GCN10: A High-Resolution Global Curve Number Dataset and a Parallelized Framework For Large-Scale Raster Operations

The Computational Lab for Advanced Water Resources Informatics and Modeling

Abdullah Azzam[1] (abdazzam@nmsu.edu), Abdul Raheem Siddiqui[2], Huidae Cho[1]
[1]Department of Civil Engineering, New Mexico State University, Las Cruces, NM 88003, [2]ERT Inc.

BE BOLD. Shape the Future.®
College of Engineering
NM STATE

We present GCN10, a new 10-meter global Curve Number dataset developed through processing of land cover and soil data at planetary scale and a parallelized raster processing framework and its implementation for simulating global rainfall-runoff using SCS CN methodology, optimized and benchmarked across MPI, OpenMP, and hybrid approaches.

Introduction

- Reliable runoff estimation requires accurate hydrologic inputs.
- There's a gap: no **high-resolution, global, open-access** CN dataset.
- Curve Number (CN) method is widely used but global datasets are either too coarse (GCN250m) or geographically limited (CUSCN30)

Objectives

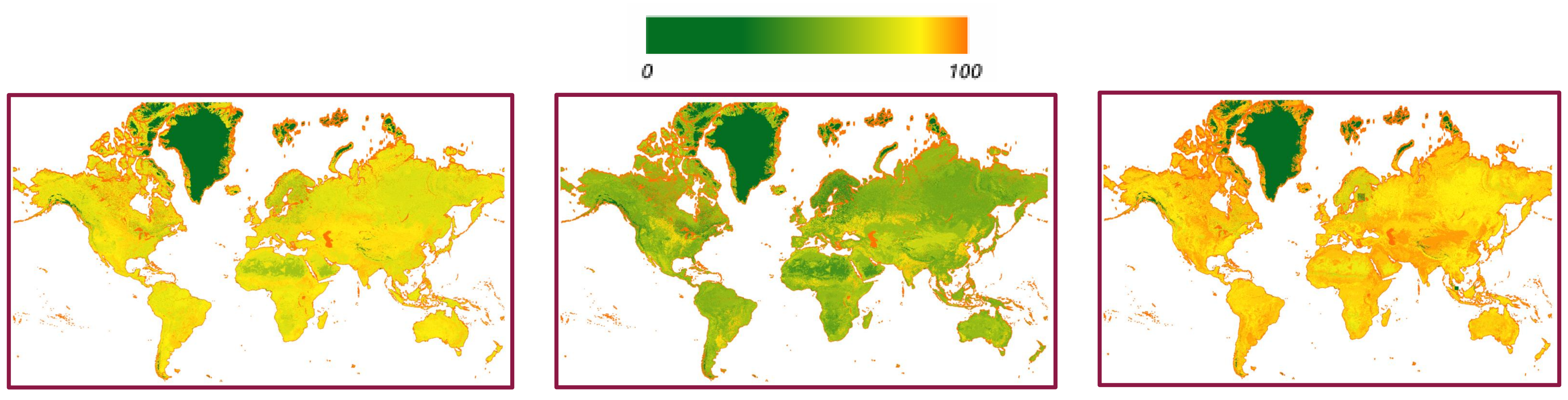
- **Create** a 10-meter resolution global CN dataset (GCN10).
- **Develop** a Parallelized, scalable raster processing framework for global runoff simulation.
- **Validate** by comparing $Runoff_{GCN10}$ with $Runoff_{GCN250}$.
- **Open Access** and community usability.

I. Global Curve Number GCN10

Methodology & Scale

ESA WorldCover 2021 + HYSOGs250m + Lookup Table = GCN10
Land Cover | **Hyd. Soil Group** | **Reference Values**

- *2651 blocks* of WorldCover (each ~ 117 GB) processed individually.
- Each block generated *18 rasters*—3 HCs, 3 ARCs, and Dual HSGs.
- 1/18 global raster contains *3.4 Trillion cells*.
- Raw processing would require ~ *12 TB RAM* if done at once.
- Final compressed outputs occupy ~ *3.4 TB*.

