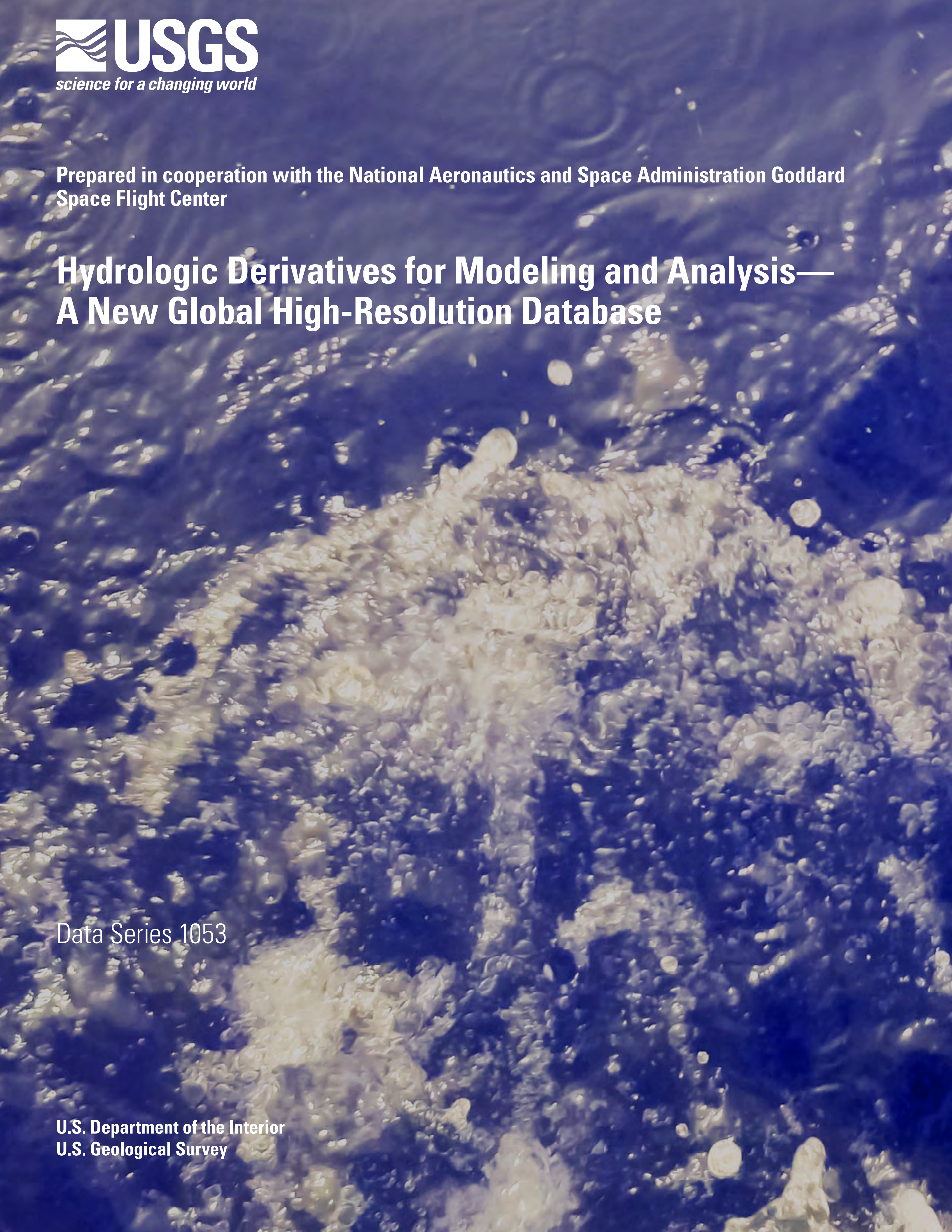


USGS
science for a changing world

Prepared in cooperation with the National Aeronautics and Space Administration Goddard Space Flight Center

Hydrologic Derivatives for Modeling and Analysis—A New Global High-Resolution Database

Data Series 1053

U.S. Department of the Interior
U.S. Geological Survey

Hydrologic Derivatives for Modeling and Analysis—A New Global High-Resolution Database

By Kristine L. Verdin

Prepared in cooperation with the National Aeronautics and Space Administration
Goddard Space Flight Center

Data Series 1053

U.S. Department of the Interior
U.S. Geological Survey

U.S. Department of the Interior
RYAN K. ZINKE, Secretary

U.S. Geological Survey
William H. Werkheiser, Acting Director

U.S. Geological Survey, Reston, Virginia: 2017

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment—visit https://www.usgs.gov or call 1–888–ASK–USGS.

For an overview of USGS information products, including maps, imagery, and publications, visit https://store.usgs.gov.

Suggested citation:
Verdin, K.L., 2017, Hydrologic Derivatives for Modeling and Analysis—A new global high-resolution database: U.S. Geological Survey Data Series 1053, 16 p., https://doi.org/10.3133/ds1053.

ISSN 2327-638X (online)

Contents

Figures

Tables

Conversion Factors

International System of Units to U.S. customary units

Multiply	By	To obtain
	Length	
meter (m)	3.281	foot (ft)
kilometer (km)	0.6214	mile (mi)
	Area	
square kilometer (km^2)	0.3861	square mile (mi^2)

Datum

Vertical coordinate information is referenced to the Earth Gravitational Model 1996 (EGM96) geoid.

Horizontal coordinate information is referenced to the World Geodetic System (WGS84).

Abbreviations

CTI	compound topographic index
DEM	digital elevation model
GIS	geographic information system
GMTED2010	Global Multi-resolution Terrain Elevation Data 2010
HDMA	Hydrologic Derivatives for Modeling and Analysis
HydroSHEDS	Hydrological data and maps based on SHuttle Elevation Derivatives at multiple Scales
SRTM	Shuttle Radar Topography Mission
USGS	U.S. Geological Survey

Hydrologic Derivatives for Modeling and Analysis—A New Global High-Resolution Database

By Kristine L. Verdin

Abstract

The U.S. Geological Survey has developed a new global high-resolution hydrologic derivative database. Loosely modeled on the HYDRO1k database, this new database, entitled Hydrologic Derivatives for Modeling and Analysis, provides comprehensive and consistent global coverage of topographically derived raster layers (digital elevation model data, flow direction, flow accumulation, slope, and compound topographic index) and vector layers (streams and catchment boundaries). The coverage of the data is global, and the underlying digital elevation model is a hybrid of three datasets: HydroSHEDS (Hydrological data and maps based on SHuttle Elevation Derivatives at multiple Scales), GMTED2010 (Global Multi-resolution Terrain Elevation Data 2010), and the SRTM (Shuttle Radar Topography Mission). For most of the globe south of 60°N., the raster resolution of the data is 3 arc-seconds, corresponding to the resolution of the SRTM. For the areas north of 60°N., the resolution is 7.5 arc-seconds (the highest resolution of the GMTED2010 dataset) except for Greenland, where the resolution is 30 arc-seconds. The streams and catchments are attributed with Pfafstetter codes, based on a hierarchical numbering system, that carry important topological information. This database is appropriate for use in continental-scale modeling efforts. The work described in this report was conducted by the U.S. Geological Survey in cooperation with the National Aeronautics and Space Administration Goddard Space Flight Center.

Introduction

The extraction of hydrologic information from digital elevation models (DEMs) has a long history (Jenson and Domingue, 1988; Tarboton, 1997; Lehner and others, 2008) and a well-established scientific utility (Verdin and Verdin, 1999; Kumar and others, 2000; Hall and others, 2004; Verdin and others, 2007; Verdin and Worstell, 2008; Beighley and others, 2009). On a global level, however, the quality and consistency of the extracted information has been limited by the available underlying DEMs. To date (2017), the only globally consistent hydrologic derivative database with a full suite of derivatives has been the HYDRO1k database (Verdin and Jenson, 1996). The HYDRO1k database consists of a suite of six raster and two vector data layers, including common terrain derivatives used in hydrologic analysis. The HYDRO1k raster data layers are the DEM and the hydrologic derivatives of flow direction, flow accumulation, slope, aspect, and the compound topographic index (CTI) (Moore and others, 1991). The vector data layers are DEM-derived streamlines and catchment polygons, topologically coded according to the Pfafstetter coding system (Verdin and Verdin, 1999). The HYDRO1k database was developed from the Global 30 Arc-Second Elevation (GTOPO30) DEM (Gesch and others, 1999), which has a 30-arc-second resolution (about 1 kilometer at the equator). With the completions of the Shuttle Radar Topography Mission (SRTM) DEM (Farr and Kobrick, 2000) and the Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010) dataset (Danielson and Gesch, 2011), production of a high-resolution hydrologic derivative database for the entire globe became possible.

The HydroSHEDS (Hydrological data and maps based on SHuttle Elevation Derivatives at multiple Scales) dataset (Lehner and others, 2008) provides two of the same layers as the HYDRO1k at the higher resolution of the SRTM DEM (3 arc-seconds or about 90 meters [m] at the equator) for most of the Earth. The HydroSHEDS dataset includes a 3-arc-second hydrologically conditioned DEM (one for which spurious sinks have been filled and drainage has been enforced along mapped hydrographic networks) and a flow direction grid for most land areas of the globe from 60°S. to 60°N. at the 3-arc-second resolution of the SRTM data, but it does not include equivalents for the other HYDRO1k derivatives. The HydroSHEDS dataset also lacks coverage of all landmasses north of 60°N. and a number of island areas (including Hawaii). The processing done to develop the hydrologically conditioned DEM for the HydroSHEDS dataset is described in detail in the HydroSHEDS technical documentation (Lehner and others, 2006).

