The QuakeSim Project: Numerical Simulations for Active Tectonic Processes

Andrea Donnellan
Jay Parker
Greg Lyzenga
Robert Granat
Jet Propulsion Laboratory
4800 Oak Grove Drive
Pasadena, CA 91109

Geoffrey Fox
Marlon Pierce
Indiana University
501 N. Morton, Suite 224
Bloomington, IN 47404

John Rundle
University of California
Davis, CA 95616

Dennis McLeod
University of Southern California
Los Angeles, CA 90089

Lisa Grant
University of California
Irvine, CA 92697

Terry Tullis
Box 1846
Brown University
Providence, RI 02912

Abstract-*In order to develop a solid earth science framework for understanding and studying of active tectonic and earthquake processes, this task develops simulation and analysis tools to study the physics of earthquakes using state-of-the-art modeling, data manipulation, and pattern recognition technologies. We develop clearly defined accessible data formats and code protocols as inputs to the simulations. These are adapted to high-performance computers because the solid earth system is extremely complex and nonlinear resulting in computationally intensive problems with millions of unknowns. With these tools it will be possible to construct the more complex models and simulations necessary to develop hazard assessment systems critical for reducing future losses from major earthquakes.*

I. INTRODUCTION

We are building a new Problem Solving Environment (QuakeSim) for use by the seismological, crustal deformation, and tectonics communities for developing an understanding of active tectonic and earthquake processes. One of the most critical aspects of our system is supporting interoperability given the heterogeneous nature of data sources as well as the variety of application programs, tools, and simulation packages that must operate with data from our system. Interoperability is being implemented by using distributed object technology combined with development of object API's that conform to emerging standards. The full objective over this three-year program is to produce a system to fully model earthquake-related data. Components of this system include:

- A database system for handling both real and simulated data
- Fully three-dimensional finite element code (FEM) with an adaptive mesh generator capable

of running on workstations and supercomputers for carrying out earthquake simulations

- Inversion algorithms and assimilation codes for constraining the models and simulations with data
- A collaborative portal (object broker) for allowing seamless communication between codes, reference models, and data
- Visualization codes for interpretation of data and models
- Pattern recognizers capable of running on workstations and supercomputers for analyzing data and simulations

Project details and documentation are available at the QuakeSim main web page at http://quakesim.jpl.nasa.gov.

II. Objectives

Our objectives in the second year of this project were twofold: to build a working prototype of the QuakeSim problem solving environment, and to implement and measure performance and scalability of high end computing simulations of earthquakes using the GeoFEST and PARK codes.

We have developed a prototype environment that demonstrates the fault database, mesh-generation, GeoFEST simulation and RIVA visualization for solving a typical crustal deformation problem. This was successfully completed and documented in Milestone I *Interoperability Prototype* at:
http://quakesim.jpl.nasa.gov/milestones.html.

QuakeSim is developing three high-end computing simulation tools: GeoFEST, PARK and Virtual California. We are working to demonstrate parallel scaling and efficiency for the GeoFEST finite element code and the PARK boundary element code for unstable slip. Both these codes were enhanced by integration with high performance computing tools: GeoFEST with Pyramid adaptive meshing, and PARK with a fast multipole library.

III. Application Programs

Several software applications are being implemented in this system for individual use, or for interaction with other codes in the system. From the Web services point of view, each of the applications is wrapped by an application web service proxy that encapsulates both the internal and external services needed by an application component. Internal services include job submission and file transfer, for example, which are needed to run the application on some host in isolation. External services are used for communication between running codes and include a) the delivery of the messages/events about application state ("Code A has generated an output file and notifies Code B") and b) file transfer services ("Transfer the output of Code A to Code B"). Notification and stateful Web services are key parts of the Open Grid Services Architecture, which we will leverage. Events and message delivery mechanisms are one of our group's areas of strength (e.g. NaradaBrokering from Indiana University). Basic code descriptions in the system are as follows:

A. disloc

Handles multiple arbitrarily dipping dislocations (faults) in an elastic half-space to produce surface displacements based on Okada's 1985 paper.

B. simplex

Inverts surface geodetic displacements for fault parameters using simulated annealing downhill residual minimization. Is based on disloc and uses dislocations in an elastic half space.

