

HYDRO1k Documentation ReadMe

Table of Contents

1.0 Introduction.....	1
2.0. Data Layers	2
3.0. Data Set Development	2
3.1 Data Processing Procedures.....	2
3.1.1. Project the DEM	2
3.1.2. Identify Natural Sink Features.....	3
3.1.3. Filling the DEM	3
3.1.4. Verification of the DEM	3
3.2. Generation of Derivative Raster Data Sets	4
3.2.1. Aspect.....	4
3.2.2. Flow Directions	4
3.2.3. Flow Accumulations.....	4
3.2.4. Slope	5
3.2.5. Compound Topographic Index.....	5
3.3 Generation of Derivative Vector Data Sets	5
3.3.1. Drainage Basin Boundaries.....	5
3.3.2. Stream Lines.....	6
4.0. Data Formats	6
4.1. Vector Data Formats.....	6
4.2. Raster Data Formats.....	6
4.2.1. Image File (.bil).....	7
4.2.2. Header File (.hdr)	7
4.2.3. World File (.blw file)	8
4.2.4. Statistics File (.stx file).....	8
5.0. Data Distribution.....	8
6.0. Notes and Hints for HYDRO1k Users.....	9
7.0. Summary	9
8.0 References.....	10
9.0 Disclaimers	10

1.0 Introduction

HYDRO1k, developed at the U.S. Geological Survey's (USGS) [EROS Data Center](#), is a geographic database providing comprehensive and consistent global coverage of topographically derived data sets. Developed from the USGS' recently released 30 arc-second digital elevation model (DEM) of the world ([GTOPO30](#)), HYDRO1k provides a standard suite of geo-referenced data sets (at a resolution of 1 km) that will be of value for all users who need to organize, evaluate, or process hydrologic information on a continental scale.

Constructive comments from users of the HYDRO1k data sets are welcomed. Please send your comments to lta.usgs.gov.

2.0. Data Layers

The HYDRO1k data sets are being developed on a continent by continent basis, for all landmasses of the globe with the exception of Antarctica and Greenland. The HYDRO1k package provides, for each continent, a suite of six raster and two vector data sets. These data sets cover many of the common derivative products used in hydrologic analysis. The raster data sets are the hydrologically correct DEM, derived flow directions, flow accumulations, slope, aspect, and a compound topographic (wetness) index. The derived streamlines and basins are distributed as vector data sets.

3.0. Data Set Development

The HYDRO1k data sets are the result of the cooperative project at the U.S. Geological Survey's (U.S.G.S.) EROS Center. The goal of the project is the development of a globally consistent hydrologic derivative data set. The effort has been led by U.S.G.S. scientists in collaboration with the United Nations Environment Programme/Global Resource Information Database ([UNEP/GRID](#)) located in Sioux Falls, South Dakota.

Development of the HYDRO1k database was made possible by the completion of the 30 arc-second digital elevation model at EROS in 1996, entitled GTOPO30. This data set, with its nominal cell size of 1 km, has been and will continue to be applied by many scientists and researchers to hydrologic and land form studies. Inevitably, these studies require development, at a minimum, of a standard suite of derivative products. In the past, users would obtain the DEM data, process the data, extract the derivative information, use the derived products in their studies and, perhaps, share the derived information with others. In an attempt to reduce repetition of these procedures by every user of the data set, the HYDRO1k data base aims to provide these standard products, developed in a consistent fashion for the entire globe and make them available for the entire user community.

3.1 Data Processing Procedures

The basis of all of the data layers available in the HYDRO1k database is the hydrologically correct DEM. This DEM is, of course, based on the GTOPO30 data set. However, to ensure that the DEM is able to reproduce the correct movement of water across its surface, the DEM is processed to remove elevation anomalies that can interfere with hydrologically correct flow. The procedures followed in development of this DEM are iterative. Some of the techniques used in the DEM development are documented in Danielson (1997).

3.1.1. Project the DEM

In order to properly perform area calculations on the DEM, the data are projected into an equal area projection. The Lambert Azimuthal Equal Area projection was selected for this database. (Steinwand et al, 1995). The cell size for all continents is 1,000 meters and the radius of the sphere of influence is 6,370,997 meters for all continents except Africa. Africa has a radius of the sphere of influence of 6,378,137 meters. Projection parameters that vary by continent are given in the following table. Other geo-referencing information is available in the projection file that is included with each continental data set.

