



Estimating cell-to-cell land surface drainage paths from digital channel networks, with an application to the Amazon basin

Emilio Mayorga^{a,*}, Miles G. Logsdon^{a,1}, Maria Victoria R. Ballester^{b,2},
Jeffrey E. Richey^{a,1}

^a*School of Oceanography, University of Washington, Box 355351, Seattle, WA 98195-5351, USA*

^b*Centro de Energia Nuclear na Agricultura, Universidade de São Paulo, Piracicaba, SP 13416-000, Brazil*

Received 18 May 2004; revised 28 February 2005; accepted 18 March 2005

Abstract

Cell-to-cell surface flow paths are commonly derived from gridded digital elevation models (DEM) by choosing the direction of steepest descent to one of the eight surrounding cells. However, adequate DEMs often are not available. We developed a topography-independent method for creating gridded, land and stream drainage direction maps based on corrected vector river networks. We applied it to the Digital Chart of the World river network in the Amazon basin gridded at 0.005° resolution; in this basin, low relief and poor topographic data have prevented the effective use of DEM-based methods. We geo-registered 224 hydrographic gages against the processed network and compared extracted vs. published drainage areas. Drainage areas ranged from 227 to 4,620,000 km². Median relative error was 4.5%, increasing in smaller basins to 94% in basins ≤ 2000 km². The effective limit of reliability may differ from 2000 km² across the basin. The drainage direction map and derivative datasets represent an improvement over existing datasets for regional research in the Amazon basin. Methods exploiting vector networks complement terrain approaches, and combined they may yield advances in the automated extraction of drainage maps and handling of topologically realistic river systems.

© 2005 Elsevier B.V. All rights reserved.

Keywords: River networks; Drainage basins; Flow direction; Watershed extraction; Geographic information systems; Amazonia

1. Introduction

River channel networks are fundamental land surface features that integrate diverse and distant

regions through the transport of water, sediments, and chemical constituents. The dynamics at a given channel reach can be understood only in the context of the characteristics of its drainage area and the upstream river network. A systematic linkage of the land to its drainage channels involves the delineation of flow paths from each element on the land surface to a channel reach and down the river network to the mouth. Such integration is required in many significant earth science and resource management

* Corresponding author. Fax: +1 206 685 3351.

E-mail addresses: emiliom@u.washington.edu (E. Mayorga), mlog@u.washington.edu (M.G. Logsdon), vicky@cena.usp.br (M.V.R. Ballester), jrichey@u.washington.edu (J.E. Richey).

¹ Fax: +1 206 685 3351.

² Fax: +55 19 3429 4610.

applications, including the validation of land surface hydrology models using observed river discharge (Lohmann et al., 1998) and studies of the effects of basin characteristics and land use on river nutrient and carbon fluxes (Krysanova et al., 1998; Ludwig et al., 1996; Smith et al., 1997).

Surface flow paths are commonly derived from a gridded digital elevation model (DEM), where the elevation values are stored in regular grid cells. Several algorithms exist to determine the direction of water flow from cell to cell based on terrain analysis and DEMs (Martz and Garbrecht, 1992; Moore et al., 1991). The most convenient and widely used method is the D8 algorithm (Hogg et al., 1997; Maidment, 1993), in which the direction of flow out of each cell corresponds to the direction of steepest descent to one of the eight surrounding cells. A consequence of this method is that water and materials in a cell can flow to only one of its neighbors, a simplification that may result in drainage artifacts (Costa-Cabral and Burges, 1994). Nevertheless, other important parameters can be derived easily and unambiguously once unique flow direction paths are delineated; examples include flow accumulation grids (the number of cells draining to any given cell), distance along the network, river networks at a desired scale, and drainage basins (e.g. Vörösmarty et al., 2000a,b). Utilities to calculate these parameters from flow direction grids have been incorporated in most Geographical Information Systems (GIS) software, such as ArcInfo (ESRI, 1997) and GRASS (Neteler and Mitasova, 2004).

The use of DEMs and unique cell-to-cell flow paths greatly facilitates the extraction of surface flow characteristics, river networks, and drainage basins, and the development of hydrological and biogeochemical applications. Nevertheless, DEMs of sufficient resolution and accuracy do not exist in large areas of the world, especially in developing countries and in low-relief forested regions such as the Amazon lowlands; low-relief areas are specially challenging even when high-resolution DEMs are available. In such cases, alternative methods are needed to derive cell-to-cell surface flow paths covering the entire land surface, at a resolution high enough to support a variety of applications. Döll and Lehner (2002) discuss additional limitations of deriving drainage directions from DEMs.

Vector maps of channel networks are widely available, detailed, and verifiable sources of relevant information. Compared to elevation, channels are more easily and accurately extracted from aerial and satellite remote sensing, and errors more easily identified and corrected. However, available digital networks are often used only for visual display or proximity analysis (river buffers); datasets supporting primarily such applications may include a number of topological network complexities, including disconnected reaches, two-bank channel representation, and flow bifurcation. For small watersheds, topologically corrected networks are often used in network analysis, and in hydraulic channel routing models based on lumped sub-watersheds or hydrological representative units, where watershed boundaries may be extracted manually (Fread, 1992; Maidment, 1993).

The use of digitized channel networks to assist in delineating basin-wide land and river surface flow paths has been limited. In the 'stream burning' method, a river network from an external source is incised into a DEM by artificially lowering the elevation of cells that are part of this network, forcing flow paths to drain into and follow the imposed network (Graham et al., 1999; Liang and Mackay, 2000; Renssen and Knoop, 2000). This scheme reduces the impact of inaccuracies in DEMs, but still requires an adequate DEM while ensuring realistic flow paths only at the larger scales for which a river network is available. Sekulin et al. (1992) used a channel network and a nearest-neighbor method only to extract gridded drainage areas without the use of a DEM. More recently, Olivera and Raina (2003) developed the Network Tracing Method (NTM) to generate coarse gridded river networks from topologically corrected and simplified vector networks, independently of DEMs. This novel scheme combines network gridding and upscaling while preserving original channel lengths more consistently than previous upscaling methods (e.g. Fekete et al., 2001; Olivera et al., 2002). However, it neglects cells not containing channels in the vector dataset; such cells may represent streams missing from the vector network or areas dominated by subsurface flow.

We present an approach related to but more exhaustive than the one used by Sekulin et al. (1992). Like NTM, it relies on topologically corrected and

simplified channel networks to derive gridded cell-to-cell surface flow paths. However, the resulting flow paths encompass both true rivers and the entire land surface. Rather than incorporating a network upscaling procedure, this approach creates a base gridded flow-direction dataset using the finest cell resolution that is computationally practical. It is intended for use in regions where existing DEMs are inadequate, but detailed channel network maps are available. With a complete land and river flow direction grid created, surface flow path analysis functions can be applied in the same manner as with flow direction grids derived from DEMs. After describing the method, we illustrate and evaluate it with a large-scale application to the entire Amazon basin using a corrected version of the Digital Chart of the World (Danko, 1992) river network; the derived datasets are freely available and may be used in hydrological and biogeochemical modeling and in studies linking river observations to their corresponding drained areas, particularly in connection with the Large-scale Biosphere-atmosphere experiment in Amazonia (LBA).

2. Methods

Fig. 1 shows a general description of the steps involved in the approach presented here to create a complete flow direction grid from river network maps. These steps are grouped into three stages, described below. All data processing and algorithm implementation was carried out using a combination of custom C programs and ArcInfo software.

