A Catchment Conflation Tool for Comparing DEM Derived River Networks

Arthur M Read
CSIRO
Canberra, Australia
arthur.read@csiro.au

Abstract— **This paper describes a tool for identifying topologically matched locations in vector stream networks extracted from different DEMs of the same catchment. The tool has applications in providing quantitative comparisons of stream network position and topological differences as well as providing a scalable and robust method for conflating stream networks to enable transfer of attribute information between networks. The tool is demonstrated using a case study in the upper Murrumbidgee River in Australia showing how matching stream links are identified and topological differences identified.**

INTRODUCTION

Stream networks and catchments are a common and important product derived from digital elevation models (DEMs). Different DEMs of the same spatial area will produce different realisations of the river network and catchments. The amount of difference in the derived network and catchments will depend on the different data sources, DEM processing methods and on the topography. For example the vertical resolution and accuracy of the DEM can have a strong impact on flow direction in low relief terrain while horizontal resolution and accuracy have a stronger impact in high relief terrain.

A comparison of the network and catchments from different sources is an important assessment of a DEMs quality and can indicate the DEMs suitability for a particular purpose. Comparing the networks spatially is difficult and most techniques rely on measurement of distance between network features. When the drainage density is high relative to the difference in the spatial location of streams distance measures tend to result in comparing streams from different parts of the network and thus under-estimating the positional error by ignoring the networks topology.

A related issue is the conflation of vector stream segments to enable the transfer of attribute data from one river network to another representation of the same physical network. Any method that relates stream segments but ignores their topological position in the network can only work in simple cases and often leaves the user with a manual checking and fixing process.

Network comparison and network conflation are both examples of the more general problem of identifying corresponding features based on topological similarity rather than spatial proximity.

Previous Work

It is well established that different DEM sources and processing steps lead to differences in extracted networks and catchments. These differences can be identified visually but quantification is more difficult. Some researchers have indirectly quantified network differences by investigating the effect of different DEMs on the output of a hydrological model [1]. Others have looked at the stream network more directly. Gatziolis and Fried [2] used a Euclidean distance between stream cells in the raster networks being compared as the basis for their work. Hengl and Reuter [3] also compare the distance between stream lines but on a vector stream network. Hengl et al [4] and Lindsay [5] compare a large number of stream network realisations by looking at the probability of any DEM cells being part of a stream network. This provides a measure of sensitivity to possible stream location error not a direct analysis of any pair of networks.

While the measures used in previous work are all valid they potentially under-estimate the differences in the stream networks being compared. There are 2 key cases where this can occur:

1) When the density of the stream network being analysed is high relative to the spatial error in the stream location, measures of distance of collocation can end up comparing very different network features.

2) When the DEMs differ in ways that make the networks topologically different such as a stream that flows down different valleys in the different DEMs resulting in a very different catchment structure for parts of the network downstream.

These cases are important when using the network for further analysis and therefore it is important that a metric of the quality of stream location from a DEM is able to capture this information. These differences can be identified visually on close inspection but an automated method would provide repeatability, consistency and the ability to scale to large networks. If locations in the stream networks could be matched using position within the catchment structure, not just spatial proximity, then this would strengthen any analysis of error in streams extracted from DEMs.

METHOD

Catchment Conflation Method

The new work that this paper presents is a method for matching (conflating) features in 2 river networks and thus allowing an improved and detailed understanding of how and where the 2 networks differ in terms of spatial location and topological connectivity. By identifying congruent locations in each network the tool is able to generate a lookup table that maps feature to feature between the two networks.

Network locations are defined as congruent based on how closely their entire catchment polygons overlay spatially. Because this is performed by comparing the entire catchment upstream of the network, location differences in the shape and position of the streamlines and watershed boundaries have little effect. The greater the catchments' size the less impact small watershed differences have. By using a catchment comparison the topology of the network is implicitly captured in the comparison.

The tool has been developed using Python 2.7 (64 bit) [6] and makes use of a number of additional modules: GDAL [7] for data IO, NetworkX [8] for network tracing, and Shapely [9] for vector spatial overlay operations. The tool has been written in such a way that computationally intensive parts are applied in parallel using IPython [10] and could be easily adapted to other ways of running parallel code.

