Designing a Dynamic Data Driven Application System for Coastal & Environmental Modeling

Gabrielle Allen
Center for Computation & Technology
Dept Computer Science
Louisiana State University
gallen@cct.lsu.edu
http://www.cct.lsu.edu/~gallen

Louisiana Coastal Area

- 1927 flood, levees, loss of wetlands, growing crisis, social impact
  - 25% lost wetlands in last century, future predictions dire; increases flooding, risk of surge
- Pam (2004), Katrina, Rita (2005)
- Important problems: ecological, hurricane, algal bloom/salinity forecasting, restoration, evacuation, emergency response strategies
- Rich dynamic environment for modeling: coupled models, multi-scale, realtime data (sensors, satellites)
- Role of HPC, Models, Grids, Community
<table>
<thead>
<tr>
<th>Hurricane KATRINA</th>
<th>Hurricane RITA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date : 23-30 Aug, 2005</td>
<td>Date : 18-26 Sep, 2005</td>
</tr>
<tr>
<td>Category 3 landfall (Peak Winds : 145 mph) on 29th Aug , 6:10 am CDT, near Buras LA</td>
<td>Category 3 landfall (Peak Winds : 120 mph ) on 24th Sep, 2:40 am CDT, Texas Louisiana Border</td>
</tr>
<tr>
<td>Voluntary evacuation New Orleans : 37 hours before landfall. Mandatory evacuation : 19 hours before landfall</td>
<td>Mandatory evacuation Galveston : 19 hours before landfall</td>
</tr>
<tr>
<td>Human Casualties : 1836 approx</td>
<td>Property Damage : 35 Billion</td>
</tr>
<tr>
<td>Property Damage : 120 Billion, New Orleans population reduced by 50% from before Katrina</td>
<td>10% population displaced from Houston and Galveston</td>
</tr>
<tr>
<td>Storm Size (width across) at landfall - 460 miles</td>
<td>Storm Size (width across) at landfall - 410 miles</td>
</tr>
<tr>
<td>Radius of Hurricane Force winds at landfall - 125 miles</td>
<td>Radius of Hurricane Force winds at landfall - 85 miles</td>
</tr>
<tr>
<td>Coastal Storm surge - 18 -22 Feet</td>
<td>Coastal Storm surge - 15 -20 Feet</td>
</tr>
<tr>
<td>Third most powerful Hurricane to hit U.S Coast, Most Expensive</td>
<td>---</td>
</tr>
</tbody>
</table>

**Image:** Katrina: 29th August 2005, Image: MODIS Rapid Response Gallery
Beyond Katrina: Land Loss

Critical contributor to problem

North Breton Island, La. looking Northwest.
South and East of Mouth of Mississippi River

After Katrina
Models, Data, Grids for …

• Hurricane forecasts
• Emergency preparedness
• Wetland reconstruction
• Ecological studies and fish populations
• Oilspill behaviour
• Levee design
• Rescue
• Hypoxia “Dead Zone”
• Algae blooms
• 24/7/365 shipping forecasts

Hurricane Tracks
Model-Model-Data Coupling

Varied Data Sources

- Simulation data (forecast, nowcast, hindcast)
  - 2D, 3D
- Sensor data (time series)
- Remote imaging
- Aerial photography
- LIDAR
- Weather satellites
- GIS
Data Products

WAVCIS: Wave-Current-Surge Information System for Coastal Louisiana

- Program Director: Greg Stone, CSI, LSU
- Provide wave information (sea state) including wave height, period, direction of propagation, water level, surge, near surface current speed and direction and meteorological conditions on a real time basis around the entire Louisiana coast.
SURA Coastal Ocean Observing Program (SCOOP)

- Integrating data from regional observing systems for realtime coastal forecasts in SE
- Coastal modelers working closely with computer scientists to couple models, provide data solutions, deploy ensembles of models on the Grid, assemble realtime results with GIS technologies.
SCOOP Prototype Lab

External feeds
GFOL
COAMPS
NAM
others

All data for archive
SCOOP net

Translated products
Windgen

In situ data
MM5
SWAN
WAM

Analysis, storage, cataloging, visualization of output

Result Dissemination

Archive
Verfication
Visualization

Ensemble of models run across distributed resources

Ensemble wind fields from varied and distributed sources

Synthetic Wind Ensembles

Regional Archives

Wind Forcing

NCEP
MM5
NCAR

Select region and time range

Transform and transport data

Wave and/or Surge Models

ADDCirc
EICirc
WAM
SWAN

Center for Computation & Technology at Louisiana State University
Hurricane Forecasts

