The Rio Chagres: A Multidisciplinary Profile of a Tropical Watershed

R. Harmon (Ed.), Springer/Kluwer, p.83-95

GIS-based Stream Network Analysis for The Chagres Basin, Republic of Panama

David Kinner[1,2], Helena Mitasova[3,4], Robert Stallard[1], Russell Harmon[4], and Laura Toma[5]

[1]U. S. Geological Survey Water Resources Division, Boulder, CO, [2]INSTAR, University of Colorado, Boulder CO; [3]Dept. of Marine, Earth, and Atmospheric Sciences, North Carolina State University, Raleigh, NC, [4]Army Research Office, Army Research Laboratory, Research Triangle Park, NC, [5]Dept. of Computer Science, Duke University, Durham, NC

ABSTRACT

To support a number of projects focused on diverse biological and physical aspects of Chagres Basin a detailed stream network was extracted from digital elevation data obtained by interferometric radar survey. The elevation data represented the earth surface plus a varying forest canopy height, therefore different algorithms for stream network extraction were qualitatively evaluated in terms of their capability to extract accurate stream location from this challenging type of elevation data. The program based on shortest path algorithm and iterative linking provided stream location that was closer to on-ground GPS measurements than the tools based on filling the depressions and iterative linking. The effect of different spatial resolutions on network structure and orientation was also explored.

2.1. INTRODUCTION

The Chagres Basin study includes investigators with a diverse set of interests and goals. The common objective of participants is to provide information about hydrology, geology, and ecosystems within the basin. This region of the Panama Canal Watershed was mapped at an extremely coarse resolution until the mid 1990s (1:2,000,000). The relative inaccessibility of the upper Rio Chagres and a persistent cloud cover over the basin's western edge have inhibited mapping of this region. A recent Interferometric Synthetic Aperture Radar Elevation (IFSARE) survey by the United States Army, LANDSAT imagery, and subsequent field investigations have provided new, more detailed information about the topography, land cover and streams in the basin. To support continuing research, the paper describes existing and new geospatial data integrated within a common Geographic Information System (GIS) database. This integration creates a spatial framework for the field observation data obtained along the main channel of the Rio Chagres during the 2002 field season.

The GIS database provided an environment for deriving secondary map layers with valuable information both for the fieldwork and modelling. One of the most important derived map layers for a hydro-geologic study is stream network. In a tropical forested landscape, it is usually impossible to map the locations of streams from satellite or aerial imagery, particularly those of a lower-stream order. The only recourse is to apply terrain analysis algorithms to a Digital Elevation Model (DEM) to determine the channel network. Various approaches were developed to perform this task, most of them designed for the bare ground 30m USGS DEM product (see O'Callahan and Mark (1984), Peckham (1998), Garbrecht and Martz (1997) and Jenson and Domingue (1988)). However, IFSARE-based elevation models have different properties than the DEMs used for the development of the stream extraction tools. It is therefore necessary to evaluate different methods and identify the one most suitable for the data available and the type of terrain in the study area. Georeferenced stream and geology data were collected at selected sites along the the Rio Chagres and its tributaries, providing on-ground data for verification of extracted stream network and evaluation of flow routing algorithms and the quality of the IFSARE data set. The modelled and observed data are compared qualitatively in this work.

2.2. STREAM NETWORK ANALYSIS

Generating river networks using digital-elevation data is a multi-step, interpretive process. The assumptions that resolve flow directions on DEMs will ultimately effect the locations of the derived-stream network. Properties like network structure, length, and bifurcation ratios may be affected by choice of stream delineation algorithm. Accuracy in stream network position is particularly critical in the case of the Chagres River Basin because there has not been extensive mapping of the river channels and the analysis output provides the "best-available" map of the river network. Moreover, the elevation data – the 10m resolution digital surface model (DSM) based on the IFSARE survey - represents the surface of the earth with vegetation rather than bare ground, posing additional challenges for stream extraction.

