Petascale Hydrologic Modeling: Opportunities and Challenges

Fred L. Ogden
Craig C. Douglas
Scott N. Miller
Ye Zhang
CI-WATER Project

- NSF Cyberinfrastructure Cooperative Agreement joint between Utah and Wyoming EPSCoR jurisdictions.

- Focused on acquisition of hardware, development of software, capacity building, education, and outreach.
CI-WATER Project

- Enhance access to high-performance computing.
Project Objectives

1. Enhance cyberinfrastructure facilities
2. Enhance access to data- and computationally-intensive modeling
3. Advance high-resolution multi-physics watershed modeling
4. Promote STEM learning and water science engagement
Team

Univ. of Utah
- Data Storage
- UEN/Outreach
- Urban Hydrology
- Climate

UWyo
- Hydrologic Modeling
- Software Engineering
- HPC, Wyoming Outreach

BYU
- Geospatial data models
- Integrated modeling software

Utah State U.
- Hydrologic Modeling
- Hydrologic Information Systems
- Water Resources Decision Support

US Army Corps ERDC DoD HPC
- Research Applications Lab.

NWSC
- Team
Gridded Surface/Subsurface Hydrologic Analysis (GSSHA) model
**Fundamental Equations**

**Darcy’s Law**

\[
Q = -KA \frac{(h_1 - h_2)}{l},
\]

*unit: \(L^3/T\)

"The velocity of flow is proportional to the hydraulic gradient"

\[K \ldots \text{hydraulic conductivity, or effective permeability, coefficient of permeability, seepage coefficient (L/T)}\]

---

**Green's Theorem**

\[
\int_C (L \, dx + M \, dy) = \int \int_D \left( \frac{\partial M}{\partial x} - \frac{\partial L}{\partial y} \right) \, dA.
\]

---

**Richards' Equation**

\[
\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial z} \left[ K(\theta) \left( \frac{\partial \psi}{\partial z} + 1 \right) \right]
\]
GSSHA is a Giga-scale Engineering Hydrologic/Hydraulic Model

- Square Grid (1m – 100m grid size)
- Multi-solver (mixed explicit and implicit)
- Multi-physics (different fundamental PDE's)
- Finite volume, mass conservative
- 2D overland flow and groundwater flow
- 1D channel routing
- Richards eqn. or Green-Ampt with redistribution coupling between overland flow and groundwater
- Sediment/contaminant/nutrient transport
- Hydraulic structures, lakes, reservoirs, detention basins, wetland, storm and tile drains
- Evapotranspiration/root zone moisture accounting
- OpenMP (multi-core) and MPI (supercomputer)
Discussion starter: who among us uses HPC for hydrologic modeling?

- Many hydrologists do not even do Terascale!
- Single CPU Gigascale modeling common.

High Performance Computing is a new frontier.

To consider the Petascale in hydrology, think BIG.
A big watershed problem:

• Upper Colorado River Basin: 280,000 km$^2$
• High resolution important in mountains, where slope, aspect, vegetation, and wind, drive snow redistribution, sublimation, and melt.
• Low resolution in broad and extensive basins, where hydrologically interesting things seldom happen.
• Square grid model structure is very inefficient for large watersheds where process scales vary.
CI-WATER Component 3 Objective

Develop a high-resolution, large-scale hydrologic model to answer three questions:

• What are the potential impacts of climate change on the long-term yield of water from the upper Colorado River basin?

• How will future land-use changes due to development and natural causes such as fire or pine bark beetle affect water supplies?

• What are the effects of trans-basin diversions and increases in water consumptive use on the water storage in Lake Powell in 50 years?
Research Goals

• Increase accessibility of high performance computing to water resources researchers, engineers, and managers.

• Produce a set of modeling tools that allow consideration of future conditions in a modeling and probabilistic framework.

• Engage the wider community by releasing the code developed for research, development, and testing.
Law of the River, Colorado River Compact, 1922

Lees Ferry, AZ, is the legal dividing point between Upper and Lower Basin

Lower Basin (CA, AZ, NV) guaranteed 7.5 MAF/y

International: Mexico- 1.5 MAF/y

Note: 1 AF = 1.23 Ml
Glen Canyon Dam:
The Upper Basin States' bank account

- Pre-1963 average inflows 12,963,000 AF
- Post-1963 average inflows 10,701,000 AF
Upper Colorado River Basin

Basin Area: 288,000 km²

Streams: 467,000 km

Population: 900,000 (USBR)

Area above 2700 m: 14.5% (9,000 ft)