2.1. Preparation and gridding of input vector datasets

The need to create a basin-wide land and river flow direction grid depicting unique cell-to-cell paths imposes certain requirements and limitations on the vector network: (a) the entire network must be interconnected; (b) channels that form loops or are split into parallel braided channels must be simplified to single channels following single, well defined directions; and (c) channels represented as polygons or with individual river bank vectors must be simplified to single-curve centerlines to avoid multiple parallel gridded flow paths. The Amazon delta region in Northeastern Brazil illustrates these network

complexities (Fig. 2). Correction or simplification of such features may be facilitated through automated methods in a GIS, but ultimately must rely on visual inspection and manual corrections. Errors in drainage network configuration must be checked and corrected against ancillary data such as atlases and air photography. Commonly available network datasets will typically require extensive pre-processing.

Two other datasets are required. First, unique cell-to-cell flow paths require a single outlet. Rivers in estuaries and deltas are often split into multiple channels and mouths (see Fig. 2), and this complexity is depicted faithfully in vector channel networks. A single mouth point must be chosen manually; all other draining channels may either be forced to reverse their direction of flow or may be cut-off at arbitrary points. Reliance on DEMs and D8 to extract flow directions and river networks sidesteps the issue because the use of elevation gradients and unique cell-to-cell paths enforces an arbitrary unique mouth point, even when additional processing is involved. Second, the basin boundary corresponding to the chosen mouth point is needed. This polygon may be created manually by visual inspection or may originate in other datasets or methods; however, it must be consistent with the channel network map. The accuracy of this basin boundary may be difficult to quantify and is dependent on the background information used to extract the boundary—DEM, atlases, or the river network dataset itself.

At this point, the river network can be gridded using an extent that encompasses the basin boundary. The choice of cell resolution involves a balance between smaller cells to accurately represent channel sinuosity, avoid connecting adjacent channels, and avoid breaking-off meander loops at their necks; and larger cells to minimize processing time and file sizes. The optimal cell size is selected through trial and error. Finally, the gridded river network and basin boundary must be evaluated carefully to ensure that artifacts such as improperly connected channels are not introduced during gridding.

2.2. Flow direction algorithm

The goal of the algorithms is to create a basin-wide, land and river flow direction grid. The input datasets are the gridded versions of the pre-processed river network,

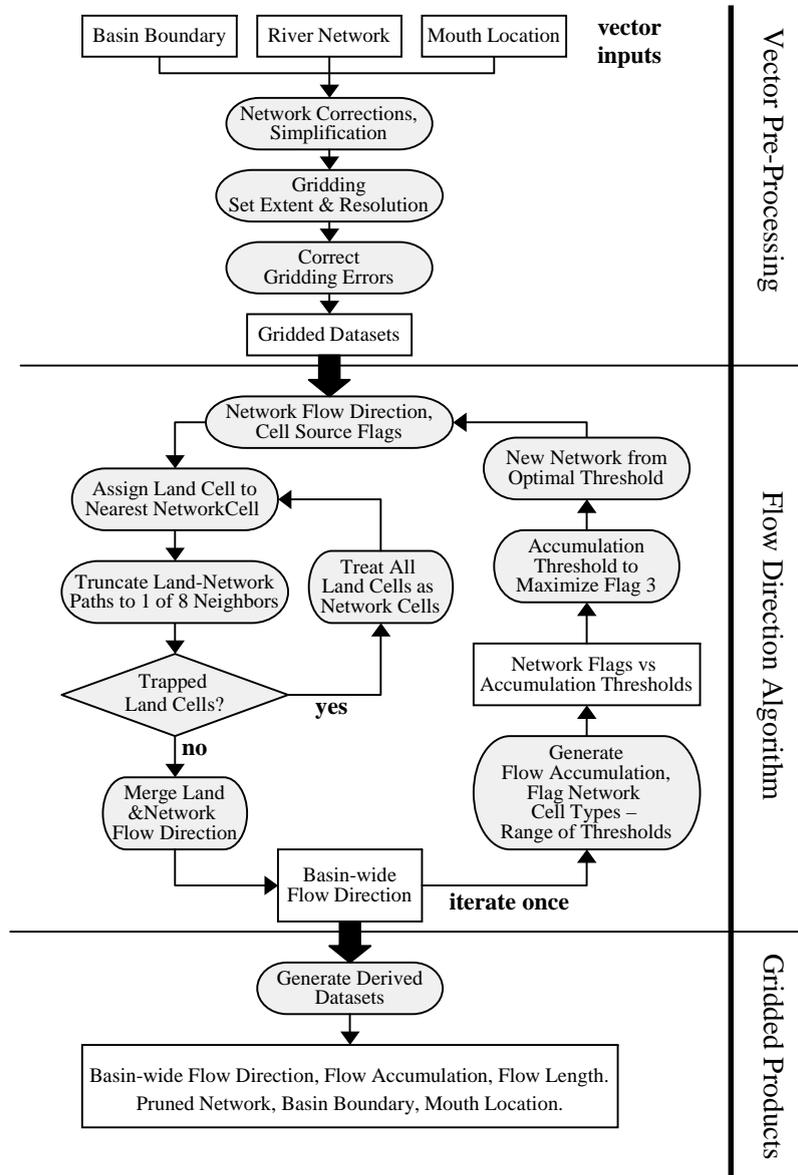


Fig. 1. Overview of steps to generate basin-wide drainage paths from vector river network and associated datasets.

basin, and mouth coordinates. For convenience, the network and basin are first combined into a single grid, assigning different integer cell values or flags to each. In the first step, in-channel cell-to-cell flow paths are assigned, leading to the mouth. These paths are determined as the shortest-path from each channel cell to the mouth, traversing the river network; flow can move from a cell to any one of its eight possible

neighbors within the network. An undesired consequence of this approach is that meanders and general curvature are often smoothed or shortened, thus reducing total channel length (Fig. 3). At this stage the mouth cell itself is assigned an arbitrary direction of flow pointing away from the basin.

Next, each 'land' cell in the basin—one not in the channel network—is allocated to the channel cell

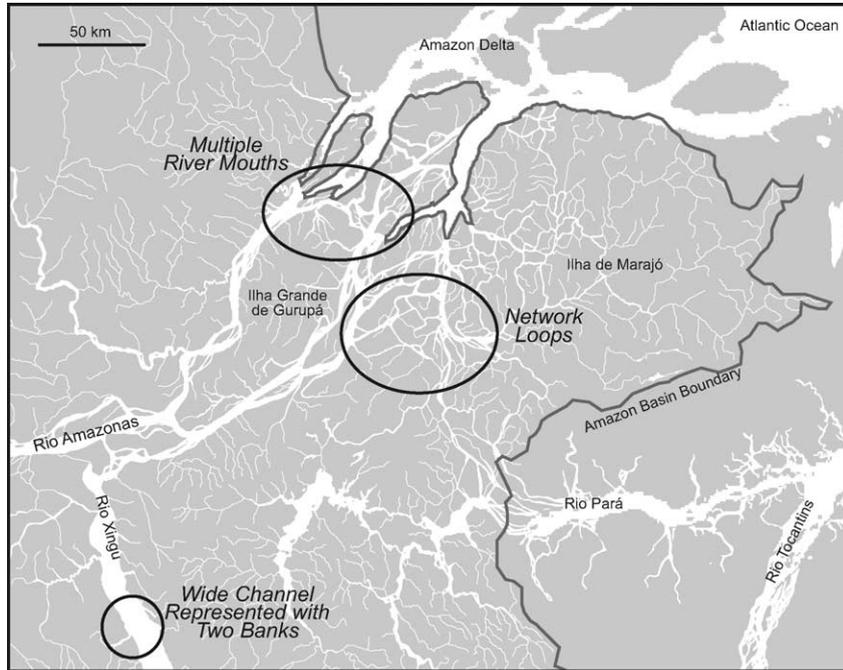


Fig. 2. Amazon River mouth region illustrating challenges encountered when using a vector river network dataset to derive drainage direction maps constrained to unambiguous, unique cell-to-cell paths. The ovals highlight three common types of problems.