This report describes the Hydrologic Derivatives for Modeling and Analysis (HDMA) database, a compilation of DEM data drawn from portions of the HydroSHEDS dataset, a void-filled SRTM DEM (International Centre for Tropical Agriculture, 2004), and the GMTED2010 dataset used to produce, first, a globally consistent DEM and, second, the same derivatives as included in the HYDRO1k database (with the exception of aspect). The raster data were produced at 3-arc-second resolution for most areas south of 60°N., 7.5-arc-second resolution for most areas north of 60°N., and 30-arc-second resolution for Greenland. The derived streams and catchments are globally seamless and have attributes following the topological Pfafstetter coding system.

Data

The primary DEM used in the development of the HDMA database was from the HydroSHEDS dataset (Lehner and others, 2008). The HydroSHEDS DEM was derived from the SRTM DEM (Farr and Kobrick, 2000) by using several ancillary layers to hydrologically condition the DEM (filling spurious sinks and enforcing drainage along previously mapped hydrographic networks). The HydroSHEDS DEM covers most land masses of the globe south of 60°N. (except Antarctica) and has a resolution of 3 arc-seconds.

For the land masses north of 60°N., the GMTED2010 dataset (Danielson and Gesch, 2011) was used. The GMTED2010 dataset itself is a hybrid of the best global DEM data available at the time of its creation and has a resolution of 7.5 arc-seconds, with the exception of Greenland where the resolution is 30 arc-seconds. Because derivation of drainage basins was one of the main requirements for the HDMA database, the HydroSHEDS and GMTED2010 data were joined along drainage divide lines. For this reason, some HDMA areas slightly south of 60°N. were developed using GMTED2010 elevation data though SRTM data were available for these areas.

The third DEM data source was the void-filled SRTM data available from the CGIAR[1] Consortium for Spatial Information (International Centre for Tropical Agriculture, 2004). These data were used to fill in gaps in the HydroSHEDS database south of 60°N. In general, these gaps were islands that had not been included in the HydroSHEDS continental data-development effort. Figure 1 depicts the spatial distribution of DEM sources used in the HDMA development effort.

Data-Layer Development

The landmasses of six continents (all but Antarctica) were subdivided into processing units based on hydrologic divides to accommodate data processing limitations. The processing units were modeled on the Pfafstetter Level 1 (Verdin and Verdin, 1999) basin subdivisions found in the HYDRO1k dataset (Verdin and Jenson, 1996). Each unit is labeled with the first two digits of the Pfafstetter code (the leading continental digit and the Level 1 Pfafstetter ID), as shown in figure 2. The leading digits for continental identification in the Pfafstetter code are taken from Verdin and Verdin (1999) and are based on the relative sizes of the continents (table 1). Additional subdivision of units was necessary to facilitate processing of very large units (indicated by a hyphenated suffix) and when the units crossed the boundary between the HydroSHEDS and GMTED2010 source data. Although the resolution of the DEM for Greenland was too low to adequately create vector streams and catchments for the landmass, raster layers were created, and Greenland was assigned the leading digit “7” for use in the database.

[1]Formerly the Consultative Group on International Agricultural Research.

Derivation of the initial data layers (DEM, flow direction, flow accumulation, slope, and CTI) was at the full resolution of the data: 3-arc-second resolution for the HydroSHEDS and SRTM data, 7.5-arc-second resolution for most of the GMTED2010 data, and 30-arc-second resolution for the Greenland GMTED2010 data.

Development of a Hydrologically Conditioned DEM

Though the HydroSHEDS DEM was already hydrologically conditioned, the GMTED2010 and void-filled SRTM DEMs were not. Both of these datasets were conditioned using standard geographic information system (GIS) processing tools. ArcGIS (Esri, 2016) tools were used to fill the DEM by removing sinks. Edge-matching between the HydroSHEDS and GMTED2010 datasets was achieved by enforcing the boundary of the higher-resolution HydroSHEDS DEM as a “wall” when conditioning the GMTED2010 DEM. Edge-matching in this manner ensured near-seamless agreement between the two datasets; gaps between the two resolutions of the data are less than the size of one pixel. The GMTED2010 data were processed at 7.5-arc-second resolution, and the HydroSHEDS and void-filled SRTM data were kept at the full resolution of 3 arc-seconds. The DEM of Greenland (derived from the GMTED2010 dataset) was included in the database, but its resolution is lower (30 arc-seconds) and quality poorer than that of the other data sources.

Development of the Flow Direction Grid

As with the hydrologically conditioned DEM, the HydroSHEDS database already contained a flow direction grid (raster) that was incorporated into the HDMA. The flow direction grids for the areas covered by the GMTED2010 and the SRTM DEMs were produced as part of the HDMA hydrologic conditioning process. Standard GIS processing techniques were used to fill the sinks in the DEM and extract the flow direction grid by using an eight-direction flow-direction algorithm (Jenson and Domingue, 1988).

Development of the Flow Accumulation Grid

The flow accumulation grid was derived for all areas, irrespective of the source of the underlying DEM data (HydroSHEDS, SRTM, or GMTED2010). The flow accumulation grid describes, for any location on the landscape, the number of upstream raster cells (sometimes called pixels) that contribute flow to that particular location. In equal-area projected DEMs, the flow accumulation grid can be simply translated into upstream area (in square kilometers [km^2]) because each pixel has the same areal dimension. Because the DEM data used in this effort are in a geographic spatial reference system, and because the pixel size varies with location on the globe,

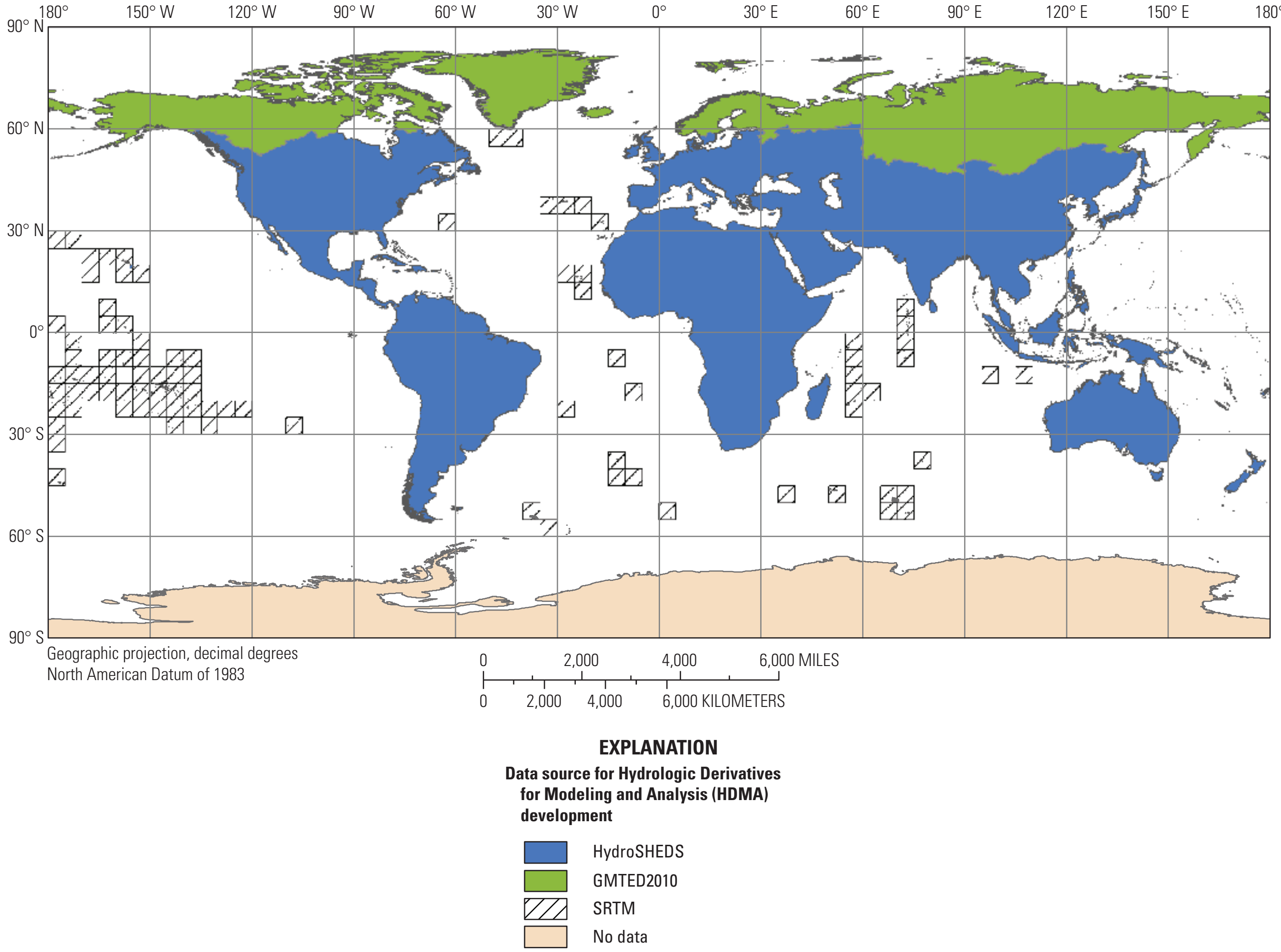


Figure 1. Digital elevation model data sources used in the development of the Hydrologic Derivatives for Modeling and Analysis database. (HydroSHEDS, Hydrological data and maps based on SHuttle Elevation Derivatives at multiple Scales [Lehner and others, 2008]; GMTED2010, Global Multi-resolution Terrain Elevation Data 2010 [Danielson and Gesch, 2011]; STRM, Shuttle Radar Topography Mission [International Centre for Tropical Agriculture, 2004])

a "weight grid" developed from tables in Robinson and Sale (1969) was used to calculate the flow accumulation grid in units of square kilometers (Verdin and Worstell, 2008). The weight grid defines the area of each pixel as a function of latitude and longitude and was used as the "input weight raster" when running the Flow Accumulation tool within ArcGIS (Esri, 2016). The distances in the x direction vary considerably, with one degree of longitude at the equator equaling 111,321 m, whereas at lat 59°N. or 59°S., one degree of longitude equals only 57,478 m. The variation of pixel length in the y direction is less dramatic, varying from 111,567 m at the equator to 111,406 m at lat 59°N. or 59°S.

