



CyberGIS Center Technical Report



CYBERGIS-TR-2016-005

07/27/2016

A CyberGIS Approach to Generating High-resolution Height Above Nearest Drainage (HAND) Raster for National Flood Mapping

Authors: Yan Y. Liu, David R. Maidment, David G. Tarboton, Xing Zheng, Ahmet Yildirim, Nazmus S. Sazib, and Shaowen Wang

Revision history

Revision	Date	Description
1	07/27/2016	Technical report finalized and presented at the Third International Conference on CyberGIS and Geospatial Data Science
0.5	06/12/2016	After edits/comments from Shaowen, Dave, and Xing
0.1	06/08/2016	First draft

Cite as: Liu, Y. Y., D. R. Maidment, D. G. Tarboton, X. Zheng, Ahmet Yildirim, N. S. Sazib and S. Wang, (2016), "A CyberGIS Approach to Generating High-resolution Height Above Nearest Drainage (HAND) Raster for National Flood Mapping," *CyberGIS 16, The Third International Conference on CyberGIS and Geospatial Data Science*, Urbana, Illinois, July 26-28. <http://dx.doi.org/10.13140/RG.2.2.24234.41925> .

A CyberGIS Approach to Generating High-resolution Height Above Nearest Drainage (HAND) Raster for National Flood Mapping

Yan Y. Liu¹, David R. Maidment², David G. Tarboton³, Xing Zheng², Ahmet Yildirim¹, Nazmus S. Sazib³, and Shaowen Wang¹

¹ University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, USA

² University of Texas at Austin, Austin, Texas 78712, USA

³ Utah State University, Logan, Utah 84322, USA

yanliu@illinois.edu, maidment@utexas.edu, dtarb@usu.edu, zhengxing@utexas.edu, ayild@illinois.edu, sazibap25@gmail.com, shaowen@illinois.edu

ABSTRACT

This paper describes the computational experience of generating the first 10m-resolution Height Above Nearest Drainage (HAND) raster data for the conterminous U.S. using the National Science Foundation (NSF) CyberGIS Facility based on the ROGER supercomputer. The HAND experiment is a joint effort with the NSF HydroShare project and the National Water Center for developing national-scale inundation mapping methodologies and data.

Keywords

CyberGIS, computational scalability, cyberinfrastructure, high-performance computing, hydrological analysis, inundation mapping, workflow

1. INTRODUCTION

The National Flood Interoperability Experiment (NFIE) is a project for innovating national-scale flood forecasting and mapping capability, conducted by the U.S. National Water Center in partnership with pertinent research communities. One of the goals of NFIE is to produce high-resolution forecasts of water surface elevation and maps of flood inundation for each of the approximately 2.67 million stream reaches in the context of the National Hydrography Dataset (NHDPlus, <http://www.horizon-systems.com/nhdplus/>). The proposed methodology for achieving this goal is to use weather and hydrologic models to generate runoff at the scale of each NHDPlus catchment, and then route this hydraulically through the stream network to determine discharge, depth and velocity in each reach. The digital elevation model (DEM) and NHDPlus flowline datasets then provide the information needed to generate the height of each grid cell above the nearest drainage (HAND; Nobre 2011), which serves as a basis for mapping reach-scale flood inundation and for determining corresponding hydraulic properties used in the hydraulics (Figure 1). This paper describes a novel cyberGIS (aka geographic information science and systems based on advanced cyberinfrastructure) (Wang 2010) approach to deriving national-scale HAND maps from the 10m

resolution US Geological Survey (USGS) 3DEP national elevation dataset (<http://nationalmap.gov/3DEP>) and NHDPlus streams for the continental US (CONUS).

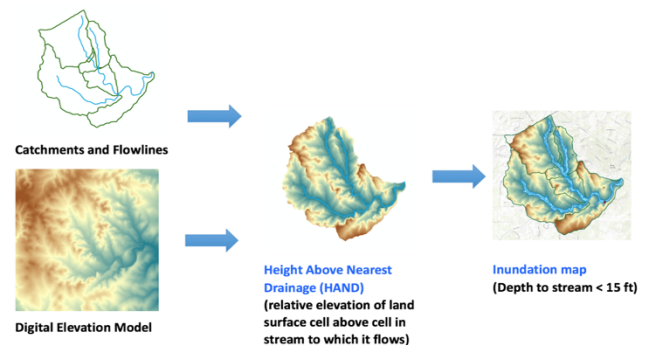


Figure 1. Conceptual methodology for flood inundation mapping in NFIE.