closest to it along a straight line (the shortest Euclidean distance). To accomplish this, cells surrounding the land cell are scanned along progressively larger one-cell-wide square ‘rings’ in a clockwise

direction starting at the upper left corner of the ring. The distance and angle between the center of the land cell and that of its allocated river cell is recorded. A land-river line crossing cells outside the basin

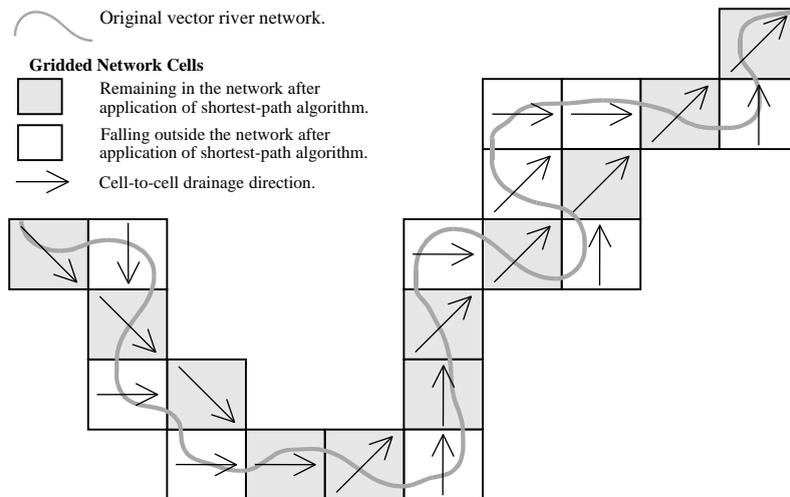


Fig. 3. Gridding of a vector river network reach and application of a shortest cell-to-cell path distance scheme to extract in-channel drainage paths on the gridded network. This example illustrates meander smoothing and overall shortening of network path lengths when applying the shortest-distance scheme.

implies flow paths leaving and re-entering the basin. Such physically impossible flow paths are invalid. Scanning proceeds until the first valid river cell assignment is made and such path is confirmed to be the shortest possible cell-to-cell distance. The invalidation of flow paths crossing the pre-defined basin boundaries can have an undesirable effect once all land cells have been processed. A land cell near a basin boundary can become ‘trapped’ by the boundary when no valid straight-line path to the center of a channel cell can be allocated. Such unallocated cells are identified and re-processed as a special case. In this step, all land cells successfully allocated are treated as channel cells together with real channel cells, and the allocation mechanism is repeated with this much larger set of targets. Identification and re-processing of unallocated cells continues until every land cell has been assigned a valid path. Finally, the angle of the land-stream connecting line is truncated to the nearest direction pointing from the center of the cell to the center of one of its eight neighboring cells. For example, a shortest-distance path with an angle of 60° NE will be rounded off to

a cell-to-cell path at 45° NE. Truncation is a key step that in effect changes the interpretation of the path direction from a long-distance nearest neighbor angle into a cell-to-cell flow direction.

At this point, a valid drainage direction map exists that can trace the path followed by a parcel of water from any point on the basin to the outlet. However, some artifacts result from gridding a vector network (Congalton, 1997) and applying a shortest-distance network path algorithm. These include short parallel drainage paths adjacent to the channel, and small clusters of pixels that are no longer part of the channel proper. Drainage paths from land cells may be diverted through these artifacts before they reach valid channel cells. To minimize this problem, we make use of the flow direction map generated above to identify potential artifact cells. The premise is that these cells cluster around the main derived channel cells and drain a relatively small number of cells, according to the flow accumulation map derived from the preliminary flow direction map. The steps are illustrated in Fig. 4. First, a support area threshold is chosen to extract a new river network from

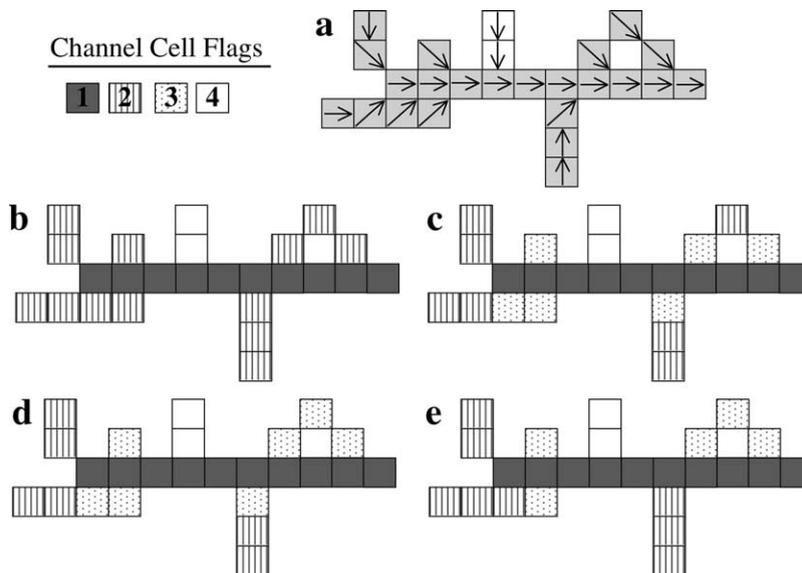


Fig. 4. Use of flow-accumulation thresholds to identify and minimize artifact channel cells resulting from gridding and channel flow direction steps: (a) schematic river network reach showing cell flow direction paths; dark cells are original gridded channels and white cells are former ‘land’ cells with a flow accumulation exceeding the current threshold. (b) Segmentation of the set of cells from (a) into three subsets or flags: original network cells meeting the support area threshold (1), original network cells not meeting the threshold (2), and former land cells that meet the threshold (4). (c) First pass to identify potential artifacts (flag 3) from flag-2 cells. (d) Second pass to identify further potential artifacts. (e) Identification of flag-3 cells likely to be valid and recoding to 2, to retain in the network.

the drainage direction map (Hogg et al., 1997). This new network is made up of both cells in the original gridded network and others previously considered land cells (Fig. 4a); the latter are artificially generated channels which are assigned a flag value of four and disregarded (Fig. 4b). Conversely, the flow accumulation threshold splits the original network into ‘channel’ (above the threshold) and ‘non-channel’ (below the threshold) cells. From this set of all original-network cells, channel cells are flagged as a value of one, and non-channel cells are flagged with a value of two (Fig. 4b). Next, flag-2 cells are further divided into valid original network cells and potential artifacts using a set of kernel calculations centered on each below-threshold cell from the original network. In the first step, flag-2 cells adjacent and perpendicular to a flag-1 cell are identified as potential artifacts and flagged as 3 (Fig. 4c). Second, a 3×3 kernel is used to flag additional potential artifacts. If a flag-2 cell has no flag-2 neighbors in the kernel (it is isolated) and has at least one flag-3 neighbor (it has adjacent potential artifacts), it is flagged as 3 (Fig. 4d). The final step identifies below-threshold cells that connect flag-1 and flag-2 cells but were incorrectly flagged as potential artifacts. If the cell is flagged as 3 and is adjacent and perpendicular to a flag-2 cell, it is recoded back to 2 (Fig. 4e).

Network cell flagging is carried out for a large range of flow accumulation thresholds. We select the threshold value that maximizes the number of artifact channel cells identified (flag 3). Cells flagged as 1 or 2 using this optimal support area are then selected to form a new, cleaned base river network. A number of small, valid stream channels will still be dropped, while some artifacts will remain. Finally, the flow direction algorithm described earlier is applied to this new network (Fig. 1). The resulting drainage direction map is the final, cleaned product.

