9-2012

# A new global river network database for macroscale hydrologic modeling

Huan Wu

John S. Kimball
*University of Montana - Missoula*

Hongyi Li

Maoyi Huang

L. Ruby Leung

*See next page for additional authors*



## Recommended Citation

Wu, H., J. S.Kimball, H.Li, M.Huang, L. R.Leung, and R. F.Adler (2012), A new global river network database for macroscale hydrologic modeling, Water Resour. Res., 48, W09701, doi:10.1029/2012WR012313

# A new global river network database for macroscale hydrologic modeling

Huan Wu,[1,2] John S. Kimball,[3,4] Hongyi Li,[5] Maoyi Huang,[5] L. Ruby Leung,[5] and Robert F. Adler[1,2]

Received 23 April 2012; revised 10 July 2012; accepted 14 August 2012; published 28 September 2012.

[1] Coarse-resolution (upscaled) river networks are critical inputs for runoff routing in macroscale hydrologic models. Recently, Wu et al. (2011) developed a hierarchical dominant river tracing (DRT) algorithm for automated extraction and spatial upscaling of river networks using fine-scale hydrography inputs. We applied the DRT algorithms using combined HydroSHEDS and HYDRO1k global fine-scale hydrography inputs and produced a new series of upscaled global river network data at multiple (1/16° to 2°) spatial resolutions. The new upscaled results are internally consistent and congruent with the baseline fine-scale inputs and should facilitate improved regional to global scale hydrologic simulations.

**Citation:** Wu, H., J. S. Kimball, H. Li, M. Huang, L. R. Leung, and R. F. Adler (2012), A new global river network database for macroscale hydrologic modeling, *Water Resour. Res.*, *48*, W09701, doi:10.1029/2012WR012313.

## 1. Introduction

[2] River networks at coarse resolutions are critical inputs to macroscale hydrologic models for representing lateral-movement processes, including flow path delineations for runoff routing and flow accumulation. There have been increasing efforts over the past decade to develop automatic algorithms for river network upscaling from relatively fine scale hydrography data [e.g., *Fekete et al.*, 2001; *Döll and Lehner*, 2002; *Olivera and Raina*, 2003]. However, these methods tend to underutilize baseline fine-scale hydrography information and tend to promote distortions, which generally require intensive manual corrections to avoid potential significant, negative impacts on hydrologic modeling. *Wu et al.* [2011, hereiafter referred to as W2011] recently proposed a hierarchical dominant river tracing (DRT) algorithm for fully automatic upscaling of river networks that addresses many of the limitations of earlier methods. The DRT algorithm was initially applied to produce a series of global hydrography data sets from 1/16° to 2° spatial scales using HYDRO1k (U.S. Geological Survey (USGS), http://eros.usgs.gov/#/Find_Data/Products_and_Data_Available/gtopo30/hydro) fine-scale hydrography inputs (W2011). A detailed description of the DRT algorithms and DRT-derived product accuracy in relation to other available methods, including NSABE [*Fekete et al.*, 2001] and DDM30 [*Döll and Lehner*, 2002] approaches, is provided by W2011. This study also reported more comprehensive and global validations of the DRT results against the baseline fine-scale hydrography inputs, indicating that the DRT has robust performance independent of spatial scale and geographic region.

[3] Here we report on a new multiscale global hydrography data set derived using the current version of the DRT algorithms and improved baseline hydrography inputs combined from HydroSHEDS (<60°N) and HYDRO1k (≥60°N). The resulting global land products are provided in a consistent (WGS84) projection and range of spatial scales from 1/16° to 2°, and include flow direction, river network, upstream drainage area, and river length delineations.

## 2. Data and Methodology

[4] The current version (i.e., version 1.1) of the DRT algorithms has been updated from W2011 to improve computing efficiency. The HYDRO1k database has limitations over some regions (e.g., relatively flat lowlands). As the successor of HYDRO1k, HydroSHEDS is now available for many regions and is purported to provide superior scale and quality relative to its predecessor [*Lehner et al.*, 2008]. As HydroSHEDS currently does not include high-latitude areas (i.e., regions above 60°N), we combined the HydroSHEDS and HYDRO1k fine resolution (i.e., 30 arc sec or ~1 km) databases to create merged global baseline DRT inputs by using the northern portion of HYDRO1k to fill areas currently not covered by HydroSHEDS (Figure 1, top). Manual corrections were performed during the baseline data integration process to ensure consistent flow paths across boundaries between the two data sets. Hereafter, the combined 1 km resolution global HydroSHEDS/HYDRO1K hydrography is referred to as the combined baseline.

