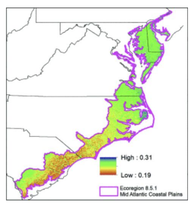
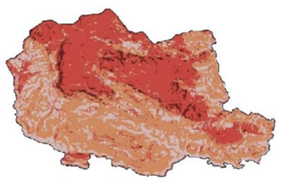
Advancing the GEOtiled Framework Through Scalable Terrain Parameter Computation

Gabriel Laboy (Student)[1], Paula Olaya (Advisor)[1], Jack Marquez (Advisor)[1], Michael Sutherlin (Advisor)[1], Rodrigo Vargas (Advisor)[2], and Michela Taufer (Advisor)[1]

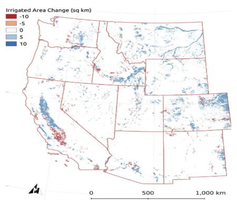
Motivation

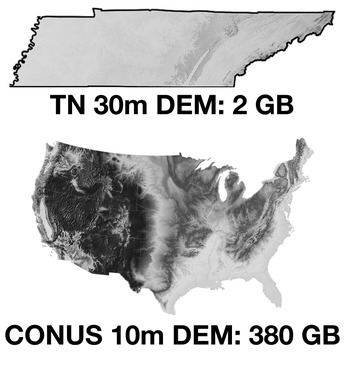
Scientists use **terrain parameters** (e.g. elevation, slope, aspect) in their analysis of various Earth Science studies, such as soil moisture prediction, fire propagation zoning, and hydrology mapping.





Propagation risk - aspect





TN 30m DEM: 2 GB

CONUS 10m DEM: 380 GB

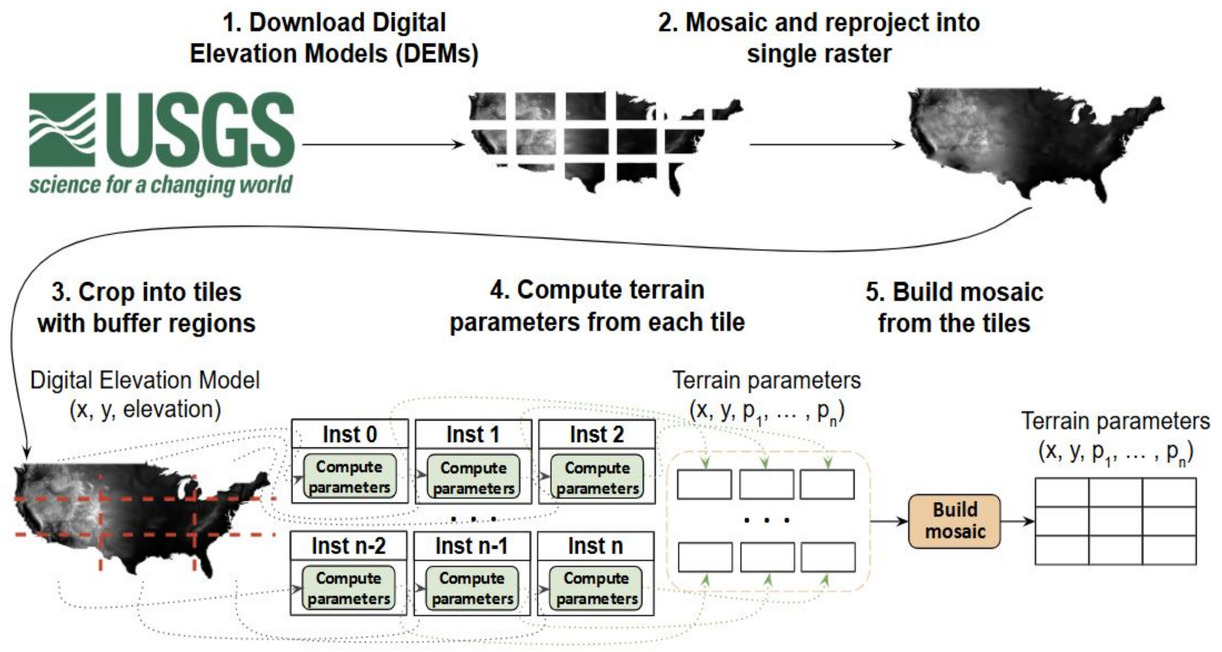
Problem: Higher resolution DEMs have larger file sizes, causing existing GIS libraries to compute parameters much slower.

