WRF Hydro Routing Grid Pre-Processor Python Toolbox Documentation-version2.0

Prepared By: Kevin Sampson and Dave Gochis, NCAR/Research Applications Laboratory

Last Updated: April 9, 2014

Purpose

This document describes the function and use of a stand-alone pre-processing utility that is designed to assist users in the creation of WRF-Hydro routing grids ('data layers') using an ArcGIS Geographic Information System (GIS) from esri®.

Overview

The processing that will soon be available on the WRF Hydro website (http://www.ral.ucar.edu/projects/wrf_hydro/) via web processing service (WPS) is available to users as an ArcGIS Python Toolbox. Python toolboxes were implemented in ArcGIS Desktop version 10.1 as a way to create custom geoprocessing tools directly from Python scripts. All ArcGIS 10.1 installations come with Python installed by default and will be able to view this Python toolbox. The only additional requirement is for the user to have the ArcGIS Spatial Analyst Extension activated:

For More information on how to activate Spatial Analyst Extension please visit:

> http://resources.arcgis.com/en/help/main/10.1/index.html#/Using_Extensions_in_ArcGIS/0189
> 00000002000000/

Spatial Analyst: http://www.esri.com/software/arcgis/extensions/spatialanalyst

Summary of Software Requirements:

- ArcGIS Desktop 10.1
- Spatial Analyst Extension

Few tools are infallible and this is true for the WRF-Hydro pre-processing tool. There may be instances when the output from the tool is not precisely what was intended for a given application. The most common issue that arises is related to how stations are mapped (or 'snapped') to a channel network and, in turn, how watersheds get defined from those locations. Should you run into this issue, please refer to the section below entitled 'Manual specification of station points' for guidance.

Tool Purpose:

The purpose of this tool is to create the data layers for terrestrial overland flow, subsurface flow and channel routing processes required by WRF Hydro. These output data layers and the input requirements are all described below.

Additional Utility Tools:

A few additional 'Utility' scripts have been created to facilitate working with geogrid files in ArcGIS and to aid in the preparation or review of output from the WRF-Hydro pre-processing tools. Those Utilities are contained in the 'Utilities' toolset within the Python Toolbox. These Utility tools are described in the 'Miscellaneous Topics in the use of the WRF-Hydro Pre-processor' section below.

Sample Input and Output Data:

Sample input data and output data are provided so that users can see both how to format input data for the WRF-Hydro pre-processing tool and so they can see what the output should look like. These sample data are contained within the 'Standalone_Test_Data' zipfile and embedded directory in the distributed tarfile.

Using the tool:

To view and open the toolbox, unzip the package to a location on your disk. From ArcCatalog or the ArcCatalog tab within ArcMap, navigate to the location Standalone_Tool folder (See Fig. 1). The Python toolbox exists as a *.pyt* file, and will behave like a regular ArcGIS toolbox (*.tbx*). The "GEOGRID_STANDALONE" toolbox contains a "Processing" toolset, containing a single "ProcessGeogridFile" script tool. There is also a "Utilities" toolset, containing three script tools for performing additional functions. Users may feel free to customize the toolbox and add any scripts or additional functionality necessary for processing WRF Hydro spatial data.
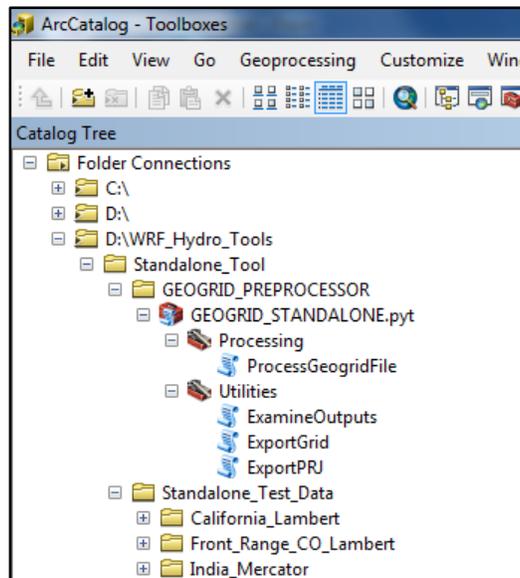
Figure 1. Catalog Tree view of the Python Toolbox in ArcCatalog.