2.3. Final dataset products

Additional products can be derived from the final flow direction map, base network, basin mask, and location of the basin outlet. We apply standard network traversal methods on the flow direction map to create flow accumulation and flow length gridded maps. In the flow length product, each cell holds the distance along the complete drainage path from

the cell to the basin outlet; the length of a horizontal or vertical path is equal to cell width, while the length of a diagonal path is equal to 1.4142 times cell width. Together, these maps and the final gridded network make up the complete, final dataset.

3. Results and discussion

The CAMREX research group (Carbon in the AMazon River Experiment) has studied the biogeochemistry, hydrology, and geomorphology of Amazonian rivers over the last two decades (Devol and Hedges, 2001; Dunne et al., 1998; Richey et al., 2002, 1989). Sampled rivers range from 1st order streams to the mainstem Amazon a few hundred kilometers above the mouth; systematic analysis of the extensive dataset used by CAMREX requires co-registration of sampling sites with a detailed basin-wide river network and extraction of watershed boundaries at a wide range of scales. The CAMREX group is also engaged in spatially distributed hydrological and biogeochemical modeling for the entire basin. These activities dictate the need for a single drainage dataset consisting of a high-resolution basin-wide river network and an associated drainage direction map to facilitate water routing modeling and watershed boundary extraction. The highest-resolution dataset currently available that spans the entire Amazon basin is HYDRO1k, a 1-km product derived from the GTOPO30 DEM (USGS, 1998; Verdin and Verdin, 1999). However, GTOPO30 and HYDRO1k are impacted by very low relief and poor data quality in the lowlands, and some artifacts in the highlands. A preliminary evaluation suggested that it was effectively unusable in the central Amazon (see comparison below). The next best datasets (Costa et al., 2002; Döll and Lehner, 2002; Graham et al., 1999; Vörösmarty et al., 2000b) are derivatives of ETOPO5 and TerrainBase, 5-min resolution global DEMs (NGDC, 1988, 1997). However, these datasets suffer from similar problems and the resolution was too coarse for our research requirements. Instead, we used the Digital Chart of the World (DCW, Danko, 1992) vector river network and the method presented in this paper to create a new, high-resolution flow direction map and derived datasets at 0.005° resolution (decimal degrees; approx. 500 m) for

the Amazon basin proper, excluding the Rio Tocantins basin. In this section, we describe the procedures used to prepare these datasets and the creation of a flow direction map. Finally, we evaluate the resulting datasets against HYDRO1k and a database of drainage areas for hydrographic gages co-registered with the DCW-derived river network.

3.1. Data sources and pre-processing

This dataset was processed in geographic coordinates (degrees latitude and longitude). A resolution of 0.005° was chosen through trial and error as an acceptable balance between excessive file size and presence of river network gridding errors. The Earth's curvature leads to changes with latitude in the projected, planar side length of each 0.005° cell; a projected coordinate system would eliminate such distortions. However, the Amazon basin straddles the equator, minimizing distortions, while the use of geographic coordinates facilitates integration with other continental datasets for modeling and analysis. We used a constant factor of $1^\circ = 111$ km to convert to planar area and length.

3.1.1. Basin boundary

We delineated the boundary of the Amazon basin (excluding the Tocantins basin) by manually tracing mountain ridges and watershed divides in ETOPO5, using the DCW networks and published paper maps as reference. This basin boundary was created before GTOPO30 became available. ETOPO5's coarse resolution and associated topographic distortion may have resulted in considerable uncertainty in the accuracy of the boundary.

3.1.2. River network

The original DCW global river network vector dataset is distributed as two subsets: lakes and large rivers represented as polygons by their two banks, and smaller rivers represented as centerline arcs only. Using the previously created Amazon basin boundary, we first selected the DCW polygons and arcs within the basin. The two subsets were then combined while eliminating topological errors. The polygonal dataset was processed first. Isolated polygons were assumed to represent small lakes and in-channel islands and were eliminated. Remaining polygons were then gridded at a resolution

finer than 0.005° , reduced to single-cell thickness channels using ArcInfo's 'thin' function, and converted back to vectors. Next, the main task was to combine this new centerline representation of large channels with the DCW centerline vector subset. The polygonal channels were originally connected to channels in the centerline subset at their banks, but this connection was broken when polygonal channels were simplified to centerlines. We developed a semi-automatic procedure to extend the terminal end of connecting arcs in the original centerline-only subset to the new channel centerlines, to re-establish connectivity. The combined vector dataset was then extensively evaluated against the original DCW network and paper reference maps. The manual corrections carried out include: addition of missing network branches and reaches, correction of wrong confluence locations, and break up of network bifurcations. Finally, when gridding created an incorrect connection between two parallel channels, the smaller channel was moved away to increase the separation.

3.1.3. Mouth location

The Amazon delta is a complex region without a single clear mouth. We chose as the mouth a terminal point of the northern mainstem sub-channel that starts out below the confluence of the Amazon mainstem and Rio Xingu and drains north of Ilha de Marajó (Marajó island) and Ilha Grande de Gurupá, at 51.4175° W, 0.4275° S (see Fig. 2). The stream network in the delta area was greatly simplified to force all flow into this single mouth point.

3.2. Flow direction algorithm and final products

Applying the procedures described in Section 2 to these Amazon basin datasets, we first obtained a set of network flag counts for a sequence of flow accumulation thresholds ranging from 5 to 10,000 cells (Fig. 5). Artifact cells identified reached a maximum at a threshold of 200 cells. This threshold was used to create a new base network for the flow direction algorithm (Fig. 1). Finally, we created flow accumulation and flow length maps based on the resulting flow direction map, using standard algorithms. We used flow accumulation to derive a new river network that combines original network cells with synthetic flow paths not present in the original network. Cells that receive no drainage from other network cells were

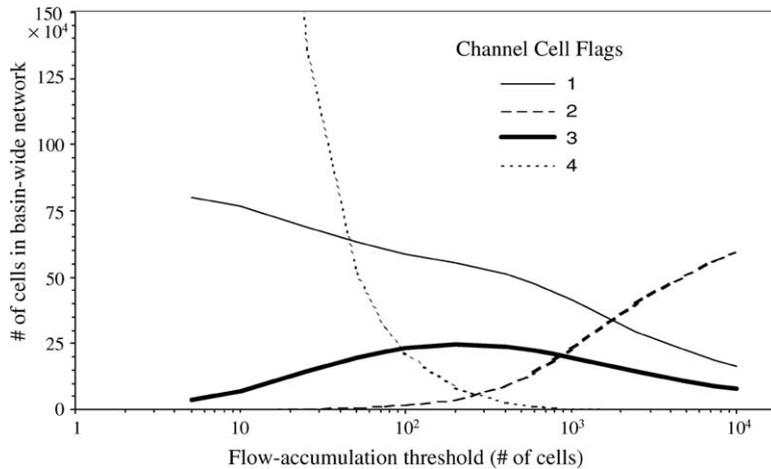


Fig. 5. Network cell flags vs. flow-accumulation threshold (log-scale) for the Amazon basin. There were 838,008 0.005° cells in the original gridded network. See text and Fig. 4 for explanations of cell flag values.

assumed to be Strahler first-order streams. This assumption is probably appropriate in most humid, lowland forested parts of the basin but is problematic in more arid regions. Nevertheless, it allowed us to assign a Strahler stream order to all river reaches in the basin and to group reaches of similar scale when necessary. Using this scheme, the mainstem below the confluence of Rios Solimões and Negro near Manaus to its mouth is a 10th order reach. For comparison, this reach is 6th order in the STN-30p global 0.5° drainage direction dataset (Vörösmarty et al., 2000b).

