HydroSHEDS

Technical Documentation

Version 1.2

July 2013

Conservation Science Program
World Wildlife Fund US
Washington, DC 20037

Prepared by Bernhard Lehner

1. Overview

HydroSHEDS (**Hydro**logical data and maps based on **SH**uttle **E**levation **D**erivatives at multiple **S**cales) provides hydrographic information in a consistent and comprehensive format for regional and global-scale applications. HydroSHEDS offers a suite of geo-referenced data sets in raster and vector format, including stream networks, watershed boundaries, drainage directions, and ancillary data layers such as flow accumulations, distances, and river topology information.

HydroSHEDS is derived primarily from elevation data of the Shuttle Radar Topography Mission (SRTM) at 3 arc-second resolution. The original SRTM data have been hydrologically conditioned using a sequence of automated procedures. Existing methods of data improvement and newly developed algorithms have been applied, including void-filling, filtering, stream burning, and upscaling techniques. Manual corrections were made where necessary. Preliminary quality assessments indicate that the accuracy of HydroSHEDS significantly exceeds that of existing global watershed and river maps.

The goal of developing HydroSHEDS was to generate key data layers to support regional and global watershed analyses, hydrological modeling, and freshwater conservation planning at a quality, resolution and extent that has previously been unachievable. Available resolutions range from 3 arc-second (approx. 90 meters at the equator) to 5 minute (approx. 10 km at the equator) with seamless near-global extent.

HydroSHEDS has been developed by the Conservation Science Program of World Wildlife Fund (WWF), in partnership or collaboration with the U.S. Geological Survey (USGS); the International Centre for Tropical Agriculture (CIAT); The Nature Conservancy (TNC); McGill University, Montreal, Canada; the Australian National University, Canberra, Australia; and the Center for Environmental Systems Research (CESR), University of Kassel, Germany. Major funding for this project was provided to WWF by JohnsonDiversey, Inc. and Sealed Air Corporation.

HydroSHEDS data are free for non-commercial and commercial use. For specific restrictions and use requirements see the License Agreement provided in Appendix A.

For more information on HydroSHEDS please visit

http://www.hydrosheds.org (official homepage of HydroSHEDS)

or follow the links via

http://hydrosheds.cr.usgs.gov (former data download site, now discontinued) or
http://worldwildlife.org/pages/hydrosheds

Citations and acknowledgement of the HydroSHEDS database should be made as follows:

Lehner, B., Verdin, K., Jarvis, A. (2008): New global hydrography derived from spaceborne elevation data. Eos, Transactions, 89(10): 93-94. Data is available at www.hydrosheds.org.

2. Data sources

This chapter briefly describes the main data sources that have been used in the generation of HydroSHEDS. The actual processing steps are addressed in chapter 3. Please also refer to the flowchart of Figure 1.

2.1 Elevation data from the Shuttle Radar Topography Mission (SRTM)

The primary data source of HydroSHEDS is the digital elevation model (DEM) of the Shuttle Radar Topography Mission. SRTM elevation data were obtained by a specially modified radar system that flew onboard the Space Shuttle Endeavor during an 11-day mission in February of 2000. The SRTM project is a collaborative effort by the National Aeronautics and Space Administration (NASA), the National Geospatial-Intelligence Agency of the U.S. Department of Defense (NGA), as well as the German Aerospace Center (DLR) and the Italian Space Agency (ASI). NASA’s Jet Propulsion Laboratory (JPL) managed the mission, and the Earth Resources Observation and Science Data Center of the U.S. Geological Survey (USGS EROS Data Center) has the responsibility of hosting, distributing and archiving the final SRTM data products. A general description of the SRTM mission can be found in Farr and Kobrick (2000).

2.1.1 SRTM elevation data, Version 1 (SRTM-1 and SRTM-3 unfinished data)

The raw SRTM data have been processed into an initial research quality DEM by JPL. No further editing has been performed, resulting in a data set that may contain numerous voids and other spurious points such as anomalously high (spike) or low (well) values. Since water surfaces produce very low radar backscatter, water bodies are generally not well defined and appear quite “noisy”. Coastlines, as well, are not clearly defined. For areas outside of the conterminous United States (CONUS), the original 1 arc-second data (SRTM-1; cell size approximately 30 meters at the equator) were aggregated into 3 arc-second data (SRTM-3) by averaging, i.e. each 3 arc-second data point is generated by averaging the corresponding 3x3 kernel of the 1 arc-second data. For more details see NASA/JPL (2005).

2.1.2 SRTM elevation data, Version 2 (DTED-2 and DTED-1 finished data)

After JPL completed the raw processing of the SRTM-1 and SRTM-3 data, NGA performed quality assurance checks and then carried out several additional finishing steps to comply with the required data standards of the Digital Terrain Elevation Data (DTED®) format (NASA 2003). Spikes and wells in the data were detected and voided out. Small voids were filled by interpolation of surrounding elevations. Large voids, however, were left in the data. The ocean was set to an elevation of 0 meters. Lakes of 600 meters or more in length were flattened and set to a constant height. Rivers of more than 183 meters in width were delineated and monotonically stepped down in height. Islands were depicted if they had a major axis exceeding 300 meters or the relief was greater than 15 meters. All finishing steps were performed at the original 1 arc-second resolution, resulting in DTED Level 2 data products. DTED-2 was then aggregated into 3 arc-second DTED-1 data. Unlike SRTM-3, however, DTED-1 has been generated by subsampling, i.e. each 3 arc-second data point is generated by assigning the value of the center pixel of the corresponding 3x3 kernel of the 1 arc-second data. For more details see NASA/JPL (2005).

2.1.3 SRTM tiling format and data availability

SRTM elevation data have been processed in a systematic fashion and mosaicked into approximately 15,000 one-degree by one-degree tiles. Following the DTED convention, the names of the individual data tiles refer to the latitude and longitude of the lower-left (southwest) corner of the tile. For example, the coordinates of the center of the lower-left pixel of tile n40w118 are 40 degrees north latitude and 118 degrees west longitude. In the case of DTED-1 and SRTM-3 data, a single tile consists of 1201 data rows and 1201 data columns. Due to the definition via pixel centers, the four edges of a tile each exceed the assigned coordinates by half a pixel and the outermost rows and columns of adjacent tiles are overlapping. For more details see NASA/JPL (2005).

Outside of the CONUS, the 1 arc-second products (SRTM-1 and DTED-2) are only available upon request for scientific purposes. The 3 arc-second products (SRTM-3 and DTED-1) are public domain and have thus been used as a basis for HydroSHEDS.

2.1.4 Void-filled SRTM data provided by CIAT

The original SRTM data is characterized by exhibiting a number of data voids, i.e. “no-data” pixels at locations where the original radar backscatter could not be interpreted properly. Small data voids can be interpolated rather easily, e.g. by applying nearest neighbor methods, but larger voids pose a problem for many applications. A team of researchers from the International Center for Tropical Agriculture (CIAT), Colombia (A. Jarvis and E.) and later from the Joint Research Center (JRC) of the European Commission (H.I. Reuter and A. Nelson) have further processed the original SRTM DEMs to fill in these no-data voids (Jarvis et al., 2008). This involved the production of vector contours and points, and the re-interpolation of these derived contours back into a raster DEM. The void-filled DEM is offered as a seamless near-global coverage (up to 60 degrees north and south), and is available for download as 5 degree x 5 degree tiles in geographic coordinate system – WGS84 datum. For the production of HydroSHEDS, Version 2 of the CIAT void-filled SRTM data was used (CIAT 2004). The data is available from the CGIAR-CSI SRTM 90m Database at http://srtm.csi.cgiar.org.

2.2 SRTM Water Body Data (SWBD)

SRTM Water Body Data files are a by-product of the data editing performed by NGA to produce the finished SRTM DTED-2 data. Ocean, lake and river shorelines were identified and delineated from the 1 arc-second DTED-2 data (for details see NASA 2003) and were saved as vectors in ESRI 3-D Shapefile format (ESRI 1998). There are approximately 12,000 SWBD files since only those SRTM tiles that contain water have a corresponding SWBD shapefile.

