Pure and Applied Geophysics

QuakeSim and the Solid Earth Research Virtual Observatory

Andrea Donnellan,[1] John Rundle,[2] Geoffrey Fox,[3] Dennis McLeod,[4]
Lisa Grant,[5] Terry Tullis,[6] Marlon Pierce,[3] Jay Parker,[1] Greg Lyzenga,[1]
Robert Granat,[1] and Margaret Glasscoe[1]

Abstract—We are developing simulation and analysis tools in order to develop a solid Earth Science framework for understanding and studying active tectonic and earthquake processes. The goal of QuakeSim and its extension, the Solid Earth Research Virtual Observatory (SERVO), is to study the physics of earthquakes using state-of-the-art modeling, data manipulation, and pattern recognition technologies. We are developing clearly defined accessible data formats and code protocols as inputs to simulations, which are adapted to high-performance computers. 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. We are using Web (Grid) service technology to demonstrate the assimilation of multiple distributed data sources (a typical data grid problem) into a major parallel high-performance computing earthquake forecasting code. Such a linkage of Geoinformatics with Geocomplexity demonstrates the value of the Solid Earth Research Virtual Observatory (SERVO) Grid concept, and advances Grid technology by building the first real-time large-scale data assimilation grid.

Key words: Earthquake, simulation, deformation, GPS, QuakeSim, web services.

Introduction

QuakeSim is a Problem Solving Environment for 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

[1] Science Division, Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, USA. E-mail: donnellan@jpl.nasa.gov
[2] Center for Computational Science and Engineering, University of California, Davis, California 95616, USA. E-mail: rundle@geology.ucdavis.edu
[3] Community Grid Computing Laboratory, Indiana University, IN 47404, USA. E-mail: gcf@indiana.edu
[4] Computer Science Department, University of Southern California, Los Angeles, CA 90089, USA. E-mail: mcleod@usc.edu
[5] Environmental Health, Science, and Policy, University of California, Irvine, CA 92697, USA. E-mail: lgrant@uci.edu
[6] Brown University, Providence, RI 02912, USA. E-mail: terry_tullis@brown.edu

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 application program interfaces (APIs) that conform to emerging standards. The full objective 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.

This project will result in the necessary applied research and infrastructure development to carry out efficient performance of complex models on high-end computers using distributed heterogeneous data. The system will enable an ease of data discovery, access, and usage from the scientific user point of view, as well as provide capabilities to carry out efficient data mining. We focus on the development and use of data assimilation techniques to support the evolution of numerical simulations of earthquake fault systems, together with space geodetic and other datasets. Our eventual goal is to develop the capability to forecast earthquakes in fault systems such as in California.

Integrating Data and Models

The last five years have shown unprecedented growth in the amount and quality of space geodetic data collected to characterize geodynamical crustal deformation in earthquake-prone areas such as California and Japan. The Southern California Integrated Geodetic Network (SCIGN), the growing EarthScope Plate Boundary Observatory (PBO) network, and data from Interferometric Synthetic Aperature Radar (InSAR) satellites are examples. Hey and Trefethen (http://www.grid2002.org) (FOX and HEY, 2003) stressed the generality and importance of Grid applications exhibiting this "data deluge."

Many of the techniques applied here grow out of the modern science of *dynamic data-driven complex nonlinear systems*. The natural systems we encounter are complex in their attributes and behavior, nonlinear in fundamental ways, and exhibit properties over a wide diversity of spatial and temporal scales. The most destructive and largest of the events produced by these systems are typically called *extreme events*, and are the most in need of forecasting and mitigation. The systems that produce these extreme events are *dynamical systems*, because their configurations evolve as forces change in time from one definable state of the system in its *state space* to another. Since these events emerge as a result of the rules governing the temporal evolution of the system, they constitute *emergent* phenomena produced by the dynamics. Moreover, extreme events such as large earthquakes are examples of *coherent space-time structures*, because they cover a definite spatial volume over a limited time span, and are characterized by physical properties that are similar or coherent over space and time.

We project major advances in the understanding of complex systems from the expected increase in data. The work here will result in the merging of parallel complex system simulations with federated database and datagrid technologies to manage heterogeneously distributed data streams and repositories (Fig. 1). The objective is to have a system that can ingest broad classes of data into dynamical models that have predictive capability.