The tool takes two sets of vector stream lines and catchment polygons as input data. The generation of these from the DEMs can be performed in a variety of ways in many different GIS packages. The user has the flexibility to use which ever method they see as appropriate. A unique identifier links the stream lines to the catchment polygons in each realisation of the network.

One possible use of this tool is comparing a DEM-derived network with a digitised network. This is possible if the digitised network is first enforced into a DEM (AGREE [11], stream burning [12], ANUDEM [13]) to allow the extraction of catchments and ensure correct network structure. The tool is capable of working with bifurcations in the networks.

The tool uses the following processing steps:

1. Load data: The software reads in the two sets of input data and builds two directed acyclic graphs using NetworkX where each edge represents a stream link and its local subcatchment.

2. Produce catchments: Each river network is then traversed from the top of the graph (and river network) to the bottom and for each edge performing a geometric union of the local subcatchment and the upstream catchments such that each edge is attributed with a polygon that represents its entire catchment

3. Search for congruent locations: For each edge/catchment in one network (networkA) search for the edge/catchment in the other network (networkB) to find the catchment that is most spatially similar i.e. both catchments have a high proportion of their area intersected by the other catchment. This produces a table of one to one relationships between then downstream end of each edge in networkA with the downstream end of an edge in networkB and a measure of how well each pair of points match.

4. Produce relationship tables: A conversion process is then applied to the point relationship table to provide a table of relationships between stream lines and another for catchments. These may have a one-to-many relationship where the networks have different densities. This step also produces a list identifying locations in the network where topological inconsistencies exist.

The search for congruent locations (step 3) is the part of the process that determines the quality of the conflation between the networks. It is also the most computationally intensive part of the process. Key to this step is the search through networkB for a match for each catchment in networkA. The area of the intersection alone is not sufficient to find a matching catchment. Because of the nested nature of the catchments a small catchment can be completely intersected by a much larger catchment but there is likely to be a better match further upstream where both catchments intersect by a high percentage.

There are a number of ways to perform the search through networkB to find the best match. A brute force comparison with all edges would work but would not scale well to large networks. Another approach is to find one match at the outlet of the network and then step up the network using the downstream result as a starting point. This works if the networks are topologically identical but becomes more difficult it the networks have even minor inconsistencies. This method is also difficult to apply in parallel as the downstream solution is required before an edge can be processed.

The approach taken to the search is, for each catchment in networkA, to start at the outlet of networkB and work upstream only taking a single path that has the best match. This "best first" search quickly leads to the matching catchment and avoids the issues with the other approaches. Most importantly searches are independent and can be applied in parallel.

The output from step 4 provides all the information required to transfer attribute data between the two networks and investigate spatial and topological differences between the networks. The output of this tool is the basis for this spatial analysis but further work is required to take the output and perform the required analysis.

Study Area

To demonstrate the tool a study of the upper Murrumbidgee River catchment in Australia was performed. The first stream network (networkA) used is the Australian Hydrological Geospatial Fabric [14] which for the purposes of this study provides a vector stream network and catchments that have been derived from the 9 second DEM of Australia [15] to match the 1:250000 stream mapping. The second network (networkA) has been extracted from the 1 second DEM-S [16] which is a processed version of the SRTM data [17]. A flow accumulation threshold was used to extract the streams so that the network extents closely matched the 1:25000 stream mapping. These networks have different stream densities which will allow a demonstration of how the tool is able to deal with this situation.

The methods used for the extraction of the networks from the DEM are not important in this demonstration. The tool is independent of these steps and can in fact be used to understand the effect of different processing steps on the network produced.

Results

The networks have 2352 and 39786 stream links respectively. Running on 2.27Ghz CPUs with the search step run in series the whole process takes approx 3 hours 10 minutes. When the search process is run in parallel on 64 threads the total run time drops to approximately 10 minutes.

The output relationship tables allow identifiers and attributes to be moved between the networks. Fig 1 shows the links of networkA randomly coloured and networkBs links coloured to match. It is worth noting that each coloured link in networkB is actually made up of many links as it is a much finer network. One limitation of the method is that first order streams in networkA can only be identified for the first link in networkB as finding the correct path further upstream requires an upstream catchment conflation. A similar relationship can be made at the sub catchment level as shown in Fig 2.

