Title
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

Journal
Water Resources Research, 35(2)

ISSN
00431397

Authors
Graham, S. T
Famiglietti, J. S
Maidment, D. R

Publication Date
1999-02-01

DOI
10.1029/1998WR900068

WATER RESOURCES RESEARCH, VOL. 35, NO. 2, PAGES 583–587, FEBRUARY 1999

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

S. T. Graham and J. S. Famiglietti

Department of Geological Sciences, The University of Texas at Austin

D. R. Maidment

Center for Research in Water Resources, The University of Texas at Austin

Abstract. A major shortcoming of the land surface component in climate models is the absence of a river transport algorithm. This issue becomes particularly important in fully coupled climate system models (CSMs), where river transport is required to close and realistically represent the global water cycle. The development of a river transport algorithm requires knowledge of watersheds and river networks at a scale that is appropriate for use in CSMs. These data must be derived largely from global digital topographic information. The purpose of this paper is to describe a new data set of watersheds and river networks, which is derived primarily from the TerrainBase 5′ Global DTM (digital terrain model) and the CIA World Data Bank II. These data serve as a base map for routing continental runoff to the appropriate coast and therefore into the appropriate ocean or inland sea. Using this data set, the runoff produced in any grid cell, when coupled with a routing algorithm, can easily be transported to the appropriate water body and distributed across that water body as desired. The data set includes watershed and flow direction information, as well as supporting hydrologic data at 5′, 1/2°, and 1° resolutions globally. It will be useful in fully coupled land-ocean-atmosphere models, in terrestrial ecosystem models, or in stand-alone macroscale hydrologic-modeling studies.

1. Introduction

In this paper we describe a digital data set of watersheds and river networks for the entirety of Earth's land surfaces. These data can be used in a broad range of Earth system science studies, including investigations of regional and global hydrology, climate, biogeochemistry, and ecosystem dynamics. The data are presented in gridded format at three different resolutions (5′, 1/2°, and 1°) and were derived primarily from the National Geophysical Data Center TerrainBase Global DTM, version 1.0 [*Row et al.*, 1995], and the CIA World Data Bank II [*Gorney and Carter*, 1987]. The following data layers are included at each resolution: flow direction, flow accumulation, a delineation of the global river network, a discretization of the land surface into 55 watersheds, a discretization of internally draining basins, a coarser discretization of Earth's surface into 19 drainage regions (aggregated watersheds) and associated ocean basins into which they drain, the same 19 drainage regions excluding ocean basins, lake delineations, and a land/sea mask. Additionally, annual average runoff data is provided for all of the 55 basins. The data set presented is similar to that recently described at 1° resolution by *Oki and Sud* [1998] but extends that work by providing higher resolution 1/2° and 5′ components as well as additional information including streamflow and lake data. In addition to enabling higher resolution modeling studies, the 1/2° and 5′ data will also enhance lower resolution studies, since they will provide detailed information from which the parameterization of subgrid features can be derived.

Paper number 1998WR900068.
0043-1397/99/1998WR900068$09.00

The flow direction data layer describes the paths followed by the global river network system by delineating the directions that water will flow across the entirety of the land. This data layer is required to implement cell-to-cell river-routing algorithms [*Vörösmarty et al.*, 1989; *Kite et al.*, 1994; *Liston et al.*, 1994; *Miller et al.*, 1994; *Sausen et al.*, 1994; *Coe*, 1997]. Flow accumulation describes the number of upgradient grid cells, or those cells that drain through a particular grid cell (determined from the flow direction data layer), and therefore the likelihood of a river existing at that cell. By specifying a threshold value of flow accumulation, river networks can be digitally extracted from the digital elevation model (DEM). In this work specific river courses were derived from the flow direction and flow accumulation information using the threshold values described below. While the primary data layer required for river transport implementation is flow direction, knowledge of the location of major river reaches may aid in the development of variable velocity transport models which utilize differing intra- and extra-channel flow equations.