The guiding principle for the development of SWBD was that water must be depicted as it was in February 2000 at the time of the Shuttle flight. In most cases, two orthorectified SRTM image mosaics were used as the primary source for water body editing. A landcover water layer and medium-scale maps and charts were used as supplemental data sources. Since the landcover water layer was derived mostly from Landsat 5 data collected a decade earlier than the Shuttle mission and the map sources had similar currency problems, there were significant seasonal and temporal differences between the depiction of water in the SRTM data and the ancillary sources. For more details see NASA/NGA (2003) and NASA (2003).

2.3 Digital Chart of the World (DCW) global vectorized river network

The Digital Chart of the World (ESRI 1993) is a global vector map at a resolution of 1:1 million that includes a layer of hydrographic features such as rivers and lakes. DCW (also known as VMAP-0) is generally considered to provide the most comprehensive and consistent global river network data currently available. It is based on the US DMA (now NGA) Operational Navigation Charts (ONC) whose information dates from the 1970s to the 1990s (Birkett and Mason 1995). The positional accuracy of DCW varies considerably between regions, and there is no distinction between natural rivers and artificial canals.

2.4 ArcWorld global vectorized river network

The ArcWorld data set (ESRI 1992) includes a global vector map of surface water bodies at a resolution of 1:3 million. As part of its classification scheme, it distinguishes linear rivers into natural (perennial and intermittent) or artificial (canals) and provides approximately 7000 polygons of large open water bodies (including rivers and lakes). Although digitized at a coarser scale, ArcWorld seems to include some corrections and updates as compared to DCW and provides a consistent focus on major rivers and lakes of the world.

2.5 Global Lakes and Wetlands Database (GLWD)

The Global Lakes and Wetlands Database (Lehner and Döll 2004) combines a variety of existing global lake and wetland maps (at 1:1 to 1:3 million resolution) into one consistent coverage. It provides shoreline polygons of approximately 250,000 lakes and reservoirs worldwide, including their surface areas and other attributes. As for lakes and reservoirs, GLWD is largely based on DCW and ArcWorld but also includes various updates and data corrections.

2.6 Various regional datasets used for reference and quality control

2.6.1 Atlas of Canada, 1:1,000,000 National Frameworks Data, Hydrology

The National Scale Frameworks Hydrology data (Natural Resources Canada, 2006) consists of area, linear and point geospatial and attribute data at a scale of 1:1 million for Canada's hydrology at a national extent. It provides a representation of Canada's surface water features, and data completeness largely reflects the original source (VMAP0, revision 4, hydrographic features) for which additional editing has been performed. The data includes stream lines, lakes, and watershed boundaries.

2.6.2 National Atlas of the United States, Two-Million-Scale Streams and Waterbodies

The Two-Million-Scale Streams and Waterbodies map layer (National Atlas of the United States, 2005) shows the major water features of the United States, Puerto Rico, and the U.S. Virgin Islands that can be represented at a map scale of 1:2,000,000, including streams and rivers, canals, aqueducts, lakes, reservoirs, marshes, glaciers, waterfalls, and dams. The map layer was compiled by the National Atlas of the United States, and the U.S. Geological Survey collected information on water features to support its production.

2.6.3 Global Map Australia 1M 2001, Drainage layer (Water courses)

Global Map Australia 1M 2001 (Geoscience Australia, 2007)is a digital dataset covering the Australian landmass and island territories, at a 1:1 million scale. Vector data for Global Map Australia 1M 2001 was produced by generalizing Geoscience Australia's GEODATA TOPO 250K Series 1 data. It consists of a series of layers of information, including attributed drainage features (stream lines).

3. Database development

This chapter provides an overview of the applied processing steps for the generation of HydroSHEDS and discusses its main technical specifications. For a discussion regarding the suitability of HydroSHEDS data for specific applications, the user is referred to chapter 4.

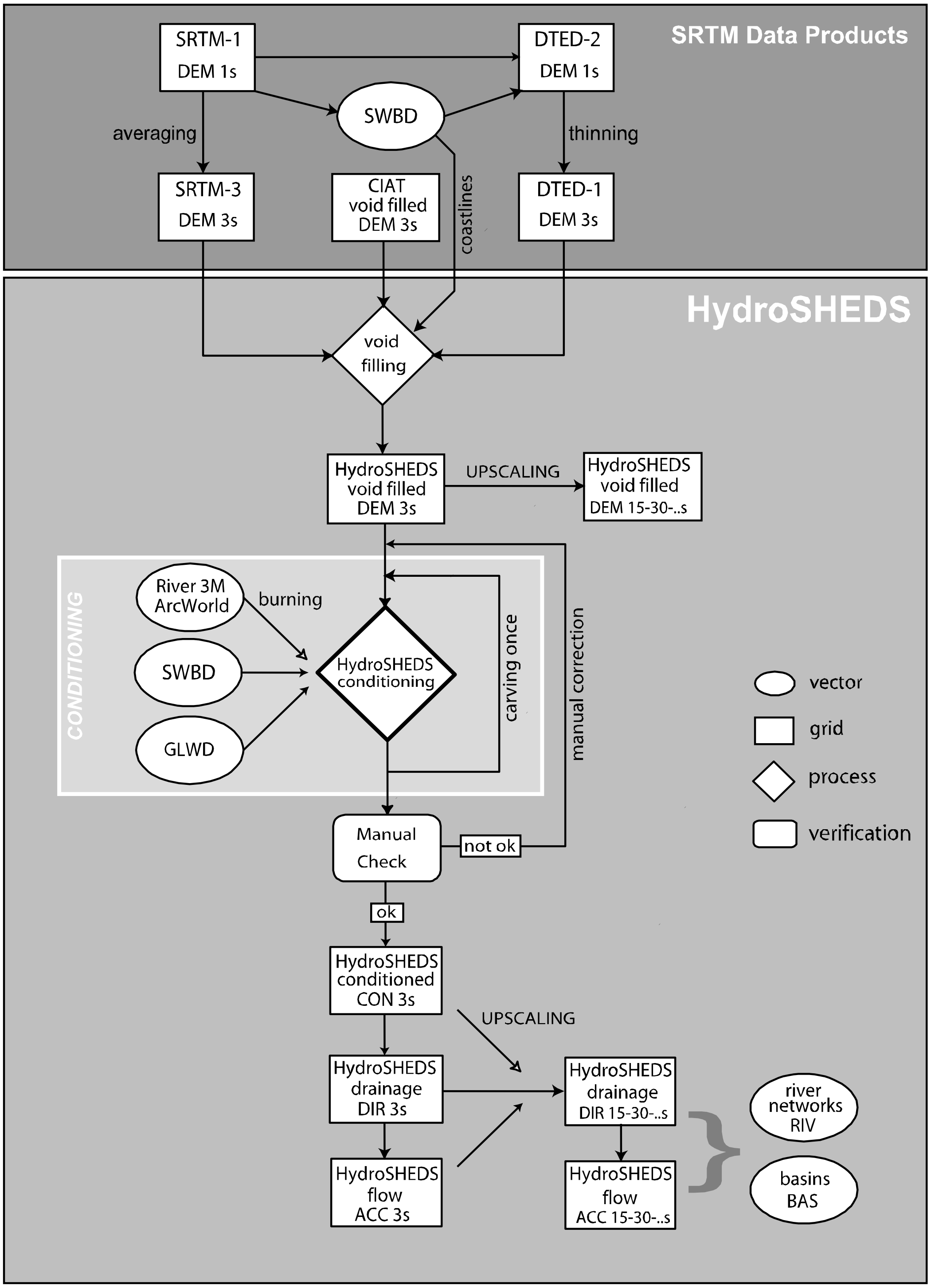


Figure 1: Flowchart of the generation of HydroSHEDS; for abbreviations and further explanations see text.

