Objective 3: Develop High-Resolution Large Watershed Hydrologic Model

Fred L. Ogden, Robert C. Steinke, Wencong Lai, Craig C. Douglas, Scott N. Miller, Ye Zhang

In cooperation with:
Petascale??

- HPC hydrologic modeling is in its infancy.
- We seldom do terascale modeling!
- We often do single CPU gigascale modeling.
- High Performance Computing is a new frontier for watershed modeling.

To consider the petascale in hydrology, one must think BIG.
Our Collaborators

- U.S. Army Corps of Engineers, Engineer Research and Development Center, Vicksburg, Mississippi, Coastal & Hydraulics and Information Technology Laboratories

- National Center for Atmospheric Research, Research Applications Laboratory
Gridded Surface/Subsurface Hydrologic Analysis (GSSHA) model
GSSHA

- Square grid (5 to 90 m typical)
- Multi-solver: different approximations of full PDE's, finite-difference and finite-volume.
- Multi-physics: different PDE's, or hybrid equations (mixed overland/groundwater)
- 2D overland flow and groundwater flow
- 1D channel routing with hydraulic structures
- Richards or Green-Ampt Redistribution coupling between overland flow and groundwater
- Sediment/contaminant/nutrient transport
A big watershed problem:

- Upper Colorado River Basin: 280,000 km²
- High resolution important in mountains, where slope, aspect, vegetation, wind, drive snow redistribution, sublimation, and melt.
- Low resolution in broad and extensive basins, where runoff is infrequently produced.
- **Square grid model structure is very inefficient for large watersheds where process scales vary.**
Colorado River Basin
**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, 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 30-50 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.
Data Needs:

**Fire and land use changes:**

**Snowfall and redistribution:**

---

**Table 1. State Population Growth as Dramatic as Municipal Growth**

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Colorado</td>
<td>539,700</td>
<td>1,325,089</td>
<td>3,294,394</td>
<td>4,301,261</td>
<td>4,864,155</td>
</tr>
<tr>
<td>Arizona</td>
<td>122,931</td>
<td>749,587</td>
<td>3,665,228</td>
<td>5,130,632</td>
<td>6,338,755</td>
</tr>
<tr>
<td>California</td>
<td>1,485,053</td>
<td>10,586,223</td>
<td>29,760,021</td>
<td>33,871,648</td>
<td>36,553,217</td>
</tr>
<tr>
<td>Utah</td>
<td>276,749</td>
<td>688,862</td>
<td>1,722,850</td>
<td>2,233,189</td>
<td>2,645,330</td>
</tr>
<tr>
<td>Nevada</td>
<td>42,335</td>
<td>160,083</td>
<td>1,201,833</td>
<td>1,998,257</td>
<td>2,565,382</td>
</tr>
<tr>
<td>New Mexico</td>
<td>195,310</td>
<td>681,187</td>
<td>1,515,069</td>
<td>1,819,046</td>
<td>2,499,481</td>
</tr>
<tr>
<td>Wyoming</td>
<td>92,551</td>
<td>290,529</td>
<td>453,588</td>
<td>493,782</td>
<td>532,668</td>
</tr>
</tbody>
</table>

**Source:** U.S. Census Bureau.
**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)
High Altitude Complexity
Test Area: Green River Basin in Wyoming

Darker blue areas are those above 2700 m elevation (9000 ft) where most snow melt occurs.
We are not starting from scratch (thanks to our collaborators)

- USACE-ERDC providing:
  - finite element computational kernel derived from the ADaptive Hydraulics (ADH) model
  - Computational model builder (CMB)
  - ezVIZ HPC visualization tools
  - ezHPC user interface toolkit
Computational Model Builder

- Designed for large complex domains & HPC
- No licensing fees
- Cross platform
- User-configurable
- Built as several complimentary, independent tools
Computational Model Builder
Work Flow
**Interrupted Sinusoidal Projection**

- Preserves area perfectly
- Lines of latitude are horizontal lines
- Longitudes converge towards the pole
- Can describe Amazon basin with minimal distortion
Upper Colorado River Stream Network

- National Hydrography Data Set

- Use geomorphological cross-section predictors

- Almost 500,000 km of streams

- River data set impossible to create manually
TauDEM vs. NHDPlus
TauDEM Accuracy Assessment

Compared TauDEM generated stream network to National Hydrography Data Set (NHD)
**TauDEM vs. NHDPlus**

Picked TauDEM threshold to match stream density of NHDPlus.

