ELSEVIER

Journal of Hydrology 258 (2002) 214–231

Journal of Hydrology

www.elsevier.com/locate/jhydrol

Validation of a new global 30-min drainage direction map

Petra Döll*, Bernhard Lehner

Center for Environmental Systems Research, University of Kassel, Kurt Wolters Strasse 3, D-34109 Kassel, Germany

Received 26 March 2001; revised 12 September 2001; accepted 29 October 2001

Abstract

Digital drainage direction maps are a prerequisite for analyzing the flow of water on the land surface of the Earth. For continental or global studies, the most appropriate and most frequently used resolution for such data sets appears to be 30′ (longitude-by-latitude). In this paper we present the new global drainage direction map DDM30, a 30′ raster map of surface drainage directions, which organizes the Earth's land area into drainage basins and provides the river network topology. DDM30 was generated by first upscaling two drainage direction maps (DDMs) at higher resolutions. The resulting map was then extensively corrected in an iterative manner by comparison against vectorized, high resolution river maps and other geographic information. Finally, it was co-referenced to the locations of 935 gauging stations (provided by the Global Runoff Data Centre GRDC), which again involved manual corrections. DDM30 was validated against drainage basin areas from the literature, against the given upstream areas of the GRDC stations and, most importantly, against information from HYDRO1k, a data set based on a hydrologically corrected 1-km digital elevation model which is thought to afford the best information on surface drainage currently available at the global scale. In the course of the validation, the quality of DDM30 was compared to three other 30′ DDMs. The validation results show that DDM30 provides a more accurate representation of drainage directions and river network topology than the other 30′ DDMs.

Keywords: Drainage networks; Drainage basins; Rivers; Global map; Discharge

1. Introduction

In the 21st century human well-being will be affected by various global change issues, such as climate change, disturbance of biogeochemical cycles, marine pollution, food security and, in particular, freshwater scarcity and pollution. For improving our understanding of these issues, a better knowledge of the macro-scale transport of water on the land surface of the Earth is essential. The fraction of the precipitation that does not evapotranspirate locally recharges the groundwater or runs off superficially, and then is transported towards the ocean or an inland sink. This water carries dissolved and particulate materials, e.g. nutrients or sediments. Macro-scale drainage occurs mainly in the form of river flow, as groundwater flow is slow, and discontinuous at larger scales.

The lateral transport of water organizes the Earth's land areas into drainage basins. It is now well established that the appropriate spatial unit for an integrated management of water and land resources is the drainage basin. The downstream/upstream condition of an area within the drainage basin strongly influences its specific situation. On the one hand, the quantity of water available downstream is mostly larger than upstream, and the technical transport of water to a water-poor upstream area is expensive. On the other hand, downstream water users strongly

* Corresponding author. Fax: +49-561-804-3176.
E-mail address: doell@usf.uni-kassel.de (P. Döll).

0022-1694/02/$ - see front matter
PII: S0022-1694(01)00565-0

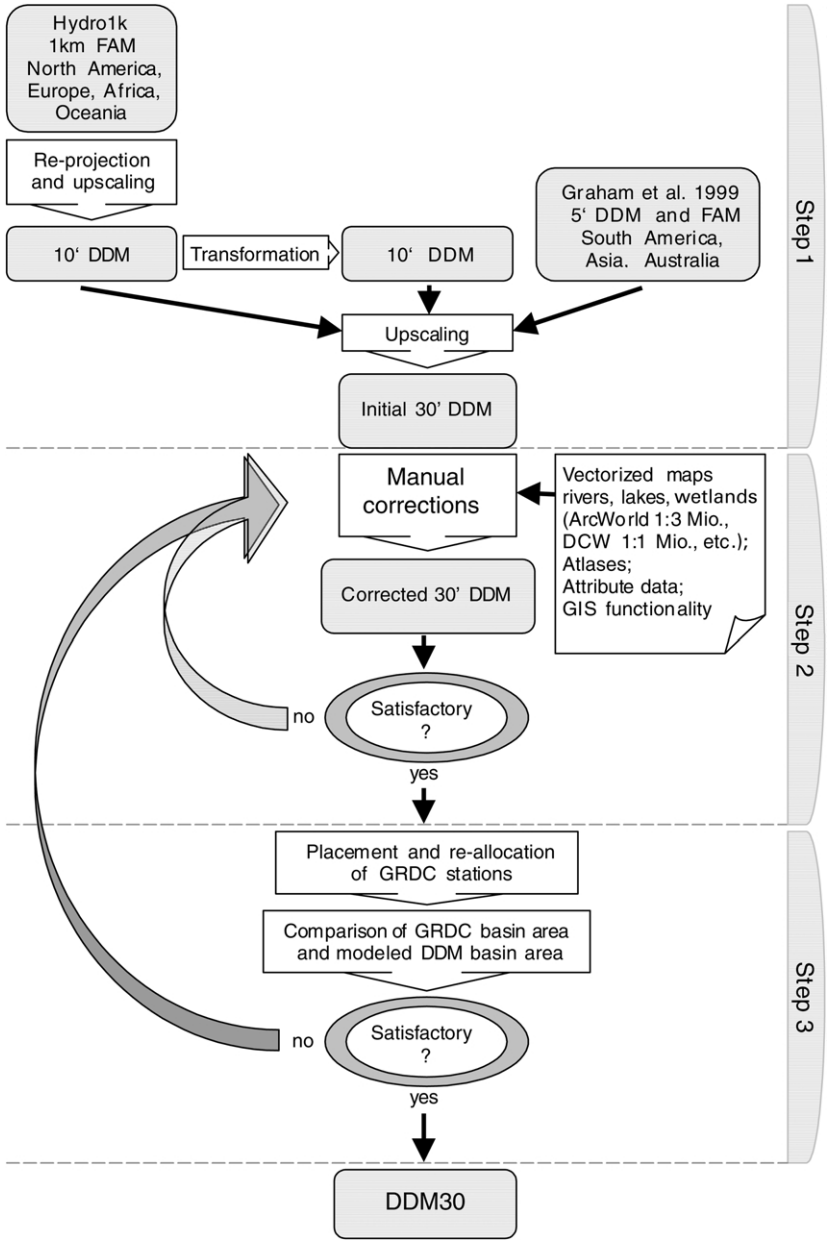

Fig. 1. Overview of map generation process (FAM: Flow accumulation map, DDM: Drainage direction map).

depend—with respect to water quantity and quality—on the activities of the upstream land and water users. The drainage paths and directions within the drainage basin are therefore important for analyses of the interactions of man and nature. Furthermore, drainage integrates over complex distributed land surface processes. The river discharge measured at a station reflects all the processes in the upstream areas, and therefore it offers the opportunity to check models that partition precipitation into evapotranspiration and runoff, i.e. hydrological models and land surface parameterizations of climate models.

Table 1
Data used to construct the global drainage direction map DDM30 (DEM: digital elevation model, FAM: digital raster map of flow accumulations)

Data	Reference	Comments
HYDRO1k FAM, 1-km resolution, for Europe, North America, Africa and Oceania	USGS (2000)	The FAM defines the number of upstream 1-km cells draining into each 1-km cell; it is based on a version of the 30″ global DEM GTOPO30, in which elevation anomalies were removed that interfere with hydrologically correct flow. It was only used for those continents for which HYDRO1k was available in December 1999
DDM and FAM, 5′ resolution, for Asia, South America, Australia	Graham et al. (1999)	The 5′ DDM and FAM were derived from the TerrainBase 5′ Global DEM, taking into account the location of major rivers as given in the CIA World Data Bank II. They were generated using automated procedures, followed by only basic manual corrections; hence they often do not coincide with information from high resolution vectorized or analog river maps
Digital vector maps of rivers and open water bodies: ArcWorld 1:3 million, Digital Chart of the World 1:1 million	ESRI (1992, 1993)	
Digital vector map of major wetlands, classified into 23 types	WCMC (1999)	
Locations and attribute data of approximately 1700 large lakes and reservoirs, based on satellite data	Birkett and Mason (1995)	
Locations of approximately 700 large reservoirs	Vörösmarty et al. (1997)	Linked to corresponding attribute data of ICOLD (1998)
Analog atlases		Times Atlas, Britannica World Atlas, Knaurs Weltatlas, Dierke Weltatlas
Locations and basin areas of 935 gauging stations	GRDC (1999)	

Consequently, a global drainage direction map has many useful applications. In particular, it is required to implement cell-to-cell river routing algorithms in large-scale hydrological models, terrestrial biogeochemical models or climate models. Currently, a 30′ (longitude-by-latitude) resolution can be regarded as appropriate for a global drainage direction map. This resolution has become widely applied for global and continental studies of water resources (e.g. Arnell, 1999; Döll et al., 1999). As it is the highest resolution at which a global gridded data set of long time series of monthly climate variables exists, it is therefore inefficient to model lateral water transport at a higher resolution. (Only if daily data were available, a routing at a higher resolution than that of the climate data might be appropriate, e.g. for flood computations.) At a resolution of less than 30′, however, it is very difficult to capture the drainage directions even for the large river basins of the Earth (compare the global 1° drainage direction map of Oki and Sud, 1998).

