Pitney Bowes Software

Discover Hydrology

Performance Strategies

Introduction

Hydrology is a computational expensive exercise in raster processing. To work an input DEM through pitfill, flow, stream and watersheds involves processing 10+ grids each the size of the input grid. Search "Tarbotan Raster Hydrology" to understand the process in detail.

Most of the 10+ grids must be open concurrently. The steps can't be absolutely serialized.

Processing Phase	Approximate Grids Used
Pitfill	6+ grids
Flow, Stream, Watershed	5+ grids
Other flow,distance etc	Yet more grids

If the input DEM is huge (larger than available RAM) it will cause the raster processing to become disc bound and provide slow performance.

This paper will help the user understand performance limitations and strategies.

Discover 2013.0.3 patch – RAM

Hydrology now uses close to the 32bit process 2Gb limit of RAM, which isn't much more than previously used but the way the RAM is allocated may make a difference.

Previously each grid opened concurrently was allocated 128Mb of RAM, even if the user only processed one step say pitfill (which actually needs 6 grids).

Now, the number of grids being opened is counted and the total available RAM for the process is divided evenly amongst the grids. This can make more RAM available to each grid if less steps are executed.

By default, the total RAM used by the process is 1920Mb, just inside the 32bit 2Gb limit.

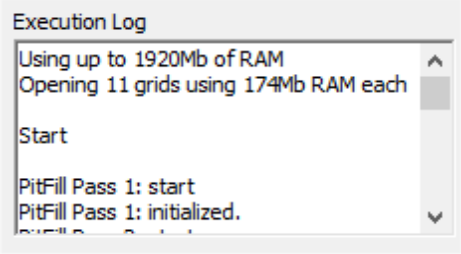
NOTE: As a safety mechanism the amount may be optionally configured in the registry setting (we only recommend this being touched by users experienced with registry settings. Do so at your own risk!!):

//HKEY_CURRENT_USER/Software/Discover/Plugins/GridDrainage/MemoryMb

The value supplied should be an integer of the number of megabytes to use and must be in the range of 200 to 2100.

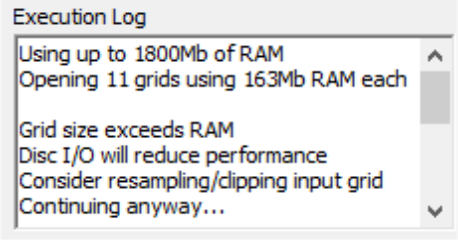
Log output

The output log now indicates how much RAM is being used.



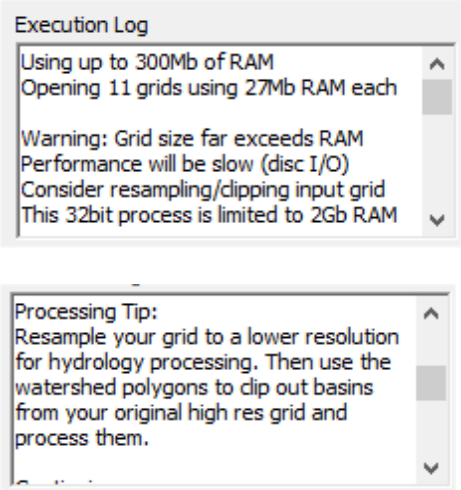
Default 1920Mb of RAM. Processing a small grid completely in memory.

When the input grid and number of steps will require more RAM than allocated, the Grid I/O system will start paging grid tiles to and from disc. This will slow performance and the log will indicate it.



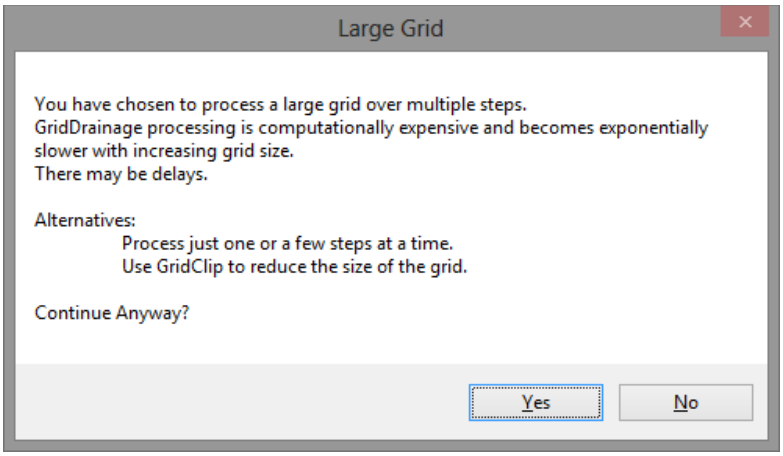
Processing a large grid. A hint that disc I/O will slow performance.

