Environment and organisms - Task #262

Organize and document ASTER input DEM files

08/04/2011 12:10 PM - Jim Regetz

Status:	Closed		**Start date:**	08/04/2011
Priority:	Normal		**Due date:**	
Assignee:	Natalie Robinson		**% Done:**	100%
Category:	Terrain		**Estimated time:**	0.00 hour
Target version:				
Activity type:				

Description

Similar to ticket #225, we need to definitively assess and document the ASTER DEM data used as inputs to the global DEM layer.

Ultimately we need to be confident in the provenance of all the data, at least beginning at the point of download from official data providers. This is not now the case, as tiles were obtained via a diverse set of (largely undocumented?) searches conducted over many weeks using some combination of the WIST and Japanese ERSDAC portals, and in some cases local files may have been manually modified/renamed/reformatted/replaced by Reeves.

Additional ASTER assessment issues:

1. Based on file names (which encode lat-lon origin of tile), do we have complete coverage? If not, we need to get the missing tiles, or determine definitively that they do not exist.
    - in all terrestrial areas where SRTM does not exist
    - in a sufficiently large zone of SRTM overlap near the boundary, for blending purposes
2. For all tiles, does the internally specified spatial location match the file name? If not, why not? Is it just a matter of downloading the offending tile again?
3. For all tiles, does the elevation data seem consistent (i.e., with surrounding tiles)? Same followup questions as above.
4. Are nodata values properly encoded, such that they are handled properly in our resampling, mosaicking, and fusion steps?

Rick mentioned having noticed instances of the first three issues, although it's unknown whether the problems were introduced post-download; for what it's worth, I haven't seen any chatter on the web about these sorts of problems with ASTER tiles. The fourth issue has been bothering me in that I think I'm seeing nodata (-9999) values in raw ASTER tiles but not in resampled versions.

History

#1 - 08/04/2011 01:51 PM - Jim Regetz

- Status changed from New to In Progress

Regarding assessment issue 2 above, I wrote up (and committed r119) a quick R script that extracts the lower left X and Y from a given ASTER DEM GeoTIFF using gdalinfo, uses this to construct the expected filename, and compares it to the actual filename. Based on our current holdings in ~organisms/DEM/asterDEM (6677 DEM tifs), there are 7 mismatches:

```
filename               expected
ASTGTM_N59E069_dem.tif   N63 E109
ASTGTM_N63E113_dem.tif   N69 E107
ASTGTM_N63E117_dem.tif   N69 E113
ASTGTM_N64E098_dem.tif   N70 E117
ASTGTM_N65E104_dem.tif   N73 E084
ASTGTM_N65E111_dem.tif   N73 E098
ASTGTM_N65E117_dem.tif   N66 E130
```

I did a little extra inspection for the first one above, ASTGTM_N59E069_dem.tif. We do have a (correctly named) tile at the specified location, i.e. ASTGTM_N63E109_dem.tif, but it does not contain the same data values. However, while examining tiles surrounding both locations, I got lucky and noticed that the offending tile ASTGTM_N59E069_dem.tif contains elevation values identical to tile ASTGTM_N63E110_dem.tif, i.e. the one sitting to the *east* of where the offending tile's internal reference information claims. In summary:

- File name suggests: N59 E069
- Internal spatial reference info suggests: N63 E109
- Actual elevation values match those in adjacent tile: N63 E110

Presumably we need to go back and download a correct N59 E069 tile.

Haven't yet looked at any of the others.

#2 - 08/23/2011 02:33 PM - Natalie Robinson

- Assignee changed from Jim Regetz to Natalie Robinson

There are a few issues with quality control files (num.tif) as well.

1) One file ("DEM/asterGdem/ASTGTM_N49E007_num.tif") contains origin information: origin.x=0, origin.y=0. dem tile for this file ("DEM/asterGdem/ASTGTM_N49E007_dem.tif") is OK. This will be moved to new folder (incorrectNumAug23) in "asterGdem".

