**Final Report**
**High Resolution Topographic Analysis for GMAO's Catchment LSM**

**Kristine Verdin**
April 27, 2013

**Abstract**

The Catchment land surface model used by the Global Modeling and Assimilation office in its GEOS-5 modeling system relies on parameters derived from topographical data. The topographical data underlying the current operating version of the model was processed a decade ago and is of relatively course resolution (1 km) relative to that of the data available today (0.1 km). The purpose of this work was to update the topography-based parameters used by the Catchment land surface model (LSM) through the processing of the higher resolution topography data. The current topographic parameters, based on coarse resolution data, are inadequate for current GMAO needs.

This technical report summarizes the work done to development high-resolution topographic data layers for the globe. The USGS processed the best-available topography data for the globe into 1 minute x 1 minute arrays of derived parameters critical for Catchment LSM function: (i) Pfafstetter code, (ii) compound topographic index (CTI: mean, standard deviation, skew, minimum, and maximum), and (iii) elevation (mean and standard deviation).

**Data**

All of the derivative data layers have as their source a digital elevation model (DEM). A hybrid of several existing global DEMs was used in the development of the six derivative data layers.

The primary DEM used was the HydroSHEDS dataset (Lehner and others, 2008). This dataset was derived from the SRTM DEM (Farr and Kobrick, 2000) along with several ancillary layers to assist in hydrologically-conditioning the DEM. This dataset covers most land masses of the globe south of 60-degrees North and has a resolution of 3-arc-seconds.

For the land masses north of 60-degrees, the GMTED2010 dataset (Danielson and Gesch, 2011) was used. This dataset is a hybrid dataset of the best-available global data and has a resolution of 7.5-arc-seconds. Since one of the primary uses of the DEM data was derivation of drainage basins, the boundary delineating the divide between usage of the HydroSHEDS data and that of the GMTED2010 was a drainage divide. For this reason, some of the areas slightly south of 60-degrees north were also developed using the GMTED2010.

The third DEM data source used was the void-filled SRTM available from the Consultative Group on International Agricultural Research (CGIAR) Consortium for Spatial Information (International Centre for Tropical Agriculture (CIAT), 2004). These data were used to fill in the holes, south of 60-degrees north, in the HydroSHEDS database. These holes were, in general, islands which were not included in the

HydroSHEDS continental data development effort.  Figure 1 depicts the sources used in the data development effort.

## Methods

In order to develop the Pfafstetter codes and statistics of the elevation and compound topographic index (CTI), development of several preliminary data layers was necessary.  To accommodate the size of the global dataset and to insure seamless data development, the globe was subdivided into processing units.  The processing units were based on the Pfafstetter Level 1 subdivisions found in the HYDRO1k dataset (Verdin and Jenson, 1996), with additional subdivision of overly large units to facilitate processing.  The processing units are shown in Figure 2.  Additional subdivision of units was necessary when the processing units crossed the boundary between the HydroSHEDS and GMTED2010 source data.  The leading continental digits are taken from Verdin and Verdin, 1999 and are based on the relative sizes of the continents.  While Greenland was not included in the initial proposed subdivision, it was assigned the leading digit "7" for inclusion in this database.

All raster procedures done in the development of the initial data layers were done at the full-resolution of the data; 3-arc-second resolution for the HydroSHEDS and CGIAR data, 7.5-arc-second resolution for the most of the GMTED2010 data and 30-arc-second resolution for the GMTED2010 data for Greenland.

Base Layer Development:

1. Development of a hydrologically-conditioned DEM.  While the HydroSHEDS database contains a hydrologically-conditioned DEM, the GMTED2010 and CGIAR datasets do not.  Both of these datasets were conditioned using standard GIS processing tools to remove spurious sinks.  Edge-matching between the HydroSHEDS and GMTED2010 datasets was achieved through use of the HydroSHEDS DEM as a "wall" in the conditioning of the GMTED2010.  This ensured seamless (subpixel) agreement between the two datasets.  The GMTED2010 data were processed at 7.5-arc-second resolution while the HydroSHEDS and CGIAR data was kept at the full-resolution of 3-arc-seconds.  The DEM of Greenland was included in the database; however, both the resolution (30-arc-second) and quality of the DEM, derived from the GMTED2010 database, is lower than the other sources.
2. Development of the flow direction grid.  As with the hydrologically-conditioned DEM, the HydroSHEDS database contains a flow direction grid.  The flow direction grids for the GMTED2010 and the CGIAR DEMs were produced as part of the hydrologic-conditioning process.
3. Development of the flow accumulation grid.  The flow accumulation grid was developed for all sources of DEM data (HydroSHEDS, CGIAR and GMTED2010).  The flow accumulation grid is used to describe, for any location on the landscape, the upstream contributing area.  This contributing area is required for development of the CTI layer.  In projected DEMs, the flow accumulation grid can be easily translated into upstream area since each pixel is

assumed to be the same size.  Since the DEM data used in this effort are in a
geographic framework and the size of the pixels varies with latitude, the flow
accumulation grid was calculated in units of square kilometers through use of a
"weight grid" (Verdin and Worstell, 2008).  This grid defined the size of each
pixel as a function of latitude and was used as the weight when running the flow
accumulation function.  The weight grid was developed using the tables in
Robinson and Sales, 1969. The distances in the X-direction vary considerably,
with one degree of longitude at the equator equaling 111,321 meters, whereas at
59 degrees north or south latitude, the same one degree of longitude equates to
only 57,478 meters.  The variation of lengths in meters in the *Y*-direction are not
as dramatic, varying from 111,567 meters at the equator to 111,406 meters at 59
degrees north or south.

4. Development of the slope dataset.  The slope data were developed using
RiverTools software (Rivix, 2012).  RiverTools accommodates large, geographic
datasets and does latitude-specific adjustments on-the-fly.  The slope was
calculated as "rise over run" using a partial quartic surface which is fitted locally
to determine the slope according to the method described by Zevenbergen and
Thorne, 1987.

