7-1-2001

Scaling gridded river networks for macroscale hydrology: Development, analysis, and control of error

Balazs M. Fekete
University of New Hampshire, Durham

Charles J. Vorosmarty
University of New Hampshire, Durham, charles.vorosmarty@unh.edu

Richard B. Lammers
University of New Hampshire, Durham, richard.lammers@unh.edu

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

WATER RESOURCES RESEARCH, VOL. 37, NO. 7, PAGES 1955–1967, JULY 2001

Scaling gridded river networks for macroscale hydrology: Development, analysis, and control of error

Balázs M. Fekete, Charles J. Vörösmarty, and Richard B. Lammers
Institute for the Study of Earth, Oceans, and Space, University of New Hampshire, Durham, New Hampshire

Abstract. A simple and robust river network scaling algorithm (NSA) is presented to rescale fine-resolution networks to any coarser resolution. The algorithm was tested over the Danube River basin and the European continent. Coarse-resolution networks, at 2.5, 5, 10, and 30 min resolutions, were derived from higher-resolution gridded networks using NSA and geomorphometric attributes, such as river order, shape index, and width function. These parameters were calculated and compared at each resolution. Simple scaling relationships were found to predict decreasing river lengths with coarser-resolution data. This relationship can be used to correct river length as a function of grid resolution. The length-corrected width functions of the major river basins in Europe were compared at different resolutions to assess river network performance. The discretization error in representing basin area and river lengths at coarser resolutions were analyzed, and simple relationships were found to calculate the minimum number of grid cells needed to maintain the catchment area and length within a desired level of accuracy. This relationship among geomorphological characteristics, such as shape index and width function (derived from gridded networks at different resolutions), suggests that a minimum of 200–300 grid cells is necessary to maintain the geomorphological characteristics of the river networks with sufficient accuracy.

1. Introduction

The advent of high-resolution gridded river networks having global coverage offers new opportunities for the study of regional, continental, and global-scale hydrological processes. Gridded river networks typically are derived from digital elevation models (DEM) using maximum downhill (decreasing) elevation gradient search procedures [*Jenson and Domingue*, 1988]. Unfortunately, gridded networks derived from currently available DEM are often inaccurate since most DEM were not designed with the intent of representing river flow patterns.

Automated methods and algorithms have been proposed to correct DEM or derived gridded networks [*Band*, 1993; *Hutchinson*, 1989; *Jenson and Domingue*, 1988]. Gridded networks derived from resampled DEM typically are fragmented owing to spurious local depressions [*Hutchinson*, 1989]. Automated methods to derive gridded networks are most sensitive over flat terrain where the maximum downhill elevation gradient search algorithm is more sensitive to DEM errors due to the lack of pronounced gradients. At global scales and coarser spatial resolutions, average gradients between grid cells are lower, and as a result, river networks derived at the coarser resolutions are subject to substantial error. An automated procedure with manual correction and careful validation against several independent sources has recently been applied to the global network of rivers [*Vörösmarty et al.*, 2000b].

Numerous gridded networks at various spatial resolutions, yielding continental and global coverage, have been released recently [*U.S. Geological Survey*, 1998; *Renssen and Knoop*, 2000; *Vörösmarty et al.*, 2000b; *Graham et al.*, 1999; *Oki and Sud*, 1998]. The most ambitious project is HYDRO1k, which seeks to deliver a 1 km resolution gridded network with global coverage. At present, its global-scale implementation at the full 1 km resolution would be far too computer intensive for most global applications. Clearly, a derived river network at suitably coarser spatial resolutions would be better suited to continental- and global-scale applications. Effective procedures for rescaling the otherwise useful high-resolution river networks, however, have not been developed.

Typical grid resolutions for global hydrologic modeling are of the order of 10′ (~10 km) to 2° or 3° (few hundred kilometers) [*Coe*, 1998; *Hagemann and Dümenil*, 1998; *Fekete et al.*, 1999]. It has been argued that the 30° resolution is suitable for a broad variety of global hydrological modeling and emerging constituent flux research [*Vörösmarty et al.*, 2000a, 2000b]. Finer-resolution networks would be needed for the representation of individual or small local basins and tributaries. Since the development of high-quality gridded networks is generally a time-consuming process, a procedure for streamlining the reaggregation of finer-resolution networks would be valuable to a broad suite of land surface hydrology, climate dynamics, and water resources studies. The availability of finer-resolution gridded networks potentially offers new opportunities for the development of coarser-resolution networks, provided the problems associated with automated rescaling of the fine-resolution grids can be solved. One network grid aggregation algorithm developed thus far [*O'Donnell et al.*, 1999] is not generic enough to use in applications where the original fine-resolution grid needs to be projected before deriving flow routing at a coarser resolution.

We demonstrate here a method for converting fine-resolution river networks to coarser resolutions while preserving the topology and key geomorphometric properties of the original data set. The proposed method has been successfully applied to all continents covered by HYDRO1k (Africa, Asia,

Paper number 2001WR900024.
0043-1397/01/2001WR900024$09.00

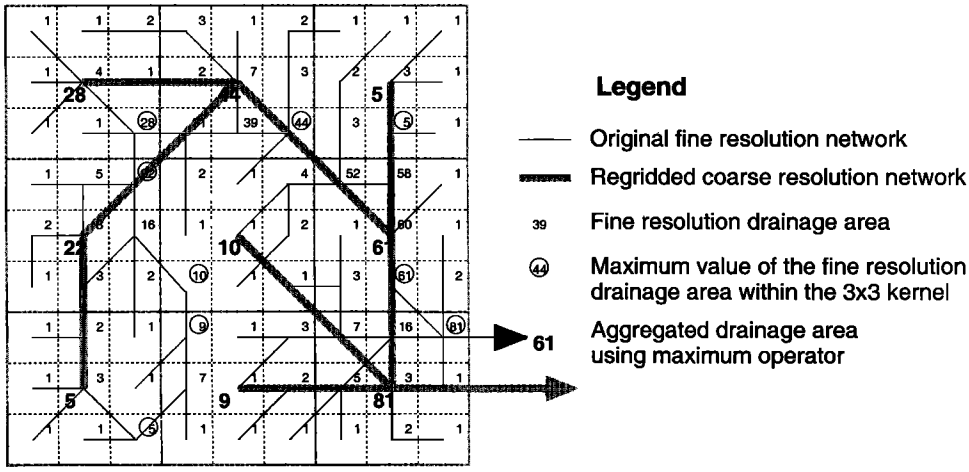

Figure 1. Network scaling algorithm (NSA) using a maximum value operator to aggregate 3 × 3 grid cells. Small numbers represent drainage area (number of cells) of the fine-scale network (small grids), and the larger numbers are the drainage areas for the coarse-scale grid. One coarse-scale grid cell equals nine fine-scale grid cells. The coarse-scale cell drainage areas are set to the maximum drainage area of the nine fine-scale cells. The coarse-scale river network is then recreated using a maximum "uphill" (or increasing) drainage area gradient search algorithm.

Australia, Europe, North America, and South America) to derive 0.1° (6′) resolution networks. In the current paper, however, we present results for Europe only. The remainder of the paper presents network scaling algorithm (NSA) methodology, where model accuracy is assessed by quantifying sources of error in the aggregation procedure. Changes in geomorphometric attributes across grid resolutions from 30″ (1 km) to >30′ (50 km) are also presented.

2. Method

2.1. Basic Algorithm

Our overall strategy for reconfiguring fine-resolution river networks into coarse-scale flow paths relies on drainage area calculated for all grid cells at the source resolution. This attribute, which is easily computed, is an integrated, conservative measure and a key property of drainage systems. At any scale, drainage area can be used as an input to automated procedures for rebuilding the aggregated river network. Figure 1 shows the aggregation of a fine-resolution grid using a 3 × 3 kernel. In order to preserve the high drainage area values along mainstems the finer-resolution grid needs to be aggregated using a maximum-value search within the aggregation kernel. When the rescaling involves both grid aggregation and projection, the high drainage area values can be preserved by projecting the drainage area grid with sufficient oversampling before the aggregation of the oversampled drainage area grid.

Grid projection is normally performed as sampling, where the procedure steps through the cells of the projected grid and calculates the coordinates of each grid cell center on the original projection. The procedure then samples the original surface at the projected locations. The sampling either (1) assigns the value of the nearest grid cell (nearest neighbor method) to the projected grid cell or (2) searches for several nearby cells within a specific radius and interpolates from these neighboring grid cells using a distance-weighted, bilinear, spline, or similar technique. Since the drainage area grid is a highly nonsmooth surface (i.e., the high drainage area values along main stems are surrounded with low values), any interpolation would distort these high values through smoothing. The nearest neighbor technique does not alter the sampled grid values, thereby preserving the high drainage area values.