C. geofest (coupled with a mesh generator)

Three-dimensional viscoelastic finite element model for calculating nodal displacements and tractions. Allows for realistic fault geometry and characteristics, material properties, and body forces.

D. VC (VirtualCalifornia)

Program to simulate interactions between vertical strike-slip faults using an elastic layer over a viscoelastic half-space.

E. park

Boundary element program to calculate fault slip velocity history based on fault frictional properties.

F. DAHMM

Time series analysis program based on Hidden Markov Modeling. Produces feature vectors and probabilities for transitioning from one class to another.

G. PDPC

Time series analysis pattern recognition program to calculate anomalous seismicity patterns and probabilities.

H. RIVA

RIVA (Remote Interactive Visualization and Analysis) System will be use to visualize the vertical and horizontal surface deformation generated by GeoFEST simulation and the Virtual California simulation. The vertical deformation will be represented as displacement on top of digital elevation model and the horizontal deformation will be represented as a semi-opaque overlay on top of the LandSat images. Animation of time-varying deformation data at 1km resolution for Southern California will be produced. RIVA will be integrated with GeoFEST and VC using the standardized data format to be defined by the project.

I. PARVOX

The 3D stress and strain data generated by GeoFEST will be visualized by ParVox (Parallel Voxel Renderer). ParVox is a parallel 3D volume rendering system capable of visualizing large time-varying, multiple variable 3D data sets. We will add new rendering capabilities in ParVox to visualize unstructured grid data generated by GeoFEST. ParVox will be integrated into the proposed web portal as a 3D visualization tool using the web service interface.

RIVA and PARVOX are NASA CT project in house efforts developed at JPL, primarily by Peggy Li.

IV. INTEROPERABILITY/PORTAL

We have demonstrated a web-services problem-solving environment prototype that links together diverse earthquake science applications on distributed computers. We can now use the prototype to build a model with faults and layers from the fault database, automatically generate a finite element mesh, solve for crustal deformation and produce a full color animation of the result integrated with remote sensing data. This portal environment is rapidly expanding to include many more applications and tools.

Our approach is to build a three-tiered architecture system (Figure 1). The tiers are: 1) A portal user interface layer that manages client components; 2) A service tier that provides general services (job submission, file transfer, database access, etc.) that can be deployed to multiple host computers; and 3) Backend resource, including databases and earthquake modeling codes.

The user interacts with the system through the Web Browser interface (top). The web browser connects to the aggregating portal, running on the User Interface Server (http://complexity.ucs.indiana.edu in the testbed). The "Aggregating Portal" is so termed because it collects and manages dynamically generated web pages (in JSP) that may be developed independently of the portal and run on separate servers. The components responsible for managing particular web site connections are known as portlets. The aggregating portal can be used to customize the display, control the arrangement of portlet components, manage user accounts, set access control restrictions, etc.

The portlet components are responsible for loading and managing web pages that serve as clients to remotely running Web services, shown in the lower figure. For example, a DB service runs on a host, job submission and file management services on another machine (typically running on danube.ucs.indiana.edu in the testbed) and visualization services on another (such as RIVA, running on the host jabba.jpl.nasa.gov). We use Web Services to describe the remote services and invoke their capabilities. Generally, connections are SOAP over HTTP. We may also use Grid connections (GRAM and GridFTP) to access

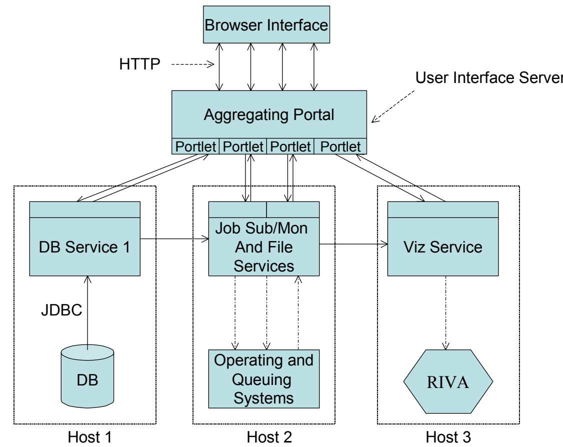

Figure 1. Portal and service architecture

our applications. Database connections between the Database service and the actual database are handled by JDBC (Java Database Connectivity), a standard technique.