Integration of multi-disciplinary models is a critical goal for both physical and computer science in all approaches to complexity, which one typically models as a heterogeneous hierarchical structure. Moving up the hierarchy, new abstractions are introduced and a process that we term coarse graining is defined for deriving the parameters at the higher scale from those at the lower. Multi-scale models are derived by various methods that mix theory, experiment and phenomenology and are illustrated by multigrid, fast multipole and pattern dynamics methods successfully applied in many fields including Earth Science (RUNDLE *et al.*, 2003). Explicitly recognizing and supporting coarse-graining represents a scientific advance but also allows one to classify the problems that really require high-end computing resources from those that can be performed on more cost effective loosely coupled Grid facilities such as the averaging of fine grain data and simulations.

Multiscale integration for Earth science requires the linkage of data grids and high performance computing (Fig. 1). Data grids must manage data sets that are either too large to be stored in a single location or else are geographically distributed by their nature (such as data generated by distributed sensors). The computational requirements of data grids are often loosely coupled and thus are embarrassingly parallel. Large-scale simulations require closely coupled systems. QuakeSim and SERVO support both styles of computing. The modeler is allowed to specify the linkage of descriptions across scales as well as the criterion to be used to decide at which level to represent the system. The goal is to support a multitude of distributed data sources, ranging over federated database (SHETH and LARSON, 1990), sensor, satellite data and simulation data, all of which may be stored at various locations
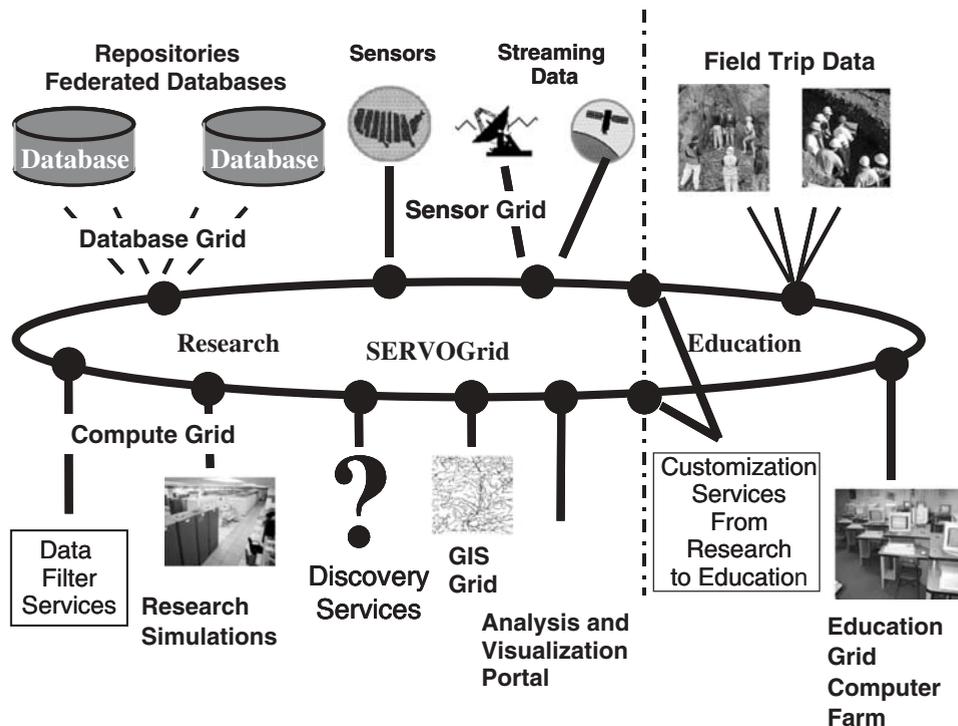
Figure 1
SERVO Grid with typical grid nodes.

with various technologies in various formats. QuakeSim conforms to the emerging Open Grid Services Architecture (TALIA, 2002).

Computational Architecture and Infrastructure

Our architecture is built on modern Grid and Web Service technology (ATKAS *et al.*, 2006) whose broad academic and commercial support should lead to sustainable solutions that can track the inevitable technology change. The architecture of QuakeSim and SERVO consists of distributed, federated data systems, data filtering and coarse graining applications, and high performance applications that require coupling. All pieces (the data, the computing resources, and so on) are specified with URIs and described by XML metadata.

Web Services

We use Web services to describe the interfaces and communication protocols needed to build our Web services. Generally defined, these are the constituent parts of an XML-based distributed service system. Standard XML schemas are used to define

implementation independent representations of the service's invocation interface, WSDL (Web Services Description Language): SOAP (Simple Object Access Protocol) messages exchanged between two applications. Interfaces to services may be discovered through XML-based repositories. Numerous other services may supplement these basic capabilities, including message level security and dynamic invocation frameworks that simplify client deployment. Implementations of clients and services can in principle be implemented in any programming language (such as Java, C++, or Python), with interoperability obtained through XML's neutrality.

