Building a Dynamic Data Driven Application System For Hurricane Forecasting

Gabrielle Allen

Center for Computation & Technology and Department of Computer Science,
Louisiana State University, Baton Rouge, LA 70803
gallen@cct.lsu.edu
http://www.cct.lsu.edu

Abstract. The Louisiana Coastal Area presents an array of rich and urgent scientific problems that require new computational approaches. These problems are interconnected with common components: hurricane activity is aggravated by ongoing wetland erosion; water circulation models are used in hurricane forecasts, ecological planning and emergency response; environmental sensors provide information for models of different processes with varying spatial and time scales. This has prompted programs to build an integrated, comprehensive, computational framework for meteorological, coastal, and ecological models. Dynamic and adaptive capabilities are crucially important for such a framework, providing the ability to integrate coupled models with real-time sensor information, or to enable deadline based scenarios and emergency decision control systems. This paper describes the ongoing development of a Dynamic Data Driven Application System for coastal and environmental applications (DynaCode), highlighting the challenges of providing accurate and timely forecasts for hurricane events.

Key words: Dynamic data driven application systems, DDDAS, hurricane forecasting, event driven computing, priority computing, coastal modeling, computational frameworks

1 Introduction

The economically important Louisiana Coastal Area (LCA) is one of the world's most environmentally damaged ecosystems. In the past century nearly one-third of its wetlands have been lost and it is predicted that with no action by 2050 only one-third of the wetlands will remain. Beyond economic loss, LCA erosion has devastating effects on its inhabitants, especially in New Orleans whose location makes it extremely vulnerable to hurricanes and tropical storms. On 29th August 2005 Hurricane Katrina hit New Orleans, with storm surge and flooding resulting in a tragic loss of life and destruction of property and infrastructure. Soon after, Hurricane Rita caused similar devastation in the much less populated area of southwest Louisiana. In both cases entire communities were destroyed.

To effectively model the LCA region, a new comprehensive and dynamic approach is needed including the development of an integrated framework for coastal and environmental modeling capable of simulating all relevant interacting processes from erosion to storm surge to ecosystem biodiversity, handling multiple time (hours to years) and length (meters to kilometers) scales. This framework needs the ability to dynamically couple models and invoke algorithms based on streamed sensor or satellite data, locate appropriate data and resources, and create necessary workflows on demand, all in real-time. Such a system would enable restoration strategies, improve ecological forecasting, sensor placement, control of water diversion for salinity, or predict/control harmful algal blooms, and support sea rescue and oil spill response. In extreme situations, such as approaching hurricanes, results from multiple coupled ensemble models, dynamically compared with observations, could greatly improve emergency warnings.

These desired capabilities are included in the emerging field of Dynamic Data Driven Application Systems (DDDAS), which describes new complex, and inherently multidisciplinary, application scenarios where simulations can dynamically ingest and respond to real-time data from measuring devices, experimental equipment, or other simulations. In these scenarios, simulation codes are in turn also able to control these varied inputs, providing for advanced control loops integrated with simulation codes. Implementing these scenarios requires advances in simulation codes, algorithms, computer systems and measuring devices.

This paper describes work in the NSF funded DynaCode project to create a general DDDAS toolkit with applications in coastal and environmental modeling; a futuristic scenario (Sec. 2) provides general needs (Sec. 3) for components (Sec. 4). The rest of this section describes ongoing coastal research and development programs aligned with DynaCode, forming a scientific research foundation:

Lake Pontchartrain Forecast System. During Hurricane Katrina storm surge water from Lake Pontchartrain flooded New Orleans via breaches in outfall canals. The Army Corp of Engineers plans to close Interim Gated Structures at canal mouths during future storms, but this takes several hours, cannot occur in strong winds, and must be delayed as long as possible for storm rain water drainage.

The Lake Pontchartrain Forecast System (LPFS), developed by UNC and LSU, provides timely information to the Army Corp to aid in decision making for gate closing. LPFS is activated if a National Hurricane Center advisory places a storm track within 271 nautical miles of the canals, and an ensemble of storm surge (ADCIRC) runs is automatically deployed across the Louisiana Optical Network Initiative (LONI, http://www.loni.org) where mechanisms are in place to ensure they complete within two hours and results are provided to the Corp.

