WRF Hydro GIS Pre-Processing Tools, Version 4.0

Documentation

Kevin Sampson and David Gochis

National Center for Atmospheric Research
Research Applications Laboratory
P.O. Box 3000
Boulder, CO 80307

Updated: May 1, 2017

Contents

Acknowledgements

This work was financially supported by the U.S. National Science Foundation, the U.S. National Weather Service, the National Aeronautics and Space Administration and the Colorado Water Conservation Board.

1 Summary

This document describes the function and use of a stand-alone Python pre-processing utility that is designed to assist users in the creation of WRF-Hydro routing grids using Esri® ArcGIS Geographic Information System (GIS) software. As of version 4, all WRF Hydro GIS Pre-processor version numbers will be linked to the WRF-Hydro version that they support.

2 Overview

The processing workflow for creating WRF-Hydro routing grids 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 and newer installations come with a version of Python installed by default and will be able to view this Python toolbox, though there may be some python toolbox parameter incompatibilities with ArcGIS 10.1. With the ArcGIS 10.2 release, the netCDF4 Python module is packaged with the version of Python installed by ArcGIS. The authors recommend using ArcGIS 10.3.1. Additionally, the user must have the ArcGIS Spatial Analyst Extension activated:

NOTE: Due to BUG-000096495, ArcGIS for Desktop versions 10.4 and 10.4.1 will not successfully execute this tool.

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

http://desktop.arcgis.com/en/arcmap/latest/extensions/spatial-analyst/enabling-the-spatial-analyst-extension.htm

Spatial Analyst:

Product page:
http://www.esri.com/software/arcgis/extensions/spatialanalyst

Web help:
http://resources.arcgis.com/en/help/main/10.2/index.html#//005900000001000000

Summary of Software Requirements:

- ArcGIS Desktop 10.3.1 Basic, Standard, or Advanced
- Spatial Analyst Extension

Few tools are infallible and this is true for the WRF-Hydro GIS 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 ‘9.1 Manual Specification of Station Points’ for guidance.

2.1 Tool Purpose

The purpose of the WRF Hydro GIS Pre-Processing Tool is to create the data layers for terrestrial overland flow, subsurface flow and channel routing processes required by WRF Hydro. The outputs from these tools are geospatial and tabular data layers. The input requirements are all described below.

2.2 Toolbox Design

The toolbox is split across two Python scripts:

\GEOGRID_STANDALONE.pyt

\ wrf_hydro_functions.py

The '*GEOGRID_STANDALONE.pyt*' script is the python toolbox that contains code that ArcGIS uses to handle and validate parameters. This tool calls functions in the '*wrf_hydro_functions.py*' script. The separation of the executing tool from the functions allows for a modular system of tools to be built that can call functions individually. Thus, many of the tools in the 'Utilities' toolset utilize single functions or a subset of the entire GIS pre-processing workflow. This design provides consistency across tools. Multiple toolboxes may be developed, all utilizing common code in the function script. The sharing of code reduces redundancy between tools and allows for rapid development of new tools and utilities.

2.3 Additional Utilities

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 GIS pre-processing tools. Those utilities are contained in the 'Utilities' toolset within the Python Toolbox. These tools are described in Section '7 Other utilities', below.

2.4 Sample 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' Zip archive.

2.5 64-bit Architecture

ArcMap and ArcCatalog are 32-bit Windows applications, and by default all geoprocessing is done in 32-bit mode. Processes may either run in the foreground or in the background. The ArcGIS Background Geoprocessing (64-bit) installation allows users to utilize 64-bit background geoprocessing. All of the WRF-Hydro GIS pre-processing tools are capable of running in background as well as foreground mode. For more on background geoprocessing, see the following links.

Foreground and Background Geoprocessing:
http://desktop.arcgis.com/en/arcmap/latest/analyze/executing-tools/foreground-and-background-processing.htm

Background Geoprocessing (64-bit):
http://desktop.arcgis.com/en/arcmap/latest/analyze/executing-tools/64bit-background.htm

ArcGIS Blog - 64-bit vs. 32-bit Python Explained:
http://blogs.esri.com/esri/supportcenter/2013/07/29/64-bit-vs-32-bit-python-explained/

In order to utilize 64-bit background geoprocessing, Background Geoprocessing (64-bit) must be installed and background geoprocessing must be enabled in the ArcMap or ArcCatalog session. The user must also ensure that the Python scripts are being run against a 64-bit Python installation. A check was put into the *Process GEOGRID File* tool and the log file created by this tool will record whether or not 64-bit processing is active. If these conditions are met, line 2 of the log file will read “64-bit: True”. It is not necessary to run the WRF Hydro pre-processing tools in 64-bit or in the background. Simply disable background geoprocessing to ensure the script tools run in 32-bit mode in the foreground.

NOTE: Occasionally with large grids, a MemoryError will occur related to memory allocation limitations in 32-bit mode.

2.6 Tool Testing

Testing was carried out on a Windows 8.1 Enterprise 2.59 GHz Intel Core i7 PC, 16GB RAM, ArcGIS 10.3.1.4959 with Background Geoprocessing (64-bit).

3 Building the Routing Grids

WRF-Hydro routing functions are executed on a different set of grids than the column land surface model (LSM) grid. The column LSM grid is the same as that of the WRF model when WRF-Hydro is run in a fully coupled mode with WRF. Because the WRF model pre-processors contain a number of functions that aid in the generation of the LSM grid users may utilize the WRF WPS 'geogrid.exe' program to generate GEOGRID files which define the domain, resolution, and other spatially varying and static grids used by the column LSM and the WRF model. For more on GEOGRID, see:

http://www2.mmm.ucar.edu/wrf/users/docs/user_guide_V3/users_guide_chap3.htm

Although WRF-Hydro routing functions are typically run on a finer grid than the land model, the two different grids must be compatible. For example, the WRF-Hydro routing grids have the same extent as the WRF grids, and the routing grids must nest exactly within the land model grid. All routing grids must share a common domain extent, cellsize, and dimensionality (rows & columns). The WRF-Hydro GIS pre-processing tools handle the building of the WRF-Hydro routing grids based on user input parameters and an existing WRF GEOGRID file.

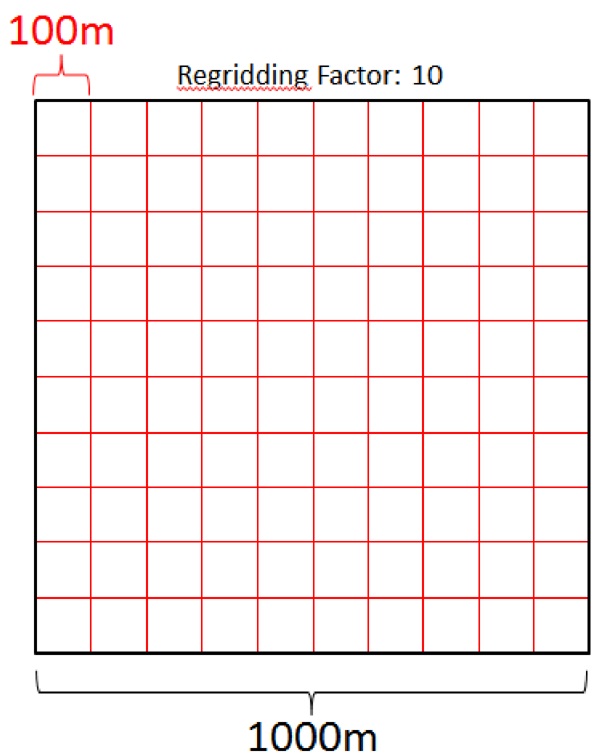

Figure 1. Nesting of WRF-Hydro routing grid cells (red boxes) within a WRF grid cell (black box).

The regridding factor shown in Figure 1 is an integer used to divide the WRF cellsize and produce the routing resolution. In this example, a WRF cellsize (DX, DY) of 1000m and a regridding factor of 10 yields 100m cells for the WRF-Hydro routing grid. Refer to Table 3 for more information.

4 Geospatial Considerations

Dealing with terrain data is a highly spatial process, to which GIS applications are well suited. Although much of the processing done on the grids is not spatial, knowledge of the coordinate reference system (CRS) and ability to translate between grids with different CRS is important. The WRF model is configured to run using a limited set of projected coordinate systems, yet many stream gages are georeferenced using a geographic coordinate system such as 'WGS84'. Further, most elevation datasets are produced using varying regional projected or geographic coordinate systems and must be re-projected to the WRF CRS in order to be used. This section will deal with the geospatial considerations used in the WRF-Hydro GIS pre-processing tool.

4.1 Coordinate Systems

The WRF model defines the CRS of its domains using global attributes in the GEOGRID file. The coordinate system is interpreted using the 'MAP_PROJ' global attribute from the GEOGRID file. Values of the MAP_PROJ attribute and the associated projected coordinate system name are given in Table 1. Once the coordinate system is identified, additional attributes are read to obtain the parameters that define the unique WRF domain projection.

MAP_PROJ	Projection Name
1	Lambert Conformal Conic
2	Polar Stereographic
3	Mercator
6	Cylindrical Equidistant

Table 1. WRF-supported projected coordinate systems

GEOGRID Attribute	Projection Parameter	Associated Projection
TRUELAT1	Standard Parallel 1	All (required)
TRUELAT2	Standard Parallel 2	Lambert Conformal Conic (optional)
STAND_LON	Central Meridian	All (required)
CEN_LAT	Latitude of Origin	Lambert Conformal Conic (required)
POLE_LAT	Pole Latitude	Cylindrical Equidistant (optional)
POLE_LON	Pole Longitude	Cylindrical Equidistant (optional)

Table 2. WRF GEOGRID projection parameters

All of the supported WRF coordinate systems may be used, with the exception of a configuration of the Cylindrical Equidistant projection which allows for the North Pole to be rotated away from 90N, 0E. This configuration is known as 'rotated pole', and is not supported by the ArcGIS Projection Engine. If encountered, any tool in the toolbox that attempts to read the GEOGRID file will produce a warning and exit.

The GEOGRID file may or may not contain all of the projection parameters listed in Table 2. In the course of generating a projection definition from the GEOGRID file, the projection type ('MAP_PROJ') is read first, and any available projection parameters are then read. An Esri® projection string is generated based on the projection and available parameters, which can be converted to a spatial reference object in the Python ArcGIS API. This spatial reference object is used to define the CRS of all WRF-Hydro grids. Examples of the Esri® and PROJ.4 definitions of the same WRF CRS are listed below:

Example Esri® Projection String:

```
PROJCS['Sphere_Lambert_Conformal_Conic',GEOGCS['GCS_Sphere',DATUM['D_Sphere',
SPHEROID['Sphere',6370000.0,0.0]],PRIMEM['Greenwich',0.0],UNIT['Degree',0.017
4532925199433]],PROJECTION['Lambert_Conformal_Conic'],PARAMETER['false_eastin
g',0.0],PARAMETER['false_northing',0.0],PARAMETER['central_meridian',-
105.0],PARAMETER['standard_parallel_1',38.25],PARAMETER['standard_parallel_2'
,41.0],PARAMETER['latitude_of_origin',39.7632102966],UNIT['Meter',1.0]];-
37188300 -29513500 10000;-100000 10000;-100000
10000;0.001;0.001;0.001;IsHighPrecision
```

PROJ.4 String:

```
+proj=lcc +lat_1=38.25 +lat_2=41 +lat_0=39.7632102966 +lon_0=-105 +x_0=0
+y_0=0 +a=6370000 +b=6370000 +units=m +no_defs
```

4.2 A note on GEOGRID projection parameter dependencies

The dimension variables 'west_east' and 'south_north' contain the X and Y dimensions, respectively, in the GEOGRID file. The global attributes 'MAP_PROJ', 'corner_lats', 'corner_lons', 'TRUELAT1', 'TRUELAT2', 'STAND_LON', and 'CEN_LAT', 'POLE_LAT', and 'POLE_LON' contain the values for the map projection, corner latitude, corner longitude, standard parallel 1, standard parallel 2, central meridian, latitude of central origin, pole latitude, and pole longitude, respectively. These are used to produce the PRJ file (*.prj*) 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 the change.