Earth's large river basins (those with areas of 5000 5′ grid cells (about 500,000 km^2 near the equator) or more) were then delineated from the river and flow direction data. The major internally draining basins of Australia, Asia, and Africa are also provided. Given these basin boundaries, hydrologic fluxes can be aggregated to determine watershed-specific responses and compared to observations. The data set also includes a division of the entirety of Earth's surface into 19 large drainage regions and the associated ocean or marginal or internal seas into which they discharge. This coarse discretization is included for use in land-ocean-atmosphere modeling studies, some of which require low-resolution coupling (via continental runoff) at the land-ocean margins [*Bryan et al.*, 1996]. A separate data layer of the 19 drainage regions without the associated water bodies is also provided, as is a land/sea mask. A discretization of lakes, as defined by the CIA World Data Bank II [*Gorney and Carter*, 1987], is included for use in global hydrologic and climate models that can accommodate lakes [*Bonan*, 1995]. Annual average streamflow observations are provided for these large watersheds as well as for several additional river basins included in the study of climate model generated runoff by *Russell and Miller* [1990].

2. Sources and Potential Uses

The majority of the data presented here have been derived from the TerrainBase [*Row et al.*, 1995] elevation data, which is a 5′ resolution comprehensive global coverage of elevation data for both the land surfaces and bathymetry of the ocean floors. Some validation and correction has been performed, specifically on the flow direction data, using the CIA World Data Bank II [*Gorney and Carter*, 1987] river and lake delineations distributed with The Generic Mapping Tools, version 3.0 [*Wessel and Smith*, 1995]. Annual average streamflow values have been taken from *Perry et al.* [1996] and the *United Nations Educational, Scientific, and Cultural Organization*, (*UNESCO*) [1974].

This data set will enable a wide range of river transport modeling activities in support of regional and global hydrological, terrestrial ecosystem, and climate modeling studies. These include the representation of river flow in climate system models (CSMs) and global climate models (GCMs) and the transport of water, sediment, and nutrients in terrestrial ecosystem models.

3. Data Analysis Methods

The following steps are required to delineate river networks and watersheds: (1) determination of a land/sea mask, (2) geolocation of rivers, (3) filling of artificial depressions, (4) calculation of flow directions, (5) calculation of flow accumulations, (6) selection and delineation of rivers, and (7) selection and delineation of watersheds. Each of these steps is described briefly below. The software employed in the analysis of these data was the Environmental Systems Research Institute, Inc. (ESRI), ARC/INFO version 7.1.2, and ARC GRID package [*ESRI*, 1997], as well as The Generic Mapping Tools, version 3.0 [*Wessel and Smith*, 1995].

1. Determination of land/sea mask. The portions of the TerrainBase DEM for hydrologic analysis were defined by creating a land-versus-sea mask and analyzing only the land areas of the globe. The mask was created by differentiating between land and ocean cells, where ocean cells were identified as those cells with an elevation less than or equal to zero.

2. Geolocation of rivers. A delineation of rivers extracted from the CIA data was used to ensure that river channels ultimately identified with the DEM would be properly geolocated. The CIA data set was assumed accurate, as it is a very high resolution vector database of river channels and streams with more detail than can be resolved by the 5′ grid cells in TerrainBase. Discrepancies arise between actual river channels and those delineated by a DEM because of the averaging that occurs when gridding elevation data. These discrepancies can be minimized by geolocating the CIA delineation on the DEM and decreasing the elevation of all DEM grid cells which underlie the rivers. Analysis of the resulting conditioned DEM, in which the differences in elevation between river cells and all other land cells has been exaggerated, yields a more realistic representation of overland flow into river channels, and hence of river channel paths across continental landscapes, since flow directions (see below) are "forced" towards the lower elevation CIA river locations.

3. Filling of artificial depressions. Artificial depressions are often created when digital elevation data is gridded. In order to ensure that runoff can make its way to the proper water body or waterway without draining into these artificial "pits," they must be filled. The method used to fill these pits was an iterative process that identified all enclosed depressions and increased the elevation of all cells within each depression until they matched the lowest adjacent outlet so that water could flow continuously and uninterrupted across the land surface. In addition to these false depressions, the large areas of Asia, Australia, and Africa that are truly internally draining (or without a river outlet to the oceans) were also identified and excluded from the infilling process.