3.1 Combination of unfinished SRTM-3 and finished DTED-1 data

3.1.1 Combining SRTM-3 and DTED-1 original data

For the generation of HydroSHEDS, the performance of the publicly available SRTM-3 and DTED-1 versions of SRTM at 3 arc-second resolution have been tested. Due to their specific characteristics, each data set showed both advantages and disadvantages for hydrological applications.

As stated earlier, SRTM-3 has been derived through averaging of 1 arc-second SRTM data, as opposed to the subsampling method of DTED-1. As averaging reduces the high frequency “noise” that is characteristic of radar-derived elevation data, it is the method generally preferred by the research community (NASA/JPL 2005).

On the other hand, SRTM-3 data does not represent open water surfaces and shorelines well. DTED-1 has been specifically corrected to represent these features. However, the correction protocol introduced some critical artifacts for hydrological applications. For example, when large rivers were identified and monotonically stepped down in height towards the ocean, it was assured that the surface of each river pixel was lower than that of the directly adjacent land pixels. Yet a slightly elevated riverbank, say due to a levee or simply caused by the interpretation of riparian vegetation in the radar image, may allow for a river reach being somewhat higher than the floodplain behind the riverbank. Since the original processing was performed at 1 arc-second resolution, the elevated riverbank can disappear in the aggregated 3 arc-second version if it is only thin (one pixel wide). The resulting effect in the derived flow direction map is a possible breakout of the river course into the floodplain.

For above reasons, and after conducting a series of local tests, it was decided to apply both SRTM-3 and DTED-1 data in combination. For each pixel the minimum value found in either SRTM-3 or DTED-1 was used to generate an initial HydroSHEDS elevation model. The minimum requirement preserves the lower of both surfaces in the combined elevation data, which is considered desirable for the later identification of drainage directions.

3.1.2 Ocean shoreline

At the ocean surface, the combined data initially shows elevation values of 0 (from DTED-1) or negative (from SRTM-3). Since land close to the shoreline can also be 0 or even negative, using elevation alone as a criterion does not allow for a clean identification of the ocean shoreline. Thus to aid in the shoreline delineation, SWBD was employed as ancillary data: where SWBD indicates “ocean”, the values of the HydroSHEDS elevation model were reclassified to no-data. The resulting shoreline was then slightly generalized in order to remove small artifacts: land was first extended by a one-pixel rim into the ocean and the boundary was then smoothed using a local cell filter. All detached ocean surfaces (e.g. small estuaries entirely surrounded by land cells) were treated as land and their elevation values were retained rather than set to no-data. Some larger rivers are defined in SWBD to extend relatively far into the ocean. In these cases, the shoreline was modified based on the shoreline of DCW. Some very small islands are missing in the source data and are thus not represented in HydroSHEDS. Finally, some minor errors were detected in SWBD in visual inspections (e.g. some incomplete island boundaries) and were individually corrected. Note that in all up-scaled HydroSHEDS layers, each cell that contains at least one land cell at 3 arc-second resolution is defined as land.

3.1.3 Data shift

Both SRTM-3 and DTED-1 original 1-degree by 1-degree data tiles are defined via the coordinates of the center of their lower-left pixel (see 2.4). This characteristic leads to overlapping edges of adjacent tiles and to some artifacts when aggregating a tile to coarser resolutions: either all adjacent tiles have to be included in the aggregation process, or overlapping edges may have to be eliminated in the result. As the processing steps for the generation of HydroSHEDS are rather complex and aggregation (scaling) plays an important role, it was decided to shift the original SRTM data by 1.5 arc-seconds to the north and east, and to remove each tile's overlapping right column and top row. This shift leads to a 3 arc-second HydroSHEDS tile having 1200 rows and 1200 columns at an extent of exactly 1-degree by 1-degree without overlaps to adjacent tiles. All other HydroSHEDS resolutions are based on the initial 3 arc-second data and, therefore, include this shift. With respect to deriving river networks, the effect of the shift on the accuracy of the data can be considered negligible, particularly when compared to the subsequently applied data manipulations as discussed below. Note, however, that the shift may lead to significant anomalies when directly comparing HydroSHEDS elevation data and original SRTM elevation data.

3.2 Void-filling

In its original release, SRTM data contains regions of no-data (voids), specifically over large water bodies, such as lakes and rivers, and in areas where radar-specific problems prevented the production of reliable elevation data. These areas include mountainous regions where the radar shadow effect is pronounced, such as the Himalayas and Andes, as well as certain land surfaces, such as bare sand or rock conditions as found in the Sahara Desert. The existence of no-data in the DEM causes significant problems for deriving hydrological products, which require continuous flow surfaces. Therefore, a void-filling procedure has been applied to provide a continuous DEM for HydroSHEDS.

Numerous methods have been developed for void-filling of SRTM data (see e.g., Gamache 2004a), but they rarely focus on specific hydrological requirements. For HydroSHEDS, two different void-filling algorithms have been applied in combination. The first has been developed by CIAT (see Jarvis 2004; in collaboration with R. Hijmans and A. Nelson). The second has been specifically developed for HydroSHEDS as introduced in this paper. While the CIAT algorithm delivers smooth interpolation surfaces, the HydroSHEDS algorithm focuses on deepening and flattening missing water surfaces. Both methods and their combination are summarized below.

3.2.1 CIAT algorithm

The CIAT algorithm fills the no-data voids by applying an interpolative technique. The original SRTM elevation data are used to produce contours at an interval of 10 meters. The contours are interpolated using the TOPOGRID algorithm in Arc/Info. TOPOGRID, based upon the established algorithms of Hutchinson (1988, 1989), is designed to use contour and point elevation data along with mapped hydrography to produce hydrologically sound DEMs. This method produces a smooth elevation surface within the no-data regions. While micro-scale topographic variation is likely to be underrepresented, most macro-scale features are captured well in small to intermediate sized voids. Jarvis et al. (2004) performed a detailed analysis of the accuracy of the interpolated elevation data for a region in Colombia and found little difference when compared to a cartographic DEM, particularly for hydrological applications. Gamache (2004b and personal communication) also analyzed the CIAT results and concluded that the void-filling algorithm is quite successful in representing broad scale patterns in topography. For the production of HydroSHEDS, Version 2 of the CIAT void-filled SRTM data was used (CIAT 2004). The data is available from the CGIAR-CSI SRTM 90m Database at http://srtm.csi.cgiar.org.

3.2.2 HydroSHEDS algorithm

The HydroSHEDS algorithm fills the no-data voids by means of an iterative neighborhood analysis. The first step fills the outermost pixel-rim of a no-data void using a combination of a 3x3 minimum and a 5x5 mean filter (the minimum filter dominates the mean filter by a factor of 3:1). Then, the next pixel-rim is filled until the entire no-data void is processed. The no-data area is finally smoothed using a 9x9 mean filter. Particularly in the case of lakes and large river surfaces the emphasis of the minimum filter results in rather low elevation values inside the voids and a relatively flat relief as small peaks are successively filtered out.

3.2.3 Combination of void-filling algorithms

The lowering effect of the HydroSHEDS algorithm for open water surfaces seems desirable for hydrological applications, as it tends to force the flow course to stay within river channels and lakes. In mountainous regions, however, the CIAT results are expected to better represent the general topography. To optimize results, both algorithms were combined. For each pixel the minimum value of either the CIAT or HydroSHEDS algorithm was used. If, however, the HydroSHEDS algorithm computed values more than 30 meters lower than CIAT, CIAT values minus 30 meters were used.

In some large no-data voids entire mountains are lost using either of the two filling methods. Therefore, starting at a distance of 0.03 degrees (approximately 3 km at the equator) from the rim of large voids, elevation values were inserted from GTOPO30, a global DEM at 30 arc-second (approximately 1 km at the equator) resolution (Gesch et al. 1999). To avoid cliff effects, the inserted values were smoothed or “feathered” in a 0.03-degree wide transition zone.