The HAND raster needs to be computed only once for CONUS. However, the computation of the HAND raster at the specified resolution (10m) for CONUS is computationally and data intensive (3DEP elevation dataset: 637212 x 280812 raster cells, ~700GB; NHDPlus: 2.67 million vectors, 18GB; NHD high-resolution (HR) dataset: ~30 million vectors, 60GB). For example, the execution of our first HAND computation would take 4.42 CPU years to finish and require 3.2TB memory at maximum to hold runtime data during analysis and 5TB disk space to hold results (see section “Results”). The limitations of existing desktop or server GIS solutions are apparent: existing desktop GIS and hydrological analysis software solutions are limited by the computing power of a single machine; server-side GIS solutions lack the capability to leverage high-performance computing power available from advanced cyberinfrastructure. Such a computational challenge calls for a scalable high-performance solution based on cyberGIS that exploits advanced cyberinfrastructure resources. The development of scalable hydrological analysis software is important to enabling efficient processing of large input datasets. Often,

DEM-based hydrological functions applied on GIS datasets are global raster functions that are computationally intensive and sophisticated to be parallelized.

The cyberGIS approach we have developed addresses the aforementioned challenges through a collaboration among NFIE, the NSF CyberGIS software project (Wang *et al.* 2013), NSF HydroShare (<http://hydroshare.org>), USGS, the NSF CyberGIS Facility based on ROGER supercomputer (Wang 2017), and the Extreme Science and Engineering Discovery Environment XSEDE (<http://xsede.org>). Our approach is built on an open source software solution that constructs a cyberGIS workflow coupling the scalable and high-performance TauDEM software (TauDEM 2016; Fan *et al.* 2014; Tesfa *et al.* 2011) for Digital Elevation Model (DEM)-based hydrological information analysis and open source geospatial software (i.e., GDAL; <http://gdal.org>) for pre- and post-processing of geospatial data. The ROGER supercomputer with a hybrid architecture was used as an integrated high-performance data handling, analysis, modeling, and visualization platform to compute a 10m-resolution HAND raster dataset for CONUS. The first CONUS HAND raster we computed was produced using 8 days of computation on ROGER. After further refining the workflow efficiency and model calibration the second CONUS HAND raster was produced using 1.5 days of computation on ROGER. The availability of this HAND raster has broad and significant research impact as it opens the door for GIScientists and hydrologists to improve the inundation mapping methodologies and achieve future NFIE goals. The NFIE experiments demonstrated the unique capability of ROGER as the first cyberGIS supercomputer for large-scale geospatial computation enabling novel scientific research.

2. SOFTWARE AND DATA INTEGRATION

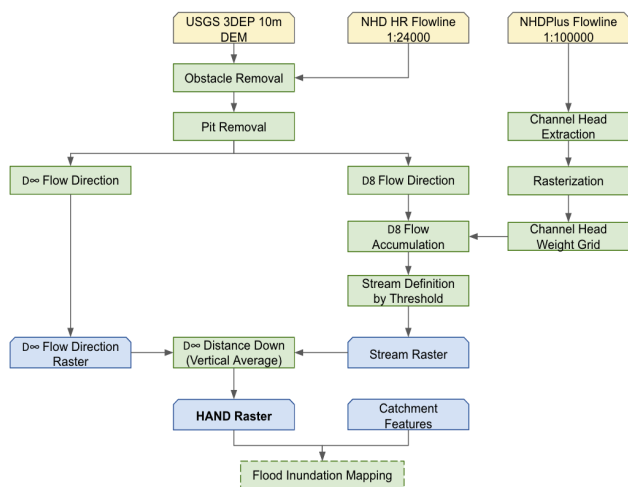


Figure 2. HAND computation workflow.