But when the grid size *far exceeds* the RAM available (more than 5 times the available RAM) the log will warn the user and print a processing tip. The processing tip is discussed below in this paper.



Processing a huge grid. A hint that disc I/O will drastically slow performance and a processing tip on how to work around it.

A message box will still be shown to the user allowing them to continue or cancel if the input DEM is huge. If the DEM is larger than 5000 rows or columns the following dialog will still be shown.



Processing Huge DEMs

It is possible to process huge grids with extra effort, and an acceptance of working within “basins”.

Three bad strategies and perhaps one acceptable strategy.

One simple strategy is to use `GridUtilities Clip` to clip your DEM down to the area of interest but this is only suitable if you're only concerned with a small part of your DEM.

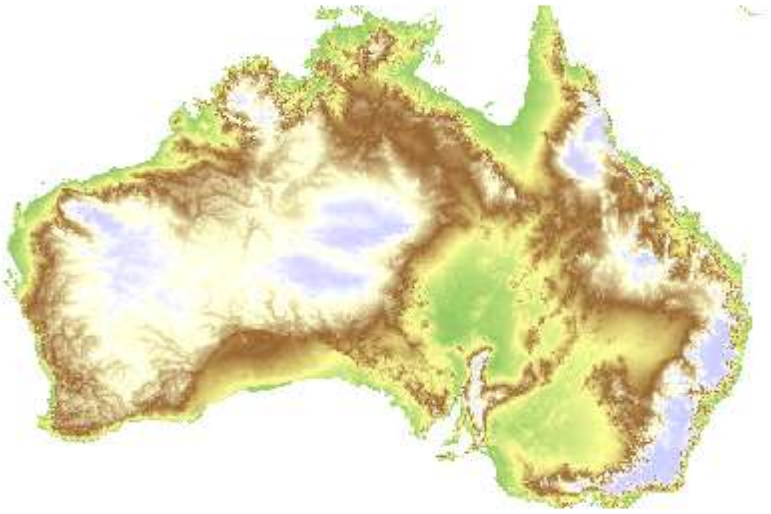
Another simple strategy is to use `GridUtilities Resample` to resample your DEM down to a lower resolution but this is only suitable if you don't mind losing resolution.

Another simple strategy (but not recommended) is to use `GridTiler` to break you're your grid into tiles and process them individually. This has a problem in that stream flows won't join up across tiles.

There is one technique that processes the full grid, works at suitable resolution and preserves flows to a degree. It is described here as was used to process Australia's SRTM grid (4.4Gb ERS file) within known flood basins.

Australia SRTM

The 4.4Gb SRTM file far exceeds the available RAM and will cause the hydrology process to be disc I/O bound.



The .ghx file shows how big the grid is

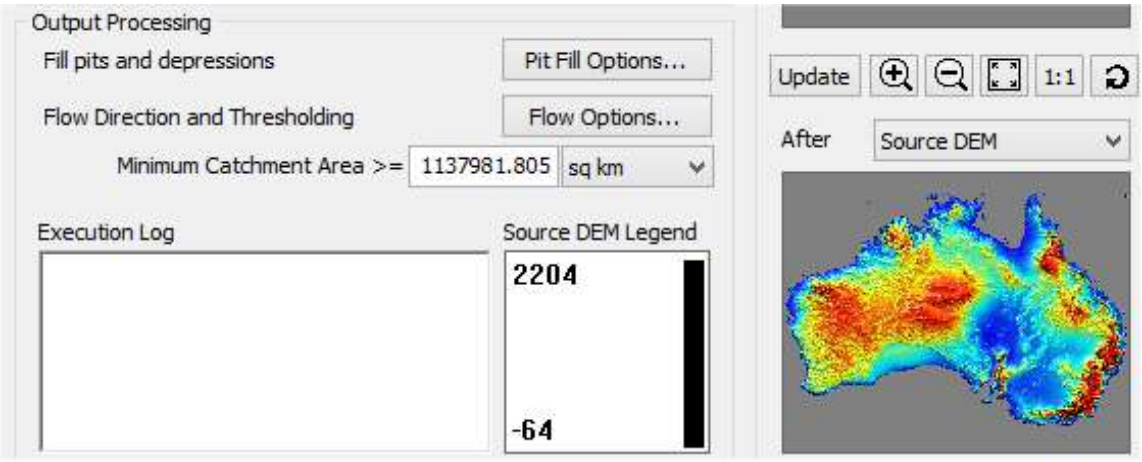
```
<rows>43669</rows>
<columns>51603</columns>
```

Resample

The first step is to resample the grid down to a manageable size (don't worry we'll do full resolution processing a bit further down). Use `GridUtilities Resample` to resample the grid down to below say 5000 by 5000 cells.