Executing the Tool:

Open the "ProcessGeogridFile" script tool by double-clicking on it, or right-click and select **Open**. A tool dialog will open, with two required inputs, as well as optional inputs and several default parameter values supplied. Once all required and desired inputs are given, simply click **OK** to execute.

Description of Tool Parameters
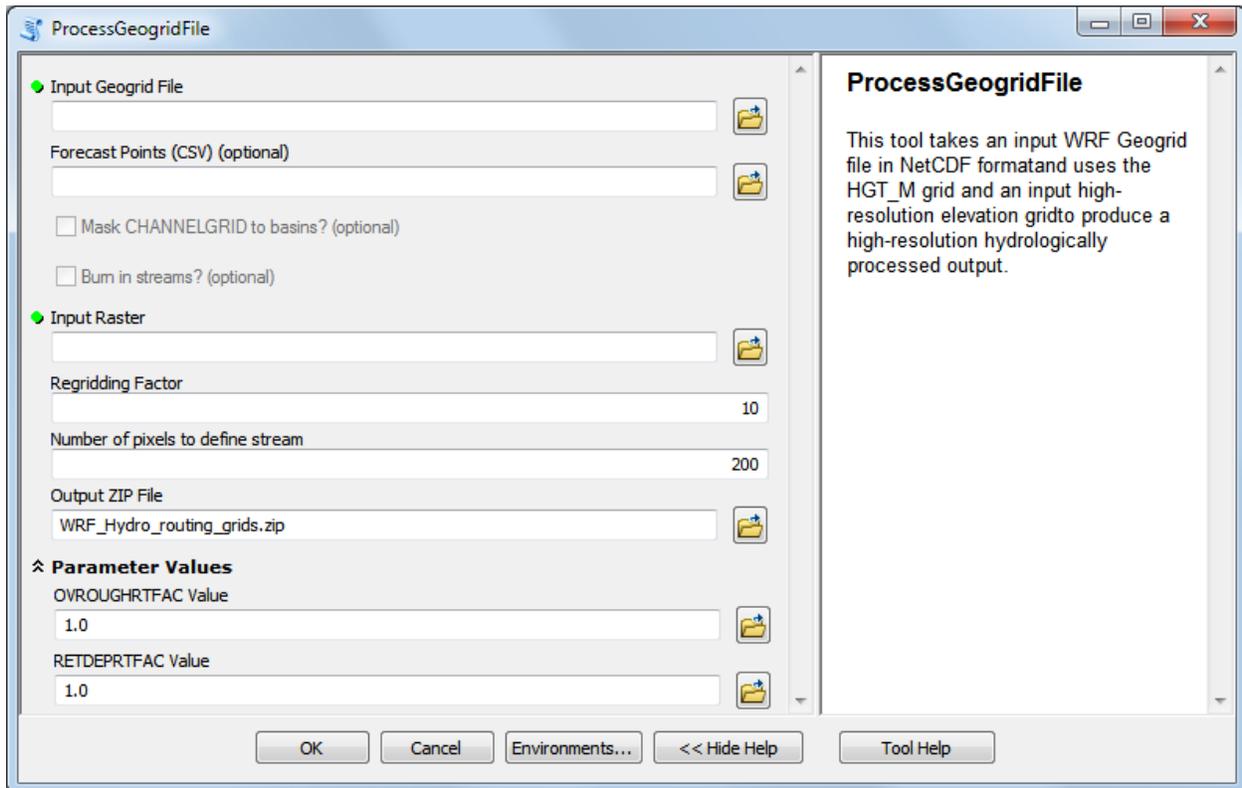
Input Geogrid File:

This is the input WRF geogrid file in NetCDF (*.nc*) format. This file must contain dimensions "west_east" and "south_north", as well as variable "**HGT_M**", which is the domain used for terrain processing. Supported projections currently include MAP_PROJ = 1 (Lambert Conformal Conic) and MAP_PROJ = 3 (Mercator).