The tool outputs statistics of catchment area and the area of the intersection of the catchments for each conflated pair. Using these values we can quantitatively assess the conflation result. One useful statistic is a ratio of catchment intersection area to catchment union area. Fig 3 shows the distribution of this ratio against catchment area. The larger catchments of the main river channel have a much higher ratio of area intersecting and therefore confidence in the conflation result. The lower ratio values tend to be first order channels and while the tool has found

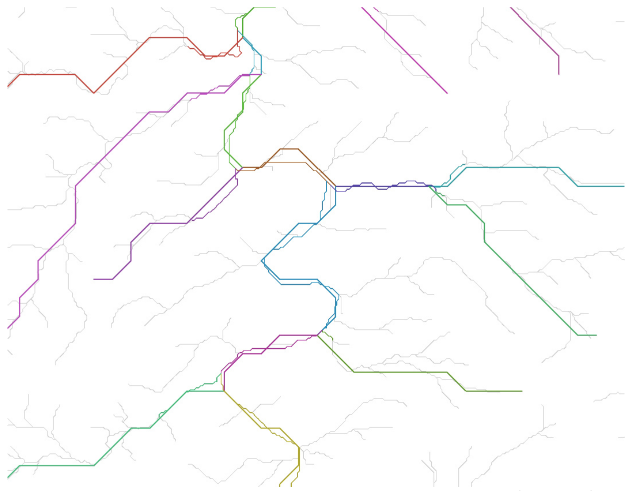

Figure 1. Randomly coloured stream links showing conflation of reaches between the two networks.

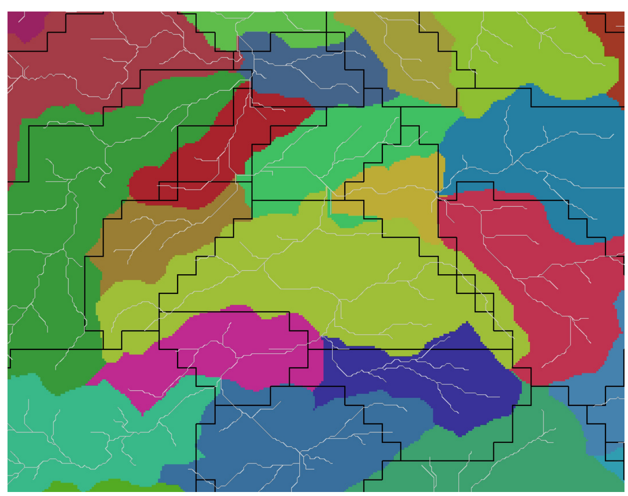

Figure 2. networkB sub catchments coloured in groups to represent conflation to networkA sub catchments (shown as dark outlines). NetworkB streams shown in light grey.

the best possible match the source DEMs differ to such a degree that the small catchments are not very similar, as can be seen in Figure 2. This leads to the possibility of using these results to assess the DEMs relative ability to represent catchments at different scales. The few catchments with unusually low values highlight locations of significant topological differences in the stream networks.

DISCUSSION

This method presents a conceptual improvement over other measures used to quantify differences in the quality of a DEM's representation of a stream network and catchments. This is achieved by matching locations on both stream networks based on their position within the catchment rather then spatial proximity. The measures of catchment overlap and difference capture both spatial and topological similarity between the two networks.

This conflation of the networks allows for comparison and analysis of any non-spatial attributes of the networks or attributes derived from the spatial structure of the network and catchments.

The conflation also produces an analysis of topological inconsistencies between the networks as well as providing a topologically sound platform for developing analytical method for comparing the hydrologic properties of two DEMs.