Note: Even if WRF does not use a particular projection parameter for the coordinate system, such as STAND_LON in the case of the Mercator projection ('MAP_PROJ'=3), ArcGIS will need a valid value.

4.3 Sphere vs. Datum

The WRF model, as with many other models, is executed assuming a spherical Earth. Generally, atmospheric circulation models use a spherical Earth model while terrestrial geospatial data uses a spheroidal earth model. In the case of WRF ARW, the Earth radius is 6,370,000m. However, most input data, such as elevation or landuse, is specified using a using a spheroidal (elliptical) datum to account for the flattened shape of the Earth, which is closer to an oblate spheroid. Appropriate care must be taken by the user to limit positional errors that may be introduced into the model because of the differing Earth models (sphere or spheroid) used in the input data. This topic has been examined by many, though David et al (2009) and Monaghan et al (2013) discuss the topic in relation to hydrological modeling and WRF, respectively. As noted in Monaghan et al (2013), users of models such as WRF should ensure that the terrestrial input data are *consistently* mapped.

A spherical datum ('D_Sphere') is used for all WRF projections in ArcGIS, and the Earth radius is always set to 6370000.0m. The input elevation and reservoirs data may be in any defined coordinate system that can be transformed to the WRF sphere. Thus, a geographic transformation method is necessary to allow for conversion of data between a spheroid and the WRF sphere. The next section, '4.4 Geographic Transformation', will address this topic.

Note: Problems may arise as a result of the WRF geogrid.exe program treating geodetic latitude as though it were geocentric latitude (Monaghan et al 2013). Although the "Process GEOGRID File" tool uses the user-supplied elevation instead of the GEOGRID file elevation, it does resample the 'landuse' grid in the GEOGRID file to the routing grid resolution. Care should be taken to note any spatial shifts between elevation-derived grids ('TOPOGRAPHY', 'CHANNELGRID') and the 'landuse' grid.

4.4 Geographic Transformation

Any time that GIS applications must convert geospatial data between datums (for example between a spherical datum such as WRF and a non-spherical datum such as WGS84), a transformation is necessary. Transformations are used in GIS processes to minimize positional error. Failing to apply an appropriate transformation between spherical and spheroidal datums, such as simply treating geodetic latitudes as geocentric, can result in significant (up to 20km North-South) locational shift in the spatial features contained in the data. This shift is greatest at 45° latitude (David et al 2009). No predefined transformation exists in ArcGIS between a sphere and spheroid, thus one must be created if necessary. As suggested by Cedric David, a 'geocentric translation' may be defined in ArcGIS, leaving all parameters '0' when translating between sphere and spheroid. For more information, see the following resources:

http://www.crwr.utexas.edu/gis/gishydro06/SpaceAndTime/SphereVsSperoid2006.htm

However, the main focus of the WRF-Hydro ArcGIS Pre-processing tool is to ensure that WRF data are mapped consistently, and ensure geospatial features in one dataset are appropriately mapped onto other datasets. This minimizes any positional error between datasets at the cost of accurate geographic positioning on the model sphere. In the case of a WRF domain created using the WRF Preprocessing System (WPS) geogrid.exe program, the terrestrial input data is assumed to be on a sphere and are thus mapped to the WRF spherical datum without transformation. Thus, for WPS-created grids, it is appropriate to use a Null Transformation in order to map other terrestrial datasets to the WRF spherical datum such that spatial features (such as mountains, river valleys) will be mapped consistently with the geospatial data used in WRF. The goal is to line up the model's terrestrial data with the other terrestrial data layers (landuse, landmask, elevation, etc.) for consistency of spatial features in the model; thus a Null Transformation is used.

When re-projecting the user-supplied high-resolution elevation data to the WRF-Hydro coordinate system, the WRF-Hydro ArcGIS pre-processing tools will create a new custom geotransformation file named ‘GeoTransform_Null_WRFHydro.gtf’ in the user’s CustomTransformations folder. By default, this is a “Null Transformation”, which will not apply a transformation between the datums. If another transformation is desired, the user may alter the string-type global variable ‘customGeoTransfm’ in the ‘wrf_hydro_functions.py’ script, which describes the desired transformation method and parameters. For more on Custom Geographic Transformations, see:

http://desktop.arcgis.com/en/arcmap/latest/tools/data-management-toolbox/create-custom-geographic-transformation.htm

5 Using the WRF-Hydro GIS pre-processing tools

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 of the standalone tool folder (see

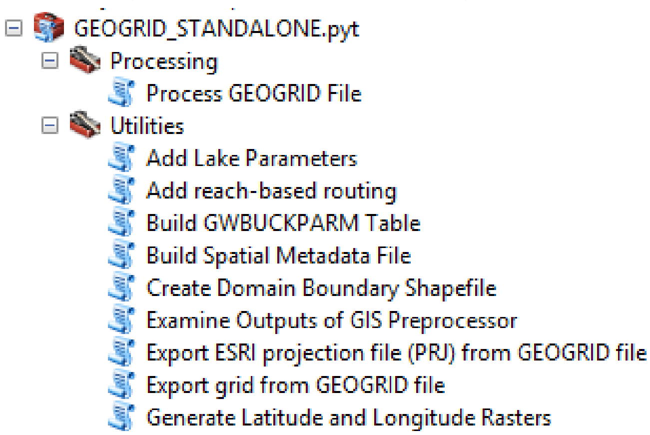

Figure 2). The Python toolbox exists as a *.pyt* file, and will behave like a regular ArcGIS ArcToolbox tool (or *.tbx* file) within ArcGIS. The ‘GEOGRID_STANDALONE’ toolbox contains both a ‘Processing’ and ‘Utilities’ toolset. The processing toolset contains a single ‘Process GEOGRID File’ script tool. The ‘Utilities’ toolset contains multiple script tools for performing additional functions. Users may customize the toolbox and add any scripts or additional functionality necessary for processing WRF-Hydro spatial data, as long as these tools conform to the existing toolbox structure. For more on Python Toolboxes, see:

http://desktop.arcgis.com/en/arcmap/latest/analyze/creating-tools/a-quick-tour-of-python-toolboxes.htm

5.1 Tool validation

The script tools provide some basic validation of input parameters, and will not run until all required inputs are entered and validated. Some validation is performed after the tool execution begins. These types of validations will cause the tool to fail if an invalid value is entered. Check the tool messages in the 'Results' tab for more information in the event of a tool failure.

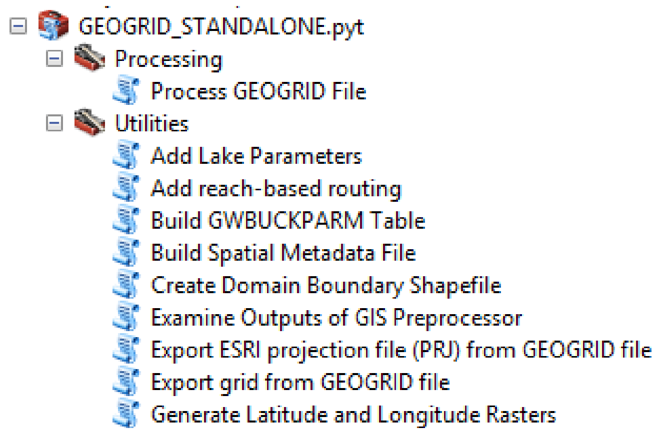

Figure 2. Catalog Tree view of the Python Toolbox in ArcCatalog.

5.2 Executing the Process GEOGRID File tool

Open the 'Process GEOGRID File' script tool by double-clicking on it, or right-click and select **Open**. A tool dialog will open, with five required inputs, as well as optional inputs and several default parameter values supplied. Once all required and any optional inputs are given, click **OK** to execute.

Although environments can be specified at the tool level by clicking on the 'Environments…' button, any settings at this level will be overridden by environment settings in the Python scripts. See the following link for more information on environment hierarchy in ArcGIS:

http://desktop.arcgis.com/en/arcmap/latest/analyze/modelbuilder/a-quick-tour-of-managing-environments.htm

The 'Process GEOGRID File' tool will create a 'scratchdir' directory in the directory provided for the Output Zip File. This scratch directory is a storage location for intermediate datasets. If the tool executes successfully, the scratch directory will be deleted. If an error causes the tool to terminate early, the user will need to manually delete the 'scratchdir' directory before re-running the tool.

5.3 Input elevation data

The selection of a high-quality and high-resolution terrain dataset is very important for proper definition of the terrain-based parameters as well as accurate location of streams. If good data goes in, good data has a better chance of coming out. The elevation data that goes into the tool should already be processed for hydrologic connectivity to ensure that no major problems exist that would break or alter the river network dramatically.

The 'Process GEOGRID File' tool will resample and regrid the input elevation data to the routing grid resolution and extent before performing the terrain processing. A large scale difference between the resolution of the input elevation grid and the routing grid resolution introduces more error into the hydrology of the output data. If possible, try to select a hydrologically processed elevation dataset that is near the resolution of the routing grid. This ensures that the hydrologic conditioning process is not hindered by artifacts introduced in the resampling step.

Hydrologically conditioned elevation datasets are often combined with hydrographic data such as shapefiles of rivers or lakes. Large elevation data collections typically exist as tiles or multiple regional rasters. Using a mulit-raster format such as the mosaic dataset in an Esri® file Geodatabase is a good way of combining multiple elevation rasters for use with the 'Process GEOGRID File' tool. For more on creating a mosaic dataset, see:

http://desktop.arcgis.com/en/arcmap/latest/manage-data/raster-and-images/what-is-a-mosaic-dataset.htm

Examples of large hydro-elevation datasets are listed below:

5.3.1 HydroSHEDS

The Hydrological data and maps based on Shuttle Elevation Derivatives at multiple Scales (HydroSHEDS; http://hydrosheds.cr.usgs.gov/index.php) has been tested for compatibility with the WRF-Hydro GIS pre-processor. HydroSHEDS data are hydrologically processed elevation data at 3 arc-second (~90m), 15 arc-second (~450m), and 30 arc-second (~900m) resolutions, that extend globally to about 60° north and south of the Equator. It is important to note that HydroSHEDS gives elevation values to a precision of 1 meter (no sub-meter elevation values).