The filled voids were then merged into the initial HydroSHEDS elevation data to provide a continuous elevation surface with no void regions. The entire process was performed for each 1-degree by 1-degree tile with a 0.25-degree overlap to the eight adjacent tiles, thus ensuring seamless transitions of topography even in areas with large voids.

The final result of steps 3.1 and 3.2 is
the HydroSHEDS void-filled elevation model (**DEM**) at 3 arc-second resolution.

3.3 Sink identification

Typically, an original DEM will show a large number of sinks or depressions. These are single or multiple pixels which are entirely surrounded by higher elevation pixels. Some of these sinks are naturally occurring on the landscape, representing endorheic (inland) basins with no outlet to the ocean. In most cases, however, the sinks are considered spurious, often caused by random and mostly small deviations in the elevation surface. These anomalies occur even in high quality DEMs and high resolutions due to DEM production methods. The spurious sinks are critical problems in hydrological applications as they interrupt continuous flow across the DEM surface. Therefore, sinks are typically removed from the DEM before deriving a river network. Standard GIS procedures have been developed to remove spurious sinks, and a common approach is to raise the elevation values within the sinks until an outflow point is encountered. Natural sinks can be forced to remain in the DEM through "seeding", e.g. by putting a no-data cell into their center.

As for HydroSHEDS, the definition of natural vs. spurious sinks has been accomplished using a GIS-assisted manual process. All sinks of the void-filled elevation model were identified in a standard GIS procedure, and their maximum depth and extent were calculated. Sinks deeper than 10 meters and larger than 10 km^2 were highlighted as "potential" natural sinks. All regions of potential natural sinks were then inspected visually and were either seeded or rejected. The decision was based on information derived from DCW, ArcWorld, GLWD, and additional atlases and maps. For example, a mapped "salt lake" with no obvious river draining from it is considered a strong indication for an endorheic basin. The visual inspections were performed at a zoom to 1-degree by 1-degree windows, and several thousand naturally occurring sinks were identified globally.

Obviously, the manual sink identification process is subjective, and in many cases the definition of natural sinks is difficult and ambiguous. Some depressions overflow periodically, following seasonal flooding cycles, others spill only occasionally. Some large, relatively dry areas may show numerous small depressions within a generally sloped surface, and flow paths are poorly developed if at all (e.g. the Argentinean Pampas and many desert areas). These depressions may or may not overflow in a rain event. In some areas of no obvious drainage only some "structural" sinks were placed at strategic locations. They do not terminate the flow at all single depressions but at a final one to indicate the endorheic character of the region. In karstic areas, rivers may disappear in surface depressions, yet they can be closely connected to a larger basin via underground pathways. In cases of large karstic depressions, sinks were introduced, as it seemed easier for a user to later remove the sinks and restore connectivity than to introduce them from scratch. Artificial sinks, however, like large pits in surface mining areas, were rejected.

3.4 Hydrologic conditioning

Besides sinks, original DEMs show a series of other characteristics, artifacts and anomalies that can cause significant problems or errors in hydrological applications. Some types of problems that are typical for the SRTM elevation model are discussed in section 4. The most significant characteristic is likely the fact that the elevation values of SRTM, being a radar-derived product, are influenced by the vegetation cover. In areas of low relief, these small deviations from the true surface elevation can cause significant errors in the derived river courses and flow directions.

In order to improve the performance of a DEM for hydrological applications, a series of GIS processes and procedures exist and are routinely applied. Yet due to the individual characteristics of different DEMs and, on a global scale, due to the regional variations in the type of errors, no one method exists that addresses all possible problems. For HydroSHEDS, a sequence of hydrologic conditioning procedures has been implemented, either adapted from standard GIS functionality, newly developed, or customized. The general focus was to strike a compromise between forcing the DEM to produce correct river network topology, particularly for the largest of rivers, while preserving as much original SRTM information as possible. Note that in any case the conditioning process alters the original elevation data and may render it unusable for other applications.

The following hydrologic conditioning procedures have been applied to the HydroSHEDS elevation data:

3.4.1 Deepening of open water surfaces

All rivers and lakes as identified in SWBD were deepened by 10 meters in order to force the derived flow to stay within these objects. As no-data voids in the original SRTM elevation data may also indicate open water surfaces (see 3.2), all void areas were lowered by 10 meters as well. The 10-meter threshold was chosen as it imparts a strong enough effect in flat areas (where the identification of river channels and lakes is particularly difficult), while producing only insignificant changes in areas with steeper slopes (where no-data voids are probably caused by radar shadow rather than open water).

3.4.2 Weeding of coastal zone

In the radar-derived elevation model, mangrove or coastal vegetation belts may be interpreted as a low but continuous embankment blocking any direct outflow to the ocean. In the derived river network model, these barriers can cause significant backwater effects. To reduce this effect, the coastal zone, i.e. a 0.02-degree wide buffer (approximately 2 km at the equator) along the ocean shoreline was “weeded” by reducing every random third cell by 5 meters. This subtle change, in combination with the following filters, forces occasional breakthroughs into the slightly elevated coastal embankments.

3.4.3 Stream burning

The most extensive conditioning process in the generation of HydroSHEDS has been the so-called “stream burning” procedure. Stream burning is an often-used process to enforce known river courses into an elevation surface. The elevation values along the rivers, as depicted e.g. in an existing vector layer, are lowered by a certain value, thus “burning” deep gorges into the elevation surface. The burning can be extended to include a buffer around the river lines in order to shape a smoother transition between the original surface and the gorge. For HydroSHEDS, only large rivers and lakes were burned into the elevation surface in order to avoid excessive alterations of the SRTM surface. All perennial and intermittent rivers and lakes of ArcWorld, as well as all rivers and lakes of GLWD were used, while the higher resolution but unclassified DCW data was omitted. Since the accuracy of the existing global maps is unknown, attempts were made to minimize the impact of these data sets on the SRTM data. After multiple tests, the burning depth for rivers was set to 12 meters, with a buffer of 0.005 degrees (approximately 500 meters at the equator) around the river courses. The burning depth was reduced, in a stepwise manner, from 12 meters at the thalweg to 2 meters at the edge of the buffer. Lakes were burned with a depth of 14 meters and a buffer distance of 0.0025 degrees. The parameter setting aimed for a noticeable forcing of the main rivers in flat areas, where otherwise the correct delineation of rivers is difficult. In steep regions, the small burning depth results in rather insignificant changes of the elevation surface, hence the SRTM data remains the dominant information for deriving drainage directions.

3.4.4 Filtering

The entire elevation surface was then filtered by applying a directional 3x3 neighborhood analysis. The elevation values of all possible straight and obtuse angle flow paths in a 3x3 kernel were averaged and the minimum value was assigned to the center cell. This filter aims to remove remaining spikes and wells while preserving and enforcing linear river courses and valley bottoms. In particular, single pixels that can block a continuous flow path are removed.

3.4.5 Molding of valley courses

Next, valley courses were depicted through a neighborhood terrain analysis and were deepened by 3 meters. The valleys were identified through a 5x5 kernel median analysis combined with a grid-thinning algorithm to detect linear features. This procedure of valley “molding” has been specifically developed to improve river delineations in tropical lowland areas by removing small obstacles in shallow valleys. Due to the small deepening of 3 meters, no significant changes occur in areas with stronger relief.

3.4.6 Sink filling

In a standard process, all spurious sinks in the elevation surface were filled. Natural sinks were seeded in order to exclude them from removal (see 3.3).

3.4.7 Carving through barriers

After sink filling, a river map was produced from the conditioned elevation surface. All main river courses, defined as rivers with an upstream catchment area of more than 1000 cells (approximately 8 km^2 at the equator), were depicted. The main rivers were then projected onto the initial HydroSHEDS elevation model, and all elevation rises along the rivers when moving downstream were identified. These rising reaches in the original elevation surface, which have obviously been removed through filtering or sink-filling in the conditioning process, may represent dams, bridges, embankments of any kind, or narrow gorges that block the flow path. In many of these cases, the sink-filling effect (i.e. the lifting and implicit flattening of the dammed area) may not be desirable as any existing relief information within the filled area is lost. To minimize this effect, a second conditioning iteration was performed: first, all rising reaches along the main river courses were leveled out in the initial elevation data by appropriately lowering their respective heights, thus effectively “carving” through the barriers. After this process, all other conditioning steps (3.4.1 to 3.4.6) were repeated.