5. Development of the streams/catchments datasets.  The procedures used for
Pfafstetter codification require definition of a stream network and corresponding
catchments.  The streams were derived using the flow accumulation grids and
applying an areal threshold of 250 square kilometers (i.e. streams are defined as
any point on the landscape that has as least 250 square kilometers of upstream
area).   Standard GIS procedures were done to extract the vector streams and
delineate a corresponding catchment for each stream segment.   For the globe,
~279,000 stream segments were defined and ~295,000 catchments (both basin
and interbasins).

Pfafstetter codification:

The streams and catchments were attributed with Pfafstetter codes which carry valuable
topological information.  The coding was done on the streams and catchments derived
from the full-resolution raster data (3-arc-second or 7.5-arc-second) and was done in a
vector/polygon processing environment.  Streams were navigated and numbered using
Pfafstetter procedures (Verdin and Verdin, 1999) and the Pfafstetter number transferred
to the corresponding catchment.  Interbasins without a corresponding stream segment
were assigned an appropriate Pfafstetter number (with all odd digits).  The Pfafstetter
codification was done on a continental basis.  Following completion of the codification, a
leading digit was added to the Pfafstetter number to indicate the continent:

| Leading digit for continental identification | Continental Area |
|---|---|
| 1 | Asia |
| 2 | Africa |

| | |
|---|---|
| 3 | North America |
| 4 | Europe |
| 5 | South America |
| 6 | Australasia |
| 7 | Greenland* |

**Table 1: Continental identification digits.  Greenland was included as Pfafstetter Unit 70, but subdivision of the landmass was not done.**

Development of the CTI layer

The Compound Topographic Index was developed using the slope, flow direction and flow accumulation base layers. The CTI is (Moore and others, 1991):

$$CTI = \ln (A_s/\tan \beta)$$

Where $A_s$ = the specific catchment area and
$\beta$ = slope

The tan of the slope was derived as one of the base layers. The specific catchment area is defined as the upslope area per unit width of contour. This is derived as a function of the flow accumulation grid and the flow direction grid. Since the size of the pixel varies with its geographic location, the unit width of contour is a function, both, of the direction of flow and the latitude/longitude of the location (Figure 3). The specific catchment area was obtained through developing a layer describing the width of the flow path and dividing the flow accumulation grid by that grid. The resulting CTI grid was created through map algebra once the specific catchment area grid was developed. An example python script used to develope the CTI for Africa is included as Appendix 1.

Summarization of layers to 1-arc-minute resolution

*Pfafstetter*

The Pfafstetter codification was done in a vector/polygon framework, with the algorithms working on the streams and catchment datasets. In order to transfer the resulting Pfafstetter codes into a raster framework with a 1-arc-minute resolution, the data were summarized using a point to raster operator. A 1-arc-minute mesh was created for the globe and the centroid of each pixel within the mesh was used to sample the underlying Pfafstetter code (from the catchment dataset). The resulting point dataset was converted to a grid using standard GIS operations (see Figure 4).

*Elevation and CTI Statistics:*

The DEM developed for this work is multiple-resolution with resolutions of 3 arc-seconds, 7.5 arc-seconds and 30 arc-seconds. The summary data layers were developed at a reduced, more globally manageable resolution of 1 arc-minute. The raster package (Hijmans and van Etten, 2012) of the R statistical software package (R Development Core Team, 2012) was used to aggregate the various data layers to the desired 1-arc-minute resolution. This involved aggregating the 3-arc-second layers using a factor of twenty (twenty pixels in both the x and y dimensions, or 400 pixels in total, were summarized to produce the final 1-arc-minute product), the 7.5-arc-second layers using a factor of 8, and the 30-arc-second layers using a factor of 2 (see Figure 5). Data were processed on a tiled basis and the resulting data layers were merged at the end of processing to assemble the global data layers.

1. Mean CTI
2. Standard deviation of the CTI
3. Skewness of the CTI
4. Mean elevation
5. Standard deviation of the elevation

## Results

There are two levels of detail in the final datasets developed in this effort. The first is the full-resolution data and the second is the data aggregated to the final 1-arc-minute resolution.

The full-resolution data include:
1. Streams data. There are 278,758 stream lines derived from the various DEM sources. The stream segments are attributed with the following attributes:
   a. Flow_acc—the contributing drainage area to the mouth of the stream in square kilometers
   b. Pf_type—an attribute indicating at which level the stream segment is identified as a basin in the Pfafstetter codification. These numbers range from 1 (indicating main stem in the first subdivision) to 11 (small headwaters streams). Figure 6 and 7 show the streams tagged with Pf_types of 1 and 2 (the first two subdivisions of the globe).
   c. Final_PfafNum—the Pfafstetter number associated with the stream segment. This field is twelve digits long with the first digit indicating the continent (Table 1). The following 11 digits are the actual Pfafstetter number.
2. Catchment data. There are 294,823 catchments delineated for the globe. Along with the catchments associated with streamlines, there are 16,065 catchments without a corresponding stream segment, which are either coastal or closed interbasins. The catchment dataset carries the 12-digit Pfafstetter number.
3. Global DEM. With units of meters, the data were assembled at the full-resolutions of 3-, 7.5- and 30-arc-seconds.
4. Global CTI. Derived from the slope and flow accumulation data, the CTI data were assembled at the full-resolutions of 3-, 7.5- and 30-arc-seconds.

Additional intermediate global datasets were developed such as flow direction, flow accumulation and slope.