2) 217 dem.tif files have no corresponding num.tif files in the directory. These will be moved to new folder (Dem_NoNumFiles) in "asterGdem".

#3 - 08/24/2011 12:11 PM - Natalie Robinson

4 of the 7 bad tiles, as identified by Jim with the aster-check R script, have been re-downloaded and checked. All 4 of these have matching origin and tilename information (yeah!).

The old "bad" tiles have been replaced with these new tiles.

#4 - 08/31/2011 07:42 AM - Natalie Robinson

All dem tiles identified as bad have been successfully replaced with good tiles.

Tiles with no or bad corresponding num.tif files are in the process of being replaced and checked (both the dem.tif and num.tif versions).

I have reorganized the aster tiles for easier visualization in QGIS. This involved:
1) Downloading separate dem files from the USGS for landcover between N40 and N80 latitude.
2) Separating the aster tiles into folders so that each folder contains a USGS dem and the aster tiles that correspond to that coverage
3) Separating the num.tif files for each aster tile into an additional folder, so that only the dem.tif files are in the main folder and the entire folder contents can be uploaded to QGIS in one step.

The final step in verifying the tiles will be to explore the USGS coverages with aster tiles overlaid to check that a) all tiles are present, and b) all tiles contain at least 1% land cover. This last step will be completed by calculating % nodata for whichever tiles contain a questionable amount of land cover, where tiles must contain less than 99% no data to be kept.

#5 - 09/15/2011 09:03 AM - Natalie Robinson

I am almost finished with the check of the Aster tiles. Since the last update I have identified ~100 tiles with less than 1% landcover- these will be removed from the main folders and stored in a separate location in case they are needed at a later time.

I am currently running one final check for data validity: comparing total landmass in the USGS coverages with landmass in the corresponding Aster tiles. Doing this has required that I increase the cell sizes of the Aster tiles to match those of the USGS tiles (attempting to make the USGS cells smaller has failed due to server crashes, error messages saying that I don't have permission to access files, etc.). The procedure I have followed is listed at the end of this update.

I am about 60% through this current task, and have so far found not more than 4.5% difference in total landmass between USGS files and corresponding Aster tiles (the number of Aster tiles corresponding to independent USGS coverages ranges from 234 to 718, so 4.5% difference is at most 11 Aster tiles). The differences in landmass may be an artifact of cell size conversion, or may be an indication that the files do not actually match. To consider is: What is an acceptable landmass difference here?

Procedure for checking landmass:

1) In QGIS: Upload USGS coverage-> Raster-> Clipper-> set extent to so that clipped file has same longitudinal coverage but is limited to latitudes of N59 to N82 (to correspond with Aster tiles).

2) In R: Run the script listed below to aggregate Aster tiles so that the cell sizes match those of the USGS files

```
LandAgg<- c(1:l) * NA
Counts_Agg<- c(1:l)*NA
  for (i in 1:l){
    Counts_Agg[i]<- count(aggregate(raster(Tiles[i]),30),0)+count(aggregate(raster(Tiles[i]),30),-9999)
    LandAgg[i]<-  ncell(aggregate(raster(Tiles[i]),30))-Counts_Agg[i]
    print( paste(round(100*i/l),"%",sep=""), quote=FALSE )
  }
```

3) In R: Use the following script to count the number of cells with land in aggregated Aster tiles and in USGS files, and compare:

```
a<-sum(LandAgg)
AstRast<- raster("DEM/asterGdem/N59to81_W180to141/w180n90/W180N90_clipped.dem")
b<-ncell(AstRast)-cellStats(AstRast, "countNA") #Total # cells- cells with no value
1-(b/a)
```

#6 - 09/21/2011 12:50 PM - Natalie Robinson

- Status changed from In Progress to Closed

- % Done changed from 0 to 100

After speaking with Jim, we decided not to remove the tiles with <1% landcover from the main file library. The Aster tiles available are thus ALL files with ANY landmass.