GDAL	SAGA
+ Fast	- Slow
- Does not scale	- Does not scale
- Computes few terrain parameters	+ Computes more terrain parameters
- Designed for data abstraction	+ Implements geoscientific algorithms

Our Solution:
We update GEOtiled – a workflow that integrates GDAL and SAGA libraries to provide rapid, scalable computation of terrain parameters by leveraging data decomposition and parallel computation – to GEOtiled 2.0. Our optimizations include:

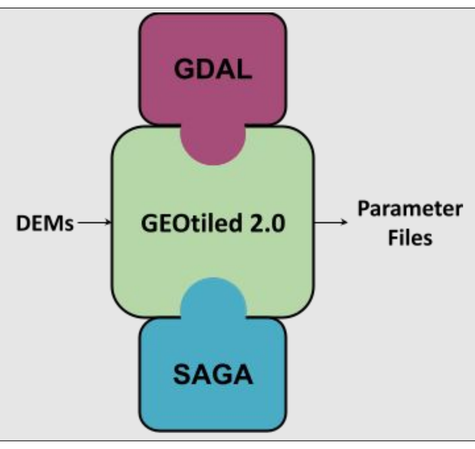
- Parallelizing the crop procedure for scalable computation
- Updating the mosaic algorithm to a better performing one
- Unifying parallel processes in the workflow to minimize overhead

GEOtiled Framework



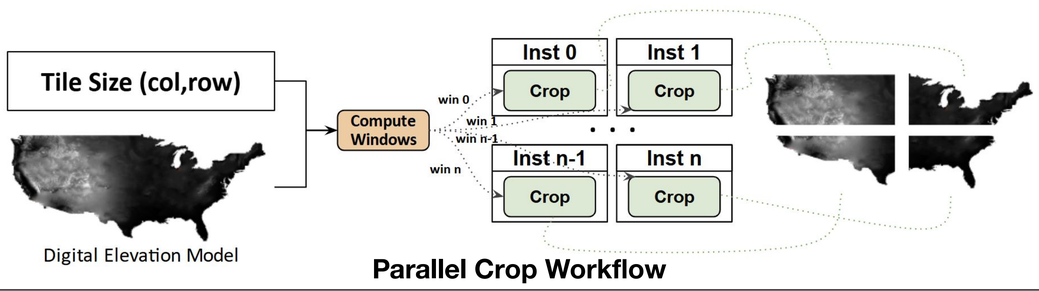
Full GEOtiled Workflow

GEOtiled 2.0



- Gives the option to choose between GDAL or SAGA for computing parameters
- Computes 5 times more parameters than GDAL with the SAGA addition
- Decreases the computational cost incurred with the SAGA library