The QuakeSim portal effort has been one of the pioneering efforts in building Computing Portals out of reusable portlet components. Pierce and Fox of the QuakeSim team collaborate with other portal developers following the portlet component approach through the Open Grid Computing Environments consortium (OGCE: Argonne National Laboratory, Indiana University, the University of Michigan, the National Center for Supercomputing Applications, and the Texas Advanced Computing Center). This project has been funded by the NSF National Middleware Initiative (Pierce, PI) to develop general software releases for portals and to end the isolation and custom solutions that have plagued earlier portal efforts. Pierce and Fox's involvement with both the QuakeSim project and the OGCE will ensure that the QuakeSim portal will benefit from the larger community of portal development: we may extend the QuakeSim portal to use capabilities developed by other groups and may share the capabilities developed by the QuakeSim portal with the portal-building community.

EARTHQUAKE FAULT DATABASE

The "database system" for this project must manage a variety of types of earthquake science data and information. There are pre-existing collections, with heterogeneous access interfaces; there are also some structured collections managed by general-purpose database management systems.

Most faults in the existing databases have been divided into characteristic segments that are proposed to rupture as a unit. Geologic slip rates are assigned to large segments rather than to the specific locations (i.e. geographic coordinates) where they were measured. These simplifications and assumptions are desirable for seismic hazard analysis, but they introduce a level of geologic interpretation and subjective bias that is inappropriate for simulations of fault behavior. Therefore, we are developing an objective database that includes primary geologic and paleoseismic fault parameters (fault location/geometry, slip rate at measured location, measurements of coseismic displacement, dates and locations of previous ruptures) as well as separate interpreted/subjective fault parameters such as characteristic segments, average recurrence interval, magnitude of characteristic ruptures, etc. The database is updated as more data are acquired and interpreted through research and the numerical simulations.

To support this earthquake fault database our database system is being developed on the public domain DBMS, MySQL. We utilize an extensible relational system. These systems support the definition, storage, access, and control of collections of structured data. Ultimately, we require extensible type definition capabilities in the DBMS (to accommodate application-specific kinds of data), the ability to combine information from multiple databases, and mechanisms to efficiently return XML results from requests.

We have developed an XML scheme to describe various parameters of earthquake faults and input data. One key issue that we must address here is the fact that such DBMSs operate on SQL requests, rather than those in some XML-based query language. Query languages for XML are just now emerging; in consequence, we initially do XML to SQL translations in our middleware/broker. With time, as XML query language(s) emerge, we will employ them in our system. To provide for the access and manipulation of heterogeneous data sources (datasets, databases), the integration of information from such sources, and the structural organization and data mining of this data, we are employing techniques being developed at the USC Integrated Media Systems Center for wrapper-based information fusion to support data source access and integration

Our fault database is searchable with annotated earthquake fault records from publications. The database team designed the fields that constituted the database records and provided a web-based interface that enables the submitting and accessing of those records. A small group of geologists/paleoseismologists searched the literature and collected annotated records of Southern California earthquake faults to begin population of the fault database.

Code Improvements

We have made major performance improvements to two of our simulation codes. The capabilities of the PARK code have increased from 15,000 fault elements and 500 time steps in just under eight hours (one processor) to 150,000 fault elements and 5,000 time steps in the same amount of time on 256 processors. This was demonstrated on the Chapman computer at Ames. Integration with the parallel fast multipole method has resulted in a one hundredfold performance improvement in solving problems of this size.

GeoFEST has had similar success. Its demonstrated capabilities have gone from solving a 55,000 finite element problem in just under fourteen hours (one processor) to 1.4 million elements in 2.8 hours (64 processors on the Thunderhead Linux computer at GSFC). Parallel implementation includes integration with the Pyramid adaptive mesh tools developed by the Computational Technologies (CT) project. GeoFEST has been ported to eight different parallel machines, including Linux, SGI and Apple. Ongoing development with Pyramid will allow use of near optimal meshes for high quality regional simulations.