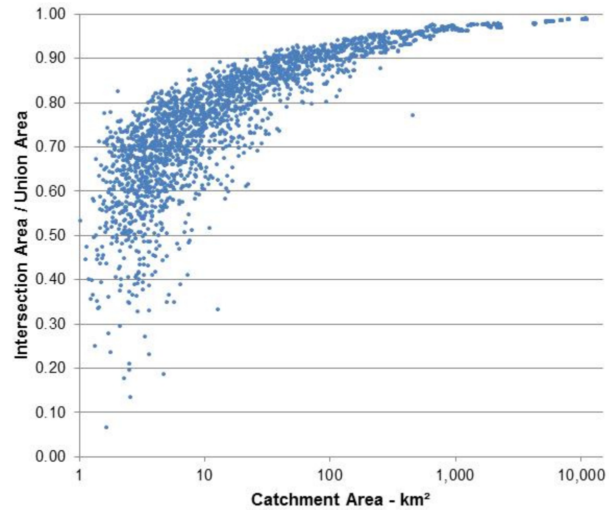

Figure 3. Distribution of the ratio of intersection area to union area against catchment area for each pair of conflated catchments.

ACKNOWLEDGMENT

This work was funded by CSIRO's Water for a Healthy Country research flagship and the Bureau of Meteorology.

REFERENCES

[1] Wise, S.: Assessing the quality for hydrological applications of digital elevation models derived from contours, Hydrol. Processes, 14, 1909–1929, 2000.

[2] Gatziolis, D., and J. S. Fried (2004), Adding Gaussian noise to inaccurate digital elevation models improves spatial fidelity of derived drainage networks, Water Resour. Res., 40

[3] Hengl, T., and H. I. Reuter. How accurate and usable is GDEM? A statistical assessment of GDEM using LiDAR data. In Geomorphometry 2011, edited by T. Hengl, I. S. Evans, J. P. Wilson and M. Gould, 45-48. Redlands, CA, 2011.

[4] Hengl, T., Heuvelink, G. B. M., van Loon, E. E., 2010. "On the uncertainty of stream networks derived from elevation data: the error propagation approach". Hydrology and Earth System Sciences, 14:1153-1165

[5] Lindsay, J. B. (2006) 'Sensitivity of channel mapping techniques to uncertainty in digital elevation data', International Journal of Geographical Information Science, 20:6, 669 – 692

[6] Python Software Foundation. (2011) Python version 2.7. Available online at: http://www.python.org

[7] GDAL – Geospatial Data Abstraction Library. Available online at: http://www.gdal.org/

[8] NetworkX (2011) Available online at: http://networkx.lanl.gov

[9] Shapely. (2011). GIS & Python & Invention. Available online at: http://trac.gispython.org/lab/wiki/Shapely

[10] F. Perez and B.E. Granger, IPython: A system for interactive scientific computing, *Computing in Science and Engineering* **9** (3) (2007), p. 21

[11] Hellweger, F., 1997. AGREE — DEM surface reconditioning system. Center for Research in Water Resources, University of Texas at Austin. Available online at: http://www.ce.utexas.edu/prof/maidment/GISHydro/home.html

[12] Saunders, W., 1999. Preparation of DEMs for use in environmental modeling analysis. ESRI User Conference, July 24–30, 1999, San Diego, California. Available online at: http://www.esri.com/library/userconf/proc99/proceed/papers/pap802/p802.htm

[13] Hutchinson MF, 2006, ANUDEM Version 5.2. Fenner School of Environment and Society, Australian National University. Available online at: http://fennerschool.anu.edu.au/publications/software/anudem.php

[14] BOM (2011) Australian Hydrological Geospatial Fabric. Available online at: http://www.bom.gov.au/water/geofabric

[15] Geoscience Australia (2008) GEODATA 9 Second Digital Elevation Model (DEM-9S) Version 3. Available online at: http://www.ga.gov.au/meta/ANZCW0703011541.html

[16] Gallant, J., Dowling, T., Read, A., Tickle, P., Wilson, N. and Inskeep, C. (2010) 1 second SRTM Derived Digital Elevation Models User Guide 1 second DSM, DEM & DEM-S 3 second DSM, DEM & DEM-S Version 1.0.3. Available online at: http://www.ga.gov.au/image_cache/GA18067.pdf

[17] Gallant, J.C. and Read, A. (2009). Enhancing the SRTM data for Australia, In Geomorphometry 2009,(Ed. R. Purves, S. Gruber, R. Straumann & T. Hengl).