Development of the Slope Layer

Slopes were calculated using the RiverTools software (Rivix, 2012). RiverTools accommodates large geographic datasets, and latitude-specific length and area adjustments are automatically computed. Slope was calculated as "rise over run" (or the tangent of the slope angle) by using a partial quartic surface fitted locally by using the method described by Zevenbergen and Thorne (1987).

Development of the Compound Topographic Index Layer

The compound topographic index (CTI) was calculated from the HDMA-derived slope and the flow direction and flow accumulation layers. The CTI is shown in equation 1 (Moore and others, 1991):

$$CTI = \ln \frac{A_s}{\tan \beta}. \tag{1}$$

where

A_s is the specific catchment area and

β is the slope angle.

The tangent of the slope angle (tan β) was derived as one of the base layers. The specific catchment area is defined as the upslope area per unit width of contour. The unit width of contour is taken as the width of the flow path, which is derived as a function of the flow accumulation grid and the flow direction grid. Because pixel sizes in both the x and y directions vary with location on the globe, the width of the flow path is a function of both the direction of flow (fig. 3) and the latitude and longitude of the location. The specific catchment area was obtained by creating a grid describing the width of the flow path and then dividing the flow accumulation grid by the width-of-the-flow-path grid. The resulting CTI grid was derived using a map algebra equivalent of equation 1 once the specific catchment area grid was developed.

Development of the Streams and Catchments Data Layers

Procedures for Pfafstetter codification required the creation of a vector stream network with corresponding catchments. A raster cell was designated as being part of a stream if its upslope contributing area exceeded 250 km^2. These stream cells were grouped into links based on confluence points, and links were assigned unique identifiers. The flow direction grid was used to identify all cells upslope of each link to define the catchment associated with that link. The stream and catchments rasters were then vectorized. For the globe, about 279,000 stream segments and about 295,000 catchments were delineated.

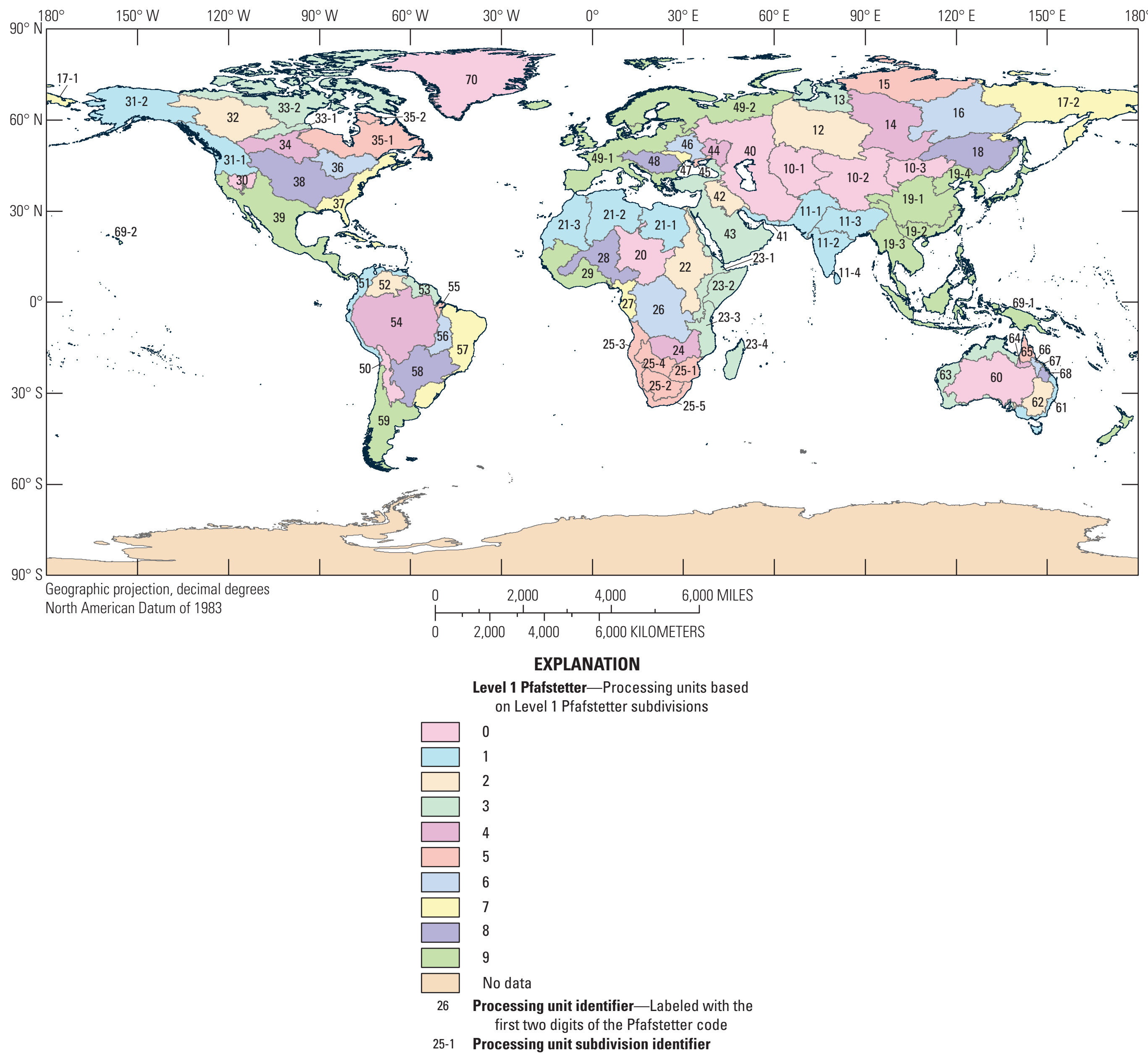


Figure 2. Processing units, based on HYDRO1k Level 1 Pfafstetter units. Leading continental digits are after Verdin and Verdin (1999). Trailing digits (hyphenated) indicate further subdivision to break up overly large processing units (for example, processing unit 11 was subdivided into units 11-1, 11-2, and 11-3).

Table 1. Leading digit for continental identification.

Leading digit for continental identification	Continental area
1	Asia
2	Africa
3	North America
4	Europe
5	South America
6	Australasia
7	Greenland[1]

[1]Greenland was included as Pfafstetter Unit 7, but the landmass was not subdivided into smaller Pfafstetter units.

Pfafstetter Codification

The streams and catchments were assigned Pfafstetter codes in a vector/polygon processing environment. Streams were navigated and numbered using Pfafstetter topological rules (Verdin and Verdin, 1999) and the Pfafstetter code transferred to the catchment corresponding to each stream. Interbasins (catchments without a corresponding stream segment) were assigned an appropriate Pfafstetter code with all odd digits (for example, interbasin 553100000000 on the South American coastline). The Pfafstetter codification was done on a continental basis; the resulting Pfafstetter codes are 12 digits in length, including the leading continental digit shown in table 1.

Use of Pfafstetter Codes for Network Navigation

The Pfafstetter codification system was founded on concepts first articulated by the late Otto Pfafstetter, an engineer with the Departamento Nacional de Obras de Saneamento, a civil works agency of the Federal government of Brazil. A full discussion of the Pfafstetter method is included in Verdin and Verdin (1999). Discussed here is the application of the Pfafstetter codes to the HDMA global database and their use in hydrographic network navigation.

To begin the continental subdivision into Pfafstetter units, the four rivers with the largest drainage areas that drain to the coast are identified for each continent. The basins for these rivers are assigned the even digits (2, 4, 6, and 8). The coastal interbasin areas between the mouths of these basins are assigned the odd digits (1, 3, 5, 7, and 9). If there is a significant closed basin on the continent, it is assigned the digit zero (0). The four rivers with the largest drainage basins in each of the six continental land masses subdivided in this work are identified in table 2.