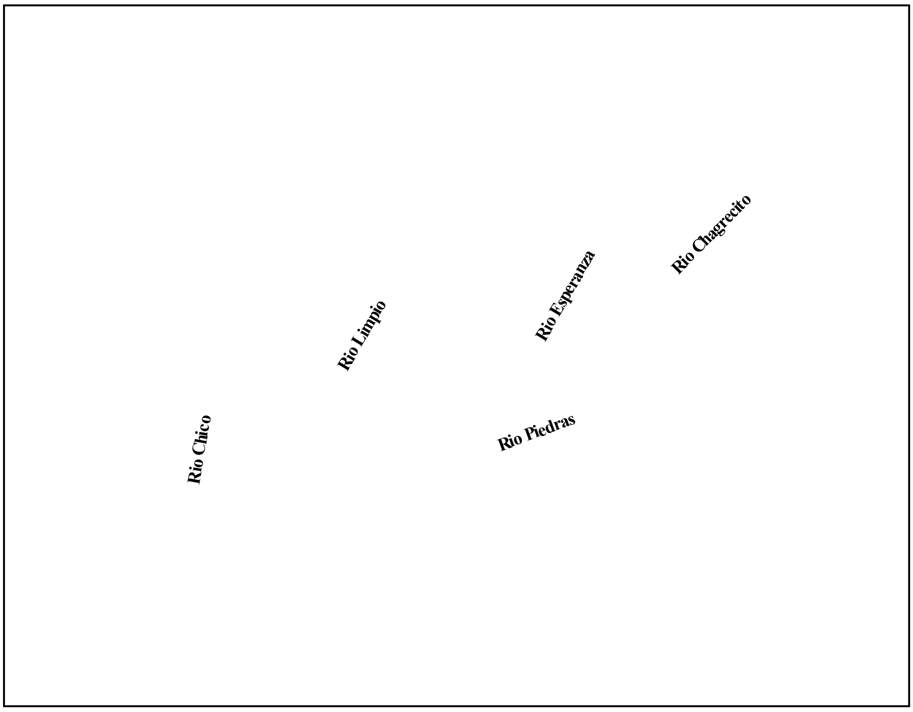


Figure 1. Map of Chagres River Basin showing contours, rivers and the basin boundary, all derived from an IFSARE DEM.

Two systems with tools for watershed analysis were used to extract the stream network:

a) Rivertools^{TM} (any use of trade, product or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.) - a commercially-available terrain analysis system for analyzing DEM-derived river basin characteristics (Rivix Limited Liability Company, 2001). It includes single and multiple flow direction (SFD, MFD) algorithms, visualization tools and tools for extracting statistical properties of river networks (i.e., bifurcation ratio, statistical similarity, fractal dimension, etc.).

b) GRASS GIS - an Open Source/Free software general purpose GIS. It includes a number of modules for basin analysis, We have used a SFD-based **r.watershed** and SFD/MFD-based r.terraflow for massive DEMs. The Panama Canal Watershed DEM with 11,000x11,000 pixels is an example of a "massive" DEM.

In general, the stream extraction algorithms compute the stream network by routing flow through the DEM and using a selected threshold of stream-order or contributing upslope area to determine which cells are stream cells.

2.1 Filling sinks in IFSARE Data

The first step used by Rivertools^{TM} and r.terraflow in determining the location of river networks is creating a depression-less landscape by removing sinks. Sinks are DEM pixels that have a lower elevation than all of the surrounding pixels. These sinks can be hydrologic features like lakes, natural depressions or sink-holes, or they could be artifacts of the DEM-construction process. For the Chagres DSM, an intermittent canopy overhanging stream channels may have generated at least some of the depressions. In some locations, the radar survey may measure the distance to the riverbed; in other locations measure the top of the canopy. This discontinuity means that areas that have no cover may appear to be sinks and, for some algorithms, they have to be filled to the level of the surrounding cells in order to route flow through topography. In a radar-based DSM, this filling may be a source of additional error in the stream delineation process.