To assist in future basin analysis, we divided the Amazon basin into major tributary subbasins and the local drainages and floodplain area around the Amazonas–Solimões mainstem. This division involved two assumptions. First, we defined the beginning of the mainstem at the confluence of the Rios Marañón and Ucayali in northeastern Perú. Traditionally, the river is named Amazonas from this point to the Brazilian border, where it becomes known as Solimões; below its confluence with Rio Negro it is known again as Amazonas. Our definition reflects the mainstem's globally significant size beginning in Perú. Second, we designated as a major tributary every subbasin whose outlet draining directly into the mainstem has a flow accumulation value of at least 40,000 cells (approx. $12,230 \text{ km}^2$). Thirty-two major subbasins were extracted, in addition to the Amazonas–Solimões mainstem drainage area. Fig. 6 shows

the resulting subbasin boundaries, together with river reaches with a Strahler order of six and higher. Table 1 lists some hydrographic properties of the major tributaries and the mainstem floodplain drainage area. The complete gridded dataset, including flow direction, flow accumulation, flow length, drainage network, and subbasins, can be downloaded from the LBA data distribution site, <http://lba.cptec.inpe.br/beija-flor/>. The 0.005° grids have a dimension of 5300 rows by 6400 columns and extend from 80.5 to 48.5° W and 20.5° S to 6.0° N. Additional datasets available include a derived vector dataset of tributary drainage boundaries and a re-processed version of the original DCW vector river network.

Like other flow direction datasets and associated river networks, the CAMREX dataset has certain limitations, some of which have already been discussed. The method presented here results in lower sinuosity and overall reduction of channel length as a consequence of shortest-path channel flow allocation (Fig. 3). Further pre-processing of the vector network before gridding may eliminate the reliance on shortest Euclidean paths. The following sequence of steps may represent a promising approach: (a) automated scanning and editing of all arcs in the network to ensure that they point downstream; (b) assignment to each arc of cumulative distance from mouth; (c) gridding the network while transferring cumulative distance values to the cells;

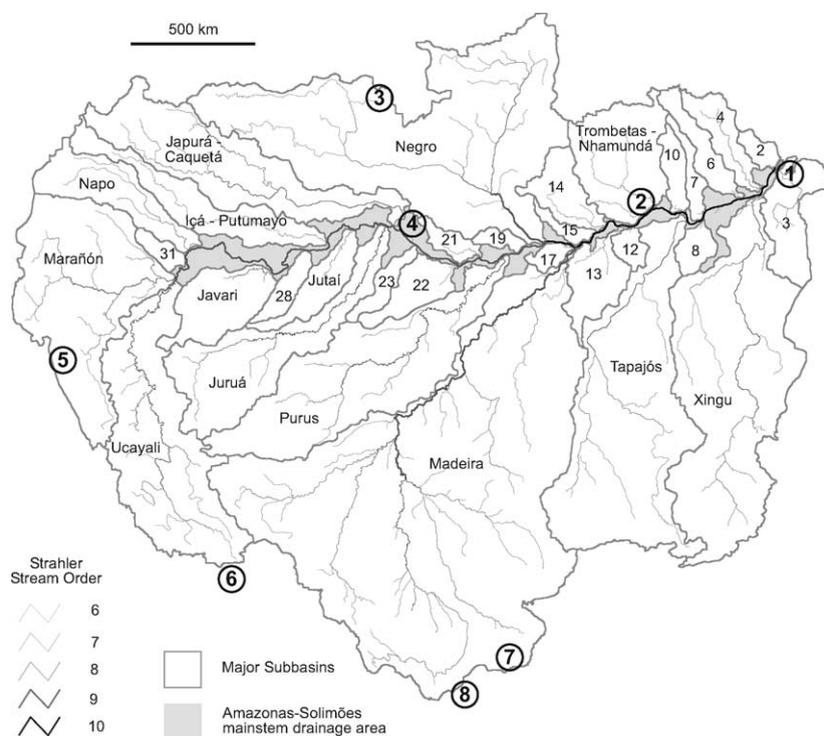


Fig. 6. Amazon major subbasins and river network extracted from the drainage direction map. River reaches with a Strahler order of six and higher are shown. Each of the larger subbasins is labeled with the name of its mainstem(s); smaller basins are labeled with their subbasin numbers (see Table 1). Circled numbers indicate prominent hydrographic ambiguities and data problems: (1) mouth of the Amazon, with multiple channels draining into the Ocean, bifurcations, small coastal catchments, and connection to the Tocantins river. (2) Combined Trombetas and Nhamundá confluence into the mainstem. (3) Linkage to the Orinoco Basin via the Negro–Casiquire connecting channel. (4) Bifurcation of the Japurá–Caquetá confluence into the mainstem. (5) Maraón headwaters mainstem dropped from the DCW network and reconstructed manually. (6) Apurimac source headwaters dropped from the basin. (7) Ambiguous hydrological connectivity in the Bañados del Izoçog wetlands. (8) Itenez headwaters in the Itonamas–Parapeti river dropped from the basin.

and (d) flow direction allocation based on cumulative distance cell values along the gridded channel network.

Additional limitations remain. Some channel distortions originally present in the DCW were not corrected. Channel bifurcations cannot be represented correctly. The Amazon and some tributary mainstems are often wider than one 500-m cell, but drainage paths are only one-cell wide. In addition, a number of small DCW streams were deleted during the initial pre-processing stage. Known, geographically specific problem areas are described in Fig. 6.

3.3. Validation

The only drainage direction map available at a scale similar to ours is HYDRO1k (Verdin and

Verdin, 1999). This dataset has the advantage of being consistent with topography. However, the river networks represented in HYDRO1k in the Amazon lowlands contain severe drainage errors as a consequence of low relief and poor quality of the source elevation data in that region. Fig. 7 shows a sample network error in the Juruá headwaters near Perú. In HYDRO1k, the entire Juruá headwater system drains incorrectly as parallel channels into the Javari basin to the North. Other portions are captured by the Purus on the East (not shown). Large errors in channel location and topology are also observed in the Rio Negro mainstem, the Amazonas mainstem near the Perú–Brazil border, and the Japurá–Caquetá headwaters. These problems would be difficult to correct without additional high-quality topographic data, unless a detailed channel network were incised into

Table 1
Major Amazon subbasins and drainage area of the mainstem

Number	Name	Area (km ²)	Length (km)	Subbasin outlet		
				Longitude (°W)	Latitude (°S)	Distance (km)
1	Amazonas–Solimões	6,020,438	3557	51.42	0.43	0
2	Macapá	16,147	318	51.45	0.50	9
3	Marajó-Pará	84,027	784	51.44	0.50	10
4	Jari	51,893	769	51.87	1.17	108
5	Xingu	515,651	2275	51.96	1.43	141
6	Paru de Este	43,640	731	52.63	1.56	221
7	Maicuru	21,917	547	53.98	2.03	398
8	Curuá-Una	24,505	315	54.08	2.34	435
9	Tapajós	498,063	2291	54.65	2.37	516
10	Curuá	28,099	484	54.80	2.08	560
11	Trombetas–Nhamundá	157,568	744	55.83	1.93	691
12	Mamuru	17,533	237	56.66	2.60	819
13	Ilha Tupinambarana	68,222	948	56.71	2.60	825
14	Uatumã	59,282	701	57.55	2.41	935
15	Urubu	13,892	405	58.06	2.84	1019
16	Madeira	1,381,696	3518	58.75	3.36	1135
17	Madeirinha	13,634	288	58.84	3.36	1146
18	Negro	719,216	2362	59.87	3.14	1274
19	Manacapuru	13,100	291	60.64	3.32	1376
20	Purus	362,981	2561	61.47	3.67	1493
21	Badajós	21,575	413	62.33	3.76	1614
22	Coari	52,715	599	63.09	4.07	1714
23	Tefé	25,877	571	64.60	3.32	1923
24	Japurá–Caquetá	267,735	1953	65.49	2.57	2066
25	Juruá	218,183	2096	65.70	2.60	2093
26	Jutai	55,826	912	66.95	2.73	2278
27	Içá–Putumayo	125,740	1770	67.93	3.13	2440
28	Jandiatuba	26,927	493	68.69	3.47	2575
29	Javari–Yavari	98,366	893	69.94	4.33	2839
30	Napo	110,378	1081	72.69	3.37	3268
31	Nanay	17,564	348	73.16	3.70	3340
32	Marañón	358,496	1561	73.52	4.48	3492
33	Ucayali	341,235	2269	73.52	4.48	3492