The oversampling results in a projected upstream area grid with a finer resolution than the original upstream area grid (i.e., values from the original grid will be present repeatedly in the projected grid). This redundancy ensures that the high drainage area values are not missed during the grid resampling. The oversampling allows the separation of the projection and the aggregation (i.e., the grid projection is done first, resulting in a projected high-resolution grid that consistently preserves all features of the original grid). The projected high-resolution grid is aggregated later. In our case, the aggregation should apply a maximum-value search since the high contributing area values are carrying the information needed for the network reconstruction.

Drainage area grids can be used to derive flow directions similar to the manner in which DEMs are used, but in contrast to the use of DEM, which use a maximum downhill (decreasing) elevation gradient search procedure, the drainage area grid instead defines flow directions based on maximum "uphill" (increasing) drainage area gradients (pathways with increasing upstream catchment area). We refer to this procedure as the NSA. This procedure can be performed with standard geographic information system (GIS) raster functions, which are commonly available in many GIS software packages.

Following construction of a reaggregated network, several network-related geomorphometric attributes, such as drainage area, distance to outlet, and stream order, can be derived for individual stream links, tributary subbasins, and entire drainage systems. Comparisons to the original fine-scale data set can then be made for each attribute. We treat the flow directionality grid and the derived network attributes as one coherent data set, the simulated topological network (STN) [*Vörösmarty et al.*, 2000b].

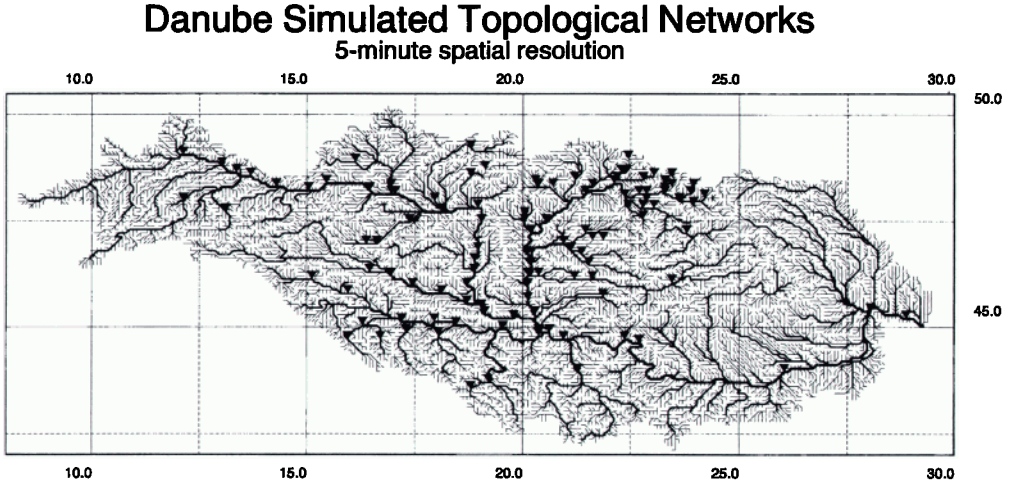

Figure 2. Manually edited Danube simulated topological networks (STN-05) at 5′ resolution and discharge gauging stations from the VITUKI (Water Resource Research Center, Budapest, Hungary) archive. The gauging stations were used to help validate and correct the network based on reported drainage area.

2.2. Input Data

The NSA was tested on HYDRO1k, which is derived from GTOPO30 30′ (~1 km) global elevation data set [*Gesch et al.*, 1999]. HYDRO1k is a hydrographically corrected DEM, wherein local depressions are removed and basin boundaries are consistent with topographic maps. Unlike other DEM, HYDRO1k includes numerous hydrology-related data layers, such as aspect, flow direction, drainage area (flow accumulation), elevation gradient, compound topographic index (wetness or topographical similarity index [*Moore et al.*, 1991; *Beven and Kirkby*, 1979]), basin and subbasin boundaries with Pfafstetter encoding [*Pfafstetter*, 1989; *Verdin*, 1997] and DEM-derived stream lines. In order to maintain uniform grid cell area the HYDRO1k data set was developed on a Lambert azimuthal equal-area projection [*Steinwand and Hutchinson*, 1995] and therefore is truly a 1 km resolution DEM.

2.3. Test Areas

We first applied NSA to the Danube basin to create a 5′ resolution network from HYDRO1k. The Danube basin and the 5′ resolution were chosen as a first test area because of the availability of a similar, carefully validated network (STN-05) we had developed (Figure 2). This network, developed before GTOPO30 and HYDRO1k became available, was derived from ETOPO5 [*Edwards*, 1989]. The initial 5′ routing was intensively edited (manually) using RiverGIS, a specialized geographic information system (GIS) tool developed at the University of New Hampshire as part of our Global Hydrological Archive and Analysis System. RiverGIS facilitates the viewing and editing of simulated gridded networks using vector river networks (e.g., the 1:3M ARC/World [*Environmental Systems Research Institute (ESRI)*, 1992a] and 1:1M digital chart of the world [*ESRI*, 1992b]) displayed in the background for guidance. RiverGIS also has the capability to quickly derive network attributes, such as drainage area, main stem length, and next station downstream at discharge gauging stations. Actual drainage area, supplied as part of the metadata from discharge gauging data sets, provides another means to identify potential errors in the simulated network. For the 5′ gridded network of the Danube basin we used reported drainage area from 113 monitoring stations contained in the VITUKI (Water Resource Research Center, Budapest, Hungary) archive to improve routing accuracy while developing STN-05.

A second step in testing the regridding algorithm examines the sensitivity of drainage area representation to network resolution. By applying this test to the European continent west of the Ural Mountains (Figure 3), assessments of the rescaled networks over a broad spatial domain and for rivers encompassing several orders of magnitude in drainage area (~100 to 3.2×10^6 km^2) can be made.

2.4. Reference Subbasins

Reference subbasins were derived from the fine-resolution networks by partitioning them into approximately equally sized subbasins. The subbasins were identified by the coordinate of their basin outlets and the corresponding subbasin grid. Network-derived attributes were calculated for each reference subbasin at its basin outlet. Both the basin outlets and the subbasin grids were projected to coarser resolutions. The outlet points of the reference subbasins were used to identify the corresponding subbasins on the coarser-resolution gridded network, in order to compare the network-derived attributes calculated at the different spatial resolutions.

3. Discussion

3.1. Analysis of NSA Errors

To understand potential limitations in the NSA procedures, we compared the network-derived attributes at the original 1 km resolution and the 5′ regridded resolution for the Danube basin. Sections 3.1.1–3.1.4 identify error sources and outline potential improvements to the NSA procedures to limit these errors.

3.1.1. Verification of the NSA procedures. The first verification of NSA was done through visual inspection of the regridded networks at the different resolutions tested (Figure 3). Qualitatively, the regridded networks retain the main river patterns and network continuity.

For quantitative evaluation of the NSA procedures the Danube basin was partitioned into 1364 reference tributaries of

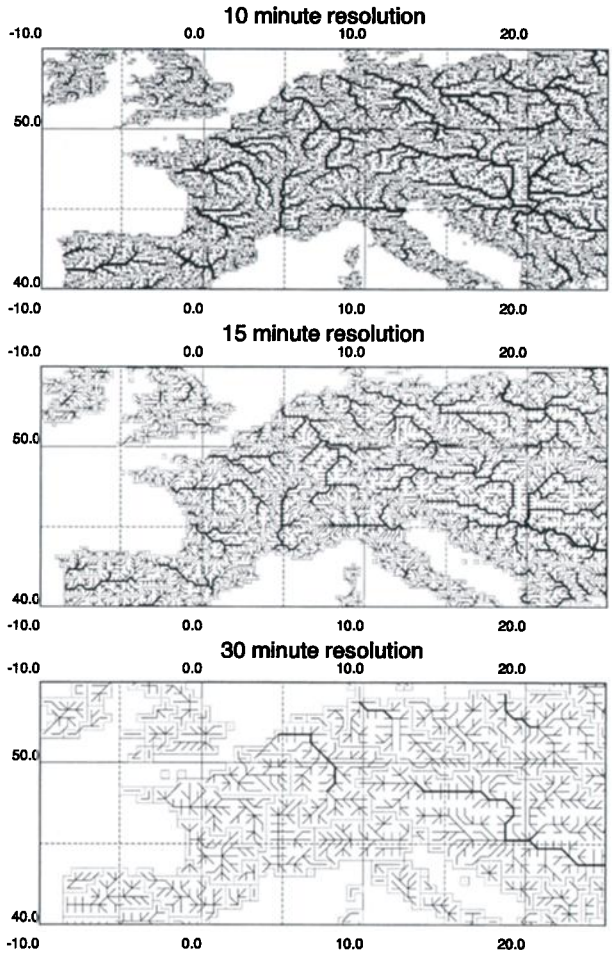

Figure 3. Simulated river network for Europe at 10′, 15′, and 30′ resolution derived from HYDRO1k using the network scaling algorithm.