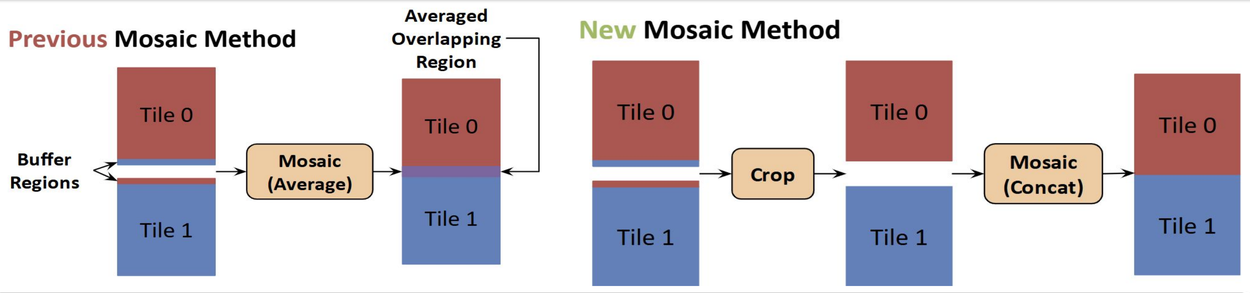
Optimization 1: Concurrent Cropping



Parallel Crop Workflow

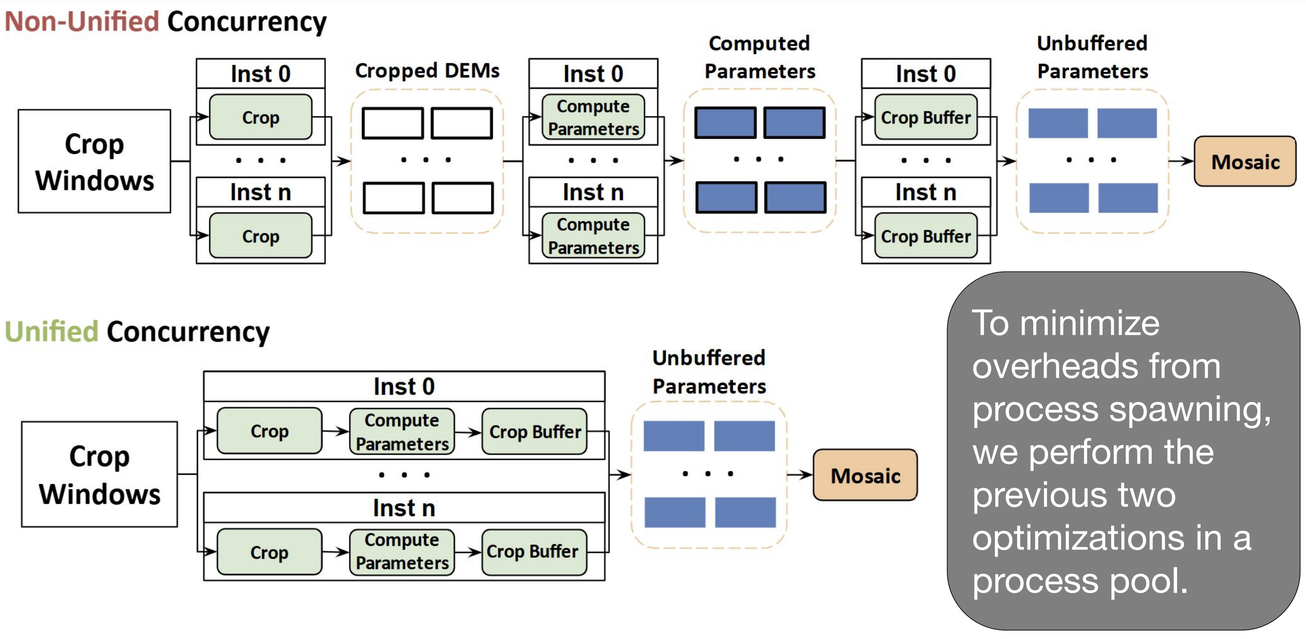
We parallelize the crop method by first computing each tile's bounds, i.e., windows. Tiles are then cropped concurrently from the input DEM for every window.

