# A CYBERGIS INTEGRATION AND COMPUTATION FRAMEWORK FOR HIGH-RESOLUTION CONTINENTAL-SCALE FLOOD INUNDATION MAPPING

*Yan Y. Liu, David R. Maidment, David G. Tarboton, Xing Zheng, and Shaowen Wang*

Senior Research Programmer (Liu), National Center for Supercomputing Applications and Department of Geography and Geographic Information Science, University of Illinois at Urbana-Champaign, 1205 W. Clark St., Urbana, Illinois 61801; Professor (Maidment), Center for Water and Environment, University of Texas, Austin, Texas 78712; Professor (Tarboton), Department of Civil and Environmental Engineering, Utah State University, Logan, UT 84322-4110; Graduate Student (Zheng), Department of Civil, Architectural and Environmental Engineering, University of Texas, Austin, Texas 78712; and Professor (Wang), Department of Geography and Geographic Information Science, University of Illinois at Urbana-Champaign, Champaign, Illinois 61820 (E-Mail/Liu: yanliu@illinois.edu)

ABSTRACT: We present a Digital Elevation Model (DEM)-based hydrologic analysis methodology for continental flood inundation mapping (CFIM), implemented as a cyberGIS scientific workflow in which a 1/3rd arc-second (10m) Height Above Nearest Drainage (HAND) raster data for the conterminous U.S. (CONUS) was computed and employed for subsequent inundation mapping. A cyberGIS framework was developed to enable spatiotemporal integration and scalable computing of the entire inundation mapping process on a hybrid supercomputing architecture. The first 1/3rd arc-second CONUS HAND raster dataset was computed in 1.5 days on the CyberGIS ROGER supercomputer. The inundation mapping process developed in our exploratory study couples HAND with National Water Model (NWM) forecast data to enable near real-time inundation forecasts for CONUS. The computational performance of HAND and the inundation mapping process was profiled to gain insights into the computational characteristics in high-performance parallel computing scenarios. The establishment of the CFIM computational framework has broad and significant research implications that may lead to further development and improvement of flood inundation mapping methodologies.

(KEY TERMS: computational methods, cyberGIS, data management, geospatial analysis, height above nearest drainage (HAND), inundation mapping, streamflow)

This is the peer reviewed version of the following article:

Liu, Y. Y., D. R. Maidment, D. G. Tarboton, X. Zheng and S. Wang, (2018), "A CyberGIS Integration and Computation Framework for High-Resolution Continental-Scale Flood Inundation Mapping," JAWRA Journal of the American Water Resources Association, 54(4): 770-784, https://doi.org/10.1111/1752-1688.12660.

# INTRODUCTION

In August 2016, the National Weather Service brought into operation the National Water Model (NWM) (U.S. National Oceanic and Atmospheric Administration (NOAA), National Water Model. Accessed March 10, 2017, http://water.noaa.gov/about/nwm), which forecasts the streamflow on approximately 2.7 million stream reaches covering about 5.2 million kilometers of rivers and streams of the conterminous United States (CONUS). The academic community is collaborating with the National Weather Service to enhance NWM in a project called the National Flood Interoperability Experiment (NFIE) (Maidment, 2016a). A key component of that project is to extend the forecasting of flood streamflow into forecasting of water depth and inundation extent at the continental scale.

Central to this methodology is a technique called Height Above Nearest Drainage (HAND) (Rodda, 2005; Rennó *et al.*, 2008; Nobre *et al.*, 2011; Tesfa *et al.*, 2011; Nobre *et al.*, 2016), which uses a Digital Elevation Model (DEM) to define the height of each cell in the land surface above the cell in the nearest stream to which the drainage from that land surface cell flows. The HAND method is applied to the stream reaches used in NWM, which themselves are derived from the medium resolution NHDPlus dataset (US Geological Survey (USGS) and Environmental Protection Agency (EPA), NHDPlus. Accessed March 10, 2017, http://www.horizon-systems.com/nhdplus). By combining NHDPlus with the USGS 3D Elevation Program (3DEP) dataset (USGS, 3DEP. Accessed March 10, 2017, https://nationalmap.gov/3DEP) at 1/3rd arc-second (about 10-meter) cell resolution, this paper shows how the HAND raster can be determined for the continental United States, and also how hydraulic geometry relationships and synthetic rating curves can be determined for each stream reach so that the forecast streamflow can be converted to forecast depth for each stream reach, and then to flood inundation extent using the HAND approach. The combination of all these techniques is here referred to as Continental Flood Inundation Mapping (CFIM).

The CFIM framework has two key components: a flood inundation model appropriate for continental-scale study and a scalable computational model that provides a platform where the methods, software, data, and results could be deployed and shared in a responsive way that fosters iterative collaboration for methodological development and validation. As well-known models such as those based on cross sections show poor scalability to continental scale (Zheng, 2015), our companion paper (Zheng *et al.*, 2017) proposes a continental flood inundation mapping methodology which estimates channel geometry properties and rating curves from high-resolution terrain data. In the companion paper, scientific challenges of CFIM are discussed. A comprehensive evaluation on model integrity and uncertainty related to space and time, stream scales, stream order, river types is also presented.

This paper addresses the computational model component. Given the significant computational challenges for conducting the inundation mapping process at CONUS scale for these massive geospatial datasets (i.e., 3DEP DEM and NHDPlus), we develop a computational model based on cyberGIS (also known as geographic information science and systems based on advanced cyberinfrastructure) (Wang, 2010) to provide a scalable integration and computation framework that is able to create HAND maps using cyberinfrastructure resources.

Computational challenges in regional- and continental-scale high resolution flood mapping have been discussed in David et al. (2013); Hodges (2013); and Tavakoly et al. (2017), and Snow *et al.* (2016). Snow *et al.* (2016) developed a computational forecast framework and a web-based visualization application to tackle similar NFIE questions. High-density ensemble national-scale stream forecasts were produced by downscaling runoff forecasts generated by ECMWF and routing the runoff using the RAPID model (David *et al.*, 2011). Streamflow forecasts are displayed using the Tethys Platform (Swain *et al.*, 2015; Swain *et al.*, 2016). In addition to obvious differences in inundation mapping methodology and focus, our computational model differs in the following aspects. First, our model is designed to achieve 1/3rd arc-second or finer inundation mapping for CONUS and covers all 2.7 million reaches in NHDPlus from the beginning. Second, high-throughput computing is the only parallel computing model used in Snow *et al.* (2016), in which HTCondor (Bockelman *et al.*, 2015) is used to employ multiple processors to compute the downscaling of ECMWF runoff forecast on multiple watersheds. We provide a more comprehensive parallelization to achieve both high throughput (via job scheduler and batch computing) and high performance (via TauDEM). Last, the cyberGIS approach for integrating massive data and computing resources and building online problem-solving environment provides an efficient and scalable hybrid supercomputing environment for tackling computational challenges in CFIM. A similar approach could also be used to enhance the ECMWF-RAPID computational forecast framework.

Our cyberGIS framework addresses CFIM computational challenges through collaboration among NFIE, the National Science Foundation (NSF) CyberGIS software project (Wang *et al.*, 2013), NSF HydroShare (Tarboton *et al.*, 2014; Horsburgh *et al.*, 2016), USGS, the NSF CyberGIS Facility (that houses the Resourcing Open Geospatial Education and Research (ROGER) supercomputer) (Wang, 2017), and the Extreme Science and Engineering Discovery Environment XSEDE (Towns *et al.*, 2014). Our open source software solution constructs a cyberGIS workflow that couples the scalable and high-performance TauDEM software (Tesfa *et al.*, 2011) for DEM-based hydrologic analysis and a collection of open source geospatial software for pre- and post-processing of geospatial data. ROGER, which has a hybrid supercomputing architecture, provides an integrated high-performance data handling, analysis, modeling, and visualization platform for CFIM by coupling high-throughput computing (HTC), high-performance computing (HPC) and cloud computing. HAND for CONUS was computed in 1.5 days on ROGER.

The following sections describe computational challenges in the CFIM methodology. We demonstrate a cyberGIS integration model and a two-level parallel computing model as a holistic computational model for tackling these challenges. The effectiveness and performance of the computational model are illustrated through the computational experience and results obtained from the HAND and inundation mapping generation workflow. We conclude with discussion of the advantages and limitations of our approach, and ideas for future work.

## DATA AND COMPUTATIONAL CHALLENGES

The development of CFIM methodology is a multidisciplinary collaboration of science communities in hydrology, hydraulics, geographic information science (GIScience), and meteorology. Computation cuts across all these disciplinaries and plays a central role that not only provides significant computing power on national cyberinfrastructure resources for the computation of NFIE experiments, but also develops an integrated solution that addresses the data, software, computation, visualization, and community collaboration challenges. The main research question raised in our work is: *Is it feasible to compute inundation maps for CONUS at 1/3rd arc-second or higher resolution and automate the computation on USGS 3DEP DEM and NHDPlus?* From a computational perspective, the following issues need to be resolved.

- Terabytes and gigabytes of high-resolution national-scale terrain, water, and weather data that are distributed by multiple data sources and vary greatly in spatiotemporal scales and resolutions need to be integrated, processed, and analyzed.