Forecast Points (CSV) (optional):

This optional parameter requires a Comma-separated Values (CSV; *.csv*) format file of gage locations in latitude/longitude coordinates (WGS84). A CSV file is a comma-separated ASCII file containing a 1 row header with gage location information on subsequent rows. The CSV file must contain a longitude field named "LON" and a latitude field named "LAT" in order to run, and no header names are allowed to contain spaces. An example is given below:

```
FID,LON,LAT,STATION,Name
15,-105.92833,40.08139,Fraser_at_Granby,9033300
18,-105.9,40.12083,COLO_nr_GRANBY,9019500
20,-106.3333,39.8803,Blue_R_blw_Grn_Mtn,9057500
```

If a CSV file is provided, the output will contain a "frxst_basns.nc" output file, which is a grid of the basins delineated from just below the gage locations.

Mask CHANNELGRID to basins? (optional)

This Boolean (TRUE/FALSE) parameter asks the user whether or not to mask the output "CHANNELGRID.nc" grid file to the basins provided in the Forecast Points CSV file. The option is only available if a CSV file is provided in the previous parameter. If TRUE (Checked), the derived gauged basins will be used to mask the CHANNELGRID file. Masking out unwanted channel systems can result in a significant computational savings.

Burn in streams? (optional)

This Boolean (TRUE/FALSE) parameter is reserved for future use and will not be available to the user in the v2 tool release.

Input Raster

This required parameter allows the user to provide a high-resolution elevation grid (digital elevation model-DEM) from which to derive the output layers. This grid must be an ArcGIS supported raster format, have the coordinate system defined, and cover the entire extent of the input geogrid file domain. (For example: The test cases included with the download package have this DEM stored in a geotiff format. Other formats such as Arc rasters are acceptable.) Additionally, the terrain processing will very likely only be successful if the input DEM has been hydrologically processed to ensure continuous flow paths. Note: The vertical units must be in meters above sea level (m). Supported datasets include the 3 arc-second (90m) USGS HydroSHEDS dataset (http://hydrosheds.cr.usgs.gov/index.php) and 1 arc-second (30m) NHDPlus elevation data derived from the National Elevation Dataset for the Continental United States (http://www.horizon-systems.com/NHDPlus/NHDPlusV2_data.php).

Regridding Factor

The required regridding factor parameter allows the user to set the output cell size for the derived datasets based on a relationship with the cell size in the input geogrid file. The output high-resolution datasets must be able to nest perfectly within the coarse geogrid resolution, and so the coarse grid resolution will be divided by the regridding factor. An example could be a 1km input geogrid file: A factor of 10 results in a 100m output, 5 results in a 200m output, and 4 in a 250m output cell size.

Number of pixels to define stream

This is the minimum number of pixels to initiate a stream segment. The smaller this number is, the higher drainage density the output CHANNELGRID file will have. When defining the minimum number of cells for defining a stream, remember that the area depends on the resolution of the grid. If the grid is 1000m x 1000m, then each cell is $1km^2$. This definition will

greatly alter the size of the smallest watersheds delineated and will control the density of the channel network to be created.

Output ZIP File

This required parameter is the directory and filename (*.zip*) of the output from the tool. This can be set to any location and filename that you choose, as long as you have read/write access to this location. The output format will be a ZIP archive.

Parameter Values

There are two parameter values: OVROUGHRTFAC and RETDEPRTFAC. The default values for these grids are 1.0 and 1.0 respectively.

Description of Output Files

All output files are in NetCDF format and contain a variable with the same name as it's filename. The x dimension = 'x' and the y dimension = 'y'. Below is the list of NetCDF output data files that are created by the tool. Please refer to the WRF-Hydro documentation for a more thorough description of these files.

