-resolution DEM in projected coordinates, and a user-defined coarse-resolution output grid in geographic coordinates. In the following, the elements of the DEM and its derivatives are called cells, whereas the elements of the coarse-resolution grid are called boxes. The coarse-resolution river network is defined by identifying the downstream box of each box of the grid.

The resolution of the input raster river network is the resolution of the DEM, usually much finer than that required by GCMs. Using DEMs as input, flow direction grids (FDR) and flow accumulation grids (FAC) can be defined. The FDR stores in each cell a pointer to the cell toward which water flows. The FDR is determined using the eight-direction pour point model, which assumes that water flows in the direction of the steepest decent [Jenson and Domingue, 1988]. The FAC stores in each cell the number of upstream contributing cells. The FAC is calculated using the FDR. As expected, FAC values increase along the river channels, reaching a maximum at the river mouth.

The coarse-resolution grid is defined in geographic coordinates because it responds to the needs of GCMs, and can have any user-defined resolution, size, and orientation. Although it is customary to define square boxes oriented North-South and East-West, the DMM supports the use of nonsquare boxes with other orientations. Once the coarse-resolution grid is defined in geographic coordinates, it is projected to the same projection as the DEM, so that it overlays the DEM, FDR and FAC.
A second coarse-resolution grid, identical to the first grid but offset half a box in both directions (see Figure 1), is generated automatically. Figure 1 shows how the offset grid (thin line) subdivides the boxes of the first grid (thick line) into four equal quadrants.

The upscaling procedure is summarized in Figures 2a–2d. In Figures 2a–2d, A refers to the boxes of the first coarse-resolution grid, and B refers to those of the offset grid. The first step is to determine where the river exits each box of the first grid. The exit location is defined by the FAC cell with the greatest value. Note that, even though most of a box area might be draining in one direction, the algorithm takes into account only the maximum FAC cell value which could correspond to a river conveying drainage from a large area upstream and which just entered the box. In Figure 2a, the exit location for box A1 is located in the southeast quadrant. Next, box B1 is selected since it contains the exit location of box A1, as seen in Figure 2b. The method for determining the exit location of box B1 is the same as that described.

Figure 1. Coarse-resolution grids. The first grid is shown in thick lines, and the offset grid is in thin lines.

Figure 2. The upscaling procedure of the DMM: (a) identification of the river exit location for each grid A box, (b) selection of the grid B box that corresponds to the exit location of the grid A box, (c) selection of the grid A box that corresponds to the exit location of the grid B box, and (d) connection of the center points of the two grid A boxes selected in Figures 2a and 2c.
for A1. The offset grid allows the algorithm to track the river after it exits the boxes of the first grid. In the third step, shown in Figure 2c, box A2 was selected since it contains the exit location determined for box B1. The final step links box A1 to its downstream box, box A2, as seen in Figure 2d. The flow direction for box A1 is toward the southeast. The same flow direction determination process is applied to each box in the first grid. The coarse-resolution river network is formed after connecting the center point of each box in the first grid to the center point of its downstream box.

[18] Tracking the river beyond the boundary of the box provides a better representation of the river’s overall flow direction, and allows for prediction of flow directions not only through the box sides but also through the corners. For example, a model that does not track the flow after it leaves the box would have predicted an east flow direction (rather than southeast) for box A1. The tracking procedure of the DMM is similar to that of O’Donnell et al. [1999]; however, it was found that dividing the boxes into four, rather than nine subboxes, decreases the tendency to predict more flow directions through the sides than through the corners. O’Donnell et al. [1999], in fact, considered different relative subbox sizes, and the case of four subboxes is a particular case of their more general case of nine subboxes.

[19] Finally, the coarse river network is projected to geographic coordinates so that it can be used with GCMs at a global scale.

3.2. Evaluation of the Double Maximum Method (DMM)

[20] Quantitative and qualitative evaluations of the resulting upscaled river networks, like the calculation of the error in the area of the predicted watersheds, the estimation of the method’s tendency to predict more flow through the sides than through the corners, and the comparison of the predicted network to observed networks and drainage areas, are necessary to assess the suitability of the method.

[21] When comparing drainage areas, it should be taken into account that inclusion in the catchment of areas that do not belong to it, leads to an overestimation of the drainage area; and similarly, exclusion of areas leads to an underestimation. Statistically, these errors tend to offset each other generating relatively small overall errors. Success in drainage area prediction therefore should not be interpreted necessarily as success in drainage area delineation, because the predicted catchment might have the correct area and an incorrect shape.