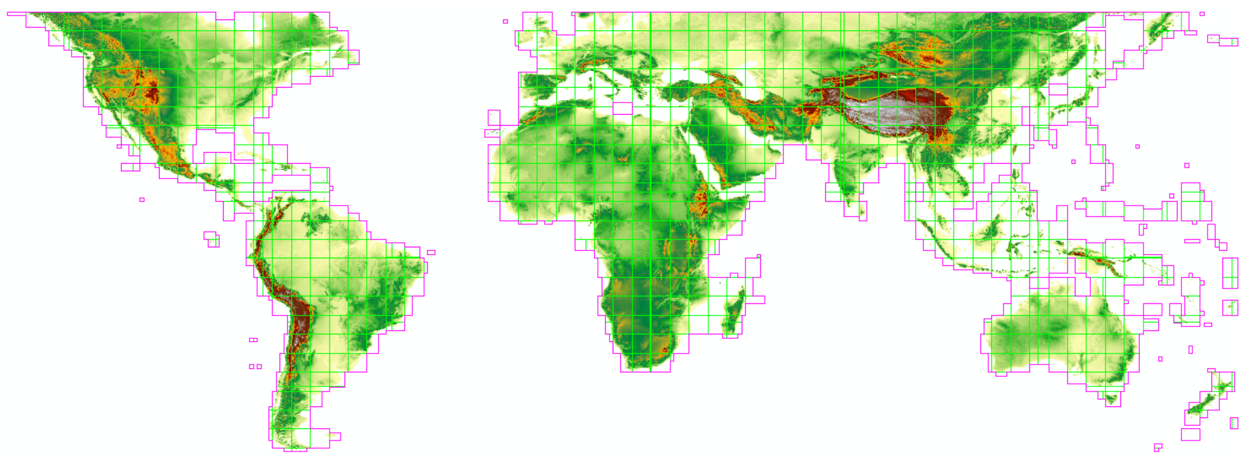

Figure 3. Mosaic of HydroSHEDS 90m elevation data.

5.3.2 NHDPlus/NED

The National Hydrography Dataset Plus, Version 2 (http://www.horizon-systems.com/nhdplus/NHDPlusV2_home.php) has been tested for compatibility with the WRF-Hydro GIS pre-processor. This elevation data covers the Continental United States at 30m resolution, as well as Hawaii and Puerto Rico at 10m resolution. The data provide elevation precision to the nearest centimeter. It is important to note that the NED snapshots used in NHDPlus Version 2 provide elevation in cm above sea level. The values must be converted to meters before being used in the 'Process GEOGRID File' tool. This can be easily done within the mosaic dataset using a Unit Conversion Raster function.

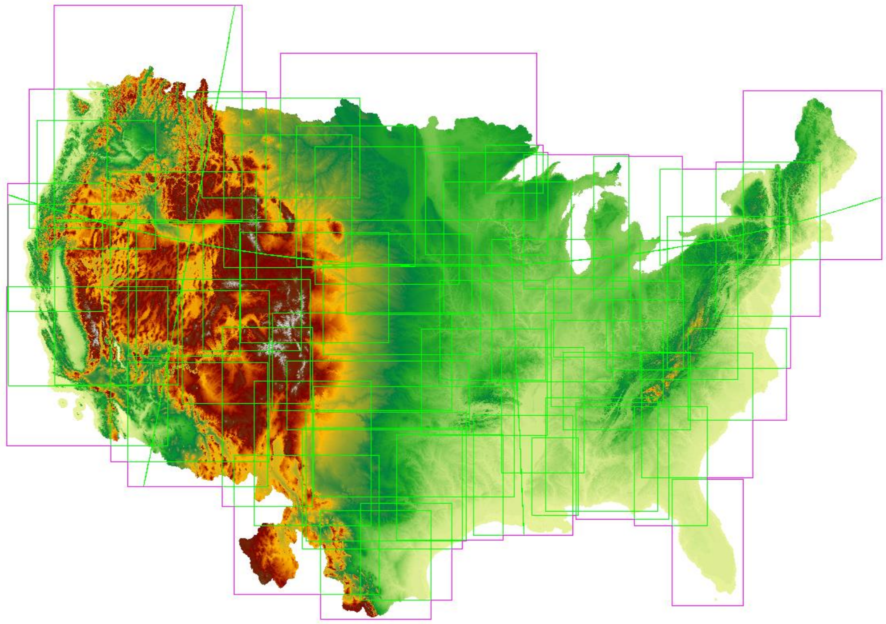

Figure 4. Mosaic of NHDPlus NED Snapshot 30m elevation data (excluding Hawaii and Puerto Rico).

5.3.3 EU-DEM

The EU-DEM elevation dataset (http://www.eea.europa.eu/data-and-maps/data/eu-dem) is not fully compatible with the WRF-Hydro GIS pre-processor. This dataset covers the geographic region of the European Union at 25m horizontal resolution and provides precision to the nearest centimeter. Preliminary testing revealed that ocean areas are given a value of 0 elevation rather than NoData, causing stream definition and channelization over ocean areas. The elevation data will need to be pre-processed and ocean values masked to NoData to eliminate elevation values over ocean areas before running the tool.

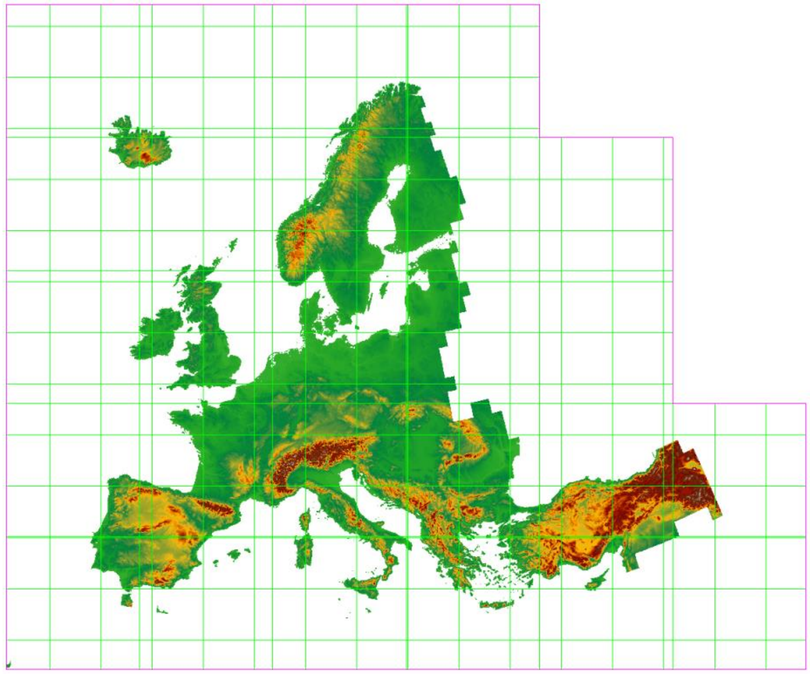

Figure 5. Mosaic of EU-DEM 25m elevation data.

5.4 Description of Process GEOGRID File tool parameters

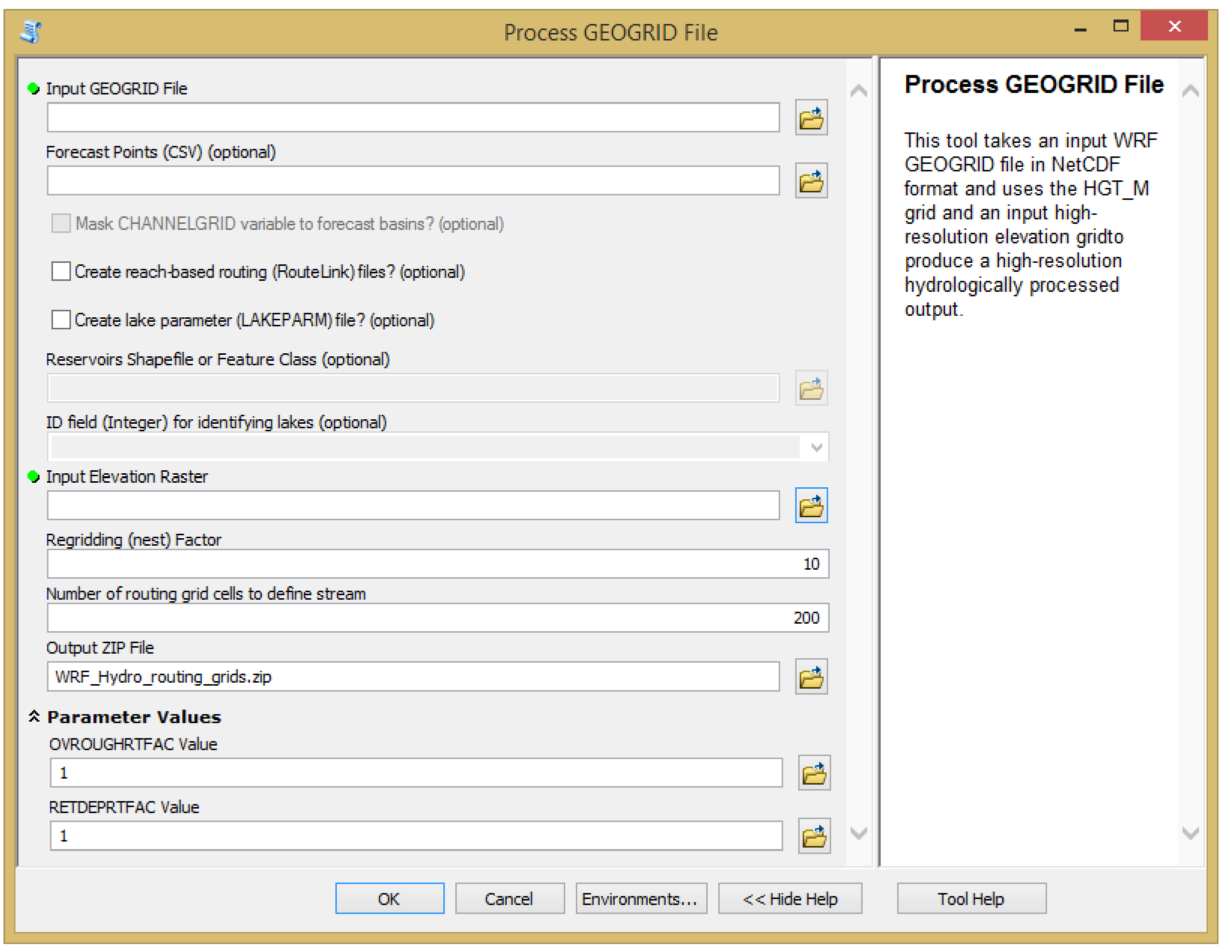

Figure 6. Tool dialog for the Process GEOGRID File tool.

5.4.1 Input GEOGRID file (required)

This is the input WRF GEOGRID file in netCDF (*.nc*) format. This file must contain dimensions 'west_east' and 'south_north', as well as the gridded variables 'HGT_M' and 'LU_INDEX' , which is the grid of elevation values and a land use grid, both on the Mass ('M') grid. The 'HGT_M' variable is the grid used to initiate all of the WRF-Hydro routing grids. The GEOGRID file is typically produced using the WRF Preprocessing System (WPS). See the following link for more information on creating the input GEOGRID file:

http://www2.mmm.ucar.edu/wrf/users/docs/user_guide_V3/users_guide_chap3.htm

5.4.2 Forecast points (CSV) (optional)

This optional parameter requires a Comma-Separated Values (CSV; *.csv*) formatted file of gage locations in latitude/longitude coordinates (WGS84). This ASCII file should contain 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' as well as an ID field named 'FID' in order to be read, and no header names are allowed to contain spaces or special characters. An example of the text in a valid input file 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 Zip archive will contain valid values in the variable 'basn_msk' variable in the 'FullDom' file, which is a grid of the basins delineated from a location just downstream of each gage (see '9.1 Manual Specification of Station Points' and '8.2 Station-based watershed delineation' for more information). The 'FullDom' file will also contain valid values in the 'frxst_pts' variable, corresponding to forecast point locations snapped to the channelgrid with a default snap tolerance of 1 pixel.