- In developing the inundation mapping workflow, computational bottlenecks at the processing, modeling, and analysis steps of the proposed inundation mapping methodology need to be identified and resolved for scalable CONUS-level computation on advanced cyberinfrastructure.
- The entire inundation mapping process needs to be automated such that the resulting software can be used to produce near real-time inundation maps from continuous NWM forecasting.
- High-performance and scalable computation is important to produce the output within a reasonable turnaround time of, say, a few hours, to match the working pace of the iterative research collaborations and the pace of NWS forecast data publishing. Taking weeks would seriously hinder the team progress. For achieving near real-time flood forecast, taking more than one day to compute the inundation map would be impractical. The current NWM, which runs in production on NOAA's Luna and Surge supercomputers (Top500 supercomputer ranking. Accessed March 10, 2017, https://www.top500.org/list/2016/11/?page=1), has a turnaround time of about 2 hours. The inundation mapping computation should not introduce additional significant delays.

Given the responsiveness requirements for research collaboration and national inundation mapping computation, we pursue a cyberinfrastructure-based computational model to address these challenges with two key foci. First, we configured an integrated computational platform on a hybrid supercomputing architecture that allows for the automation and integration of the inundation mapping workflow as an open software solution and provides a solution for collaboration, data sharing, visualization, and high-performance computing. Second, we evaluated the scalability of the computational solution to both the data size and the number of computer processors so that the turnaround time for computation can be reduced by simply adding more computing power. This model achieves a modeling environment where compute, data storage, and network resources are integrated on demand on a centralized platform for building the required online geospatial and hydrologic services. This online problem-solving environment, in turn, serves as a community platform for broader engagement and outreach of continental inundation mapping research.

The continental scale inundation mapping methodology described in the companion paper (Zheng *et al.*, 2017) takes as input DEM (e.g., 1/3 arc second USGS 3DEP elevation DEM) and hydrography data that comprises geospatial vector data of flow lines, catchments and water bodies (i.e., the NHDPlus dataset). Our computational approach uses the generalized hydrologic terrain analysis concepts from TauDEM (Tarboton, 1997; Tarboton *et al.*, 2008; 2009; Tesfa *et al.,* 2011). DEM derived streams are initiated at the sources of NHDPlus streams to produce a stream raster consistent with the DEM. A general method for calculating distance to stream in the vertical direction was used to produce HAND from this stream raster. Three output datasets for inundation mapping are produced:

1. HAND raster of the same resolution as the input DEM. The HAND value of each raster cell represents the height of each raster cell above the nearest stream along the flow path from that cell to the stream. The HAND raster represents a type of hydrologic terrain. It is a reference dataset for inundation mapping, that is produced once and only needs updating when input DEM or NHDPlus data source is updated. The size of this raster is the same as the input DEM.

2. Hydraulic property table with one record for each catchment defined in NHDPlus. This table is derived from the HAND raster and a defined input list of stage height values to compute. Each record in the table represents the hydraulic properties for a river reach in NHDPlus for each designated stage height. Attributes calculated from HAND for each reach and stage include surface area, bed area, volume, top width, wetted perimeter, cross sectional area, hydraulic radius, and uniform flow streamflow. This table serves as a lookup table to interpolate the real-time water depth given a specific river streamflow forecast from NOAA NWM under the assumption of uniform flow. This table is updated whenever HAND or NHDPlus is updated or an improved rating curve method or different roughness (Manning’s ***n***) is to be applied.
3. Inundation forecast tables and maps based on streamflow forecast data from NOAA NWM for each NHDPlus reach. Streamflow information is converted to water depth by using the hydraulic property table, which is then compared with HAND to determine inundation information at each river reach for all the catchment cells associated with the reach.

The foremost challenge of national scale inundation mapping is rooted in the data, including terrain data and open water data in space and time (Maidment, 2016b). Table 1 lists the properties of the national-scale geospatial input datasets used for producing the aforementioned outputs. The geospatial data involved in this work represent typical scientific big data in volume, variety, and velocity. Desktop-based GIS software is ill-suited for processing such big data collection.

TABLE 1. Properties of the Input Data Sources.

| *Data Source* | *Resolution & Coverage* | *Size* | *Update Frequency* |
|---|---|---|---|
| USGS 3DEP Elevation Dataset | 1/3rd arc-second (10m); Entire U.S. | 635939 x 282122 (180b cells);<br>718GB uncompressed | Partial update every 3 months |
|  | 1/9th arc-second (3m) | ~10 times larger than 1/3rd arc-second | N/A |
|  | 1/27th arc-second (1m) | ~100 times larger than 1/3th arc-second | N/A |
| NHDPlus from EPA and USGS | 1:100,000;<br>Entire U.S. | ~2.7 million reaches;<br>12 attribute layers;18GB | Version 2.1 |
|  | 1:24,000<br>Partial U.S. | ~30 million reaches;<br>~77 attribute layers in pre-release versions | Not released yet |
| NOAA NWM streamflow forecast | 1:100,000;<br>Entire U.S. | 2.7 million reaches; ~4MB each forecast<br>short range: 18 hourly forecasts<br>medium range: 80 forecasts (3-hr; 10 days) | Hourly; daily |

## COMPUTATIONAL MODEL

A high-performance and scalable cyberGIS integration and computation framework was built on the ROGER supercomputer to provide a holistic computational model for CFIM collaboration and computation. A two-level parallelization approach was developed as a scalable computing strategy to efficiently compute HAND and inundation information for CONUS.

### *CyberGIS Integration Model*

We exploited massive computing power enabled by advanced cyberinfrastructure (e.g., XSEDE) to address the integration and computational challenges presented in the NFIE CFIM project. Allocable resources on cyberinfrastructure include not only hardware (compute, memory, storage, and network) resources, but also software environment, parallel computing libraries, and higher-level services such as performance profiling and acceleration, community application development, collaborative science gateways (Lawrence *et al.*, 2015) through user support programs such as XSEDE ECSS (Wilkins-Diehr *et al.*, 2015). The cyberinfrastructure approach integrates a powerful computational platform via supercomputers that enables the development of HPC solutions for domain applications. We leverage a promising advance in supercomputing, hybrid supercomputing architecture (Qiu *et al.*, 2010), to provide a highly integrated computational platform from application to system level for the support of large research computing projects such as the NFIE CFIM. A hybrid supercomputing architecture typically couples HPC, cloud, and data-intensive computing resources together on a single physical and/or virtualized supercomputer.

In CFIM, most of the data processing and analysis employs and outputs geospatial data, which presents a typical case of the aforementioned integration and computation challenges that arise from cyberGIS problems. CyberGIS is cyberinfrastructure-enabled high-performance, integrated, and collaborative GIS (Wang, 2010). As a major approach in our framework, we employ ROGER, a dedicated cyberGIS supercomputer, as the hybrid supercomputer for CFIM development and operation. ROGER has three components: HPC (32 computing nodes with 12 equipped with GPU), cloud (13 nodes), and data-intensive computing (a Hadoop cluster of 11 nodes). The three components share 5 petabyte (PB) of usable storage via the GPFS parallel file system. This architecture makes it possible to eliminate the cumbersome intermediate steps in data transfer and software management in the HPC-only supercomputer usage model and presents an online research computing environment with direct access to and processing of input and output data by end user.

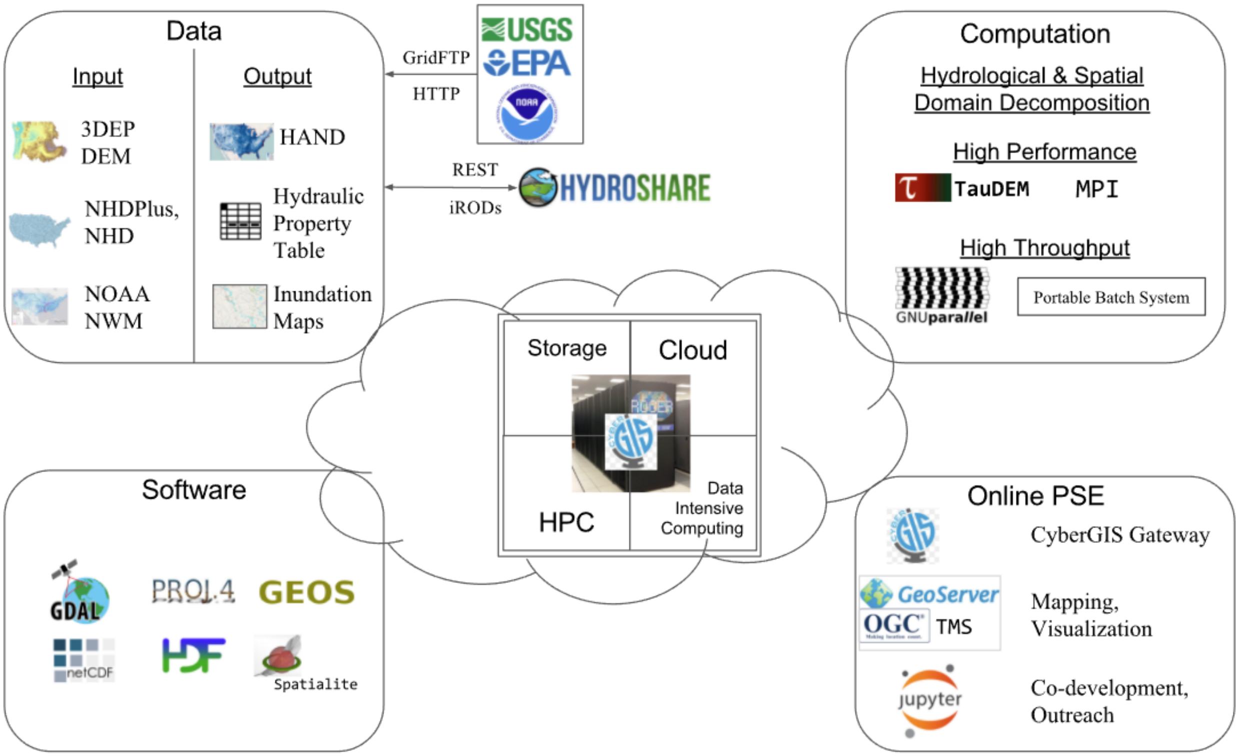


FIGURE 1. CyberGIS Integration Model. DEM, Digital Elevation Model; HAND, height above nearest drainage; HPC, high-performance computing; MPI, Message Passing Interface; PSE, problem-solving environment; TMS, Tile Map Service.