Further subdivision of each drainage basin and addition of Pfafstetter digits follow the techniques outlined in Verdin and Verdin (1999). The four largest tributaries within each subdivision are identified and assigned even digits, and the interbasins receive the odd digits; the headwaters always receive the digit 9. To illustrate the iterative subdivision of a basin and

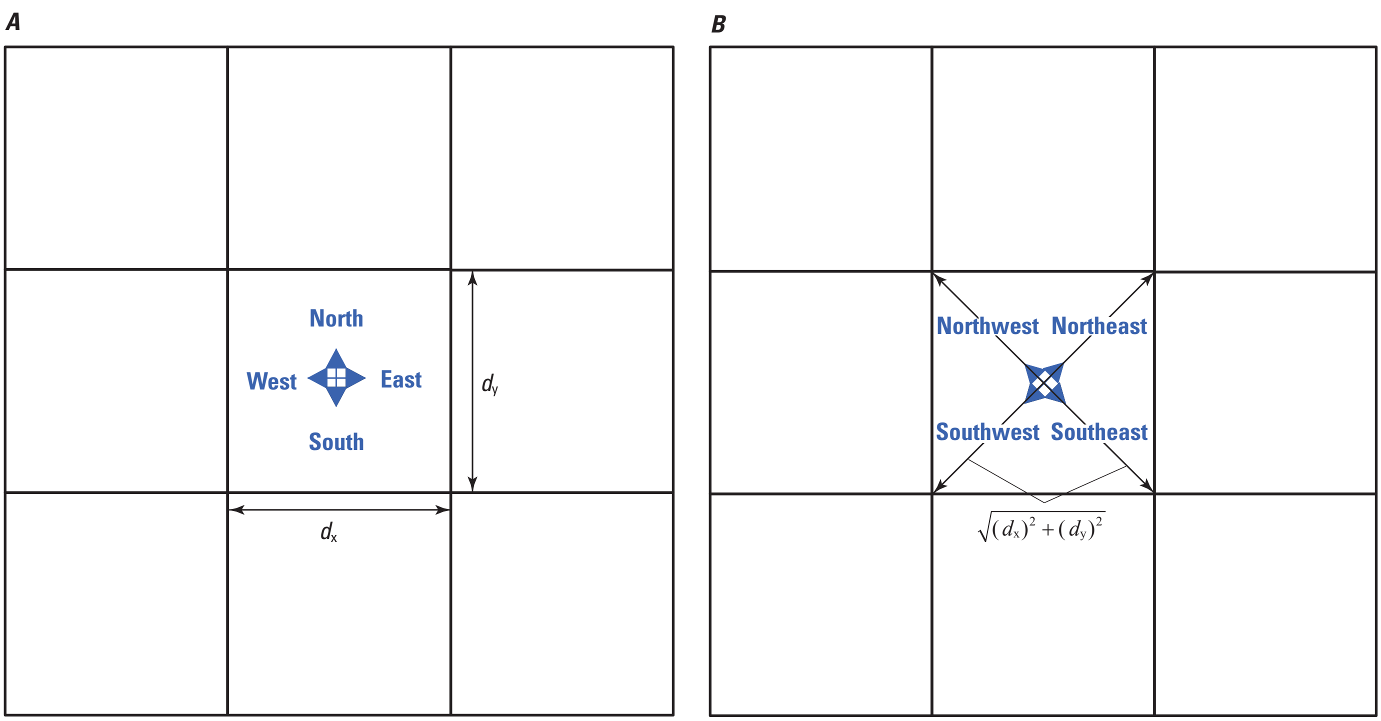


Figure 3. Schematic showing the process used to calculate flow width for use in the compound topographic index calculation. *A*, Flow widths for flow in the east-west or north-south directions are equal to the x dimension (d_x) or y dimension (d_y) of the pixel, respectively. The x and y dimensions are functions of the latitude and longitude of the pixels. *B*, Flow width for flow in the northeast, southeast, southwest, or northwest directions are equal to the length of the diagonal of the pixel, determined by d_x and d_y.

Table 2. Four largest river systems of the six continents included in the database.

Continental area	Leading digit	Basin 2 name	Basin 4 name	Basin 6 name	Basin 8 name
Asia	1	Ob	Yenisey	Lena	Amur
Africa	2	Nile	Zambezi	Congo	Niger
North America	3	Mackenzie	Nelson	St. Lawrence	Mississippi
Europe	4	Shatt al-Arab	Don	Dnieper	Danube
South America	5	Orinoco	Amazon	Tocantins	Rio de la Plata
Australasia	6	Murray-Darling	Flinders	Burdekin	Fitzroy

the topological information carried by the Pfafstetter code, the Amazon River and one of its main tributaries, the Rio Xingu, will be used as an example (fig. 4). Because the Amazon River Basin is basin 4 (table 2) on the South American continent (leading digit 5, which is not considered a level in Pfafstetter numbering; table 1), its Level 1 Pfafstetter code is 54 (fig. 4*A*). The basin for the Rio Xingu receives an even digit, so its Level 2 Pfafstetter code is 542 (fig. 4*B*). The Rio Xingu basin is then further subdivided and numbered with Level 3 (fig. 4*C*) and Level 4 (fig. 4*D*) Pfafstetter codes. The subdivision proceeds until no further tributaries require numbering.

The numbering of all the stream segments and catchments for the HDMA database required Pfafstetter coding to twelve digits. Trailing zeroes indicate no further subdivision of a particular unit. The Pfafstetter code carries important topological information that permits, by simple query, the identification of stream segments (and their associated catchments) that lie both up- and downstream of any location. The logic is different depending on whether the trace is performed up- or downstream. Each is explained below.

Upstream Tracing

The Pfafstetter code can be used to identify the streams and catchments that lie upstream from a particular location by examining the code to determine in what basin the location lies. The Pfafstetter code is examined by parsing the digits from right to left until the first even digit is encountered. The first even digit identifies the smallest basin in which the location lies.

As an example, the streams and catchments upstream from the stream segment numbered with the Pfafstetter code of 544131000000 are determined by examining the Pfafstetter code from right to left. The first even digit encountered is the digit “4” in the tenth position from the right. This means that this stream segment lies on the main stem of basin 544 identified at the second subdivision (the leading continental digit [5 in this case] should not be used in the Pfafstetter logic). To determine all the segments upstream from this location, first the main-stem basin code is incremented by one (to 545), and the remainder of the 12-digit Pfafstetter code is filled with trailing zeroes. Then, these two PFAF_CODE values are used to compose a map algebra expression that selects the segments upstream from this location.

$$\mathit{PFAF\_CODE} \geq 544131000000 \ \mathit{AND}$$
$$\mathit{PFAF\_CODE} < 545000000000$$

This query returns the 941 stream segments that are tributary to (upstream from) the stream with the Pfafstetter code of 544131000000.

Because the digit zero embedded in a Pfafstetter code indicates a closed basin, if a zero is encountered after any nonzero digit (odd or even) when parsing the digits from right to left, the parsing stops. Therefore, if no even digit was encountered before finding a zero after nonzero digits, then the basin has no upstream contributing basins. To illustrate, the basin with the Pfafstetter code of 511222010000 is located in South America (leading digit 5). Parsing from right to left, the first nonzero digit encountered is 1. The next digit to the left is 0, indicating that this basin has no upstream basins. Parsing stops at this point. However, another South American example with an embedded zero is the basin with the Pfafstetter code of 501410000000. In this case, parsing from right to left, the digit 4 in the ninth position from the right is the first even digit encountered. To build the upstream query as in the first example, the rightmost even digit is incremented by one, the digits filled in with zeroes, and the map algebra expression built.

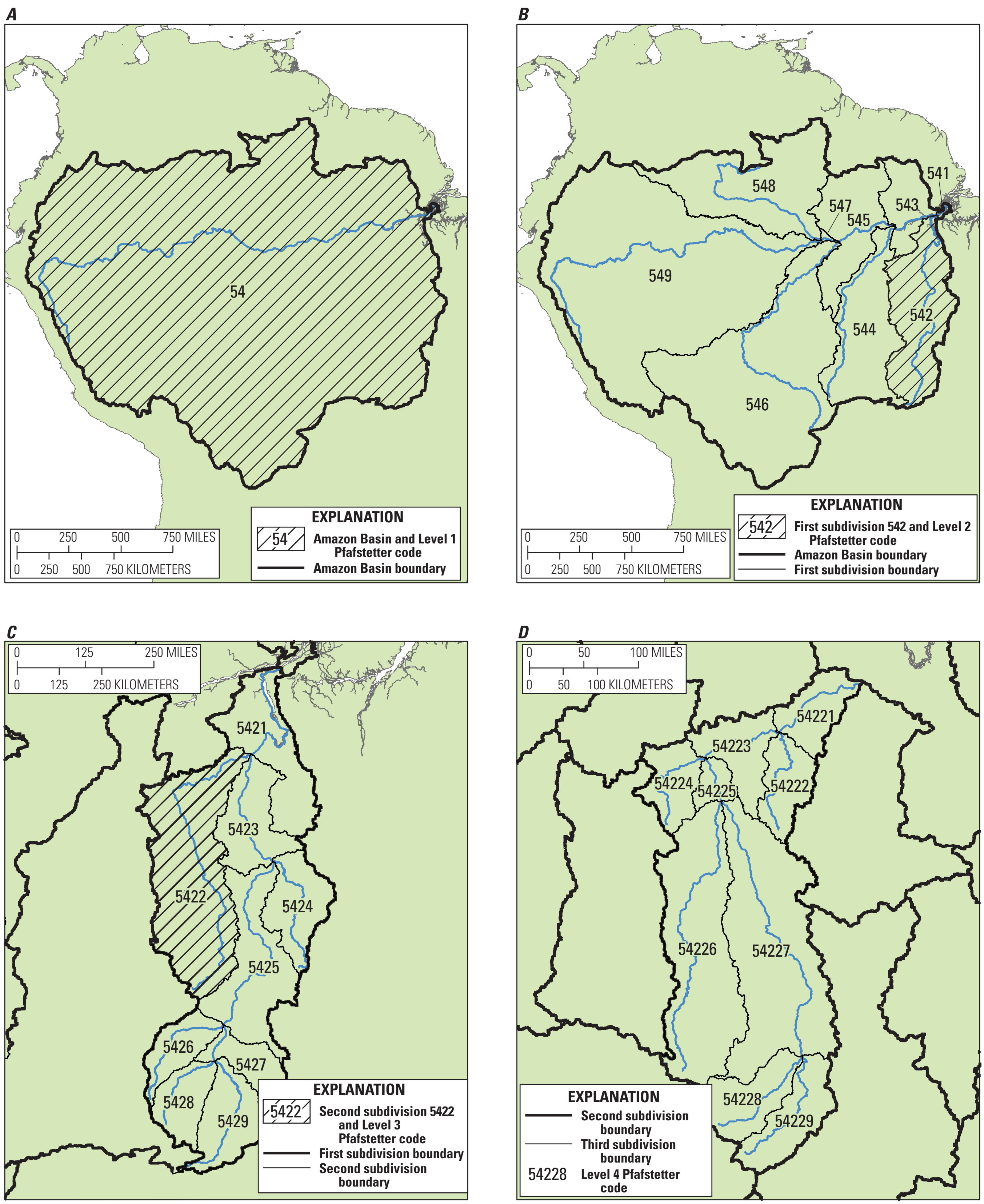
$$\mathit{PFAF\_CODE} \geq 501410000000 \ \mathit{AND}$$
$$\mathit{PFAF\_CODE} < 501500000000$$