~500 km^2 each at the original HYDRO1k resolution. The outlets of the reference tributaries defined overlapping subbasins ranging from 500 to 780,000 km^2 in size. Figure 4a shows good correspondence between the drainage areas derived from the original HYDRO1k and the rescaled 5′ network. As expected, smaller basins show larger numerical dispersion, while larger basins have less error. As a measure of this correspondence, we introduce symmetric relative error (SRE) ε_{sym} [*Fekete et al.*, 1999] as

$$\varepsilon_{sym} = \frac{X_{sim} - X_{obs}}{\max(X_{sim}, X_{obs})} 100\%, \quad (1)$$

where X_{obs} is the observed (or 1 km) value and X_{sim} is the simulated (5′) value. This error term is symmetric regardless of overestimation or underestimation of the observed value, and it ranges between −100 and 100%. Using (1), the error distribution by drainage area can be calculated (Figure 4b). The mean SRE (for all basin sizes) is −4.09%, with a standard deviation of 17.06% and mean absolute SRE of 16.56% (Figure 4b).

3.1.2. Comparing NSA-generated and manually edited 5′ networks. Comparison between the manually edited and the HYDRO1k-derived 5′ networks to digital chart of the world 1:1M [*ESRI*, 1992b] scale river networks shows good correspondence. The manually edited network, however, has much higher accuracy in the rendering of actual river courses than the uncorrected HYDRO1k-derived network (Figure 5). The large bend on the Tisza River in Hungary before its confluence with the Danube, according to the HYDRO1k-derived (uncorrected) network, is an error in HYDRO1k that was inherited from the original 1 km routing. This is one example of many which underlies the need for careful hand editing of digital river networks derived from automated methods. The development of new methods that use not only elevation but also additional information, such as digitized river networks and drainage area benchmark points, may further improve regridding methods.

Comparison of reported and simulated network-derived drainage area at discharge gauges contained in the VITUKI archive shows that the drainage area error is significantly less when manual editing is applied (Figure 6). The drainage area error, compared to a manually edited network, is within ±10% for 90% of the stations (Figure 6a), while only 68% of the stations are within ±10% when compared against the

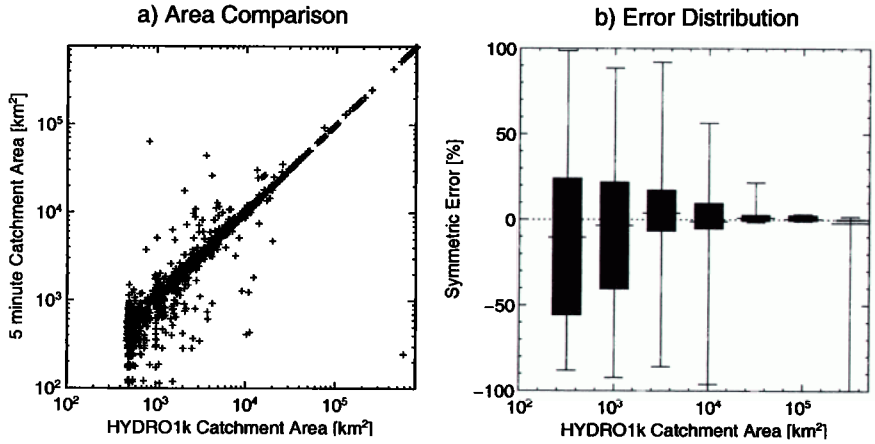

Figure 4. (a) Comparison of HYDRO1k versus NSA regridded 5′ network-derived catchment areas and (b) NSA regridded 5′ network catchment area error distribution (average, minimum, maximum, tenth, and ninetieth percentiles) by basin sizes.

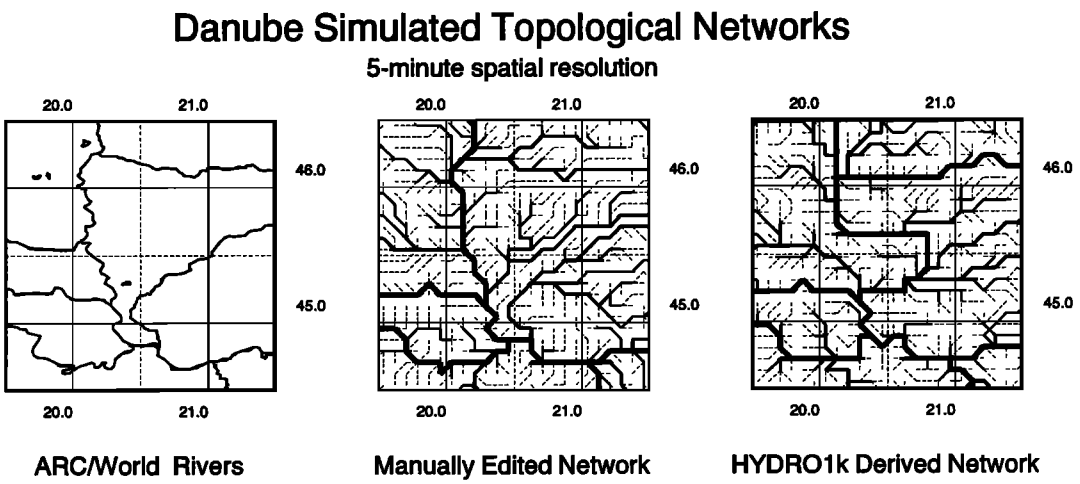

Figure 5. Danube-Tisza confluence. Comparison of manually edited and HYDRO1k-derived 5′ networks to ARC/World rivers.

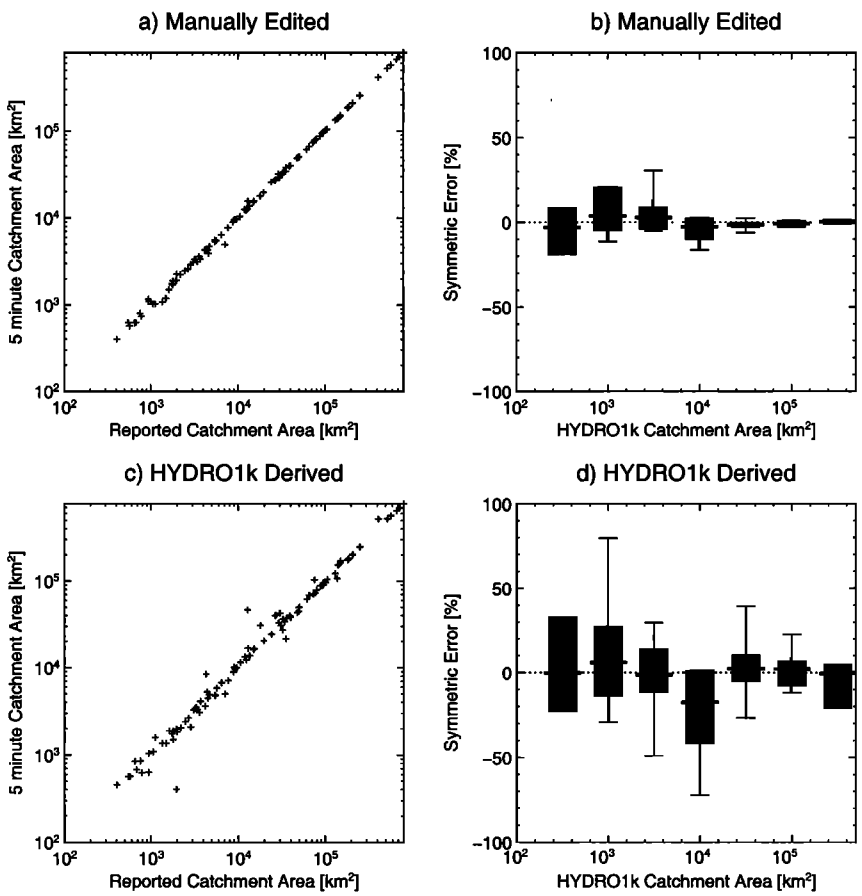

Figure 6. Comparison of reported drainage area (at 113 gauging stations from VITUKI archives) and estimated drainage area derived from manually edited STN-05 network and HYDRO1k-derived simulated networks. Figures 6a and 6c show one-to-one comparisons of the reported and derived drainage area of the manually edited and HYDRO1k-derived networks, respectively. Figures 6b and 6d show the symmetric error distributions of the manually edited and HYDRO1k-derived networks by drainage area. The differences between the HYDRO1k-derived and the manually edited networks reflects the inaccuracies in HYDRO1k rather than the errors introduced by the NSA regridding and emphasizes the need to incorporate more information (e.g., existing river networks and reported drainage areas) in the delineation of gridded networks.