Figure 1 shows the cyberGIS integration model on ROGER, which integrates data, software, computation, and an online problem-solving environment (PSE) on ROGER's cloud, HPC, and storage systems (currently, the data-intensive computing cluster is not used) in order to automate or streamline the methodological components of CFIM.

The three major national data sources for CFIM are deployed and updated on the 5 petabyte GPFS storage system on ROGER. DEM data is managed by a USGS 3DEP elevation data service - TopoLens (Hu *et al.*, 2016) for the customization of DEM data delivery. ROGER HPC alone (8 terabyte memory in total), is able to provide sufficient compute, memory, storage, and network resources for the most demanding geospatial and hydrologic functions to be applied on the entire 3DEP elevation dataset. For example, the maximum memory requirement of TauDEM functions is four times of the input DEM, i.e., 3.2 terabyte for processing the entire 1/3rd arc-second 3DEP DEM.

All of the hydrologic and geospatial software needed to handle these input datasets are open source or free libraries. They are customized and built into the ROGER geocomputation software environment. These software handle raster, vector, and raster-vector processing in CFIM computation. Raster processing functions include clipping and hydrologic information extraction. Vector functions include inlet/outlet identification, flowline analysis, multi-layer operations such as river reach-catchment join and georeferencing with non-spatial NWM data. Raster-vector processing includes vectorization, rasterization, and dynamic inundation mapping by coupling HAND raster, catchment raster, and NWM vector data.

An online problem-solving environment (PSE) is built on the ROGER OpenStack cloud to share and visualize CFIM data. A set of services is built as on-demand virtual machine images (VM) and container (e.g., Docker (Docker containerization. Accessed March 10, 2017, https://www.docker.com)) instances. As shown in Figure 1, data sharing is enabled through web-based downloading and the underlying iRODS data federation between ROGER and HydroShare's iRODS data storage at RENCI (RENCI, HydroShare project. Accessed March 10, 2017, http://renci.org/research/hydroshare/). The iRODS data management system is used to provide cross-domain data integration between ROGER and HydroShare in order to share HAND results through HydroShare data repositories. Online visualization of continental-scale raster results at 1/3rd arc-second resolution or finer is beyond the capabilities of a single mapping server. Our strategy to build the visualization tiles for multiple zoom levels, i.e., raster pyramids, is two-fold. First, the computation needed to generate the visualization data occurs on the same computing nodes that produce CFIM analysis results. Second, the need for a powerful mapping server to host massive visualization data is eliminated by publishing a raster layer as an OGC standard Tile Map Service (TMS), which only requires a web server to hold the data and supports the tile rendering using the straightforward [*tile.x, tile.y, zoom_level*] URL mapping to tile image file path. Visualization data is then rendered within a browser using a CyberGIS WebGIS module. Common geospatial datasets, such as the NHD water boundary dataset (WBD), are hosted on GeoServer, an open source mapping server based on Java.

### *Scalable Computing*

Our parallel computing framework leverages two levels of parallelism to accelerate CFIM functions and the entire workflow. These two levels of parallelisms are exploited by two types of parallel computing models: high-throughput and high-performance computing, respectively.

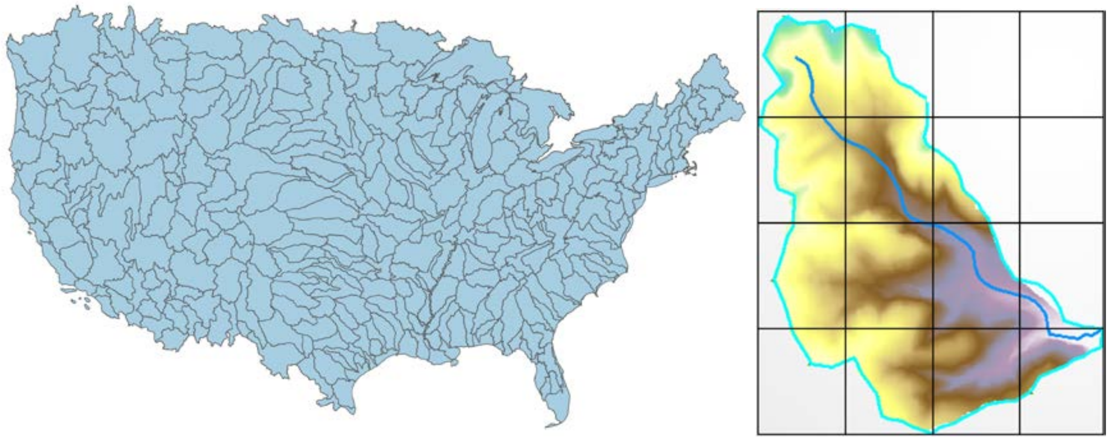


FIGURE 2. Two-level parallelization based on hydrologic and spatial domain decomposition. Left: HUC6 map of CONUS. Right: decomposition of a single HUC6 unit. HUC, Hydrologic Unit Code; CONUS, conterminous U.S.

The first level of parallelism, leveraging the Hydrologic Unit Code (HUC) system (USGS, HUC system. Accessed March 10, 2017, http://water.usgs.gov/GIS/huc.html), implements a divide-and-conquer strategy to provide a hydrologic and spatial data decomposition (Wang *et al.*, 2009) mechanism for the parallel computing of all the HUC units for CONUS, as shown on the left in Figure 2. The HUC system provides a natural spatial domain decomposition framework to divide CONUS terrain. It provides an explicit spatial granularity from which we can match and allocate the runtime computing power to the computational requirements of each HUC unit, making the batch processing of all HUC units in CONUS possible. We chose HUC6 as the basic decomposition HUC level within which we consider all the flowlines and catchments in NHDPlus. There are 336 HUC6 units on CONUS, but the five Great Lakes units are not considered. The 331 units in consideration are sent to a batch job scheduler as independent computing tasks.

The second level of parallelism is for the parallel computing of each individual HUC6 unit, shown on the right in Figure 2. Regular spatial domain decomposition (e.g., row-, column-, or block-wise decomposition) is applied to distribute input and output data domains to a set of processors. These domains form a network topology that represents their adjacency relationship. Ghost zones, which store the boundary data belonging to neighboring processors, are established for runtime data exchange via the broadly used Message Passing Interface (MPI) among participating processors. TauDEM, the well-known high-performance hydrologic information analysis software built on MPI, is employed to process individual HUC6 units.

Our two-level parallelization strategy provides a comprehensive scalable computing framework that is adaptive to data coverage, resolution (e.g., finer resolution DEMs derived from LiDAR), and the number of allocated processors. HUC6 is chosen because TauDEM exhibits the best computing efficiency at this level when experiments were conducted. As TauDEM's performance is accelerated, we can apply it to higher level HUC, which results in fewer computing jobs, but each job requires higher performance obtained by employing more processors for each job. Given sufficient computing power and TauDEM numerical performance, it is possible to compute the entire CONUS as a single computing job at 1/3rd arc-second resolution. On the other hand, when 3-meter, 1-meter and sub-1-meter DEM become available for CONUS, this framework can be applied with an appropriate HUC level decomposition that is determined by TauDEM's capability to handle a single DEM.

## HEIGHT ABOVE NEAREST DRAINAGE (HAND) COMPUTATION

In this section, we describe the data and information flow in HAND computation and discuss insights gleaned from generating the HAND at CONUS scale.