Figure 2 shows a longitudinal profile comparison of the initial and Rivertools™ filled IFSARE DSMs at a 25-m grid cell resolution. As shown in the figure, the filling process creates a step-like stream profile, but the DSM-based stream profile has many elevation spikes. Also, near the mouth of the Chagres (distance of 0-7 km on the x-axis in Figure 2), elevation errors are dampened. Land use data indicate there is bare soil, shrubs and grass through this reach-- all features that would likely cause fewer elevation errors than an overhanging forest canopy. The average filling necessary to route water through the reach in Figure 2 is 7.68 m. To put this in some context, average filling of 40-km long headwater reaches in the Nishnabotna River Basin in Iowa and Boulder Creek Watershed in Colorado are 0.48 m and 0.99 m respectively. These estimates were created using USGS 1:25,000, comparably scaled, 30m resolution grid cell DEMs that were constructed from topographic maps. The combined effects of IFSARE radar and the forest canopy appear to significantly increase the number, depth and spatial extent of sinks in the DSM.

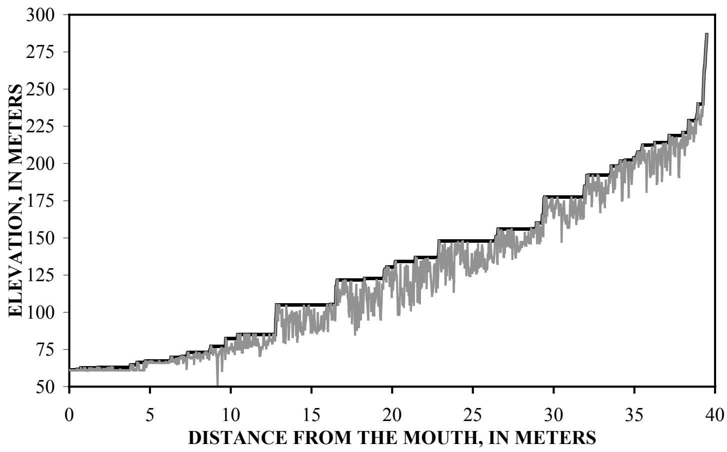


Figure 2. Comparison of filled topography (dark line) and IFSARE original topography (gray line) for a 40 km reach of the Upper Chagres River. Distance 0.0 represents Lake Madden.

It is necessary to make a cautionary note about the IFSARE-based DSM. Because there are many sinks in the landscape, slope and curvature estimates may be affected by holes in the topography. It is necessary to average stream slope estimates over a large enough area to remove the effects of these discontinuities in elevation. Smoothing of topography also may be useful to better compute average gradients and terrain characteristics.

2.2 Computing Flow Direction and Resolving Flats

A core step in stream network analysis is resolving the flow directions of water on the landscape for each elevation cell in a DEM. With the exception of r.terraflow (see section 3.4), we used a single-direction flow algorithm (SFD)with eight possible directions of flow (D-8) (O'Callahan and Mark, 1984). For a given grid-cell, the slope from the center of that cell to the center of an adjacent cell is calculated for all of the cells in an eight-cell neighborhood. "Flow" proceeds in the direction of the cell with the greatest slope. Contributing areas for each pixel are determined recursively in a down gradient direction and represent the number of cells that "flow" into a given cell.

Flow direction using the D-8 algorithm is easy to compute in areas where slope is well-defined. As the size of a river basin increases, channels become larger and the river gradient decreases. The areas where the gradient magnitude is zero are called "flats", for which the flow direction is undefined. In DEMs represented with cm to mm vertical precision areas with zero slope are rare (e.g. in the currently used floating point DEMs as opposed to older integer DEMs that had vertical precision of 1m) and most flats are created by filling large, multi-pixel depressions like the ones shown in Figure 2. Several approaches have been developed to route flow through flats, of which only iterative linking and imposed gradients are considered here.

Iterative linking (Jenson and Domingue, 1988, Peckham, 1998) defines the flow direction for all of the cells in a flat. It starts from the spill points of flat areas and iteratively assigns flow directions to the neighbours of the spill points, then to the neighbours of the neighbours, and continues recursively until all of the flow directions in a flat are defined. The final form of the network therefore depends heavily on the order that cells are defined (cells could be defined along a row for example) rather than any physically-based criteria. There is no physical basis to iterative linking, but it will resolve flow direction and can be numerically efficient.