The subbasin number of the mainstem drainage is one, and values increase upstream from the mouth. Basin names are taken from the local name of the main tributary trunk. When the trunk has different names across national boundaries, both names are included. Names in bold are subbasins that are not originally made up of a single tributary system but were simplified in order to remove network bifurcations. Outlet distance is the distance from the subbasin outlet to the mouth of the Amazon. Length is the longest drainage distance from a source point in the subbasin to its outlet into the Amazon mainstem.

the DEM to enforce correct drainage paths. Therefore, we feel HYDRO1k cannot be used reliably for validating the DCW-derived Amazon dataset.

We co-registered with the CAMREX network dataset 224 hydrographic gages that include associated drainage areas in their original metadata; these published drainage areas range from 227 km² to approximately 4,620,000 km² at Óbidos, on the Amazonas mainstem. Two hundred and ten sites are in Brazil (ANEEL, 1987) and 14 in Bolivia (Carrasco

and Bourges, 1992; Guyot, 1993; Guyot et al., 1989, 1988). Discharge or stage data are also available. Geo-reference information associated with the gages was of variable quality. For Brazil, data was received electronically. Latitude and longitude coordinates were present in most cases, but their accuracy and origin are not known. Metadata also included river, roads, or town names in most cases. Uncertainty in data sources and reliability is especially problematic near river confluences and for relatively small

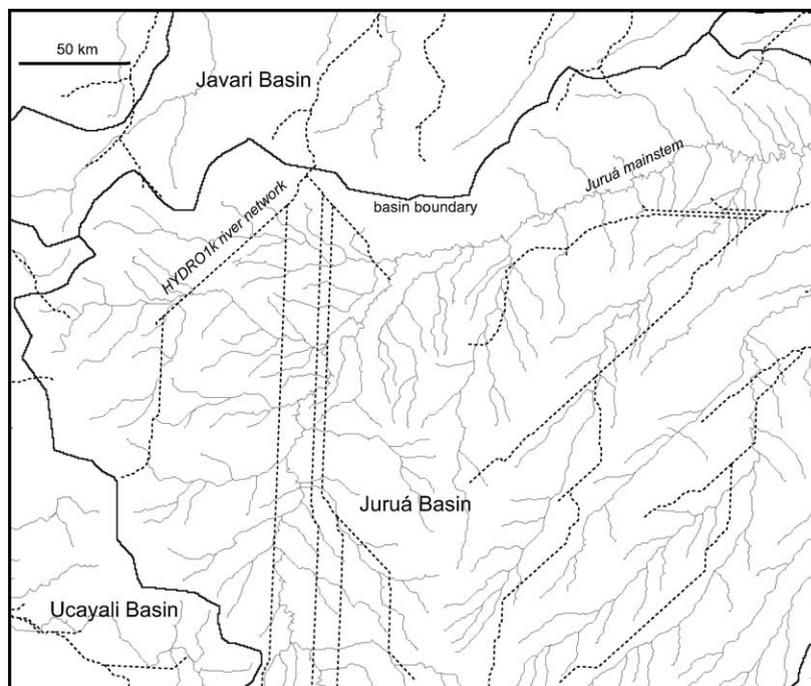


Fig. 7. Comparison of CAMREX river network (solid lines) and HYDRO1k network (dotted lines) in the Juruá headwaters, in the lowlands of the Western Amazon basin. The basin boundaries shown are the subbasin boundaries presented in Fig. 6 and created from the CAMREX drainage direction map. In this region, HYDRO1k drains the entire Juruá headwater system as parallel straight channels into the Javari basin to the North.

watersheds, where local stream names were not accessible to us. For Bolivia, gage location, drainage area, and all additional information was gleaned from publications and digitized manually. Registration of all gage sites against the river network was carried out and verified manually, using descriptive information when the coordinates yielded ambiguous locations. Even when coordinates were available, site location had to be shifted towards the digital network in most cases. Finally, drainage areas were extracted for all co-registered sites using the flow direction and flow accumulation maps created from the CAMREX network dataset.

We used this hydrographic database and extracted basin areas to evaluate the CAMREX drainage direction map. Fig. 8a presents a comparison of extracted vs. published area values. While overall there is good agreement, the relative error is larger in smaller basins. We examined the distribution of the percent error of CAMREX vs. published drainage area (A_{CAMREX} and A_{pub} , respectively), where percent error is calculated as $100 (A_{\text{CAMREX}} - A_{\text{pub}}) / A_{\text{pub}}$

(Fig. 8b). The histogram is divided into four basin area (A_{pub}) classes: ≤ 2000 , 2000–10,000, 10,000–20,000, and $> 20,000$ km². Summary results for all basins and for each area class are presented in Table 2. Overall, the median and mean errors are 4.5 and 16%, respectively. Fifty-five percent (73%) of estimated areas are within 10 (20%) of published areas. This level of agreement is similar to that shown by STN-30p when compared to a global database of published drainage areas (Vörösmarty et al., 2000b). However, 23 basins have an error $> 50\%$, with a maximum as high as 372%. Results by area class show that the error gradually increases as the size of the basin decreases, and for basins ≤ 2000 km² the majority of the areas are severely overestimated. When these 19 basins are removed, the median and mean errors are reduced to 3.6 and 7.4%, respectively (Table 2), and 60 (79%) of estimated areas are within 10 (20%) of published areas.

Two thousand square kilometers may represent an effective limit of reliability for the extraction of basin areas in this dataset. However, other considerations