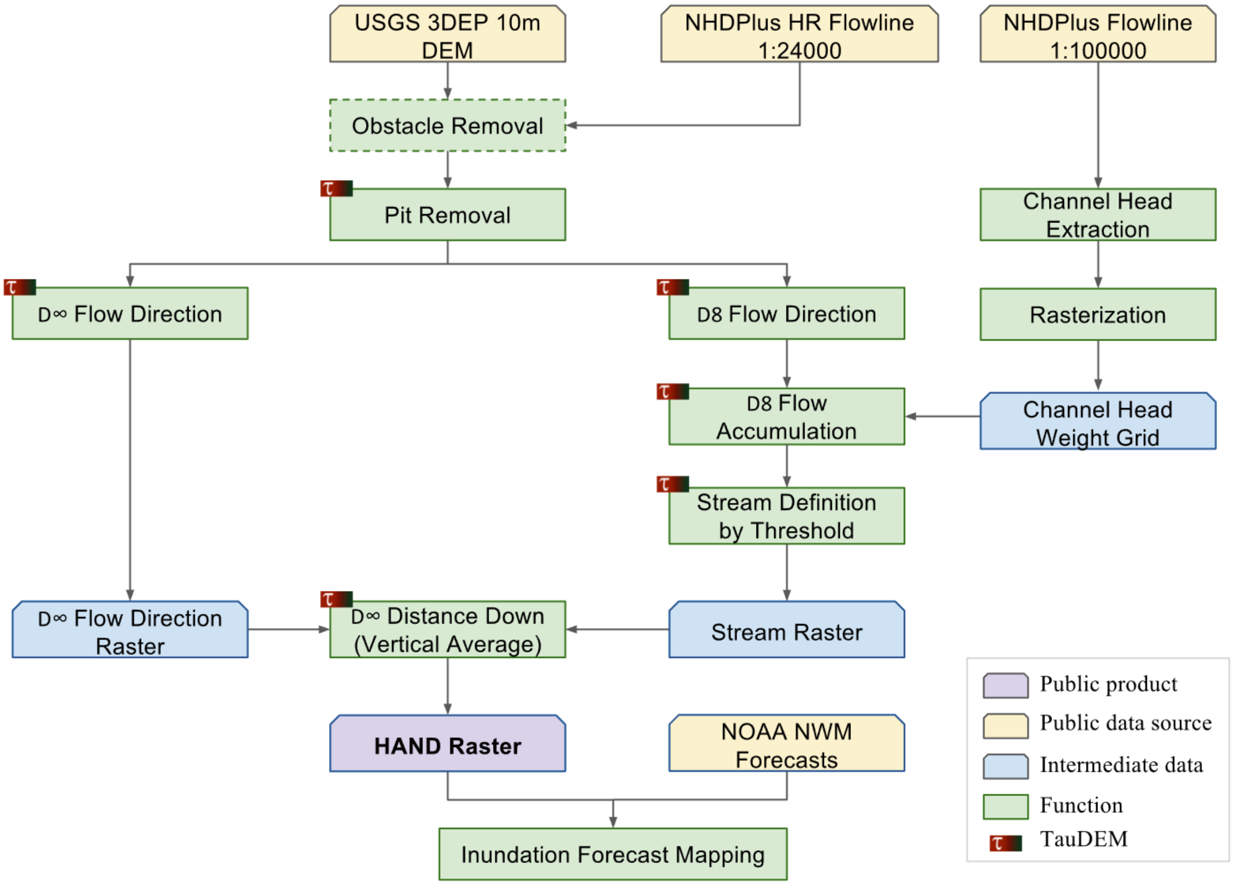


FIGURE 3. HAND computational workflow for a HUC unit. HR, high resolution

### *HAND Workflow and Computational Analysis*

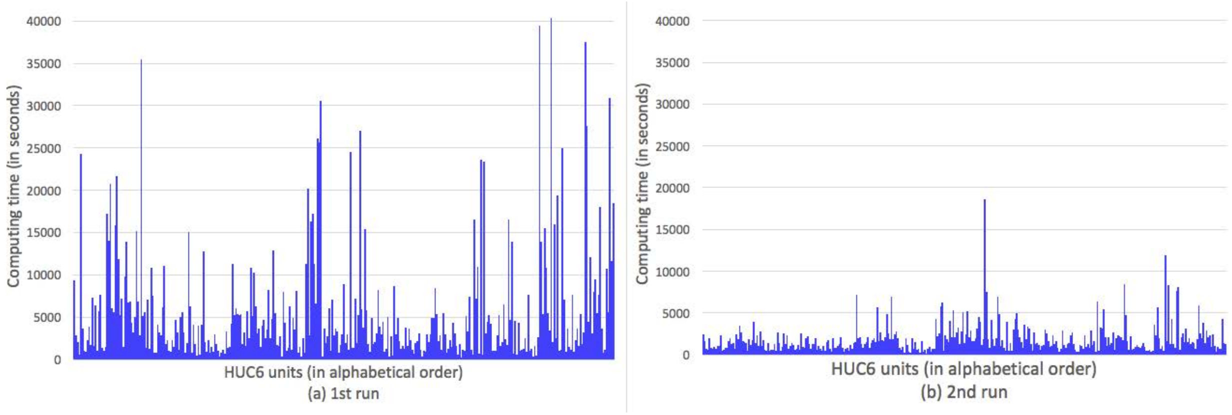
Figure 3 illustrates the HAND computational workflow. HAND, by its definition, is a geospatial raster dataset in which the value of each cell is the height above its nearest drainage. To identify the nearest drainage, we construct a flow direction grid of the same resolution and spatial extent using the $D\infty$ flow direction representation (Tarboton, 1997). The $D\infty$ grid is derived by TauDEM from the hydrologically conditioned DEM of the studied HUC unit. The hydrologic conditioning consists of two steps: obstacle removal and pit removal. The obstacle removal function adjusts the elevation of DEM cells that are on identified river flowlines but blocked by topographic structures such as roads and dams. The flowlines used for obstacle removal come from the high resolution NHDPlus dataset, which has more than 30 million reaches, in order to improve the coverage and accuracy. The methodology of obstacle removal is being prototyped and, thus, is currently integrated into the HAND workflow as an abstract interface. The pit removal function calls TauDEM's *pitremove* function (Tarboton *et al.*, 2008) which takes the raw DEM as input. The input DEM for the specified HUC unit is generated by clipping 3DEP DEM using GDAL, an open source geospatial data processing library. The clipping function uses the USGS NHD Water Boundary Dataset (WBD) to retrieve the boundary shape of a HUC unit and creates a 10-kilometer DEM clipping buffer to avoid edge effects along the HUC boundary. Since the 3DEP DEM is organized as a virtual raster in VRT format, there is no need to create a single DEM with CONUS coverage for clipping purpose.

With the $D\infty$ method, a flow routing network is constructed on a grid by analyzing the topographic data only. The next step in HAND is to compute a stream network in raster format, where streams are rasterized cells on a grid of the same resolution and spatial extent. The stream grid is also derived using the DEM model, with guidance from NHDPlus. This process includes a series of vector and raster processing functions. The vector processing step takes the Flowline layer of the medium resolution NHDPlus to identify channel heads. Each flowline feature has two attributes: FromNode and ToNode. Channel heads are identified if the FromNode of the corresponding flowline is not a ToNode (downstream) of any other flowlines. The output of this step is a point dataset, which is then rasterized to create a channel head weight grid. TauDEM's *D8* flow direction function generates a *D8* flow direction grid from the same hydrologically conditioned DEM used in calculating the $D\infty$ grid. The *D8* grid is then used by TauDEM's flow accumulation function (*aread8*) to generate weighted accumulated areas using the channel head weight grid. The *threshold* function in TauDEM is called with threshold value 1 to generate the stream grid from the *aread8* output. The result of this is a stream grid aligned with the DEM but initiated at the source of each NHDPlus stream. Taking the *D*∞ flow direction grid and the stream grid as input, TauDEM's Distance to Streams function (*DistanceDown*) produces the HAND value of each cell using the vertical distance measure. We can also use the horizontal distance or the combination of horizontal and vertical distance in this function if these additional distance grids are of interest.

We accelerate TauDEM to scale to thousands of processors and DEMs of tens of gigabytes through the XSEDE ECSS program. Through the work in Fan *et al.* (2014), Survila *et al.* (2016), and Yildirim *et al.* (2016), we have identified a set of computational bottlenecks of older TauDEM versions and improved the numerical performance and the parallel algorithms for the two flow direction functions in TauDEM by eliminating bottlenecks in file IO and runtime communication and developing more performant parallel algorithms.

### *Computational Experience*

The first HAND computation was on 331 out of the 336 HUC6 accounting units. Each unit comprises a computing job that was submitted to ROGER HPC. Each job used 60 to 180 processor cores based on a coarse estimation of computational intensity, described in the section, "Scalable Computing." The first run was completed on April 16, 2016 and consumed a total of 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 job queue. Figure 4(a) depicts the computing time of all 331 jobs. The large variation shows a heterogeneous computing profile for the 331 HUC6 units of different sizes, topographical, and hydrologic 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*∞) took, on average, 72.65% of the workflow computing time. The first run was conducted as a stress test to calibrate a more accurate computational intensity estimation for the units. Using the computation profile obtained from the first run, the workflow was adjusted for better configuration of edge contamination, DEM buffer size, and inlet identification from rivers passing through a watershed unit. This information also helped us capture CPU and memory requirements.



(a) 1st run

(b) 2nd run

FIGURE 4. Computing time distribution: comparison between the first (a) and second (b) run.

Following the first run, the second run was completed on May 29, 2016, using the calibrated workflow and newly accelerated *D8* and *D*∞ algorithms (Survila *et al.*, 2016). Figure 4(b) depicts the computing time for all of the 331 units. The second run finished in 36 hours and 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 took only 12.65% of the workflow computing time, on average. The majority, 70.57% time, was spent on GDAL commands for pre- and post-processing. The total input and output of HAND for CONUS takes about 5 terabyte disk space.