In the course of developing a global model of water availability and water use, we generated our own 30′ drainage direction map as the two maps available to us at the time (FDir by Graham et al., 1999, and RRN by Renssen and Knoop, 2000) were not of acceptable quality. Meanwhile, this new map has been applied for deriving scenarios of water stress in river basins under conditions of global change (Alcamo et al., 2000).

In this paper, we present our global drainage direction map DDM30 and evaluate its quality. We describe the map generation method and show the resulting map. We then validate DDM30, in particular against basin area sizes from HYDRO1k, a data set based on a hydrologically corrected 1-km digital elevation model (USGS, 2000). The validation includes a comparison of the quality of DDM30 and three other 30′ drainage directions maps, FDir, RRN and the more recently available STN-30p by Vörösmarty et al. (2000a,b).

2. Map description

The global drainage direction map DDM30 is a raster map representing the drainage directions of surface water at a spatial resolution of 30′ longitude by 30′ latitude. The map comprises 66,896 single grid cells covering the land surface area of the globe (with the exception of Antarctica). The individual cells are connected to each other by their respective drainage directions and are thus organized into drainage basins. Each cell either drains into one of the eight neighboring cells, or into none if the cell represents an inland sink (endorheism) or a basin outlet to the ocean (exorheism). Any flow bifurcations, like those occurring naturally in river deltas or artificially into canals, are not represented in DDM30 because at present they cannot be simulated by macro-scale hydrological models due to insufficient information on the partitioning of stream flow.

A number of derived maps were generated from DDM30 by applying standard geographical information systems (GIS) functionality. The flow accumulation map (FAM) represents the areas of all upstream cells, i.e. the area of those cells that drain through a given cell. Furthermore, vectorized maps of basin areas and stream networks have been generated based on DDM30.

3. Map generation

Fig. 1 provides an overview of the map generation procedure, while Table 1 lists the data used for the construction of DDM30. Before we describe the map generation procedure in Sections 3.2–3.4, we first discuss the limitations of deriving drainage directions from digital elevation models.

3.1. Limitations of deriving drainage directions from digital elevation models

In general, a drainage direction map (DDM) can be derived from a digital elevation map (DEM) by applying standardized and automated procedures. Many software packages, in particular GIS, provide tools to derive the drainage direction for each raster cell of a DEM by comparing the elevation of the cell to the elevation of its neighboring cells (steepest descent method). However, there are major limitations to this otherwise straightforward approach that render it unfeasible for deriving a global 30′ DDM:

1. DEMs are not free of errors. In addition to isolated local effects, this can have significant cumulative impacts when applying it in order to derive river basin topology: a single cell, wrongly directed in the downstream area of a river network, assigns the complete upstream area to the wrong basin as well.
2. Topographic gradients at a resolution of 30′ are not representative for river courses in nature. Applying the steepest descent method implies that the DDM has the same resolution as the DEM. Gradients that are computed from elevations averaged over approximately 2000–3000 km^2 do not necessarily reflect actual river courses which are often governed by gradients at much smaller scales. According to Vörösmarty et al. (2000a), ‘numerous inconsistencies with respect to the location of stream lines in nature’ resulted when a 30′ DDM was derived from a DEM that had been aggregated to a 30′ resolution.
3. Upscaling of DDMs from higher resolutions is a complex task. Instead of first averaging a DEM with a higher resolution to 30′ and then deriving the DDM by the steepest descent method, drainage directions can first be computed at the finer resolution and can then be upscaled to the 30′ resolution. This method focuses on preserving the more realistic topology derived from the higher-resolution DEMs. However, upscaling of vectors (drainage directions) is more complex than upscaling of scalars (elevation), which makes its automation difficult. If two or more rivers (e.g. derived at a 5′ resolution) leave a 30′ cell, it is necessary to determine the dominant flow direction of the 30′ cell. This can be done, for instance, by assigning the direction according to the 5′ river with the largest number of upstream cells. However, a rule that leads to the best local decision often results in a wrong representation of the overall structure of drainage in the basin. One example is the case of two parallel 5′ rivers that are located within one 30′ cell and which, by simple upscaling, are merged into one river. The better overall solution might be to re-direct one of the rivers around the merging cell in order to preserve the correct topology.
4. Deriving drainage directions strictly from

Table 2
Global 30′ drainage direction maps: data sources and methodology (DEM: digital elevation model, FAM: flow accumulation map)

Short name Authors	Basic data	Methodology	Manual corrections	Co-referencing of discharge stations
Fdir Graham et al. (1999)[a]	5′ DEM[b]	Maximum topographic gradient from DEM averaged to 30′, stream burning using vectorized rivers[c]	Not extensive	No
RRN Renssen and Knoop (2000)	5′ DEM[b]	Maximum topographic gradient from DEM averaged to 30′, stream burning using vectorized rivers[d]	Not extensive	No
STN-30p Vörösmarty et al. (2000a,b) and Fekete et al. (2000)	5–10′ DEM[e]	Maximum topographic gradient from DEM aggregated to 30′	Extensive, using vectorized rivers[f], regional maps and atlases, location of discharge monitoring stations	Yes
DDM30 (this paper)	1-km FAM or 5′ DDM/FAM[g]	Upscaling of DDM/FAM	Extensive, using vectorized rivers[f], atlases, location and basin areas of discharge monitoring stations, location and attribute data of lakes, wetlands and reservoirs	Yes

[a] Also generated a 5′ DDM using the same general methodology.
[b] TerrainBase DEM, NGDC (1997).
[c] CIA World Data Bank II (Gorney and Carter, 1987).
[d] Lowest resolution version wtr5 of ArcWorld 1:3 million, ESRI (1992).
[e] ETOPO5 DEM, NGDC (1988).
[f] ArcWorld 1:3 million, ESRI (1992), Digital Chart of the World 1:1 million, ESRI (1993).
[g] HYDRO1k of USGS (2000), 5′ FAM of Graham et al. (1999), depending on continent.

differences in land surface elevation is not necessarily representing natural 'flow' conditions. This is often the case in arid areas, where rivers disappear, i.e. reach a 'sink' before they reach the lowest terrain point, in very flat areas like swamps, inland deltas, and braided river systems, and in areas where the drainage direction depends on temporal variations of the water level. In these situations, it might be preferable to set the drainage direction according to 'average river run' as it is shown on maps (which represent knowledge about the existence of surface drainage) and not to rely solely on a DEM.

The automatic generation of a DDM from a DEM can be improved by so-called 'stream burning' in which information on the location of major rivers is taken into account by setting the elevation of cells that contain a major river (derived from vectorized lines of river maps) to an artificially low elevation (Graham et al., 1999; Renssen and Knoop, 2000). With stream burning, at least the course of the large rivers of the world can be correctly represented in an automated manner. The representation of the internal topology of drainage basins, however, cannot be improved as it becomes infeasible when more than one vectorized river is located within the cell. This is related to the resolutions of both the DEM and the river map.

From the above considerations we conclude that for a good representation of the 30′ drainage topology it is necessary to manually correct any automatically derived DDM. Nevertheless, we think it important to generate the best possible automatically derived DDM first, which then facilitates manual corrections and leads to a better final DDM.