CHANNELGRID.nc - The channel grid. Channel pixels = 0, non-channel pixels = -9999. If the "Mask Channelgrid to basins?" option is set to True (checked), the output will be masked to the gaged basins provided, while non-gaged channels are given a value of -1.

flowdirection.nc – Flow direction grid. The values for each direction from center are:

32	64	128
16		1
8	4	2

frxst_basns.nc – Delineated basins. If a gage CSV file is provided, catchments will be delineated from a point that is 3 pixels downstream of the gage's coordinates. If masking of the channelgrid is selected, this is the layer that will be used to perform the masking. Basins are numbered according to the values in the 'FID' field. NoData gridcells have values of -9999.

frxst_pts.nc – Gage location grid. Each grid will have a pixel identified where the gage is located. Gage pixels are numbered in the same way as the frxst_basns grid. NoData gridcells have values of -9999.

gw_basns.nc – Groundwater basins grid. The default values in this grid are currently identical to the frxst_basns.nc file. NoData gridcells have values of -9999. See main WRF-Hydro documentation for a description of this file.

gw_basns_gegrid.txt - This is an ASCII raster of the gw_basns grid, regridded to the geogrid file resolution. NoData gridcells have values of -9999.

LAKEGRID.nc – Lake grid. The default values in this grid are currently set to -9999. See main documentation for the procedure to define lakes in WRF-Hydro.

landuse.nc – This is the same grid as the "LU_INDEX" variable in the input geogrid file, but resampled using Nearest Neighbor assignment to the resolution of the routing grid.

latitude.nc – Grid of latitude values, in the geographic coordinate system based on a sphere of radius 6370000.0m (EMEP sphere, wkid: 104128).

longitude.nc – Grid of longitude values, in the geographic coordinate system based on a sphere of radius 6370000.0m (EMEP sphere, wkid: 104128).

ovroughrtfac.nc – OVROUGHRTFAC parameter. Currently set to 1.0.

retdeprtfac.nc – RETDEPRTFAC parameter. Currently set to 1.0.

str_order.nc – Stream order grid, calculated using the Strahler method. NoData gridcells have values of -9999.

topography.nc – High resolution elevation grid. Units are the same as the Input Raster provided.

Miscellaneous Topics in the Use of the WRF-Hydro Pre-processor

WRF-Hydro Pre-Processor Utility: 'ncks_file_concatenator.sh', concatenating individual files together:

WRF-Hydro expects all of the routing grids created above to be in a single NetCDF file. Therefore the individual data files need to be concatenated together. This is simply done use a shell script which calls several 'NetCDF command-line operators' (nco commands). The script is called 'ncks_file_concatenator.sh' and it is included in the tool ZIP package. This script must be run under the LINUX/UNIX environment. Therefore users need to copy this script and the Zipfile of individual NetCDF files to a common directory, unpack and run the script. The script will produce a single NetCDF file with the necessary data layers.

WRF-Hydro Pre-Processor Utility: 'ExamineOutputs', ArcGIS script to import pre-processor output NetCDF files to EsriGRID raster format:

The 'ExamineOutputs' utility script extracts the individual NetCDF files from the ZIP file created by the WRF-Hydro pre-processor and converts those to EsriGRID format rasters. This allows users to easily view and manually manipulate the rasters in ArcGIS. Each *.nc* file in the output ZIP archive will be converted to a raster, including the ASCII raster (gw_basns_geogrid.txt).

WRF-Hydro Pre-Processor Utility: 'ExportGrid', ArcGIS script to export gridded variables from the NetCDF geogrid file to EsriGRID or other raster format:

The 'ExportGrid' utility allows the user to import any data layer on the Mass Grid (: stagger = "M";) from a WPS-created NetCDF 'goegrid' file into any ArcGIS-writable raster format. The tool automatically determines the projection based on attribute data in the geogrid file.

WRF-Hydro Pre-Processor Utility: 'ExportPRJ', ArcGIS script to export projection file from Mass grid of the NetCDF geogrid file:

The 'ExportPRJ' utility takes the WRF geogrid file as input and automatically determines the projection based on the global attribute data to produce a *.prj* projection file. This file allows the user to build data and grids on the projection of the WRF domain using standard GIS applications. The projection information based on the Mass Grid (: stagger = "M";) from a WPS-created NetCDF 'geogrid'.