The main source of acceleration in TauDEM's flow direction algorithms (Tarboton, 1997) is illustrated in Figures 5 and 6. A computational strategy was applied to allow multiple processors to efficiently compute the flats resolving function, the most expensive function in the two flow direction algorithms. A flat is a set of contiguous cells on DEM with same elevation or zero slope value. Determining the flow direction on flat cells requires an iterative algorithm that is computationally costly. Figure 5 shows the distribution of 16,560,871 flats on a hydro-conditioned DEM of a HUC6 unit. The original TauDEM uses an implicit communication mechanism to exchange ghost zone data on the boundary of the decomposed data domains in each MPI process. This mechanism has the benefit of hiding the inter-process communication complexity with automatic ghost zone data exchange after each iteration of the flats resolving function. However, this mechanism introduces significant communication cost as more processors are used to analyze larger DEM, as demonstrated by the performance difference between the first and second HAND run. A strategy to reduce the communication cost by locating and localizing flats resolving was developed to process local flats that are fully contained in spatial domains on a MPI process without any communication. Flats whose boundary shape cross multiple processes are shared flats and processed via MPI communication functions. The identification of local and shared flats was efficiently implemented with $O$(n) computing complexity.

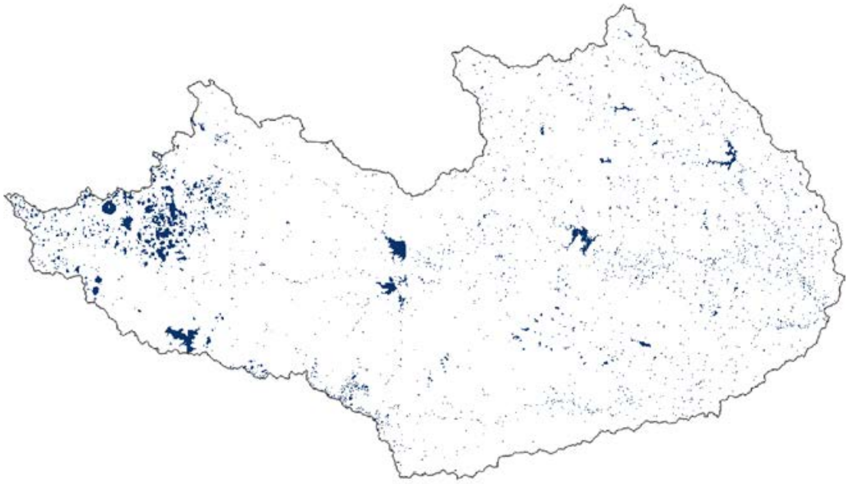


FIGURE 5. Illustration of the map of flats for HUC6 unit 120901, Middle Colorado-Concho. Flats, denoted as dark blue parts in the figure, in hydrologically conditioned DEM may include natural flats, flat surfaces in DEM (such as water surface where elevation information on the water channel beneath the water is not available), and filled pits.

With this strategy, an experiment on the *D8* algorithm was conducted on 4 computing nodes, using 1 to 32 processors, to evaluate the performance of flats resolving, the major bottleneck in both *D8* and *D∞*. TauDEM version 5.3.7, which has not yet incorporated the acceleration code, is used for comparison. For the parallel runs using 4, 8, 16, and 32 processors, respectively, 90.81%, 89.98%, 76.03%, and 66.89% local flats are identified and processed without inter-processor communication. The performance gain, measured as the time taken to finish the *D8* function, is shown in Figure 6. In TauDEM version 5.3.7, the flats resolving function takes the majority of the computing time in all cases, although both the flow direction function and the flats resolving function scale well as the number of processors increases. With the flats resolving acceleration, this function is no longer a bottleneck. The execution time of *D8* algorithm decreased from 3.4 hours to 6 minutes and 11 seconds on one processor and from 1227.22 seconds to 77 seconds on 32 processors. Using 32 processors, the flats resolving function requires only 2.36 seconds, compared to 1152.72 seconds on TauDEM version 5.3.7. In the accelerated version, the slightly worse performance of the *D8* function using 32 processors, compared to using 16 processors, indicates that the parallel IO cost outweighed the benefits from employing more than 16 processors on the 2.18GB DEM.

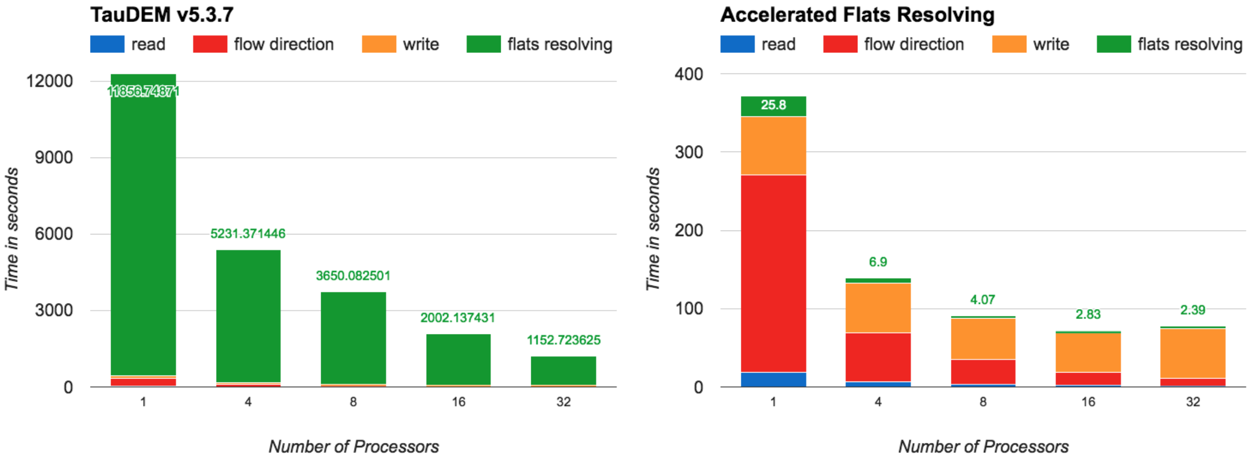


FIGURE 6. Performance of the D8 flow direction algorithm before and after the acceleration on flats resolving. HUC6 unit: 120901. DEM size: 42877 x 21711, 2.18GB. For visual purpose, the range on Y-axis is plotted based on actual maximal values obtained in the two tests, respectively.

## *Results*

Figure 7 shows the HAND map for CONUS, generated from the second run. Each HAND raster of an HUC unit is published as an OGC TMS map layer. A CONUS layer is created by merging all 331 HUC6 unit layers. The availability of this HAND dataset piqued our interest in evaluating the results and identifying methodological improvements, which is elaborated in our companion paper (Zheng *et al.*, 2017). The hydrologic and hydraulic comparison of HAND and other flood inundation mapping approaches can be found in Maidment *et al.* (2016c) and our companion paper Zheng *et al.* (2017).

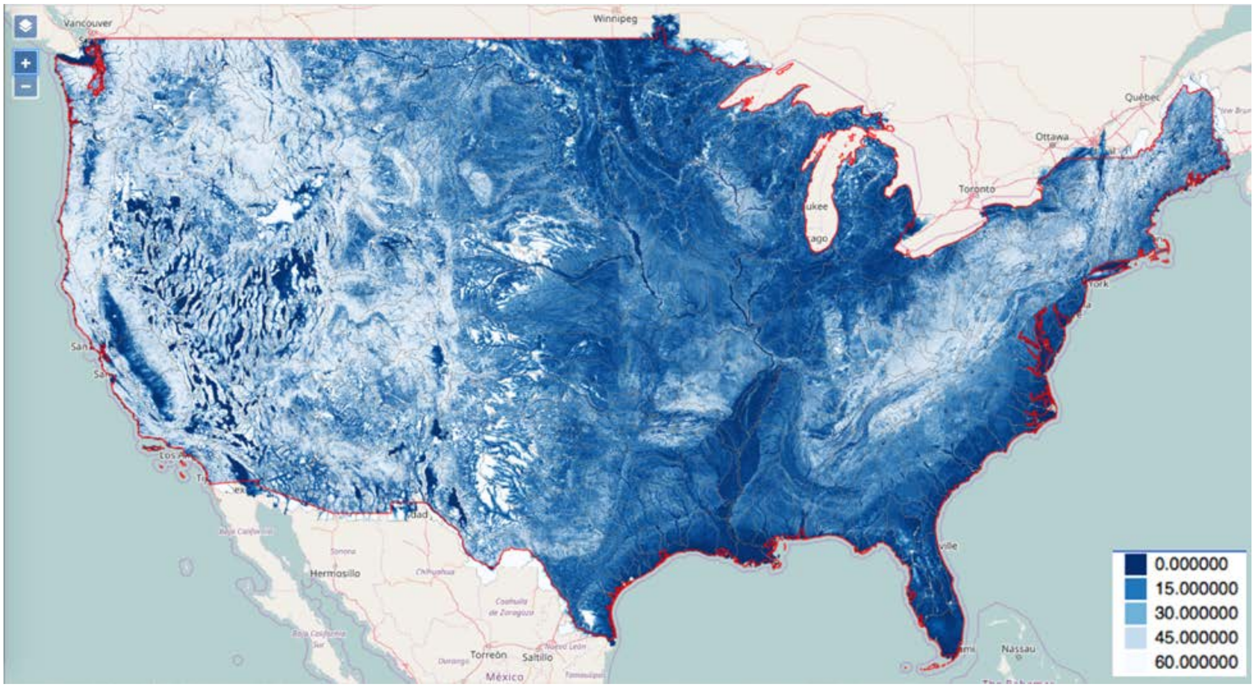


FIGURE 7. Hydrologic terrain of CONUS using the 1/3rd arc-second HAND with U.S. 5km boundary, created on 5/19/2016. Projection: Web Mercator (EPSG:3857). Coloring denotes HAND value in meters.