Final Aggregated Data Layers:

Developed from the full-resolution data, six layers were developed at 1-arc-minute resolution. The data are distributed as GeoTIFFS, have 21600 rows and 10800 columns, and are masked with the value "-9999" indicating nodata pixels. Table 2 and Figures 8 through 12 summarize the data.

| Data | Pixel Depth | Min Value | Max Value | Figure |
|---|---|---|---|---|
| Pfafstetter Code | 64-bit | 101111000000 | 699900000000 | -- |
| Mean CTI | 32-bit float | 2.79 | 27.62 | Figure 8 |
| Std Dev CTI | 32-bit float | 0 | 9.76 | Figure 9 |
| Skewness CTI | 32-bit float | -0.199 | 0.199 | Figure 10 |
| Mean Elevation | 32-bit float | -414.68 | 8053.71 | Figure 11 |
| Std Dev Elevation | 32-bit float | 0 | 688.05 | Figure 12 |

**Table 2: Details on six final layers developed at 1 arc-minute resolution**

The different sources of data are apparent in the CTI layers.  This is, in part, due to the use of the slope in the development of the CTI.  The coarser resolution of the data from the GMTED2010 results in lower calculated slopes, and decreased mean and standard deviations of the slope (Evans, 1979).  These reduced slopes from the 7.5-arc-second and 30-arc-second DEM data result in CTI values which are relatively larger than those derived from the higher-resolution, 3-arc-second SRTM data.

## References

- 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.
- Evans, I.S., 1979, An integrated system of terrain analysis and slope mapping. Final Report on DA-ERO-591-73-G0040 (Durham: University of Durham).
- Farr, T. G., and M. Kobrick (2000), Shuttle Radar Topography Mission produces a wealth of data, *Eos Trans. AGU, 81*(48), 583, 585.
- Hijmans, Robert J. and Jacob van Etten (2012). raster: Geographic analysis and modeling with raster data. R package version 2.0-12., http://CRAN.R-project.org/package=raster
- International Centre for Tropical Agriculture (CIAT) (2004), Void-filled seamless SRTM data, Cali, Colombia. (Available at http://srtm.csi.cgiar.org)
- Lehner, B., K. Verdin and A. Jarvis, 2008, New Global Hydrological Data based on SRTM Elevation Derivatives, Eos, Transactions, American Geophysical Union, Vol 89, No. 10, pp. 93-94.
- Moore, I.D., Grayson, R.B., Ladson, A.R., 1991. Digital terrain modeling: a review of hydrological, geomorphological, and biological applications. Hydrol. Processes 5, 3–30.
- R Development Core Team (2012). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org/.
- 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, Elements of Cartography, 3rd edition: New York, John Wiley & Sons, Appendix D.
- Verdin, K., and S. Jenson, 1996, Development of continental scale DEMs and extraction of hydrographic features, proceeding of the Third International Conference/Workshop on Integrating GIS and Environmental Modeling, Santa Fe, New Mexico, January 21-26, 1996, (CD-ROM), National Center for Geographic Information and Analysis, Santa Barbara, California, 93106, USA
- Verdin, K.L. and J.P. Verdin, 1999, A topological system for delineation and codification of the Earth's river basins, Journal of Hydrology, vol. 218, nos. 1-2, pp. 1-12
- Verdin, K. and B. Worstell, A Fully Distributed Implementation of Mean Annual Stream Flow Regional Regression Equations, Journal of the American Water Resources Association, Vol. 44, Issue 6, pp. 1537-1547, Dec. 2008
- Zevenbergen, L.W., and C.R. Thorne, 1987. Quantitative analysis of land surface topography, *Earth Surface Processes and Landforms*, 12:47–56.
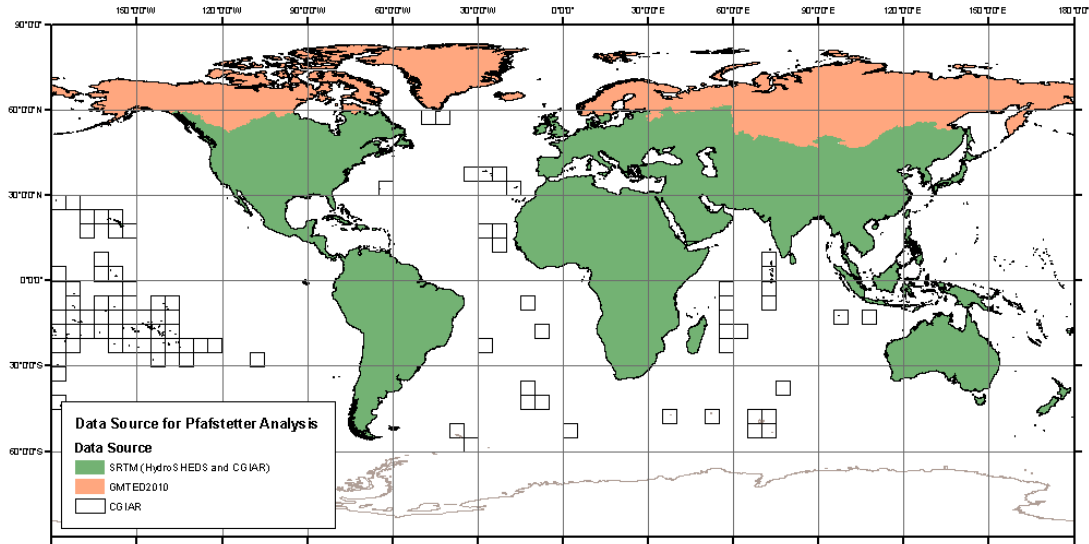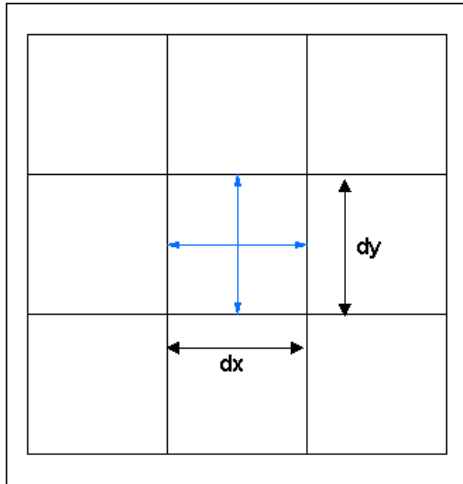
**Figure 1: DEM data sources used in the development of the database for derivation of streams, watersheds, compound topographic index and Pfafstetter codification.**
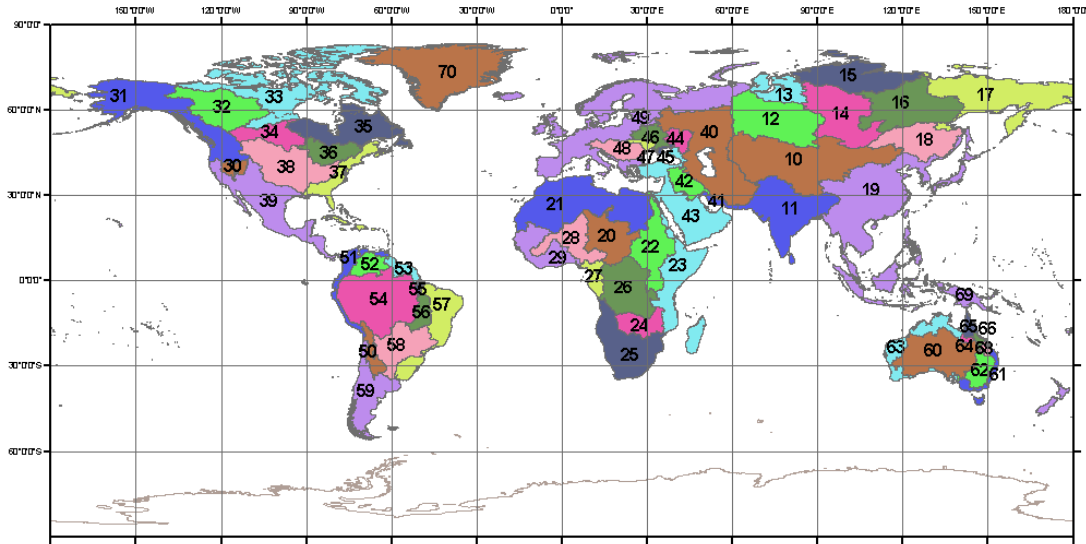
**Figure 2: Processing units based on HYDRO1k Level1 Pfafstetter units. Leading continental digits are after Verdin and Verdin, 1999.**