During the entire conditioning process, hard- and software limitations were reached due to the very large data size at 3 arc-second resolution. All steps have therefore been performed on a tile-by-tile basis, with extents between 1-degree by 1-degree and 5-degree by 5-degree. In order to avoid edge effects, appropriate overlaps to the adjacent tiles were added. In particular the sink-filling algorithm proved highly susceptible to tile sizes and edge effects and had to be implemented in an iterative approach. The processing was performed with an overlap of up to 5 degrees (approximately 500 km at the equator) to adjacent tiles to ensure seamless results without edge effects.

3.5 Manual corrections

The result of section 3.4 is a hydrologically conditioned elevation surface at 3 arc-second resolution. From this elevation surface, a new river network was derived and used for error checking. Because computation of the river network at 3 arc-second resolution is very time intensive, the data was first upscaled to 15 arc-second resolution (approximately 500 meters at the equator; for upscaling see 3.6 below). The derived river network was then compared visually to the rivers of DCW, ArcWorld, and various atlases and paper maps.

Errors occurred particularly in flat areas with varying vegetation cover (see section 4), such as floodplains and coastal zones. If the actual rivers could be visually detected in the raw elevation data, their courses were traced or adopted from the existing DCW river layer. These rivers were then added to the stream layer used in the river burning procedure of 3.4. In some areas, the given elevation values significantly misrepresented the actual flow conditions (e.g. blocked pathways due to narrow gorges, or inadequate filling of the no-data voids of the original data). In these cases, the burning depth was individually adjusted. Some other topological problems (e.g. diversions into canals or multiple spillways of reservoirs) were treated in a similar manner through introduction and adjustment of main pathways. Actual flow channels of braided rivers and large river deltas could not be topologically resolved due to the constraint of allowing only one drainage direction per cell (the single flow direction algorithm does not allow for river bifurcations). These zones have only been "cleaned" to represent the main channel properly.

After detecting the errors and preparing the corresponding correction data, all steps of 3.4 were repeated. In some areas, several iterations of manual corrections were performed. As with the sink identification process, the manual correction process is highly subjective. The visual inspections were performed at a zoom to one-degree by one-degree windows, and corrections were applied for several thousand locations globally.

The final results of steps 3.4 and 3.5 are
(1) the HydroSHEDS hydrologically conditioned elevation model (**CON**), and
(2) the HydroSHEDS drainage direction map (**DIR**) at 3 arc-second resolution.

3.6 Upscaling

All procedures described in sections 3.1 to 3.5 were performed at 3 arc-second resolution. Yet for many applications, in particular continental or global assessments, coarser resolutions are desirable as they may significantly reduce calculation times while providing acceptable accuracy. HydroSHEDS therefore delivers various resolutions, from 3 arc-second to 5 minute. The coarser resolutions are all derived from the 3 arc-second data through upscaling.

Upscaling drainage directions is not a straightforward process, as typical aggregation methods, such as averaging of neighborhood kernels, are not appropriate for directional values. A frequently applied upscaling method is to first upscale the elevation data, and then derive a new drainage direction map from this coarser DEM. This method is generally fast and easy to perform, but it often delivers low-quality results with respect to river network topology, due to the loss of significant information in the aggregation process. An alternative option is to first derive the river network at high resolution, and then to upscale this network. This option preserves the network information, which is most important for hydrological applications. However, it requires complex procedures, which are difficult to realize at a global scale and for the desired high resolutions. As a compromise, a combined method has been developed and applied for generating HydroSHEDS. The main steps in the upscaling process are as follows:

1. The void-filled DEM is upscaled from the original 3 arc-second to the desired resolution. For this process, an algorithm was applied that calculates both the mean and minimum value found within the aggregation kernel and then takes the average. The minimum value is included in the calculation to emphasize valleys. Natural sinks were preserved in the upscaling process.

2. A network of main rivers was calculated at 3 arc-second resolution. Main rivers were defined as those having an upstream catchment area of more than 1000 cells (approximately 8 km^2 at the equator). The river network was derived for five-degree by five-degree tiles with a one-degree overlap to adjacent tiles to avoid edge effects.
3. The main rivers were then burned into the upscaled elevation surface. The burning depth was defined as the sum of a constant (500 meters) and a value dependent on the size of the respective river reach (0-400 meters, proportional to the logarithm of upstream cells). The relatively large burning depth assured that the river channels were preserved in the new elevation surface. No buffering was applied.
4. Sinks were filled in the upscaled and burned elevation surface, and finally new drainage directions were calculated. Note that due to the strong burning, the elevation surface does not represent natural conditions any more. It is appropriate only for deriving drainage directions. To avoid confusion with true DEMs, the upscaled elevation surface is not offered as a standard HydroSHEDS product.

The upscaling process delivers a new drainage direction map (DIR) from which a new river network can be derived. Due to the applied stream burning, all main rivers (as defined in the upscaling process) should be in very good alignment with the original river network. Only if two close-by rivers drain through the same or adjacent upscaled cells, they may be incorrectly merged into one flow channel. Smaller rivers, for which no burning occurs, are based solely on the upscaled elevation surface. Their quality may thus differ from the river network at 3 arc-second resolution.

The final results of step 3.6 are upscaled HydroSHEDS drainage direction maps (**DIR**) at resolutions of 15 arc-second and 30 arc-second. Also, a 5 minute product is in preparation.

3.7 Derived products

Ancillary HydroSHEDS products can be derived from the individual drainage direction maps at their respective resolutions. These products include flow accumulations, flow distances, river networks, and watershed boundaries. A list of available HydroSHEDS datasets is provided in chapter 5.

4. Quality assessment

With all digital geospatial datasets, users must be aware of certain characteristics of the data, such as resolution, accuracy, method of production, and resulting artifacts, in order to be able to judge its suitability for a specific application. A characteristic that renders the data unsuitable for one application may have no relevance as a limiting factor in a different application (NASA/JPL 2005).

Despite several (uncompleted) efforts, the final data quality of the HydroSHEDS product has not been evaluated systematically. Yet regional comparisons with other global hydrographic data sets support the following conclusions:

- HydroSHEDS shows significantly better accuracy than other global river network representations derived from elevation data. In particular, due to its superior underlying digital elevation model, HydroSHEDS represents a clear improvement over HYDRO1k, an earlier and widely used global hydrographic data set at 1-km resolution (USGS 2000).
- HydroSHEDS tends to show better accuracy than the 1:1 million DCW (VMAP-0) mapping product. However, the accuracy of both data sets varies by location. In some regions where HydroSHEDS is particularly susceptible to errors, such as vegetated floodplains, the quality of DCW can be superior to HydroSHEDS.
- As a global product, HydroSHEDS does not reach the accuracy of high-resolution local river networks (e.g. those depicted on national 1:50,000 hydrographic maps). The user is thus encouraged to further improve HydroSHEDS through incorporation of local information.