Continent	Longitude of Origin	Latitude of Origin
Africa	20° 00' 00"E	5° 00' 00"N
Asia	100° 00' 00"E	45° 00' 00"N
Australasia	135° 00' 00"E	15° 00' 00"S
Europe	20° 00' 00"E	55° 00' 00"N
North America	100° 00' 00"W	45° 00' 00"N
South America	60° 00' 00"W	15° 00' 00"S

3.1.2. Identify Natural Sink Features

All continents contain some closed basins; drainage basins with no natural outlet to the sea. In processing the HYDRO1k DEM to replicate natural flow patterns, techniques were developed to (1). identify which sink features in the DEM are, indeed, natural features and (2). preserve these sink features during the processing. Identification of the natural sinks in the DEM was begun by creating a "sink layer" containing all sink features contained in the projected GTOPO30 DEM. This sink layer was then thresholded to extract only sinks with a surface area greater than a specified minimum. This was used as a "first-cut" on identification of the natural sink features.

3.1.3. Filling the DEM

To allow filling of the DEM using standard GIS techniques while still maintaining the sinks identified in step 3.1.2., the identified sinks are "seeded" by placing a NODATA point at the bottom of each sink. Since the standard GIS implementation of the hydrologic filling technique allows flow only off the edge of the DEM or to NODATA points, this procedure "tricks" the GIS into letting water flow to the sink. All spurious sinks, those not identified as potential natural features in 3.1.2, are removed.

3.1.4. Verification of the DEM

Following filling of the DEM, initial streamline and basin data sets are generated for use in the verification of the DEM. Flow direction and flow accumulation grids are generated and the vector stream lines and basin boundaries are produced. The streamlines and basins thus derived are compared against existing digital data. In most cases, the Digital Chart of the World (DCW) drainage cover was used for comparison (Defense Mapping Agency, 1992; Danko, 1992). However, all available map sources were used. Comparison of the generated streamlines with mapped hydrography allows identification of essentially two types of errors in the DEM:

(1). Errors of omission or inclusion of natural sink features. Examination of mapped hydrography often serves to identify whether or not the first pass identification of the natural sinks features was adequate. In the case of an error of omission, the newly identified sink feature is "seeded" in the DEM and in the case of inclusion, the "seeded" sink is removed ("unseeded").

(2). Errors in the DEM which prevent proper flow across its surface. These errors can be caused by the DEM generation or resampling techniques or can simply be caused by the 1-km horizontal or the 1-m vertical resolution of the DEM. Comparison with mapped hydrography serves to identify locations where the generated streamlines or basin boundary deviate. If the difference between the two sources of information proves to be the DEM, editing of the DEM is done to guarantee that flow progresses in the required direction. These type of DEM edits usually involve only small changes in the elevation of one or two pixels.

The procedures in 3.1.3. and 3.1.4. are repeated until the DEM is able to produce streamlines and basins that adequately match mapped hydrography.

3.2. Generation of Derivative Raster Data Sets

Following generation of the hydrologically correct DEM, the final versions of the additional derivative data layers are produced. Along with the hydrologically correct DEM, the following five raster data layers are developed using standard GIS techniques. All derivative raster data layers were produced using ARC/INFO's GRID module (ESRI, 1992).

3.2.1. Aspect

The aspect data set describes the direction of maximum rate of change in the elevations between each cell and its eight neighbors. It can essentially be thought of as the slope direction. It is measured in positive integer degrees from 0 to 360, measured clockwise from north. Aspects of cells of zero slope (flat areas) are assigned values of -1.

3.2.2. Flow Directions

The flow direction data layer defines the direction of flow from each cell in the DEM to its steepest down-slope neighbor. Values of flow direction vary from 1 to 255. Defined flow directions follow the convention adopted by ARC/INFO's flow direction implementation:

32	64	128
16		1
8	4	2

Cells with undefined direction of flow represent sinks and have flow directions that are simple combinations of its neighbors' flow direction values.

3.2.3. Flow Accumulations

The flow accumulation data layer defines the amount of upstream area draining into each cell. It is essentially a measure of the upstream catchment area. The flow direction layer is used to define which cells flow into the target cell. Since the cell size of the HYDRO1k data set is 1 km, the flow accumulation value translates directly into drainage areas in square kilometers. Values range from 0 at topographic highs to very large numbers (on the order of millions of cells) at the mouths of large rivers.

3.2.4. Slope

The slope data layer describes the maximum change in the elevations between each cell and its eight neighbors. The slope is expressed in integer degrees of slope between 0 and 90.

3.2.5. Compound Topographic Index

The Compound Topographic Index (CTI), commonly referred to as the Wetness Index, is a function of the upstream contributing area and the slope of the landscape. The implementation used in the HYDRO1k data set is based on Moore et al (1991). The CTI is calculated using the flow accumulation (FA) layer along with the slope as:

$$CTI = \ln (FA / \tan (\text{slope}))$$

In areas of no slope, a CTI value is obtained by substituting a slope of 0.001. This value is smaller than the smallest slope obtainable from a 1000 m data set with a 1 m vertical resolution.