## HYDRAULIC PROPERTY TABLE AND INUNDATION MAPPING

Figure 8 shows the computation workflow of inundation mapping at CONUS scale. It has three major components. First, the hydraulic property table is calculated by a series of raster and vector computing that takes the HAND data, a pre-defined stage height list, and NHDPlus as input. This computation is decomposed at the HUC6 level into 331 computing jobs. The output tables for each unit are then merged as the CONUS-level table. Second, an inundation forecast table is computed for each NWM forecast time and stored as either CSV or NetCDF4 files. This table is computed at CONUS scale directly since it does not introduce significant computing cost. Third, the inundation mapping visualization process is invoked at the HUC6 level to generate map layers for each forecast table. A CONUS view of the inundation map is generated by merging HUC6-level map tiles.

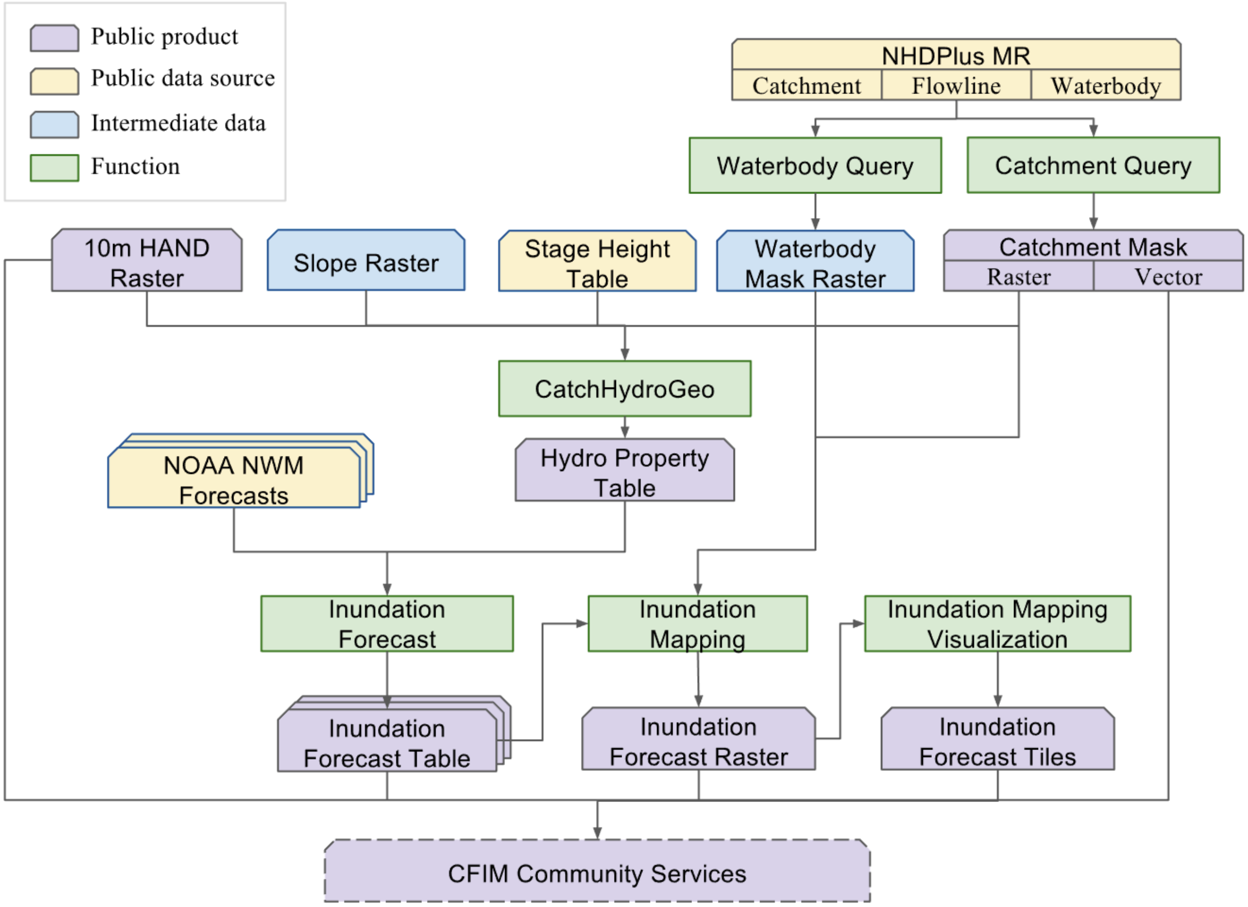


FIGURE 8. Preliminary inundation mapping workflow in continental flood inundation mapping (CFIM). MR, medium resolution.

### *Computing the Hydraulic Property Table*

A list of hydraulic properties described in section “Data and Computational Challenges” can be computed for each catchment along with corresponding stage height from HAND and NHDPlus. In the current configuration, the stage height table includes 82 heights, increasing above zero at one-foot interval. The number of records in the hydraulic property table is thus (*number_of_river_reaches* x *number_of_stages*).

A number of auxiliary attributes needed for the computation of these properties, such as catchment ID, river segment slope, and length, are available in the NHDPlus file geodatabase. In order to carry the catchment ID (COMID) information embedded in the vector format of NHDPlus into the raster computing of HAND, a catchment grid of the same spatial extent, projection, and resolution as HAND is generated first. Catchment polygons in an HUC6 unit is retrieved from the *Catchment* layer in NHDPlus and then rasterized. A high-performance raster processing function, *CatchHydroGeo*, is developed using the TauDEM parallel computing framework to derive hydraulic properties from HAND, the catchment grid, the stage height table, and the slope grid (one output of the TauDEM *D*∞ flow direction function), as shown in Figure 8. In this computational process, the query for the catchment polygon of COMIDs on the Flowline and Catchment layer, which have 2,691,344 and 2,647,454 records, respectively, is optimized to take only (*n* x *log(*2,691,344*)*) lookup operations to finish. As a result, computing the hydraulic property table for CONUS took about 2.5 hours on ROGER. The scientific evaluation of the computed river geometry and rating curve is done in Maidment *et al.* (2016c).

### *Computing for Real-time Inundation Forecast*

The availability of the hydraulic property table allows straightforward translation of water depth information from NOAA NWM streamflow forecast information. The NWM streamflow forecast of a river reach (station) is linearly interpolated to the water depth by looking up the hydraulic property table, which has the (*water depth* : *streamflow*) mapping for each stage height defined by the stage height table. In addition, other inundation criteria specific to certain communities, e.g., anomaly map for emergency management, can be incorporated in this process. The computation of the inundation forecast table queries two relational tables, i.e., the hydraulic property table and the NWM forecast, and does not involve geospatial processing.

The inundation forecast table can be used by hydrologists and others who understand the NHDPlus and NWM. For users who need to make decisions based on inundation mapping visualization, near real-time inundation mapping visualization is needed. This can be done in two ways: static or dynamic mapping. We experimented a static mapping solution, shown in Figure 8. First, an inundation map raster is generated by comparing the HAND value of a grid cell and the forecast water depth, using the catchment grid in order to identify the relevant cell. If a cell's forecast water depth is larger than the HAND value, it is marked as inundated. A masking option is available to show or hide cells covered by the masking layer. For example, all of the water body areas are masked because the CFIM model is not suitable for inundation analysis on water body objects. The inundation map is generated at the HUC6 level so that all the units can be computed in parallel. Second, the visualization function takes the GeoTIFF file of a unit's inundation map and a coloring scheme and generates a tile pyramid that covers multiple zoom levels. The tiles are published as OGC TMS map layers for visualization. Aggregating HUC6-level tiles into a single CONUS map layer is straightforward using image overlay techniques.

We conducted a CONUS scale experiment on the entire inundation forecast process on 15 computing nodes of ROGER HPC. In the experiment, the short-range NWM forecast at forecast initialization time 12:00:00am, March 23, 2017 UTC was used. Table 2 shows the time distribution of each step. The first three steps needed to generate the inundation forecast maps took 44 minutes 31 seconds on 15 computing nodes, which means using 15 or more computing nodes to generate inundation maps is sufficient to match the hourly pace of the short-range NWM forecast data streaming. The TMS-based visualization computation, however, required almost 12 hours to generate the CONUS view because of large amount of IO cost on tile creation.

TABLE 2. Execution time of the inundation map generation process, in seconds.

| | NWM Download | Forecast Table | Forecast Map | HUC6 TMS | CONUS TMS |
|---|---|---|---|---|---|
| Time | 49 | 603 | 1779 | 27,845 | 13,892 |
| Data size | 780MB unzipped (52MB x 15 forecasts) | 889MB (60MB x 15 forecasts) | 223GB (4901 maps for 331 HUC6 units) | 45GB (4,140,833 tiles; 8 zoom levels) | 35GB (2,405,624 tiles; 8 zoom levels) |

The drawback of the static mapping approach in the TMS visualization step can be resolved through a dynamic mapping process which queries HAND and a few auxiliary rasters, the hydraulic property table, and the inundation forecast table directly, and renders the inundation map on the fly. The dynamic mapping process is still being developed using Esri’s mosaic dataset and raster function techniques. Upon completion, the inundation mapping step can be completed in 11 minutes because no inundation raster maps or visualization tiles are needed.