Basins

Now open the resampled grid in the Hydrology module.

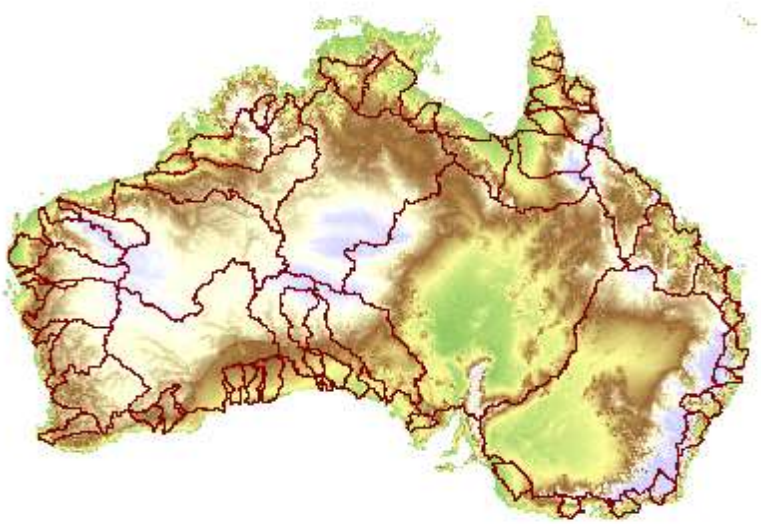


Default Minimum Catchment Area is 10% of the grid

By default, the Minimum Catchment Area displayed will be 10% of the size of the grid. This default is designed to produce a small number of watersheds in the final output. Making this number bigger produces fewer, but larger watersheds. Making it smaller could produce an unworkable number of watersheds.

We want to make this number large so that we produce just a few large watersheds which are major "basins" of drainage. The key to this technique is being able to work later at full resolution within a basin.

Edit the Minimum Catchment Area to be large and do full processing. Repeat with different sizes until you have a small number of watersheds that represent major basins.



A resampled Australia SRTM processed to a small number of watershed basins. Well actually in this example we weren't interested in the interior desert area so the number of basins is higher to get some key flooding basins like The Lockyer Valley (Brisbane) and Charlton (Victoria).

We really want to be left with a ring of basins that flow off the grid to be sure that they are isolated and independent.

Clip out full resolution basins

Now, In MapInfo, select a watershed basin and save it to a file.

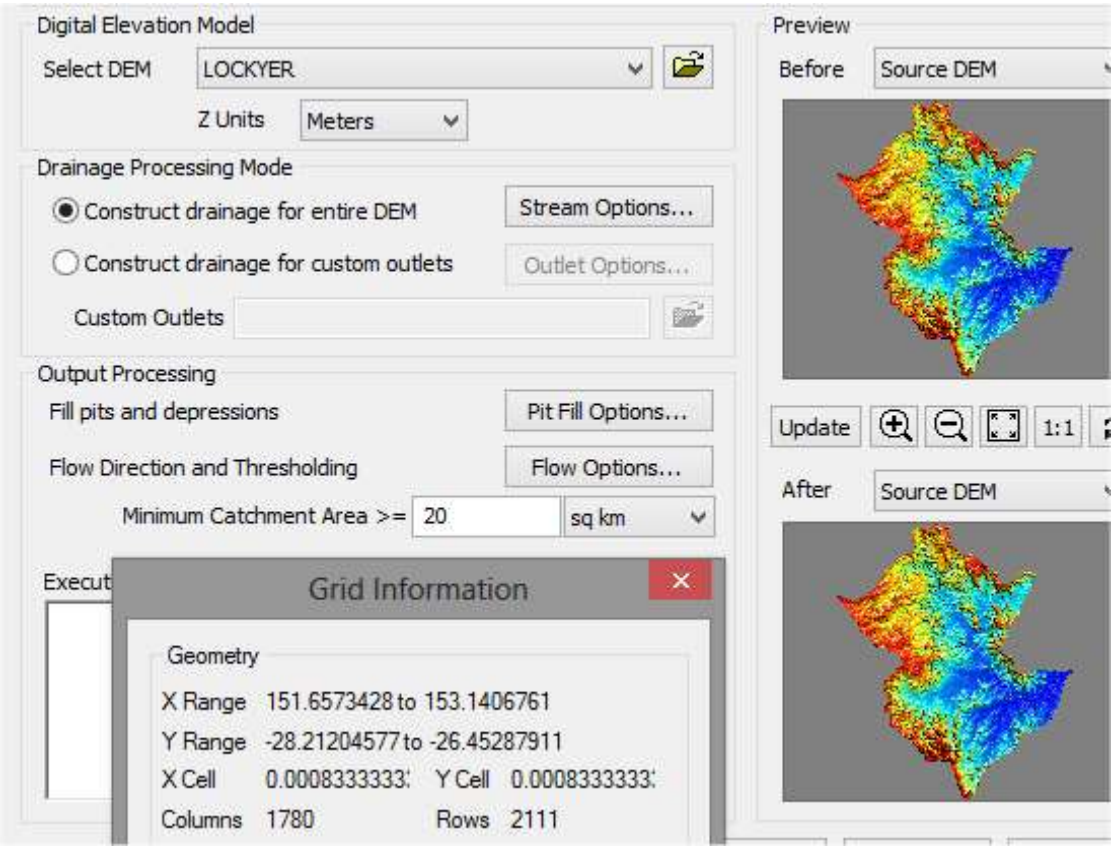
Use `GridUtilities Clip` to open your original full resolution DEM and clip out the basin, and save it to another file.