3.2. Automatic derivation of an initial 30′ DDM from DDMs and FAMs of higher resolutions (Step 1)

Considering that topographic gradients at a resolution

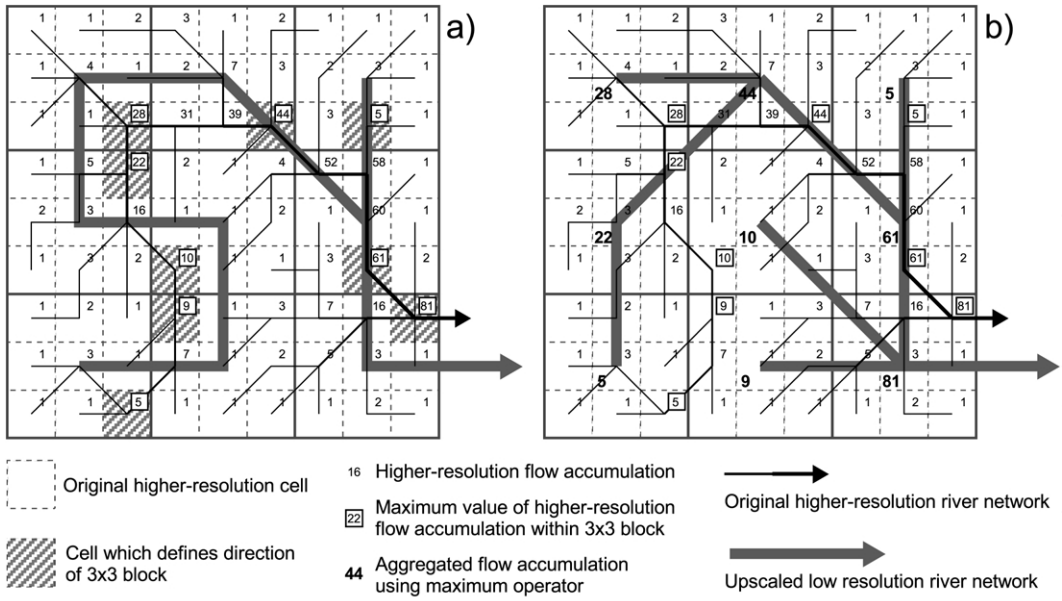

Fig. 2. Network upscaling algorithms for the example of a 3×3 block. (a) as proposed in this paper: the flow direction of the 3×3 block is defined by the direction of the higher-resolution cell with the highest flow accumulation within the block. (b) as proposed by Fekete et al. (2001, their fig. 1): the highest flow accumulation found within the 3×3 block is assigned to the block; its flow direction points to the neighboring block with the highest aggregated flow accumulation.

of $30'$ are not representative of river courses in nature, we derived our initial $30'$ DDM not by averaging a higher-resolution DEM to $30'$ (as it was done before by others, compare Table 2) but by upscaling higher-resolution DDMs (Step 1 in Fig. 1). These DDMs have a higher local accuracy of flow directions and we aimed to preserve this accuracy by applying an appropriate upscaling procedure that required both the higher-resolution DDMs and the corresponding FAMs (flow accumulation maps). We used the global DDM/FAM included in the 1-km HYDRO1k data set (USGS, 2000; Table 1) because the data set is based on the digital elevation model with the globally highest resolution (GTOPO30, resolution $30'' \times 30''$), and, in addition, has been hydrologically corrected. As HYDRO1k was only available for North America, Europe, Africa and Oceania when we constructed DDM30, we used the $5'$ DDM (and accompanying FAM) of Graham et al. (1999) for all other regions (compare Table 1).

In order to upscale the 1-km and $5'$ DDMs to a $30'$ DDM, a new network upscaling algorithm was developed using GIS. As the HYDRO1k data set is provided in an equal area projection, the DDM, i.e. the flow direction values of the 1-km grid cells, cannot easily be re-projected into geographic coordinates (degrees latitude/longitude). Different from re-projection of scalars such as elevation, which is usually done by averaging or nearest-neighbor methods, a raster-based re-projection of flow directions, i.e. topological neighborhood relations, is very complex. To avoid this complexity, only the FAM of the HYDRO1k data set was re-projected using the nearest neighbor method. The re-projection was processed at a higher resolution than given in the original grid (oversampling) in order to assure that no original grid values are lost.

The resulting grid was then aggregated into a $10'$ flow accumulation grid using standardized GIS functions (maximum filter, i.e. every $10'$ cell was assigned the maximum flow accumulation value found within the $10'$ cell). For reasons why the resolution was set to $10'$, see below. In order to re-calculate a $10'$ DDM out of the $10'$ FAM, for every cell the eight neighboring cells were inspected, and the direction was set towards the neighboring cell with the highest flow accumulation

value. If the highest value was found in the cell itself, the cell was assumed to represent a sink (either an inland sink or an outlet sink to the ocean) and was assigned no direction.

In order to upscale the 10′ DDM (based on HYDRO1k) and the 5′ DDM (provided by Graham et al., 1999) to a 30′ DDM, the following steps were taken (compare Fig. 2a): Every cell from the higher-resolution maps which falls within a cell of the 30′ DDM is inspected. If the (5′ or 10′) cell with the highest flow accumulation is on the edge of the 30′ cell it is determined to which neighboring 30′ cell its flow direction points, and the flow direction of the 30′ cell is then assigned correspondingly. If the highest flow accumulation is not found at the edge but further inside the 30′ cell, the cell is assigned to be a sink.

The HYDRO1k FAM was upscaled to a 10′ DDM rather than to a 5′ DDM in order to achieve a more even distribution of the likelihood for all eight flow directions. The fact that only 5′ or 10′ cells located at one of the corners of a 30′ cell can possibly point towards a diagonal 30′ cell (NW, NE, SW, SE) leads to the symptom that the 30′ DDM derived from the 5′ DDM shows a very low number of diagonal directions. Upscaling of a 10′ DDM, with a higher likelihood of a cell to be a corner cell, leads to more reasonable results.

Recently, Fekete et al. (2001) presented a new algorithm for upscaling higher-resolution drainage direction maps like HYDRO1k to any lower resolution. Fig. 2 compares their network upscaling algorithm to the one presented here. Both algorithms use a maximum value operator to identify, within the low-resolution cell (e.g. a 3×3 block), the higher-resolution cell with the highest flow accumulation. In the algorithm of Fekete et al. (2001) (Fig. 2b), this value is then assigned to the low-resolution cell; the low-resolution drainage direction is determined similar to the manner flow directions are derived from DEMs, but based on maximum flow accumulation gradients instead of elevation gradients ('uphill' search). While their algorithm is based on flow accumulations only, our algorithm also incorporates higher-resolution flow directions. It has the tendency to route the main river through all low-resolution cells where at least one higher-resolution cell that represents the main river is located (Fig. 2a). This leads to longer mainstem lengths than in the case of Fekete et al. (2001) algorithm, which tends to find the shortest routes between neighboring low-resolution cells.

Fekete et al. (2001) extended their network upscaling algorithm by forcing the river network to remain within basin boundaries derived from higher-resolution drainage direction maps, thus improving the quality of the resulting maps. It is beyond the scope of this paper to quantitatively assess the capability of the two network upscaling algorithms (which are both applicable at arbitrary resolutions) to represent actual drainage directions. However, drainage direction maps derived with any of the two algorithms require additional manual corrections.

3.3. Manual corrections (Step 2)

To facilitate manual corrections of the initial global 30′ DDM, the following maps were derived from the DDM by applying standard GIS functionality: (a) flow accumulation (both accumulated cells and accumulated area), (b) topologic stream network and (c) drainage basins boundaries. In addition, new GIS functionalities were developed to support the manual correction process. In particular, after each manual change of a flow direction, the structural logic of the stream network was automatically tested, i.e. crossing rivers or circles in the flow paths were indicated.