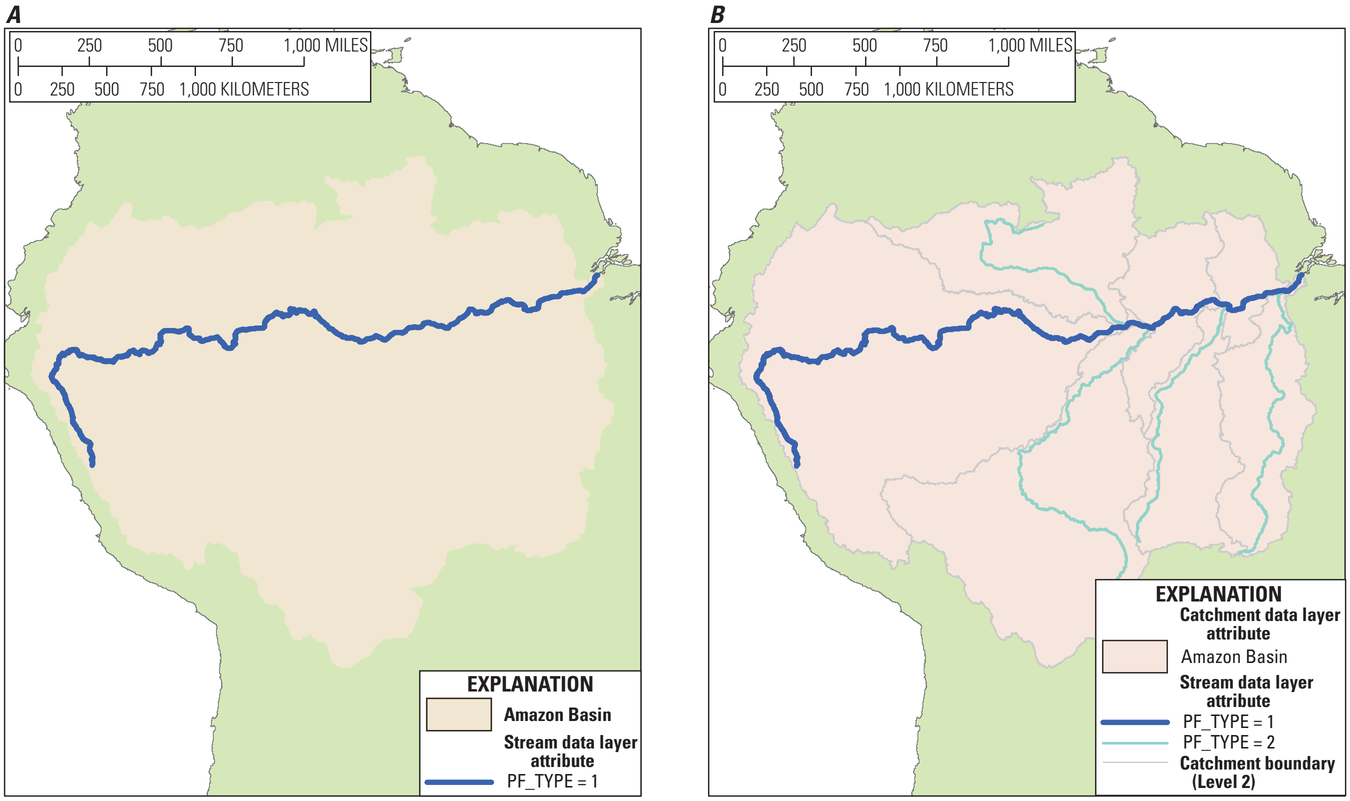
This query returns the 19 stream segments that are tributary to (upstream from) the stream with the Pfafstetter code of 501410000000.

Downstream Tracing

The Pfafstetter codes, along with an additional attribute on the streams data layer (PF_TYPE), can be used to trace the streams downstream through the network. The PF_TYPE attribute identifies at which Pfafstetter subdivision level the stream segment is identified as the main stem of one of the four basins called out in the Pfafstetter coding. Figure 5 illustrates

Figure 4 (following page). Iterative Pfafstetter subdivision of the Amazon River Basin highlighting an upper reach of the Rio Xingu. *A*, Amazon River Basin assigned the Pfafstetter code 54: 5 is the leading continental digit for South America, and 4 is the Level 1 Pfafstetter code for the Amazon Basin. *B*, Next level subdivision of the Amazon Basin: the basins for the four main tributaries identified for the Amazon River are assigned the even digits (2, 4, 6, 8), the headwaters subbasin is assigned the digit 9, and the interbasins are assigned the remaining odd digits (1, 3, 5, 7). The basin for the Rio Xingu (542) is identified. *C*, Subdivision of basin 542, the Rio Xingu, to third Pfafstetter level: even digits are assigned to the Rio Xingu’s four major tributaries, 9 to the headwaters subbasin, and the remaining odd digits to the interbasins. A tributary basin (5422) is identified. *D*, Subdivision of basin 5422 to fourth Pfafstetter level: even digits are assigned to four major tributaries, 9 to the headwaters for the 5422 tributary basin, and the remaining odd digits to the interbasins.


A
54
EXPLANATION
54
Amazon Basin and Level 1 Pfafstetter code
Amazon Basin boundary
0 250 500 750 MILES
0 250 500 750 KILOMETERS
B
541
543
547
545
548
549
544
542
546
EXPLANATION
542
First subdivision 542 and Level 2 Pfafstetter code
Amazon Basin boundary
First subdivision boundary
0 250 500 750 MILES
0 250 500 750 KILOMETERS
C
0 125 250 MILES
0 125 250 KILOMETERS
5421
5423
5422
5424
5425
5426
5427
5428
5429
EXPLANATION
5422
Second subdivision 5422 and Level 3 Pfafstetter code
First subdivision boundary
Second subdivision boundary
D
0 50 100 MILES
0 50 100 KILOMETERS
54221
54223
54224
54225
54222
54226
54227
54228
54229
EXPLANATION
Second subdivision boundary
Third subdivision boundary
54228 Level 4 Pfafstetter code



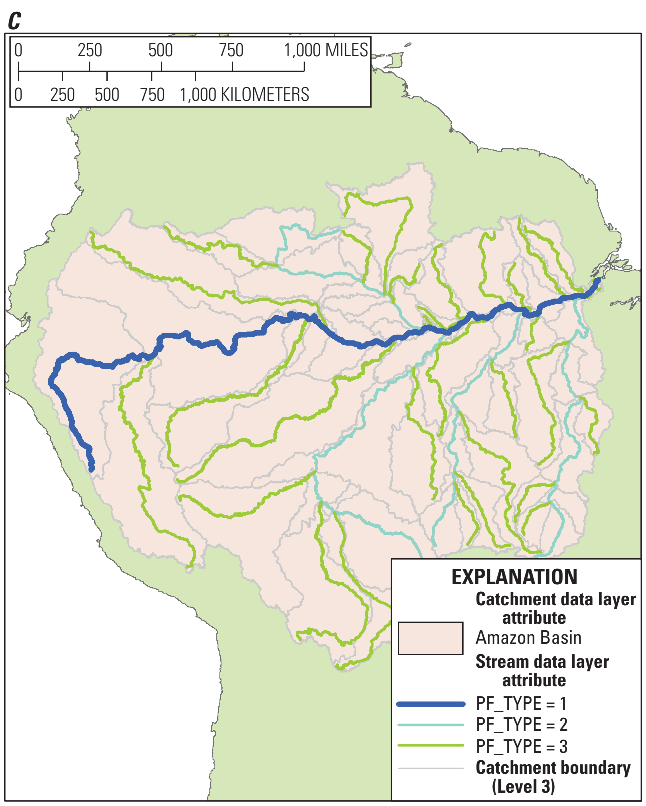


Figure 5. Explanation of the PF_TYPE attribute for the Amazon River Basin. *A*, The main stem of the Amazon River, from the mouth to the headwaters, is assigned PF_TYPE = 1 because the Amazon Basin was identified at the first subdivision and has an even Level 1 Pfafstetter code (54). *B*, The main stems of the basins identified at the second subdivision, from their headwaters to their confluences with the main stem of the Amazon, are assigned the PF_TYPE =2. *C*, The main stems of the basins identified at the third subdivision, from their headwaters to their confluences with other main stems, are assigned the PF_TYPE = 3.

the structure of the PF_TYPE for the first 3 subdivisions of the Amazon River Basin. Downstream tracing is a multistep process, stepping down through the main-stem segments, building a query for each even digit encountered.

