Fast Stream Extraction from Large, Radar-Based Elevation Models with Variable Level of Detail

M. Metz[1], H. Mitasova[2], R.S. Harmon[3]

[1]Research Group Cognitive Ethology , German Primate Center
Kellnerweg 4, 37077 Goettingen, Germany
Telephone: +49-(0)921-2614299
Email:markus_metz@gmx.de

[1]Department of Marine, Earth and Atmospheric Sciences, North Carolina State University
Campus Box 8208, Raleigh, North Carolina 27695, USA
Telephone:001- 919-513-1327
Fax:001-919-515-7802
Email:hmitaso@unity.ncsu.edu

[3]Environmental Sciences Division, Army Research Office, U.S. Army Research Laboratory,
4300 South Miami Boulevard, Durham, North Carolina, 27703-9142, USA
Telephone:001-919-549-4326
Email:russell.harmon@us.army.mil

1. Introduction

Interferometric Synthetic Aperture Radar for Elevation (IFSARE) and Shuttle Radar Topographic Mission (SRTM) surveys provide a new generation of digital surface models (DSM) in regions that have had only limited elevation data coverage. These new topographic data sets are increasingly used to improve mapping of geomorphic and hydrologic features in remote, hard to reach areas and at regional to global scales (e.g. Kinner et al. 2005, Lehner and Döll 2004, World Wildlife Fund 2009). Extraction of hydrologic features from radar-based elevation data poses several challenges: (a) elevation surfaces include tree canopy that often requires depression filling of large areas (Figures 1, 2); (b) depending on the size of the study region and resolution, data sets can be massive and require extensive processing time. Significant effort has been devoted to development of new flow tracing and watershed analysis algorithms that support efficient processing of large DSMs and address the issue of depression filling (e.g., Arge et al. 2003; Danner et al. 2007).

We present a new implementation of method for flow routing, flow accumulation, and watershed analysis based on a least-cost path search algorithm (A* Search, Hart et al. 1968; module *r.watershed* in GRASS GIS, Ehlschlaeger 1989). This implementation dramatically improves computational efficiency while preserving its high accuracy routing capabilities through nested depressions, even for a challenging triple-canopy tropical rainforest environment with tree heights of more than 30m above the land surface. The new implementation that includes both single (SFD) and multiple flow direction (MFD) routing is compared with previously developed methods in terms of performance and accuracy. The impact of mapping technology (IFSARE, SRTM) and resolution on the extracted stream networks is also analyzed.

2. Methods

2.1 Fast Least Cost Flow Routing with SFD and MFD Support

Traditional flow routing algorithms that rely on depression filling are not suitable for data that include large numbers of nested pits due to natural topography, vegetation cover (Figure 2), or man-made structures like bridges. Alternative approaches that rely on digital elevation model (DEM) carving or combined depression-filling and carving were not designed for DSMs and may require significant modifications in elevation surface in regions with variable canopy height. Therefore, this work focuses on an algorithm that does not require depression filling or DEM carving and uses the *A** least-cost search method (Ehlschlaeger 1989).

The original implementation of the *A* Search* algorithm was optimized for large datasets by both increasing the processing speed and decreasing memory consumption. The core *A* Search* algorithm was not changed. Intermediate results of the *A* Search* process are now stored in a heap data structure (Atkinson et al. 1986) instead of a linked list. The speed gain in the time required to process a given grid is not fixed but increases with the number of grid cells to be processed and thus becomes more prominent for larger datasets.

Additionally, a multiple flow direction (MFD) algorithm has been implemented in *r.watershed* that makes use of the path determined by the *A* Search* and is based on Holmgren (1994), with an option to control the strength of flow convergence. Multiple flow direction provides more realistic results for flow accumulation in terrain with low slope and when using higher resolution (<= 10m) DEMs as input. Improvement in computational performance was evaluated by comparing the efficiency and accuracy with the old version of *r.watershed,* the GRASS module *r.terraflow* (Arge et al. 2003) designed for massive data sets, and flow routing modules from SAGA and TAS, with focus on treatment of depressions.