On Demand Science

The portal is intended to enhance science activities and is available to the science community. The system allows for users to link applications, which allows for ingestion of data, more complete modeling, and visualization of data and results. Here we describe results from two science applications using the portal.

Simplex, which is implemented in the QuakeSim portal, was used to investigate the geometry of the fault ruptured by the December 22, 2003, magnitude 6.5, San Simeon earthquake. The Simplex portal tool enables scientists to rapidly find fault parameters based on surface deformation and a nonlinear least-squares optimal Okada solution in an elastic half space. Determination of the elastic rupture geometry is necessary to investigate immediate time-dependent deformation processes. Additionally, the rupture geometry is important for determining stress transfer to adjacent faults and understanding triggered seismicity. In future use, rapidly available information about the rupture parameters could be used to optimally deploy campaign GPS receivers and select InSAR mission targets, as well as assess increased seismic hazard.

The preliminary report for the San Simeon earthquake provided details about the earthquake. The report was available December 24, 2003. The San Simeon main shock was located 11 km NE of San Simeon and caused damage and casualties in Paso Robles, 39 km to the ESE. The focal mechanism derived from regional seismic networks is nearly pure thrust motion on a fault striking ESE-WNW. The estimated dip slip is 50 cm. The main shock was followed by more than 897 aftershocks greater than magnitude 1.8 in the first 48 hours. The aftershock region extends 30 km in length, with a surface projection width of 10 km. The aftershock depths rarely exceed 8 km.

The epicenter of the San Simeon earthquake is located in a region with many mapped active faults, however the fault that ruptured had not been determined at the time of the preliminary report. Additionally, no surface rupture had been identified. The preliminary report described two possible fault planes, one dipping NE and the other dipping SW. The aftershock distribution and finite-fault models suggest a NE dipping fault.

Using Simplex, the two possible fault dip directions were tested. Co-seismic displacements for 12 local stations of the Southern California Integrated GPS Network (SCIGN) provided solution constraints. Professor Tom Herring of MIT, Chairman of SCIGN, provided these preliminary displacement estimates. Initial fault guesses were located near the epicenter and parameters consistent with the reported magnitude and focal mechanism were chosen. For both the SW and NE dipping fault initial guesses, the fault orientation and geometry were allowed to vary. The SW dipping fault consistently reached lower residual values than the NE dipping fault for tests with modifications to the parameters in the input file. This fault dips 46 degrees and strikes approximately ESE. The fault reaches a depth of 8 km and does not break the surface. The dip slip magnitude in the minimized Simplex result is 50 cm. Due to the lack of sensors close to the fault, and absence of sensors SE of the epicenter, it is difficult to accurately constrain the fault location.

GPS data shows an anomalous band of compression in the northern part of the Los Angeles basin. Scientists at UC Davis used QuakeSim tools to investigate both elastic and viscoelastic models of the basin to test what set of model parameters best reproduced the observed GPS data. They used SIMPLEX for elastic models and GeoFEST for viscoelastic models.

A GeoFEST model was constructed using geologic boundary conditions, recurring fault slip and a layered viscoelastic crustal structure. The top of the model is a free surface, which produces surface velocities that can be compared to observations from SCIGN measurements.

The models that yielded the best results featured a vertically strong crust, a single fault, and a low-rigidity, anelastically deforming basin abutting the rigid San Gabriel Mountains block (see Figure 4 below). These models produced a horizontal velocity gradient that matched the observed gradient across the LA basin both in magnitude and shape. While the match is not perfect, the models suggest that the sedimentary basin may be controlling the deformation in northern metropolitan Los Angeles, rather than faults south of the Sierra Madre fault as some have claimed.

ACKNOWLEGMENT

We thank the graduate students, Anne Chen, and Shan Gao from USC for their work on the database systems. We thank Robert Shcherbakov from UC Davis for his work assessing the data assimilation methods. Charles Norton, from JPL, assisted with the code parallelization and mesh refinement. Peggy Li, from JPL, created the visualization software RIVA and PARVOX. Nick Beeler from the USGS worked on code improvements for PARK. Portions of this work were carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract with NASA.