3.3 Generation of Derivative Vector Data Sets

The stream line and basin data in the HYDRO1k data set are distributed as vector layers.

3.3.1. Drainage Basin Boundaries

The drainage basins distributed with the HYDRO1k data set are derived using the vector streamlines along with the flow direction layer. The basins are seeded following procedures first articulated by Otto Pfafstetter, a Brazilian engineer, and adapted for use in the HYDRO1k data set (Verdin, 1997) . Each polygon in the basin data set has been tagged with a Pfafstetter code uniquely identifying each sub-basin. The six-digit Pfafstetter code assigned to each basin carries basin linkage information. This permits determination of basin interconnectedness through simple examination of the Pfafstetter code.

The drainage basin polygons are attributed with the following attributes:

Level1 to Level6 = Pfafstetter units of each polygon

Slope_mean = Mean value of the slopes within the subbasin (degree)

Slope_stdev = Standard deviation of the slopes within the subbasin (degree)

Aspect_mean = Mean value of the aspects within the subbasin (degree from N)

Aspect_stdev = Standard deviation of the aspects within the subbasin (degree from N)

Dem_mean = Mean elevation value within the subbasin (m)

Dem_stdev = Standard deviation of the elevations within the subbasin (m)

3.3.2. Stream Lines

The stream line data layer distributed with the HYDRO1k data set is derived from the flow accumulation and flow direction layers. Cells with upstream drainage areas greater than 1000 km² are selected from the flow accumulation layer and processed through the STREAMLINK function. The resulting links are attributed with the maximum flow accumulation occurring within that link and the result is vectorized using the STREAMLINE function. These procedures result in a vector data layer of streamlines with each segment of stream attributed with the upstream contributing drainage area. The vector streamlines are attributed with the following fields:

Flowacc = The maximum flow accumulation value of the stream segment. This value corresponds directly with the upstream watershed contributing area. (10⁻³ km²)

Pf_type = The Pfafstetter level at which the stream segment is considered "main stem".

Level1 to Level6 = The Pfafstetter units in which the stream segment lie.

Frmelevation = The elevation value of the stream segment's from-node (m)

Toelevation = The elevation value of the stream segment's to-node (m)

Strorder = Strahler stream order of the segment

Gradient = Gradient of the stream segment calculated as the difference of the from and to-node elevations divided by the length of the segment

Frmup_flowlen = The upstream flowlength from the from-node. Calculated using ARC/INFO's FLOWLENGTH function, it is the longest path from the from-node to the drainage basin divide. (m)

Toup_flowlen = The upstream flowlength from the to-node. (m)

Frndn_flowlen = The downstream flowlength from the from-node. Again from ARC/INFO's FLOWLENGTH function, it is the length from the from-node to the ocean or a terminal sink. (m)

Todn_flowlen = The downstream flowlength from the to-node. (m)

4.0. Data Formats

4.1. Vector Data Formats

The vector data sets, stream lines and basins, distributed with HYDRO1k are being made available in a ARC/INFO Export Format (.E00 extension).

4.2. Raster Data Formats

The six raster data layers distributed for each continent are being distributed as simple binary raster data. Each raster data layer is provided as four files, with the extension of each file defining the file type.

File Extension	File Type
.bil	Raster Data File
.hdr	Header File
.blw	World File
.stx	Statistics File

4.2.1. Image File (.bil)

The raster data for each layer are provided as signed integer data in a simple binary raster format. All the data layers are 16-bit data with the exception of the flow accumulation layer, which, due to the range of values needed, is 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.).

4.2.2. 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 .hdr 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 M = Motorola byte order (most significant byte first)
LAYOUT:	organization of the bands in the file BIL: band interleaved by line (note: the raster layers are all single band images)
NROWS:	number of rows in the image
NCOLS:	number of columns in the image
NBANDS:	number of spectral bands in the image (1)
NBITS:	number of bits per pixel (16 or 32)
BANDROWBYTES:	number of bytes per band per row (twice the number of columns for a 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 a single band 16-bit image; four-times for the 32-bit image)
BANDGAPBYTES:	the number of bytes between bands in a BSQ format image (0)

4.2.3. World File (.blw file)

The world file is an ASCII text file containing coordinate information. It is used by some packages for geo-referencing of image data.