Green and blue lines show where there is no match within 100 meters.
TauDEM vs.
National Hydrography Dataset

- Qualitatively, TauDEM performs well
- We developed a quantitative algorithm to find points in one set of line segments far away from the closest point in another set of line segments
- We have submitted a paper describing the algorithm, which will become part of TauDEM
High Resolution Large Watershed Model

- Surface layer
- Interflow layer
- Groundwater layer
- Infiltration
- Percolation
- Leakage to deep groundwater
Mathematical model

surface water:
2D shallow water equations
  - dynamic wave
  - diffusive wave
  - kinematic wave

subsurface water:
3D Richards’ equation

1D vadose zone flow
2D saturated groundwater flow
Mathematical model

2D dynamic wave:
\[ \frac{\partial h}{\partial t} + \frac{\partial hu}{\partial x} + \frac{\partial hv}{\partial y} = 0 \]
(hyperbolic convective)
\[ \frac{\partial hu}{\partial t} + \frac{\partial huu}{\partial x} + \frac{\partial huv}{\partial y} = -gh \frac{\partial z}{\partial x} - \frac{gn_x^2 u \sqrt{u^2 + v^2}}{h^{1/3}} \]
\[ \frac{\partial hv}{\partial t} + \frac{\partial huv}{\partial x} + \frac{\partial hvv}{\partial y} = -gh \frac{\partial z}{\partial y} - \frac{gn_y^2 v \sqrt{u^2 + v^2}}{h^{1/3}} \]

1D vadose zone flow
(ODE)
\[ \frac{dZ}{dt} = \frac{1}{(\theta_o - \theta_i)} \left( \frac{K_s H_c}{Z} + K_s \right) \]

2D groundwater flow
(parabolic diffusive)
\[ S_y \frac{\partial H}{\partial t} = \frac{\partial}{\partial x} \left( K_x (H - z_b) \frac{\partial H}{\partial x} \right) + \frac{\partial}{\partial y} \left( K_y (H - z_b) \frac{\partial H}{\partial y} \right) + R \]
Numerical model

2D unstructured finite volume method for overland flow and saturated groundwater flow

\[
\frac{\partial U}{\partial t} + \nabla \cdot F = S
\]

\[
\int \frac{\partial U}{\partial t} d\Omega + \int \nabla \cdot F d\Omega = \int S d\Omega
\]

\[
\int \frac{\partial U}{\partial t} d\Omega + \oint F \cdot n d\Gamma = \int S d\Omega
\]

\[
\frac{U_{i}^{n+1} - U_{i}^{n}}{\Delta t} + \frac{1}{\Omega_i} \sum_{j=1}^{3} F_{ij} n_{ij} \Delta \Gamma_{ij} = S_{i}
\]

Upwind Riemann solver for convective flux in overland flow

Central difference for diffusion term in groundwater equation
1-D Unsaturated Flow model: T-O (Talbot and Ogden, 2008) infiltration and redistribution method

Infiltration:

\[
\frac{dZ_k}{dt} = \frac{K(\theta_d)}{(\theta_d - \theta_i)} \left( \frac{\Psi(\theta_d)}{Z_k} + 1 \right)
\]

Redistribution:

\[
Z'_k = V \left( \frac{\Psi(\theta_k)}{\sum_j \Psi(\theta_j)} \right) + Z_k
\]
Talbot and Ogden 1-D Infiltration (2008)

- Allows simulation of near surface ground water table without numerical solution of Richards (1931) equation:

\[
\frac{\partial \theta}{\partial t} = \frac{\partial}{\partial z} \left( K(\theta) \frac{\partial \psi(\theta)}{\partial z} + 1 \right)
\]

with:

\[
\Theta = \theta_r + \frac{\theta_e - \theta_r}{\left( 1 + \left( \frac{\alpha \psi}{\rho_w g} \right)^n \right)^m}
\]

\[
K(\Theta) = \Theta^{1/2} \left( 1 - (1 - \Theta^{1/m})^m \right)
\]

\[
m = 1 - 1/n \quad \Theta = \frac{\theta - \theta_r}{\theta_e - \theta_r}
\]
Numerical results: Hypothetical sloping plane (Sulis et al., 2010)

400 m X 320 M sloping plane;