Figure 2 illustrates the HAND computation workflow. For a watershed of interest, the input DEM is generated by clipping the 10m 3DEP elevation dataset with a buffer distance. The boundary of the watershed is extracted from the USGS Water Boundary Dataset (WBD). The DEM of the watershed needs to go through a pit removal and soft burnin process for getting hydrologically conditioned and handling atypical features such as bridges and tunnels. Two flow-direction algorithms (d8 and d-infinity) are applied to the conditioned DEM to derive flow direction at each grid cell. The d8 flow-direction grid is then used to derive the flow accumulation area. In order to make the analysis consistent with the National Hydrography Dataset (including NHDPlus and NHD HR), a weight grid generated by rasterizing the inlet points at channel heads extracted from NHDPlus is fed to the flow accumulation calculation. Finally, a stream grid is derived and fed as input to the TauDEM DistanceDown function, along with the d-infinity flow-direction grid, to generate the HAND grid using the vertical distance measure. In this cyberGIS-HAND workflow, the step functions (rectangles in figure 2) were implemented using GDAL (i.e., DEM clipping, watershed shape buffering, flowline extraction, rasterization, and visualization) and the parallel computing version of TauDEM within the CyberGIS Toolkit (Tarboton *et al.* 2008; Wang *et al.* 2015) (i.e., pit removal, flow direction conditioning (Tarboton 1997), and the calculation of flow directions, flow accumulation, thresholding, and HAND).

To enable the software and data integration for methodology development and HAND computation, we used ROGER that has a hybrid architecture with a traditional high-performance computing (HPC) component (PBS job scheduler), a data-intensive computing component (Hadoop cluster), and a cloud component (OpenStack)

(<https://wiki.ncsa.illinois.edu/display/ROGER/>). All three components share 5PB usable storage via a GPFS parallel file system. We used the HPC part of ROGER for HAND computation and established Web and mapping servers on ROGER cloud for accessing HAND results and visualization. iRODS data management system was also deployed on ROGER to provide cross-domain data integration between ROGER and HydroShare in order to share HAND results in HydroShare. All of the input data required for HAND computation were deployed on GPFS. The 10m 3DEP elevation dataset for CONUS (~180 billion raster cells) was deployed in virtual raster (VRT) format and organized as a collection of 936 1 degree by 1 degree GeoTIFF raster tiles. The NHDPlus, NHD HR, and WBD vector datasets are in Esri's FileGDB format. TauDEM v5.3.4 with GDAL IO and GDAL v2.0.2 with PROJ4, GEOS, HDF5/NETCDF4, Esri FileGDB, SpatiaLite/SQLite, Java, and Python support are deployed on ROGER. With HPC component alone (31 computing nodes, 20 cores/node, 7.936TB memory in total), ROGER is able to provide sufficient computing power for the most

demanding TauDEM functions to be applied on the entire 10m elevation dataset (those functions require memory size

of four times of the input DEM, i.e., 3.2TB).

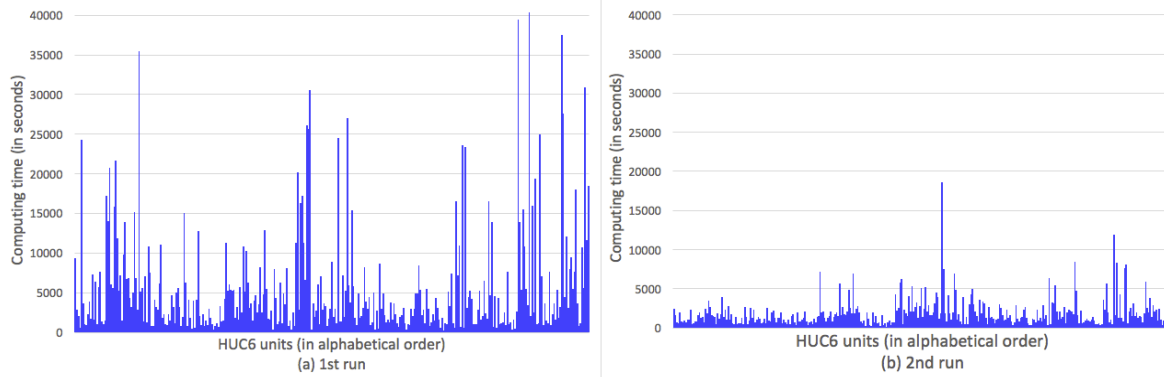


Figure 3. Computing time distribution: comparison between the first (a) and second (b) runs.