# DISCUSSION AND CONCLUSION

The HAND workflow is scalable to higher resolution DEM and NHDPlus. A preliminary scalability study using LiDAR-derived 3m DEM and NHDPlus HR is conducted at each step of the HAND workflow, shown in Figure 9. Four scenarios by combining two resolutions of DEM (i.e., 3m and 10m) with two resolutions of NHDPlus (i.e., MR and HR) are studied, shown as the columns in Figure 9. Results show that DEM resolution is the main determinant to the execution time of HAND, shown by the 3m columns (5.4GB, 80160 x 48058 cells) and the 10m columns (595MB, 26730 x 16025 cells). The most expensive functions in 3m DEM computation are the sequential DEM clipping and post-processing (i.e., removing the 10-kilometer buffer when creating HAND) by GDAL. In contrast, TauDEM performance scales well in proportion to DEM size. The major impact of using NHDPlus high-resolution dataset is on the performance of flowline retrieval because the join function operates on 30 million vectors, instead of 2.7 million in the medium resolution dataset.

While this paper focuses on the computational aspects of CFIM evaluation, a comprehensive scientific evaluation of HAND, river geometry, and inundation mapping results has been conducted at the 2016 Summer Institute of the Consortium of Universities for the Advancement of Hydrologic Science, Inc. and the National Water Center (Maidment *et al.*, 2016c). Assumptions made for the CFIM methodology can be found in our companion paper (Zheng *et al.*, 2017).

We successfully demonstrated the computational feasibility of continental-scale flood inundation mapping with the cyberGIS framework. The computation of HAND for the relevant 331 HUC6 units on CONUS achieved a turnaround time of 1.5 days on the ROGER supercomputer. An additional 2.5 hours was taken to compute the hydraulic property table and store it in NetCDF4 format. The inundation forecast process took 45 minutes for producing 15 hourly inundation tables and maps (excluding the TMS tile pyramiding step) on a short-range NOAA NWM forecast initialization time stamp for CONUS coverage.

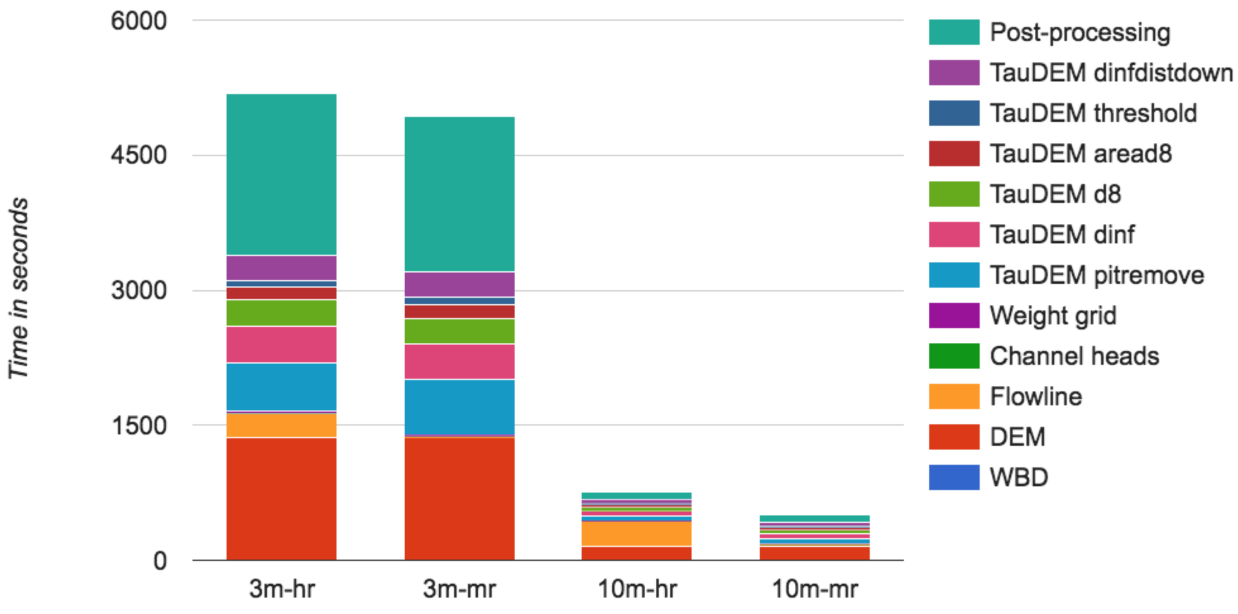


FIGURE 9. HAND scalability to 3DEP DEM and NHDPlus resolution. Study watershed: HUC6 unit 120402, Galveston Bay-Sabine Lake. DEM resolutions: 1/9th arc-second (3 m) and 1/3rd arc-second (10 m). NHDPlus resolutions: 1:100,000 (mr) and 1:24,000 (hr). Six computing nodes (120 processors) on Resourcing Open Geospatial Education and Research were used in each case. WBD, water boundary dataset.

All of the CFIM data and visualization layers are published online (NFIE CFIM data repository. Accessed January 01, 2018, https://web.corral.tacc.utexas.edu/nfiedata/HAND/) to engage further community evaluation. The availability of HAND at 1/3rd arc-second resolution and CONUS scale has auspicious, broad, and significant research implications, enabling pertinent research communities to conduct large-scale flood inundation mapping research by pertinent research communities. The CFIM collaboration resulted in significant scalability and performance improvement of cyberGIS and TauDEM software. The CFIM computational model is based on open source geospatial and hydrologic software that is able to harness massive computing power for enabling the computation of the CFIM workflow. The computation on ROGER seamlessly exploits its HPC and cloud components for workflow methodology development and CFIM workflow computation, visualization, and validation. We will continue to improve the usability of the CFIM computational framework to couple related hydrologic modeling processes for producing flood inundation forecasts at high spatial and temporal resolutions. We will build an interactive methodology building and validation environment online using CyberGIS Jupyter (Yin *et al.*, 2017) to further accelerate CFIM research, data and software integration, and computation.

## ACKNOWLEDGEMENTS

This work is part of the ECSS project (award number ENG140009) of XSEDE that is supported by NSF under grant number 1053575. This research is supported in part by USGS under grant number G14AC00244 and NSF under grant numbers 1047916 and 1343785. The work used the ROGER supercomputer, which is supported by NSF under grant number: 1429699. The authors acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin for providing HPC and storage resources that have contributed to the research results reported within this paper. HydroShare is being developed under NSF grants ACI 1148453 and 1148090. 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. The authors are grateful for the insightful discussions with Steve Kopp and Dean Djokic at Esri, and Larry Stanislawski at USGS. The authors would like to thank Dandong Yin at the University of Illinois at Urbana-Champaign for developing the HAND Jupyter notebook.

## LITERATURE CITED

Bockelman, B., T. Cartwright, J. Frey, E. M Fajardo, B. Lin, M. Selmeci, T. Tannenbaum, and M. Zvada, 2015. Commissioning the HTCondor-CE for the Open Science Grid. *Journal of Physics: Conference Series* (664).

David, C.H., D.R. Maidment, G.-Y. Niu, Z.-L. Yang, F. Habets, and V. Eijkhout, 2011. River Network Routing on the NHDPlus Dataset. *Journal of Hydrometeorology* 12(5):913-934.

David, Cédric H., Zong-Liang Yang, Seungbum Hong, 2013, Regional-scale river flow modeling using off-the-shelf runoff products, thousands of mapped rivers and hundreds of stream flow gauges. *Environmental Modelling & Software* 42: 116-132. DOI: 10.1016/j.envsoft.2012.12.011.