5.4.3 Mask CHANNELGRID variable to forecast basins (optional)

This optional boolean (TRUE/FALSE) parameter asks the user whether or not to mask the output 'CHANNELGRID' grid to the basins delineated based on the provided Forecast Points CSV file. The option is only available if a CSV file is provided in the 'Forecast Points (CSV)' parameter. If TRUE (Checked), the derived gauged basins will be used to mask the channels in the 'CHANNELGRID' grid. The resulting 'CHANNELGRID' grid will contain values of 0 for channels inside a gaged basin, values of -1 for channels outside of gaged basins, and -9999 for all other areas. Masking out unwanted channels can result in a significant computational savings because routing is not performed on channels with -1 grid values.

5.4.4 Reach-based routing (RouteLink) files (optional)

This optional boolean (TRUE/FALSE) 'Create reach-based routing (RouteLink) files?' parameter will produce additional output files that facilitate reach-based routing in WRF-Hydro. All channel elements will be vectorized and channel attributes and topology are summarized in the Route_Link.nc output file. Additional files created by this option include a 'LINKID' variable in the 'FullDom' file that gives a grid of the channels with values corresponding to the 'link' variable in the Route_Link.nc file and a streams.shp line shapefile for plotting the streams in a GIS. See '8.3 Reach-Based Routing' for more information on how reach-based routing is performed.

5.4.5 Adding reservoirs (optional)

The optional Boolean (TRUE/FALSE) 'Create lake parameter (LAKEPARM) file?' parameter will produce additional output files that facilitate the addition of reservoirs in WRF-Hydro. If TRUE (checked), the 'Reservoirs Shapefile or Feature Class' and 'ID field (INTEGER) for identifying lakes' parameters will become active and a polygon shapefile will be required and a fieldname may also be provided. Any integer type field may be selected from the input shapefile or feature class, or this parameter may be left blank. If left blank, a new ID will be created using a 1-n range of integer values and the outputs will reflect these new IDs. If a fieldname is provided, the outputs will reflect the values in this field.

This option will add a 'LAKEPARM.nc' file to the output Zip archive and the 'CHANNELGRID' variable in the 'FullDom' file will be altered to include lakes. This alteration includes adding lake outlet points to the 'CHANNELGRID' grid and masking out channel elements where they coincide with a lake pixel. The LAKEGRID will be populated with the IDs and spatial locations of the lakes as well.

Notes:

This option will fundamentally alter and break the 'CHANNELGRID' grid stream connectivity. The reservoir routing option will need to be activated in WRF-Hydro in order to use the resulting routing grids. See '8.4 Lake/Reservoir incorporation and routing' for more information on adding reservoirs to the routing grids.

The default minimum lake size is set at 0.75km^2. This can be changed by altering the 'Threshold' global variable in the 'wrf_hydro_functions.py' script.

If the field 'FTYPE' exists in the lakes shapefile (as with NHDPlus 'NHDWaterbody' shapefiles), the tool will select only features with FTYPE='LakePond' or FTYPE='Reservoir'.

5.4.6 Input Elevation Raster (required)

This required 'Input Elevation Raster' parameter allows the user to provide a high-resolution digital elevation model (DEM) from which to derive the output layers. This grid must be an

ArcGIS supported raster or mosaic dataset format, have the coordinate system defined, and cover the entire extent of the input GEOGRID domain. The test cases included with the download package have a DEM stored in a GeoTIFF format. Additionally, the terrain processing will very likely only be successful if the input DEM has been hydrologically processed to ensure continuous flow paths. See '5.3 Input elevation data' for more information on hydrologic processing and supported elevation datasets.

Note: The vertical units must be in meters above sea level (m).

5.4.7 Regridding (nest) Factor (required)

The required 'Regridding (nest) Factor' parameter allows the user to set the output cell size for the derived datasets based on a relationship to the cell size in the GEOGRID file. The output high-resolution grids must be able to nest perfectly within the coarse GEOGRID resolution, and so the GEOGRID grid resolution will be divided by the regridding factor. Several examples are given in Table 3 for a 1000m input GEOGRID resolution.

GEOGRID resolution	Regridding Factor	Routing resolution
1000m	10	100m
1000m	5	200m
1000m	4	250m

Table 3. Regridding factor effect on routing grid cellsize.

5.4.8 Number of routing grid cells to define stream (required)

The required 'Number of routing grid cells to define stream' parameter is the number of contributing cells from the routing grid used to initiate a stream segment. The number of contributing cells is given by the 'flowacc' flow accumulation grid. Smaller thresholds will yield higher drainage density in the resulting channel network ('CHANNELGRID'). See Figure 7 for examples of channel density based on this parameter.

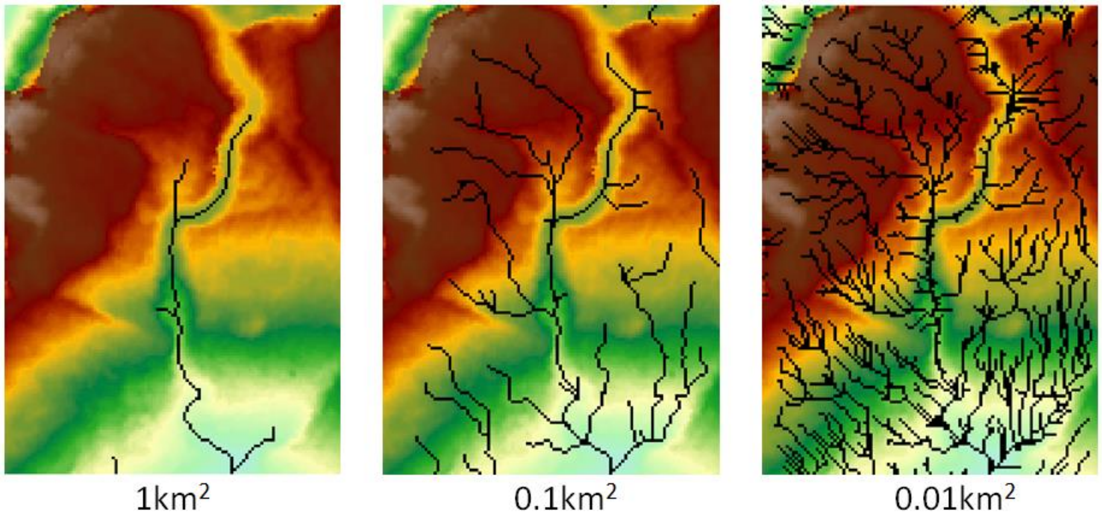

Figure 7. Effect of contributing area threshold on stream density.

When choosing the number of cells for defining a stream, remember that the area depends on the resolution of the grid. If the routing grid is 100m x 100m, then each cell is 10,000m^2 and a threshold of 100 would yield a stream below every 1km^2 of contributing area. This definition will greatly alter the size of the contributing area above each stream and will control the density of the channel network to be created. Several examples are given in the following table for a parameter value of 100 using different routing grid resolutions:

Routing resolution	Contributing cells	Contributing area
100m	100	1,000,000m^2
200m	100	4,000,000m^2
500m	100	25,000,000m^2

Table 4. Minimum contributing area for stream definition based on different routing resolutions.

5.4.9 Output ZIP File (required)

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 containing compressed versions of all output files. Be sure to include the .zip file extension.

5.4.10 Other Parameters (optional)

There are two parameter values: 'OVROUGHRTFAC' and 'RETDEPRTFAC'. The default values for these grids are 1.0. Any valid number will be taken and the resulting 'ovroughrtfac' and 'retdeprtfac' variables will contain constant-value grids with those values. Spatially-varying grids may be substituted in at a later time if desired.

5.5 Process GEOGRID File function workflow

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 (variables: HGT_M and LU_INDEX) from the input GEOGRID file. This function is called 'georeference_geogrid_file'.
2) High Resolution topography is created for the domain, based on the input elevation raster. This grid is projected and then resampled to the output resolution provided by the 'Regridding (nest) Factor' parameter. The WRF GEOGRID resolution is divided by the 'Regridding (nest) Factor' to arrive at the routing grid resolution. This function is called 'create_high_res_topogaphy'.
3) The routing 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 compatibility 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 re-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 parameter 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. Further, if the routing table or reservoir parameter table are requested, those functions are called from within this function. All output layers are converted to numpy arrays and inserted into the 'FullDom' netCDF file, compressed (*.zip*) and returned to the user. This function is called 'sa_functions'.
5) Intermediate files and the temporary 'scratchdir' directory are deleted.

5.6 A note about large domains

Many intermediate rasters are held in memory during processing, which allows 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-descriptive 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.

5.7 Debugging

If the tool fails to execute, the first place to look will be the 'Results' tab in ArcMap or ArcCatalog. In the messages section, an error message may be printed. This should yield what line of the code the error occurred on and may give a message to indicate the type of failure.

Note: If the tool fails to run, the user will be required to manually delete the '\scratchdir' folder, which is a temporary directory created to store intermediate files. This can be found in the directory the user provided for the output Zip archive.

6 Understanding Tool Outputs

6.1 Description of output files

The 'Process GEOGRID File' tool creates a Zip archive of output files according to the options provided to the tool. There will be two netCDF files, one spatial metadata file describing the LDAS grid ('GEOGRID_LDASOUT_Spatial_Metadata.nc), and one describing the routing grid ('Fulldom_hires.nc'). The 'FullDom' file will contain between 12 and 13 2-dimensional gridded netCDF variables. The output Zip file may additionally include shapefiles (*.shp* and accompanying files), netCDF 1-dimensional parameter tables (.nc), and an ASCII raster (*.txt*) file. All variables in the routing grid netCDF file ('Fulldom_hires.nc') will have the same grid dimensions. The x-dimension is always 'x' and the y-dimension always 'y'. Below is an alphabetically sorted list of gridded variables that are created by the 'Process GEOGRID File' tool, and Table 5 contains descriptions of each grid.

Fulldom_hires.nc File:

CHANNELGRID - 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, where non-gaged channels are given a value of '-1'. If lake routing is activated, lake outflow points will be identified by the lake ID value.

FLOWACC – Flow accumulation grid. This grid gives the number of contributing cells for each cell in the domain. This grid is provided for convenience and is not read by WRF-Hydro. For more information, see:

http://desktop.arcgis.com/en/arcmap/latest/tools/spatial-analyst-toolbox/flow-accumulation.htm

FLOWDIRECTION – Flow direction grid. This grid gives the direction of flow (D8) between each cell and the steepest downslope neighbor according to Jenson and Domingue (1988). The result is an integer grid with values ranging from 1 to 128. The values for each direction from center are:

32	64	128
16		1
8	4	2

Figure 8. Flow direction grid values from the analysis cell (blue).