2.2 Data

The improved flow routing and watershed analysis algorithm was evaluated by performing stream extraction for the entire country of Panama using a combination of IFSARE and SRTM data.

Countrywide elevation coverage of Panama was available as 90m resolution SRTM DSM. We selected SRTM version 2 for our study as the most reliable in terms of accuracy and minimal artifacts, after evaluating properties of the currently available SRTM products (v1, v2, v3, v4.1). SRTM v2 tiles covering all of Panama were combined and gaps in the dataset filled using the regularized spline with tension (RST) interpolation method (GRASS module *r.fillnulls*; Mitasova and Mitas 1993). The seamless SRTM coverage was then reprojected from geographic to UTM zone 17N coordinate system with 90m resolution to keep resampling modifications to a minimum. For testing purposes, the DSM was then reinterpolated to 30m resolution using the RST method. A recent IFSARE survey has provided new, more detailed information about the topography in central Panama. The original IFSARE data were collected at 2.5m resolution and processed by standard procedures into 10m resolution DSM (Kinner et al. 2005). Stream extraction was performed on each of the DSMs separately at their original resolutions and then on a seamless 30m resolution DSM. Two levels of detail were created by merging the IFSARE and SRTM DSMs and re-interpolating to 30m resolution using RST, to ensure adequate routing for rivers flowing along the borders of the two DSMs (Figure 3).

Streams digitized from LANDSAT imagery (Earthsat dataset GLS2000, year 2000 with improved orthorectification, provided by the United States Geological Survey (USGS)) and field measurements were used for accuracy assessment. Georeferenced stream data were collected in the field during the years 2002-07 at sites in the Chagres river watershed (Figure 1) and during 2005-09 at lower reaches of most major rivers across Panama.

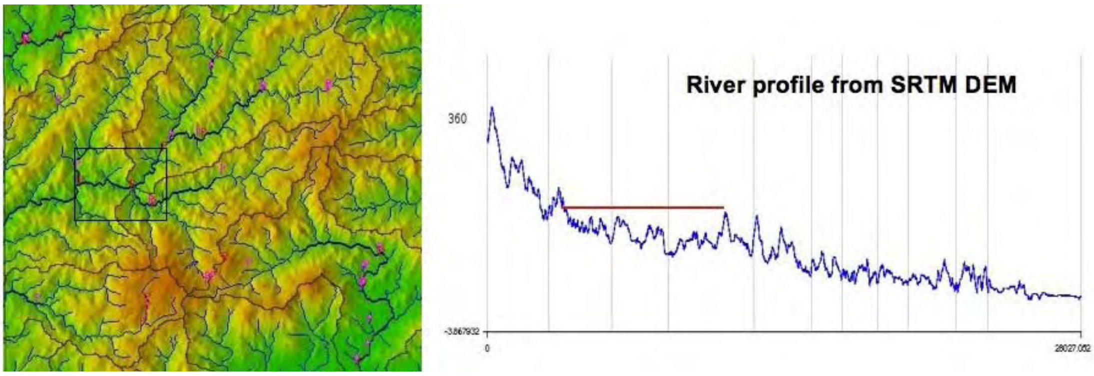

Figure 1. Left: Shaded SRTM relief with extracted streams (blue) and measured points (magenta). Right: Sample river profile, with local peaks and nested depressions.