[5] The same metrics as by W2011 were used to evaluate the new DRT upscaled basin geometry calculations against the combined baseline, including modeling efficiency (ME),

[1]Earth System Science Interdisciplinary Center, University of Maryland, College Park, Maryland, USA.
[2]NASA Goddard Space Flight Center, Greenbelt, Maryland, USA.
[3]Flathead Lake Biological Station, Division of Biological Sciences, University of Montana, Polson, Montana, USA.
[4]Numerical Terradynamic Simulation Group, College of Forestry and Conservation, University of Montana, Missoula, Montana, USA.
[5]Pacific Northwest National Laboratory, Richland, Washington, USA.

Corresponding author: H. Wu, Earth System Science Interdisciplinary Center, University of Maryland, College Park, MD 20742, USA. (huanwu@umd.edu)


0043-1397/12/2012WR012313

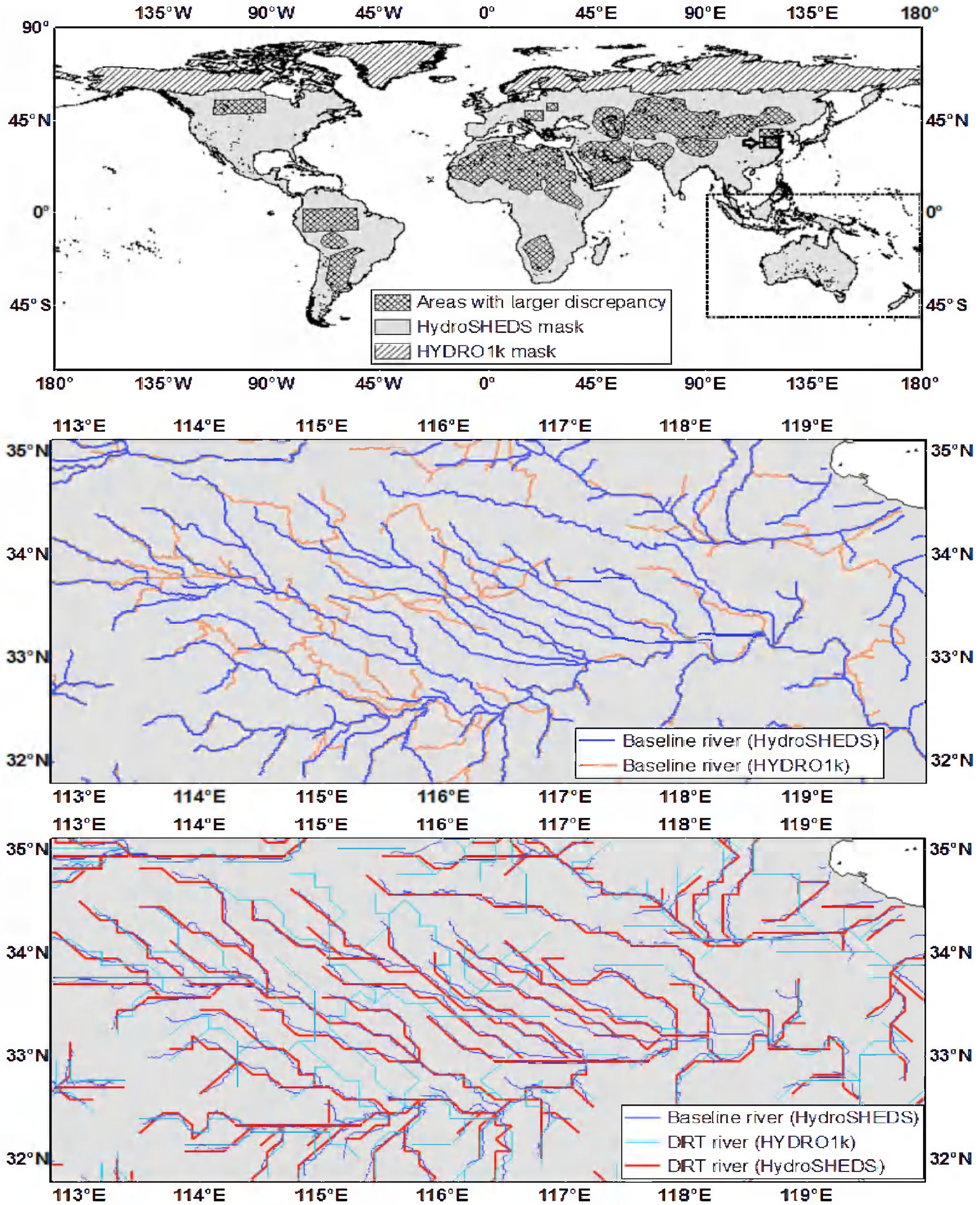


**Figure 1.** (top) The global study domain showing boundaries between HYDRO1k and HydroSHEDS areas. The crosshatched areas within the HydroSHEDS domain are the areas where the discrepancy between the two baseline data sets is relatively large. The example domain (rectangle with arrow in Figure 1 (top)) shows (middle) HYDRO1k and HydroSHEDS defined differences in baseline fine-scale river networks and (bottom) the resulting DRT upscaled river networks.

normalized RMSE (NRMSE), and mean absolute relative error (MRE) statistics (see details in W2011). We also evaluated the new DRT results against the previous HYDRO1k-based DRT upscaling database from W2011.

## 3. Results

### 3.1. Comparison of HYDRO1k and HydroSHEDS

[6] Differences between HYDRO1k and HydroSHEDS will be inherited in the DRT results during the upscaling process. Comparison of the baseline river networks from HYDRO1k and HydroSHEDS at their native 1 km (30 arc sec) resolution indicated relatively small differences over North America, but larger differences for other areas of the globe (not shown). Almost all rivers (including those of North America) from HydroSHEDS have some degree of geolocation shift in relation to the corresponding rivers from HYDRO1k (also seen from *Lehner et al.* [2008]). Although these relative distortions generally have no significant impact on the DRT upscaled results at coarser resolutions, many areas showing larger differences lead to significant differences in the DRT upscaled river networks.