Typically, river network products derived from digital elevation surfaces are susceptible to various errors, foremost in flat regions without well-defined relief. Additionally, the quality of HydroSHEDS depends on the characteristics of the SRTM-based elevation model. Being a radar product, SRTM elevation values are influenced by vegetation and other surface effects, such as roughness, wetness, low backscatter signal at open water surfaces and radar shadow (Freeman 1996). Known regions prone to errors in HydroSHEDS include:

- Areas of low or not well-defined relief, including lake surfaces.
- Areas with varying vegetation cover and low-relief topography, e.g. large river floodplains. The radar signal is, at least partly, reflected from atop and within the vegetation cover and the returned signal is a complex mix of land surface elevation and vegetation height.
- Low-relief coastal areas, in part due to the barrier effect of mangroves.
- Large-scale roads or clearings in vegetation of low-relief areas. The lack of vegetation causes artificial depressions in the elevation surface.
- Rivers less than 90 m wide enclosed by riparian vegetation. The vegetation effect can cause the river channel to appear slightly elevated.
- Braided rivers and deltas. The use of the single flow direction algorithm does not allow for depiction of river bifurcations.
- Narrow gorges. If a gorge is less than 90 m wide, it can appear closed on the elevation surface.
- Inland sinks, depressions, and karst features such as sinkholes. The hydrologic connections are often ambiguous or temporary in nature. In karst areas flow paths are not necessarily terminated at sinks due to possible underground connectivity, and artificial depressions like large-scale mining may have flow bypasses.
- Elevated “barriers” in the elevation surface that in reality have no effect on flow connectivity (e.g. bridges, high-density housing areas).
- Areas of no-data voids in the original SRTM data. The larger the void, the more uncertain is the filled surface (see 3.2).

5. Data layers and availability

The HydroSHEDS database provides a suite of raster and vector datasets, covering many of the common derivative products used in hydrological analyses. The HydroSHEDS data layers are being developed for all landmasses of the globe from 56° South to 84° North (i.e. excluding Antarctica). The data is prepared in seamless mode (no edge effects) and is provided as global or regional maps, or in 5-degree by 5-degree tile coverage.

HydroSHEDS is produced on a continental basis. The drainage direction maps—which represent the core data layer from which all other products are derived—have been completed or are scheduled for release as follows:

Original regions (56° South to 60° North)	*Timeline*
South America	Completed (May 2006)
Central America (Mexico and Caribbean)	Completed (March 2007)
Asia	Completed (March 2007)
Africa	Completed (October 2007)
Australasia	Completed (March 2008)
Europe and Middle East	Completed (October 2008)
North America (USA and Canada)	Completed (January 2009)
Appended regions (north of 60° North)	
Arctic (northern Canada)	Scheduled for July 2014
Scandinavia and Iceland	Scheduled for July 2014
Siberia	Scheduled for July 2014
Greenland	Scheduled for July 2014

At this time, HydroSHEDS provides several core data sets at 3 arc-second, 15 arc-second, and 30 arc-second resolutions, including elevation surfaces, drainage directions, river networks, and watershed boundaries. Additional data layers are under construction, and future data developments are planned. Please refer to the announcements on the HydroSHEDS website (http://www.hydrosheds.org) for further information. Currently, the following data layers and resolutions are available:

5.1 Void-filled digital elevation model

Name signature	**DEM**
Data format	Raster
Values	Elevation in meters (referenced to WGS84 EGM96 geoid)
Projection	Geographic (latitude/longitude) referenced to WGS84 horizontal datum
Available resolutions	3 arc-second, 15 arc-second, 30 arc-second

The elevation layers distributed with HydroSHEDS are based on a combination of the original SRTM-3 and DTED-1 elevation models of SRTM (for further specifications see 2.1 to 2.4 and 3.1). No-data voids have been filled using interpolation algorithms, and the data has been clipped at the ocean shoreline. Resolutions other than 3 arc-second are derived through aggregation (averaging): each upscaled data pixel is generated by averaging the corresponding neighborhood kernel of the 3 arc-second data. Note that for technical reasons HydroSHEDS elevation data show a consistent shift of 1.5 arc-seconds to the north and east as compared to original SRTM data (see 3.1.3).

SRTM elevation data is provided in geographic projection (latitude/longitude) referenced to the WGS84 horizontal datum, and EGM96 vertical datum. Note that most global positioning systems use the WGS84 vertical datum as default. This difference in vertical datums may result in absolute elevation differences between +106 m and -86 m. For precise comparison of GPS elevation data with SRTM elevation data, conversion of the vertical datum should be considered beforehand.

5.2 Hydrologically conditioned elevation

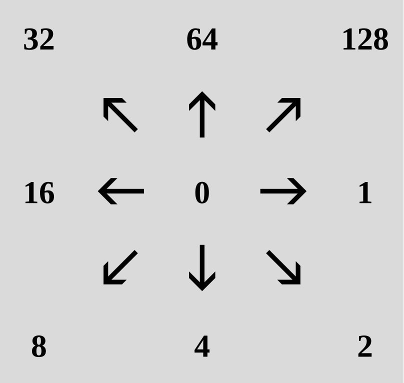
Name signature	**CON**
Data format	Raster
Values	Elevation in meters (referenced to WGS84 EGM96 geoid)
Projection	Geographic (latitude/longitude) referenced to WGS84 horizontal datum
Available resolutions	3 arc-second

The hydrologically conditioned elevation layers distributed with HydroSHEDS are the result of an iterative conditioning and correction process described in detail in section 3. For the specifics of the underlying digital elevation model see 5.1. Note that the conditioning process alters the original DEM and may render it incorrect for applications other than deriving drainage directions. Endorheic basins (inland sinks) are “seeded” with a no-data cell at their lowest point in order to terminate the flow.

5.3 Drainage directions

Name signature	**DIR**
Data format	Raster
Values	Drainage directions in ESRI format (see below)
Projection	Geographic (latitude/longitude) referenced to WGS84 datum
Available resolutions	3 arc-second, 15 arc-second, 30 arc-second

The drainage direction maps distributed with HydroSHEDS define the direction of flow from each cell in the conditioned DEM to its steepest down-slope neighbor. Values of flow direction vary from 1 to 128. All final outlet cells to the ocean are flagged with a value of 0. All cells that mark the lowest point of an endorheic basin (inland sink) are flagged with a value of -1. The flow direction values follow the convention adopted by ESRI's flow direction implementation:



5.4 Flow accumulation (number of cells)

Name signature	**ACC**
Data format	Raster
Values	Flow accumulation in number of cells (see below)
Projection	Geographic (latitude/longitude) referenced to WGS84 datum
Available resolutions	15 arc-second

The flow accumulation maps distributed with HydroSHEDS define the amount of upstream area (in number of cells) draining into each cell. The drainage direction layer is used to define which cells flow into the target cell. The number of accumulated cells is essentially a measure of the upstream catchment area. However, since the cell size of the HydroSHEDS data set depends on latitude, the cell accumulation value cannot directly be translated into drainage areas in square kilometers. A flow accumulation map reflecting true catchment areas is in preparation. Values range from 1 at topographic highs (river sources) to very large numbers (on the order of millions of cells) at the mouths of large rivers.

5.5 River network (stream lines)

Name signature	**RIV**	
Data format	Vector (lines)	
Projection	Geographic (latitude/longitude) referenced to WGS84 datum	
Resolutions	15 arc-second, 30 arc-second	
Line attributes	ID	unique identifier
	Up_cells	max. flow accumulation (number of cells) of stream reach

The river network layers distributed with HydroSHEDS are directly derived from the drainage direction layers. The flow accumulation layer is used for selection and attribution. Only rivers with upstream drainage areas exceeding a certain threshold are selected: for the 15 arc-second resolution a threshold of 100 upstream cells has been used. The vectorized river reaches are currently attributed with the maximum flow accumulation (in number of cells) occurring within each river reach. More attributes will be added in future versions.

5.6 Drainage basins (watershed boundaries)

Name signature	**BAS**	
Data format	Vector (polygons)	
Projection	Geographic (latitude/longitude) referenced to WGS84 datum	
Resolutions	15 arc-second, 30 arc-second	
Polygon attributes	ID	unique identifier
	Area_skm	surface area in square kilometers

The current drainage basin layers distributed with HydroSHEDS are showing contiguous watersheds only, i.e. without further subdivisions (thus their naming syntax is amended with “beta”). A sequence of additional layers with hierarchical, nested sub-basin delineations is under development.

6. Data formats and distribution

6.1 File name syntax

HydroSHEDS provides data in various regional extents, types, and resolutions. Information about the content is provided in the file names which follow the naming convention "*Extent_DataType_Resolution*".