3. COMPUTATION AND RESULTS

Previous HAND computation was limited to small watersheds and could not reveal insights that can only be gained at national level and high spatial resolutions. Computation for larger watersheds would take too long to finish, if possible at all, which makes related analyses impractical for researchers. On ROGER, therefore, the computational performance of the workflow is critical for achieving reasonable responsiveness in terms of turnaround time for a single HAND computation. We have developed scalable computing solution to resolve this challenge by exploiting both high-throughput and high-performance computing models.

The hydrological unit code (HUC) system organizes CONUS watershed as a hierarchical hydrological unit network (<http://water.usgs.gov/GIS/huc.html>). There are four levels of HUCs (2, 4, 6, and 8). At each level, a HUC unit can be treated relatively independently with marginal effort needed to aggregate results from multiple units. The HUC system provides a natural spatial domain decomposition framework to divide and conquer at a HUC level for HAND computation. It provides an explicit spatial granularity from which we can match and allocate runtime computational resources to the computational requirements of each HUC unit, making the batch processing of all HUC units at a level possible. Each HUC unit is composed as a computing job on ROGER HPC. Since no communication is needed among these jobs, they are put into the shared job queue for high-throughput computing as soon as idle computing nodes become available. The computation of a single HUC unit is more challenging and requires high-performance hydrological analysis software, i.e., TauDEM. Details of enhancing computational performance of TauDEM were published in Fan *et al.* 2014, Survila *et al.* 2016, and Yildirim *et al.* 2016.

We conducted the first HAND computation for 331 out of 336 HUC6 accounting units of CONUS (the five

great lakes units are excluded) on ROGER. Each job used 60 to 180 processor cores based on a rough estimation of computational intensity for completing the computation within 2 days (unit area is a primary factor in the estimation) (Wang and Armstrong 2009). In total, the first run consumed 4.42 CPU years. On average, each unit used 65.6 cores and took 1.78 hours to compute. The first run took about 8 days to finish on the shared ROGER HPC queue. Figure 3(a) depicts the computing time of all 331 jobs. The high variation shows a heterogeneous computing profile for the 331 HUC6 units of different sizes, topographical and hydrological characteristics (e.g., the number of pits, flat regions and their sizes). Among the TauDEM functions called in the workflow, the two flow direction algorithms (d8 and d-infinity) took 72.65% computing time of the workflow, on average. The first run was also conducted as a stress test to derive more accurate computational intensity estimation for the units. After looking at the HAND grid from the first run, the workflow was calibrated for better configuration of edge contamination, DEM buffer size, and inlet identification from rivers passing through a watershed unit. The knowledge of computational intensity was gained to capture the requirements for processors and memory.

Following the first run, the second run was conducted in May 2016 using the calibrated workflow and newly accelerated d8 and d-infinity algorithms (Survila *et al.* 2016). Figure 3(b) depicts the computing time of all of the 331 units. The second run finished in 36 hours. It consumed 1.34 CPU years in total. On average, each job used 65.26 cores and took 0.54 hours to compute. The two flow direction algorithms only took 12.65% computing time of the workflow, on average. The majority 70.57% time was spent on GDAL commands for pre- and post-processing.

Figure 4 shows the HAND map for CONUS, generated from the second run. The pertinent scientific communities were provided with an unprecedented opportunity to examine such high-resolution surface properties at the national level.