[22] Evaluation of the flow direction distribution consists of estimating its bias toward predicting more flow in some directions than in others. According to the eight-direction pour-point model [Jenson and Dominigue, 1988], four flow directions correspond to the cell or box sides and four to the corners. Although in some areas the topography of the terrain might induce a preference to flow in a specific direction, it is unlikely that this would happen over all the drainage area. Even more, when analyzing the entire globe and according to the law of large numbers [DeGroot, 1986], it can be assumed that water has no preference to flow in any specific direction. Thus predicted flow directions should be evenly distributed. In the case of upsampling flow direction algorithms, usually all four sides are treated equally, and preference to predict flow toward one specific side is not expected. The same concept can be claimed for the corners. However, an artificial preference to predict more flow toward the sides as opposed to the corners, or vice versa, can be generated by the upsampling algorithms.

[23] An even flow direction distribution generates, not the same number of boxes flowing through their sides and corners, but the same overall flow length through the orthogonal and diagonal paths. Note that because flow through the corners runs for the length of a box diagonal (1.41 times the box size), whereas flow through the sides for the length of a box side (1.00 times the box size), an even flow direction distribution should generate 41% (i.e., 1.41–1.00) more boxes flowing through the sides than through the corners. Thus an even distribution should have 59% (i.e., 1.41/1.41 + 1.00) of cells flowing through the sides and 41% (i.e., 1.00/1.41 + 1.00) through the corners. The relation between these percentages is called here side-to-corner ratio, which for an even distribution is equal to 59/41.
The DMM was written in Arc Macro Language (AML), the programming language of ArcGIS Workstation developed by the Environmental Systems Research Institute (ESRI).

4. Application, Results, and Discussion

The DMM was applied at a global scale using as inputs the HYDRO1K DEM and derivatives [Gesch et al., 1999] with a resolution of one kilometer and in Lambert Azimuthal Equal-Area map projection, and a coarse-resolution grid that consisted of $2.8125^\circ$ longitude $\times$ $2.8125^\circ$ latitude boxes (which subdivides the Earth into 8192 boxes, 64 rows by 128 columns). Since the map projection parameters of HYDRO1K are different for each continent, the upscaling process was performed separately for each continent, then transformed each of them into geographic map projection, and finally assembled all the pieces together into a single global data set.

For results at a global scale, the upscaled river networks of the whole world are shown in Figure 3. In Figure 3 the main river systems of the world are identifiable. Note that no attempt at manual correction of the derived river network has been made, the goal of this paper being to demonstrate the power of the new automated methodology. Clearly, manual correction will very likely be required with any upscaling algorithm to achieve an optimal fit between the observed and extracted networks.

For results at a continental scale, the upscaled river networks of the African continent, including the corresponding HYDRO1K DEM based watershed boundaries, are presented in Figure 4. Comparison of the river networks with the watershed boundaries in Figure 4 is consistent because both have the HYDRO1K DEM data set as the basic definition of the terrain. In Figure 4 the river networks for the large river systems are located mostly within the watershed boundaries, whereas small river systems with watershed areas less than one box often become incorporated into the neighboring watersheds. These inaccuracies can always be improved by decreasing the size of the coarse-resolution grid boxes.

For results at a large-watershed scale, the upscaled river networks of the Congo and Niger River systems, including their corresponding HYDRO1K DEM based watershed boundaries, their river networks in vector format [ESRI, 1992], and the coarse-resolution grid boxes included in their drainage area, are presented in Figures 5 and 6. Both Figures 5 and 6 show that the coarse-resolution river networks follow the river patterns and define the drainage boundaries well. However, it can be noted that some boxes with more than half of their area outside the basin have been
included, while some others with more than half of their area inside have been missed by the coarse network. This situation, however, should not be interpreted as a problem because it may respond to the distribution of boxes between more than two drainage systems, or to the presence of large rivers with a small drainage area in a specific box but with a significant drainage area upstream.

[29] In the following, the DMM is contrasted with the Arora and Boer [1999] upscaled global river network and with results from the O’Donnell et al. [1999] river network upscaling method. In this case, in addition to a qualitative assessment, predicted catchment areas and flow direction distributions were compared. Comparison of river lengths was not considered because length depends on the grid resolution and systematic errors in predicting lengths do not necessarily imply limitations in the upscaling methods.

[30] In Figures 7 and 8, observed river networks of the Congo and Niger River systems [ESRI, 1992] were compared to upscaled river networks developed by Arora and Boer and O’Donnell et al.. Overall, it can be seen that, given the limitations placed by the coarse-resolution, the river

Figure 7. Upscaled river networks for the Congo River system, with coarse-resolution boxes, HYDRO1K DEM based watershed boundary, and river network in vector format, according to (a) Arora and Boer [1999] and (b) O’Donnell et al. [1999].