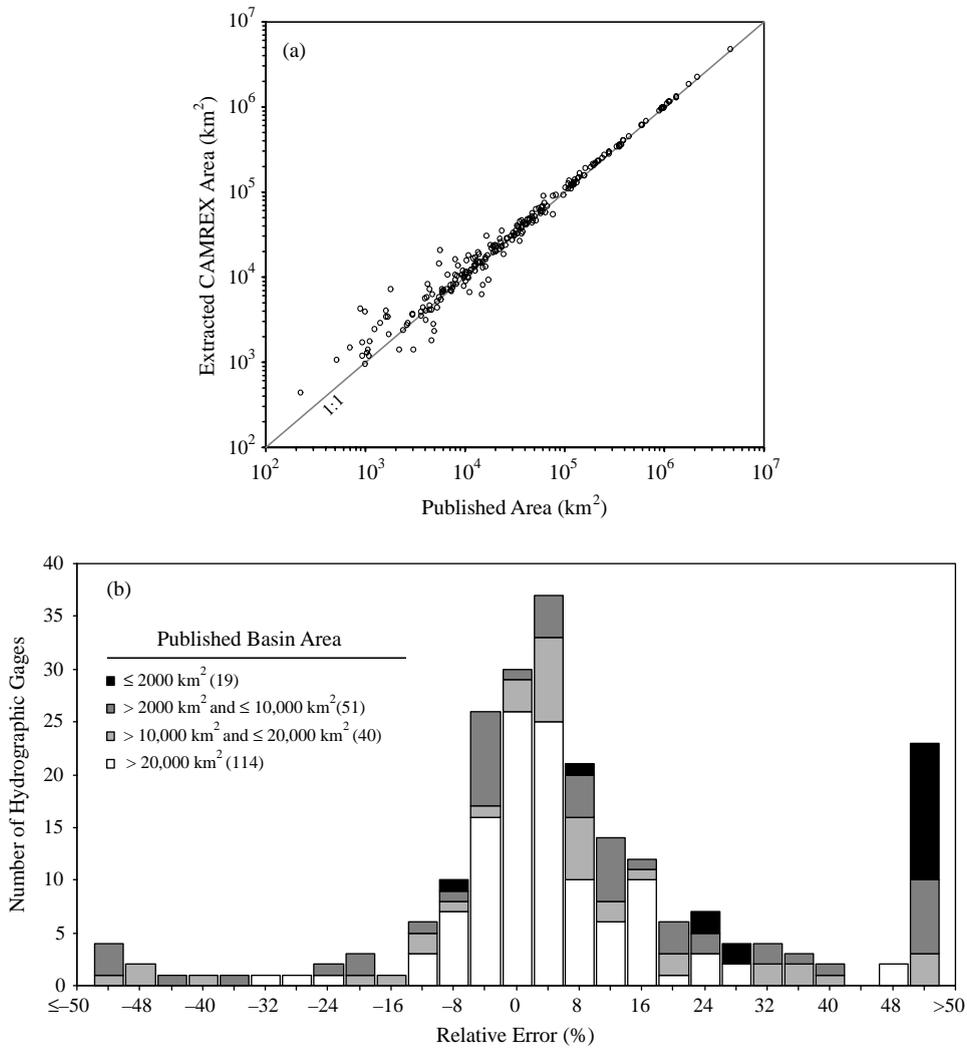


Fig. 8. Drainage basin areas derived from the CAMREX drainage direction map compared with published drainage areas from the hydrographic gage database ($n=224$). (a) Scatter plot of CAMREX vs. published areas; the 1:1 line is also plotted. (b) Frequency distribution of relative error partitioned into four drainage area classes. The number of gage stations in each area class is shown in parenthesis.

Table 2

Summary statistics of the relative error (%) between extracted and published drainage areas, by published area class

	Drainage area classes (km ²)				All sites	> 2000
	≤ 2000	2000–10,000	10,000–20,000	> 20,000		
Count	19	51	40	114	224	205
Minimum	-6.9	-61	-59	-30	-61	-61
Maximum	372	272	82	49	372	272
Median	94	7.8	5.5	2.2	4.5	3.6
Mean	108	16	6.8	3.8	16	7.4

The last column includes only sites with drainage area greater than 2000 km².

must be taken into account. First, the accuracy of published basin areas is unknown. Second, determining the exact location of hydrographic gages on the CAMREX network was more difficult for smaller basins due in part to the inaccessibility of local river names. Third, the stream density of any given area in the original DCW network in the Amazon basin appears to reflect the ease of interpretation from air photos and degree of effort spent digitizing that area, rather than an inherent stream density. Consequently, the limit of reliability for basin extraction is unlikely to be a uniform threshold but will vary across the basin to values smaller or larger than 2000 km². More tests are needed to estimate the reliability throughout the basin. Additional analysis is also required to determine the reason for the observed positive bias in small basins, to assess the contribution of the flow direction method and inaccuracies in published basin areas, location of gages, and the DCW network.

We did not attempt to reduce the disagreement by adjusting the locations of sites with larger errors. Future work may involve adjustments of site location as well as improvements in channel representation. A potential source of more accurate river networks is radar imagery from the JERS-1 sensor spanning almost the entire basin (Siqueira et al., 2000); Muller et al. (1999) have demonstrated the potential of such data for channel extraction in the Rio Negro subbasin of the Amazon.

4. Conclusion

We developed a method for creating a gridded, land and stream drainage direction map based on a vector river network and independent of topography. This scheme is intended for use in situations where an adequate digital elevation model is not available. It may be applied at a wide range of scales, as its reliability is dependent primarily on the accuracy and resolution of the river network dataset. The most time-consuming step involves the pre-processing of the vector river dataset to create a connected and topologically simple gridded network. While an obvious weakness of this method is its potential inconsistency with topographic data, the ability to create flow direction maps from vector networks can expand the use of river transport modeling

and automatic watershed extraction, and foster the co-registration of field measurements against a consistent dataset. An additional strength is the ability to easily improve the accuracy of all or parts of a basin through seamless insertion of improved river networks when these become available.

We applied the method to the Digital Chart of the World river network in the Amazon basin at a nominal 0.005° resolution. This dataset required extensive pre-processing to create a consistent, usable network. The derived CAMREX basin-wide flow direction map was used to identify 32 major subbasins draining into the mainstem. We geo-registered 224 hydrographic gages against the processed network, and used the published drainage area for each site to evaluate the CAMREX drainage direction map. Published drainage areas for these sites ranged from 227 km² to approximately 4,620,000 km². The overall level of agreement is comparable to that shown by global-scale datasets when compared to global databases for hydrographic sites (Vörösmarty et al., 2000b). However, the relative error increases in smaller basins and becomes very large in basins smaller than 2000 km². While this basin area may be seen as an effective limit of reliability for the CAMREX dataset, we suggest that such a threshold in fact will vary across the basin depending on the accuracy of the original DCW network and of our co-registration of hydrographic sites against the network.

The CAMREX drainage direction map represents an improvement over currently available alternatives (Costa et al., 2002; Döll and Lehner, 2002; Graham et al., 1999; Verdin and Verdin, 1999; Vörösmarty et al., 2000b) for regional research in the Amazon basin. The complete gridded dataset and associated vector datasets can be downloaded from the LBA data distribution web site. We have also co-registered 280 CAMREX and other biogeochemical sampling sites from mountain and lowland rivers in Brazil, Bolivia, and Perú; these sites are linked to a relational database holding corresponding biogeochemical measurements. Future updates may involve improvements to the network with more accurate and finer-scale networks; corrections to the Amazon basin boundary; and adjustments to the site location to reduce the relative error. The availability of high-quality 90 and 30-m DEMs from the Shuttle Radar Topography

Mission (SRTM) in the coming years (Farr and Kobrick, 2000) and subsequent extraction of a drainage direction map will bring dramatic improvements over this and other existing datasets. Nevertheless, approaches that fully exploit vector river networks such as the one presented here and NTM (Olivera and Raina, 2003) complement DEM-based schemes and should prove invaluable for improving the gridded representation of topologically realistic river systems.

Acknowledgements

Harvey Greenberg and Amy Miller carried out much of the original basin boundary delineation and DCW network pre-processing. The Brazilian hydrographic gage data was provided by Reynaldo Victoria (CENA, Piracicaba, Brazil), Tom Dunne (University of California at Santa Barbara) and Brad Newton (UCSB). This work was supported by NSF-DEB and NASA EOS and LBA funding to the CAMREX group, and by a NASA Earth System Science Fellowship to E.M. This paper was greatly improved by comments from Bernhard Lehner, Balasz Fekete, and an anonymous reviewer. This is CAMREX publication 127.