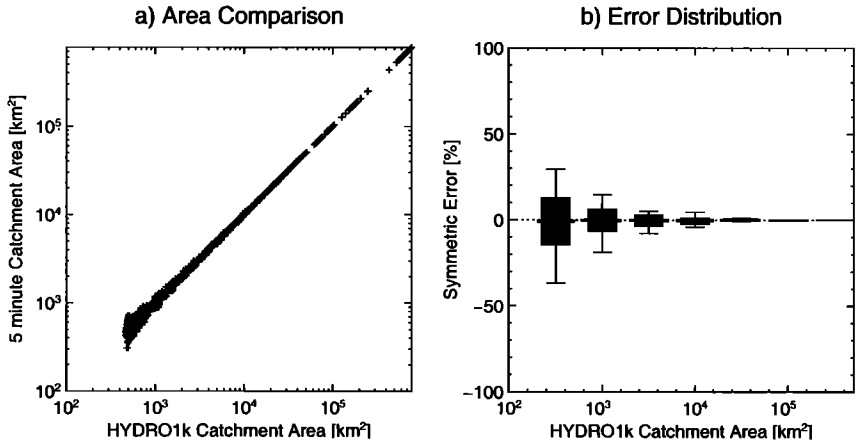

Figure 7. (a) Comparison of the HYDRO1k versus the 5′ regridded drainage areas at the outlets of the 1364 subbasins and (b) regridded drainage area error distribution (average, minimum, maximum, tenth, and ninetieth percentiles) by basin sizes.

HYDRO1k-derived network (Figure 6b). It should be noted that the manually edited network was heavily optimized to represent accurately these discharge gauging stations. Considering other stations not included in the manual editing may result in a smaller difference between the manually edited and the HYDRO1k-derived networks' performance. We would argue, however, that strong benchmarking of manually edited networks to a fairly dense set of discharge gauging stations along with careful editing to actual river networks should result in better performance between the gauging stations.

3.1.3. Analyzing the sources of the NSA errors. The NSA yields two sources of errors. The first error is due to generic grid resampling, which does not involve the river network. The minimum grid resampling error can be assessed by considering a rounding error analogy. We express the maximum rounding error of the subbasins' grid representation as $\varepsilon = \Delta A/2$, where ΔA is the cell area. Since the rounding error of any basin at the 5′ resolution ($\Delta A = \sim 60$ km^2) should be less than $\sim$30 km^2, the average of the absolute rounding errors should be approximately half of that. This can be tested by projecting the HYDRO1k-derived reference subbasin grid using simple grid resampling and then comparing the projected areas of the reference subbasins with the original subbasins at the HYDRO1k resolution. Figure 7 shows the comparison of HYDRO1k catchment area and the regridded accumulated subbasin area at the 1364 subbasins outlets for the Danube. The mean absolute error was 92.3 km^2, which shows that the generic resampling method introduces errors in addition to the rounding error of the 5′ grid representation.

The mean SRE of the basin regridding is −0.28%, with a standard deviation of 4.96% and mean absolute SRE of 4.41%. This is substantially less than the NSA overall error (Figure 4) and represents an upper bound on the achievable accuracy of the NSA algorithm.

The second source of error arises from the use of a drainage area grid as the surface to derive a river network. Although this method is simple and robust enough to preserve the major flow patterns, it has limitations. Networks derived from a particular drainage area surface do not result in exactly the same flow pattern as the original network. Typically, the differences between the original network (which is used to define the drainage area surface) and the derived networks are small (Figure 8). The error can be more severe, however, when two major flow lines fall close to each other.

3.1.4. Network scaling algorithm with basin enhancement. The most important limitation of the NSA rescaling algorithm is that the simulated network, derived from drainage area, may not precisely match the original network used to generate the initial drainage area surface. Improvements can be realized by limiting the largest of the potential differences to ensure that the regridding algorithm maintains the subbasin configuration as accurately as possible.

An approach toward improving NSA is to incorporate subbasins in the regridding procedure. The subbasins derived from the original HYDRO1k, projected and resampled to the target network resolution, can be used in a modified maximum "uphill" (increasing) drainage area gradient search procedure. On a first pass the modified procedure (while searching for the maximum drainage area gradient) considers only those neighboring cells which fall into the same subbasin region as the cell for which the flow direction is to be determined. Should the procedure fail to find any flow direction (i.e., the cell is the outlet of a subbasin) on the first pass, a second pass then extends the search into neighboring subbasins.

The network scaling algorithm with basin enhancement (NSABE) was applied to the Danube basin, with the 1364 subbasins used to guide the algorithm in deriving a 5′ network. The use of subbasins helped to improve NSA performance (Figure 9); the mean SRE dropped from −4.09 (NSA) to −0.55% and approaches the −0.28% basin regridding error. The standard deviation of the SRE also decreased substantially from 17.06 to 7.74%, approaching the 4.96% standard deviation of the regridding SRE. Similarly, the mean absolute SRE decreased from 16.56 to 9.12%, which is about twice as high as the 4.41% mean absolute SRE from regridding.

This procedure described above can be extended to consider any number of hierarchically nested subbasin partitions, starting with the finest set of subbasins (which partitions the network to the smallest subcatchments) and continuing the search to larger subbasins. HYDRO1k, with different Pfafstetter encoding levels [*Pfafstetter*, 1989; *Verdin*, 1997], provides an excellent set of hierarchical subbasins for NSABE rescaling. This more general NSABE with HYDRO1k hierarchical subbasins

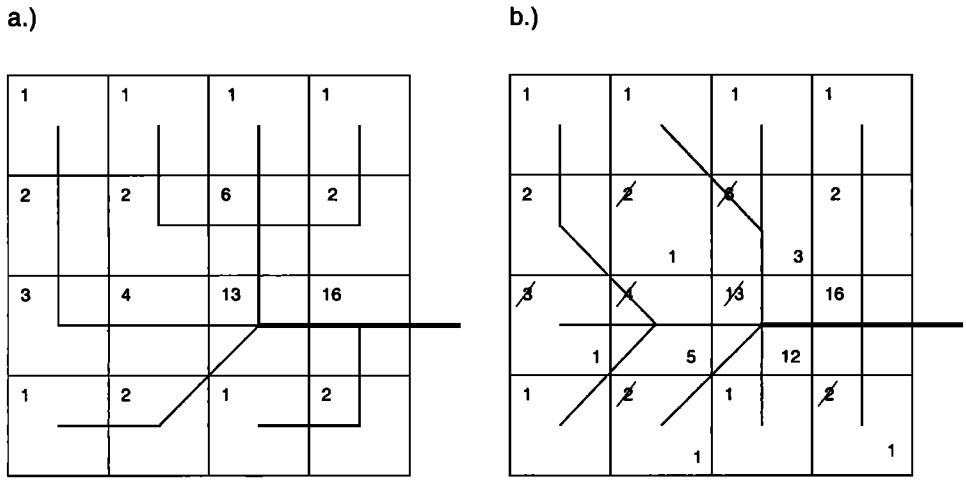

Figure 8. Regridding error due to reconstructing flow routing from drainage area. (a) the original network and (b) the reconstructed network from drainage area. The reconstructed network is not exactly the same as the original network. Therefore the drainage area surface derived from the reconstructed network differs from the original drainage area surface. The small numbers in both Figures 8a and 8b represent the drainage area in grid cells, and the slashed numbers in Figure 8b show where the reconstructed network and the drainage area derived from the reconstructed network differs from the original network. This error is a deficiency of the NSA and occurs without any aggregation projection.

was used to rescale the river networks of Europe to different resolutions.

3.2. Impact of Resolution on Gridded Network Performance

The NSA and NSABE allow us to derive comparable networks over the same domains at different resolutions to study the impact of resolution on network-derived basin characteristics, such as stream order, catchment area, mainstem length, and other geomorphometric properties. Such analysis can give a better understanding of how to optimize networks for particular applications.

3.2.1. Deriving 2.5′, 5′, 10′, 15′, and 30′ networks from HYDRO1k. Simulated topological networks at five different resolutions (2.5′, 5′, 10′, 15′, and 30′) were derived from HYDRO1k using NSABE with Pfafstetter encoded subbasins supplied as part of HYDRO1k. The 2.5′ network served as a reference data set for the comparison of network performance at different resolutions.