For more information, see: http://desktop.arcgis.com/en/arcmap/latest/tools/spatial-analyst-toolbox/flow-direction.htm

frxst_pts – Gage location grid. If a gage CSV file is provided, the grid will have a cell identified at the location of each gage in the gage CSV file. If no input CSV gage location file is provided, this grid will be uniform with values of '-9999'. Gage pixels are numbered in the same way as the 'basn_msk' grid. NoData cells are given a value of '-9999'.

basn_msk – Groundwater basins grid. If a CSV gage location file is provided, catchments are delineated from a point that is up to 3 pixels (3 * cellsize) downstream of the gage coordinates. This distance can be modified by altering the 'walker' global variable in the 'wrf_hydro_functions.py' script. If masking of the 'CHANNELGRID' is selected, this layer is the mask. Basins are numbered according to the values in the 'FID' field of the input gage CSV file. If no gage location file is provided, this grid will be uniform with values of '-9999'. NoData cells are given a value of '-9999'.

LAKEGRID – Lake grid. If the option to create the lake parameter file is TRUE (checked) and a lake shapefile is provided, this grid will contain ID values for each lake identified by the lake shapefile, the ID field provided (optional), and lake size threshold criteria. If FALSE, this grid will be uniform with values of -9999. See '8.4 Lake/Reservoir incorporation and routing' for more information on how the LAKEGRID is created.

landuse – This is the same data as the 'LU_INDEX' variable in the GEOGRID file, but resampled using Nearest Neighbor assignment to the resolution of the routing grid. For more information on the resampling method, see: http://desktop.arcgis.com/en/arcmap/latest/tools/data-management-toolbox/resample.htm

LATITUDE – Grid of latitude value at the center of each grid cell, in a geographic coordinate system based on a sphere of radius 6370000.0m (EMEP sphere, wkid: 104128). For more information on how latitude is derived, see '7.8 Generating the Latitude and Longitude Rasters'.

LINKID – The channel ID grid. This grid provides a unique integer identifier for each channel segment that is defined in the 'Route_Link.nc' file and 'streams' shapefile. The 'LINKID' grid will only be created if the option 'Create reach-based routing files?' is TRUE (checked).

LONGITUDE – Grid of longitude value at the center of each grid cell, in a geographic coordinate system based on a sphere of radius 6370000.0m (EMEP sphere, wkid: 104128). For more information on how longitude is derived, see '7.8 Generating the Latitude and Longitude Rasters'.

OVROUGHRTFAC – OVROUGHRTFAC parameter. Currently set to a default of 1.0. This default value may be changed by providing an alternate value to the 'OVROUGHRTFAC Value' parameter (under 'Parameter Values') of the 'Process GEOGRID File' tool.

RETDEPRTFAC – RETDEPRTFAC parameter. Currently set to a default of 1.0. This default value may be changed by providing an alternate value to the 'RETDEPRTFAC Value' parameter (under 'Parameter Values') of the 'Process GEOGRID File' tool.

STREAMORDER – Stream order grid, calculated using the Strahler method (Strahler 1957). NoData cells have values of '-15'. More information on derivation of stream order can be found here:

http://desktop.arcgis.com/en/arcmap/latest/tools/spatial-analyst-toolbox/stream-order.htm

TOPOGRAPHY – Elevation grid. The units of elevation are the same as the 'Input Elevation Raster' dataset, which should be in meters (m). This grid is derived from the elevation values in the 'Input Elevation Raster', but has been converted to floating point, resampled to the routing grid resolution, and filled according to the method described here:

http://desktop.arcgis.com/en/arcmap/latest/tools/spatial-analyst-toolbox/fill.htm

See '8.1.1 Pit filling' for more information on the filling process.

Other files:

GEOGRID_LDASOUT_Spatial_Metadata.nc - This is a netCDF format file that provides the spatial metdata associated with the GEOGRID variables. By default, no 2-dimensional grids are written to the file. This file is used for appending geospatial metadata to the LDAS model output, if necessary.

gw_basns_geogrid.txt - This is an ASCII raster of the 'basn_msk' grid, but regridded to the GEOGRID file resolution. NoData cells are given a value of '-9999'. This file is only created if a CSV gage location file is provided.

LAKEPARM.nc – Lake parameter table. This netCDF format file is created if the option to create the lake parameter file is TRUE (checked) and a lake shapefile is provided. The table will contain a record for each lake feature in the FullDom 'LAKEGRID' variable, and contain derived and default parameters for each lake. See '8.4 Lake/Reservoir incorporation and routing' for more information on how the lake parameter table is created.

Route_Link.nc – The reach-based routing parameter file. This netCDF format file is created if the 'Create reach-based routing (RouteLink) files?' parameter is TRUE (checked). The table contains a record for each stream segment. The stream segments in this table are also identified by the unique 'LINKID' values in the 'LINKID' variable in the FullDom file, and values in the 'ARCID' field of the output 'streams' shapefile. This table contains derived and default stream segment parameters that are calculated based on the vector stream network and topology in the 'streams' shapefile. For more information see '8.3 Reach-Based Routing'.

Streams.* - Streams shapefile. Shapefiles are a collection of files with the same name but different extensions. Together these files make up the shapefile. The 'streams' shapefile has one feature for each stream segment in the domain. This file is meant to accompany the 'Route_Link.nc' file and FullDom 'LINKID' variable. The 'streams' shapefile is only created when the option 'Create reach-based routing files?' is TRUE (checked). Note: the 'FROM_NODE' and 'TO_NODE' fields refer to node IDs, not the 'ARCID' of the stream segment. The 'ARCID' field is the unique identifier for each stream segment, and corresponds to the 'link' variable of the 'Route_Link.nc' file and the 'LINKID' variable in the 'Fulldom_hires.nc' file. The geometry of the stream segments in this shapefile informs many of the parameters in the 'Route_Link.nc' file.

File Name	Variable Name	Data Type	Default Value	NoData
Fulldom_hires.nc	CHANNELGRID	Integer	-	-9999
Fulldom_hires.nc	FLOWACC	Float	-	-
Fulldom_hires.nc	FLOWDIRECTION	Integer	-	-15
Fulldom_hires.nc	frxst_pts	Integer	-	-9999
Fulldom_hires.nc	basn_msk	Integer	-	-9999
gw_basns_geogrid.txt	-	Integer	-	-9999
Fulldom_hires.nc	LAKEGRID	Integer	-	-9999
Fulldom_hires.nc	landuse	Float	-	-
Fulldom_hires.nc	LATITUDE	Float	-	-
Fulldom_hires.nc	LINKID	Integer	-	
Fulldom_hires.nc	LONGITUDE	Float	-	-
Fulldom_hires.nc	OVROUGHRTFAC	Float	1.0	-
Fulldom_hires.nc	RETDEPRTFAC	Float	1.0	-
Fulldom_hires.nc	STREAMORDER	Integer	-	-15
Fulldom_hires.nc	TOPOGRAPHY	Float	-	-9999

Table 5. Output grids and data types from Process GEOGRID File tool.

6.2 Log File

A log file is initiated using the filename provided to the 'Output ZIP File' parameter of the pre-processing tool. This log file collects all messages provided by the tool during execution and writes them to an ASCII file in the same directory as the output Zip archive, with the '.log' extension appended to the archive filename. If the tool fails to execute successfully, however, the log file will be created but nothing will be written. In this case, consult the geoprocessing 'Results' tab in ArcMap or ArcCatalog for more information (Figure 9).

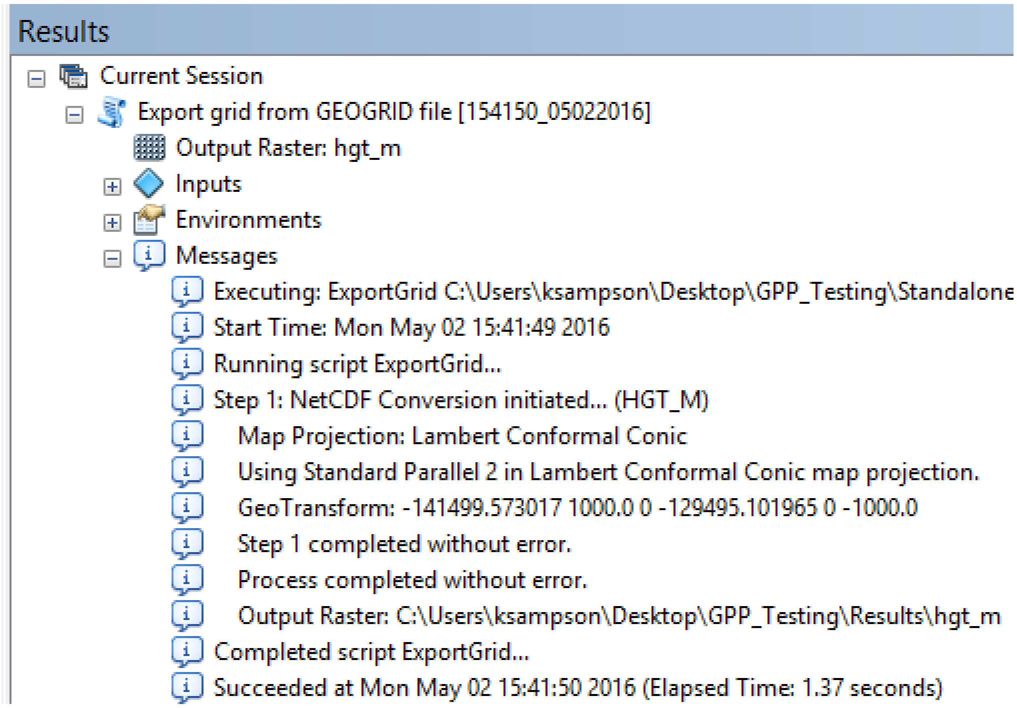

Figure 9. Messages returned by the ExportGrid tool as viewed in the 'Results' tab of ArcMap.

6.3 Preparing the WRF-Hydro 'routing stack' file

The output files from the WRF-Hydro GIS pre-processing tool are intended to a) produce the required set of routing grids necessary to run WRF-Hydro, and b) assist in the GIS processing, visualization, and analysis of data related to the WRF-Hydro configuration. Thus, the output archive from 'Process GEOGRID File' contains files that will compose the final 'routing stack' as well as additional data. It is here that the user may substitute any custom grids they have. See '6.3.1 Using alternate grids in WRF-Hydro ' for more information on this option.

6.3.1 Using alternate grids in WRF-Hydro

The 'Process GEOGRID File' tool will create all of the necessary grids for WRF-Hydro. There may be instances, however, when a user wishes to use an alternate grid. An example is provided for using a spatially varying 'OVROUGHRTFAC' grid. Rather than using the static, single-value version of the 'OVROUGHRTFAC' grid given by 'Process GEOGRID File', the user may wish to use an alternate grid. If the desired grid is formatted to have the same dimensions, extent, and cellsize as the routing grid, it may be used in exchange for the grids provided by the 'Process GEOGRID File' tool as long as it has the same variable name and dimensions as the corresponding grid that is being replaced. Simply substitute the replacement grid using netCDF tools such as 'ncks'.