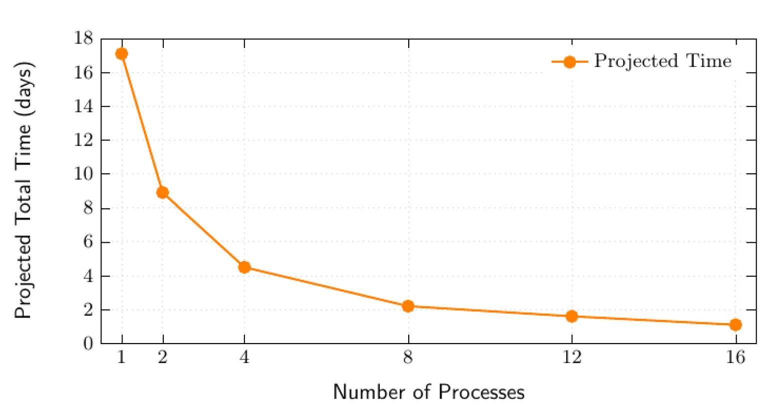


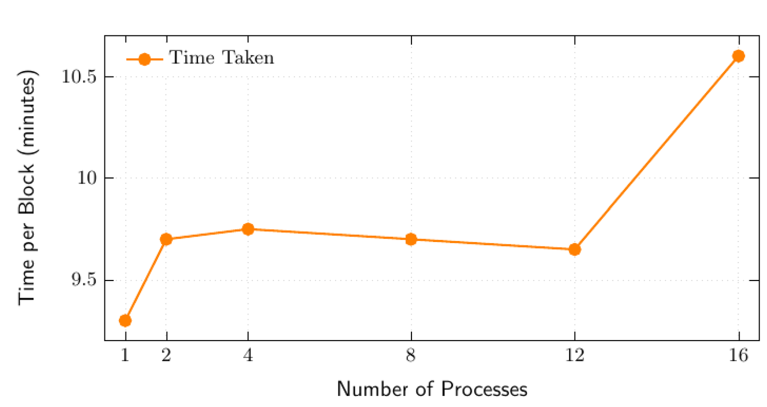
HC:Poor, ARC: I | *HC:Fair, ARC: III* | *HC:Good, ARC: III*

Parallel CN Generation

Python Multiprocessing – No GIL bottleneck, distributed tasks.

```
1. Read (ID, geometry) from blocks shapefile
2. Create multiprocessing.pool with `n' processes
3. For each block in pool:
    a. Extract block ID and geometry
    b. Clip and load landcover and soil rasters
    c. For condition in {drained, undrained}:
        For hydrologic state in {poor, fair, good}:
            For ARC in {I, II, III}:
                - Generate and save CN raster
4. Wait for children and parent to finish
5. Return "done"
```





OpenMP implementation

```
# Pragma omp parallel for
for (int i = 0; i < num_cells; i++) {
    cn_raster[i] = lookup_cn(landcover[i], soilgroup[i]);
}
```

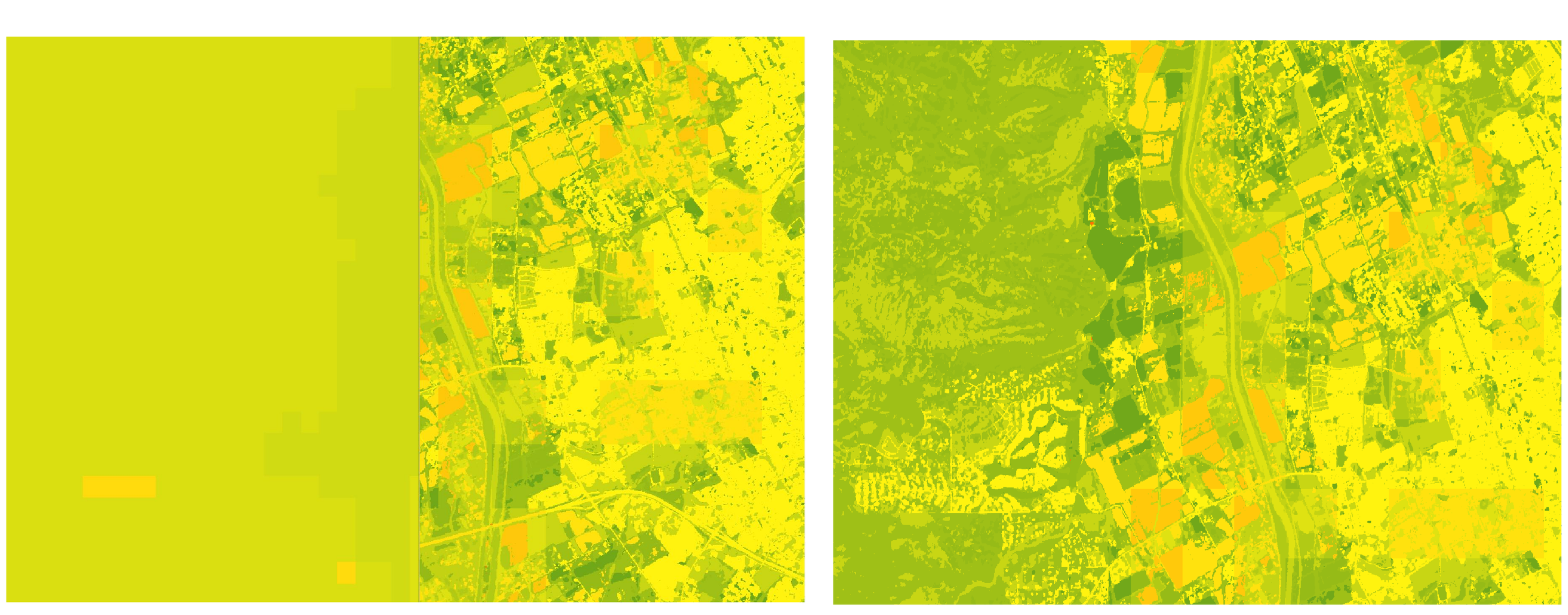
- OpenMP/Pool Speedup = 3.14

GCN10 vs GCN250

GCN10 offers a **25x finer resolution** (10 meters), allowing:

- Accurate representation of urban hydrology and impervious surfaces.
- Improved modeling for small watersheds and headwater basins.
- Better capture of fine-scale floodplain dynamics and flood risk zones.
- Enhanced support for infrastructure planning and disaster resilience studies.

This granularity is critical for Next-Gen hydrologic modeling at both local and global scales.



GCN250 compared to GCN10 for Las Cruces, NM

II. Global Parallelized Raster Processing Framework

Rainfall-Runoff Implementation

A global-scale raster processing framework was developed to simulate rainfall-runoff using the SCS Curve Number method on different Curve Number products (GCN10, GCN250).

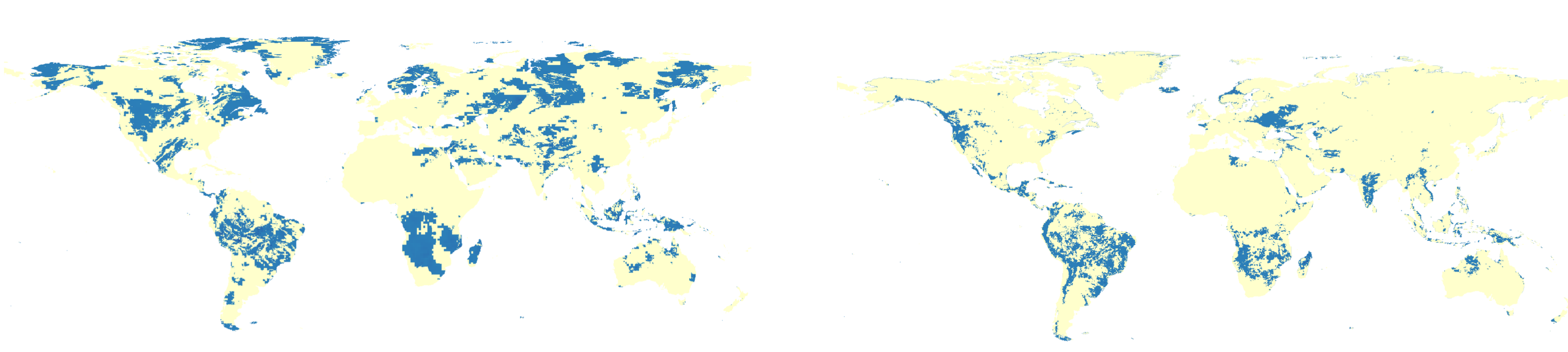
Parallelization Strategies:

- **Blockwise MPI:** The raster domain is partitioned into spatial tiles; each MPI process independently computes runoff for its assigned tile.
- **Intra-block OpenMP:** Within a single tile, pixel-level runoff computations are parallelized using OpenMP threads.
- **Hybrid MPI+OpenMP:** Combines blockwise distribution across MPI processes with multithreaded intra-tile processing.

Ongoing performance benchmarking compares MPI, OpenMP, and Hybrid approaches in terms of runtime, scalability, and resource efficiency.

GCN10 Validation

- Validation of the framework is being performed using GLDAS 3-hourly rainfall forcing to drive runoff simulations.
- Simulated runoff outputs are compared against GLDAS surface runoff products over matching time periods.
- The MPI-based tile-parallelized approach is currently used for validation, forking multiple processes to compute runoff across tiles simultaneously.
- Preliminary results indicate significant acceleration in global runoff computation without loss of geospatial integrity.



GLDAS 3-hr rainfall and runoff for 03:00 Jan 1, 2020

Next Steps

- Complete performance comparison to select the best-suited parallelization scheme.
- Expand the framework for rainfall-runoff simulations at global scale.
- Release the processed runoff products and framework scripts as open-access, GIS-agnostic tools for the broader hydrologic and geospatial communities.

Acknowledgement

I acknowledge the New Mexico Water Resources Research Institute (NM WRRI) and the New Mexico State Legislature for their support and resources, which were essential for this project. This work is funded by the NM WRRI and the New Mexico State Legislature under the account number NMWRRI-SG-FALL2024.

Software Development

The **QGIS Curve Number Generator Plugin** (Global ESA & ORNL Algorithm) automates the generation of GCN10 rasters for a vector AOI and is freely available for public use.

QGIS

Disclaimer

The views and conclusions are those of the authors and should not be interpreted as representing the opinions or policies of the NMWRRI.