Optimization 2: Updated Mosaic Procedure



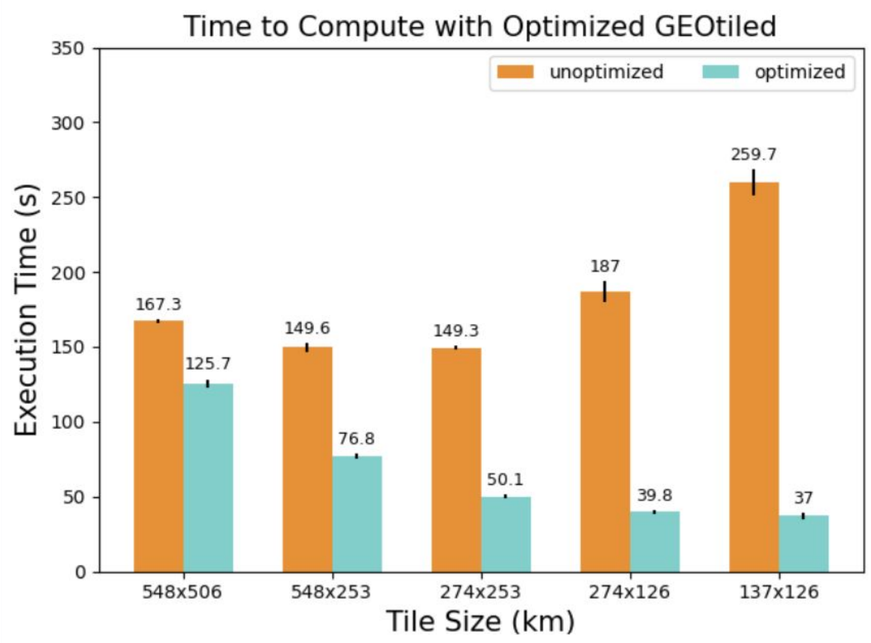
The previous mosaic process averages overlapping regions, but since these buffers have copied and not shared data, we crop the regions in parallel and concatenate the tiles together.

Optimization 3: Unified Parallel Processes



To minimize overheads from process spawning, we perform the previous two optimizations in a process pool.

Optimization Results



Experimental Setup

Tile Size (km)	Concurrent Processes
548 x 506	2
548 x 253	4
274 x 126	8
274 x 126	16
137 x 126	32

Parameter: Slope
Input Data Size: 2 GB

Our optimizations give GEOtiled 2.0 consistent **scaling performance** and up to **~7x speedup** compared to no optimization.

Future Work

- Implement SAGA's remaining terrain parameter algorithms into GEOtiled 2.0, giving it greater applicability to other Earth Science domains
- Optimize memory usage by mitigating the file I/O in the intermediary steps (i.e., Steps 3 and 4) of GEOtiled 2.0

References

- Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., Wehberg, J., Wichmann, V., and Bohner, J. (2015). System for automated geoscientific analyses(saga) v. 2.1.4. *Geoscientific Model Development*, 8(7):1991–2007.
- D. Rorabaugh, M. Guevara, R. Llamas, J. Kitson, R. Vargas and M. Taufer, "SOMOSPIE: A Modular SOil MOisture SPatial Inference Engine Based on Data-Driven Decisions," *2019 15th International Conference on eScience (eScience)*, San Diego, CA, USA, 2019, pp. 1-10, doi: 10.1109/eScience.2019.00008.
- Ketchum, D., Jencso, K., Maneta, M. P., Melton, F., Jones, M. O., & Huntington, J. (2020). IrrMapper: A Machine Learning Approach for High Resolution Mapping of Irrigated Agriculture Across the Western U.S. *Remote Sensing*, 12(14), 2328. https://doi.org/10.3390/rs12142328.
- Roa, C., Olaya, P., Llamas, R., Vargas, R., and Taufer, M. (2023). Geotiled: A scalable workflow for generating large datasets of high-resolution terrain parameters. *In Proceedings of the 32nd International Symposium on High-Performance Parallel and Distributed Computing*, HPDC '23, page 311–312, New York, NY, USA. Association for Computing Machinery.
- Santana Neto, Vicente & Soares, David & Camargos, Thais & Torres, Fillipe. (2023). Assessment of two methods on zoning wildfire propagation in Itacolomi State Park, Minas Gerais State, Brazil. *Pesquisa Florestal Brasileira*. 43. 1-12. 10.4336/2023.pfb.43e202102227.
- Warmerdam, F. (2008). The Geospatial Data Abstraction Library, pages 87–104. Springer Berlin Heidelberg, Berlin, Heidelberg.

GEOtiled software is available on GitHub:

[1] University of Tennessee, Knoxville, TN, USA
[2] University of Delaware, Newark, DE, USA

The authors acknowledge the support of the National Science Foundation (NSF) through awards #2103845, #2138811, #2331152, and #2334945