Figure 8. Upscaled river networks for the Niger River system, with coarse-resolution boxes, HYDRO1K DEM based watershed boundary, and river network in vector format, according to (a) Arora and Boer [1999] and (b) O’Donnell et al. [1999].
networks developed by Arora and Boer and O’Donnell et al. follow well the river patterns and drainage area boundaries. The borders of the river network’s boxes shown in Figures 5–8 represent the coarse-resolution watershed boundary. In terms of catchment area, the actual areas of the Congo and Niger River basins are 3,730,000 km² and 2,260,000 km², respectively [Revenga et al., 1998]. However, the predicted areas are 4,080,000 km² (error of +9%) and 2,660,000 km² (+11%) according to the DMM; and 3,800,000 km² (+2%) and 2,180,000 km² (−9%) according to Arora and Boer, as shown in Table 1. An explanation for the consistency in the overprediction of catchment areas is that upscaling algorithms tend to incorporate small drainage systems into the large ones, where small is defined in relation to the size of the coarse-resolution boxes. Catchment areas according to O’Donnell et al. are not included because their algorithm was applied only to those DEM cells located within a previously delineated basin.

The upscaled river networks generated by the DMM had side-to-corner ratios of 68/32 for the globe, 64/36 for Africa, and 62/38 for both the Congo and Niger River systems. Likewise, river networks generated by Arora and Boer had 72/28 for Africa, 79/21 for the Congo and 70/30 for the Niger, while those generated by O’Donnell et al. (using a ratio of noncorner subbox side to box side of 0.01) had side-to-corner ratios of 74/26 for the Congo and 72/28 for the Niger, as shown in Table 2. It can be noted that each of the tested river networks has more side flow directions than diagonals by more than the expected 59/41 side-to-corner ratio. This bias is a limitation of all three data sets, although the DMM shows the smallest bias.

To more clearly illustrate the bias of the DMM to predict more flow through the sides than corners, it was applied to the African and South American continents with the same fine-resolution DEM data, but with the coarse grid rotated 45°; thus switching the side and corner directions (i.e., north, east, south and west are now corners instead of sides, and northeast, southeast, southwest and northwest are now sides instead of corners). The resulting flow direction distribution showed similarly side-biased results, which suggests that the DMM itself creates a biased distribution. After rotating the coarse grid, the side-to-corner ratio changed only slightly from 64/36 to 68/32 for Africa and from 63/37 to 62/38 for South America. Note that these results imply that the coarse grid orientation, and not only the terrain topography defined by the DEM, is determining the upscaled flow directions. The rotated river network for Africa is shown in Figure 9.

5. Conclusions

The land surface component of the hydrologic cycle is an integral part of GCMs. Although global DEMs will be available in the future at increasingly finer resolutions, due to the complexity of the processes represented in GCMs, coarse-resolution river networks are necessary. Therefore an algorithm and a computer program to resample fine-resolution flow direction data into coarse resolution were developed. The algorithm was called double maximum method (DMM). In this paper, these algorithm and computer program are described and evaluated. The resulting river networks of the DMM reproduce well the river network pattern and the watershed boundaries of observed fine-resolution data. Compared to other coarse-resolution river networks, the DMM produces similar or better results.

One advantage of the DMM with respect to other methods is that it extracts, from fine-resolution flow direction data defined in a projected environment, river networks on a coarse grid defined in a nonprojected environment (i.e., longitude and latitude); thus allowing for interaction between surface water and global climate modelers. Another advantage of the DMM is that it advances in reducing the bias toward predicting more flow through the sides than corners. However, it still shows a bias, though lower than that observed in other methods.

Acknowledgments. This research was funded by NSF grants DGE-9454099 and ATM 9617980. This support is gratefully acknowledged. The authors would also like to thank B. Nijssen with the Department of Civil Engineering and Engineering Mechanics and the Department of Hydrology and Water Resources of the University of Arizona, and A. Berg

| Watershed | Observed | DMM | Arora and Boer |
|-----------|----------|-----|----------------|
| Congo     | 3.73     | 4.08| 3.80           |
| Niger     | 2.26     | 2.66| 2.18           |

| Area       | Even     | DMM     | Arora and Boer | O’Donnell et al. |
|------------|----------|---------|----------------|------------------|
| Globe      | 59/41    | 68/32   | 72/28          |                  |
| Africa     | 59/41    | 64/36   | 79/21          |                  |
| Congo      | 59/41    | 62/38   | 70/30          | 74/26            |
| Niger      | 59/41    | 62/38   | 70/30          | 72/28            |

Figure 9. Upscaled African river network developed with the DMM, with coarse-resolution grid rotated 45°.
with the Department of Geological Sciences of the University of Texas at Austin, for their comments and input.