A subbasin partitioning with drainage area of ~5000 km^2 (with corresponding basin outlets) was derived from the 2.5′ network. The 5000 km^2 threshold is well below the minimum catchment area that can be represented well by a 30′ network [*Vörösmarty et al.*, 2000b] but still larger than the average cell area at this resolution, ensuring that basin outlets will not fall into the same grid cell. Network-derived attributes at subbasin outlet points were computed based on the 2.5′ network to serve as a basis for assessing NSABE performance at four spatial resolutions. The basin outlets were georegistered to the 5′, 10′, 15′, and 30′ resolution using the 3 × 3 kernel search for the cell with the best fitting drainage area. Network attributes

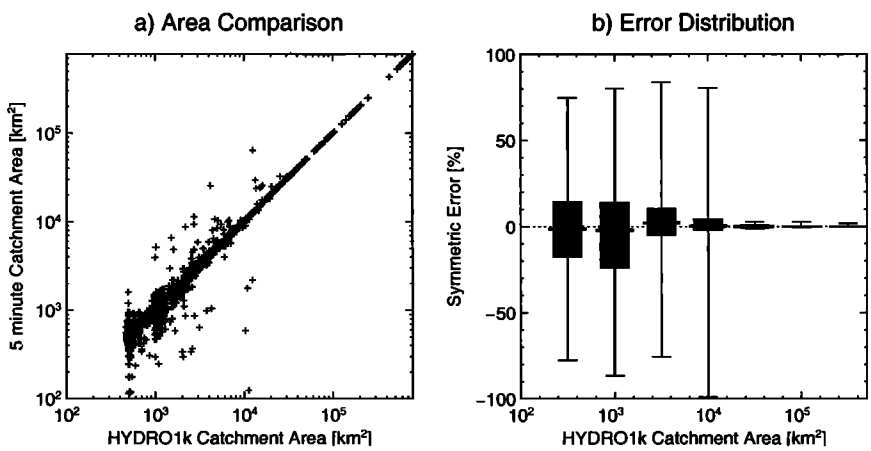

Figure 9. (a) Comparison of HYDRO1k versus NSA with basin enhancement (NSABE) regridded 5′ network-derived drainage areas and (b) NSABE regridded 5′ network drainage area error distribution (average, minimum, maximum, tenth, and ninetieth percentiles) by basin sizes. Figure 9 may be compared to Figure 4 to see the decrease in error resulting from the incorporation of basin enhancement.

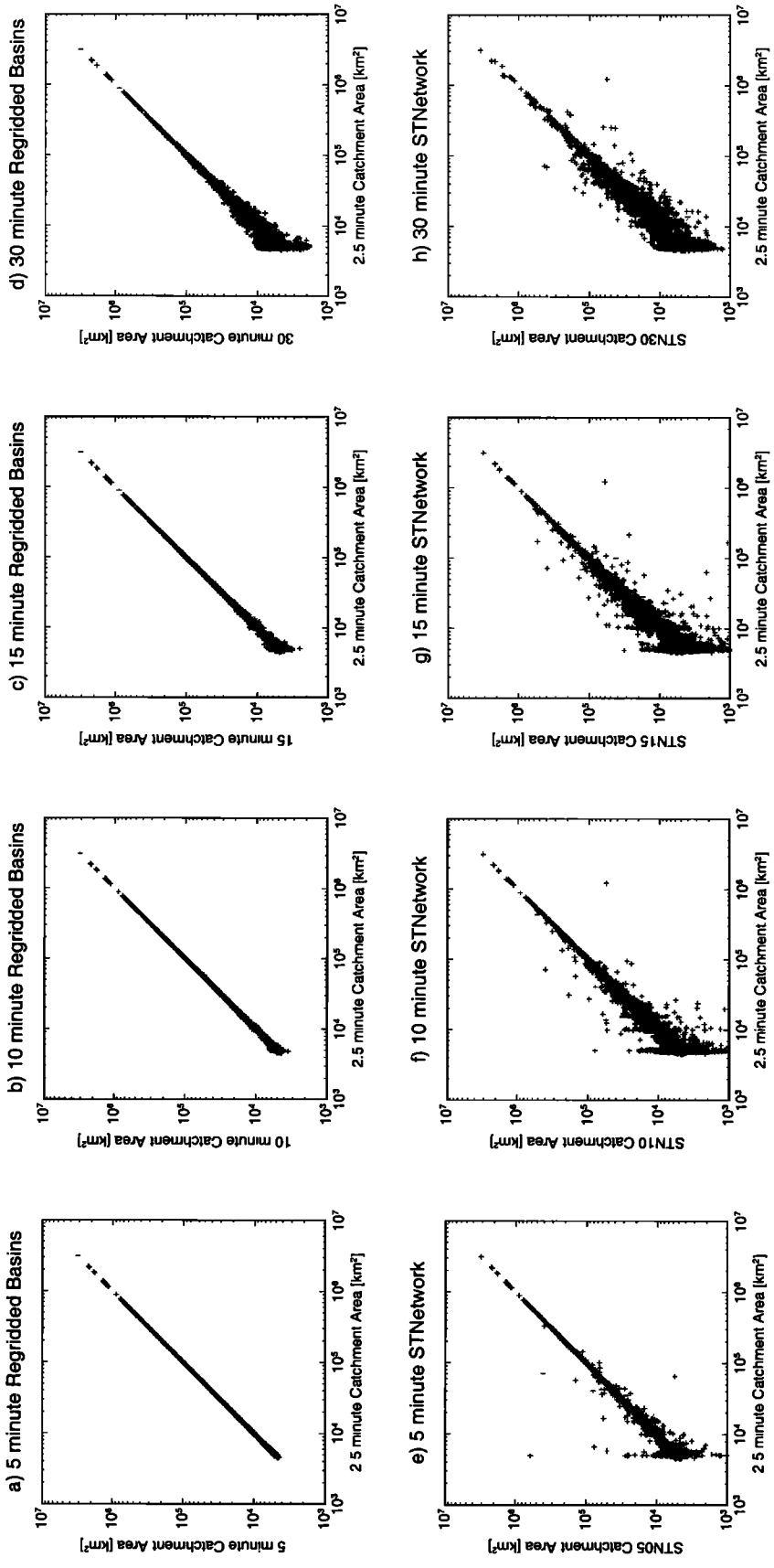

Figure 10. Comparison of gridded network performance in terms of drainage area representation. (a–d) The regridding error (i.e., regridding the 2.5′ subbasin grid at different resolutions) and (e–h) the total NSABE rescaling error.

Table 1. Grid Comparisons

Resolution	Mean Cell Area, km^2	Mean Cell Length, km	Number of Cells	Number of Stream Segments	Number of Basins	Mean Number of Cells Per Basin	Highest Order
2.5′	14	4	1,205,647	818,075	11,571	104.2	9
5′	55	9	301,437	207,144	3,835	78.6	8
10′	221	17	75,355	52,530	1,452	51.9	7
15′	497	25	33,467	23,614	838	39.9	7
30′	1994	50	8,360	5,986	327	25.6	6

were then derived from the coarser-resolution networks. Figure 10 shows the performance (in terms of resolving drainage area) of the different networks derived from HYDRO1k, with error in drainage area clearly increasing as resolution decreases.

3.2.2. Comparison of gridded networks at different resolutions. The network statistics for the five European gridded networks at 2.5′, 5′, 10′, 15′, and 30′ resolutions are presented in Table 1. Some of the statistics, such as average cell area, average cell length, and number of cells, are predictable by simply considering the grid resolution. The number of Strahler stream segments follows the same logarithmic trend as the number of cells. The average number of cells per basin does not decrease at the same rate as the number of cells in the whole network because the coarser-resolution networks tend to preserve the larger basins while the smaller basins are integrated into the large basins as the resolution is degraded.

Basin shape indices were calculated at the basin outlets using the 5000 km^2 subbasin partitioning. The shape index [*Vörösmarty et al.*, 2000a] is defined as

$$S = L/\sqrt{A} \tag{2}$$

where S is shape, L is mainstem length [km], and A is basin area [km^2]. Figure 11 shows the shape index error as a function of number of grid cells. The shape index error increases dramatically under 300 grid cells and has an increasing negative bias at coarser resolutions (Figure 11). This negative bias indicates that regridding to coarser resolutions tends to result in more rounded basins. A review of the data sets reveals that this tendency toward more rounded basins is not due to the changing outline of the basins but is, instead, a function of decreasing mainstem length at coarser resolutions.