Note: replacement of certain grids may affect the model behavior. For example, substituting the GEOGRID 'landuse' or 'landmask' variable will affect model behavior if the 'topography' variable does not respect the new layers. For example, the model may produce erroneous results or fail to run if there is no elevation gradient (or a negative slope) between areas classified as 'land' and 'ocean'. Be sure that areas with valid elevation values correspond only with 'land' grid cells from the coarse grid 'LANDMASK' variable in the GEOGRID file. This must be done because the model cannot route water through areas classified as ocean in the LANDMASK.

7 Other utilities

In addition to the main 'Process GEOGRID File' script, other script tools are provided in the 'Utilities' toolset (Figure 10) of the WRF-Hydro GIS pre-processing toolbox. Below is a description of each utility.

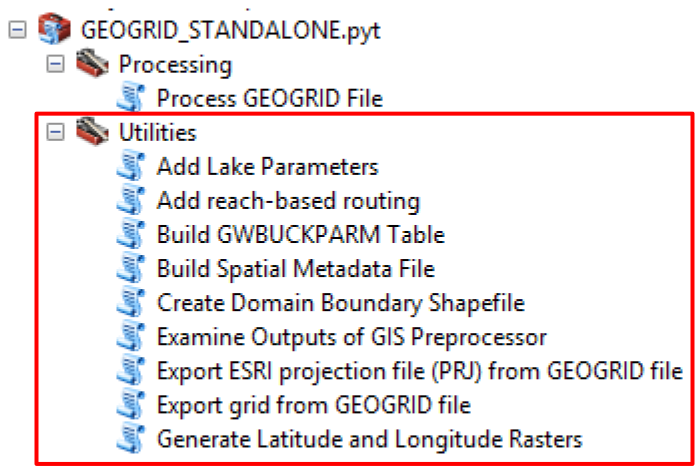

Figure 10. Utilities toolset as viewed from ArcCatalog.

7.1 Add Lake Parameters

The 'Add Lake Parameters' tool will add the lake parameters to an existing routing stack .zip file. The user will submit a lakes polygon shapefile, the field containing the lake ID (optional), and an output .zip file to store the new routing stack. The LAKEGRID variable will be updated to include lakes, the CHANNELGRID variable will also be masked to the lakes in LAKEGRID, the LAKEPARM.nc lake routing table will be created, and the lake shapefile will be saved to the output .zip.

7.2 Add reach-based routing

The 'Add reach-based routing' tool will add reach-based routing parameters to an existing routing stack .zip file. No inputs are necessary other than the input routing stack .zip file and the output routing stack .zip file. The 'CHANNELGRID' variable is used to define the stream network. A LINKID variable is added to the FullDom_hires.nc file, a streams.shp polyline shapefile is also added to the output, as is the reach-based routing parameter table (Route_Link.nc) in netCDF format.

7.3 Build Spatial Metadata File

The 'Build Spatial Metadata File' tool will create netCDF format spatial metadata files from the GEOGRID file (for LDASOUT) and the FullDom_hires.nc routing grid file. These spatial metadata files contain CF-compliant georeference information for the grids, as well as some additional metadata that assists GIS applications like ArcGIS to interpret the projection and datum associated with each grid. Associating WRF-Hydro output with these spatial variables may be useful for visualizing the output in GIS.

7.4 Create a Domain Boundary Shapefile

The ‘Create Domain Boundary Shapefile’ utility will create a simple single-feature polygon shapefile using the GEOGRID file as input (required). The extent of the grid is determined from the grid dimensions, cellsize, and georeferencing information, and a polygon is generated using the WRF CRS. The geometry is saved to shapefile format. This shapefile may be used, for example, to display the boundary of a WRF domain on a map (Figure 11).

Figure 11. Polygon (red) representing the extent of the WRF domain.

7.5 Examinine the Outputs of the WRF-Hydro GIS Pre-processor

The ‘Examine Outputs of GIS Preprocessor’ utility script extracts the individual datasets from the ZIP archive created by ‘Process GEOGRID File’ and converts those files to formats that are easily consumed by a Desktop GIS. Each 2D variable in the netCDF file in the output ZIP archive will be converted to a raster, as will the ASCII raster ‘gw_basns_geogrid.txt’. Additional layers, such as shapefiles and single-dimension netCDF files will be uncompressed. The ‘Input ZIP File’ parameter requires the ZIP archive that is created by the ‘Process GEOGRID File’ tool. The ‘Output Folder’ parameter is required and should be the name and location of a folder on disk that does not yet exist. The tool will create the folder, de-compress the archive, and save converted grids to this location.

7.6 Creating a projection definition file from GEOGRID

The 'Export ESRI projection file (PRJ) from GEOGRID file' tool takes the WRF GEOGRID file as input (required) and automatically determines the projection based on the global attribute data to produce a *.prj* projection file. The file is simply an ASCII file containing the text that describes the coordinate reference system parameters. This file allows the user to build data and grids in the projection of the WRF domain using standard GIS applications. The projection information is based on the Mass Grid (: stagger = 'M';) from a WPS-created GEOGRID file. See the following link for more information on the ESRI® PRJ format:

http://vsp.pnnl.gov/help/Vsample/ESRI_PRJ_File.htm

See '4.1 Coordinate Systems' for more information on how the projection definition is acquired from the GEOGRID file.

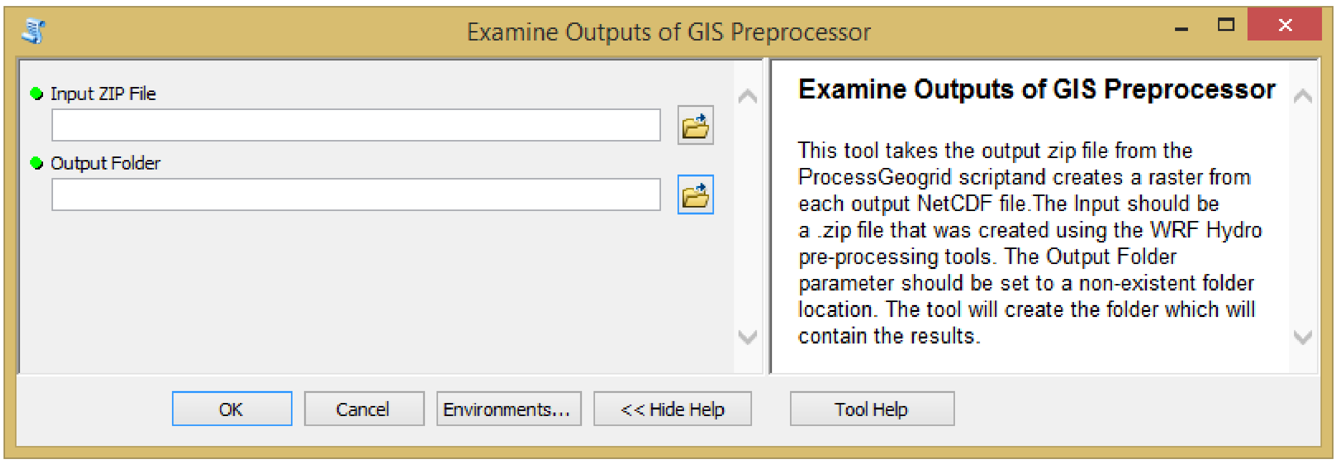

Figure 12. Dialog box for the ExamineOutputs utility.

7.7 Exporting a grid raster from a GEOGRID variable

The 'Export grid from GEOGRID file' utility takes as input a WRF GEOGRID file (required), and saves it as a raster. This tool allows the user to quickly plot any Mass grid variable in a GIS, with proper georeferencing. Once a valid GEOGRID file is input, internal tool validation is performed to gather a list of all variables on the Mass grid (: stagger = 'M') from a WPS-created netCDF GEOGRID file. The user must select one of the variables from the 'Variable Name' drop-down menu (Figure 13). The tool automatically determines the projection based on attribute data within the GEOGRID file. See '4.1 Coordinate Systems' for more information on georeferencing a GEOGRID file. The required 'Output Raster' may be any format of raster supported by ArcGIS.

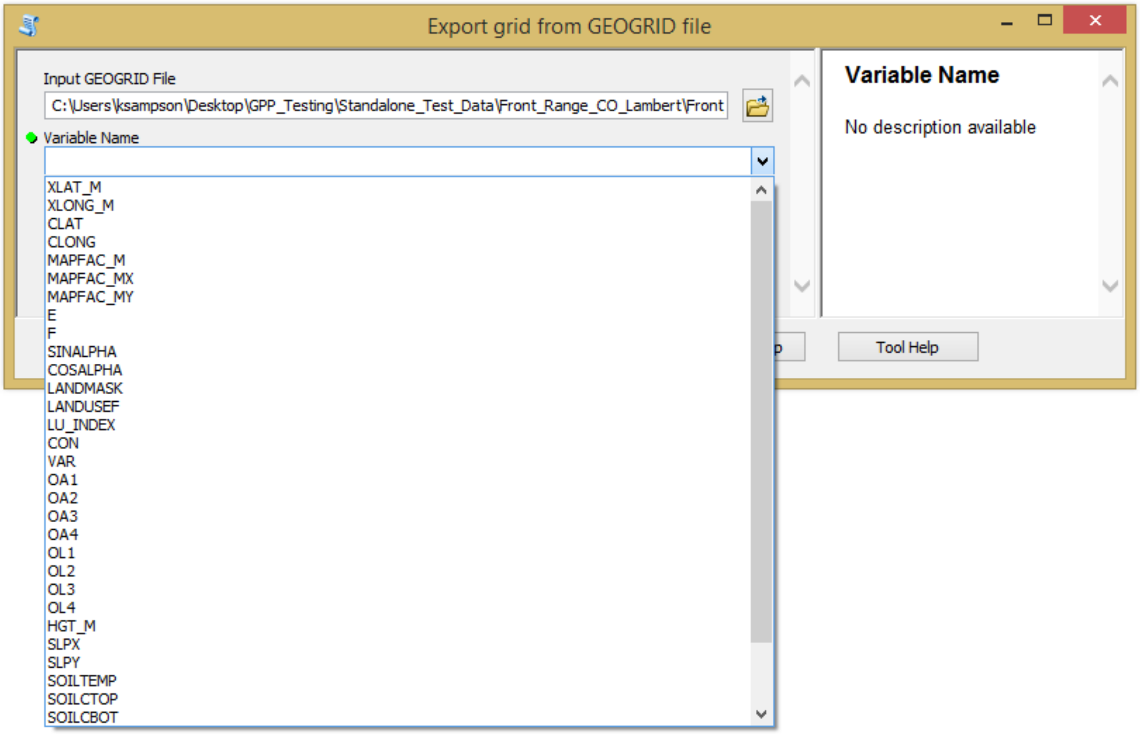

Figure 13. Dialog box for the 'Export grid from GEOGRID file' utility.

7.8 Generating the Latitude and Longitude Rasters

The 'Generate Latitude and Longitude Rasters' tool is a general utility for creating a grid of latitude and longitude values. This function is necessary for WRF-Hydro, but is not an easily found function in ArcGIS. The tool takes any raster as input, with a valid coordinate system defined, and produces as output the 'latitude' and 'longitude' Esri GRID rasters in the output directory. The required 'Output Folder' parameter is a directory and must exist on disk.