Louisiana CLEAR Program. The Coastal Louisiana Ecosystem Assessment and Restoration (CLEAR) program is developing ecological and predictive models to connect ecosystem needs with engineering design. CLEAR has developed a modeling tool to evaluate restoration alternatives using a combination of modules that predict physical processes, geomorphic features, and ecological succession. In addition, simulation models are being developed to provide an ecosystem forecasting system for the Mississippi Delta. This system will address questions

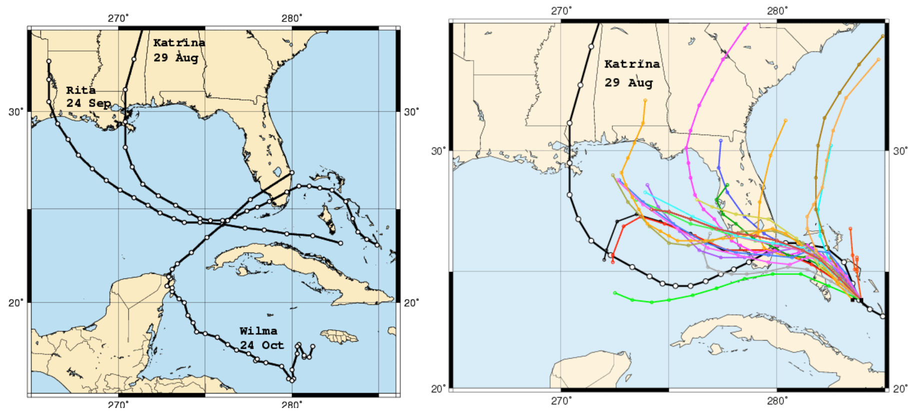

Fig. 1. Timely forecasts of the effects of hurricanes and tropical storms is imperative for emergency planning. The paths and intensity of the devastating hurricanes Katrina, Rita and Wilma [left] during 2005, as with other storms, are forecast from five days before expected landfall using different numerical and statistical models [right]. Model validity depends on factors such as the storm properties, location and environment.

such as what will happen to the Mississippi River Deltaic Plain under different scenarios of restoration alternatives, and what will be the benefits to society?

SURA Coastal Ocean Observing and Prediction (SCOOP): The SCOOP Program [1] (http://scoop.sura.org) involves a diverse collaboration of coastal modelers and computer scientists working with government agencies to create an open integrated network of distributed sensors, data and computer models. SCOOP is developing a broad community-oriented cyberinfrastructure to support coastal research activities, for which three key scenarios involving distributed coastal modeling drive infrastructure development: 24/7 operational activities where various coastal hydrodynamic models (with very different algorithms, implementations, data formats, etc) are run on a daily basis, driven by winds from different atmospheric models; retrospective modeling where researchers can investigate different models, historical data sets, analysis tools etc; and most relevant for DDDAS, hurricane forecasting. Here, severe storm events initiate automated model workflows triggered by National Hurricane Center advisories, high resolution wind fields are generated which then initiate ensembles of hydrodynamic models. The resulting data fields are distributed to the SCOOP partners for visualization and analysis, and are placed in a highly available archive [2].

2 Data Driven Hurricane Forecast Scenario

When advisories from the National Hurricane Center indicate that a storm may make landfall in a region impacting Louisiana, government officials, based on information provided by model predictions (Fig. 1) and balancing economic and social factors, must decide whether to evacuate New Orleans and surrounding towns and areas. Such advisories are provided every six hours, starting from some five days before the storm is predicted to make landfall. Evacuation notices for large cities like New Orleans need to be given 72 hours in advance.