4. Calculation of flow directions. The direction that surface water flows must be calculated so that river networks can be identified. The direction of flow was calculated on the basis of the direction of steepest descent. Antarctica was treated differently to ensure valid flow directions. If flow was not to the north, northwest, or northeast originally, it was assigned a value of north [*Oki and Sud*, 1998].

5. Calculation of flow accumulations. Flow accumulation is the number of cells that drain through a given cell and is determined from the flow direction data. These data are then used in step 6 in order to determine the location of river channel cells.

6. Selection and delineation of rivers. Rivers were delineated as those cells that met a certain flow accumulation criterion. A file of river channels corresponding to grid cells that drain 500 5′ grid cells as a threshold, or about 50,000 km^2, has been included for the watersheds described in step 7. Smaller threshold values, which would include an increasing number of grid cells with smaller flow accumulation values, are more appropriate for finer resolution investigations and larger threshold values for coarser resolution studies when smaller stream networks are not as important.

7. Selection and delineation of watersheds. Watersheds were delineated as those cells that drained to a particular river outlet, the source cell, as delineated in step 6. The watershed source cells were defined as those cells which corresponded to the rivers determined in step 6, as well as the complement of the 33 rivers suggested by *Russell and Miller* [1990] for use in global climate modeling studies, for a total of 55 watersheds. The 19 larger drainage regions were also derived in a similar manner; the source cells in this case, however, were a set of coastal cells.

In addition to rivers and watersheds, lakes were also delineated from the CIA data and gridded at a 5′ resolution. All cells containing lakes were considered to be 100% covered by lakes at the 5′ resolution. Annual streamflow observations are included for the 55 rivers described above and were extracted from the data of *Perry et al.* [1996] and from *UNESCO* [1985].

All of the analyses described above were first conducted with the 5′ DEM (Plate 1). The DEM was then resampled to 1/2°