Manual corrections were mainly performed based on visual checks. Various overlays were produced to compare the modeled stream network and basin boundaries to digital maps of vectorized rivers, wetlands and lakes as well as analog atlases (Table 1, and Step 2 in Fig. 1). The digital river maps were available at different scales, which allowed identification of major and minor objects. We checked the initial 30′ DDM against the highest-resolution global vector maps (Digital Chart of the World 1:1 million) on a cell-by-cell basis, applying the following set of rules:

- The locations of major rivers, lakes and wetlands (shown on the ArcWorld 1:3 million map, Table 1) have the highest priority, i.e. their pathways are represented in best conformity with the vector maps. If the conformity is restricted by the resolution of the 30′ DDM (e.g. two major rivers flow through the same cell), then overall structural accuracy is more important than local spatial accuracy:

it is considered to be more important to represent a major river, even if it has to be shifted by one cell, than to loose it due to resolution limits.
- Minor rivers, lakes and wetlands are principally treated like the major ones, but with a lower priority.
- If the decision for a local flow direction is ambiguous, priority is given to preserve the correct basin and subbasin areas. For example, if a few cells at the margin of a subbasin are assigned to the subbasin although each of them also belongs partly to another subbasin, the total upstream area of the subbasin is likely to be overestimated. Concurrently, the neighboring basins will show a tendency to be underestimated. In this case, the flow directions of the ambiguous cells are set such that all subbasins and basins concerned are optimally represented.
- Major confluences are represented in a structurally correct way. It is checked whether each single confluence is located in the right cell (which is not necessarily the cell covering its actual geographic location). If there are multiple confluences within the same cell but more than approximately half a cell length apart, one of the confluences is shifted to a neighboring cell if possible. This is done to allow a more flexible separation of subbasins.
- Major lakes and reservoirs (approximately 1700 lakes and 700 reservoirs, Table 1) are drained correctly, i.e. by the right river. Pertinent attribute data (e.g. information on the drainage area of the lake or reservoir or on whether a lake has an outflow or represents an inland sink) is taken into consideration.

Special rules and considerations apply in the following circumstances:

- In arid regions, maps often do not indicate rivers, i.e. evidence of surface drainage, and any drainage direction is only potential. For some applications it might be of interest where the water would flow if there were enough rain (e.g. climate scenarios). Then again, rivers disappear due to transmission losses to the groundwater, which is very difficult to simulate by any hydrological model. As a compromise, all areas with no further information about the actual or potential existence of river paths (e.g. most parts of the Sahara) are assigned drainage directions according to elevation information only (from DEM or atlases, Table 1). Whenever a clear indication for the termination of a surface drainage path (e.g. a symbol for a depression or a salt lake in one of the atlases) is found, a sink is introduced, even if the DEM does not show a minimum.
- In glacierized regions and some arid basins with a very low relief (e.g. Greenland, Takla Makan Desert in China, Pampa and Gran Chaco in Argentina), only a general tendency of flow direction can be given. In order to clearly show this uncertainty, parallel flow directions are assigned.
- In case of some major wetlands (e.g. inland deltas of the Niger and the Okavango, Sudd swamps), the main river is routed in a zigzag way to simulate its actual spreading over a large area.
- Artificial channels and diversions as well as large river deltas with an extent of more than one $30'$ cell (e.g. Amazon, Nile, Ganges) cannot be represented as each cell can only drain into one neighboring cell (no flow bifurcations).
- Drainage directions within large lakes, which cover one or more cells completely, are set such that the inflow and outflow locations are represented correctly. Inside the lake, the shortest path is assumed.

Manual corrections were performed by looking at 5° by 5° windows at a time, taking into account the whole river basin in the case of larger basins. After each block of corrections, the new stream network and basin boundaries of the $30'$ DDM were calculated and visualized. If the corrected version was still not satisfactory, manual corrections inside the window were continued through another loop in Step 2 (Fig. 1).

3.4. Co-referencing of discharge gauging stations (Step 3)

The manually corrected $30'$ DDM was checked against information on drainage basin areas of gauging stations provided by the Global Runoff Data Center (GRDC, 1999). Thousand twenty-four stations were selected based on the size of the basin area (more than 9000 km^2) and the length of the

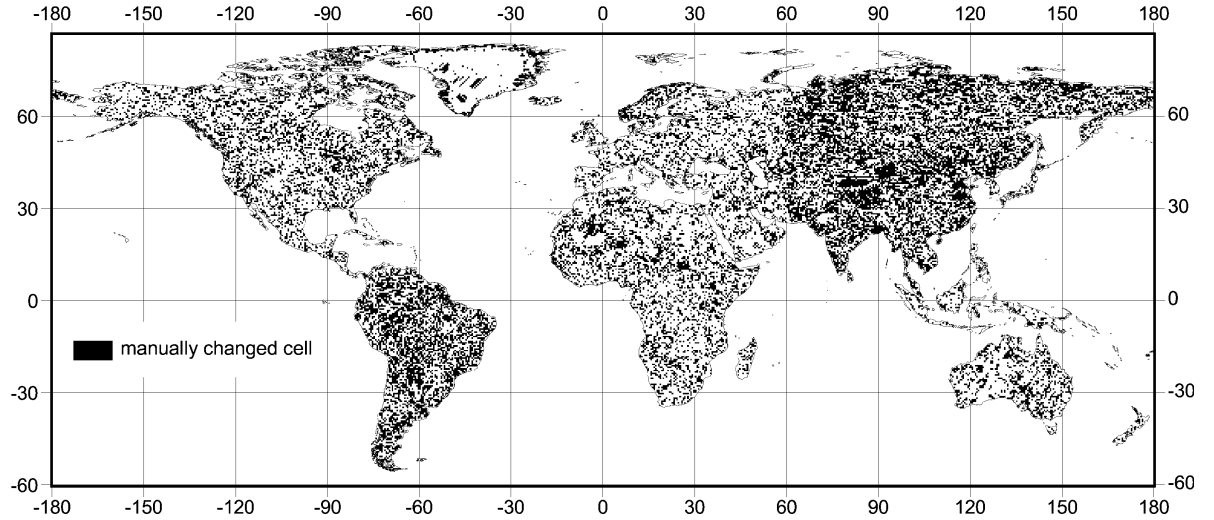

Fig. 3. Manual correction: grid cells for which the drainage direction was manually corrected.

respective discharge series (more than 3 years of monthly discharge values). Nine hundred and thirty five of these stations were found to have plausible information on both location and drainage area, while the others were obviously incorrect. For the comparison of basin areas, it was necessary to co-reference the stations to the 30′ DDM, i.e. to assign each station to a 30′ grid cell.

Co-referencing was achieved by the following procedure. First, the stations were automatically placed onto the respective grid cells, using their given coordinates. Then, GRDC and simulated basin areas were compared, and the geographic location of each station was checked. Due to the resolution of the 30′ DDM, some stations were not placed onto the correct river stretch, but, for example, onto a cell that belongs to a tributary or a neighboring river basin, resulting in a very different drainage basin area. Considering that an automated re-allocation of those stations would not necessarily lead to a good representation of the actual situation, the stations were manually re-allocated.

After re-allocation, GRDC and simulated basin areas (from the corrected DDM, Step 2 in Fig. 1) were again compared. Where the areas differed by more than 5%, the drainage directions within the whole basin were checked again (Step 2), following the guidelines of Section 3.3. The intention was not to 'force' the 30′ DDM to result in basin areas equal to the (uncertain) GRDC basin areas. Manual corrections were only made if they were consistent with the guidelines. The repeated manual check helped to improve ambiguous areas and to detect errors in the first round of manual corrections. After the second manual correction process, the GRDC stations were re-allocated again, starting at their original locations, and compared to the basins of the latest 30′ DDM. If the results were still not satisfactory, the relevant basins went through another correction loop. Typically, Steps 2 and 3 were repeated twice. When all continents were processed in the earlier described fashion, the final version of the 30′ DDM, called DDM30, was completed.

In total, the drainage directions of 35% of all 66,896 cells of DDM30 were manually corrected. Fig. 3 shows the location of these cells. The fraction of corrected cells is particularly high in South America and Asia, two continents for which the HYDRO1k data set was not available for the automatic derivation of the initial 30′ DDM (compare Table 1). This indicates that the automatic upscaling procedure preserved the more accurate information that the HYDRO1k data set provided for the other continents.