Imposed gradients is a second flat resolution approach described by Garbrecht and Martz (1997). In this method, artificial topography is created over flat areas by adding micrometer increments to the elevations of the flat. This addition is completed both longitudinally and laterally such that topography slopes downstream and from the sides of the flats towards the outlet. In our experience, imposed gradients tends to translate the shape of the valley walls to the shape of the channel.

The r.watershed flow direction algorithm differs from the D-8 approach in several fundamental respects (Ehlschaleger, 1989; Ehlschaleger, 1991). Topography is not filled in the r.watershed program; rather flow is routed through sinks. The r.watershed flow direction algorithm is based on the A^T search algorithm for finding the "least cost" path between a upstream cell and a downstream cell. The sum of the elevations along a given path represents the "cost" of the path. The algorithm begins at the watershed outlet or any internal sinks, like a lake, and begins working upslope (Ehschlaeger, 1989).

In the case of D-8 algorithms like those implemented in Rivertools™, flow can only be routed to cells of a lower elevation. Because of this requirement, sinks must be filled so that the slope from any grid cell is at least zero. Negative slopes are not permitted along a flow path. In the A^T method, flow directions can proceed in the direction of negative slope. Because sinks are not removed from the dataset, more of the original information from the DEM is preserved in deriving final flow grid. This preservation may enhance the accuracy of flow paths in areas that are filled by other algorithms because some of the filled data could be close to the actual elevations of stream channels.

3.2.3 Algorithm Comparison

Figure 3 includes a comparison of channels generated using iterative linking, imposed gradient and the shortest path methods with GPS point locations of the Rio Chagres channel for the section labeled “study reach” in Figure 1. All three methods seem to track the observed channel fairly well for most of the measured reach. However, both the imposed gradients and the shortest path methods track the channel better than iterative linking in the area of filled depression. The difference suggests that qualitatively, the combined use of filling sinks and iterative linking reduces the accuracy of stream extraction a complex DSM with spatially variable vegetation cover.

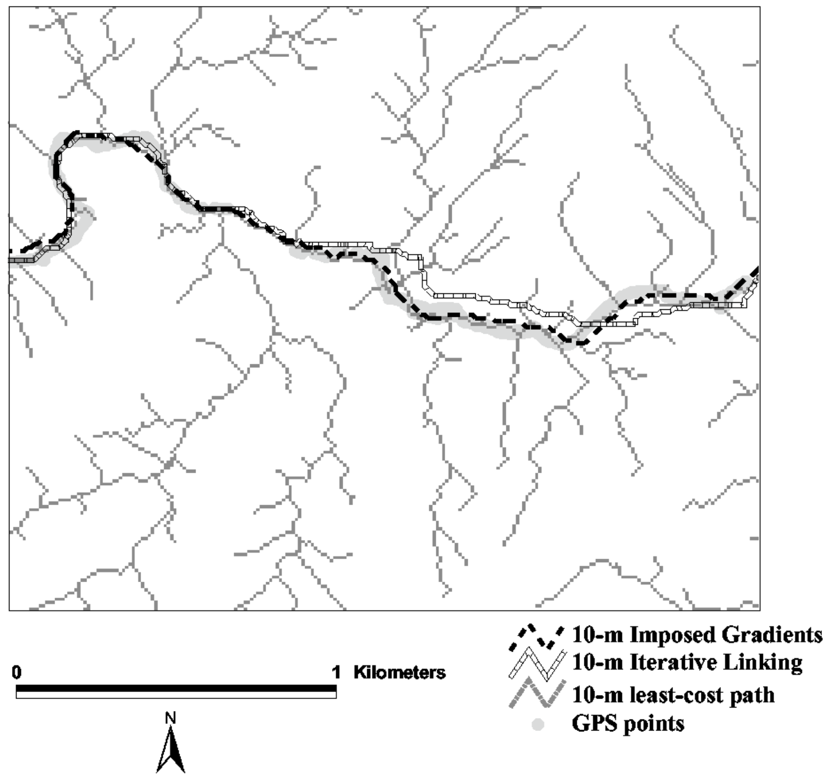
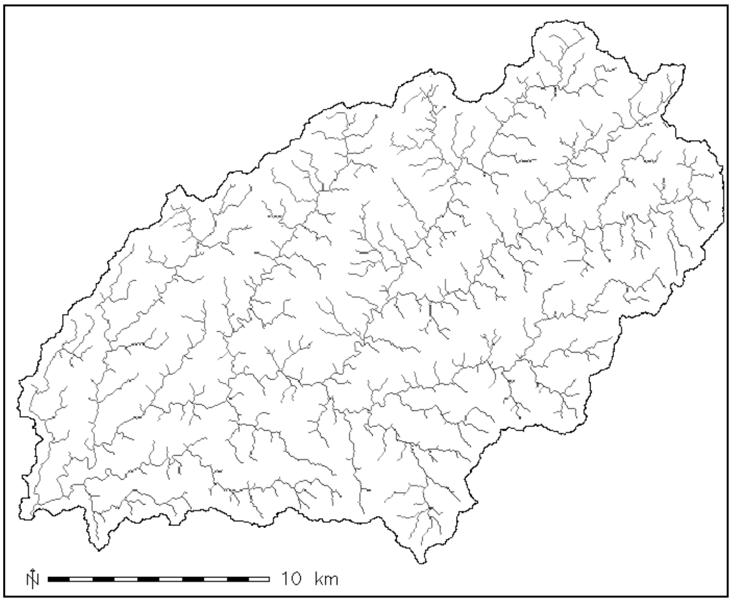