Flow path for flow directions in the East/West or North/South directions is the x-dimension or y-dimension of the pixel. The x- and y-dimensions are functions of the latitude and longitude of the pixels.
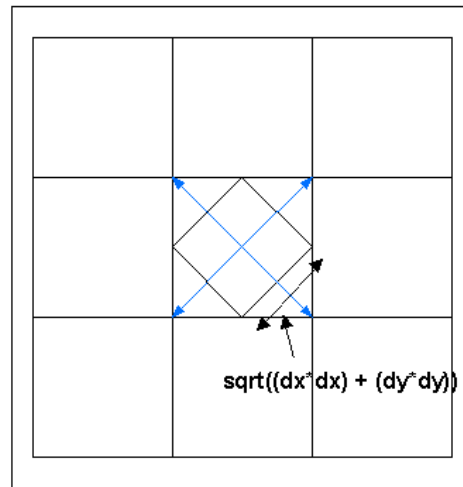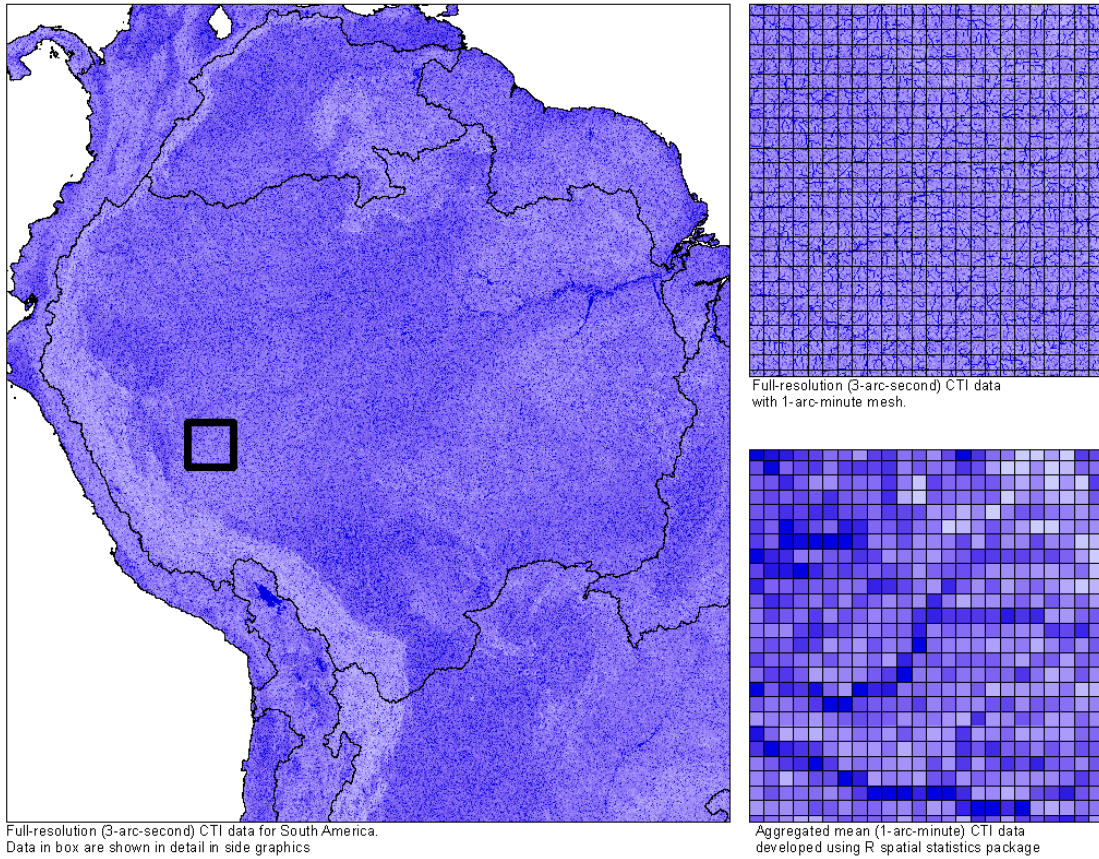
Flow path for flow directions in the NE, SE, SW or NW directions is the dimension of the diagonal.
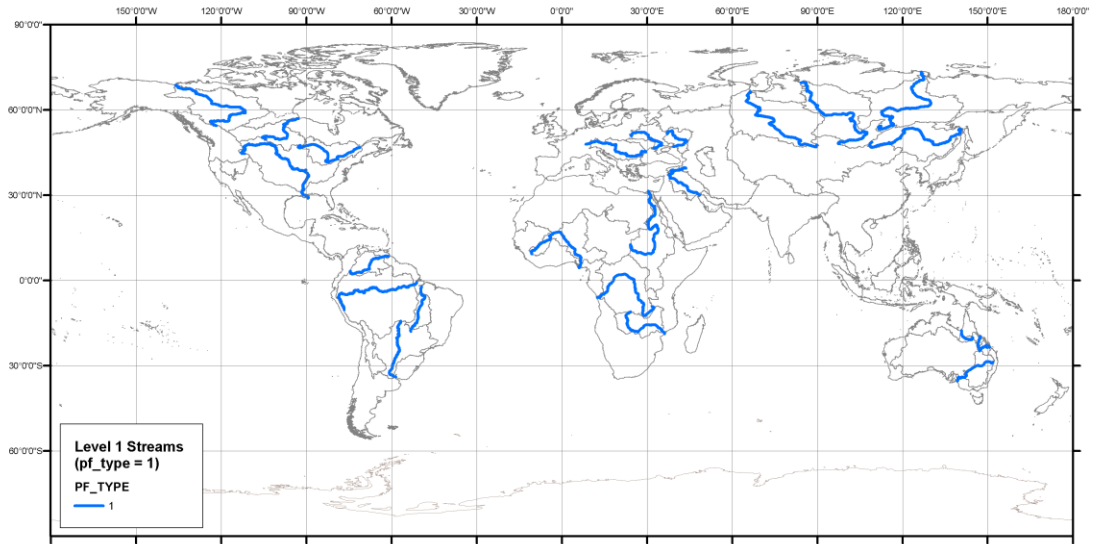
$sqrt((dx*dx) + (dy*dy))$