4. Results and discussion

The global drainage direction map DDM30 is

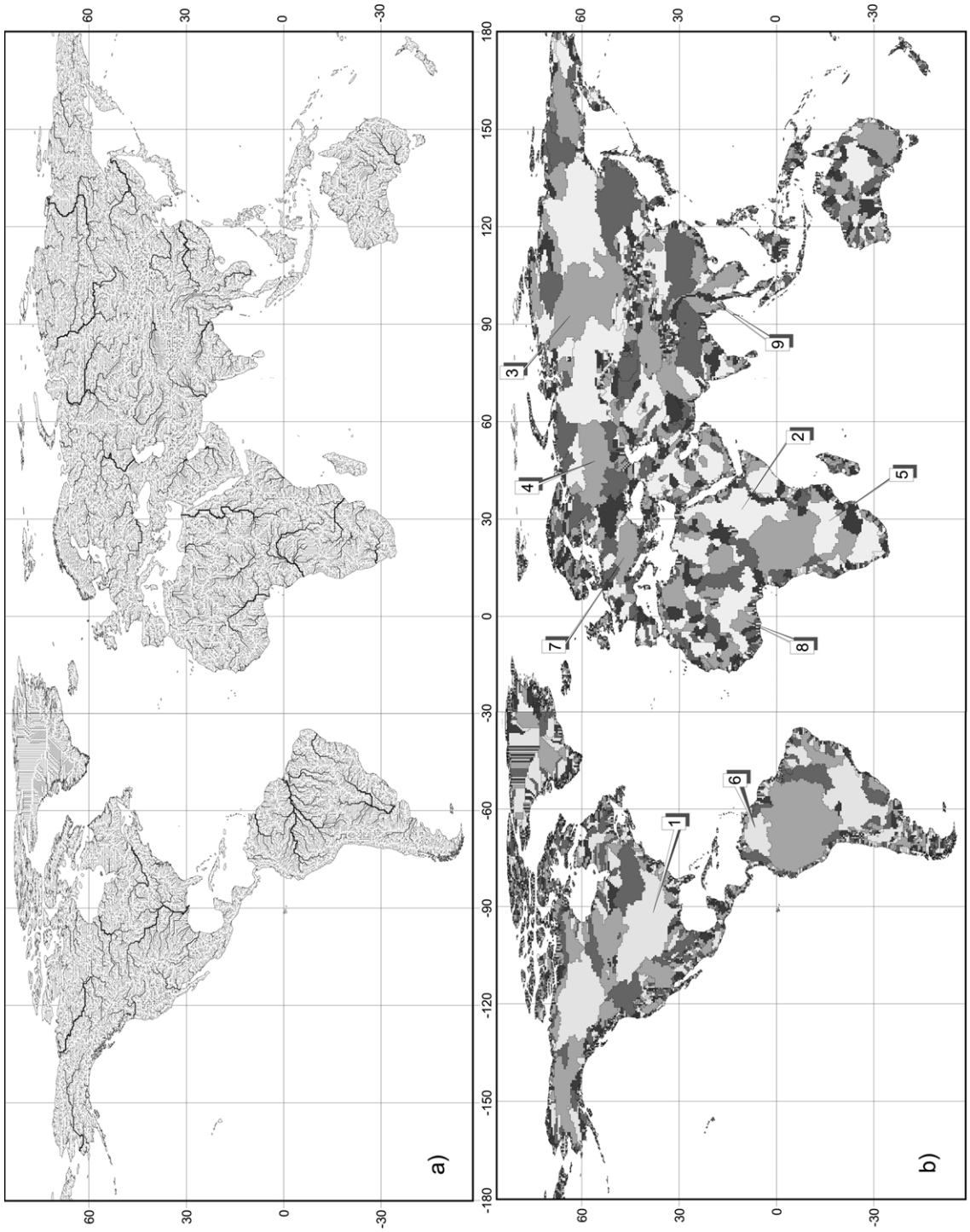

Fig. 4. The global 30′ drainage direction map DDM30: (a) stream network representation (line thickness proportional to upstream area), (b) drainage basins, either draining to the ocean or into an inland sink; numbers identify the basins listed in Table 3.

visualized in Fig. 4. In the stream network representation (Fig. 4a), the world's longest rivers show clearly, as the line thickness increases with increasing upstream areas. In DDM30, 10,832 drainage basins are represented which either drain into the ocean or into an inland sink (Fig. 4b). Six thousand eight hundred and ninety-eight are small basins consisting of only one cell. Of the 3934 drainage basins covering 2 or more cells, 663 drain to inland sinks. Nine hundred and thirty-five GRDC discharge measurement stations are co-referenced to the map such that the available observed discharges can easily be used for the validation of hydrological models. For 80% of the stations, the upstream drainage areas of DDM30 do not differ by more than 5% from the areas given by GRDC (symmetric error as defined by Eq. (1) in Section 4.2). All drainage basins larger than 15,000 km^2 differ by less than 10%. However, the accuracy of the GRDC basin areas is not known as the information sources are not documented (Th. Mauer, Head of GRDC, personal communication, 2001).

In the following sections, the quality of DDM30 will be compared to the quality of three other global 30′ DDMs: the 30′ version of FDir by Graham et al. (1999), RRN by Renssen and Knoop (2000) and STN-30p by Vörösmarty et al. (2000a,b). Table 2 provides an overview of the methods and data used to derive the four different maps.

4.1. Validation against literature data

The quality of DDM30 can only be assessed if reliable independent information on drainage directions or basin sizes exists. In the literature, information on the size of the large river basins of the world is available. Renssen and Knoop (2000), who compared the basin sizes resulting from their DDM with estimates from nine different sources, found that published data on basin areas differ considerably. The same observation was made by Vörösmarty et al. (2000a, their Table 1). They concluded that the wide range of published estimates arises (1) from the repetition of original, incorrect estimates and (2) from difficulties of consistently defining the basin boundaries, e.g. related to the position of the river mouth (vs. estuary), the inclusion of non-discharging portions of the basins in arid regions, the existence of engineered waterways or complex drainage patterns. Therefore, the mismatch between simulated basin areas from the 30′ DDMs and published values does not reliably indicate the DDM error.

The hydrologically corrected global 1-km HYDRO1k data set (USGS, 2000) provides a better benchmark against which the performance of the 30′ DDMs can be tested. After finalization of DDM30, it became available for all world regions except Australia. HYDRO1k can be regarded as the best global representation of drainage basins available at present. As an additional advantage, HYDRO1k provides information on upstream areas for each point of a basin, not only at the outlet or for discharge gauging stations.

Table 3 lists river basin areas as derived from the four global 30′ DDMs under investigation and from HYDRO1k as well as the range of basin areas as given in the literature (compiled by Renssen and Knoop, 2000). The river basins in Table 3 are selected to show that both FDir and RRN suffer from significant errors even in case of large river basins. While discrepancies in the case of the Nile may be attributed to different treatment of non-discharging areas, both FDir and RRN clearly underestimate the size of the Mississippi basin, and overestimate the size of the Volta and Salween basins, in comparison to both HYDRO1k and literature data. In addition, FDir shows strong discrepancies in the case of the Yenisei, the Volga, the Zambezi, the Orinoco and the Danube, and RRN in the case of the Zambezi and Orinoco. The simulated areas of DDM30 and STN-30p are relatively close to each other and to the HYDRO1k data, except for the Nile and the Zambezi. In the latter case, STN-30p includes the inland drainage basin of the Okavango even though it is connected to the Zambezi basin only in years of high flows.