Fan, Y., Y. Liu, S. Wang, D. Tarboton, A. Yildirim, and N. Wilkins-Diehr, 2014. Accelerating TauDEM as a Scalable Hydrological Terrain Analysis Service on XSEDE. In: *Proceedings of the 2014 Annual Conference on Extreme Science and Engineering Discovery Environment*, Atlanta, GA, USA, July 13-18. ACM, New York. DOI: 10.1145/2616498.2616510

Hodges, B., 2013. Challenges in Continental River Dynamics. *Environmental Modelling and Software* 50:16-20. DOI: 10.1016/j.envsoft.2013.08.010

Horsburgh, Jeffery S., Mohamed M. Morsy, Anthony M. Castronova, Jonathan L. Goodall, Tian Gan, Hong Yi, Michael J. Stealey, and David G. Tarboton, 2016. HydroShare: Sharing Diverse Environmental Data Types and Models as Social Objects with Application to the Hydrology Domain. *Journal of the American Water Resources Association (JAWRA)* 52(4):873–889. DOI: 10.1111/1752-1688.12363

Hu, H., X. Hong, J. Terstriep, Y. Liu, M. Finn, J. Rush, J. Wendel, and S. Wang. 2016, TopoLens: Building A CyberGIS Community Data Service for Enhancing the Usability of High-resolution National Topographic Datasets. In: *Proceedings of the 2016 Annual Conference on Extreme Science and Engineering Discovery Environment (XSEDE'16)*. Miami, Florida. July 17-21. ACM, New York, pp. 39:1--39:8. DOI: 10.1145/2949550.2949652

Lawrence, Katherine A., Michael Zentner, Nancy Wilkins-Diehr, Julie A. Wernert, Marlon Pierce, Suresh Marru, Scott Michael, 2015. Science gateways today and tomorrow: positive perspectives of nearly 5000 members of the research community. *Concurrency and Computation: Practice and Experience* 27: 4252–4268. DOI: 10.1002/cpe.3526

Maidment, David R., 2016a. Conceptual Framework for the National Flood Interoperability Experiment. *Journal of the American Water Resources Association (JAWRA)* 53(2): 245-257. DOI:

10.1111/1752-1688.12474

Maidment, David R., 2016b. Open Water Data in Space and Time. *Journal of the American Water Resources Association (JAWRA)* 52(4):816–824. DOI: 10.1111/1752-1688.12436

Maidment, David R., Adnan Rajib, Peirong Lin, Edward P. Clark, 2016c. National Water Center Innovators Program Summer Institute Report 2016. Consortium of Universities for the Advancement of Hydrologic Science, Inc. and the National Water Center. https://www.cuahsi.org/uploads/library/cuahsi_tr13_8.20.16.pdf

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. DOI: 10.1016/j.jhydrol.2011.03.051

Nobre, Antonio Donato, Luz Adriana Cuartas, Marcos Rodrigo Momo, Dirceu Luís Severo, Adilson Pinheiro, and Carlos Afonso Nobre. 2016. HAND contour: a new proxy predictor of inundation extent. *Hydrological Process*es 30: 320–333. DOI: 10.1002/hyp.10581

Qiu, Judy, Jaliya Ekanayake, Thilina Gunarathne, Jong Youl Choi, Seung-HeeBae, Hui Li, Bingjing Zhang, Tak-Lon Wu, Yang Ruan, Saliya Ekanayake, Adam Hughes, and Geoffrey Fox. 2010. Hybrid cloud and cluster computing paradigms for life science applications. *BMC Bioinformatics* 11(12):S3.

Rennó, Camilo Daleles, Antonio Donato Nobre, Luz Adriana Cuartas, João Vianei Soares, Martin G. Hodnett, Javier Tomasella, and Maarten J. Waterloo. 2008. HAND, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazonia. *Remote Sensing of Environment* 112(9) 3469-3481. DOI: 10.1016/j.rse.2008.03.018

Rodda, Harvey J. E. 2005. The Development and Application of a Flood Risk Model for the Czech Republic. *Natural Hazards* 36(1): 207-220. DOI: 10.1007/s11069-004-4549-4

Snow, Alan D., Scott D. Christensen, Nathan R. Swain, E. James Nelson, Daniel P. Ames, Norman L. Jones, Deng Ding, Nawajish S. Noman, Cedric H. David, Florian Pappenberger, and Ervin Zsoter, 2016. A High-Resolution National-Scale Hydrologic Forecast System from a Global Ensemble Land Surface Model. *Journal of the American Water Resources Association (JAWRA)* 52(4): 950-964, DOI: 10.1111/1752-1688.12434

Survila, K., A.A. Yildirim, T. Li, Y. Liu, D.G. Tarboton, and S. Wang, 2016. A Scalable High-performance Topographic Flow Direction Algorithm for Hydrological Information Analysis. In: *Proceedings of the 2016 Annual Conference on Extreme Science and Engineering Discovery Environment (XSEDE'16)*. Miami, Florida. July 17-21. ACM, New York, pp. 11:1--11:7. DOI: 10.1145/2949550.2949571

Swain, Nathan R., K. Latu, Scott D. Christensen, Norman L. Jones, E. James Nelson, Daniel P. Ames, Gustavious P. Williams, 2015. A Review of Open Source Software Solutions for Developing Water Resources Web Applications. *Environmental Modeling & Software* 67: 108-117.

Swain, N. R., S. D. Christensen, A. D. Snow, H. Dolder, G. Espinoza-Dávalos, E. Goharian, N. L. Jones, E. J. Nelson, D. P. Ames and S. J. Burian, 2016. A new open source platform for lowering the barrier for environmental web app development. *Environmental Modelling & Software* 85: 11-26.

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): 309-319. DOI: 10.1029/96WR03137

Tarboton, D.G. and M.E. Baker, 2008. Towards an algebra for terrain-based flow analysis. In: *Representing, modeling and visualizing the natural environment: innovations in GIS* 12, NJ Mount, GL Harvey, P. Aplin and G. Priestnall (Editors). CRC Press, Florida. pp: 167-194. DOI:

10.1201/9781420055504.ch12

Tarboton, D. G., K. A. T. Schreuders, D. W. Watson, and M. E. Baker, 2009. Generalized terrain-based flow analysis of digital elevation models. In: *18th World IMACS Congress and MODSIM09 International Congress on Modelling and Simulation*, July 2009. R. S. Anderssen, R. D. Braddock and L. T. H. Newham (Editors). Modelling and Simulation Society of Australia and New Zealand and International Association for Mathematics and Computers in Simulation, pp. 2000-2006. URL: http://www.mssanz.org.au/modsim09/F4/tarboton_F4.pdf

Tarboton, D. G., R. Idaszak, J. S. Horsburgh, J. Heard, D. Ames, J. L. Goodall, L. Band, V. Merwade, A. Couch, J. Arrigo, R. Hooper, D. Valentine and D. Maidment, 2014. HydroShare: Advancing Collaboration through Hydrologic Data and Model Sharing. In: *Proceedings of the 7th International Congress on Environmental Modelling and Software*, San Diego, California, USA. D. P. Ames, N. W. T. Quinn and A. E. Rizzoli (Editors). International Environmental Modelling and Software Society (iEMSs), ISBN-13: 978-88-9035-744-2

Tavakoly, Ahmad A., Alan D. Snow, Cédric H. David, Michael L. Follum, David R. Maidment, Zong-Liang Yang, 2017. Continental Scale River Flow Modeling of the Mississippi River Basin Using High-Resolution NHDPlus Dataset. *Journal of the American Water Resources Association* 53(2):258–279. DOI: 10.1111/1752-1688.12456

Tesfa, Teklu K., David G. Tarboton, Daniel W. Watson, Kimberly A.T. Schreuders, Matthew E. Baker, and Robert M. Wallace, 2011. Extraction of hydrological proximity measures from DEMs using parallel processing, *Environmental Modelling & Software* 26(12): 1696-1709. DOI: 10.1016/j.envsoft.2011.07.018.

Towns, John, Timothy Cockerill, Maytal Dahan, Ian Foster, Kelly Gaither, Andrew Grimshaw, Victor Hazlewood, Scott Lathrop, Dave Lifka, Gregory D. Peterson, Ralph Roskies, J. Ray Scott, and Nancy Wilkins-Diehr, 2014. XSEDE: Accelerating Scientific Discovery. *Computing in Science & Engineering* 16(5): 62-74. DOI:10.1109/MCSE.2014.80

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

Wang, Shaowen, 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, Shaowen, Luc Anselin, Budhendra Bhaduri, Christopher Crosby, Michael F. Goodchild, Yan Liu, and Timothy L. Nyerges, 2013. CyberGIS Software: A Synthetic Review and Integration Roadmap. *International Journal of Geographical Information Science* 27(11): 2122-2145

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

Wilkins-Diehr, N., S. Sanielevici, J. Alameda, J. Cazes, L. Crosby, M. Pierce, and R. Roskies, 2015. An Overview of the XSEDE Extended Collaborative Support Program. In: *High Performance Computer Applications 6th International Conference, ISUM 2015*. Mexico City, Mexico, March 9-13. Gitler, Isidoro, Klapp, Jaime (Eds.) Springer International Publishing. ISBN-13: 978-3-319-32243-8, pp: 3-13. DOI: 10.1007/978-3-319-32243-8

Yildirim, A.A., D.G. Tarboton, Y. Liu, N.S. Sazib, S. Wang, 2016. Accelerating TauDEM for Extracting Hydrology Information from National-Scale High Resolution Topographic Dataset. In: *Proceedings of the 2016 Annual Conference on Extreme Science and Engineering Discovery Environment*

*(XSEDE'16)*. Miami, Florida. July 17-21. ACM, New York, pp. 3:1--3:2. DOI: 10.1145/2949550.2949582

Yin, D., Y. Liu, A. Padmanabhan, J. Terstriep, J. Rush, S. Wang, 2017. A CyberGIS-Jupyter Framework for Geospatial Analytics at Scale. In: *Proceedings of the 2017 Practice & Experience in Advanced Research Computing (PEARC'17)*. New Orleans, LA. July 9–13.

Zheng, X., 2015. Hydraulic fabric: an information framework for river channel cross section data. Master Dissertation, University of Texas at Austin, Austin, Texas.

Zheng, Xing, David Tarboton, David R. Maidment, Yan Y. Liu, and Paola Passalacqua, 2017. River Channel Geometry and Rating Curve Estimation Using Height Above the Nearest Drainage. *Journal of the American Water Resources Association (JAWRA)*, Draft for review.