**Figure 3: Calculation of the flow width for use in the CTI calculation.**
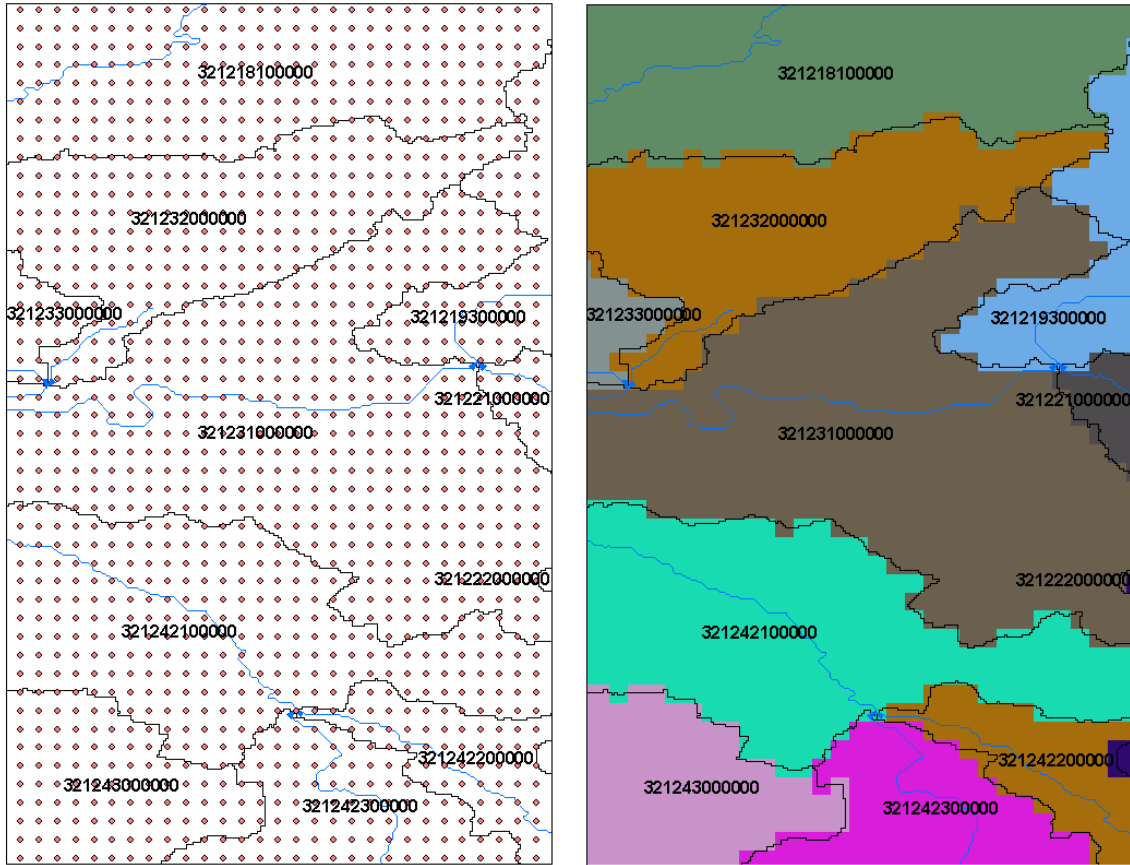
**Figure 4: Aggregation of Pfafstetter Codes from polygon to 1-arc-minute grid. Area shown is a small portion of North America. Pfafstetter codes are extracted from polygon using 1-arc-minute mesh labels and converted to 1-arc-minute raster.**