4.2. Validation against upstream areas of GRDC gauging stations

As GRDC stations have been independently co-referenced to both STN-30p and DDM30 by their respective authors (Table 2), the listed drainage areas of the stations can be compared to the areas simulated by STN-30p and DDM30. For the subset of 625 stations co-referenced to both maps, Fig. 5 shows the correspondence between the simulated

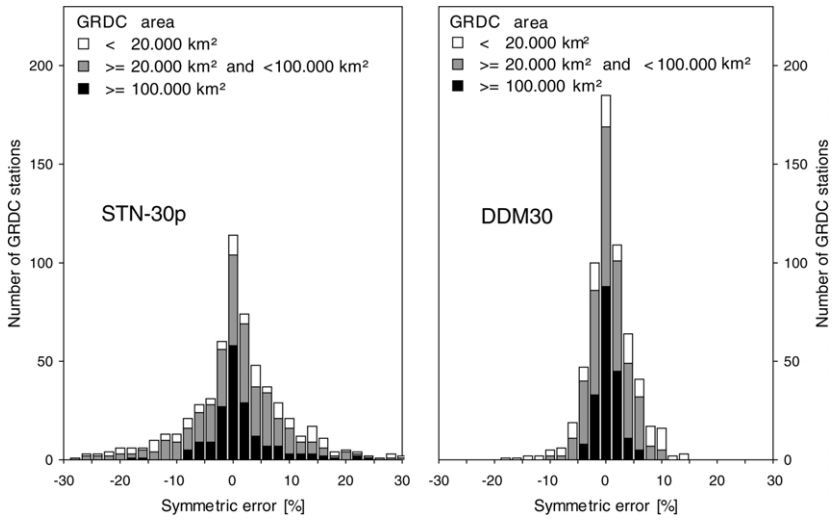

Fig. 5. Drainage basin areas of 625 GRDC discharge gauging stations: symmetric error of simulated DDM30 and STN-30p basin areas with respect to values provided by GRDC (1999). The stations are co-referenced to both STN-30p and DDM30. Please note that in case of STN-30p, 31 stations have an error larger than 30%.

and GRDC drainage basin areas. The (lack of) correspondence is expressed as the symmetric error (Fekete et al., 1999)

$$\varepsilon = 100\frac{A_{\text{sim}} - A_{\text{GRDC}}}{\max(A_{\text{sim}}, A_{\text{GRDC}})} \quad (1)$$

where A_{GRDC} is the drainage basin of station as provided by GRDC (1999). and A_{sim} is the upstream drainage basin area of $30'$ DDM cell to which the GRDC measurement station is co-referenced.

Overall, the median and mean value of the absolute symmetric errors between simulated and GRDC basin areas are 4.5 and 9.8% for STN-30p and 2.1 and 2.9% for DDM30. In particular for smaller basins (or subbasins) below 100,000 km^2, DDM30 basin areas agree better with GRDC values than the STN-30p basin areas do. However,

Table 3
Drainage basin areas of the four global $30'$ DDMs compared to the HYDRO1k areas (USGS, 2000) and the range of basin areas as given in the literature. River basins are selected to show that both FDir and RRN suffer from significant errors even in case of large river basins (Basin areas in 1000 km^2)

Sno.	River basin	Literature[a]		HYDRO1k	DDM30	STN-30p	RRN	Fdir
		Avg.	(Min.–Max.)					
1	Mississippi	3240	(3220–3270)	3197	3231	3203	2968	2946
2	Nile	2830	(1900–3030)	3078	2900	3826	2780	3575
3	Yenisei	2600	(2530–2700)	2557	2531	2582	2428	2719
4	Volga	1380	(1350–1420)	1390	1386	1463	1434	945
5	Zambezi	1320	(1200–1420)	1388	1387	1989	1469	2313
6	Orinoco	980	(880–1090)	941	965	1039	826	1186
7	Danube	810	(800–820)	780	795	788	817	713
8	Volta	390	(380–390)	414	407	398	510	492
9	Salween	310	(280–330)	258	290	273	452	981

[a] As listed in Renssen and Knoop (2000).

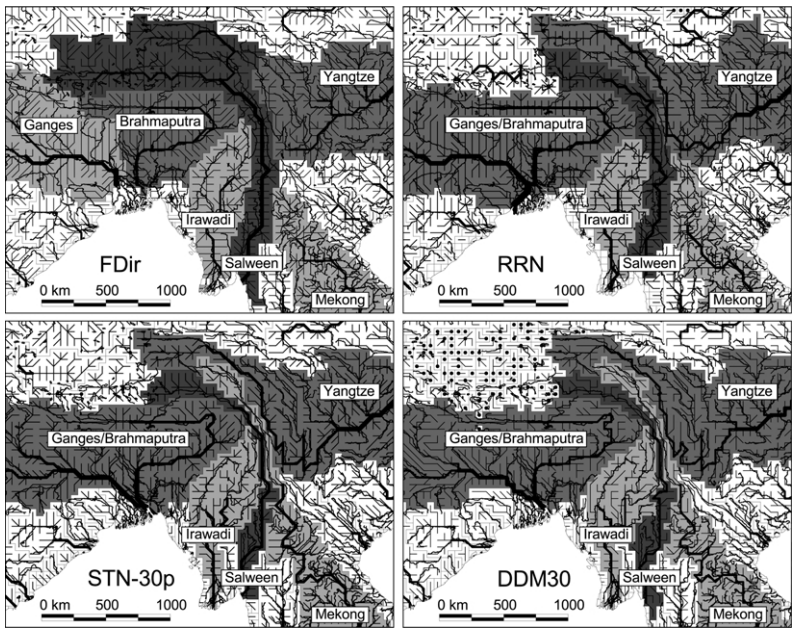

Fig. 6. Comparison of four 30′ DDMs for a region in Southern Asia. In FDir (Graham et al., 1999), the upper Yangtze and the upper Mekong are incorrectly routed into the Salween, while in RRN (Renssen and Knoop, 2000), the upstream basin of the Mekong is missing. Both STN-30p (Vörösmarty et al., 2000a) and DDM30 (this paper) represent the major river basins well, but individual drainage directions can differ considerably.

please note that, different from DDM30, STN-30p was not tuned to all of the compared 625 stations but to stations provided by other sources (Vörösmarty et al., 2000a).

4.3. Validation by visual inspection at a regional scale

Fig. 6 compares the stream networks of the four 30′ DDMs for a region in Southern Asia and illustrates

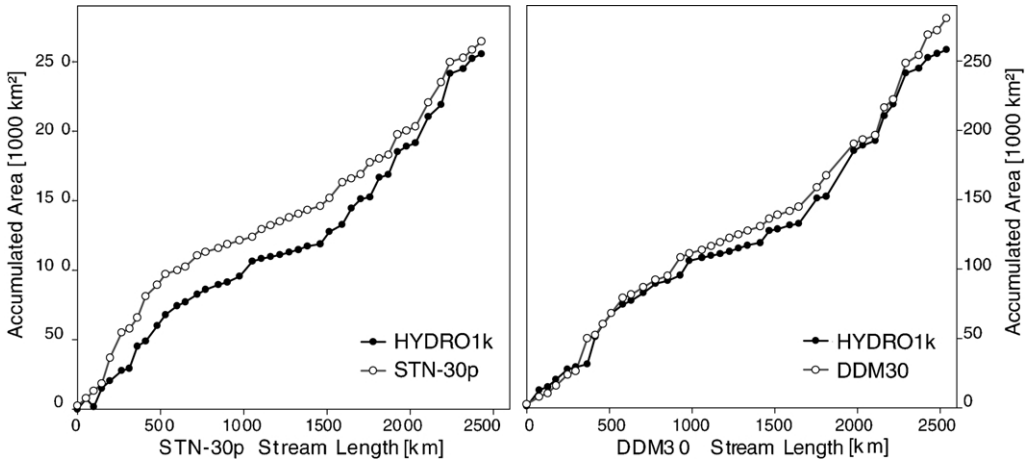

Fig. 7. Upstream areas of 30′ DDM cells along the river Salween, as represented by STN-30p (left) and DDM30 (right), compared to the respective upstream areas from HYDRO1k. The highest value of all HYDRO1k upstream areas that occurs within each 30′ cell has been chosen for comparison.

that FDir and RRN do not represent even large river basins in a satisfactory manner. In FDir, the upper Yangtze, the upper Mekong and part of the Tibetan Plateau are incorrectly routed into the Salween, while in RRN, the upstream basin of the Mekong is missing. Both STN-30p and DDM30 represent the major river basins well, but individual drainage directions can differ considerably (compare, for example, the eastern part of the Ganges/Brahmaputra basin). A characteristic difference between STN-30p and DDM30 can be observed in Tibet, where DDM30 shows more sinks than STN-30p. The reason for this is that STN-30p represents potential (and not contemporarily active) river networks (Vörösmarty et al., 2000a), which may only become effective if due to climate change more discharge occurs.