[7] We performed comparisons between HydroSHEDS and HYDRO1k in terms of the numbers of basins and rivers, as they will be reflected in the DRT results. The comparisons excluded northern land areas (>60°N) where HydroSHEDS is unavailable and portions of Australia and Southern Asia (dashed rectangle in Figure 1 (top)) where HYDRO1k is unavailable. We compared the numbers of the basins with variable sizes (basin sizes in Table 1), numbers of basin main stem rivers at variable lengths (stem river lengths in Table 1) and the number of tributaries at variable river orders (river orders in Table 1) between the two baseline data sets. The Strahler river orders are defined starting from headwater cells (i.e., grid cells without upstream cells), which are coded as "first-order" rivers and are not included in the statistics. From Table 1, HydroSHEDS defines more river basins, main stem rivers, and major tributaries (i.e., rivers with orders less than ninth order) than HYDRO1k. For basins with drainage areas >500 km$^2$, the number of basins by HydroSHEDS (i.e., 13,286) is almost twice of that from HYDRO1k (7379; Table 1) for the same domain. However, HYDRO1k tends to define a greater number of larger basins and more higher-order rivers (i.e., order 9, Table 1).

[8] Figure 1 (top) shows regions with the most significant differences between HYDRO1k and HydroSHEDS. Larger discrepancies seem to occur more often in flat areas (e.g., Sahara desert), probably due to HydroSHEDS being based on a superior digital elevation model than HYDRO1k [*Lehner et al.*, 2008]. The DRT-derived results show significant differences in these areas correspondingly. For example, Figure 1 (bottom) shows the derived upscaled (1/8°) river networks for the region indicated as a rectangle with an arrow in Figure 1 (top), where there are large differences in the baseline river networks between HYDRO1k and HydroSHEDS (Figure 1, middle). Such discrepancies in DRT results resulted from the differences in baseline hydrography data sets (Figure 1, bottom) tend to increase as the upscaling spatial resolution becomes higher.

### 3.2. Global Evaluation of the New Upscaled Hydrography Results

[9] We followed the same method from W2011 to evaluate the new upscaled results against the combined baseline inputs, which are referred to as the baseline "observation". The same rules as in W2011 were used to select basins, stem rivers and major tributaries for comparisons. Table 2 shows all of the metrics calculated for the results in this study.