Figure 3 .Comparison of observed GPS data (light gray symbols in background) and flow networks derived using the r.watershed shortest path approach (light gray), Rivertools^{TM} imposed gradient (darker gray), and Rivertools^{TM} iterative linking (darkest gray).

The comparison presented here supports work completed on Boulder Creek in Colorado (Kinner, 2003) that indicated qualitatively that the imposed gradient method of Garbrecht and Martz (1997) performed better than iterative linking. However, several Chagres Basin subwatersheds are excluded from the river basin when we used imposed gradients. We believe that this is because the topography constructed on the large flats becomes higher than the surrounding terrain and creates sinks in the already filled topography. The imposed gradients method was designed for USGS 1-m scale DEMs and not the floating point, millimeter precision DEMs (Peckham, personal communication). This problem does not appear to affect the delineation of the main channel in Figure 3.

2.3 Extracting the stream network using Rivertools and r.watershed

In this work, two different methods were used to differentiate grid cells that represent streams and those that are part of the surrounding landscape. After completing the flow direction step in Rivertools^{TM}, we used the flow paths to compute Strahler stream order. This process starts at ridge cells and follows flowpaths down gradient. The flow paths that begin at ridges are considered first order streams; if two streams of the same order n converge, then Strahler ordering rules indicate that the stream below the convergence has a stream order of $n+1$ (Strahler, 1957). This process continues in a downstream direction until all streams receive a Strahler order. We used Rivertools^{TM} to prune the stream network by cutting all streams below a given Strahler order threshold, which for the 10-m DEM, was a third-order stream.