The query is built using both the Pfafstetter code and the PF_TYPE attribute. For example, the stream segment coded with the PFAF_CODE 543647300000 is in a tributary of the Amazon River. It has a PF_TYPE value of 4, which is consistent with the fact that the first even digit (the digit 4) identified in parsing the PFAF_CODE from right to left is in the fourth position from the left (disregarding the leading continental digit). To determine the stream segments downstream from this location, then, one builds logic using all the even digits found in the Pfafstetter code along with the corresponding PF_TYPE value. The downstream trace steps down the network through each of the basin levels. Parsing from right to left, the first even digit encountered is the 4 in the fourth position from the left (PF_TYPE = 4, ignoring the leading continental digit), the next even digit is the 6 in the third position from the left (PF_TYPE = 3), and the last even digit is the 4 in the first position (PF_TYPE = 1). Using this information, to step down from basin 543647300000 to the confluence with the next main stem, the following logic is used, with the digits to the right of the rightmost even digit (4) filled in with zeroes for the first step.

$$PFAF\_CODE \leq 543647300000 \ AND$$
$$PFAF\_CODE \geq 543640000000 \ AND \ PF\_TYPE = 4$$

From this location (543640000000) to the next confluence, a similar logic is built, filling in with zeroes to the right of the next even digit (6).

$$PFAF\_CODE \leq 543640000000 \ AND$$
$$PFAF\_CODE \geq 543600000000 \ AND \ PF\_TYPE = 3$$

And from this location (543600000000) to the coast, a similar logic is built, filling in with zeroes to the right of the last even digit (4).

$$PFAF\_CODE \leq 543600000000 \ AND$$
$$PFAF\_CODE \geq 540000000000 \ AND \ PF\_TYPE = 1$$

Data Availability

The entire dataset is available at https://doi.org/10.5066/F7S180ZP (Verdin, 2017). The dataset comprises five raster grids (DEM, flow direction, flow accumulation, slope, and compound topographic index) and three vector layers (streams, catchments, and the boundaries used to subdivide raster processing [processing units]).

The raster grids are distributed by processing unit because of the large size of the raster datasets. Processing units are based on the Level 1 Pfafstetter code and are subdivided into smaller units where needed. All raster grids were assembled at full resolution: 3 or 7.5 arc-seconds for most data and 30 arc-seconds for Greenland.

1. *DEM*.—Horizontal units of decimal degrees and vertical units of meters.
2. *Flow direction*.—Unitless. The flow-direction convention adopted is that used by the ArcGIS implementation of the flow direction algorithm (Esri, 2016).
3. *Flow accumulation*.—Units of square kilometers.
4. *Slope*.—Expressed as a percent (rise over run, or tangent of the slope angle) but converted to integers (by multiplying by 100) to minimize the size of the distribution files. The values are in units of percent.
5. *Compound Topographic Index (CTI)*.— A floating point layer derived from the slope and flow accumulation grids but converted to integers (by multiplying by 100) to minimize the size of the distribution files. The values are unitless.

The vector data are packaged for distribution by continent; each vector continental package contains the streams, catchments, and processing units for the continent.

6. *Streams*.—Globally, there are 278,758 streamlines derived from the various DEM sources. The stream segments carry the following attributes.
   A. FLOW_ACC—The contributing drainage area to the mouth of the stream, expressed in square kilometers.
   B. PF_TYPE—An attribute indicating at which Pfafstetter subdivision level the stream segment is identified as a basin. These numbers range from 1 (indicating main stem in the first subdivision) to 11 (small headwaters streams).
   C. PFAF_CODE—The Pfafstetter code associated with the stream segment. This field is 12 digits long with the first digit indicating the continent (table 1). The following 11 digits are the Pfafstetter code.

   The streams datasets are distributed by continent.
7. *Catchments*.—There are 295,335 catchments delineated for the globe. Of these, 16,577 catchments, all with odd Pfafstetter digits following the leading continental digit, do not correspond to a stream segment; they are either coastal or closed interbasins. The catchment data layer carries the 12-digit Pfafstetter code (PFAF_CODE). The catchment datasets are distributed by continent.
8. *Processing units*.—The processing units depict the units used in the processing and distribution of the raster grids. They were defined by hydrologic divide and carry only an attribute defining the number of the processing unit (UNIT). These layers are distributed by continent, along with the streams and catchments.

The final processing and distribution units for the continents are shown in figures 6 through 11. The landmass of Greenland was not subdivided into streams and catchments, but raster grids are available for Greenland (leading continental digit, 7) as a whole.

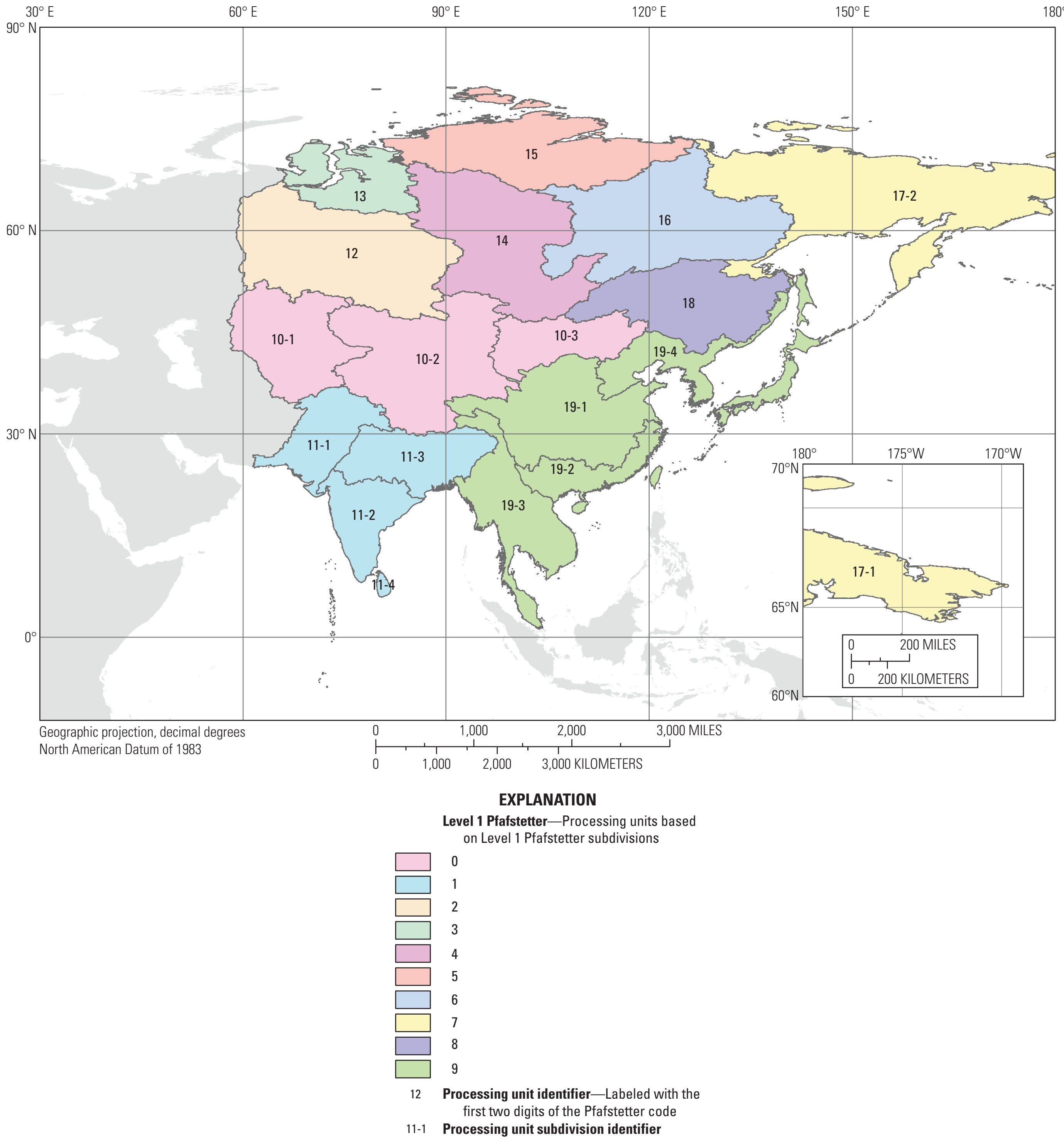


Figure 6. Processing and distribution units for Asia (leading continental digit, 1).