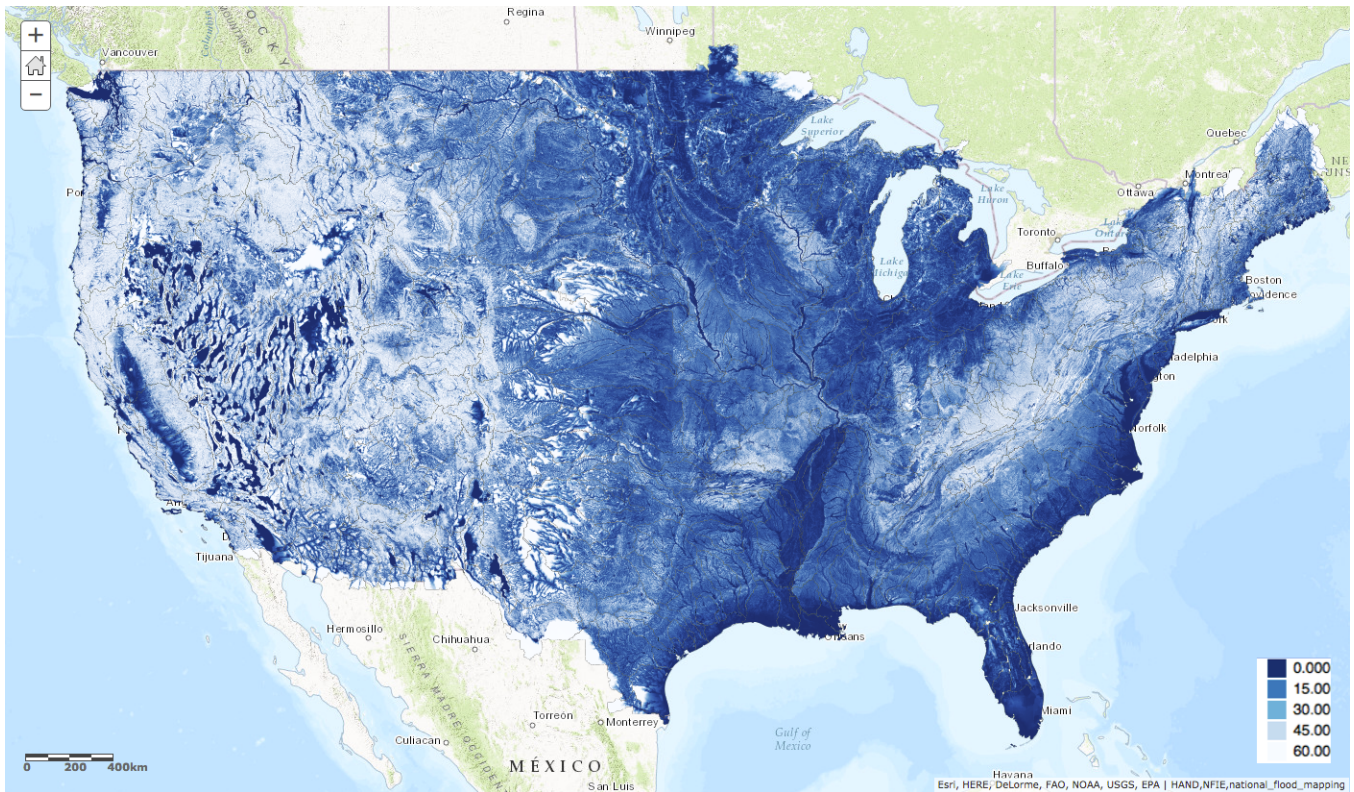


Figure 4. The 10m HAND map for CONUS resulted from the 2nd run on May 29, 2015. HAND unit: meters.

4. CONCLUSION

A novel cyberGIS workflow was developed to generate high-resolution and national-scale HAND raster for national flood mapping. This workflow is based on open source geospatial and hydrological software that is able to harness massive cyberGIS computing power for enabling the computation of the HAND raster. The computation on the first cyberGIS supercomputer – ROGER – exploits its hybrid architecture as an integrated platform for workflow methodology development and HAND computation, visualization, and validation. The first 10m HAND raster at national scale was successfully generated by using the high-throughput computing model for HUC6 units and improving computational scalability of TauDEM software. The HAND raster and the workflow provide unprecedented capabilities to conduct large-scale flood inundation mapping research by pertinent research communities. Our ongoing work continues to improve the cyberGIS workflow to couple related hydrological modeling processes for producing flood inundation forecasts at high spatial and temporal resolutions. The first 10m HAND raster described in this paper is a preliminary result of the cyberGIS-enabled HAND methodology. This result exhibited a rich set of observations as well as issues that were not recognized before. The researchers are working closely to improve related methodologies and conduct rigorous validation processes to quantify uncertainties of the approach from both GIScience and hydrology perspectives. We expect the further development of our approach and

future data products to engage broader communities for serving various cutting-edge research and education purposes.