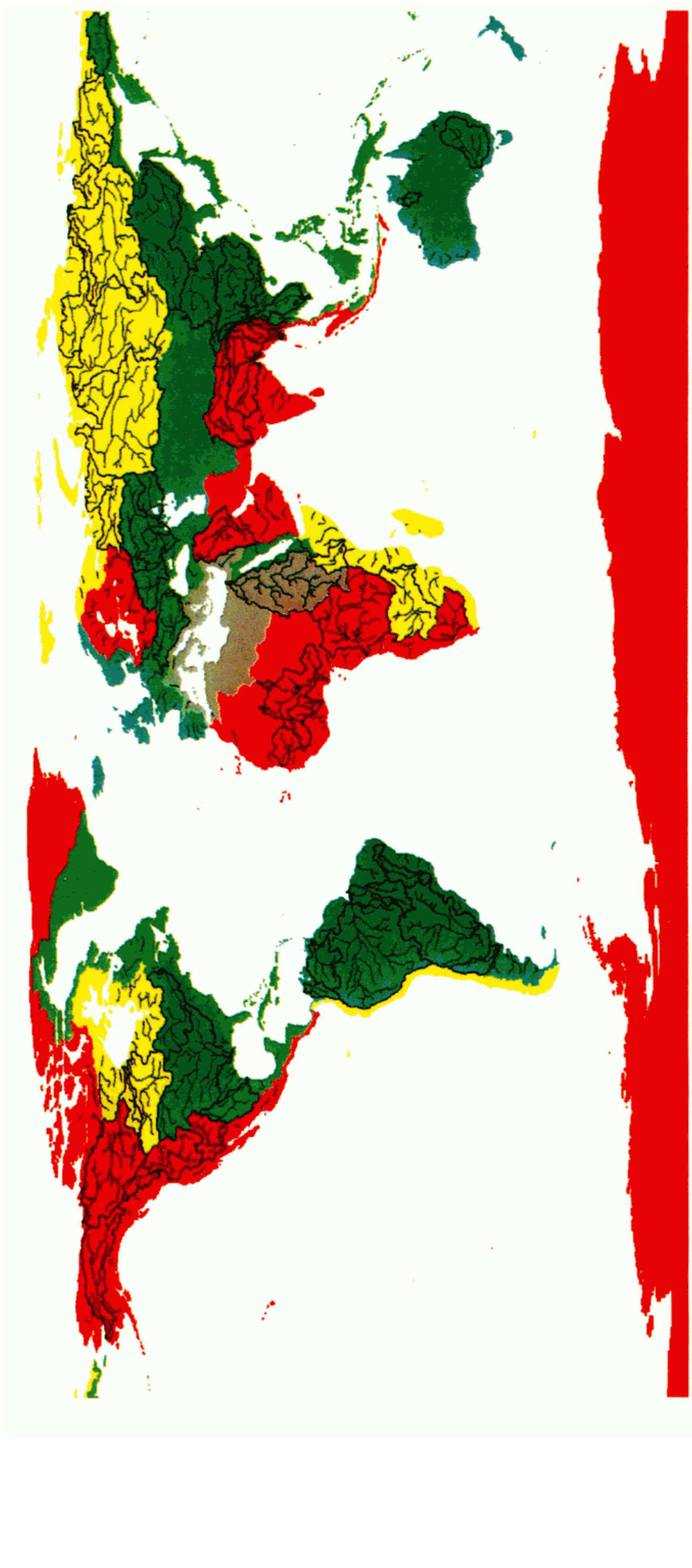

Plate 1. Nineteen global drainage regions (in color) with boundaries of the 55 large watersheds (in black) and a river delineation (in blue) using a 500 grid cell threshold. All presented at a 5′ resolution.

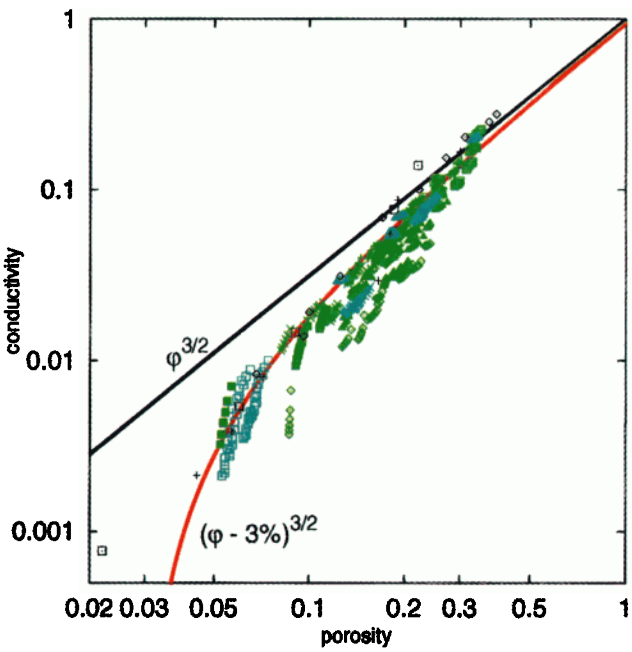

Plate 1. The electrical conductivity (relative to the brine conductivity) as a function of porosity. The data points are taken from *Hausenblas* [1995] (green and blue points) and from *Wong et al.* [1984] (black points). The Hausenblas data include measurements in four different outcrop sandstones (blue) and nine different reservoir sandstones (green). Archie's law and its modification are given by equations (1) and (3), in which $a = 1$ and $m = 3/2$.

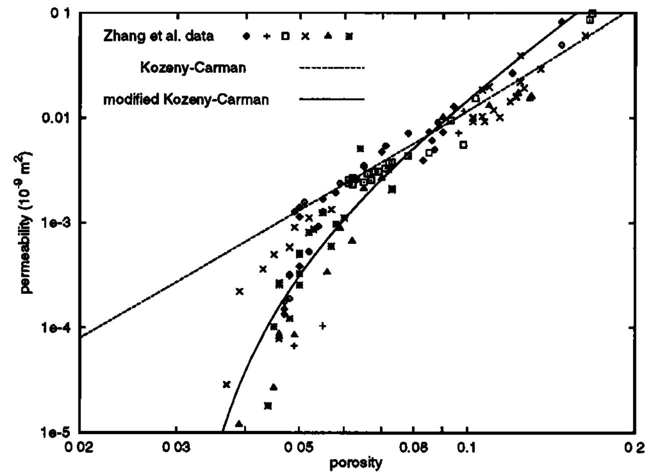

Figure 2. The permeability as a function of porosity. The data points are taken from *Zhang et al.* [1994], the Kozeny-Carman equation and its modification are given by equations (2) and (4), in which $D = 1.3$ μm and $D = 2.5$ μm were used, respectively.