References

- ANEEL, 1987. Anon., 1987. Inventário das estações fluviométricas, Agência Nacional de Energia Elétrica, Brasília, Brazil 1987.
- Carrasco, N., Bourges, L.M., 1992. Estudio del régimen del escurrimiento superficial en la cuenca Amazónica del río Beni. In: Roche, M.A., Bourges, J., Salas, E., Diaz, C. (Eds.), Seminario sobre el Programa Hidrológico y Climatológico de la Cuenca Amazónica de Bolivia (PHICAB), La Paz, Bolivia, pp. 41–49.
- Congalton, R.G., 1997. Exploring and evaluating the consequences of vector-to-raster and raster-to-vector conversion. *Photogrammetric Engineering and Remote Sensing* 63, 425–434.
- Costa-Cabral, M.C., Burges, S.J., 1994. Digital elevation model networks (DEMON): a model of flow over hillslopes for computation of contributing and dispersal areas. *Water Resources Research* 30, 1681–1692.
- Costa, M.H., et al., 2002. A macroscale hydrological data set of river flow routing parameters for the Amazon Basin. *Journal of Geophysical Research* 107 doi:10.1029/2001JD000309.
- Danko, D.M., 1992. The digital chart of the world. *GeoInfo Systems* 2, 29–36.
- Devol, A.H., Hedges, J.I., 2001. Organic matter and nutrients in the mainstem Amazon River. In: McClain, M.E., Victoria, R.L., Richey, J.E. (Eds.), *The Biogeochemistry of the Amazon basin*. Oxford University Press, New York, pp. 275–306.
- Döll, P., Lehner, B., 2002. Validation of a new global 30-min drainage direction map. *Journal of Hydrology* 258, 214–231.
- Dunne, T., Mertes, L.A.K., Meade, R.H., Richey, J.E., Forsberg, B.R., 1998. Exchanges of sediment between the flood plain and channel of the Amazon river in Brazil. *Geological Society of America Bulletin* 110, 450–467.
- ESRI, 1997. ARC/INFO version 7.1.1. ESRI Press, Redlands, CA.
- Farr, T., Kobrick, M., 2000. Shuttle radar topography mission produces a wealth of data. *EOS Transactions AGU* 81, 583–585.
- Fekete, B.M., Vörösmarty, C.J., Lammers, R.B., 2001. Scaling gridded river networks for macroscale hydrology: development, analysis, and control of error. *Water Resources Research* 37, 1955–1967.
- Fread, D.L., 1992. Flow routing. In: Maidment, D.R. (Ed.), *Handbook of Hydrology*. McGraw-Hill, New York, pp. 10.1–10.36.
- Graham, S.T., Famiglietti, J.S., Maidment, D.R., 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. *Water Resources Research* 35, 583–587.
- Guyot, J.L., 1993. Hydrogeochimie des fleuves de l'Amazonie bolivienne, PhD Thesis, Editions de l'ORSTOM, Paris.
- Guyot, J.L., et al., 1988. Exportation de matieres en suspension des Andes vers l'Amazonie par le Rio Beni. Bolivie. *IAHS Publ.* 174, 443–451.
- Guyot, J.L., 1989. Transport of Suspended Sediments to the Amazon by the Andean River: the River Mamore, Bolivia, Fourth International Symposium on River Sedimentation. ISRS, Beijing, China, pp. 106–113.
- Hogg, J., McCormack, J.E., Roberts, S.A., Gahegan, M.N., Hoyle, B.S., 1997. Automated derivation of stream-channel networks and selected catchment characteristics from digital elevation models. In: Mather, P.M. (Ed.), *Geographical Information Handling: Research and Applications*. Wiley, Chichester, pp. 211–235.
- Krysanova, V., Müller-Wohlfeil, D.-I., Becker, A., 1998. Development and test of a spatially distributed hydrological/water quality model for mesoscale watersheds. *Ecological Modelling* 106, 261–289.
- Liang, C., Mackay, D.S., 2000. A general model of watershed extraction and representation using globally optimal flow paths and up-slope contributing area. *International Journal of Geographical Information Science* 14, 337–358.
- Lohmann, D., et al., 1998. The Project for Intercomparison of Land-surface Parameterization Schemes (PILPS) Phase 2(c) Red-Arkansas River basin experiment: 3. Spatial and temporal analysis of water fluxes. *Global and Planetary Change* 19, 161–179.
- Ludwig, W., Probst, J.L., Kempe, S., 1996. Predicting the oceanic input of organic carbon by continental erosion. *Global Biogeochemical Cycles* 10, 23–41.

- Maidment, D.R., 1993. GIS and hydrologic modeling. In: Goodchild, M.F., Parks, B.O., Steyaert, L.T. (Eds.), *Environmental Modeling with GIS*. Oxford University Press, New York, pp. 147–167.
- Martz, L.W., Garbrecht, J., 1992. Numerical definition of drainage network and subcatchment areas from digital elevation models. *Computers and Geosciences* 18, 747–761.
- Moore, I.D., Grayson, R.B., Ladson, A.R., 1991. Digital terrain modeling: a review of hydrological, geomorphological, and biological applications. *Hydrological Processes* 5, 3–30.
- Muller, F., Seyler, F., Guyot, J.L., 1999. Utilisation d'imagerie radar (ROS) JERS-1 pour l'obtention de réseaux de drainage, Example du Rio Negro (Amazonie), *Hydrological and Geochemical Processes in Large Scale River Basins*. HiBAm, Manaus, Brazil, p. 112.
- Neteler, M., Nitasova, H., 2004. *Open Source GIS: A GRASS GIS approach*, 2nd Edition. Kluwer, Dordrecht, The Netherlands, 424pp.
- NGDC, 1988. ETOPO5 Digital relief of the surface of the Earth, <http://www.ngdc.noaa.gov>. NOAA, National Geophysical Data Center, Boulder, Colorado.
- NGDC, 1997. TerrainBase global terrain model, <http://www.ngdc.noaa.gov>. NOAA, National Geophysical Data Center, Boulder, Colorado.
- Olivera, F., Raina, R., 2003. Development of large scale gridded river networks from vector stream data. *Journal of the American Water Resources Association* 39, 1235–1248.
- Olivera, F., Lear, M.S., Famiglietti, J.S., Asante, K., 2002. Extracting low-resolution river networks from high-resolution digital elevation models. *Water Resources Research* 38, 1231 doi:10.1029/2001WR000726.
- Renssen, H., Knoop, J.M., 2000. A global river routing network for use in hydrological modeling. *Journal of Hydrology* 230, 230–243.
- Richey, J.E., et al., 1989. Sources and routing of the Amazon River flood wave. *Global Biogeochemical Cycles* 3, 191–204.
- Richey, J.E., Melack, J.M., Aufdenkampe, A.K., Ballester, M.V., Hess, L.L., 2002. Outgassing from Amazonian rivers and wetlands as a large tropical source of atmospheric CO₂. *Nature* 416, 617–620.
- Sekulin, A.E., Bullock, A., Gustard, A., 1992. Rapid calculation of catchment boundaries using an automated river network overlay technique. *Water Resources Research* 28, 2101–2109.
- Siqueira, P., et al., 2000. A continental-scale mosaic of the Amazon basin using JERS-1 SAR. *IEEE Transactions on Geoscience and Remote Sensing* 38, 2638–2644.
- Smith, R.A., Schwarz, G.E., Alexander, R.B., 1997. Regional interpretation of water-quality monitoring data. *Water Resources Research* 33, 2781–2798.
- USGS, 1998. HYDRO1k, <http://edcwww.cr.usgs.gov/landdaac/gtopo30/hydro/>. US Geological Survey, Earth Resour. Obs. Syst. Data Center, Sioux Falls, SD.
- Verdin, K.L., Verdin, J.P., 1999. A topological system for delineation and codification of the Earth's river basins. *Journal of Hydrology* 218, 1–12.
- Vörösmarty, C.J., Fekete, B.M., Meybeck, M., Lammers, R.B., 2000a. Geomorphometric attributes of the global system of rivers at 30-minute spatial resolution. *Journal of Hydrology* 237, 17–39.
- Vörösmarty, C.J., Fekete, B.M., Meybeck, M., Lammers, R.B., 2000b. Global system of rivers: Its role in organizing continental land mass and defining land-to-ocean linkages. *Global Biogeochemical Cycles* 14, 599–621.