The Locker Valley Basin

The Locker Valley Basin DEM clipped out of the original full resolution DEM

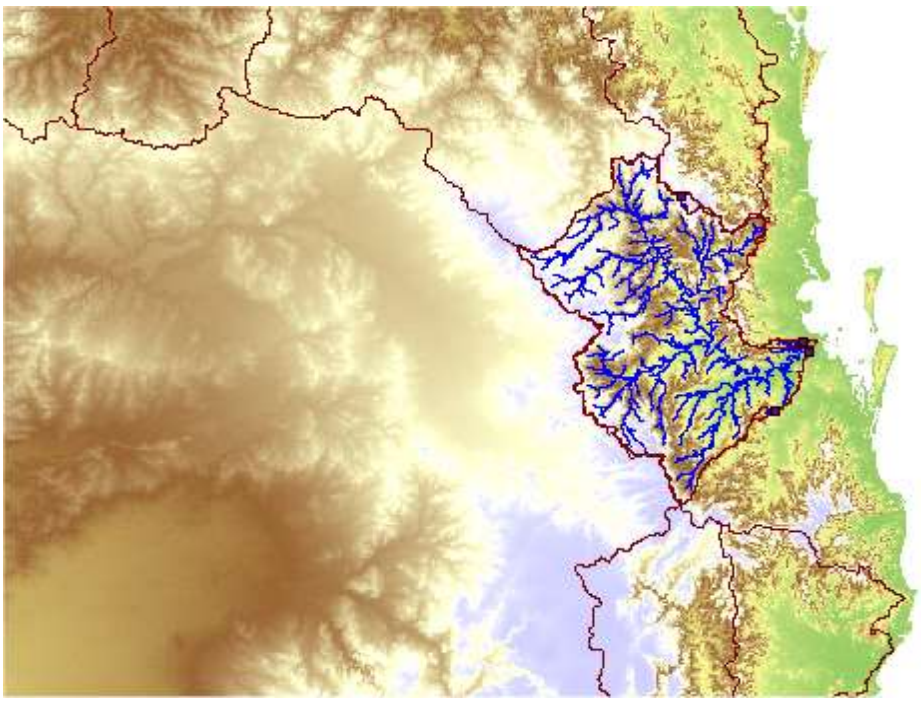
Process the basin

In the Hydrology module process the basin.



The clipped out basin is a smaller more manageable grid (1780 x 2111). Process it at a smaller Minimum Catchment Area (20km).

The Locker Valley Basin streams and outlets. It all drains through Brisbane except for a few small artefacts cause by the prior low resolution processing and clip.



The basin within the original grid.

The streams and watersheds of all the basins will be independent and non-overlapping, but flow accumulation values won't grow across basin boundaries. This leaves a final conclusion.

Conclusion

If flow accumulation values at a grid cell are important across the whole grid, then resample your whole grid down to a manageable size (say less than 5000x5000) and process it fully.

But if stream resolution is important, then use the above technique to clip out a basin and process it fully at a high resolution.

Todd Spurling, Software Engineer

Pitney Bowes Software