Here we outline a complex DDDAS scenario which provides hurricane predictions using ensemble modeling: *A suddenly strengthening tropical depression tracked by satellite changes direction, worrying officials. An alert is issued to state researchers and an advanced autonomic modeling system begins the complex process of predicting and validating the hurricane path. Realtime data from sensor networks on buoys, drilling platforms, and aircraft, across the Gulf of Mexico, together with satellite imagery, provide varied resolution data on ocean temperature, current, wave height, wind direction and temperature. This data is fed continuously into a ensemble modeling tool which, using various optimization techniques from a standard toolkit and taking into account resource information, automatically and dynamically task farms dozens of simulations, monitored in real-time. Each simulation represents a complex workflow, with closely coupled models for atmospheric winds, ocean currents, surface waves and storm surges. The different models and algorithms within them, are dynamically chosen depending on physical conditions and required output sensitivity. Data assimilation methods are applied to observational data for boundary conditions and improved input data. Validation methods compare data between different ensemble runs and live monitoring data, with data tracking providing additional information for dynamic decisions. Studying ensemble data from remotely monitored simulations, researchers steer computations to ignore faulty or missing input data. Known sensitivity to uncertain sensor data is propagated through the coupled ensemble models quantifying uncertainty. Sophisticated comparison with current satellite data is made with synthesized data from ensemble models to determine in real-time which models/components are most reliable, and a final high resolution model is run to predict 72 hours in advance the detailed location and severity of the storm surge. Louisiana's Office of Emergency Preparedness disseminates interactive maps of the projected storm surge and initiates contingency plans including impending evacuations and road closures.*

3 System Requirements

Such scenarios require technical advances across simulation codes, algorithms, computer systems and measuring devices. Here we focus on different technical issues related to the various components of a DDDAS simulation toolkit:

– **Data Sources & Data Management.** Data from varied sources must be integrated with models, e.g. wind fields from observational sources or computer models. Such data has different uncertainties, and improving the quality of input data, on demand, can lower forecast uncertainty. The ability to dynamically create customized ensembles of wind fields is needed, validated and improved with sensor data, for specified regions, complete with uncertainty functions propagated through models. Sensor data from observing systems must be available for real-time verification and data assimilation. Services for finding, transporting and translating data must scale to complex workflows of coupled interacting models. Emergency and real-time computing scenarios demand highly available data sources and data transport that is fault tolerant with guaranteed quality of service. Leveraging new optical networks requires mechanisms to dynamically reserve and provision networks and data scheduling capabilities. Metadata describing the huge amounts of distributed data is also crucial. and must include provenance information.

- **Model-Model Coupling and Ensembles.** Cascades of coupled circulation, wave and transport models are needed. Beyond defining interfaces, methods are needed to track uncertainties, create and optimize distributed workflows as the storm approaches, and invoke models preferentially, based on algorithm performance and features indicated by input data.
  Cascades of models, with multiple components at each stage, lead to potentially hundreds of combinations where it is not known *a priori* which combinations give the best results. Automated and configurable ensemble modeling across grid resources, with continuous validation of results against observations and models, is critical for dynamically refining predictions. Algorithms are needed for dynamic configuration and creation of ensembles to provide predictions with a specifiable, required accuracy. In designing ensembles, the system must consider the availability and "cost" of resources, which may also depend on the threat and urgency e.g. with a Category 5 Hurricane requiring a higher quality of service than a Category 3 Hurricane.
- **Steering.** Automated steering is needed to adjust models to physical properties and the system being modeled, e.g. one could steer sensor inputs for improved accuracy. The remote steering of model codes, e.g. to change output parameters to provide verification data, or to initiating the reading of new improved data, will require advances to the model software. Beyond the technical capabilities for steering parameters (which often requires the involvement of domain experts), the steering mechanism must require authentication, with changes logged to ensure reproducibility.
- **Visualization and Notification.** Detailed visualizations, integrating multiple data and simulation sources showing the predicted effect of a storm are important for scientific understanding and public awareness (e.g. of the urgency of evacuation, or the benefits of building raised houses). Interactive and collaborative 3-D visualization for scientific insight will stress high speed networks, real-time algorithms and advanced clients. New visualizations of verification analysis and real-time sensor information are needed. Notification mechanisms to automatically inform scientists, administrators and emergency responders must be robust and configurable. Automated systems require human intervention and confirmation at different points, and the system should allow for mechanisms requiring authenticated response with intelligent fallback mechanisms.
- **Priority and Deadline Based Scheduling.** Dealing with unpredictable events, and deploying multiple models concurrently with data streams, pro-

vides new scheduling and reservation requirements: priority, deadline-based, and co-scheduling. Computational resources must be available on demand, with guaranteed deadlines for results; multiple resources must be scheduled simultaneously and/or in sequence. Resources go beyond traditional computers, including archival data, file systems, networks and visualization devices. Policies need to be adopted at computing centers that enable event-driven computing and data streaming; computational resources of various kinds need to be available on demand, with policies reflecting the job priority.

