

Accelerating TauDEM for Extracting Hydrology Information from National-Scale High Resolution Topographic Dataset

[Extended Abstract]

Ahmet Artu Yıldırım^{1,2,3}
ayild@illinois.edu

David Tarboton⁴
dtarb@usu.edu

Yan Liu^{1,2,3}
yanliu@illinois.edu

Nazmus Shams Sazib⁴
sazibap25@gmail.com

Shaowen Wang^{1,2,3}
shaowen@illinois.edu

¹CyberGIS Center for Advanced Digital and Spatial Studies

²National Center for Supercomputing Applications

³University of Illinois at Urbana-Champaign, Champaign, IL, USA

⁴Utah Water Research Laboratory, Utah State University, Logan, UT, USA

ABSTRACT

We present performance improvements on parallel hydrology algorithms in TauDEM suite that allowed us to process the 10m USGS 3DEP topographic dataset (667GB, 180 billion raster cells) by adopting a block-wise data decomposition feature and maximizing the disk parallelism.

Categories and Subject Descriptors

C.2.4 [Distributed Systems]: Distributed Applications

Keywords

Parallel pitremove algorithm; TauDEM; HPC

1. INTRODUCTION

Digital elevation model (DEM) is widely used digital model to store elevations of a terrain. TauDEM (Terrain Analysis Using Digital Elevation Models) is a suite of DEM tools for high performance extraction and analysis of hydrologic information. Computer programs in TauDEM take advantage of parallel processing using the message passing interface (MPI). TauDEM is broadly used by hydrologists in desktop and cluster computing environments through direct software downloads, community software deployed on XSEDE, as well as cyberinfrastructure-based online services (e.g. TauDEM services integrated in Open Topography (<http://opentopography.org>) and CyberGIS gateway (<http://sandbox.cigi.illinois.edu>)). Accelerating TauDEM performance has been the focus of an XSEDE Extended Collaborative Support Services (ECSS) effort since 2013 [1].

This paper shares our experience in scaling TauDEM's communication and IO module to large DEMs such as the

10m USGS 3DEP topographic dataset (667GB, 180 billion raster cells) for the conterminous U.S. Removing depressions (pits) of a digital elevation model (DEM) is an essential step (referred to as *pitremove*) in many geoscience workflows [2]. We found that this TauDEM function was unable to fill the depressions of the entire 10m 3DEP DEM in 2 days using 4,096 processor cores on Stampede supercomputer. TauDEM only supports row-wise data decomposition of input DEM data without considering the 2D sizes of the input data domain. In TauDEM's current communication module, edge data are exchanged between adjacent vertical neighbors, as a result of using row-wise data decomposition. When the number of processes is large, such data decomposition results in thin and long stripes of data, distributed across processes, which leads to two issues. First, these long data stripes lead to significantly large message size for exchanging boundary data values in inter-process communication. Second, most of TauDEM algorithms are iterative and the number of iterations requiring inter-process communications highly depends on how data is decomposed across processes. In pitremove function, for example, the algorithm needs to communicate if cells in a pit are located on more than one process. If the raster data on each process is thinner vertically and longer horizontally, it is more likely that cells in a pit are distributed on multiple processes. As a result, more iterations need inter-process communications which are implemented as collective MPI calls in TauDEM. Therefore, the row-wise decomposition in current TauDEM exhibits poor data locality for pitremove and results in high communication cost. Third, TauDEM creates a single GeoTIFF file using GDAL IO routines where each MPI process writes its own partition to the same file in order by its rank number. Such an approach exhibited poor performance in writing big raster files because only one process performs writing operation at a time. We addressed these issues by providing a block-wise data decomposition feature and an embarrassingly parallel IO module for each process to write its own output raster through the virtual raster concept. After the proposed improvements, TauDEM pitremove successfully filled the depressions of the entire 3DEP DEM within 2 hours using 400 processors on ROGER supercomputer.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

XSEDE16 July 17-21, 2016

© 2016 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-4755-6/16/07.

DOI: <http://dx.doi.org/10.1145/2949550.2949582>

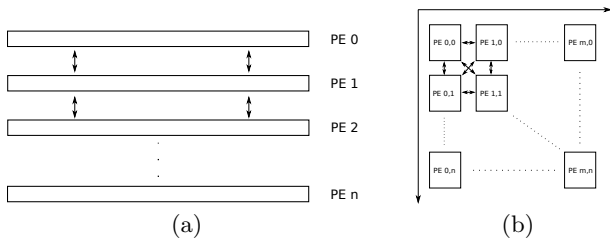


Figure 1: Data Decomposition Approaches; arrows shows communication between processors. (a) Row-wise data decomposition (b) Block-wise data decomposition