#### 3.2.1. Evaluation of DRT-Derived Basin Area

[10] The number of global basins evaluated ranged from 907 (2° resolution) to 65,289 (1/16° resolution) using the combined baseline. The DRT results indicate that basin areas are effectively preserved across all spatial scales relative to the combined baseline for all upscaling resolutions ($R^2 \geq 0.97$; $p < 0.0001$). The NRMSE differences between the DRT upscaled results and combined baseline ranged from 0.04% (1/8° resolution) to 0.47% (2° resolution), while MRE differences ranged from 7.9% (1/16° resolution) to 2.1% (2° resolution). Both NRMSE and MRE terms vary subtly across all upscaling levels (Table 2). From the global statistics, 20,212 of the 65,289 selected basins

**Table 1.** Global Comparison of HYDRO1k Versus Combined Baseline Hydrography in Terms of Number of Basins and Rivers

| | Basin Size | | | | | |
|---|---|---|---|---|---|---|
| Comparison of Basins | >10,000 km$^2$ | >5000 km$^2$ | >1000 km$^2$ | >500 km$^2$ | >100 km$^2$ | >50 km$^2$ |
| Number of basin outlets in combined | 1400 | 2456 | 8267 | 13,286 | 34,097 | 48,888 |
| Number of basin outlets in HYDRO1k | 1022 | 1659 | 4778 | 7379 | 20,567 | 32,696 |
| | Stem River Length | | | | | |
| Comparison of Basin Stem Rivers | >1000 km | >500 km | >100 km | >50 km | >25 km | >10 km |
| Number of stem rivers in combined | 106 | 357 | 3563 | 8018 | 15,106 | 26,719 |
| Number of stem rivers in HYDRO1k | 96 | 279 | 2289 | 4746 | 8804 | 16,061 |
| | Strahler River Order | | | | | |
| Comparison of All Rivers | 11–12 | 9–10 | 8 | 7 | 6 | 2–5 |
| Number of stem rivers in combined | 19,782 | 216,303 | 378,828 | 794,431 | 1,636,410 | 48,993,489 |
| Number of stem rivers in HYDRO1k | 21,429 | 229,647 | 356,435 | 753,488 | 1,508,962 | 44,030,464 |

**Table 2.** Global Comparison of HydroSHEDS/HYDRO1k Combined Baseline Hydrography Versus DRT-Derived Basin Area, Lengths of Stem Rivers and Major Tributaries, and Basin Shapes

| | 2° | 1° | 1/2° | 1/4° | 1/8° | 1/16° |
|---|---|---|---|---|---|---|
| *Comparison of Basin Areas* | | | | | | |
| Basins with variable sizes | | | | | | |
| Basin size | >10,000 km$^2$ | >5000 km$^2$ | >1000 km$^2$ | >500 km$^2$ | >100 km$^2$ | >50 km$^2$ |
| Number of basins | 907 | 2072 | 6469 | 13,277 | 35,469 | 65,289 |
| NRMSE | 0.47% | 0.17% | 0.07% | 0.08% | 0.04% | 0.2% |
| MRE | 2.1% | 3% | 2.8% | 4.0% | 4.4% | 7.9% |
| ME | 1.00 | 1.00 | 1.00 | 1.00 | 1.00 | 0.97 |
| Basins with drainage area between 5000 and 50,000 km$^2$ | | | | | | |
| Number of basins | 523 | 1651 | 2360 | 2508 | 2447 | 2486 |
| NRMSE | 2.7% | 4.5% | 3.6% | 3.2% | 4.0% | 4.5% |
| MRE | 0.7% | 2.8% | 3.9% | 3.4% | 2.5% | 3.6% |
| ME | 0.98 | 0.96 | 0.97 | 0.98 | 0.96 | 0.95 |
| Number of resolvable basin outlets | 879 | 2004 | 6392 | 12,628 | 33,266 | 60,065 |
| Number of basin outlets in baseline | 1742 | 3147 | 10,981 | 17,850 | 48,116 | 71,691 |
| *Comparison of River Lengths* | | | | | | |
| Rivers (major tributaries) with variable lengths | | | | | | |
| River length | >100 km | >50 km | >50 km | >20 km | >15 km | >10 km |
| Number of rivers | 2085 | 7411 | 18,646 | 75,405 | 178,865 | 339,237 |
| NRMSE | 0.7% | 0.3% | 0.3% | 0.03% | 0.02% | 0.02% |
| MRE | 1.6% | 1.5% | 1.9% | 2.5% | 4.0% | 5.3% |
| ME | 0.99 | 1.00 | 0.99 | 1.00 | 1.00 | 1.00 |
| Rivers (major tributaries) with length between 20 and 200 km | | | | | | |
| Number of rivers | 1856 | 7561 | 26,880 | 72,033 | 133,283 | 167,989 |
| NRMSE | 8.1% | 4.8% | 3.5% | 1.9% | 1.2% | 0.7% |
| MRE | 0.6% | 1% | 1.7% | 2.6% | 3.5% | 3.1% |
| ME | 0.95 | 0.96 | 0.97 | 0.99 | 0.99 | 1.00 |
| *Comparison of Basin Shapes* | | | | | | |
| Basins with drainage area greater than 1000 km$^2$ | | | | | | |
| Number of basins | 866 | 1882 | 5776 | 7960 | 8440 | 8920 |
| MRE | 1.9% | 2.8% | 2.7% | 4.2% | 3.0% | 2.3% |
| ME | 0.98 | 0.96 | 0.94 | 0.91 | 0.96 | 0.99 |