4. Theoretical Investigations

Several investigators have attempted to describe the low-porosity behavior of granular systems on theoretical grounds. *Schwartz and Banavar* [1989], for instance, performed random walk simulations to calculate conductivity and permeability in models of consolidated granular systems. The results compared well with measurements from *McLachlan et al.* [1987] and *Guyon et al.* [1987]. Further fluid flow computations were done by [*Martys et al.*, 1994] using a finite difference scheme for flow in packings of nonoverlapping and overlapping spheres. They introduced a percolation threshold of 3% as an Ansatz, or "working hypothesis," for their scaling law to describe the data. *Zhu et al.* [1995] divided the experimental data into two regimes, namely, above and below an experimentally determined "crossover" porosity. They formulated a network model

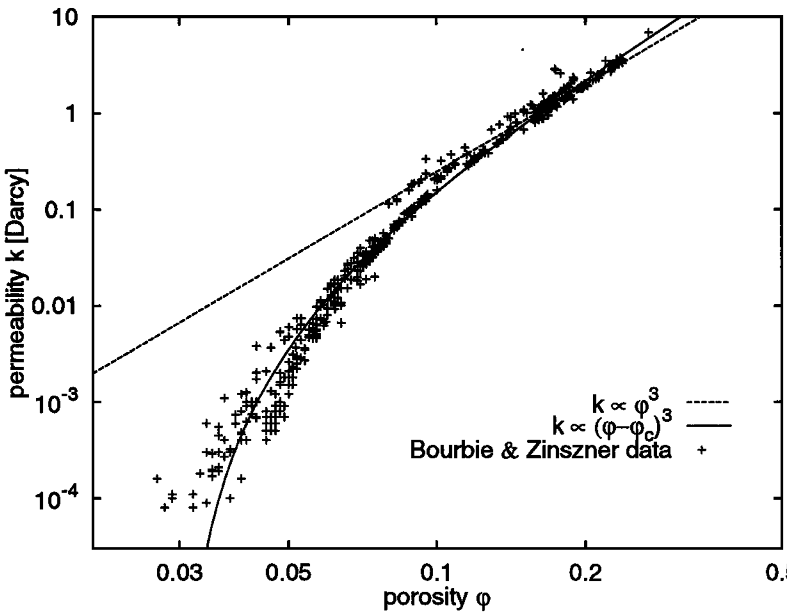

Figure 1. The permeability as a function of porosity. The data points are from *Bourbie and Zinszner* [1985], plus additional data points measured by the same researchers since then. The dashed and solid line are given by the Kozeny-Carman equation (equation (2)) and its modification (equation (4)). In the first $D = 210$ μm was used; in the second $D = 280$ μm.

nomena. Most importantly, it will provide a template for the implementation of river transport schemes in regional and global climate and hydrological models. The inclusion of river transport in climate models effectively closes the global water cycle and will provide a framework for understanding the role of rivers and the outflow of freshwater from the continents on Earth's climate system, as well as potential climate change impacts on the redistribution of the world's freshwater resources. Including river transport in large-scale climate and hydrologic models will also provide for the implementation of sediment and nutrient transport algorithms, allowing for more realistic simulation of erosion and the delivery of biogeochemical materials to the ocean. Finally, the watershed and river network data, combined with the streamflow data, provide a means for validating climate and hydrologic model simulations of terrestrial hydrology.