5. ACKNOWLEDGMENTS

This research is supported in part by NSF under grant numbers: 1047916, 1343785, and 1443080. The work used the ROGER supercomputer, which is supported by NSF under grant number: 1429699. HydroShare is being developed under NSF grants ACI 1148453 and 1148090. This work is part of the ECSS project (award number ENG140009) of XSEDE that is supported by NSF under grant number 1053575. TauDEM was enhanced to support parallel computing and integrate with GDAL libraries with support from the US Army Corps of Engineers contract numbers W912HZ-11-P-0338 and W91238-15-P-0033 and XSEDE ECSS allocation EAR130008.

6. REFERENCES

- Fan, Y., Liu, Y., Wang, S., Tarboton, D., Yildirim, A., and Wilkins-Diehr, N. 2014. "Accelerating TauDEM as a Scalable Hydrological Terrain Analysis Service on XSEDE," *Proceedings of the 2014 Annual Conference on Extreme Science and Engineering Discovery Environment*, Atlanta, GA, USA, July 13-18, ACM, Extended Abstract, <http://dx.doi.org/10.1145/2616498.2616510>.
- Nobre, A. D., L. A. Cuartas, M. Hodnett, C. D. Rennó, G. Rodrigues, A. Silveira, M. Waterloo and S. Saleska. 2011.

"Height Above the Nearest Drainage – a hydrologically relevant new terrain model," *Journal of Hydrology*, 404(1–2): 13-29

Survila, K., Yildirim, A.A., Li, T., Liu, Y., Tarboton, and D., Wang, S. 2016. "A Scalable High-performance Topographic Flow Direction Algorithm for Hydrological Information Analysis". *Proceedings of the 2016 Annual Conference on Extreme Science and Engineering Discovery Environment (XSEDE'16)*. July 17-21. Miami, Florida. accepted.

TauDEM. <http://hydrology.usu.edu/taudem>. Accessed on June 12, 2016.

Tarboton, D.G. 1997. "A new method for the determination of flow directions and upslope areas in grid digital elevation models." *Water resources research* 33-2 (1997): 309-319.

Tarboton, D.G., and Baker, M.E. 2008. "Towards an algebra for terrain-based flow analysis." *Representing, modeling and visualizing the natural environment: innovations in GIS* 13 (2008): 167-194.

Tesfa, T. K., Tarboton, D. G., Watson, D. W., Schreuders, K. A. T., Bake, M. E., and Wallace R. M. 2011. "Extraction of hydrological proximity measures from DEMs using parallel processing," *Environmental Modelling & Software*, 26(12): 1696-1709, <http://dx.doi.org/10.1016/j.envsoft.2011.07.018>.

Wang, S., and Armstrong, M. P. 2009. "A Theoretical Approach to the Use of Cyberinfrastructure in Geographical Analysis." *International Journal of Geographical Information Science*, 23 (2): 169-193

Wang, S. 2010. "A CyberGIS Framework for the Synthesis of Cyberinfrastructure, GIS, and Spatial Analysis." *Annals of the Association of American Geographers*, 100(3): 535-557

Wang, S., Anselin, L., Bhaduri, B., Crosby, C., Goodchild, M. F., Liu, Y., and Nyerges, T. L. 2013. "CyberGIS Software: A Synthetic Review and Integration Roadmap." *International Journal of Geographical Information Science*, 27(11): 2122-2145

Wang, S., Liu, Y., and Padmanabhan, A. 2015. "Open CyberGIS Software for Geospatial Research and Education in the Big Data Era". *SoftwareX*, DOI:10.1016/j.softx.2015.10.003

Wang, S. 2017. "CyberGIS". *The International Encyclopedia of Geography*. Edited by Douglas Richardson, Noel Castree, Michael F. Goodchild, Audrey Kobayashi, Weidong Liu, and Richard A. Marston. John Wiley & Sons, Ltd. DOI: 10.1002/9781118786352.wbieg0931

Yildirim, A.A., Tarboton, D., Liu, Y., Sazib, N.S., Wang, S. 2016. "Accelerating TauDEM for Extracting Hydrology Information from National-Scale High Resolution Topographic Dataset". *Proceedings of the 2016 Annual Conference on Extreme Science and Engineering Discovery Environment (XSEDE'16)*. July 17-21. Miami, Florida. Accepted.