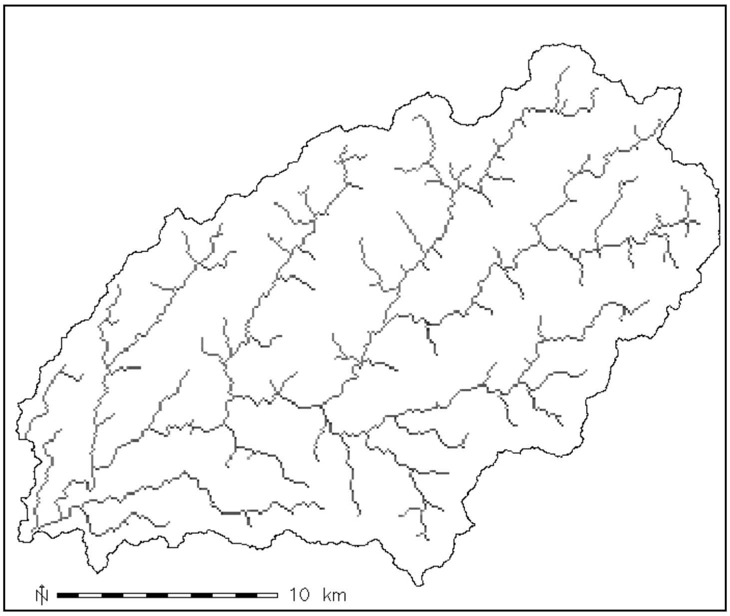
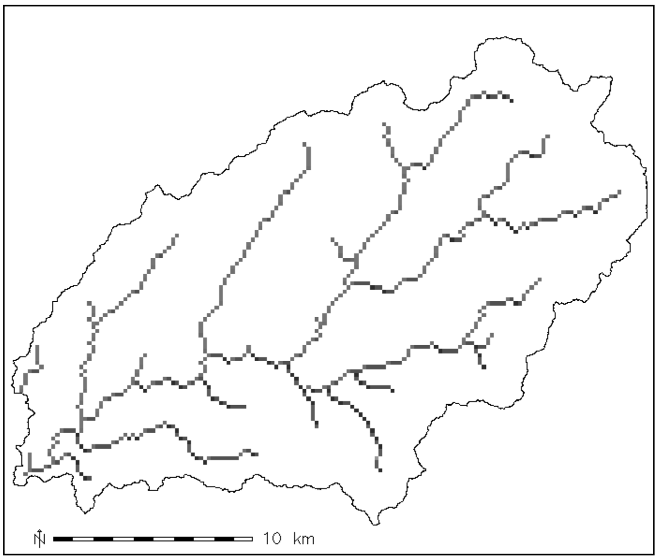


Figure 4. a): Stream network derived using r.watershed and 50 m grid cells. *4b)*: Stream network derived using r.watershed and 100 m grid cells.



Figure, 4*c):* Stream network derived using r.watershed and 200 m DEM.

The location of streams can be set by a threshold size of upslope contributing area in r.watershed. We have used this approach with the threshold area equal to 100 cells. To compare the impact of resolution on the level of detail and structure of the extracted stream network we performed the analysis at 10, 25, 50m, 100m and 200m resolutions, so the selected threshold area has increased with increased cell size, leading to effective generalization of the resulting network. The selection of the 100 unit threshold was arbitrary, based on cartographic rather than physical criteria. This leads to fewer major streams being extracted at lower resolutions. The resulting series of stream networks extracted from DSM at 50m, 100m and 200m resolutions is in Figure 4.

Extraction of stream networks at a series of decreasing resolutions (Figure 4) reveals the stream network geometry at different scales and levels of detail. The coarser-scale resolution highlights the difference between the northern half-basin with clearly parallel stream orientation and the southern half-basin with greater variability of stream orientation and lengths. It is unclear what may control the fundamental differences in stream structure between the northern and southern halves of the river basin. However, the coarser data more readily resolve the underlying network structure.

The density and structure of stream network derived from the 10m resolution DSM and 100 cells = 10,000m^2 threshold by r.watershed was practically identical with the stream network extracted by Rivertools with the Strahler stream order 4 at the same resolution; the most significant differences were in the areas with large filled sinks. New field data will allow us to set the threshold more accurately and to determine whether contributing area, stream order or curvature provides the best method to determine where channels begin. The vector representation of the stream network was then created by converting the raster stream network to vector format.