6.1.1 Regional extent

To facilitate electronic distribution, all raster data at 3 arc-second resolution have been divided into 5-degree by 5-degree tiles. For example, the area of South America is covered by 103 individual tiles. The tile names are defined by a 7-digit identifier which refers to the latitude and longitude of the lower-left (southwest) corner of the tile. For example, the coordinates of the lower-left corner of tile s15w065 are 15 degrees south latitude and 65 degrees west longitude.

File names of layers with continental extent are defined by a two-digit identifier:

Identifier	*Region*
Af	Africa
As	Asia
Au	Australasia
Ca	Central America (Mexico and Caribbean)
Eu	Europe and Middle East
Na	North America (USA and Canada)
Sa	South America

6.1.2 Data type

Identifier	*Type of data*
DEM	Digital elevation model (void-filled)
CON	Hydrologically conditioned elevation
DIR	Drainage directions
ACC	Flow accumulation (number of cells)
RIV	River network (stream lines)
BAS	Drainage basins (watershed boundaries)

6.1.3 Resolution

Identifier	*in sec/min*	*in degree*	*in meters/km*
3s	3 arc-second	0.000833333333333	approx. 90 m at the equator
15s	15 arc-second	0.0041666666666667	approx. 500 m at the equator
30s	30 arc-second	0.0083333333333333	approx. 1 km at the equator
5m	5 minute	0.0833333333333333	approx. 10 km at the equator

Please note that all data provided in 5-degree by 5-degree tiles is in 3 arc-second resolution. However, the extension "3s" is omitted in order to shorten the file names.

6.2 Data formats

6.2.1 Vector data format – ESRI Shapefile format

The vector data sets distributed with HydroSHEDS, i.e. river lines (RIV) and basin polygons (BAS), are being made available in ESRI Shapefile format (ESRI 1998). A HydroSHEDS shapefile typically consists of five main files (.dbf, .sbn, .sbx, .shp, .shx). Additionally, basic metadata information is provided both in XML format (.xml) and in HTML format (.htm) following the FGDC standard. Projection information is provided in an ASCII text file (.prj). All shapefiles are in geographic (latitude/longitude) projection, referenced to datum WGS84.

6.2.2 Raster data formats

The raster data layers distributed with HydroSHEDS, i.e. the void-filled elevation model (DEM), the hydrologically conditioned elevation model (CON), drainage directions (DIR), and flow accumulation (ACC), are being made available in two different formats: (a) in ESRI GRID format, and (b) as binary raster image in ESRI Band Interleaved by Line (BIL) format. The simple BIL format should allow for easy ingest into most popular image processing and geographic information systems packages. Further information on the contents of the raster files is provided below.

(a) ESRI Grid format

The ESRI GRID format is supported within ESRI's ArcGIS software environment, but it can also be imported into various independent GIS packages. Each raster coverage is provided in a root folder (named after the raster layer) which contains two subfolders: the grid data folder (again named after the raster layer) and a corresponding "info" folder. In the grid folder, basic metadata information is provided both in XML format (.xml) and in HTML format (.htm) following the FGDC standard. Projection information is provided in an ASCII text file (prj.adf). All raster data are in geographic (latitude/longitude) projection, referenced to datum WGS84.

Note that it is not possible to store two or more grids in the same root folder by simply copying the according grid folders and merging the content of the "info" folders into one. To store multiple grids in one root folder, the grids have to be copied or moved from within an ESRI compatible software application.

(b) ESRI BIL format

Each raster image is provided as four main files, with the extension of each file defining the file type (see table below). Additionally, basic metadata information is provided both in XML format (.xml) and in HTML format (.htm) following the FGDC standard. Projection information is provided in an ASCII text file (.prj). All raster data are in geographic (latitude/longitude) projection, referenced to datum WGS84.

Extension	*File type*
.bil	Raster data file
.hdr	Header file
.blw	World file
.stx	Statistics file

Raster data file (.bil)
The raster data for each layer are provided as unsigned integer data in a simple binary raster format (either 8-bit, 16-bit, or 32-bit). There are no header or trailer bytes embedded in the image. The data are stored in row major order (all the data for row 1, followed by all the data for row 2, etc.).

Header file (.hdr)
The raster data header file is an ASCII text file containing size and coordinate information for the layer. Many standard software packages require the header file to provide important geo-referencing information for the image. The following keywords are used in the header file:

BYTEORDER	byte order in which image pixel values are stored I = Intel byte order (least significant byte first)
LAYOUT	organization of the bands in the file BIL = band interleaved by line
NROWS	number of rows in the image
NCOLS	number of columns in the image
NBANDS	number of spectral bands in the image (1 for all raster layers)
NBITS	number of bits per pixel (8, 16 or 32)
BANDROWBYTES	number of bytes per band per row (twice the number of columns for the 16-bit image; four-times for the 32-bit image)
TOTALROWBYTES	total number of bytes of data per row (twice the number of columns for the 16-bit image; four-times for the 32-bit image)
BANDGAPBYTES	the number of bytes between bands in a BSQ format image (0 for all raster layers)
NODATA	Value used for masking purposes (-9999 for all elevation images DEM and CON; -9 for all drainage direction maps DIR; 0 for all flow accumulation maps ACC)
ULXMAP	longitude of the center of the upper-left pixel (decimal degrees)
ULYMAP	latitude of the center of the upper-left pixel (decimal degrees)
XDIM	x-dimension of a pixel in geographic units (decimal degrees)
YDIM	y-dimension of a pixel in geographic units (decimal degrees)

World file (.blw)
The world file is an ASCII text file containing coordinate information. It is used by some packages for geo-referencing of image data. The following parameters are provided in the world file:

XDIM	x-dimension of a pixel in geographic units (decimal degrees)
Rotation term	always zero
Rotation term	always zero
Negative YDIM	negative y-dimension of a pixel in geographic units (decimal degrees)
XMIN	longitude of the center of the upper-left pixel (decimal degrees)
YMAX	latitude of the center of the upper-left pixel (decimal degrees)

Statistics file (.stx)
The statistics file is an ASCII text file which lists the band number, minimum value, maximum value, mean value, and standard deviation of the values in the raster data file.

Important note for using BIL files:

Because the BIL image data (.bil) are stored in binary format, users must be aware of how the bytes are addressed on their computers. The data are provided in Intel byte order, which stores the least significant byte first (“little endeian”). Most PCs use the Intel byte order. However, systems such as Sun SPARC and Silicon Graphics workstations use the Motorola byte order, which stores the most significant byte first ("big endian"). Users with systems that address bytes in the Motorola byte order may have to "swap bytes" of the BIL data unless their application software performs the conversion during ingest. The statistics file (.stx) provided for each data set gives the range of values in the image file, so that users can check if they have the correct values stored on their system.

Because the BIL images are provided in unsigned binary format, negative values are stored in reverse order starting at the top end of the possible value range for each data file. For example, HydroSHEDS drainage direction maps (DIR) are stored as unsigned 8-bit data, allowing for values between 0 and 255, and the value -1 is stored as 255, -2 is 254, etc. HydroSHEDS elevation layers (DEM and CON) are stored as unsigned 16-bit data, allowing for values between 0 and 65535. Also, ocean areas have been masked in HydroSHEDS, using NoData flags of -9999 (DEM and CON layers) and -9 (DIR layer). Many software applications cannot directly interpret these negative values. An easy fix can be accomplished by converting the original BIL image using a formula like the following (from ArcInfo):

For 8-bit data:	out_grid = con(in_grid >= 128, in_grid - 256, in_grid)
For 16-bit data:	out_grid = con(in_grid >= 32768, in_grid - 65536, in_grid)

The converted grid will then have the negative values properly represented, and the statistics of the grid should match those listed in the .stx file. If desired, the -9999 or -9 ocean mask values in the grid can then be set to NoData. The HydroSHEDS flow accumulation maps (ACC), stored as unsigned 32-bit data, do not include negative values and the value 0 masks ocean areas.