Script Customization:

Users may modify the Python script as they wish though it is a good idea to keep a copy of the original version should your modification efforts not fare so well. To open the script from within ArcGIS, right click on the toolbox and select **edit**. The default Python editor will open (IDLE) and the script may be edited. You may have to customize your Python installation in order to recognize *.pyt* files (see http://blogs.esri.com/esri/arcgis/2012/12/14/how-to-debug-python-toolboxes-in-3-easy-steps/ for more information).

The input LSM geogrid file currently contains the variable "HGT_M" with is a grid of coarse elevation data necessary to initiate watershed processing. The dimension variables "west_east" and "south_north" contain the X and Y dimensions, respectively. The global attributes "corner_lats", "corner_lons", "TRUELAT1", "TRUELAT2", "STAND_LON", and "CEN_LAT" contain the values for the corner latitude, corner longitude, standard parallel 1, standard parallel 2, central meridian, and latitude of central origin, respectively that are used to produce the

projection file and properly project the raster. If the names of any of these elements changes in the future, the script will need to be modified to reflect that change.

Manual specification of station points:

This section describes the steps needed to improve the 'snapping' of station points to the derived channel grid. Because of inaccuracies in the spatial location information of station data and digital elevation models and because of small errors introduced when projecting geospatial data from one coordinate system to another, the situation occasionally occurs where streamflow observation/gauging points specified in the station *.csv* file do not get properly assigned to the desired location on the channel network. The result of this error is the generation of erroneous basins or watersheds and erroneous extraction of flow values from the WRF-Hydro channel routing components. Because these errors are often random, or non-systematic, there is no generalizable way to automate the correction procedure with a high degree of fidelity and manual manipulation or specification of the data is often required. This situation is very common in hydrographic data processing and well known to hydrologists. The approach presented here to deal with this issue seeks to present a relatively simple way to determine the correct locations of the stations for the channel network that is created from the WRF-Hydro pre-processing tools.

The steps to improve the assignment of station locations in ArcGIS are as follows:

1. Import the original *.csv* station file using the File->Add Data-> Add XY Data drop-down tool. Upon ingest, be sure to define the projection as Geographic WGS84. Export the imported data layer as a shapefile.

2. Close and re-open ArcGIS to re-initialize the coordinate framework.

3. Assuming you have already executed the pre-processing tool once with a user-specified set of station observation points CSV file and successfully created a Zipfile with the individual NetCDF data layers, you'll need to extract the topography (topography.nc), basin mask (gw_basns.nc) and channel network (CHANNELGRID.nc) NetCDF files from that Zipfile. **VERY IMPORTANT:** Users can also use the '**ExamineOutputs**' utility script described above to automatically convert all NetCDF output layers to EsriGRID rasters.

4. Import these NetCDF datafiles into ArcGIS using the Multidimension->Make NETCDF Raster Layer tool. The default projection will be the same as that in the geogrid file. **VERY IMPORTANT:** Users can also use the '**ExamineOutputs**' utility script described above to automatically convert all NetCDF output layers to EsriGRID raster format.

5. Add the station point shapefile created in #1 above to the map.

6. Open a copy of the original *.csv* station file in Microsoft Excel or your preferred spreadsheet editor.

7. Zoom into each station's location that you need to modify. If not already set, you'll need to set the displayed coordinates (i.e. the location information in the lower right-hand corner of the ArcGIS map window) to 'Decimal Degrees' [This can be done by right-clicking on the 'Layers' icon in the ArcGIS Table of Contents window and navigating as follows: Properties->General->Units>Display->Decimal Degrees]. As you zoom into your 'problem' station location you'll likely notice that the station is not on the channel network where you wish it to be.

8. Pan your cursor over the exact location of where you want the station to be on the derived channel network. Again, due to subtle errors associated with projection of geospatial data this new location may not be the exact location your original data specified your station to be. Such is life with GIS, but the point is to make sure the station location is where you want it on the *derived* channel network. Once you have determined that location, insert those exact coordinates into the LAT and LON fields of the *.csv* file you have open for editing. An easy way to get these coordinates is to use the 'Identify' cursor option in ArcGIS which provides information when you click on a grid-point. Simply click on the gridpoint where you desire to have the station, then select, copy, and paste the latitude and longitude coordinates from the "Location" field in the 'Identify' window into your new *.csv* file. You will then use this new CSV file when you re-run the WRF-Hydro pre-processing tool.

