H451

Domain and Formulation Scalability of the Next Generation National Water Model Framework

Donald Johnson, Robert Bartel, Nels J Frazier, Fred L. Ogden, Shengting Cui

National Water Center, National Weather Service, NOAA, ERT

OWP | OFFICE OF WATER PREDICTION

Background

.

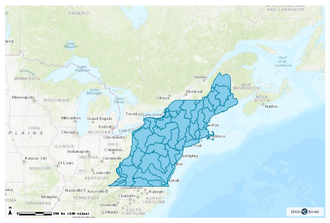
The Next Generation Framework (NGen) is a new modeling framework that will be the basis of the 4th Major Version of the National Water Model. Among its goals are supporting of modeling at scale ranging from simple basins to full continental simulations. It is intended to usable for both operational forecasting and academic work, and should be deployable without access to HPC system.

Datasets and Methods

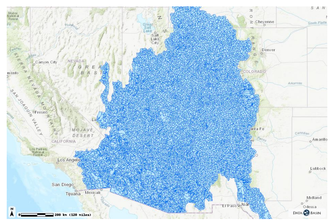
Data was collected by using the Linux time utility to extract measures of user and real time, for runs using different physics formulation in each of the target datasets.

Datasets

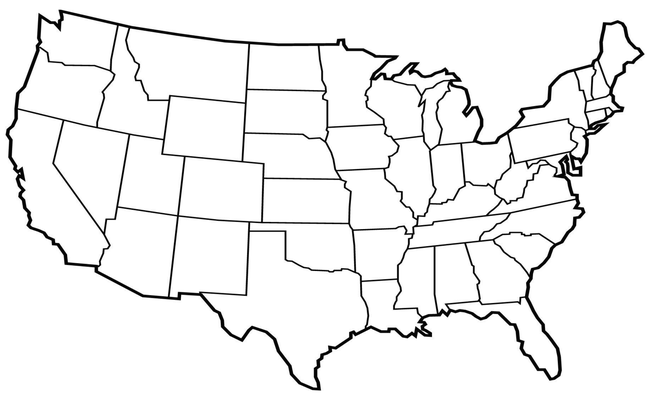
Timing data was acquired for 3 data sets



Huc 01, **20,544** Catchments

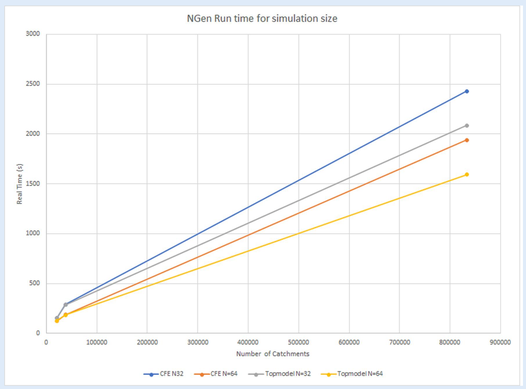


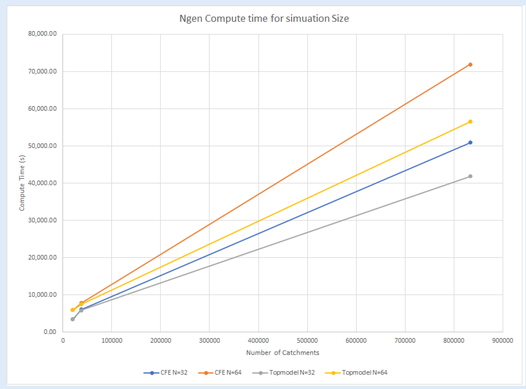
Huc 12, **37,714** Catchments



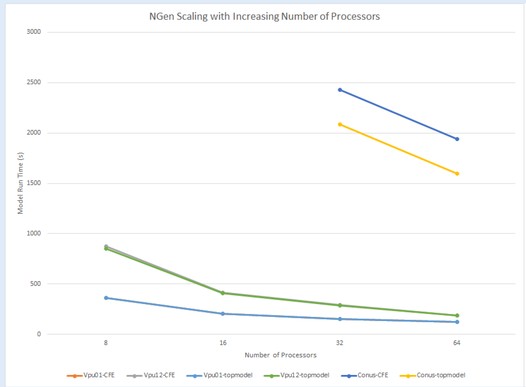
Conus, **833,074** Catchments

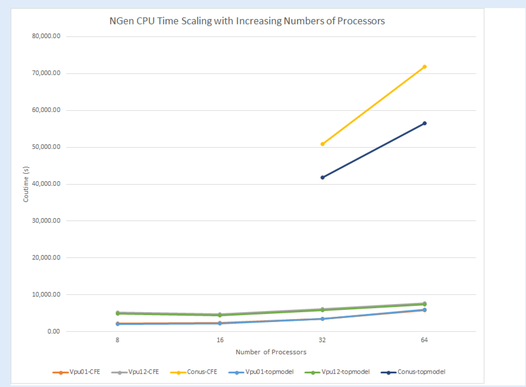
The NGen Framework is highly scalable system that allows simulations to be run, at local, regional, and continental scales, without requiring a HPC system.





Left: Runtime increases with number of catchments but the increase for CONUS is at a lower rate. Test with more processors are consistently faster showing positive scaling. **Right:** Compute cost increase with number of processors with more processor runs increasing faster. This shows the increased cost of using more processors.





Left: Runtime decreases as the number of processors involved increases. There is a clear difference in the time required by dataset but the tested physics had minimal effect for none CONUS runs. **Right**: CPU time is near constant with increasing processors for non-CONUS datasets. This shows that there was minimal increase in cost as the number of processors increases. The notable increase in CPU time for CONUS can be attributed to more communication in the larger dataset

Key Points

- The NGen Framework allows execution of user selected physics code in simulation regions CONUS. Larger simulations are possible but CONUS is the operational target for the National Water Center.
- At all tested scales increasing the number of processors used increased the performance of the system. Very small simulations will have limited or no improvement from parallel execution, when there are not enough regions that can be run simultaneously.
- The current scaling is good but not perfect, it is likely that performance can be further improved with improved load balancing.
- The fact that CPU Time increases more quickly for the CONUS dataset than for the smaller datasets is expected, communication cost are expected to increase with both datasets complexity and degree of attempted parallelism.
- This testing was done without use of a supercomputer and with the exception of CONUS the test domains can be run on laptops.

Future Work

- While the NGen framework has made great progress there is room for improvements. In particular load balancing is an issue with large datasets, additions of both profile guided optimization and dynamic load balancing will address this issue.
- When supporting either large datasets or long simulation times, further optimization of I/O systems is possible
- Support for physics formulations with internal parallelization is also an upcoming goal.

ACKNOWLEDGEMENTS:

My thanks to Mike Johnson for providing the dataset files necessary for this work .

REFERENCES:

Catchment Maps from databasin.org

CONTACT

Website: *https://water.noaa.gov*
Email: *Donald.W.Johnson@noaa.gov*

View my poster and other AGU materials