Re: the issue mentioned above with % landcover calculations differing between USGS coverages and Aster tiles:

```
Jim and I discussed this, and it seems to be an issue to do with the amount of coastline present in a set of A
ster tiles/USGS
     coverage and the resolution at these coastlines (higher resolution files will pick up more landcover and
thus differ in this
     calculation from lower resolution coverages). Despite aggregating the Aster tiles to match the cell size
of the USGS coverages,
     landcover calculations were consistently higher for Aster tile sets (which are higher resolution) than fo
r the corresponding USGS
     coverages. This was especially true where a large amount of coastline existed in the USGS coverage and co
rresponding Aster tiles.
       *In all, USGS gTopo30 coverages differed from Aster tiles by 3.5% in calculated landcover.
```

As for this task as a whole, I have finished checking the Aster files and all are present and appear to be accurate.

The final step in this process was to do a visual check of Aster tiles overlaid on USGS gTopo30 coverages to ensure that all Aster files were present and showed matching landcover to that shown in the USGS files. This visual check did draw my attention to a few tiles that seemed to be duplicates or incorrect. I replaced these tiles, checked for accuracy in their dem.tif and num.tif subfiles, and rechecked them visually in QGIS. In all cases the replacement was successful and the final visualization looked accurate.

<u>Project Summary</u>

```
Steps: 1) Organize Aster tiles for easier use
         2) Use R script (AsterCheck_demAndnum.r) to verify that lower lefthand coordinate of tiles matches f
ilenames (this is how the files
             are supposed to be named) in both dem.tif and num.tif files.
         3) Download USGS gTopo30 coverages for global landmass above N49 degrees latitude
         4) In QGIS, clip gTopo30 coverages to extent: N59 to N82 degrees lat.
         5) Use R script (PctNoLand.r) to calculate the percentage of each Aster tile that is designated "oce
an" (this did not differ greatly
             from percentage of each tile that was designated "ocean" AND nodata- so only % ocean was calcu
lated to save time)
         6) Use same R script to
             a) aggregate Aster tiles by a factor of 30 so that cell sizes matched USGS gTopo30 coverage cel
l sizes
             b) calculate the total landmass for these aggregated tiles
             c) calculate the landmass in the corresponding USGS gTopo30 coverages
             d) calculate the percentage difference in total landmass between the USGS gTop030 coverage and
corresponding Aster tiles
         7) Upload clipped USGS gTopo30 coverages and corresponding Aster tiles into QGIS for final visual ch
eck of tiles compared to outside
             dem coverages

Organization:  The asterGdem file contains Aster tiles for all landmass falling above N49 degrees lat.
             Tiles for landmass below N59 and above N81 degrees lat. can be found in separate folders wit
hin the asterGdem folder
             Tiles for landmass falling between N59 and N81 degrees lat. are separated into different fil
e folders for easier browsing,
                 loading into QGIS, etc.
             Each of these folders contains:
               1) One folder containing a USGS gTopo30 dem coverage, and a clipped version of this covera
ge that only shows landmass
                   between N59 and N82 degrees latitude.
               2) Aster tiles (_dem.tif) corresponding to this clipped USGS dem
               3) One folder containing num.tif files for all Aster tiles in the main folder
               4) A QGIS project file showing the clipped USGS coverage with all Aster (_dem.tif) files o
verlaid
             Another folder contains Aster tiles that were found to be faulty by Rick Reeves in July 2011
             Another folder contains R scripts I used for this project
             Additional documents in the main asterGdem folder provide project documentation

Documentation: The asterGdem file contains the following sources of documentation for this project:
                 PCTNonLandCoverCalcs.ods- List of all Aster tiles and the % of each that is classified "
ocean."
                 AsterTileLog.ods- Complete list of all Aster tiles with which a problem ever occurred du
ring the quality check process,
                     what that problem(s) was/were, when the tile was replaced, and the outcome
                 LandCoverDiffs_AsterVsGtopo30.ods- List of total landcover calculated for all Aster tile
s and USGS gTopo30 coverages, by
```

coverage extent, with % difference between them and number of Aster tiles included in calculation.

Three .txt and one .sh file that were created by Rick Reeves before my involvement in the project