4.4. Validation against HYDRO1k for a selected river basin

If the total area of a drainage basin is correct, this does not necessarily mean that the stream network inside the basin is properly represented. A good representation of the internal stream network, however, is essential for any modeling in which river routing is performed and/or water availability and water demand is compared. As an example, the internal stream network of the Salween basin as represented by STN-30p and DDM30 (shown in Fig. 6) is checked by comparing the upstream basin areas of all $30'$ cells along the mainstem with the basin areas derived from HYDRO1k. As no information from HYDRO1k was used to derive the Asian part of DDM30 (it was not available at the time), the comparison of the performance of STN-30p and DDM30 should not be biased.

Due to the scale discrepancy, such a comparison is not straightforward. In order to determine the corresponding HYDRO1k area, the highest upstream area value of any 1-km cell located within the respective $30'$ cell of either STN-30p or DDM30 was chosen. Fig. 7 shows that STN-30p overestimates the drainage area along the first 500 km of the Salween, while the basin representation of DDM30 agrees well with HYDRO1k. Looking at the Salween basin in Fig. 6, one can observe that the uppermost part of the basin is wider according to STN-30p than according to DDM30. Close to the mouth of the Salween, however, DDM30 differs considerably from HYDRO1k. This is due to the inclusion of some small tributaries which in reality drain directly into the estuary and not into the river Salween. With respect to macro-scale modeling of discharge and water availability, such a discrepancy does not have significant consequences. It is more important to represent the upstream part of a basin in an optimal manner, as runoff that is generated there contributes to the discharge at each downstream measurement location.

4.5. Validation against HYDRO1k at the global scale

The most comprehensive assessment of the quality of the four $30'$ DDMs is performed by comparing the upstream drainage areas of all their $30'$ cells to those from HYDRO1k. Here, like in the comparison for the internal drainage pattern of the Salween (Fig. 7), the highest upstream area value of any 1-km cell located within the respective $30'$ cell is chosen. With this approach, extreme discrepancies can occur if, for example, a mainstem cell of a large meandering river in the 1-km HYDRO1k resolution is located just outside the mainstem cell in the $30'$ DDM. To exclude these cases which would lead to meaningless extreme discrepancies between the drainage areas, only those cells are taken into account, for which the upstream drainage areas of the $30'$ DDM and of HYDRO1k differ by less than a factor of 3. Besides, the upstream basin area of the $30'$ cells had to be larger than 10,000 km^2 because scale effects become dominant for smaller spatial units. The correspondence between the drainage basin areas of the individual $30'$ cells and the HYDRO1k basin areas is shown in Fig. 8, separately for FDir, RRN, STN-30p and DDM30. DDM30 exhibits the best correspondence to HYDRO1k, followed by STN-30p. This is particularly obvious for cells with a small upstream area between 10,000 and 100,000 km^2 (compare log-scaled zooms in Fig. 8), where DDM30 shows the strongest clustering around the 1:1 line, while FDir and RRN show no clustering at all. The degree of correspondence is quantified by the modeling efficiency ME (Janssen and Heuberger, 1995), which is equivalent to the Nash–Sutcliffe coefficient and measures the

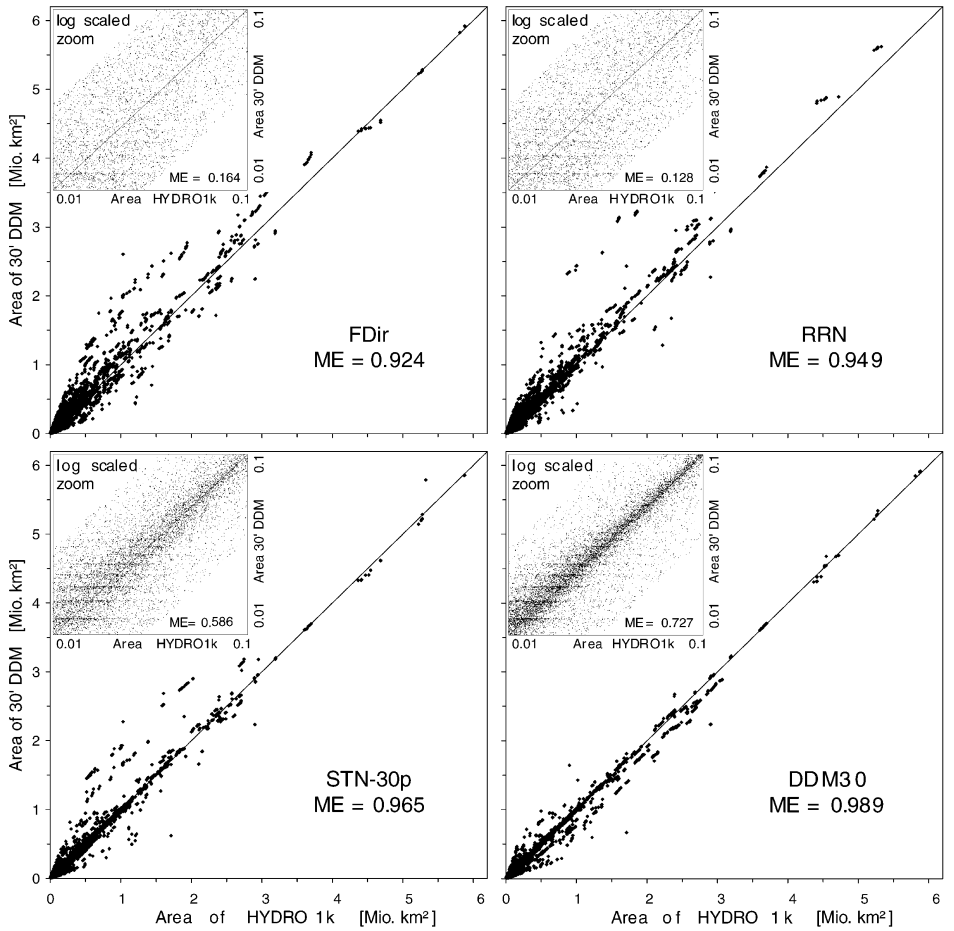

Fig. 8. Drainage basin areas of individual 30′ cells as represented by four 30′ DDMs plotted against HYDRO1k drainage basin areas. The highest value of all HYDRO1k upstream areas that occurs within each 30′ cell has been chosen for comparison. Only those cells are taken into account, for which (1) the upstream drainage areas of the 30′ DDM and of HYDRO1k differ by less than a factor of 3, and (2) the 30′ upstream area is larger than 10,000 km^2.

goodness-of-fit to the line-of-perfect-fit (the 1:1 line):

$$ME = \frac{\sum_{i=1}^{N}\left(A_{\mathrm{HYDRO1k},i} - \overline{A_{\mathrm{HYDRO1k},i}}\right)^2 - \sum_{i=1}^{N}\left(A_{\mathrm{sim},i} - A_{\mathrm{HYDRO1k},i}\right)^2}{\sum_{i=1}^{N}\left(A_{\mathrm{HYDRO1k},i} - \overline{A_{\mathrm{HYDRO1k},i}}\right)^2} \quad (2)$$

where $A_{\mathrm{HYDRO1k},\,i}$ is the drainage basin area of cell i according to HYDRO1k, $\overline{A_{HYDRO1k,i}}$ the average of all drainage basin areas according to HYDRO1k, and $A_{\mathrm{sim},\,i}$ is the drainage basin area of cell i according to the 30′ DDM.