References

Arora, V. K., and G. J. Boer, A variable velocity flow routing algorithm for GCMs, J. Geophys. Res., 104(D24), 30,965–30,979, 1999.

Arora, V. K., F. Seglenieks, N. Kouwen, and E. Soulis, Scaling aspects of river flow routing, Hydrool. Processes, 15(3), 461–477, 2001.

Coe, M. T., Simulating continental surface waters: An application to Holocene Northern Africa, J. Clim., 10, 1680–1689, 1997.

Coe, M. T., A linked global model of terrestrial hydrologic processes: Simulation of modern rivers, lakes, and wetlands, J. Geophys. Res., 103(D8), 8885–8899, 1998.

Fekete, B. M., C. J. Vorosmarty, and R. B. Lammers, Scaling gridded river networks for macroscale hydrology: Development, analysis, and control of error, Water Resour. Res., 37(7), 1955–1967, 2001.

Gesch, D. B., K. L. Verdin, and S. K. Greenlee, New land surface digital elevation model covers the Earth, Eos Trans. AGU, 80(6), 69–70, 1999.

Graham, S. T., J. S. Famiglietti, and D. R. Maidment, Five-minute, 1/2°, and 1° data sets of continental watersheds and river networks for use in regional and global hydrologic and climate system modeling studies, Water Resour. Res., 35(2), 583–587, 1999.

Hagemann, S., and L. Dümenil, A parameterization of the lateral waterflow for the global scale, Clim. Dyn., 14, 17–31, 1998.

Henderson-Sellers, A., Z. L. Yang, and R. E. Dickinson, The project for intercomparison of land-surface parameterization schemes, Bull. Am. Meteorol. Soc., 74, 1335–1349, 1993.

Jenson, S. K., and J. O. Dominigue, Extracting topographic structure from digital elevation data for Geographic Information System analysis, Photogramm. Eng. Remote Sens., 54, 1593–1600, 1988.

Kassens, H., I. Dmierenko, V. Rachold, J. Thiede, and L. Timokhov, Russian and German scientists explore the Arctic’s Laptev sea and its climate system, Eos Trans. AGU, 79(27), 1998.

Liston, G., Y. Sud, and E. Wood, Evaluating GCM land surface hydrology parameterizations by computing river discharges using a runoff model: Application to the Mississippi Basin, J. Appl. Meteorol., 33, 394–405, 1994.

Miller, J., and G. Russell, The impact of global warming on river runoff, J. Geophys. Res., 97(D3), 2757–2764, 1992.

Miller, J., G. Russell, and G. Caliri, Continental scale river flow in climate models, J. Clim., 7, 914–928, 1994.

Nijssen, B., D. P. Lettenmaier, X. Liang, S. U. Wörlz, and E. F. Wood, Streamflow simulation for continental-scale river basins, Water Resour. Res., 33(4), 711–724, 1997.

O’Donnell, G., B. Nijssen, and D. P. Lettenmaier, Simple algorithm for generating streamflow networks for grid-based, macroscale hydrological models, Hydrol. Processes, 13, 1269–1275, 1999.

Oki, T., and Y. C. Sud, Design of total runoff integrating pathways (TRIP)—A global river channel network, Earth Inter., 2, 1998. (Available at http://EarthInteractions.org)

Olivera, F., J. S. Famiglietti, and K. Asante, Global-scale flow routing using a source-to-sink algorithm, Water Resour. Res., 36(8), 2197–2207, 2000.

Olesen, H., and J. Knoop, A global river routing network for use in hydrological modeling, J. Hydrol., 230, 230–243, 2000.

Revenga, C., S. Murray, J. Abramowitz, and A. Hammond, Watersheds of the world, Water Resour. Inst., Washington, D.C., 1998.

Sausen, R., S. Schubert, and L. Dümenil, A model of river runoff for use in coupled atmosphere-ocean models, J. Hydrol., 155, 337–352, 1994.

Vörösmarty, C. J., B. Moore III, A. L. Grace, M. P. Gildea, J. M. Melillo, B. J. Peterson, E. B. Rastetter, and O. A. Steudler, Continental scale models of water balance and fluvial transport: An application to South America, Global Biogeochem. Cycles, 3(3), 241–265, 1989.

Weatherly, J. W., and J. E. Walsh, The effects of precipitation and river runoff in a coupled ice-ocean model of the Arctic, Clim. Dyn., 12, 785–798, 1996.