7. Summary

A data set of continental watersheds and river networks has been presented at three resolutions (5′, 1/2°, and 1°). At each resolution the data set contains a layer of flow direction, flow accumulation, a delineation of the global river network, a discretization of the land surface into 55 watersheds, a discretization of internally draining basins, a coarser discretization of Earth's surface into 19 drainage regions (aggregated watersheds) and associated ocean basins into which they drain, the same 19 drainage regions excluding ocean basins, lake delineations, and a land/sea mask. A separate file of average annual discharge data for the 55 watersheds is also provided. The data will be useful for implementing river transport and more realistic hydrology in climate models, validating climate model output, as well as in a range of modeling and analysis studies in Earth system science.

Acknowledgments. This work was sponsored by NSF grant ATM-9617980 and by NASA Earth System Science Fellowship NGT5-30008. This research support is gratefully acknowledged. We also thank Francisco Olivera for his helpful comments and assistance with data processing.

References

Bonan, G. B., Sensitivity of a GCM simulation to inclusion of inland water surfaces, *J. Clim.*, *8*, 2691–2704, 1995.

Bryan, F. O., B. G. Kauffman, W. G. Large, and P. R. Gent, The NCAR CSM flux coupler, *Tech. Note NCAR/TN-424+STR*, Natl. Cent. for Atmos. Res., Boulder, Colo., 1996.

Coe, M. T., Simulating continental surface waters: an application to Holocene Northern Africa, *J. Clim.*, *10*(7), 1680–1689, 1997.

Environmental Systems Research Institute, Inc., ARC/INFO version 7.1.2, Redlands, Calif., 1997.

Gorney, A. J., and R. Carter, *World Data Bank II General User's Guide*, Cent. Intel. Agency, Washington, D. C., 1987.

Hornberger, G. M., Data and analysis note: A new type of article for *Water Resources Research*, *Water Resour. Res.*, *30*(12), 3241–3242, 1994.

Kite, G. W., A. Dalton, and K. Dion, Simulation of streamflow in a macroscale watershed using general circulation model data, *Water Resour. Res.*, *30*(5), 1547–1559, 1994.

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

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

O'Donnell, G., B. Nijssen, and D. P. Lettenmaier, A simple algorithm for generating streamflow networks for grid-based, macroscale hydrological models, *Hydrol. Process.*, in press, 1999.

Oki, T., and Y. C. Sud, Design of Total Runoff Integrating Pathways (TRIP): A global river channel network, *Earth Inter.*, *2*, EI013, 1998. (Available at http://EarthInteractions.org.)

Perry, G. D., P. B. Duffy, and N. L. Miller, An extended data set of river discharges for validation of general circulation models, *J. Geophys. Res.*, *101*(D16), 21,339–21,349, 1996.

Row, L. W., D. A. Hastings, and P. K. Dunbar, *TerrainBase Worldwide Digital Terrain Data Documentation Manual*, Natl. Geophys. Data Cent., Boulder, Colo., 1995.

Russell, G. L., and J. R. Miller, Global river runoff calculated from a global atmospheric general circulation model, *J. Hydrol.*, *117*, 241–254, 1990.

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

United Nations Educational Scientific and Cultural Organization, *Discharge of Selected Rivers of the World*, vols. I, II, III (parts I, II, III, IV), Paris, France, 1985.

U.S. Geological Survey, GTOPO30, http://edcwww.cr.usgs.gov/landdaac/gtopo30/gtopo30.html, Earth Resour. Obs. Syst. Data Cent., Sioux Falls, S. D., 1996.

U.S. Geological Survey, HYDRO1k, http://edcwww.cr.usgs.gov/landdaac/gtopo30/hydro/, Earth Resour. Obs. Syst. Data Cent., Sioux Falls, S. D., 1998.

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

Wessel, P., and W. H. F. Smith, The Generic Mapping Tools (GMT) version 3.0 technical reference and cookbook, Natl. Oceanic and Atmos. Admin., Silver Spring, Md., 1995.

J. S. Famiglietti and S. T. Graham, Department of Geological Sciences, The University of Texas, Austin, TX 78712. (jfamiglt@maestro.geo.utexas.edu)

D. R. Maidment, Center for Research in Water Resources, The University of Texas, Austin, TX 78712.

(Received May 4, 1998; revised October 15, 1998; accepted October 20, 1998.)