3.2.3. Network performance at different resolutions. Designing gridded networks for particular applications needs to

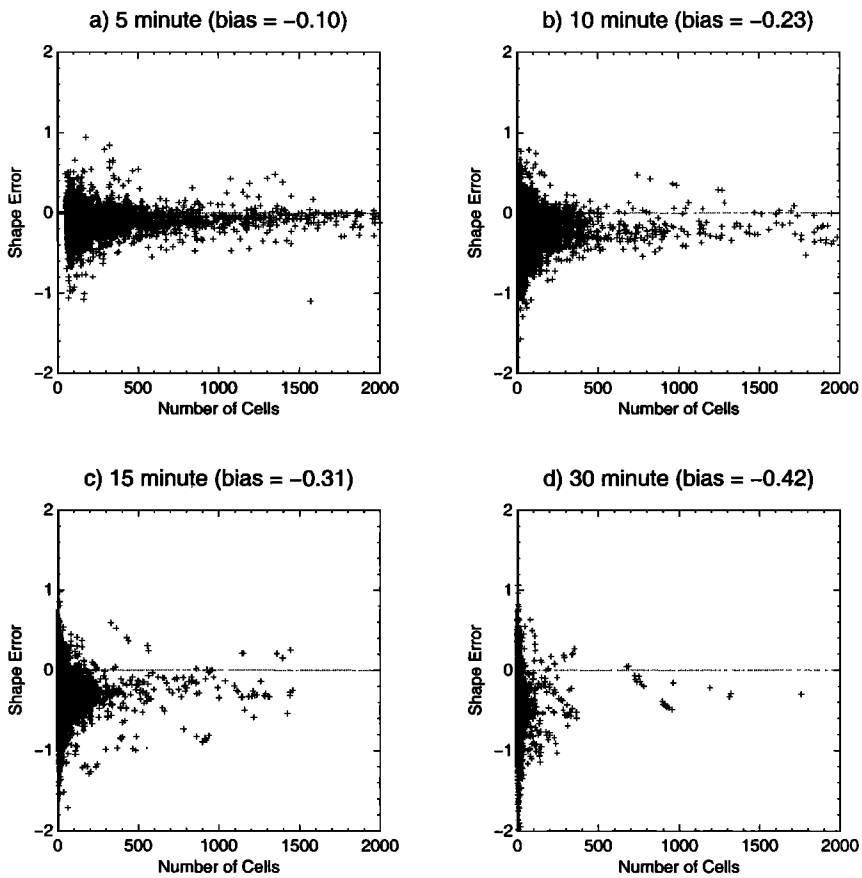

Figure 11. Basin shape index error (defined as basin shape index at the regridded resolution minus basin shape index at 2.5′ resolution) by drainage area at 5′, 10′, 15′, and 30′ resolutions.

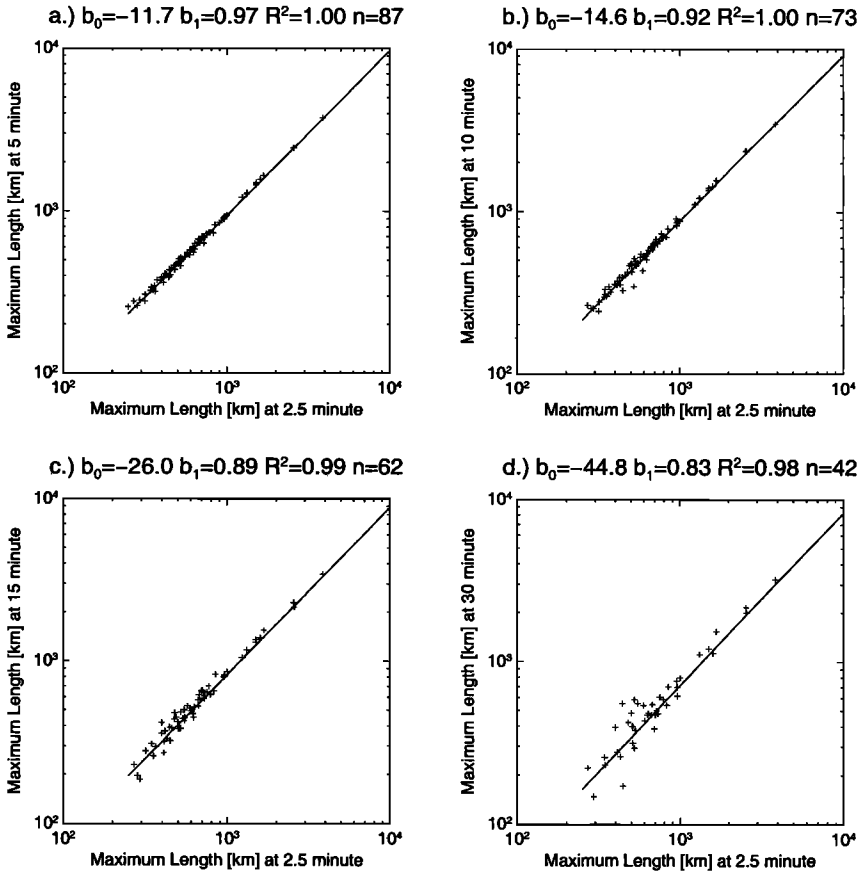

Figure 12. Comparison of the maximum river length at 2.5′ versus 5′, 10′, 15′, and 30′ resolutions. Only those basins where the drainage area estimate at 2.5′ resolution and the regridded resolution agreed within ±10% error were included in the comparison. These criteria were applied to ensure that the compared basins have reasonably similar representation at all resolutions and that the differences in the morphomentric characteristics are due to the resolution differences only and are not affected by the errors introduced through rescaling. Linear regression coefficients including the intercept b_0, slope b_1, and correlation coefficient R^2 are shown.

strike a balance between the higher accuracy of fine-resolution networks versus the increasing difficulties of developing and using those networks in associated flow-routing schemes. The first step in assessing network performance of gridded networks at different resolutions was to identify 88 basins with drainage area >25,000 km^2 for each resolution. The 25,000 km^2 basin size was found to be the minimum that could be represented in a 30′ network [*Vörösmarty et al.*, 2000b], which was the coarsest resolution in our experiment.

We now assess the fidelity of simulated river networks by comparing the maximum lengths of the 88 basins at 5′, 10′, 15′, and 30′ resolutions (Figure 12). As expected, the regression lines show a systematic decreases in the maximum lengths at coarser resolutions (Figure 13). This systematic length decrease explains the increasing negative bias in the basin shape index described in section 3.2.2. The intercept terms of the regression lines are negligible compared to the maximum basin lengths. The slope terms of the regression lines, however, show an important linear trend in the decrease of the maximum length at coarser resolution (Figure 13). This decrease is due to the inability of the coarser-resolution networks to represent the sinuosity of the real rivers.

When slopes of the maximum-length regression lines at different resolutions as a function of the resolution differences (i.e., the cell size ratio) (Figure 13) are plotted, a log linear relationship between cell size ratio and stream length ratio is apparent. This relationship can be expressed as

$$\frac{L_f}{L_c} = 1.024 - 0.077 \ln\left(\frac{\Delta L_f}{\Delta L_c}\right), \tag{3}$$

where L_f is stream length [km] at fine resolution, L_c is stream length [km] at coarse resolution, ΔL_f is cell size [km] at fine resolution, and ΔL_c is cell size [km] at coarse resolution.

Equation (3) can be used to predict the decrease in stream length for coarser-resolution networks or adjust the stream lengths calculated at a coarser-resolution network. We constructed the width function of the Danube basin from the original HYDRO1k network and the 5′ regridded network (Figure 14). The width function is either normalized [*Rinaldo et al.*, 1995] or nonnormalized [*Veneziano et al.*, 2000]. In this paper we use the width function $W(L)$ from *Veneziano et al.* [2000], defined such that $W(L)dL$ is the drainage area increase for a drainage area located between distance L and $L + dL$ from basin mouth. The width functions of the Danube at different resolutions show that the 5′ network without length correction is systematically shorter than the 1 km network (Figure 14a). By applying the length correction computed from (3) the coarser-resolution network can be adjusted to match

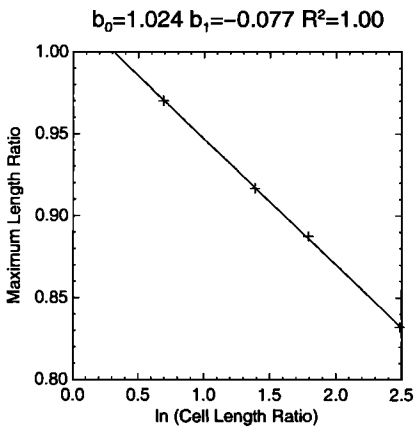

Figure 13. Maximum river length ratio as a function of cell size ratio. The slope term in the linear regression of maximum river length (Figure 12) can be viewed as the ratio of the fine- and the coarse-resolution maximum river lengths. Relating these maximum length ratios to the logarithm of the corresponding cell size ratios shows a linear trend. This relationship can be used to predict the shortening of river lengths due to decreasing grid resolution.

the width function derived from the fine-resolution network (Figure 14b).