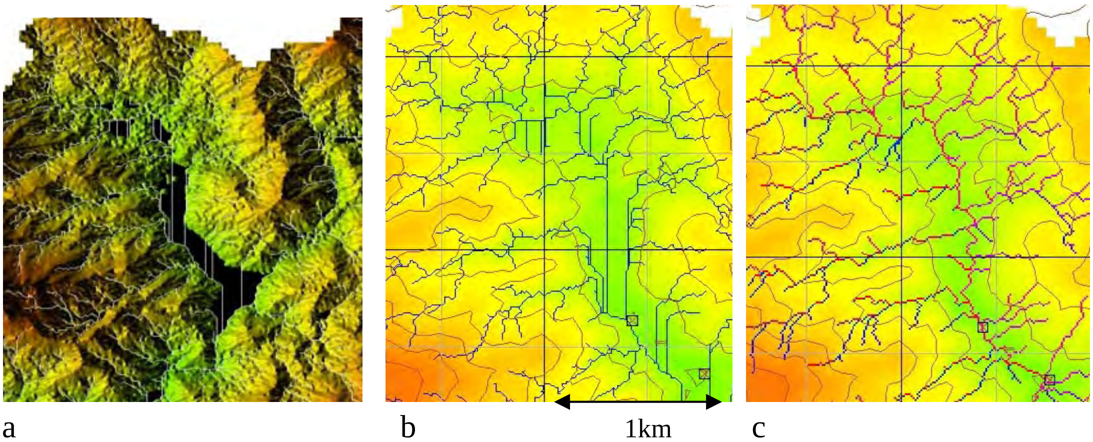

Figure 2. Impact of depression filling on stream extraction from IFSARE-based DSM in tropical forest environment: (a) extent of depression filling (black area), (b) artificial stream geometry extends over 1.5 km, (c) improved result - least cost path algorithm.

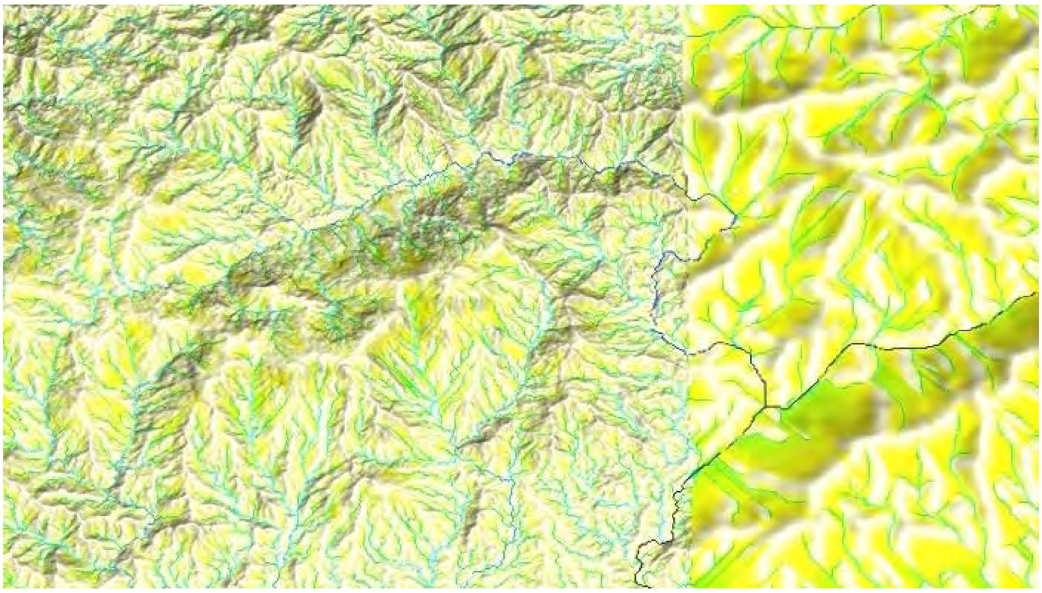

Figure 2. Seamless extraction of flow accumulation at 30m resolution with two levels of detail from merged IFSARE and SRTM DSMs.