7.9 Build Groundwater Bucket Parameter Table and Grid (GWBUCKPARM)

The 'Build GWBUCKPARM Table' tool is a general utility for creating a parameter table of groundwater buckets used in the groundwater scheme in WRF-Hydro. Groundwater buckets may be defined in many different ways, but the most common may be delineated above forecast points, generating contributing basins for each reach in a river network, or providing a polygon shapefile defining groundwater buckets. No matter which method is employed, the groundwater buckets must be resolved first on the coarse LSM grid (the grid defined by the GEOGRID file). The outputs of this tool are a groundwater bucket ASCII raster, and a parameter table in both netCDF and ASCII format. Currently, WRF-Hydro will read the .TBL ASCII bucket parameter table and the ASCII raster. In the future, WRF-Hydro may be capable of reading one or both directly from netCDF.

The tool requires the Fulldom_hires.nc netCDF file, as well as the GEOGRID file to define the model domains. A method must be selected from the 'Method for deriving groundwater basins' drop-down menu.

Methods to define groundwater basins:

Currently, the user may select ‘FullDom basn_msk variable’ to use the ‘basn_msk’ variable in the FullDom file, as long as the ‘Process GEOGRID file’ tool was run using forecast locations. This method is provided as a convenience, as an ASCII raster ‘gw_basns_geogrid.txt’ will have already been created from these forecast locations during the ‘Process GEOGRID file’ tool execution. If the user substitutes a new ‘basn_msk’ grid into the FullDom file, this method may be used to re-generate the ASCII grid and .TBL parameter tables against this grid.

The user may also choose ‘FullDom LINKID local basins’ to generate contributing basins for each reach in the CHANNELGRID network. This may be selected regardless of whether or not reach-based routing outputs were generated.

Currently, providing a shapefile of groundwater basins is not supported.

Tool outputs:

The output of this tool will include a ‘gw_basns_geogrid.txt’ ASCII raster in the output directory, and optionally a ‘GWBUCKPARM.TBL’ or ‘GWBUCKPARM.nc’ file, or both depending on the ‘Output table type’ option selected.

8 GIS Methods Descriptions

Many GIS methods are used in the WRF-Hydro GIS pre-processing tools. A description of some of the more important methodological choices is given below.

8.1 Terrain Processing

The workflow and methods used in processing the terrain data are largely borrowed from the early portions of the Basic Dendritic Terrain Processing workflows of ArcHydro (http://downloads.esri.com/archydro/archydro/Tutorial/Doc/Arc%20Hydro%20GP%20Tools%202.0%20-%20Tutorial.pdf, page 47). These processes occur in the wrf_hydro_functions.py file, in the 'sa_functions' function.

8.1.1 Pit filling

The input elevation data, once regridded to the routing grid domain and resolution, are first converted to floating point, then processed to derive terrain-based hydrologic parameters. The elevation data must be pit-filled, which is a process to remove depressions. The elevation of the area of the depression is increased until the depression fills.

Note: It may occur that the elevation resampling process may introduce pits. In some landscapes, filling these pits will cause large areas upstream of the pit to 'fill', or have the elevation raised to the level that eliminates the depression. This is noticeable because the output steam channels behave as they would in lakes or in areas of extremely low relief. Manual stream-burning methods may be required on the input elevation data before processing.

The 'z_limit' global variable in the wrf_hydro_functions.py script can be used to control the maximum fill depth from the pit-filling process. This value is set to 1000.0m by default.

8.1.2 NoData assignment

WRF-Hydro requires NoData values in grids to be specified by a value of '-9999'. However, GIS applications allow NoData in gridded datasets. Throughout the workflow, NoData is frequently converted to -9999.

8.2 Station-based watershed delineation

If the user provides a CSV file with gage locations as input to the 'Forecast Points (CSV)' parameter in the Process GEOGRID File tool, additional processing is performed to delineate basins based on those points. For more on the format of the CSV file, see '5.4.2 Forecast points (CSV) (optional)', above. The 'sa_functions' function in the 'wrf_hydro_functions.py' script will handle the delineations. The input CSV file is used as input to the 'MakeXYEventLayer' tool, and fields 'LAT' and 'LON' are used to define the geographic locations (in WGS84). These points are snapped to the nearest grid cell center. Iteration may be required to make sure that station gage locations will coincide with channel elements. S '9.1

Manual Specification of Station Points' for a description of how to ensure that station points will be delineated properly.

The 'frxst_pts' FullDom variable will contain a grid with gage locations snapped to the nearest grid cell. The basins delineated from the gage locations must be 'walked' downstream by a certain distance so that they contain the gage location. Thus, the gage locations are snapped again to the flow accumulation grid, but this time with a tolerance of 3x the resolution of the routing grid. Basins in 'basn_msk' variable are delineated from the downstream location (see Figure 14). The 'gw_basns_geogrid.txt' ASCII raster is the same as the 'basn_msk' grid, except that it has been resampled to the coarser resolution of the WRF LDAS grid.

If the 'Mask CHANNELGRID variable to forecast basins?' parameter is TRUE (checked), the 'basn_msk' grid will be used to mask channels to these basins, such that channel cells inside the delineated basins will equal '0' and channel cells outside of these basins are given a value of '-1'.

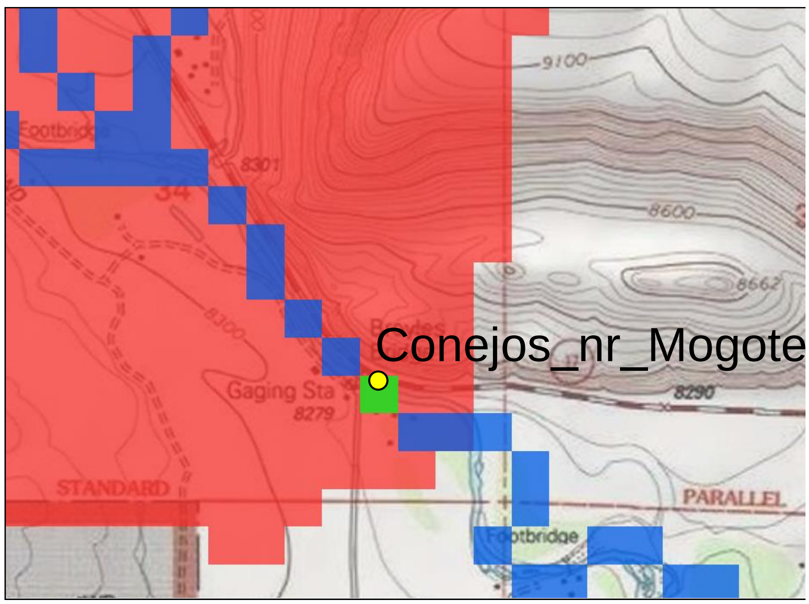

Figure 14. Gage point (yellow circle) given by in the gage CSV file, the corresponding gage location grid cell (green square), and the resulting watershed (shaded red).

8.3 Reach-Based Routing

The efficiency of routing of runoff may be significantly improved by performing a reach-based routing approach. This is facilitated through building a table of stream segment attributes on which to perform the routing. This process is only executed if the 'Create reach-based routing (RouteLink) files?' option to the Process GEOGRID File tool is TRUE (checked). Within the 'sa_functions' function, a separate function called 'Routing_Table' is called which performs the vectorization and prepares the Route_Link.nc file. This section provides information on the process of building this table.

8.3.1 Vectorizing the stream network and describing connectivity

The Spatial Analyst > Hydrology > Stream to Feature tool converts gridded channel information (CHANNELGRID) to vector-based stream segments. Each segment is given a unique identification (ARCID), and a basic network topology is defined in the form of a to:from system based on nodes. Although the nodes are not explicit, the relationship of each node to the upstream or downstream end of a segment is known. Using a mapping from segment ID to node ID, a segment based topology (to:from) is established (Figure 15). For more on how this process is performed in ArcGIS, see:

http://desktop.arcgis.com/en/arcmap/latest/tools/spatial-analyst-toolbox/stream-to-feature.htm

The vector-based stream segments by default extend to the edge of the domain. A simple process is used to find and trim all segments such that the lines terminate at the cell center of grid cells along the outer edge of the domain.

8.3.2 Route_Link.nc reach-based routing table

The Route_Link.nc file describes spatial and non-spatial characteristics of each river reach. The 'link' variable in this table corresponds to the 'ARCID' field in the 'streams.shp' shapefile. To visualize any of the attributes in this table on a map, simply join the streams shapefile to this table using the 'link' and 'ARCID' fields, respectively. Converting the 'Route_Link.nc' table to an ArcGIS readable table requires using 'Make NetCDF Table View' tool in the Multidimension Tools toolbox. Many of the fields in this table are given default attributes, which may be changed in the global attributes section of the 'wrf_hydro_functions.py' file. Other attributes are calculated from the stream segment geometry (for example 'length'), or a combination of geometry and topographic attributes (for example 'slope').

Note: The order of records in this table is extremely important. The segments must be ordered such that no segment is entered before all contributing segments are entered. A topological sorting is performed ('sort_topologically_stackless' function nested in the 'Routing_Table' function of 'wrf_hydro_functions.py' script) that begins with all order 1 streams and continues to add elements to the table until all downstream segments are entered.

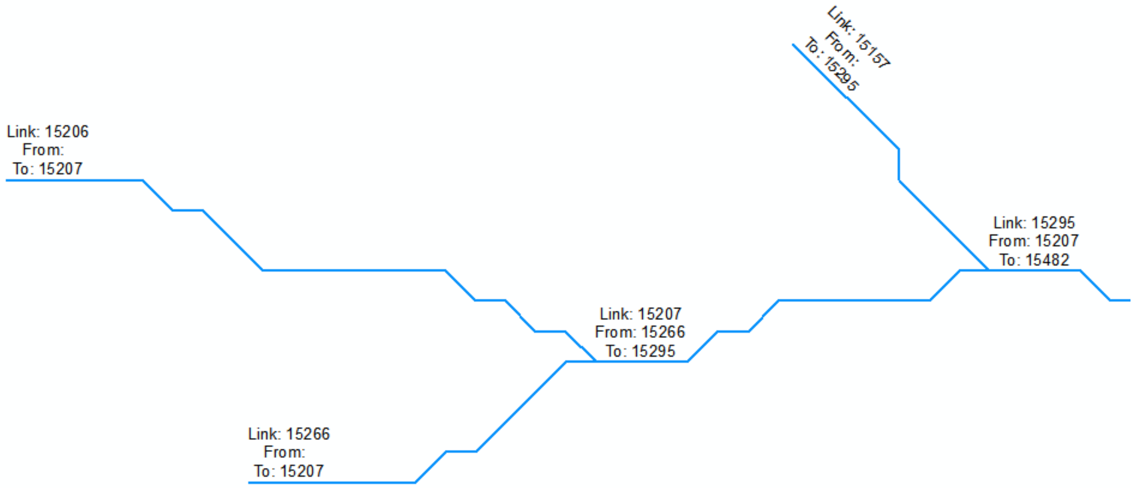

Figure 15. Example of simple network topology in reach-based routing.

Variable	Parameter name	Field type	Default value
link	Link ID	Integer	-
from	From Link ID (not used)	Integer	-
to	To Link ID	Integer	-
lon	Start node longitude	Float	-
lat	Start node latitude	Float	-
alt	Start node elevation (m)	Float	-
order	Stream order (Strahler)	Integer	-
Qi	Initial Flow in Link (cms)	Float	0
MusK	Muskingum routing time (s)	Float	3600
MusX	Muskingum weighting coefficient	Float	0.2
Length	Stream length (m)	Float	-
n	Manning's roughness	Float	0.035
So	Slope (%; drop/length)	Float	-
ChSlp	Channel side slope (%; drop/length)	Float	0.05
BtmWdth	Bottom width of channel (m)	Float	5
time	Time of measurement (not used)	Float	-
Kc	Channel conductivity (mm/hr)	Integer	0

* See definition of orifice and lake elevations, above.