- Starting 5 days from expected landfall, Hurricane advisory from NHC provides best guess of track and intensity, along with other tracks
  - RSS & Email for advisory, ftp site for tracks
- Surge/Wave models prepared (spin-up initial conditions with hindcasts or use existing data from regular forecasts
  - Data location, transport, translation, maybe model deployment
- Locate wind forcing either from gridded products (MM5, GFDL, …) or calculate “analytic” wind fields from tracks
  - Data location
- Run ensemble of models
  - Data transport, translation, resource brokering, application deployment
- Archive model results
  - Data archive, transport
- Calculate products, e.g. MOM, MOF, MEOW and compare with observations
- Disseminate results
  - OpenGIS, visualization, notification

Current Models

- Variety of meshes and included properties.
- Models are suited for different conditions: e.g. shallow water, deep ocean.
- Basic variables are water level and current, and wave height, frequency and direction.
- All require forcing (wind, tides), boundary conditions, initial conditions (“hotstart” or “spin-up”)
- Currently porting, benchmarking, optimizing, etc.
## SCOOP Models

<table>
<thead>
<tr>
<th>Model</th>
<th>Description</th>
<th>Domains</th>
<th>Mesh</th>
</tr>
</thead>
<tbody>
<tr>
<td>ADCIRC</td>
<td>2D Surge and current</td>
<td>Entire SE</td>
<td></td>
</tr>
<tr>
<td>CH3D</td>
<td>3D velocities, temperature, salinity, surge</td>
<td>Western Florida</td>
<td></td>
</tr>
<tr>
<td>EICIRC</td>
<td>3D Baroclinic circulation across rivers to oceans</td>
<td>Chesapeake Bay</td>
<td></td>
</tr>
<tr>
<td>WW3</td>
<td>Water depth &amp; wave propagation</td>
<td>Entire SE, parts of Gulf of Mexico</td>
<td></td>
</tr>
<tr>
<td>WAM</td>
<td>Water depth &amp; wave propagation</td>
<td>Gulf of Mexico</td>
<td></td>
</tr>
<tr>
<td>SWAN</td>
<td>Water depth &amp; wave propagation</td>
<td>Coastal Louisiana</td>
<td></td>
</tr>
</tbody>
</table>

## Restart Files

[Diagram showing restart files process]

CENTER FOR COMPUTATION & TECHNOLOGY AT LOUISIANA STATE UNIVERSITY
3SCOOP Surge

Dr. Brian Blanton -- UNC

2SCOOP Surge

Dr. Brian Blanton -- UNC
SCOOPO Surge

Dr. Brian Blanton -- UNC

DynaCode (DDDAS)
- Hurricane ensemble modeling
  - Coupling ocean circulation, storm surge, wave generation models for the Gulf
  - Notifications from NHC trigger customized ensemble hurricane models (surge/wind/wave), sensors verify, guide dynamic ensembles
  - Event driven, dynamic component framework with algorithm selection, optimization tools, workflow, data assimilation, result validation with sensor/satellite.
- Ecological restoration and control
  - Breton Sound diversion, control structure to allow Mississippi to flow into wetlands
  - Coupled models (hydrodynamic, salinity, geomorphic, sediment) control diversion, sensors/wind fields inject real time data.
Requirements

- Data translation services
- Metadata services to
  - Describe models
  - Describe data
- Provenance information
  - Model versions and configurations
  - Data history
  - Hindcasts, nowcasts, forecasts
- Visualization & Notification
  - For different purposes, PR (e.g. evacuation), scientific insight, levee design
  - Realtime
- Validation and verification

- Workflow & model-model coupling
  - Legacy codes vs. framework
- Scheduling
  - Dynamic prioritization based on scenario and resources
  - On-demand
  - Co-scheduling and advanced reservation
- Monitoring of applications
- Reliability
- Data archive
- Clients usable by science community

DDDAS new capabilities

- dynamically invoke more accurate models and algorithms as hurricane approaches coast,
- choose appropriate computing resources for needed confidence levels
- compare model results with observations to feedback into running simulations
- realtime data assimilation
- adaptive multi-scale simulations
- dynamic component recomposition
- simulation needs steer sensors and data input

- Provide lots of requirements for numerical libraries, HPC scheduling, policies etc.
System Components

- General task farming infrastructure to support dynamic ensembles (starting)
- Grid Application Toolkit for abstract access to grid capabilities
- SCOOP coastal data archive and clients
- Application monitoring (starting)
- SPRUCE for priority event-driven computing on compute resources
- GridSphere based portal
Grid Application Toolkit (GAT)