2. ACCELERATION

We implemented block-wise data decomposition to overcome the issues of row-wise data decomposition in TauDEM. Each processor is assigned to a distinct 2D block instead of a row of data. The block-wise decomposition fits well with how geographically pits are distributed on earth and, thus, is able to improve pit data locality on each process. We defined a Cartesian topology where each processor is identified by a unique 2D cartesian coordinate and is responsible of processing a distinct block of data. Block-wise data decomposition, however, changes the current TauDEM communication pattern that we used non-blocking MPI routines. With row-wise decomposition, a process communicates with at most two neighbors. With block-wise decomposition, one process can communicate with up to 8 processes. Figure 1 illustrates the communication patterns in row-wise and block-wise data decomposition strategies, respectively.

VRT is a the virtual raster format that enables composition of a virtual raster dataset of multiple files in which each data file is stored separately on disk. Metadata information is stored as an XML VRT index file. The current version of TauDEM writes results into a single file leading to low disk throughput. We adopted the VRT model for writing TauDEM output to maximize the disk parallelism. Each processor writes its own result data to a single GeoTIFF file using GDAL IO routines to take advantage of fast parallel writing capabilities of the file system. Then, the master node writes the XML-formatted VRT index file.

3. EVALUATION

We implemented the block-wise decomposition and VRT output option in TauDEM without the need to change the algorithm of each TauDEM function, attributed to the modular software architecture of TauDEM. Although the use of non-blocking communication may change the order of message arrival, the implementation handles it correctly. Thorough tests against the TauDEM Test Suite data and results verified the correctness of our implementation.

We conducted performance evaluation experiments on the Tennessee DEM of size approximately 11 GB to measure the effectiveness of the implemented improvements on TauDEM pitremove program on the Stampede supercomputer. Table 1 shows the experimental results used to compare the performance between the previous version and the new version with the improvements. TauDEM pitremove iteratively fills the depressions so computation time is somewhat dependent on the topography and distribution of pits over the DEM. Experiments show that when less than 128 processors are

Results/Processors	32	64	128	256
Compute Time (prev.)	152.43	109.30	92.43	86.98
Compute Time (impr.)	202.35	137.16	89.27	59.76
# of Iterations (prev.)	309	337	393	515
# of Iterations (impr.)	285	287	289	293
I/O Write Time (prev.)	58.16	54.00	52.50	55.24
I/O Write Time (impr.)	20.84	4.60	4.52	2.95

Table 1: Compute times, I/O write times and number of iterations of TauDEM pitremove program using 32, 64, 128 and 256 processors for the previous version (prev.) and the version with improvements (impr.); times in seconds

utilized, pitremove with row-wise data decomposition takes less computation time. One reason of that might be that there are fewer communication interactions with neighboring processors in the previous version, however the improved version starts outperforming the previous version with more processors, demonstrating better scalability in the computations. We obtained significant performance gains in writing results to disk using the VRT approach that we implemented. The test on the entire 10m 3DEP DEM fills pits on a 637,212 (columns) x 280,812 (rows) input raster. In 116-minute running time, computation of the pit filling part is completed within 114 minutes, and reading and writing of entire DEM (IO) are completed within 38 seconds and 71 seconds, respectively using 400 processors on ROGER supercomputer.

4. CONCLUSION

We have improved the computational scalability of TauDEM to a new level that is capable of handling the entire 3DEP DEM at 10m resolution for the conterminous U.S. in our pitremove evaluation. We are working with TauDEM team to verify similar scalability results in all TauDEM functions. The capability of handling the entire 10m 3DEP DEM in pitremove opens the door to new hydrological studies at national scale. A new TauDEM version will be released soon to provide the new capability to community users.

Acknowledgments

This work is part of the ECSS project (award number EAR130008) of XSEDE, which is supported by National Science Foundation (NSF) grant number ACI-1053575. This material is based in part upon work supported by the NSF under grant number 1047916. The work used the ROGER supercomputer, which is supported by NSF under grant number: 1429699. We also acknowledge the Texas Advanced Computing Center (TACC) for the Stampede system.

5. REFERENCES

- [1] Y. Fan, Y. Liu, S. Wang, D. Tarboton, A. Yildirim, and N. Wilkins-Diehr. Accelerating taudem as a scalable hydrological terrain analysis service on xsede. In *Procs. of the 2014 Annual Conf. on Extreme Sci. and Eng. Disc. Env.*, XSEDE '14, pages 5:1–5:2. ACM, 2014.
- [2] C. Wallis, R. Wallace, D. G. Tarboton, D. W. Watson, K. A. T. Schreuders, and T. K. Tesfa. Hydrologic terrain processing using parallel computing. pages 2540 – 2545. 18th World IMACS / MODSIM Congress, 2009.