2.3 Evaluation of Performance and Accuracy

Assessment of stream extraction accuracy was performed by computing distances between the extracted stream segments and (a) on-ground GPS measurements, (b) points digitized on rivers that could be clearly identified on LANDSAT imagery. The extracted streams were vectorized using the threshold of a minimum upstream catchment area of 100,000 square meters. General statistics of the measured distances was used to assess the accuracy of a particular method, dataset and spatial resolution. Deviations of derived streams from rivers visible on LANDSAT could be the result of either errors in the DSM or failure of the algorithm in difficult terrain, e.g. low topography or flow blocked by dams or bridges.

3. Results

Stream and watershed boundaries extraction was done individually at the original resolutions of 10m for IFSARE (156 million cells) and 90m for SRTM (27 million cells) and on the seamless 30m DSM (241 million cells) for the entire country of Panama.

To illustrate the improvement in computational performance of the new implementation of the *A* Search* method, a comparison was made between the processing time needed by the new and old version of *r.watershed* for flow accumulation computation at 90m resolution for all of Panama and for the central Panama subregion (IFSARE spatial extent). The new version was 350 times faster than the old version for central Panama represented by a relatively small DSM with 2 million grid cells at 90m resolution. The improvement was even more dramatic for the countrywide application with 27 million grid cells: the new version was 1940 times faster. The resulting flow accumulation raster maps from the old and the new version are identical for SFD; MFD is supported in the new version only.

For large regions represented by DSMs that fit into memory, the processing time of the new *r.watershed* in ram mode (all in memory) is shorter than that of *r.terraflow*. If data do not fit into memory, *r.watershed* uses segmented processing with intermediate data stored on disk, leading to longer processing time than for *r.terraflow* which uses I/O efficient algorithm specially designed for this case (Arge et al. 2003). The size of intermediate data created by *r.watershed* is about 16% of the size of intermediate data created by *r.terraflow*. Apparently, the segmented mode of r.*watershed* needs further optimization.

Accuracy assessment based on ground control points was done for 10m IFSARE, 30m IFSARE, 30m SRTM, and 90m SRTM, processed with *r.watershed* once in SFD and once in MFD mode, and processed with *r.terraflow* in MFD mode only. The 10m IFSARE DEM processed with the *A* Search* method and MFD provided the most accurate results and ground control points were closest to extracted streams.

Accuracy assessment based on rivers digitized from LANDSAT imagery revealed the known problems of the depression-filling method. Multiple flow direction produced more accurate results in areas with low topography. These areas are often also areas where depression-filling would be necessary, and where the combination of *A* Search* and MFD delivered the best results. Figure 4 shows LANDSAT GLS2000 imagery as backdrop and streams extracted once with *A* Search* and MFD and once with depression-filling and MFD. Additional comparisons will be performed using methods in SAGA and TAS.

4. Conclusion

We have presented a method for fast hydrological analysis specially designed for DSMs with large, nested depressions and evaluated it against other commonly used methods. We also provided insight into accuracy of stream extraction from widely available SRTM data using IFSARE data, ground control points and LANDSAT satellite imagery in a triple canopy tropical forest environment and coastal plain setting.