(31%) are smaller than 100 km$^2$ at 1/16° resolution and show a MRE of 10.7%. In contrast, 9963 basins are larger than 1000 km$^2$ at 1/16° resolution, account for 15% of the total selected basins and have a MRE of 3.9%. For all basins with areas between 5000 and 50,000 km$^2$, the comparisons of basin areas between DRT upscaled and combined baseline show the largest MRE (3.9%) at 1/2° resolution and the lowest MRE (0.7%) at 2° resolution (Table 2).

[11] The number of basin outlets defined at the baseline fine resolution, and according to the basin area thresholds (basin size in Table 2) are shown under the number of basin outlets in baseline category in Table 2. The upscaling process should maximize the preservation of these outlets (basins). However, it is not possible to preserve all of these basins (basin outlets) during the upscaling process, especially at coarser spatial resolutions, because when multiple river outlets (mouths) defined from the baseline inputs are located in the same coarse-resolution grid cell, this grid cell can only be assigned to a single upscaled basin (thus a single outlet) consistent with the D8 single flow method (W2011). Hence, when multiple rivers end in a same coarse cell, the DRT defines the coarse grid cell as the outlet cell of the river with the largest drainage area because larger rivers are prioritized over smaller rivers (W2011). The number of resolvable basin outlets in Table 2 shows the number of coarse cells that contain all of the outlets defined from the fine-scale baseline hydrography, which are smaller than the number of basin outlets in baseline results, particularly for relatively coarser resolutions. For example, globally there are 1742 basin outlets with drainage areas >10,000 km$^2$ defined at the baseline fine-scale resolution (1 km) while all of these basin outlets are located in only 879 grid cells at the 2° resolution. However, the DRT is able to preserve relatively more basins (number of basins in Table 2) during spatial upscaling by reverse tracing of secondary rivers within outlet cells to recover some river mouths and sinks when the associated river basins are important/large enough (W2011).

### 3.2.2. Evaluation of DRT-Derived River Lengths

[12] We conducted global comparisons of the DRT-derived river lengths for basin main stem rivers and major tributaries in the selected basins (section 3.2.1) across all scales relative to the combined baseline. The number of selected main stem rivers and tributaries ranged from 339,237 (1/16° resolution) to 2085 (2° resolution). The global comparison (Table 2) indicates that the total lengths of DRT upscaled rivers and tributaries are well preserved across all spatial scales relative to the combined baseline ($R^2 \geq 0.99$; $p < 0.0001$), with the NRMSE < 1% for all upscaling levels, while MRE differences range from 1.5 to 5.3 percent. For all rivers with lengths between 20 and 200 km, the comparisons of river lengths between the DRT and combined baseline results indicate consistent DRT performance across all spatial scales in this size category, with the largest MRE (3.5%) at 1/8° resolution from 133,283

rivers selected, and the lowest MRE (0.6%) at 2° resolution from 1856 rivers selected (Table 2).

### 3.2.3. Evaluation of DRT-Derived Basin Shapes

[13] Basin shape indices were calculated for the same set of selected basins (section 3.2.1) and compared with the combined baseline. These results (Table 2) indicate favorable DRT performance in preserving basin shapes for basins with drainage areas >1,000 km$^2$ for all spatial scales, with MRE differences ranging from 1.9% (1° resolution) to 4.2% (1/4° resolution).