6.3 Data distribution

HydroSHEDS data is available electronically in compressed zip file format from http://www.hydrosheds.org. [*Please note that the former data download site at the EROS Data Center of USGS at http://hydrosheds.cr.usgs.gov is now discontinued.*] To use the data files, the zip files must first be decompressed. Each zip file includes a copy of the HydroSHEDS Technical Documentation (but note that some files include the outdated version 1.0 of the Technical Documentation). Multiple zip files that contain data in ESRI Shapefile or BIL format can be decompressed into one folder. Multiple zip files that contain data in ESRI Grid format are by default decompressed into separate subfolders. To store multiple grids in one root folder, the grids have to be copied or moved from within an ESRI compatible software application.

Below are some estimates for file sizes of the compressed zip files. Decompressed, the various file sizes may exceed the zipped sizes by 10-fold and more.

Zipped file	*Estimated size range*
Five-degree by five degree tiles	10-50 MB
Continental grids at 15s resolution	30-100 MB
Continental grids at 30s resolution	10-50 MB
Continental vector files	10-50 MB

7. Frequently Asked Questions

This section provides answers to Frequently Asked Questions for users of HydroSHEDS and may be updated on the HydroSHEDS webpage (http://www.hydrosheds.org) as needed.

Q 01: If there is additional local information available in form of high-quality river maps, how can I incorporate this information in HydroSHEDS?

A 01: First, compare your local maps with the HydroSHEDS river network. If there are only a few differences, you can depict the respective river lines from the local map and burn them into the *hydrologically conditioned* DEM of HydroSHEDS. Then run a sink-fill process and derive a new drainage direction map and river network from the corrected and conditioned DEM. If the local river map shows generally better quality at many locations, or if the conditioning process of HydroSHEDS introduced an error that is not apparent in the raw SRTM data, you can use the local river map and burn it into the *void-filled* (un-conditioned) DEM of HydroSHEDS. However, you will then have to apply your own individual conditioning processes (at a minimum you will have to apply a sink-fill procedure) in order to produce a new hydrologically conditioned DEM.

Q 02: How can I remove an incorrect sink from HydroSHEDS and restore continuous flow?

A 02: Use the *hydrologically conditioned* DEM of HydroSHEDS. Reclassify the no-data cell that represents the incorrect sink with an appropriate elevation value. [There are different ways to achieve this, for example: make a new point theme; place a point at the location of the no-data cell and enter the desired elevation value into the attribute table; convert the point theme to a grid and merge it into the DEM.] Then run a sink-fill process and derive a new drainage direction map and river network from the corrected and conditioned DEM.

8. Disclaimer and acknowledgement

8.1 License agreement

HydroSHEDS data (as defined in Appendix A) are free for non-commercial and commercial use. For all regulations regarding license grants, copyright, redistribution restrictions, required attributions, disclaimer of warranty, indemnification, liability, waiver of damages, and a precise definition of licensed materials, please refer to the ***License Agreement*** as provided in Appendix A.

8.2 Acknowledgement and citation

We kindly ask users to cite HydroSHEDS in any published material produced using this data. If possible, online links to the HydroSHEDS website should be provided as follows:

http://www.hydrosheds.org

Scientific citations should be made as follows:

Lehner, B., Verdin, K., Jarvis, A. (2008): New global hydrography derived from spaceborne elevation data. Eos, Transactions, 89(10): 93-94.

9. References

Birkett, C.M., Mason, I.M. (1995): A new global lakes database for a remote sensing program studying climatically sensitive large lakes. Journal of Great Lakes Research 21(3): 307-318.

CIAT – International Centre for Tropical Agriculture (2004): Void-filled seamless SRTM data V2. Data available from the CGIAR-CSI SRTM 90m Database at http://srtm.csi.cgiar.org

ESRI – Environmental Systems Research Institute (1992): ArcWorld 1:3 Mio. Continental Coverage. ESRI, Redlands, CA. Data available on CD.

ESRI – Environmental Systems Research Institute (1993): Digital Chart of the World 1:1 Mio. ESRI, Redlands, CA. Data available on 4 CDs or at http://www.maproom.psu.edu/dcw/

ESRI – Environmental Systems Research Institute (1998): ESRI Shapefile Technical Description - An ESRI white paper. Available at http://www.esri.com/library/whitepapers/pdfs/shapefile.pdf

Farr, T.G., M. Kobrick (2000): Shuttle Radar Topography Mission produces a wealth of data. American Geophysical Union, Eos 81: 583-585.

Freeman, T. (1996): What is Imaging Radar? Jet Propulsion Laboratory. Document available at http://southport.jpl.nasa.gov/desc/imagingradarv3.html

Gamache, M. (2004a): Free and low cost datasets for international mountain cartography. Documentation for the Alpine Mapping Guild, 42 pp. Available at http://www.icc.es/workshop/abstracts/ica_paper_web3.pdf

Gamache, M. (2004b): Various documents including evaluations of CIAT's void-filling methodology, available at: http://www.icc.es/workshop/presentacions/s5/1_martin/presentation.pdf and http://www.alpinemapguild.com/ICA_CMC2004/Gamache final web.pdf

Geoscience Australia (2007): Global Map Australia 1M 2001, Drainage layer (Water courses). Geoscience Australia, Australian Government, Canberra, Australia. Available at http://www.ga.gov.au/topographic-mapping/digital-topographic-data.html

Gesch, D.B., Verdin, K.L., Greenlee, S.K. (1999): New land surface digital elevation model covers the Earth. Eos. Transactions, American Geophysical Union 80(6): 69-70.

Hutchinson, M. (1988): Calculation of hydrologically sound digital elevation models. Third International Symposium on Spatial Data Handling, Columbus, Ohio, International Geographical Union.

Hutchinson, M. (1989): A new procedure for gridding elevation and stream line data with automatic removal of spurious pits. Journal of Hydrology 106: 211-232.

Jarvis, A., J. Rubiano, A. Nelson, A. Farrow, Mulligan M. (2004): Practical use of SRTM data in the tropics: comparisons with digital elevation models generated from cartographic data. Working Document no. 198. Cali, International Centre for Tropical Agriculture (CIAT): 32 pp.

Jarvis, A. (2004): SRTM Data Processing Methodology. Available at http://srtm.csi.cgiar.org/SRTMdataProcessingMethodology.asp

Jarvis, A., Reuter, H.I., Nelson, A., Guevara, E. (2008): Hole-filled SRTM for the globe, Version 4. Available from the CGIAR-CSI SRTM 90m Database (http://srtm.csi.cgiar.org).

Lehner, B., Döll, P. (2004): Development and validation of a global database of lakes, reservoirs and wetlands. Journal of Hydrology 296(1-4): 1-22. Data available at http://www.wwfus.org/science/data/globallakes.cfm

NASA (2003): SRTM Data Editing Rules, Version 2.0. 23 pp. Available at ftp://e0srp01u.ecs.nasa.gov/srtm/version2/Documentation/SRTM_edit_rules.pdf or at http://edc.usgs.gov/products/elevation/swbdedit.doc

NASA/JPL (2005): SRTM Topography (SRTM documentation). 8pp. Available at ftp://e0srp01u.ecs.nasa.gov/srtm/version2/Documentation/SRTM_Topo.pdf

NASA/NGA (2003): SRTM Water Body Data Product Specific Guidance, Version 2.0. 4pp. Available at http://edc.usgs.gov/products/elevation/swbdguide.doc

National Atlas of the United States (2005): Two-Million-Scale Streams and Waterbodies. Available at http://www.nationalatlas.gov/mld/hydrogm.html

Natural Resources Canada (2006): Atlas of Canada 1,000,000 National Frameworks Data, Hydrology (v. 6.0). Government of Canada, Natural Resources Canada, Earth Sciences Sector, Mapping Information Branch, Sherbrooke, Quebec, Canada. Available at http://geogratis.cgdi.gc.ca/

USGS – U.S. Geological Survey (2000): HYDRO1k Elevation Derivative Database. USGS EROS Data Center, Sioux Falls, SD. Available at http://edc.usgs.gov/products/elevation/gtopo30/hydro/