Full-resolution (3-arc-second) CTI data for South America.
Data in box are shown in detail in side graphics

Full-resolution (3-arc-second) CTI data
with 1-arc-minute mesh.

Aggregated mean (1-arc-minute) CTI data
developed using R spatial statistics package

**Figure 5: Example of aggregation of the CTI data to the final 1-arc-minute resolution**

**Figure 6: Streams segments identified as "basins" in the first Pfafstetter subdivision of the globe (PF_TYPE = 1)**
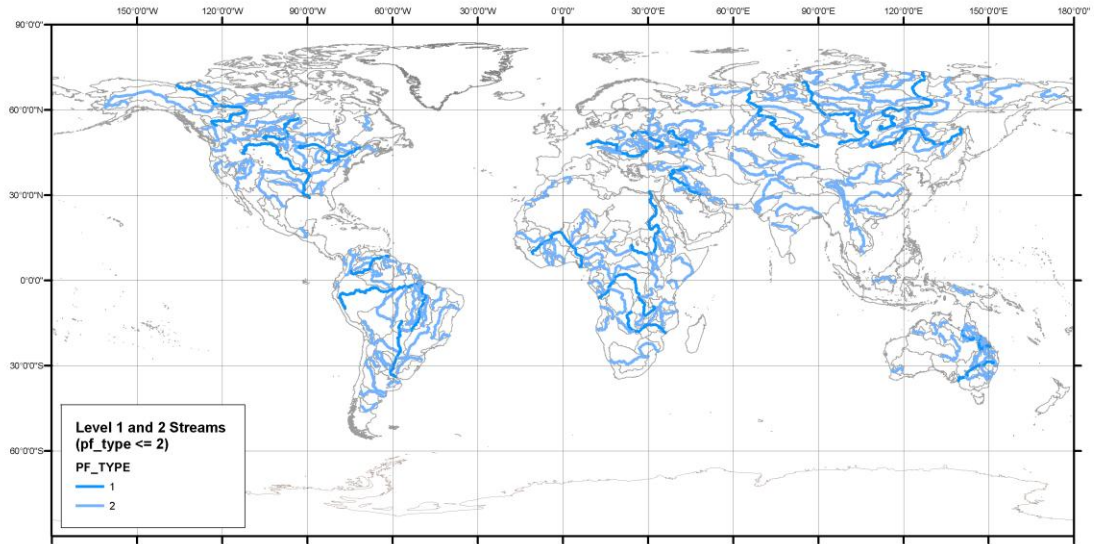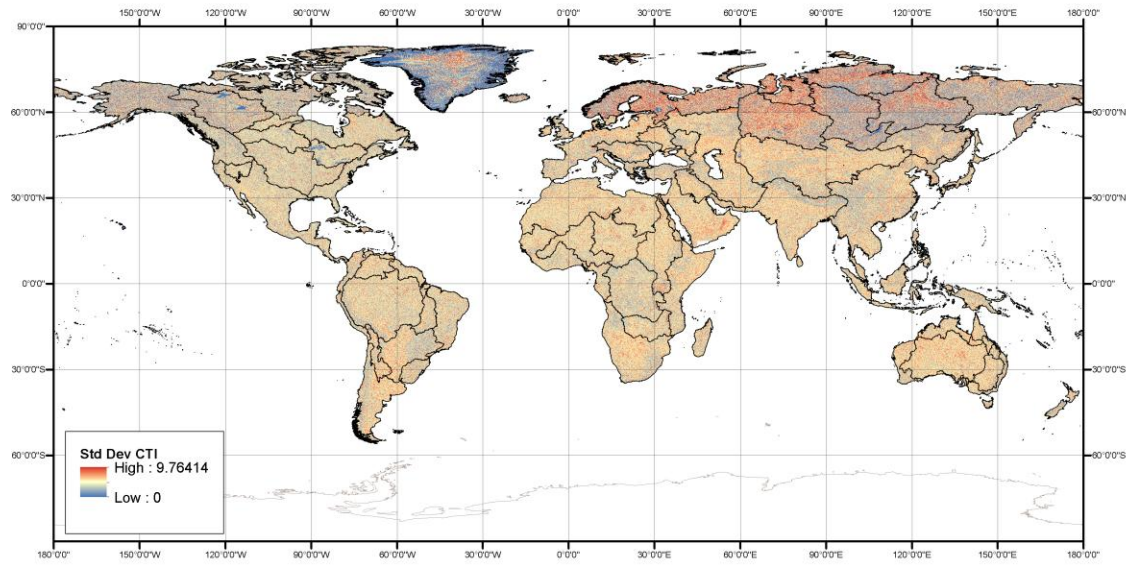
**Figure 7: Streams identified as "basins" in the second Pfafstetter subdivision (PF_TYPE <= 2)**
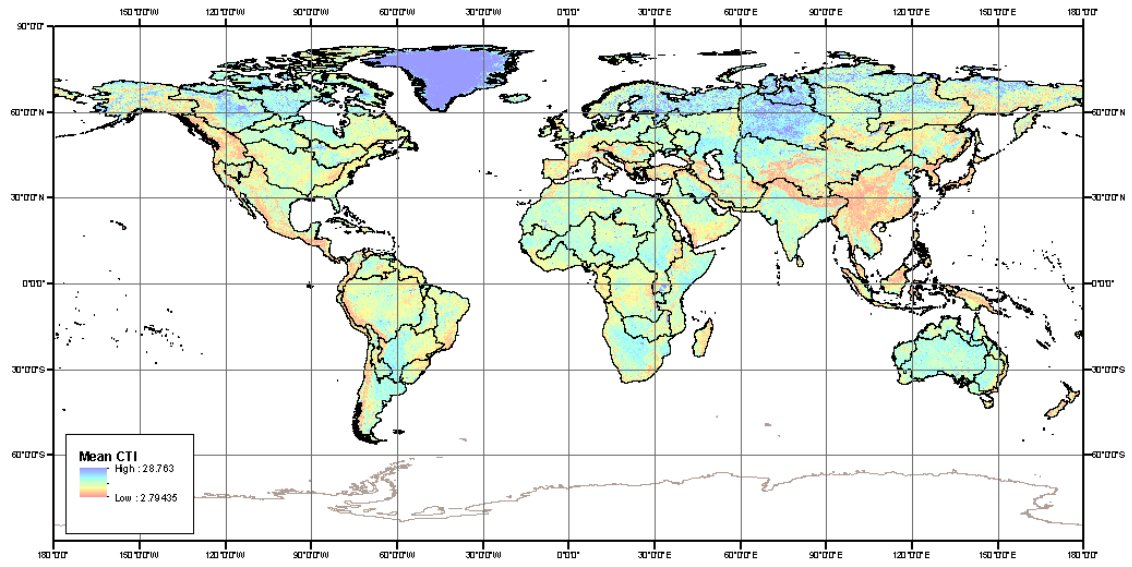
**Figure 8: Mean of Compound Topographic Index (1-arc-minute)**