Table 6. Variable definitions for the Route_Link.nc reach-based routing table.

8.4 Lake/Reservoir incorporation and routing

When the optional Boolean 'Create lake parameter (LAKEPARM) file?' parameter is TRUE (checked) and a lakes shapefile is provided to the Process GEOGRID File tool, then several layers are altered to take reservoir routing into account.

Lake processing is performed in the 'add_reservoirs' function in the 'wrf_hydro_functions.py' script. There are two global variables in the script that may be altered to produce desired results; 'Threshold' and ‘walker’. The ‘Threshold’ variable controls the minimum size of lake polygons to be considered, and is given in square kilometers. The area in square kilometers is calculated directly from the lake polygon geometry. The default value is $0.75km^2$. The ‘walker’ variable gives the distance (in number of pixel-widths) downstream of each lake to travel in order to determine the minimum elevation value of the lake (altitude of the lake bottom).

If no value is provided in the ‘ID field (Integer) for identifying lakes’ parameter, the features in the input shapefile are renumbered from 1 to n.

Note: The input lakes shapefile is clipped to the extent of the GEOGRID file. This means that any lakes that cross the edge of the domain will be clipped and only the area inside the model domain will be calculated.

The lake polygons are rasterized according to the “MAXIMUM_AREA” algorithm in order to create the ‘LAKEGRID’ output grid (LAKEGRID.nc). Each lake outlet is identified using the largest flow accumulation value (from the flow accumulation grid, ‘flowacc’) overlying the lake. The CHANNELGRID pixel at the location of maximum flow accumulation is given the ID of the lake, while all channel elements within the lake are eliminated. The lake minimum elevation is gathered from a location downstream of the lake outlet, with the distance given by the number of pixels in the ‘Threshold’ global variable. For example, a routing grid resolution of 100m and threshold of 3 yields a snapping tolerance of 300m downstream of the lake outlet.

Zonal Statistics is performed on the altered lakes to gather area and maximum lake elevation. The topography layer is used to gather elevation, while area and lake centroid are calculated directly from the altered geometry.

Note: the lakes have at this point been gridded (resolved on the routing grid), and this area will differ slightly from the area of each lake in the input lakes shapefile. See Figure 16 for an example.

For more information on Zonal Statistics, see:

http://desktop.arcgis.com/en/arcmap/latest/tools/spatial-analyst-toolbox/zonal-statistics.htm

The ‘Orifice elevation’ and ‘lake elevation’ are calculated based on the minimum and maximum lake elevations, defined above. A fraction is applied to of the elevation range is used, as defined below:

1) The Orifice elevation is the minimum lake elevation plus 1/3 of the vertical distance between the lake minimum and maximum elevations.
2) The Lake elevation is the minimum lake elevation plus 2/3 of the vertical distance between the lake minimum and maximum elevations.

The calculated lake parameters are written to the ‘LAKEPARM.nc’ file (Table 7). The ‘lake’ variable contains each lake identifier, and corresponds to the grid cell values in LAKEGRID.nc. The CHANNELGRID.nc grid is also altered so that the lake outlet pixel contains the lake identifier as well. Default lake routing parameters (OrificeC, OrificeA, WeirC, WeirL) can be altered in the global variables of the ‘wrf_hydro_functions.py’ script.

Variable	Parameter name	Field type	Default value
lake_id	Lake ID	Integer	-
LkArea	Gridded lake area (km^2)	Double	-
LkMxH	Maximum lake elevation (m ASL)	Double	-
WeirC	Weir Coefficient	Double	0.4
WeirL	Weir Length (m)	Double	0.0
WeirH	Weir Height (m ASL)	Double	
OrificeC	Orifice Coefficient	Double	0.1
OrificeA	Orifice Cross-sectional area (m^2)	Double	1.0
OrificeE	Orifice Elevation (m ASL)*	Double	-
lat	Centroid latitude (WGS84)	Float	-
lon	Centroid longitude (WGS84)	Float	-
alt	Lake elevation (m ASL) *	Float	-
time	Time of measurement	Double	
Discharge	Default Discharge (CMS)	Double	

* See definition of orifice and lake elevations, above.

Table 7. Field definitions for the LAKEPARM.nc lake parameter file.

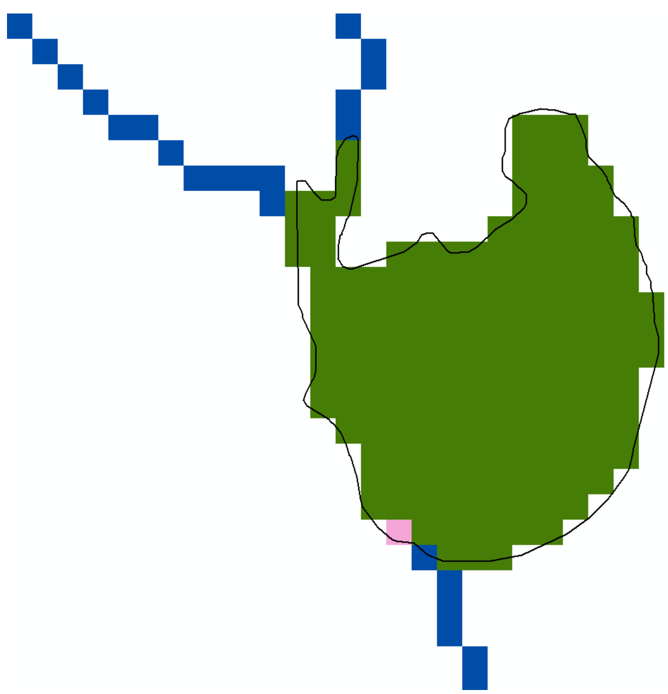

Figure 16. Lake/Reservoir routing grid example. CHANNELGRID layer contains channels in blue and lake outlet in pink, the LAKEGRID layer shows a lake in green, and lake outline from the input lakes shapefile is outlined in black.

9 Miscellaneous Topics in the Use of the WRF-Hydro pre-processor

9.1 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 menu. Upon ingest, be sure to define the projection as Geographic WGS84. Remember that the coordinates of your station points should be in WGS84. Export the imported data layer as a shapefile by right-clicking on the event layer and selecting > Data > Export Data.
2. 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 netCDF data variables, you'll need to extract the topography (variable 'TOPOGRAPHY'), basin mask (variable 'basn_msk') and channel network (variable 'CHANNELGRID') from the 'Fulldom_hires.nc' file. **VERY IMPORTANT:** Users can also use the '**Examine Outputs of GIS Preprocessor**' utility script described above to automatically convert netCDF variables to Esri® GRID raster format.
3. Import these netCDF variables from the Fulldom_hires.nc file 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 '**Examine Outputs of GIS Preprocessor**' utility script described above to automatically convert all netCDF variables to Esri® GRID raster format.
4. Add the station point shapefile to the map that was created in step #1 above.
5. Open a copy of the original *.csv* station file in Microsoft Excel or your preferred spreadsheet editor.
6. 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.

7. 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. You will need 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 GIS Pre-processing tool.
8. Repeat this search-edit process for each station in question.
9. 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.
10. Re-run the WRF-Hydro GIS Pre-processing 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 desired. If not, you'll need to inspect the locations again more closely and make further adjustments as necessary.

10 Script Customization

Users may have certain customizations that they would like to add into the WRF-Hydro GIS pre-processing workflow. A thorough understanding of Python and the ArcGIS API for Python (arcpy) will be useful. The following notes may also assist users in customizing the scripts. For more information on authoring Python Toolboxes in ArcGIS, see:

http://desktop.arcgis.com/en/arcmap/latest/analyze/creating-tools/a-quick-tour-of-python-toolboxes.htm

10.1 Structure of the scripts in the pre-processing tool

The WRF-Hydro pre-processing 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. There are also .xml files for each script tool that describe the tool and any tool parameter descriptions. The scripts must be located in the same directory at all times.

10.2 Editing the scripts

Users may modify the Python script as they wish according to the terms of the license, 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).

10.3 After making script changes

ArcGIS and ArcMap use XML (*.xml*) files to describe the Python toolbox. Deleting these changes will delete helpful parameter descriptions used in the tool dialog. After script changes are made, right click on the toolbox from within ArcMap or ArcCatalog and 'refresh'. This should allow ArcGIS to recognize the script changes. Testing revealed that the tool behavior may not change until this step has been completed. Further, periodically deleting files in the user's TEMP directory may be necessary.

11 Future Compatibilities

11.1 Compatibility with TauDEM software

The use of more sophisticated terrain processing techniques is being explored. In the future, terrain processing may include options for processing using Terrain Analysis Using Digital Elevation Models (TauDEM) version 5. 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 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', which is currently under development. Additional functionality includes the ability to choose output from the D-Infinity (Tarboton 1997) flow direction or Peuker-Douglas stream definition algorithms (Peuker and Douglas 1975), and check for edge contamination in the output CHANNELGRID and frxst_basns 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. Contact the author for more information.

11.2 Stream burning

One method of informing the channel derivation process to known stream locations is by using a hydrography dataset to 'burn' stream channels into the hydro DEM before performing flow direction calculations. This method of burning in streams is used as a pre-processing step in a variety of hydrological GIS applications, including ArcHydro and TauDEM. Users may want to pre-process their terrain data in this way before running the Process GEOGRID File tool, or alternatively, use an elevation dataset that has already been hydrologically enforced such as HydroSHEDS or NHDPlus. The authors have explored incorporating burn-in methodology as an option in the terrain processing steps. Every geographic location is different and there is no one-size-fits-all stream burning method that will work well in a global fashion. Although it may be a possibility in the future, the current release of the Process GEOGRID File tool does not include a stream burning component.

References

David, C. H, D. J. Gochis, D. R. Maidment, W. Yu, D. N. Yates, and Z. Yang , 2009, "Using NHDPlus as the Land Base for the Noah-distributed Model", Transactions in GIS 13(4): 363-377.

Jenson, S. K., and J. O. Domingue, 1988, "Extracting Topographic Structure from Digital Elevation Data for Geographic Information System Analysis", Photogrammetric Engineering and Remote Sensing 54(11): 1593–1600.

Monaghan, A. J., M. Barlage, J. Boehnert, C. L. Phillips, and O. V. Wilhelmi, 2013, "Overlapping Interests: The Impact of Geographic Coordinate Assumptions on Limited-Area Atmospheric Model Simulations", Monthly Weather Review 141: 2120-2127.

Peuker, T. K. and D. H. Douglas, 1975, "Detection of surface-specific points by local parallel processing of discrete terrain elevation data", Comput. Graphics Image Process., 4: 375-387.

Strahler, A. N., 1957, "Quantitative Analysis of Watershed Geomorphology", American Geophysical Union Transactions, 38: 913-920.

Tarboton, D. 1997, "A New Method for the Determination of Flow Directions and Contributing Areas in Grid Digital Elevation Models", Water Resources Research, 33(2): 309-319