- Abstract programming interface between applications and Grid services
- Designed for applications (move file, run remote task, migrate, write to remote file)
- Led to GGF Simple API for Grid Applications (SAGA)

<table>
<thead>
<tr>
<th>Default Adapters</th>
<th>Basic Functionality, will work on single isolated machine (e.g. cp, fork/exec)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Globus Adapters</td>
<td>Core Grid functionality: GRAM, MDS, GT-Relay, GridFTP</td>
</tr>
<tr>
<td>GridLab Adapters</td>
<td>GRMS, Mercury, Delphi, iGird</td>
</tr>
<tr>
<td>Under Develop</td>
<td>Ssh, DRMAA, Condor, SGE, SRB, Curl, RPT</td>
</tr>
</tbody>
</table>

Main result from GridLab project

www.gridlab.org/GAT

Grid Application Toolkit

- Standard API and Toolkit for developing portable Grid applications independently of the underlying Grid infrastructure and available services
- Implements the GAT-API
  - Used by applications (different languages)
- GAT Adaptors
  - Connect to capabilities/services
  - Implement well defined CPI (mirrors GAT-API)
  - Interchangeable adaptors can be loaded/switched at runtime
- GAT Engine
  - Provides the function bindings for the GAT-API
GAT API Subsystems

<table>
<thead>
<tr>
<th>File Subsystem</th>
<th>Monitoring and Event Subsystem</th>
<th>Information Exchange Subsystem</th>
</tr>
</thead>
<tbody>
<tr>
<td>GATFile</td>
<td>GATRequestListener</td>
<td>GATAdvertiseable</td>
</tr>
<tr>
<td>GATEndpoint</td>
<td>GATRequestNotifier</td>
<td>GATAuditService</td>
</tr>
<tr>
<td>GATPipeListner</td>
<td>GATMetric</td>
<td></td>
</tr>
<tr>
<td>GATPipe</td>
<td>GATAction</td>
<td></td>
</tr>
<tr>
<td>GATLogialFile</td>
<td>GATMetricEvent</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Resource Management Subsystem</th>
</tr>
</thead>
<tbody>
<tr>
<td>GATSoftwareDescription</td>
</tr>
<tr>
<td>GATJobDescription</td>
</tr>
<tr>
<td>GATJob</td>
</tr>
<tr>
<td>GATResourceDescription</td>
</tr>
<tr>
<td>GATResourceBroker</td>
</tr>
<tr>
<td>GATResource</td>
</tr>
<tr>
<td>GATReservation</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Utility Subsystem</th>
</tr>
</thead>
<tbody>
<tr>
<td>GATSelf</td>
</tr>
<tr>
<td>GATStatus</td>
</tr>
<tr>
<td>GATContext</td>
</tr>
<tr>
<td>GATPreferences</td>
</tr>
<tr>
<td>GATSecurityContext</td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

GAT++ = SAGA

- Global Grid Forum Research and Working Groups to develop a “Simple API for Grid Applications” (growing community support)
  - Focus on scientific and engineering applications
  - Focus on simplicity
- Similar to GAT
  - Better thought out API (based on use cases)
  - Asynchronous calls, bulk operations, QoS
  - More functional areas (e.g. streaming)
- Implementations under development now.
SAGA (GAT) & OGSA?

• Complementary and not competitive approaches
  – OGSA defines interfaces at the service and middleware level, SAGA at the application level
  – OGSA is primarily an architecture, SAGA is an API
  – SAGA implementations interface OGSA-compliant services (but not limited to)
  – OGSA is for middleware developers, SAGA is for application developers

SCOOP Portal
Louisiana Optical Network (LONI)

State initiative ($40M) to support research:
- 40 Gbps optical network
- Connects 7 sites
- Grid resources (IBM P5) at sites LIGO/CAMD

New possibilities:
- Dynamical provisioning and scheduling of network bandwidth
- Network dependent scenarios
- "EnLIGHTened Computing (NSF)"

On-Demand Computing

- For the hurricane scenarios need to be able to rapidly secure and use large amounts of resources for a short time
- Needs improved policies as well as technologies (e.g. SPRUCE)
  - Preemptive queues and priorities
  - Authorization
  - Ability to change reservations
- Checkpoint running jobs to make room
  - System level (e.g. IBM AIX level, LPAR, Meiosys)
  - Application level (e.g. Cactus)
- Reschedule preempted jobs at other sides
  - Migration
  - Data transport/shared file system (e.g. GPFS)
The checkpoint queues allow using the whole machine while the non-checkpoint queues are limited to the number excluding the sum of nodes assigned to on-demand queues.

High priority spruce queues can be used to provide next to run status to jobs by using the appropriate token.
Questions?