We applied the correction coefficient to the 5′, 10′, 15′, and 30′ networks in order to compare the width function of the 88 basins in Europe at different resolutions. Figure 15 shows the width functions for different resolutions of the larger European basins. The ability of the width function to replicate the finer scales becomes limited for smaller drainage area basins. However, this result is expected, given the increased error in the shape index at coarser resolutions (Figure 11).

3.3. A Priori Estimation of Error Characteristics

The rounding error discussed in section 3.1.3 offers a means to assess the expected accuracy of a gridded network at a given resolution. Two equations derived in Appendix A relate the desired area accuracy ε_A, mean length accuracy ε_L, and the smallest area A expected to be represented at that accuracy to the minimum number of grid cells n needed to maintain those accuracies as $n = 1/(2\varepsilon_A)$ and $n = 1/(4\varepsilon_L^2 S_m^2)$, respectively. The S_m term in the length accuracy equation is the mean length shape index of the basin, which is similar to the shape index (equation (2)). S_m, however, relates the mean river length to the square root of the basin area instead of main stem length. The mean river length is typically half of the main stem length, and therefore the mean length shape index is also half of the traditional shape index. While the values of the traditional shape index span 1.0–3.6 [*Vörösmarty et al.*, 2000a], the mean length shape index S_m varies between 0.5 and 2.0.

On the basis of the area accuracy equation, if the desired area accuracy is $\varepsilon_A = 0.1$, then $\Delta A = 0.2A$ and at least five grid cells are needed to maintain 10% accuracy. This corresponds well with the 10,000–25,000 km^2 minimum basin size that can be represented in 30′ networks (~2000 km^2 average grid cell size) [*Vörösmarty et al.*, 2000b; *Lammers et al.*, 2000]. Similarly, if the desired length accuracy is $\varepsilon_L = 0.1$ and we assume $S_m = 0.5$ as a worst case scenario, the length accuracy equation yields $n = 1/\varepsilon_L^2$. Therefore, to maintain 10% accuracy, a minimum of 100 grid cells is required. This estimate does not take into account the cell length variation in gridded networks and any additional error contained within a gridded network. Considering these uncertainties, a minimum of 200–300 grid cells may be required to represent the underlying river topology, which is consistent with our finding that the shape error dramatically increases below 300 grid cells (Figure 11). This result is important in applying gridded networks in flow-routing schemes since the failure to maintain the geomorphometric characteristics of river basins may result in substantial flow-routing errors.

4. Conclusion

The availability of high-resolution global, gridded network data sets such as HYDRO1k offers new opportunities in large-scale river flow and constituent transport modeling. However, the use of these data sets in continental- or global-scale routing schemes is not yet feasible due to limitations in currently available computer resources. Nonetheless, the macroscale modeling community can benefit from the availability of these high-resolution river networks by using them to derive coarser-resolution data sets.

The present paper presented two simple and robust algorithms: the network scaling algorithm (NSA) and a modified version of NSA with basin enhancement (NSABE), which can be used to rescale fine-resolution gridded networks to coarser

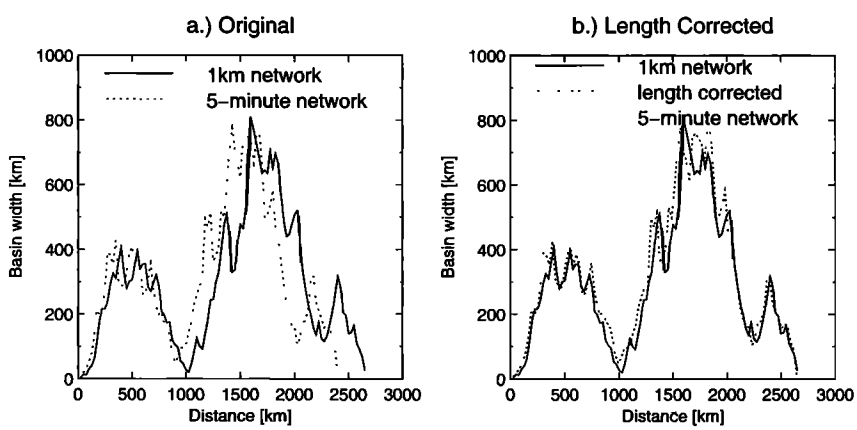

Figure 14. Width function of the Danube basin at 1 km and 5′ resolution (a) without length correction and (b) with length correction.

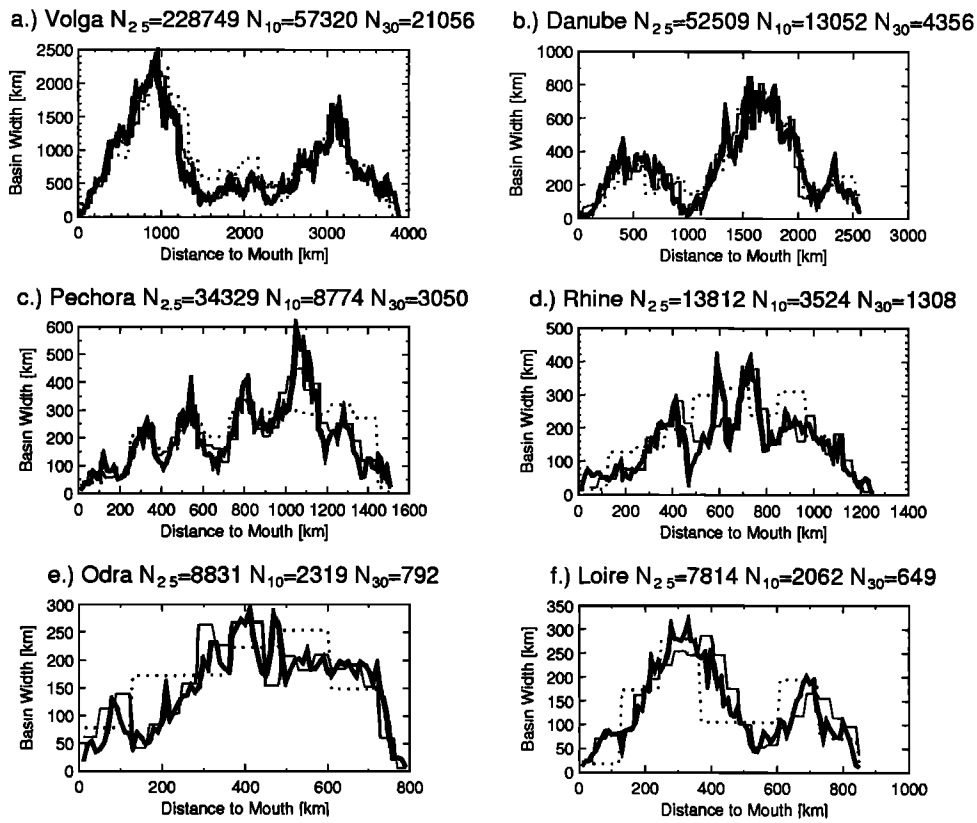

Figure 15. Width functions of major European rivers with the length correction at 2.5′ (thick solid line), 10′ (thin solid line), and 30′ (dashed line) resolutions. Solid line shows 2.5′, dotted line is 10′, and dot-dashed line is 30′ resolution.

resolutions. The NSA uses projected and aggregated drainage area surfaces derived from a fine-resolution data set to delineate river networks at coarser resolutions. The NSABE algorithm is an improvement on the NSA algorithm which incorporates subbasin partitioning derived from the fine-resolution network to guide the delineation of the river network at the coarser resolution. The performance of both algorithms was demonstrated by comparing a 5′ gridded network derived from HYDRO1k to a highly accurate gridded network of the Danube basins at the same resolution. Two major sources of error were identified. The first is error introduced by regridding, which was found to be higher than the rounding error as a result of the coarser discretization. The second source of error arises from the use of drainage area as a surface to derive flow routing. This error cannot be completely eliminated. It can, however, be reduced with the basin enhanced algorithm, which considers not only the drainage area surface but the subbasin partitioning as well.