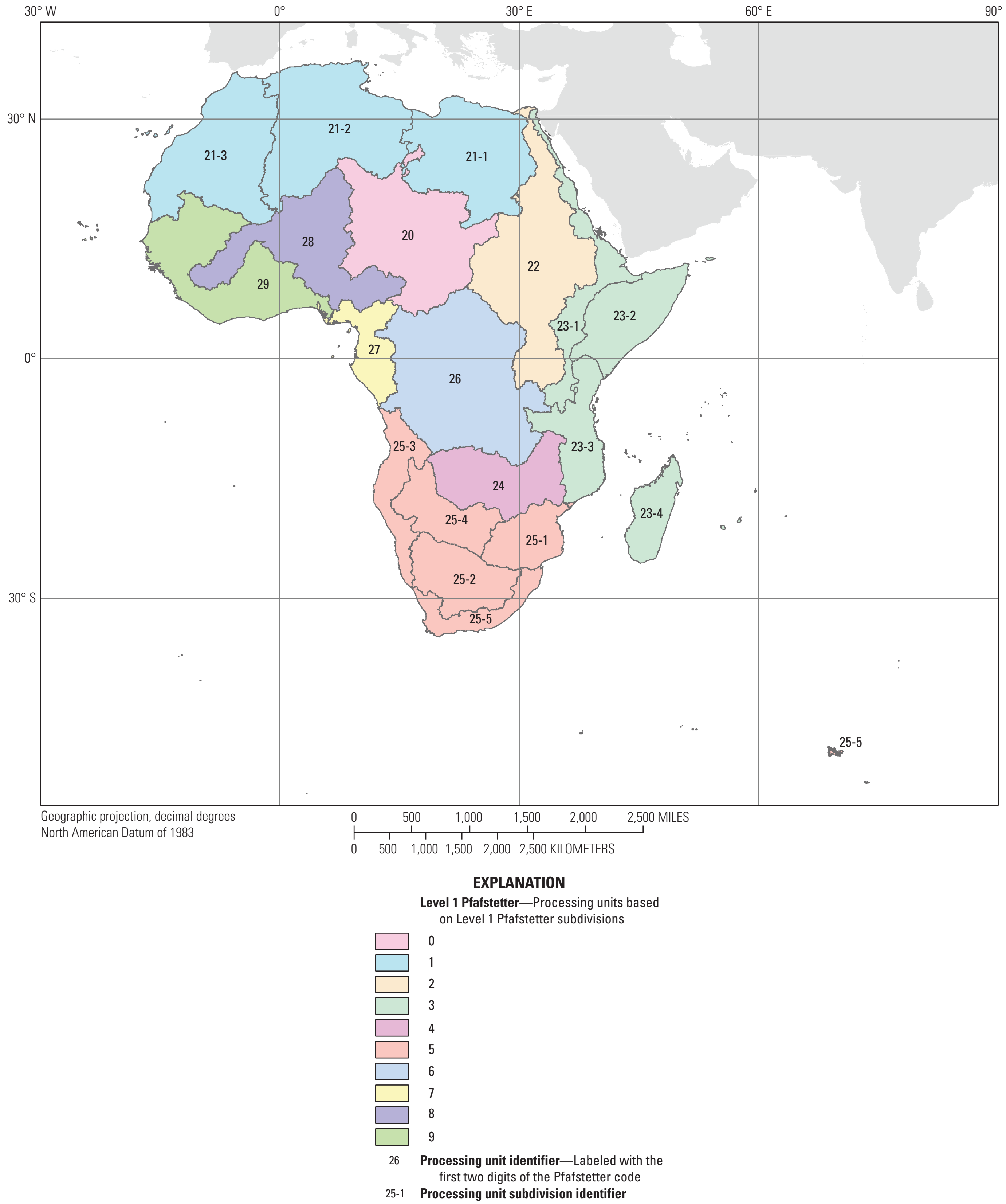


Figure 7. Processing and distribution units for Africa (leading continental digit, 2).

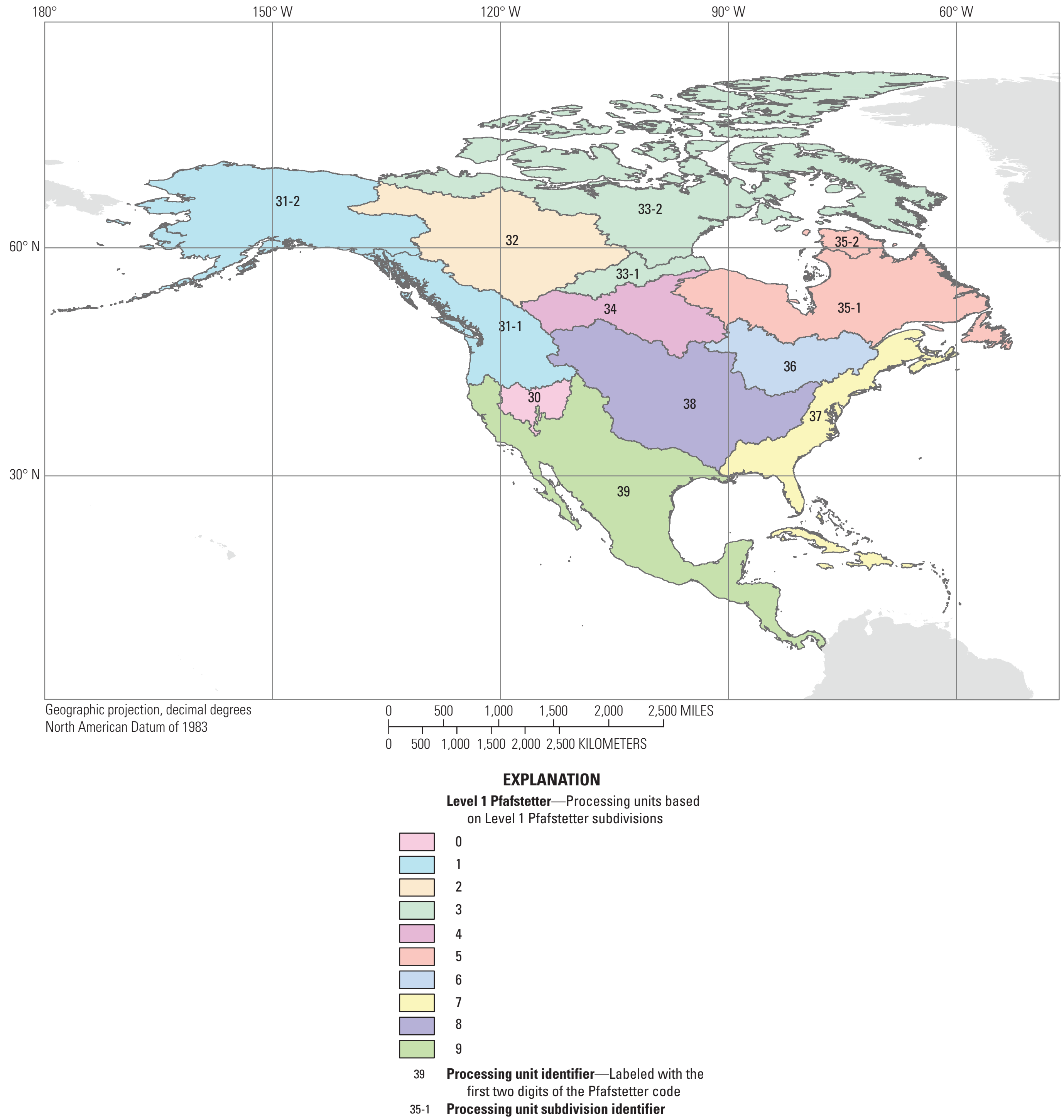


Figure 8. Processing and distribution units for North America (leading continental digit, 3).

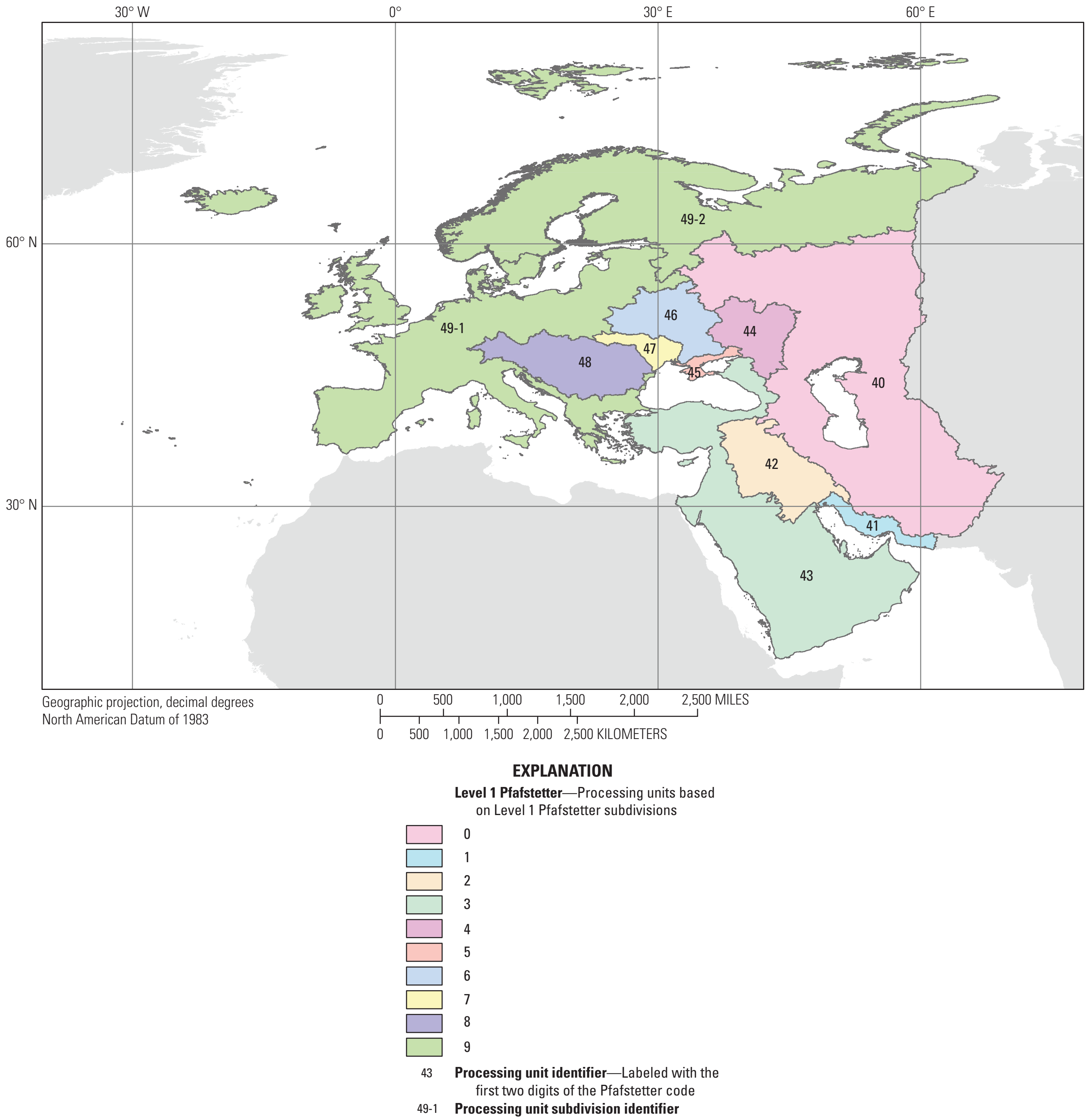


Figure 9. Processing and distribution units for Europe (leading continental digit, 4).