XDIM:	X-dimension of a pixel (1000)
Rotation term:	Always zero
Rotation term:	Always zero
Negative YDIM:	Negative Y-dimension of a pixel (-1000)
XMIN:	X-location of center of upper-left pixel (projected meters)
YMAX:	Y-location of center of upper-left pixel (projected meters)

4.2.4. Statistics File (.stx file)

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

5.0. Data Distribution

HYDRO1k data for each continent are distributed electronically as tar files. The data files are identified by the two-digit continental identifier according to the following scheme:

Two-digit Identifier	Continent
AF	Africa
AS	Asia
AU	Australasia
EU	Europe
NA	North America
SA	South America

Users have the option of obtaining the entire HYDRO1k data set for a continent (all eight data layers) or selectively choosing layers for download. In either case, the data are distributed as tar

files. In the case of raster data sets, the .bil files have been compressed with the gzip function before creation of the tar file. The vector data export files have been compressed (gzipped) as well prior to creation of the tar file. As an example of the naming convention used, the North American data sets that are available are:

Na.tar	A tar file containing all the North American data layers along with README
Na_asp.tar	Tar file containing the aspect data layer (compressed bil file, three ancillary files and README)
Na_bas.tar	Vector basin data layer in compressed ARC/INFO Export format along with README
Na_cti.tar	Tar file with CTI data layer (compressed bil file, three ancillary files and README)
Na_dem.tar	Tar file with DEM data layer (compressed bil file, three ancillary files and README)
Na_fd.tar	Tar file with flow direction data layer (compressed bil file, three ancillary files and README)
Na_fa.tar	Tar file with flow accumulation data layer (compressed bil file, three ancillary files and README)
Na_slope.tar	Tar file with slope data layer (compressed bil file, three ancillary files and README)
Na_str.tar	Vector streams data layer in compressed ARC/INFO Export format along with README

6.0. Notes and Hints for HYDRO1k Users

Because the image (.bil) data are stored in a 16-bit binary format, users must be aware of how the bytes are addressed on their computers. The data are provided in Motorola byte order, which stores the most significant byte first ("big endian"). Systems such as Sun SPARC and Silicon Graphics workstations use the Motorola byte order. The Intel byte order, which stores the least significant byte first ("little endian"), is used on DEC Alpha systems and most PCs. Users with systems that address bytes in the Intel 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.

Users of ARC/INFO or ArcView can display the image data directly. However, if a user needs access to the actual pixel values for analysis in ARC/INFO the image must be converted to an ARC/INFO grid with the command IMAGEGRID. IMAGEGRID does not support conversion of signed image data, therefore the negative 16-bit image values will not be interpreted correctly. After running IMAGEGRID, an easy fix can be accomplished using the following formula in GRID:

```
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 ocean mask values in the grid could then be set to NODATA with the SETNULL function.

7.0. Summary

The HYDRO1k data set provides many of the derivative products useful in earth science applications. The hydrologically correct DEM and ancillary data layers are useful in studies of earth systems including watershed analysis, landform studies and global change scenarios. Development of a standard set of data layers minimizes duplication of effort and will provide consistent global coverage.

8.0 References

Danielson, J.J., 1996. Delineation of drainage basins from 1 km African digital elevation data. In: Pecora Thirteen, Human Interactions with the Environment - Perspectives from Space, Sioux Falls, South Dakota, August 20-22, 1996.

Danko, D.M., 1992. The digital chart of the world. *GeoInfo Systems*, 2:29-36.

Defense Mapping Agency, 1992, *Development of the Digital Chart of the World*: Washington, D.C., U.S. Government Printing Office

ESRI, 1992, "Cell based modeling with GRID", ESRI, Inc., Redlands, California.

Moore, I.D., R.B. Grayson and A.R. Ladson, 1991, Digital Terrain Modelling: A Review of Hydrological, Geomorphological and Biological Applications. In: *Hydrological Processes An International Journal*, January - March, 1991, pp. 3 - 30.

Steinwand, D.R., Hutchinson, J.A., and Snyder, J.P. ,1995, Map projections for global and continental data sets and an analysis of pixel distortion caused by reprojection: *Photogrammetric Engineering and Remote Sensing*, v. 61, p. 1,487-1,497.

Verdin, K.L., and Greenlee, S.K., 1996. Development of continental scale digital elevation models and extraction of hydrographic features. In: *Proceedings, Third International Conference/Workshop on Integrating GIS and Environmental Modeling*, Santa Fe, New Mexico, January 21-26, 1996. National Center for Geographic Information and Analysis, Santa Barbara, California.

Verdin, K.L., A System for Topologically Coding Global Drainage Basins and Stream Networks. In: *Proceedings, 17th Annual ESRI Users Conference*, San Diego, California, July 1997.

9.0 Disclaimers

Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government. Please note that some U.S. Geological Survey (USGS) information contained in this data set and documentation may be preliminary in nature and presented prior to final review and approval by the Director of the USGS. This information is provided with the understanding that it is not guaranteed to be correct or complete and conclusions drawn from such information are the sole responsibility of the user.