4 DynaCode Components

In the DynaCode project this functionality is being developed by adapting and extending existing software packages, and building on the SCOOP and LPFS scenarios. Collectively, the packages described below form the basis for a "DDDAS Toolkit", designed for generic DDDAS applications, with specific drivers including the hurricane forecast scenario:

- **Cactus Framework.** Cactus [3], a portable, modular software environment for developing HPC applications, has already been used to prototype DDDAS-style applications. Cactus has numerous existing capabilities relevant for DDDAS, including extreme portability, flexible and configurable I/O, an inbuilt parameter steering API and robust checkpoint/restart capabilities. Cactus V5, currently under development, will include additional crucial features, including the ability to expose individual 'thorn' (component) methods with web service interfaces, and to interoperate with other framework architectures. Support for the creation and archiving of provenance data including general computational and domain specific information is being added, along with automated archiving of simulation data.
- **User Interfaces.** An integrated, secure, web user interface developed with the GridSphere portal framework builds on existing relevant portlets [4]. New portlets include Threat Level and Notification, Ensemble Monitoring, Ensemble Track Visulization and Resource Status. A MacOSX "widget" (Fig. 2, left) displays hurricane track information and system status.
- **Data Management.** Through the SCOOP project a highly reliable coastal data archive [2] has been implemented at LSU, with 7TB of local storage and 7TB of remote storage (SDSC SRB) for historical data. This archive was designed to ingest and provide model (surge, wave, wind) and observational (sensor, satellite) data, and is integrated with a SCOOP catalogue service at UAH. To support dynamic scenarios a general trigger mechanism was added to the archive which can be configured to perform arbitrary tasks on arrival of certain files. This mechanism is used to drive ensemble configuration and deployment, notification and other components. DynaCode is partnered with the NSF funded PetaShare project which is developing new technologies for distributed data sharing and management, in particular data-aware storage systems and data-aware schedulers.

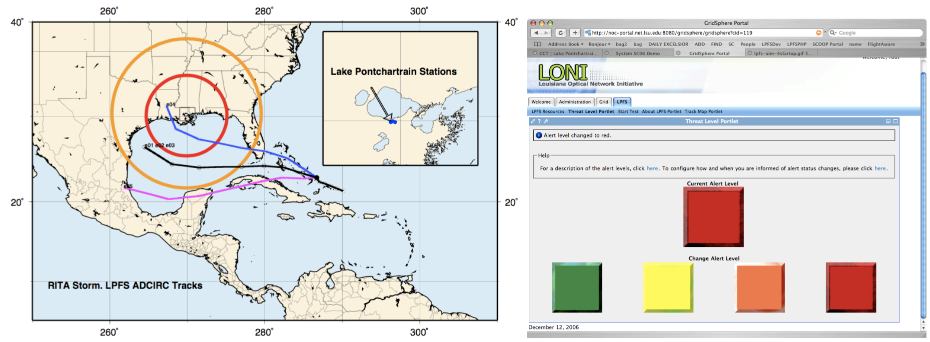

Fig. 2. [Left] The LPFS system is activated if any ensemble member places the hurricane track within 271 nautical miles of the canal mouths (inside the inner circle). [Right] A threat level system allows the threat level to be changed by trusted applications or scientists. This triggers the notification of system administrators, customers and scientists and the setting of policies on compute resources for appropriately prioritized jobs. The diagram shows a portal interface to the threat level system.