Aggregated river networks across Europe at 2.5′, 5′, 10′, 15′, and 30′ resolutions were developed using NSABE, and the geomorphometric properties of each were then compared. Analysis showed a systematic basin length decrease at coarser resolutions. Predictions based on the cell size ratios (equation (3)) allowed scaling corrections to be made to the river lengths at coarser resolutions. Comparison of the shape index and width functions at different resolutions and theoretical considerations suggest that 200–300 grid cells are necessary to maintain the geomorphometric characteristics of individual basins. Strahler order 4–5 basins typically have approximately the same number of grid cells. These results, although based on the European landmass, are applicable to all other continents, although the cell size relationship in (3) may require testing in other areas.

Appendix A

The grid representation of any quantity introduces discretization error similar to rounding error, where the quantity X is approximated by a finite number n of discrete elements (ΔX), $X^* = n\Delta X$, and X^* is the approximation of X. The maximum rounding error can be expressed as $\varepsilon = \Delta X/2$, and the relative error then becomes $\varepsilon = \varepsilon/X$ or $\varepsilon = \Delta X/2X$. The required resolution as a function of a desired accuracy ε can be expressed as

$$\Delta X = 2\varepsilon X. \tag{A1}$$

In the gridded river network context the two most important quantities are the catchment area and distance to the basin outlet. Applying (A1) to grid cell area, we find that to achieve a desired area accuracy ε_A, the grid cell area has to be smaller than

$$\Delta A = 2\varepsilon_A A, \tag{A2}$$

where A is the area of the smallest basin or subbasin we expect the gridded network to represent with ε_A accuracy. Therefore the minimum number of grid cells within the smallest subbasin equals

$$n = 1/(2\varepsilon_A). \quad \text{(A3)}$$

Since the distances to the basin outlet vary within the basin, it is harder to apply the same criteria to river lengths. We seek to represent the mean river $\bar{L}$ (the average distance from any point within the basin to the basin outlet) at some accuracy ε_L. Applying (A1) to mean river length $\bar{L}$, the necessary resolution ΔL becomes

$$\Delta L = 2\varepsilon_L \bar{L}. \quad \text{(A4)}$$

The catchment area of the smallest basin of interest is more often known than the mean length. We therefore relate ΔL resolution to catchment area by using a modified version of the basin shape (equation (2)), which relates main stem length to catchment area. Mean length shape S_m is

$$S_m = \bar{L}/\sqrt{A}, \quad \text{(A5)}$$

where A is the basin/subbasin area [km] and $\bar{L}$ is mean river length.

Using (A5), we can define $\bar{L} = S_m\sqrt{A}$. Substituting $\bar{L}$ in (A4) yields

$$\Delta L = 2\varepsilon_{\bar{L}} S_m \sqrt{A}. \quad \text{(A6)}$$

In (A6) the resolution was expressed as the distance (ΔL) between the adjacent grid cells. We approximate grid cell area as $\Delta A = \Delta L^2$. Therefore the resolution expressed as grid cell area ΔA becomes

$$\Delta A = 4\varepsilon_L^2 S_m^2 A. \quad \text{(A7)}$$

The minimum number of grid cells ($n = A/\Delta A$) becomes

$$n = 1/(4\varepsilon_L^2 S_m^2). \quad \text{(A8)}$$

Equations (A3) and (A8) represent two criteria for the minimum number of grid cells needed to maintain ε_L length and ε_A area accuracy. Assuming that the desired length and area accuracies are equal ($\varepsilon_L = \varepsilon_A = \varepsilon$), we find that for rounded (low shape value) basins the length criteria is typically more strict than the area criteria. We also note that the length criteria are also more strict when higher accuracy is required. For practical purposes (<10–20% error), the satisfaction of the length criteria requires more grid cells.

Acknowledgments. We acknowledge J. Famiglietti and F. Olivera for creative discussions on the NSA concept. We would like to thank Kristine Verdin, Francesco Olivera, Theodore Endreny, David Tarboton, and one anonymous reviewer who provided insightful reviews of our first draft. This work was supported by NSF (grant ATM-9707953), NASA-EOS (grant NAG5-6137), NASA-ESIP (grant NCC5-304), NASA-Altimetry (grant NAG5-8683), NASA-SeaWiFS (grant NAG5-6452), and DOE (grant DE-FGO2-92ER61473). We also recognize the valuable assistance of Franziskaner Weizenbier in developing the NSA algorithm.

References

Band, L. E., Extraction of channel networks and topographic parameters from digital elevation data, in *Channel Network Hydrology*, edited by K. Beven and M. J. Kirkby, pp. 13–42, John Wiley, New York, 1993.

Beven, K., and M. J. Kirkby, A physically base, variable contributing area model of basin hydrology, *Hydrol. Sci. Bull.*, *24*, 43–69, 1979.

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

Edwards, M. O., Global gridded elevation and bathymetry (ETOPO5), digital Raster data on a 5-minute geographic (lat × lon) 2160 × 4320 (centroid-registered) grid, Natl. Geophys. Data Cent., Natl. Oceanic and Atmos. Admin., Boulder, Colo., 1989.

Environmental Systems Research Institute (ESRI) ARCWorld [CD-ROM], Redlands, Calif., 1992a.

Environmental Systems Research Institute (ESRI), Digital chart of the world for use with ARC/INFO, data dictionary, 1:1M [CD-ROM], Redlands, Calif., 1992b.

Fekete, B. M., C. J. Vörösmarty, and W. Grabs, Global, composite runoff fields based on observed river discharge and simulated water balances, *Tech. Rep. 22*, Global Runoff Data Cent., Koblenz, Germany, 1999.

Gesch, D. L., 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*, 583–587, 1999.

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

Hutchinson, M. F., A new procedure for gridding elevation and stream line data with automatic removal of spurious pits, *J. Hydrol.*, *106*, 211–232, 1989.

Jenson, S. K., and J. O. Domingue, Extracting topographic structure from digital elevation data for geographic information system analysis, *Photogramm. Eng. Remote Sens.*, *54*, 1593–1600, 1988.

Lammers, R. B., A. I. Shiklomanov, C. J. Vörösmarty, B. Fekete, and B. J. Peterson, Assessment of contemporary Arctic river runoff based on observational discharge records, *J. Geophys. Res.*, *106*, 3321–3334, 2001.

Moore, I. D., R. B. Grayson, and A. R. Ladson, Digital terrain modelling: A review of hydrological, geomorphological and biological applications, *Hydrol. Processes*, 3–30, 1991.

O'Donnell, G., B. Nijssen, and D. Lettenmaier, A 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).

Pfafstetter, O., Classification of hydrographic basin: coding methodology, Dep. Nac. de Obras e Saneamento, Rio de Janeiro, Brazil, 1989. (Translated by J. P. Verdin, U.S. Bureau of Reclamation, Denver, Colo., 1991.)

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

Rinaldo, A., K. V. Gregor, R. Rigon, and I. Rodriguez-Iturbe, Can one gauge the shape of a basin, *Water Resour. Res.*, *31*, 1119–1127, 1995.

Steinwand, D. R., and J. A. Hutchinson, Map projections for global and continental data sets and analysis of pixel distortion caused by reprojection, *Photogramm. Eng. Remote Sens.*, *61*, 1487–1497, 1995.

U.S. Geological Survey, HYDRO 1K: Elevation derivative database, http://edcwww.cr.usgs.gov/landdaac/gtopo30/hydro/index.html, USGS EROS Data Center, Sioux Falls, S. D., 1998.

Veneziano, D., G. Moglen, P. Furcolo, and V. Iacobellis, Stochastic model of the width function, *Water Resour. Res.*, *36*, 1143–1157, 2000.

Verdin, K. L., A system for topologically coding global drainage basins and stream networks, paper presented at 17th Annual ESRI Users Conference, Environ. Syst. Res. Inst., San Diego, Calif., July 1997.

Vörösmarty, C. J., B. M. Fekete, M. Meybeck, and R. B. Lammers, Geomorphometric attributes of the global system of rivers at 30-minute spatial resolution, *Hydrol. Processes*, *237*, 17–39, 2000a.

Vörösmarty, C. J., B. M. Fekete, M. Meybeck, and R. B. Lammers, Global system of rivers: Its role in organizing continental land mass and defining land-to-ocean linkages, *Global Biogeochem. Cycles*, *14*(2), 599–621, 2000b.

B. M. Fekete, R. B. Lammers, and C. J. Vörösmarty, Institute for the Study of Earth, Oceans, and Space, University of New Hampshire, Morse Hall, 39 College Road, Durham, NH 03824, (balazs.fekete@unh.edu; richard.lammers@unh.edu; charles.vorosmarty@unh.edu)

(Received July 21, 2000; revised January 8, 2001; accepted January 11, 2001.)