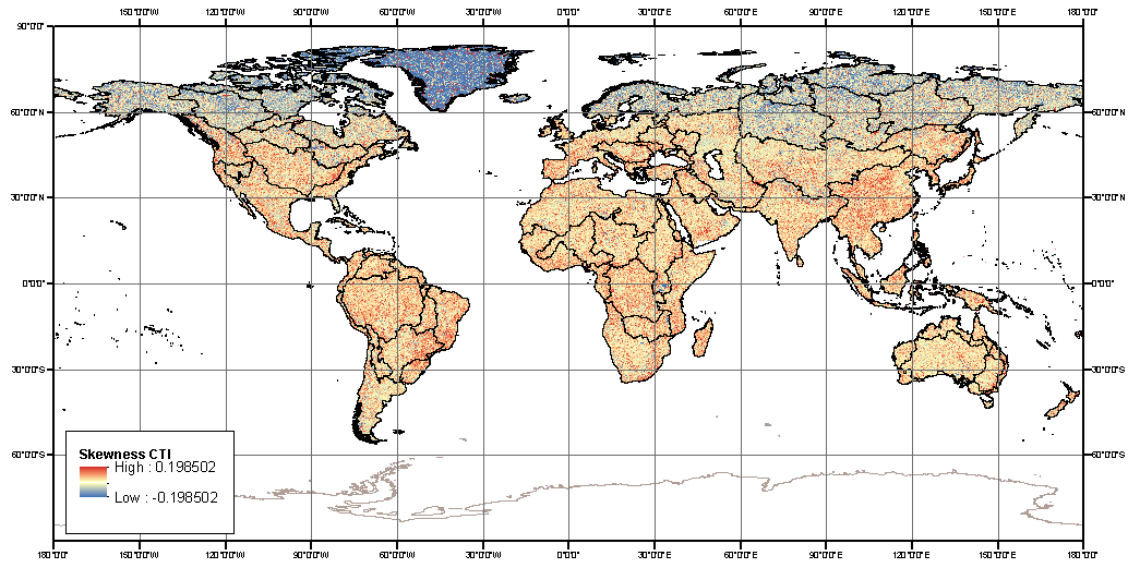**Figure 9: Standard Deviation of Compound Topographic Index (1-arc-minute)**