- **Ensembles.** Ensembles for DynaCode scenarios are currently deployed across distributed resources with a management component executed on the archive machine. The system is designed to support rapid experimentation rather than complex workflows, and provides a completely data-driven architecture with the ensemble runs being triggered by the arrival of input wind files. As ensemble runs complete, results are fed to a visualization service and also archived in the ROAR Archive [2]. Metadata relating to the run and the result set is fed into the catalog developed at UAH via a service interface.
- **Monitoring.** The workflow requires highly reliable monitoring to detect failures and prompt corrective action. Monitoring information (e.g. data transfers/job status) is registered by the various workflow components and can be viewed via portal interfaces. A spool mechanism is used to deliver monitoring information via log files to a remote service, providing high reliability and flexibility. This ensures the DynaCode workflow will not fail due to unavailability of the monitoring system. Also, the workflow executes faster than a system where monitoring information is transported synchronously.
- **Notification.** A general notification mechanism sends messages via different mechanisms to configurable role-based groups. The system currently supports email, instant messaging, and SMS text messages, and is configurable via a GridSphere portlet. The portlet behaves as a messaging server that receives updates from e.g. the workflow system and relays messages to subscribers. Subscribers can belong to different groups that determine the information content of messages they receive, allowing messages to be customized for e.g. system administrators, scientists or emergency responders.
- **Priority Scheduling & Threat Levels.** Accurate forecasts of hurricane events, involving large ensembles, need to be completed quickly and reliably

with specific deadlines. To provide on-demand resources the DynaCode workflow makes use of policies as well as software. On the large scale resources of LONI and CCT, the queues have been configured so that it is possible to preempt currently running jobs and free compute resources at extremely short notice. Large queues are reserved for codes that can checkpoint and restart. These queues share compute nodes with preemptive queues that preempt jobs in the 'checkpoint' queues when they receive jobs to run. Software such as SPRUCE (http://spruce.teragrid.org/) is being used to provide elevated priority and preemptive capabilities to jobs that hold special tokens reducing user management burden from system administrators.
A "Threat Level" service has been developed; trusted applications or users can set a global threat level to red, amber, yellow or green (using web service or portal interfaces), depending on the perceived threat and urgency (Fig. 2). Changes to the threat level triggers notification to different role groups, and is being integrated with the priority scheduling system and policies.

– **Co-allocation.** DynaCode is partnering with the NSF Enlightened Computing project which is developing application-enabling middleware for optical networks. The HARC co-allocator, developed through the Enlightened-DynaCode collaboration, can already allocate reservations on compute resources and optical networks, and is being brought into production use on the LONI network to support DynaCode and other projects.

Acknowledgments. The author acknowledges contributions from colleagues in the SCOOP, CLEAR and LPFS projects and collaborators at LSU. This work is part of the NSF DynaCode project (0540374), with additional funding from the SURA Coastal Ocean Observing and Prediction (SCOOP) Program (including ONR Award N00014-04-1-0721, NOAA Award NA04NOS4730254). Computational resources and expertise from LONI and CCT are gratefully acknowledged.

References

1. Bogden, P., Allen, G., Stone, G., Bintz, J., Graber, H., Graves, S., Luettich, R., Reed, D., Sheng, P., Wang, H., Zhao, W.: The Southeastern University Research Association Coastal Ocean Observing and Prediction Program: Integrating Marine Science and Information Technology. In: Proceedings of the OCEANS 2005 MTS/IEEE Conference, Sept 18-23, 2005, Washington, D.C. (2005)
2. MacLaren, J., Allen, G., Dekate, C., Huang, D., Hutanu, A., Zhang, C.: Shelter from the Storm: Building a Safe Archive in a Hostile World. In: Proceedings of the The Second International Workshop on Grid Computing and its Application to Data Analysis (GADA'05), Agia Napa, Cyprus, Springer Verlag (2005)
3. Goodale, T., Allen, G., Lanfermann, G., Massó, J., Radke, T., Seidel, E., Shalf, J.: The Cactus framework and toolkit: Design and applications. In: High Performance Computing for Computational Science - VECPAR 2002, 5th International Conference, Porto, Portugal, June 26-28, 2002, Berlin, Springer (2003) 197–227
4. Zhang, C., Dekate, C., Allen, G., Kelley, I., MacLaren, J.: An Application Portal for Collaborative Coastal Modeling. Concurrency Computat.: Pract. Exper. **18** (2006)