5 m deep soil with water table 1 m below;

Uniform rainfall intensity of 33 mm/min for 200 minutes;
Numerical results: Hypothetical sloping plane (Sulis et al., 2010)
Published laboratory results: sloping plane (Smith and Woolhisser, 1971)

1.22 m deep soil with 3 layers of fine sand with different porosities

Uniform rainfall intensity of 250 mm/hr for 15 minutes;
Numerical results: Laboratory sloping plane (Smith and Woolhisser, 1971)
Numerical results: Laboratory sloping plane (Smith and Woolhisser, 1971)
Objectives for next 12 months

- Incorporate lakes, reservoirs, bathymetry in channel network
- Develop comparison simulations using ADH, PARFLOW, GSSHA
- Add evapotranspiration to ADHydro
- Add channel routing to ADHydro
- Incorporate Utah Energy Balance snowmelt model in ADHydro
- Run ADHydro in Green River headwaters catchment (May, 2013)
- Communicate data needs/input structures/work flow to Utah
- Add needed solvers to ADH parallel code, and set up partitioner
- Release code and establish user community, December, 2013.
- Collaborate with USBR and upper Basin water managers in developing reservoir simulation model.
- Incorporate irrigated areas, begin developing irrigation simulator, early 2014
Thank you
“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.”
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 = 1.233 MI
Glen Canyon Dam: The Upper States' bank account

- Pre-1963 average 12,963,000 AF
- Post-1963 average 10,701,000 AF
The NCAR-Wyoming Supercomputing Center (NWSC) provides dedicated *petascale* capabilities for geosciences.

For more information visit, www.nwsc.ucar.edu

NWSC Partners:

[Logos of UNR University of Wyoming, Wyoming Business Council, NSF, NCAR, Cheyenne LEADS, UCAR, Cheyenne Light Fuel & Power]

Architects, Contractors and Consultants:

H+L Architecture | Saunders Construction, Inc. | California Data Center Design Group | Rumsey Engineers | RMH Group
Martin & Martin Consulting Engineers | Rider Levett Bucknall | Reliable Resources | E Cube, Inc.
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
  - The largest resource controlled by a university in the US

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

- **Data assimilation**
  - How do we collect enough data to keep a Petascale computer busy? Just inventing data through interpolation is not acceptable.
  - We need a tsunami of data from inexpensive sensors or high-resolution simulations.
  - Satellite images 1-2 times per day in composite (incomplete) JPEG files. This is not necessarily high enough resolution and cloud cover is a problem.
  - *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.
  - Finally, how do we afford massive data collection?
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.
  • Dynamic steering of a computation is essential to make this work and can be done as postprocessing.

• Load balancing
  • This is a preprocessing step in the major computations.
  • First generate base meshes of interest and store them.
  • Generate a series of domain decompositions for different representative numbers of cores and store them.
  • Similar to the ocean modeling community meshes.
HPC Time Stepping

Implicit methods
- Implicit time stepping allows larger time steps while maintaining stability.
- With massively parallel computers, an implicit method requires using massively parallel solvers from one time step to the next, while many common algorithms today just do not scale to O(100K) cores, unfortunately.

Explicit methods
- Time steps usually limited by stability conditions to $\Delta t < C(\Delta x)^2$, where $C$ is a positive real constant.
- A new set of algorithms has recently been developed that are stable on given time steps, but use intermediate time steps (where stability may be violated) so that the stability condition is $\Delta t < C\Delta x$ instead (different $C$). Hence, vastly larger time steps are possible.
- Massively parallel computations are straightforward with explicit methods.
HPC Time Stepping

- **Hybrid explicit-implicit methods**
  - On the boundaries of the subregions use an explicit method to approximate the solution on the next time step.
  - Use an implicit method in each subregion, where the size of the subregions is small enough so that the algorithm used to get to the next time step scales well.
  - Possibly iterate on the boundary points to improve accuracy.

- **Hybrid implicit-explicit methods**
  - Downscale the problem to only the boundaries of the subregions and use an implicit method to approximate the solution on the next time step. This can be done in parallel based on subregions.
  - Use an explicit method in each subregion.
  - Possibly iterate on the boundary points to improve accuracy.

- **Implications for Petascale computing**
  - Both hybrid methods should scale and be fast.
  - Need to analyze which hybrid method works best for CI-WATER.