Figure 10. Processing and distribution units for South America (leading continental digit, 5).

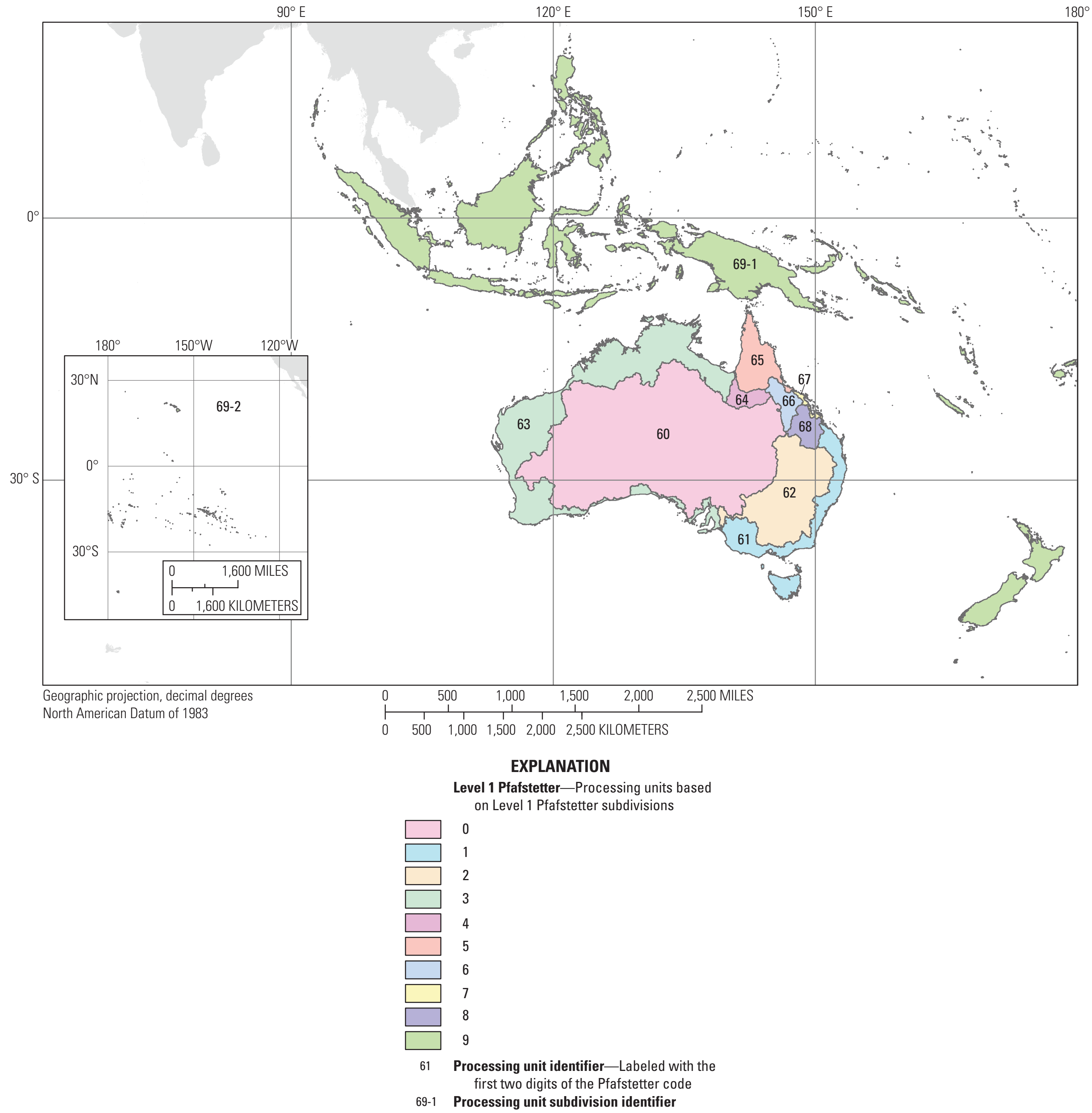


Figure 11. Processing and distribution units for Australasia (leading continental digit, 6). Main continent is west of the international dateline; islands (including Hawaii) shown in inset (69-2) are east of the dateline.

References Cited

Beighley, R.E., Eggert, K.G., Dunne, Thomas, He, Y., Gummadi, Venkat, and Verdin, K.L., 2009, Simulating hydrologic and hydraulic processes throughout the Amazon River Basin: Hydrological Processes, v. 23, no. 8, p. 1221–1235.

Danielson, J.J., and Gesch, D.B., 2011, Global Multi-resolution Terrain Elevation Data 2010 (GMTED2010): U.S. Geological Survey Open-File Report 2011–1073, 26 p.

Esri, 2016, ArcGIS for Desktop, release 10.3.1: Redlands, Calif., Environmental Systems Research Institute.

Farr, T.G., and Kobrick, Mike, 2000, Shuttle radar topography mission produces a wealth of data: Eos, Transactions, American Geophysical Union, v. 81, no. 48, p. 583–585.

Gesch, D.B., Verdin, K.L., and Greenlee, S.K., 1999, New land surface digital elevation model covers the Earth: Eos, Transactions, American Geophysical Union, v. 80, no. 6, p. 69–70.

Hall, D.G., Cherry, S.J., Reeves, K.S., Lee, R.D., Carroll, G.R., Sommers, G.L., and Verdin, K.L., 2004, Water energy resources of the United States with emphasis on low head/low power resources: U.S. Department of Energy, Office of Energy Efficiency and Renewable Energy, Wind and Hydropower Technologies, DOE/ID-11111, 52 p.

International Centre for Tropical Agriculture, 2004, SRTM [Shuttle Radar Topographic Mission] 90m digital elevation data—Void-filled, seamless for Cali, Columbia: Consortium for Spatial Information (CGIAR–CSI) dataset, accessed May 15, 2017, at http://srtm.csi.cgiar.org.

Jenson, S.K., and Domingue, J.O., 1988, Extracting topographic structure from digital elevation data for geographic information system analysis: Photogrammetric Engineering and Remote Sensing, v. 54, no. 11, p. 1593–1600.

Kumar, Praveen, Verdin, K.L., and Greenlee, S.K., 2000, Basin level statistical properties of topographic index for North America: Advances in Water Resources, v. 23, no. 6, p. 571–578.

Lehner, Bernhard, Verdin, Kris, and Jarvis, Andy, 2006, HydroSHEDS technical documentation, version 1.0: World Wildlife Fund, accessed November 2016 at http://hydrosheds.cr.usgs.gov/webappcontent/HydroSHEDS_TechDoc_v10.pdf.

Lehner, Bernhard, Verdin, Kristine, and Jarvis, Andy, 2008, New global hydrography derived from spaceborne elevation data: Eos, Transactions, of the American Geophysical Union, v. 89, no. 10, p. 93–94.

Moore, I.D., Grayson, R.B., Ladson, A.R., 1991, Digital terrain modeling—A review of hydrological, geomorphological, and biological applications: Hydrologic Processes, v. 5, p. 3–30.

Rivix, LLC, 2012, RiverTools 3.0.3: Broomfield, Colo., Rivix, LLC, accessed April 2013 at http://rivix.com.

Robinson, A.H., and Sale, R.D., 1969, Geographical tables, appendix D *of* Elements of cartography (3d ed.): New York, John Wiley & Sons, p. 362–364.

Tarboton, D.G., 1997, A new method for the determination of flow directions and upslope areas in grid digital elevation models: Water Resources Research, v. 33, no. 2, p. 309–319.

Verdin, K.L., 2017, Hydrologic Derivatives for Modeling and Analysis: U.S. Geological Survey data release, accessed July 2017 at https://doi.org/10.5066/F7S180ZP.

Verdin, K.L.; Godt, Jonathan; Funk, Chris; Pedreros, Diego; Worstell, Bruce; and Verdin, James, 2007, Development of a global slope dataset for estimation of landslide occurrence resulting from earthquakes: U.S. Geological Survey Open-File Report 2007–1188, 25 p.

Verdin, K.L., and Jenson, S.K., 1996, Development of continental scale DEMs and extraction of hydrographic features, *in* Third International Conference/Workshop on Integrating GIS and Environmental Modeling, Santa Fe, N. Mex., January 21–26, 1996, Proceedings [CD-ROM]: Santa Barbara, Calif., National Center for Geographic Information and Analysis, p. 397–402.

Verdin, K.L., and Verdin, J.P., 1999, A topological system for delineation and codification of the Earth's river basins: Journal of Hydrology, v. 218, no. 1–2, p. 1–12.

Verdin, K.L., and Worstell, Bruce, 2008, A fully distributed implementation of mean annual streamflow regional regression equations: Journal of the American Water Resources Association, v. 44, no. 6, p. 1537–1547.

Zevenbergen, L.W., and Thorne, C.R., 1987, Quantitative analysis of land surface topography: Earth Surface Processes and Landforms, v. 12, p. 47–56.

Publishing support provided by:
Denver Publishing Service Center, Denver, Colorado

For more information concerning this publication, contact:
Director, USGS Colorado Water Science Center
Box 25046, Mail Stop 415
Denver, CO 80225
(303) 236-4882

Or visit the Colorado Water Science Center website at:
https://co.water.usgs.gov/

This publication is available online at:
https://doi.org/10.3133/ds1053

USGS

Verdin, K.L.—**Hydrologic Derivatives for Modeling and Analysis—A New Global High-Resolution Database**—Data Series 1053

ISSN 2327-638X (online)
https://doi.org/10.3133/ds1053