The modeling efficiency ME of DDM30 is 0.989, for all selected cells, and 0.727 for all cells with a drainage area between 10,000 and 100,000 km^2. Both values show that DDM30 is a better representation of the HYDRO1k drainage topology than STN-30p (with MEs of 0.965 and 0.586, respectively), and a much better representation than RRN and FDir. For the latter DDMs, the MEs of the small basins indicate that there is barely any correspondence with HYDRO1k. The differences in ME become even more significant when considering that in case of RRN and FDir, about 50% less cells than in case of DDM30 could be selected for the comparison with HYDRO1k (‘number of cells’ in Table 4), as the

Table 4
Upstream drainage areas of individual cells as represented by the 30′ DDMs compared to the respective areas of HYDRO1k, differentiated according to the continents for which HYDRO1k is currently available. The highest value of all HYDRO1k upstream areas that occurs within each 30′ cell has been chosen for comparison. For computation of the modeling efficiency ME, only those cells are taken into account ('number of cells'), for which (1) the upstream drainage areas of the 30′ DDM and of HYDRO1k differ by less than a factor of 3, and (2) the 30′ upstream area is larger than 10,000 km^2

	Number of cells	Modeling efficiency ME						
		World	N. America	S. America	Africa	Asia	Europe	Oceania
All cells with a drainage area larger than 10,000 km^2								
DDM30	14840	0.989	0.996	0.994	0.983	0.986	0.980	0.998
STN-30p	12234	0.965	0.991	0.994	0.957	0.936	0.973	0.997
RRN	7739	0.949	0.964	0.983	0.969	0.901	0.960	0.996
Fdir	7398	0.924	0.945	0.986	0.872	0.926	0.832	0.997
Only cells with a drainage area between 10,000 and 100,000 km^2								
DDM30	11435	0.727	0.784	0.561	0.670	0.754	0.801	0.930
STN-30p	9312	0.586	0.689	0.540	0.490	0.588	0.594	0.844
RRN	5171	0.128	0.066	0.314	0.014	0.065	0.335	0.535
Fdir	4912	0.164	0.260	− 0.006	0.214	0.167	0.087	0.562

basin areas of the other cells differed by more than a factor of three from the HYDRO1k basin areas. Thus, we conclude that the overall spatial accuracy of FDir and RRN is poor. In the case of STN-30p, about 18% less cells than in the case of DDM30 could be selected for comparison, both for all selected cells and for the cells with drainage basin areas below 100,000 km^2 (Table 4).

It could be argued that DDM30 is bound to have a better correspondence with HYDRO1k than the other three 30′ DDMs as its initial automatic version was based on HYDRO1k for all world regions except Asia, South America and Australia. Therefore, a separate comparison of the goodness-of-fit for Asia and South America is indicated (Table 4; HYDRO1k is not yet available for Australia). For Asia and South America, as well as for the other world regions, DDM30 shows the highest modeling efficiencies, both for all basins and for the small basins. In South America, the overall MEs for STN-30p and DDM30 are equal, but for the drainage basin areas below 100,000 km^2, DDM30 has a higher ME than STN-30p. Hence, the comparison among the four 30′ DDMs is not invalidated by a bias of DDM30 with respect to HYDRO1k, and the fact that DDM30 shows the best correspondence to HYDRO1k can be interpreted as an indication of its good quality.

5. Conclusions

The global drainage direction map DDM30 represents the global pattern of surface water drainage at a resolution of 30′. It organizes the land area of the world into drainage basins, simulates the river networks and describes for each 30′ cell, where the water flowing into the cell comes from and towards which direction the water leaves the cell.

DDM30 provides a better representation of the actual drainage directions and river networks than the other three existing 30′ DDMs. This conclusion is based on a comparison of simulated drainage basin areas with values from the literature and from the Global Runoff Data Centre, but primarily on a comparison with information from the hydrologically corrected global 1-km data set HYDRO1k. This data set is thought to provide the best information on surface drainage currently available. The 30′ DDM RRN by Renssen and Knoop (2000) as well as the 30′ version of FDir by Graham et al. (1999) represent some of the large river basins incorrectly, while for basins smaller than 100,000 km^2, they generally perform quite poorly. The unsatisfactory quality of these two data sets is caused by the low degree of manual correcting. The generation of both STN-30p (Vörösmarty et al., 2000a,b) and DDM30 (this paper) included extensive manual corrections, which lead to a quality that makes both data sets appropriate for

macro-scale modeling of the terrestrial water cycle, including climate change studies. DDM30, however, shows a better correspondence to GRDC and HYDRO1k basin areas than STN-30p, particularly for basins smaller than 100,000 km^2.

At present, a 30′ DDM is the appropriate basis for continental and global-scale simulations. With improved data availability and computing power, the application of higher-resolution DDMs in macro-scale modeling will become feasible and efficient. Then, upscaling of the 1-km HYDRO1k data set or an even more highly resolved data set to, for instance, the 5′ resolution will be of interest. If the upscaling to any resolution below 30′ is performed by algorithms that only take into account local factors (like the upscaling algorithms we used), extensive manual corrections are likely to be necessary. As the manual corrections become more time-consuming with increasing resolution, we recommend the development of innovative automatic upscaling algorithms that result in a better representation of the overall structure of a drainage basin and thus reduce the necessary extent of manual corrections.

DDM30 is available for free from the authors. Please contact us at the corresponding address or through http://www.usf.uni-kassel.de.

Acknowledgements

We acknowledge the Global Runoff Data Centre, Koblenz, Germany, for providing information on river discharge measurement stations. We would like to thank Balazs Fekete for his inspiring work on drainage directions maps and his helpful review of our first draft.

References

Alcamo, J., Henrichs, T., Rösch, T., 2000. World Water in 2025—Global modeling and scenario analysis for the World Commission on Water for the 21st Century. Kassel World Water Series 2. Center for Environmental Systems Research, University of Kassel, Germany.

Arnell, N.W., 1999. A simple water balance model for the simulation of streamflow over a large geographic domain. J. Hydrol. 217, 314–335.

Birkett, C.M., Mason, I.M., 1995. A new global lakes database for a remote sensing programme studying climatically sensitive large lakes. J. Great Lakes Res. 21 (3), 307–318 publication available online at http://wwwcpg.mssl.ucl.ac.uk/orgs/un/glaccd/html/mgld.html.

Döll, P., Kaspar, F., Alcamo, J., 1999. Computation of global water availability and water use at the scale of large drainage basins. Mathematische Geol. 4, 111–118.

ESRI—Environmental Systems Research Institute, 1992. ArcWorld 1:3 M Continental Coverage.

ESRI—Environmental Systems Research Institute, 1993. Digital Chart of the World 1:1 M.

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

Fekete, B.M., Vörösmarty, C.J., Grabs, W., 2000. UNH/GRDC Composite Runoff Fields V1.0. (http://www.grdc.sr.unh.edu).

Fekete, B.M., Vörösmarty, C.J., Lammers, R.B., 2001. Scaling gridded networks for macroscale hydrology: development, analysis, and control of error. Water Resour. Res. 37, 1955–1968.

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

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 Resour. Res. 35 (2), 583–587 available online at http://www.geo.utexas.edu/faculty/famiglietti/grahamdata/.

GRDC—Global Runoff Data Centre, 1999. GRDC Station Catalog (available online at http://www.bafg.de/html/internat/grdc/download.html).

ICOLD—International Commission on Large Dams, 1998. Word Register of Dams (CD-ROM), Paris, France.

Janssen, P.H.M., Heuberger, P.S.C., 1995. Calibration of process-oriented models. Ecol. Model. 83, 55–66.

NGDC—National Geophysical Data Center, 1988. ETOPO5. Data Announcement 88-MGG-02, Digital relief of the Surface of the Earth. NOAA, National Geophysical Data Center, Boulder, Colorado (available online at http://www.ngdc.noaa.gov).

NGDC—National Geophysical Data Center, 1997. TerrainBase Global Terrain Model. NOAA, National Geophysical Data Center, Boulder, Colorado (available online at http://www.ngdc.noaa.gov).

Oki, T., Sud, Y.C., 1998. Design of total runoff integrating pathways (TRIP)—A global river channel network. Earth Interact., 2 available online at http://hydro.iis.u-tokyo.ac.jp/~taikan/trip-data/tripdata.html.

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

USGS—United States Geological Survey, 2000. HYDRO1k (available online at http://edcdaac.usgs.gov/gtopo30/hydro/).

Vörösmarty, C.J., Sharma, K., Fekete, B., Copeland, A.H., Holden, J., Marble, J., J, A., Lough, J.A., 1997. The storage and aging of continental runoff in large reservoir systems of the world. Ambio 26, 210–219.

Vörösmarty, C.J., Fekete, B.M., Meybeck, M., Lammers, R., 2000a.

Global system of rivers: its role in organizing continental land mass and defining land-to-ocean linkages. Global Biogeochem. Cycles 14 (2), 599–621.

Vörösmarty, C.J., Fekete, B.M., Meybeck, M., Lammers, R., 2000b. Geomorphic attributes of the global system of rivers at 30-min spatial resolution. J. Hydrol. 237, 17–39.

WCMC—World Conservation Monitoring Centre, 1999. Wetlands Dataset, Cambridge, UK.