9. Repeat this search-edit process for each station in question.

10. When you have finished editing the station points, close all data layers in ArcGIS (or simply restart ArcGIS) and your newly edited CSV station file.

11. Re-run the WRF-Hydro preprocessing tool exactly as specified above EXCEPT make sure you use the new, modified CSV station file whose LAT/LON locations have been edited to match the desired locations on the derived channel network. The new set of data layers created should produce the station forecast points and watersheds that are desire. If not, you'll need to inspect the locations again more closely and make further adjustments as necessary.

Note about large domains

Many intermediate rasters are held in memory during processing for quick access. However, the ability to store raster data in memory is limited by the memory capacity of the machine performing the processing. Some large domains will require substantial resources when high resolution elevation datasets are used over very large domains. Be aware that this limitation may manifest itself through non-descript error messaging upon failure of the process. Always check the tool's results and messages by viewing the 'Results' tab in ArcMap or ArcCatalog for more information.

Structure of the scripts in the Pre-Processing Tool

The WRF-Hydro preprocessing tool is split between two python scripts. The first is the GEOGRID_STANDALONE.pyt file, which contains the classes and syntax necessary to construct the Python Toolbox and handle parameters. The second script, wrf_hydro_functions.py, contains all of the functions necessary to process the geogrid file. The scripts must be located in the same directory at all times.

The WRF-Hydro preprocessing tool workflow proceeds as follows:

1) The input WRF geogrid file variables and attributes are accessed and a spatial reference object is constructed from the projection parameters, along with a properly georeferenced grid (variable: HGT_M) from the input file. This function is called "georeference_geogrid_file".

2) High Resolution topography is created for the domain, based on the input elevation raster. A high resolution raster layer is created from the input elevation raster and projected and then resampled to the output resolution provided by the "Regridding Factor" parameter. The WRF geogrid resolution is divided by the "Regridding Factor" to arrive at the routing grid resolution. This function is called "create_high_res_topogaphy"

3) The WRF grid is used to create grids of both latitude and longitude, according to the geographic coordinate system defined on the WRF sphere (radius = 6370000.0m). A reprojection is used to convert from the projected WRF coordinate system to a geographic coordinate system on the same sphere. This process uses the ArcGIS 9.3 python module "arcgisscripting", which is available through backward compatability from the "arcpy" module. This function is called "create_lat_lon_rasters"

4) Spatial analysis functions are performed on the high resolution elevation grid (created in step 2). The high-resolution elevation data is hydrologically processed in this step. First, the elevation layer has all depressions filled using the "Fill" algorithm. Next, flow direction and flow accumulation are calculated. The basin size threshold is used to create the channelgrid layer. Stream order is then calculated from the channelgrid layer. A variety of rasters of constant value are created from input and default parameters. If a gages CSV file is provided, additional layers are calculated after applying the "Snap Pour Point" and "Watershed" Spatial Analyst tools. Finally, all layers are converted to NetCDF format, compressed (*.zip*) and returned to the user. This function is called "sa_functions".

5) Intermediate files and the temporary 'scratchdir' directory are deleted.

Future Compatibility with TauDEM Funcitons

The use of more sophisticated terrain processing techniques is being explored. In the future, terrain processing may include options for processing using TauDEM version 5 (Terrain Analysis Using Digital Elevatoin Models). TauDEM is a suite of digital elevation model (DEM) tools for the extraction and

analysis of hydrologic information from topography as represented by a DEM (http://hydrology.usu.edu/taudem/taudem5/index.html). These functions will require the TauDEM tools to be installed on the local machine executing the WRF-Hydro pre-processing tool (in addition to the ArcGIS requirements). The functions are called by the WRF-Hydro pre-processing Python Toolbox, but are contained in a separate script, "TauDEM_py.py". Additional functionality includes the ability to choose output from the D-Infinity or Peuker-Douglas algorithms, and check for edge contamination in the output channelgrid and basin grids. The many additional customization parameters for tuning output of TauDEM functions are set in the header of the "TauDEM_py.py" script, and can be modified by altering the values of those variables.