### 3.2.4. Evaluation of the New DRT Results Against the W2011 Results

[14] The metrics in Table 2 are directly comparable to Table 6 in W2011 and the latter was previously derived based on HYDRO1k. Overall, the above evaluation metrics (Table 2) are similar to those derived for the earlier HYDRO1k based DRT results (W2011). From W2011, for all rivers with lengths between 20 and 200 km, the MRE of river lengths between the DRT and HYDRO1k baseline results ranges from 0.54% at 2° resolution to 3.5% at 1/16° resolution (Table 6 in W2011), while the MREs of basin shape for basins with drainage areas greater than 1000 km$^2$ are between 2.41% (1° resolution) and 4.63% (1/4° resolution).

[15] However, the numbers of basins and rivers/major tributaries selected for evaluation from the new DRT results are significantly larger than that from the HYDRO1k-based DRT results, which are predominantly due to (1) a larger number of basins and rivers represented from HydroSHEDS relative to HYDRO1k (Table 1) and (2) additional inclusion of the Australia/Southern Asia domain.

## 4. Conclusions

[16] A new set of global coarse-resolution river networks have been defined at multiple spatial scales (from 1/16° to 2°) by applying the DRT upscaling algorithms (W2011) using combined fine-scale baseline hydrography inputs from HydroSHEDS [*Lehner et al.*, 2008] and HYDRO1k (USGS, http://eros.usgs.gov/#/Find_Data/Products_and_Data_Available/gtopo30/hydro). The new upscaled global hydrography data set includes upscaled flow direction, river network, upstream drainage area, and river length parameters required for runoff routing and river discharge calculations in macroscale hydrological modeling. The new DRT upscaled results were globally evaluated against the combined HydroSHEDS/HYDRO1K baseline fine-scale (1 km resolution) hydrography. The results indicate robust DRT performance relative to the baseline hydrography; the DRT algorithm preserves the baseline hydrography including river shape and length, basin shape and area, and internal drainage structure, with globally consistent performance across the different spatial scales.

[17] Improved baseline hydrography inputs enable greater accuracy in DRT upscaled river networks, which in turn would facilitate better accuracy in regional and macroscale hydrological model simulations that utilize these data [W2011; H. Li et al., A physically based runoff routing model for land surface and Earth system models, submitted to *Journal of Hydrometeorology*, 2012]. The new DRT results translate these improvements in HydroSHEDS into more accurate upscaled hydrography layers relative to an earlier DRT record defined from HYDRO1k (W2011). The improvements include the quality of upscaled flow direction, drainage area, and river length calculations. These improvement may be potentially beneficial to other parameters that are critical to hydrological models such as drainage density, channel geometry, Manning's roughness coefficient etc. The DRT algorithm is largely automated and can be efficiently applied using any baseline hydrography information; additional updates to the DRT global data sets may occur as higher quality baseline hydrography data become available. The DRT upscaled global hydrography data sets generated from this study are available through the UMT online data archives (ftp://ftp.ntsg.umt.edu/pub/data/DRT/).

[18] **Acknowledgments.** This work was conducted at the University of Montana (UMT) and Earth System Science Interdisciplinary Center (ESSIC), University of Maryland, with financial support from the Gordon and Betty Moore Foundation, the NASA Applied Sciences Program (Michael Goodman), and the integrated earth system modeling (iESM) project funded by the DOE Earth System Modeling Program. PNNL is operated for the U.S. DOE by Battelle Memorial Institute under contract DE-AC06-76RLO1830. The authors would like to thank John Lucotch (UMT) for assistance in data processing.

## References

Döll, P., and B. Lehner (2002), Validation of a new global 30-min drainage direction map, *J. Hydrol.*, *258*, 214–231.

Fekete, B. M., C. J. Vörösmarty, and R. B. Lammers (2001), Scaling gridded river network for macroscale hydrology: Development and analysis, *Water Resour. Res.*, *37*(7), 1955–1967.

Lehner, B., K. Verdin, and A. Jarvis (2008), New global hydrograhy derived from spaceborne elevation data, *Eos Trans. AGU*, *89*(10), 93–94.

Olivera, F., and R. Raina (2003), Development of large scale gridded river network from vector stream data, *J. Am. Water Resour. Assoc.*, *39*(5), 1235, 199–200.

Wu, H., J. S. Kimball, N. Mantua, and J. Stanford (2011). Automated upscaling of river networks for macroscale hydrological modeling. *Water Resour. Res.*, *47*, W03517, doi:10.1029/2009WR008871.