2.4 Computing the stream network for Chagres Basin and entire Panama using r.terraflow

As apparent from the multi-agorithmic comparison, the program r.watershed works fast and produces reasonable matches to observed data when applied to small datasets. However, this algorithm is not efficient on larger datasets. For example, the computation of the stream network for the Chagres watershed at 10m resolution (3200x3600 pixels) took 12 hours. Computation of the stream network for all of the Panama Canal Watershed (~11,000X11000 pixels) is unfeasible using r.watershed. The inefficiency is likely due to r.watershed design that assumes that data are small enough to fit in the internal memory of the computer. If datasets are large, it does not fit in the main memory of the computer and reside on disk instead. Because disks are much slower than main memory, the bottleneck during analysis of large datasets is typically the movement of information between main memory and disk, and not the CPU computation time.

The r.terraflow module was recently implemented in GRASS GIS for computing flow direction and flow accumulation. It was designed and optimized for massive digital elevation data and was tested here as an efficient alternative to r.watershed. R.terraflow uses the standard ideas for computing flow direction and flow accumulation mentioned above but encodes them using scalable Input/Output (I/O)-efficient algorithms. As a result r.terraflow is scalableto very large datasets. For instance, it computes flow direction and flow accumulation for the entire Panama Canal Watershed dataset in approximately 3 hours.

The module r.terraflow can compute both SFD and MFD flow routing. It can also use a combination of the two by switching to SFD if contributing area exceeds a user-defined threshold. In flat areas, r.terraflow uses the iterative linking process proposed by Jenson and Domingue (1988) that was incorporated into Rivertools™. In depressions, r.terraflow routes flow by computing the lowest-height path from each cell outside the terrain. For each cell, the height of its lowest-path outside the terrain and corresponds to the steady-state level of water at that point. It can be formally shown (Arge et. al, in press) that computing lowest-height paths for all cells in the terrain can be done using a bottom-up plane sweep of the terrain which simulates the process of flooding the terrain (or rather of uniformly raising a water table).

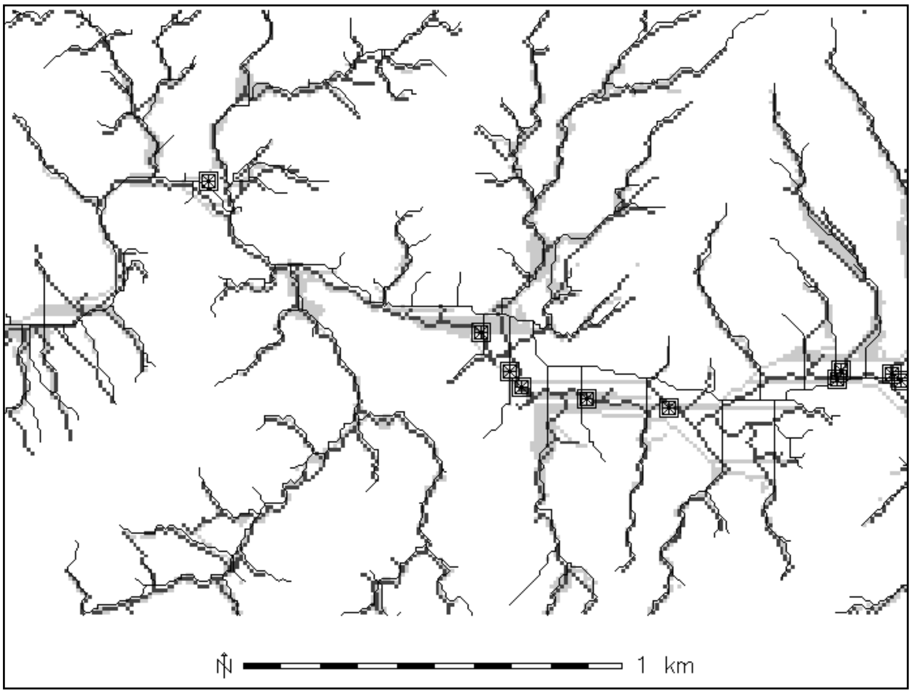


Figure 5: Comparison of stream channel output from r.terraflow (thick light gray), r.watershed (dark gray) and observed data (square symbols). The streams were extracted from 10m resolution DSM.