Area above 3050 m: 3.2% (10,000 ft) - this is where most of the water comes from.
High Altitude Complexity
Colorado River Basin is Highly Managed
Test Area: Wyoming Green River Basin

Darker blue areas are those above 2700 m elevation (9000 ft) where most snow melt occurs.
The NCAR-Wyoming Supercomputing Center (NWSC) provides dedicated Petascale capabilities for geosciences.
Wyoming’s 20% Share of NWSC's 72,300 cores represents a huge increase in EPSCoR HPC capabilities…

- On the latest (6/11) Top500 list of fastest supercomputers, Wyoming’s share on NWSC-1 alone is estimated to be…
  - The 28th fastest computer in the world
  - The 14th largest supercomputer in the US
  - The largest system in an EPSCoR state outside of Department of Energy facilities

Reference: http://www.top500.org
HPC Data Issues

• Data assimilation (BIG DATA Problem to solve!)
  ➔ How do we acquire enough data to keep a Petascale computer busy? Just inventing data through interpolation is not acceptable.
  ➔ Future: large numbers of inexpensive networked sensors.
  ➔ We need a massive number of remote, on ground sensors, not just a massive quantity of data from a relatively few sensors.
  ➔ We need a symbiotic relationship between smart sensors and computational models, e.g., a dynamic data-driven application system, so that we get the right amount of data for the right scales while computing.
**HPC Numerical Algorithms**

- **Multiscale methods**
  - We use a base resolution with an average or median mesh size.
  - We can *upscale* to compute on a coarser mesh much quicker than on the base mesh.
  - We can *downscale* to compute on a finer mesh in a subregion of the entire domain to pick up features that are not visible on the base mesh. If the subregion is small enough, this is both computationally feasible and scientifically useful.

- **Load balancing**
  - This is a preprocessing step in the major computations.
  - First generate a catalog of base meshes and store them.
  - Generate a series of domain decompositions for different representative numbers of cores and store them.
  - Use weather input data such as temperature and precipitation to estimate computational demands.
We are not starting from scratch 
(thanks to our collaborators)

• USACE-ERDC providing:
  • Finite element computational kernel derived from ADH model
  • Computational model builder (CMB)
  • ezVIZ HPC visualization tools
  • ezHPC user interface toolkit
Conclusions

- The CI-WATER project is joint cooperative agreement between NSF and the Utah and Wyoming NSF EPSCoR jurisdictions.

- The project aims to increase accessibility of HPC resources for water resources research and management in the Upper Colorado River Basin.

- CI-WATER has components related to data, cloud computing, accessibility, education, outreach, and diversity.
Thank you
Computational Model Builder

- Designed for large complex domains and HPC
- No licensing fees
- Cross platform
- User-configurable
- Built as several complimentary, independent tools
Computational Model Builder Data Flow
User Interface Toolkit
Application Program Interface (API)

• Supports Novice to Expert Users
• Central Access to HPC Resources
• Custom Productivity Clients
  • Complete Job Stream Management
  • Fast Large File Transfers
  • Secure Authentication
**User Interface Toolkit – ezHPC**

**Tabbed Functions**
- MOTD and system news @ HOME Tab
- Monitor Jobs & Queue Status on all machines
- Job Management
  - Script generator & editor
  - Allocation and Utilization viewer
- Fast large file transfers
- Easy access to custom scripts

**Monitor Kerberos Ticket Session Time**

**Easy Access to on-line documentation**
Establishing a Petascale Collaboratory for the Geosciences: Scientific Frontiers

• “A PCG will enable the simulation of the full spectrum of interactions among physical, chemical, and biological processes in coupled Earth system models.
• Land-atmosphere property fluxes are forced by surface ecosystem heterogeneity on scales of 1 m or less. The forcing is the result of a huge array of interacting biological, chemical, and geological processes
• Understanding the integrated effects of these processes is necessary for predicting ecosystem change and water availability.”

A Report to the Geosciences Community.
UCAR/JOSS. 80 pp., 2005
Conclusions

• To get to Petascale, we need the following:
  • New ways of thinking about the hydrology model.
  • More complex and massive data collection through remote, intelligent sensors.
  • Multi-physics hydrologic process models at very fine scales.
  • Much more complex numerical algorithms in both time and space that are stable and new for O(100K) cores.
  • Much more complex software developed for O(1K) to O(10K) cores that is being extended to O(100K) to O(10M) cores.
  • By the time we are done getting to Petascale, Exascale will be threatening us.
  • This is not your daddy’s hydrology or computing model anymore.

• Think BIG??? Maybe EXTREMELY MASSIVE is more appropriate.