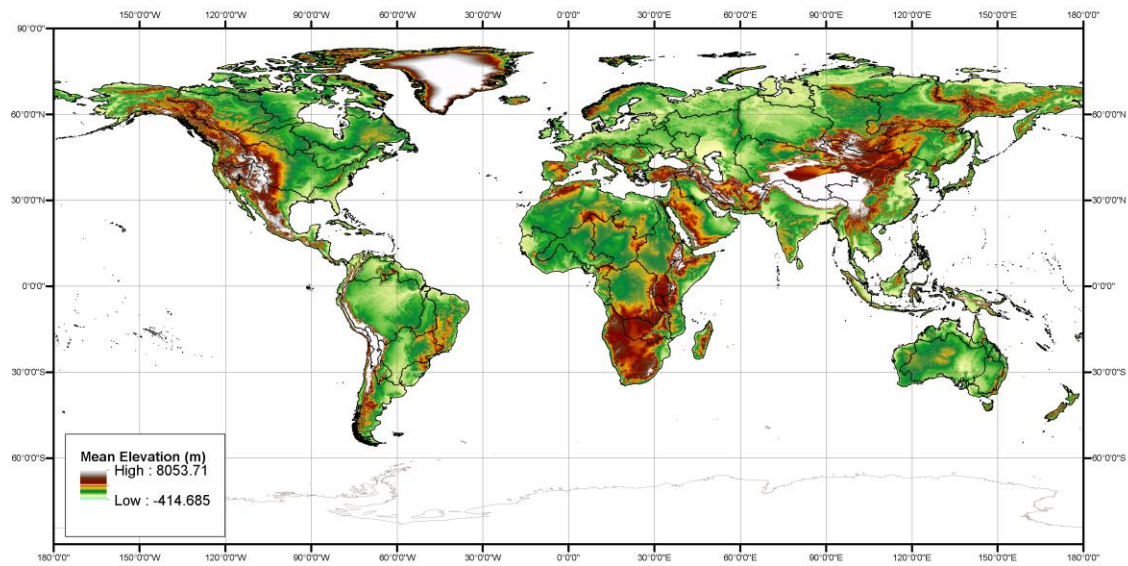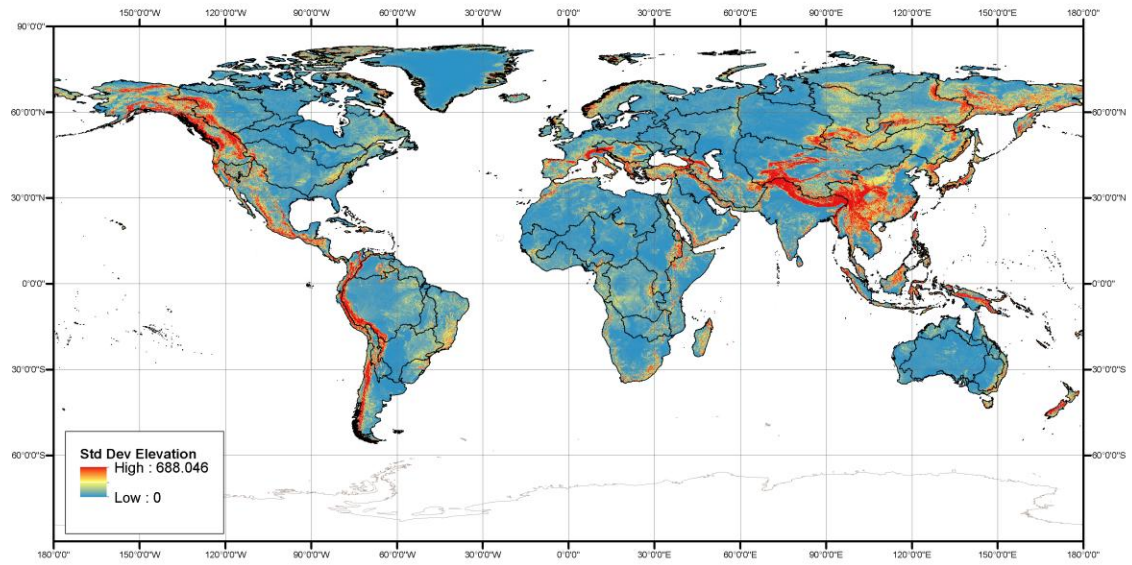**Figure 10: Skewness of Compound Topographic Index (1-**



**Figure 11: Mean Elevation (meters) at 1-arc-minute**

**Figure 12: Standard Deviation of Elevation (1-arc-minute)**

**Appendix 1: Example python script for development of the CTI layer**

```python
# Import system modules
import arcpy, os, time
from arcpy import env
from arcpy.sa import *

timeStart = time.time()
osTimes1 = os.times()

# overwrite
env.overwriteOutput = "TRUE"

# Check out ArcGIS Spatial Analyst extension license
arcpy.CheckOutExtension("Spatial")

# GMTED2010 data
xlength = "f:\\nasa_cti\\basedata\\x_length_75as"
ylength = "f:\\nasa_cti\\basedata\\y_length_75as"
wtGrid = "f:\\NASA_CTI\\BaseData\\wt_grid_75as"

dataPrefix = "f:\\nasa_cti\\fa_processing\\AS\\level"

numberList = ["2","3","4","5","6","7_1","7_2"]

for iNumber in numberList:

    # set environment stuff
    slopeWork = "G:\\NASA_CTI\\AS_Slope\\Level" + iNumber
    print "processing " + slopeWork + " ..."
    env.workspace = slopeWork
    dataPath = dataPrefix + iNumber
    fd = dataPath + "\\flow_dir"
    fa = dataPath + "\\flow_acc"
    tanbeta = slopeWork + "\\tan_beta"
    env.extent = fd
    env.cellSize = fd
    env.snapRaster = fd

    print "Create the diagonal grid..."
    outDiagLength = SquareRoot(Square(xlength) + Square(ylength))
    outDiagLength.save("DiagLength")

    print "Reclassify the FD grid into three classes..."
    # 1 = use y_length
    # 2 = use x_length
    # 3 = use diagonal Length
    outReclass1 = Reclassify(fd,"VALUE",RemapValue([[0,1],[1,1],[16,1],[4,2],[64,2],[2,3],[8,3],[32,3],[128,3]]))
    outReclass1.save("fd_reclass")
    print "Assign the appropriate flow width using Con..."
    outFlowWidth = Con ( outReclass1 == 1, ylength , Con ( outReclass1 == 2 , xlength , outDiagLength ) )
    outFlowWidth.save("flow_width")

    print "Adjust the FA by the area of the upstream pixel and divide by the flow width to get contributing area..."
    outContribArea = Divide ( Plus ( fa , wtGrid ) , outFlowWidth )
    outContribArea.save("ContribArea")

    print "Calculate CTI..."
    outCTI = Ln ( outContribArea / tanbeta)
    arcpy.BuildPyramidsandStatistics_management(outCTI)

    print "Convert to Integer..."
```

```
    intCTI = Int(outCTI * 100 + 0.5 )
    arcpy.BuildPyramids_management(intCTI)
    intCTI.save("cti_int")

timeEnd = time.time()
osTimes2 = os.times()
# print result
utime = osTimes2[0] - osTimes1[0]
stime = osTimes2[1] - osTimes1[1]

print "Clock time was %s seconds" % (timeEnd - timeStart)
print "utime = %s, stime = %s" % (utime, stime)
```