A detailed analysis of the accuracy of r.terraflow is under investigation. Figure 5 indicates that for some parts of the DSM, r.terraflow is remarkably similar to the other algorithms. Because it uses iterative linking, it has many of the same issues with flats, including the parallel flow paths indicated in Figure 5. Additionally, r.terraflow often diverges from the more accurate r.watershed-generated stream channels. The stream networks produced by extracting cells with flow accumulation larger than 100 cells look different for MFD and SFD flow routing. With either method the stream network contains big wet areas in the flat parts of the terrain. The output of r.terraflow are therefore more suitable for identifying areas susceptible to floods, than to narrowly delineating the stream network. Further work is necessary to incorporate into r.terraflow a procedure for refining the stream network.

3. CONCLUSIONS

For most of the studied basin, all tested algorithms produced a realistic stream network in locations that were identical or very close to the observed streams. The results were surprisingly accurate given the fact that the flow was traced on top of vegetation cover that was several meters high in most of the studied area, rather than on the bare ground. The most significant artifacts, in the form of long straight stretches of stream channels located tens of meters away from the observed stream locations, were produced by the tools that fill in the depressions caused by gaps in tree cover creating large artificial flats and then use iterative linking to overcome these flats. The algorithms that use the shortest path or imposed gradients generated more accurate results for these areas.

The tested software tools also differed in computational efficiency. While r.watershed provided the most accurate results it was slower than r.terraflow. R.terraflow computed the stream network for the Chagres Basin at 10m resolution almost ten times faster than r.watershed. Further, r.terraflow extracted the stream network from the DSM for the entire Panama Canal Watershed in 3 hours where the other r.watershed had not completed running after a day. The efficiency of algorithms in Rivertools™ have not been systematically compared to the GRASS GIS algorithms and this comparison remains the subject of future work.

The identification of stream origins and first order streams also remains an open research question, to some extent due to the fact that it is a dynamic phenomenon. Therefore the resulting 10m resolution stream network based on $10{,}000m^2$ upslope area threshold (or third order Strahler threshold) should be interpreted as a map of all potential streams, rather than actual streams. Field data are needed for sound selection of thresholds defining the stream origin instead of using the arbitrary threshold methods reported in this work.

REFERENCES

Arge, L., Chase, J., Halpin, P., Toma, L., Urban, D., Vitter, J. S., Wickremesinghe, R., in press. Flow computation on massive grid terrains, Geoinformatica.

Ehlschlaeger, C.R., 1989. Using the AT Search Algorithm to Develop Hydrologic Models from Digital Elevation Data. Proceeding of the International Geographic Information System (IGIS) Symposium, Baltimore, MD, 275-281.

Garbrecht, J. and Martz, L. W., 1997. The assignment of drainage directions over flat surfaces in raster digital elevation models: Journal of Hydrology, v. 193, no. 1-4, p. 204-213.

Jenson, S. K., and Domingue, J. O., 1988. Extracting topographic structure from digital elevation data for geographic-information system analysis: Photogrammetric Engineering and Remote Sensing, v. 54, no. 11, p. 1593-1600.

Kinner, D. A., 2003, Chapter 2-Delineation and Characterization of Boulder Creek Watershed and its Sub-basins, in, Murphy, S. F., Verplanck, P. L. and Barber, L. B., eds., *Comprehensive Water Quality of the Boulder Creek Watershed, Colorado, During High-Flow and Low-Flow Conditions, 2000*, U. S. Geological Survey Water-Resources Investigations Report, 03-4045.

O'Callahan, J., and Mark, D., 1984, The extraction of drainage networks from digital elevation data: Comput. Vision and Graph. And Image Processes, 28

Peckham, S.D., 1998, Efficient extraction of river networks and hydrologic measurements from digital elevation data, in Barndorff-Nielsen and others, eds., Stochastic Methods in Hydrology: Rain, Landforms and Floods: Singapore, World Scientiifc, p. 173-203

Rivix Limited Liability Company, 2001, RiverTools™ User's Guide release 2001: Boulder, CO, Research Systems, Inc., 202 p.

Strahler, A.N., 1952, Equilibrium theory of erosional slopes approached by frequency distribution analysis, American Journal of Science 248, 673-696 and 800-814.