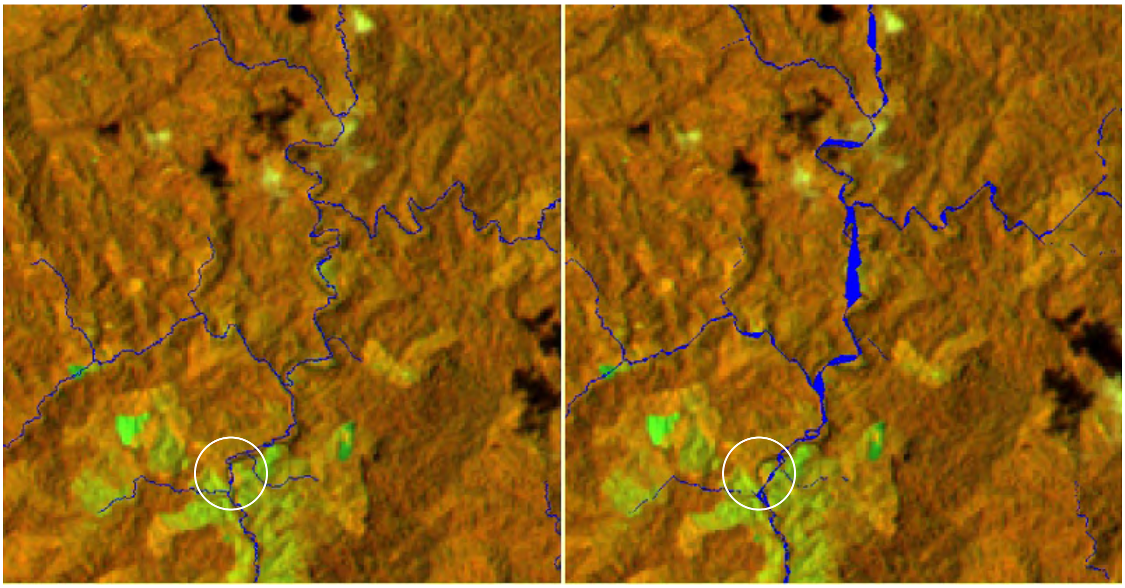

Figure 4. LANDSAT with bands 453 as RGB overlaid with A: River streams extracted with A * *Search* and MFD, and B: River streams extracted with depression-filling and MFD. Dense green vegetation appears dark orange, water bodies and cloud shadows appear dark, extracted streams are blue.

Acknowledgements

U.S. Army Research Office support for this research is gratefully acknowledged.

References

Arge L, Chase JS, Halpin PN, Toma L, Vitter JS, Urban D and Wickremesinghe R, 2003, Flow computation on massive grid terrains. *GeoInformatica* 7(4):283-313.

Atkinson MD, Sack J-R, Santoro N and Strothotte T, 1986, Min-max heaps and generalized priority queues. *Programming techniques and Data structures, Comm. ACM*, 29(10): 996-1000.

Danner A, Yi K, Moelhave T, Agarwal PK, Arge L and Mitasova H, 2007, TerraStream: From Elevation Data to Watershed Hierarchies, *Proc. ACM GIS 2007*

Ehlschlaeger C, 1989, Using the A* Search Algorithm to Develop Hydrologic Models from Digital Elevation Data. *Proceedings of International Geographic Information Systems (IGIS) Symposium*, Baltimore, MD, USA, 275-281.

Hart P E, Nilsson N J, Raphael B, 1968, A Formal Basis for the Heuristic Determination of Minimum Cost Paths, *IEEE Transactions on Systems Science and Cybernetics* 4 (2):100-107.

Hofierka J, Mitasova H, Neteler M, 2008, Geomorphometry in GRASS GIS. In: Hengl T and Reuter HI (eds), *Geomorphometry: Concepts, Software, Applications. Developments in Soil Science*, Elsevier, 33:387-410.

Holmgren P, 1994, Multiple flow direction algorithms for runoff modeling in grid based elevation models: An empirical evaluation. *Hydrological Processes*, 8(4):327-334.

Kinner D, Mitasova H, Harmon R, Toma L, Stallard R, 2005, GIS-based Stream Network Analysis for The Chagres River Basin, Republic of Panama. In: Harmon R (ed) *The Rio Chagres: A Multidisciplinary Profile of a Tropical Watershed*, Springer/Kluwer, 83-95.

Lehner B and Döll P, 2004, Development and validation of a global database of lakes, reservoirs and wetlands. *Journal of Hydrology* 296(1-4): 1-22.

Mitasova H and Mitas L, 1993, Interpolation by Regularized Spline with Tension: I. Theory and Implementation. *Mathematical Geology*, 25:641-655.

Neteler M and Mitasova H, 2008, *Open Source GIS: A GRASS GIS Approach*. Third Edition. Springer New York.

World Wildlife Fund, 2009, HydroSHEDS, http://hydrosheds.cr.usgs.gov/index.php