One of the basic attributes of Web services is their loose integration. One does not have to use SOAP, for example, as the remote method invocation procedure. There are obviously times when this is desirable. For example, a number of protocols are available for file transfer, focusing on some aspect such as reliability or performance. These services may be described in WSDL, with WSDL ports binding to appropriate protocol implementations, or perhaps several such implementations. In such cases, negotiation must take place between client and service.

Our approach to Web services divides them into two major categories: Core and application. Core services include general tasks such as file transfer and job submission. Application services consist of metadata and core services needed to create instances of scientific application codes. Application services may be bound to particular host computers and core services needed to accomplish a particular task.

Two very important investigations are currently underway under the auspices of the Global Grid Forum (GANNON et al., 2002). The first is the merging of computing grid technologies and Web services (i.e., grid Web services). The current focus here is on describing transitory (dynamic, or stateful) services. The second is the survey of requirements and tools that will be needed to orchestrate multiple independent (grid) Web services into aggregate services.

XML-Based Metadata Services

In general, SERVO is a distributed object environment. All constituent parts (data, computing resources, services, applications, etc.) are named with universal resource identifiers (URIs) and described with XML metadata. The challenges faced in assembling such a system include a) resolution of URIs into real locations and service points; b) simple creation and posting of XML metadata nuggets in various schema formats; c) browsing and searching XML metadata units.

XML descriptions (schemas) can be developed to describe everything: computing service interfaces, sensor data, application input decks, user profiles, and so on. Because all metadata are described by some appropriate schema, which in turn derive from the XML schema specification, it is possible to build tools that dynamically create custom interfaces for creating and manipulating individual XML metadata pieces. We have taken initial steps in this direction with the development of a "Schema Wizard" tool.

After metadata instances are created, they must be stored persistently in distributed, federated databases. On top of the federated storage and retrieval systems, we are building organizational systems for the data. This requires the development of URI systems for hierarchically organizing metadata pieces, together with software for resolving these URIs and creating internal representations of the retrieved data. It is also possible to define multiple URIs for a single resource, with URI links pointing to the "real" URI name. This allows metadata instance to be grouped into numerous hierarchical naming schemes.

Federated Database Systems and Associated Tools

Our goal is to provide interfaces through which users transparently access a heterogeneous collection of independently operated and geographically dispersed databases, as if they formed a large virtual database (SHETH and LARSON, 1990; ARBIB and GRETHE, 2001). There are five main challenges associated with developing a meta-query facility for earthquake science databases: (1) Define a basic collection of concepts and interrelationships to describe and classify information units exported by participating information providers (a *"geophysics meta-ontology"*), in order to provide for a linkage mechanism among the collection of databases (WIEDERHOLD, 1994). (2) Develop a "meta-query mediator" engine to allow users to formulate complex meta-queries. (3) Develop methods to translate meta-queries into simpler derived queries addressed to the component databases. (4) Develop methods to collect and integrate the results of derived queries, to present the user with a coherent reply that addresses the initial meta-query. (5) Develop generic software engineering methodologies to allow for easy and dynamic extension, modification, and enhancement of the system.

We use the developing Grid Forum standard data repository interfaces to build data understanding and data mining tools that integrate the XML and federated database subsystems. Data understanding tools enable the discovery of information based upon descriptions, and the conversion of heterogeneous structures and formats into SERVO compatible form. The data mining in SERVO focuses on insights into patterns across levels of data abstraction, and perhaps even to mining or discovering new pattern sequences and corresponding issues and concepts.

Interoperability Portal

QuakeSim demonstrates a web-services problem-solving environment that links together diverse earthquake science applications on distributed computers. For example, one can use QuakeSim 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 Grid portal system (Fig. 2) that is compatible with Web Service Architecture (BOOTH *et al.*, 2004) principals. These components are 1) a "User Interface Server" (center of the figure) that is responsible for managing user interactions and client components of the system by creating and managing the life cycles of Web Service Requester Agents; 2) a distributed collection of Web Service Provider Agents that implement various tasks (code submission and management, file transfer, database access) that are deployed on various host computers (Hosts 1, 2, and 3 of Fig. 2); and 3) various backed resources, including databases (such as QuakeTables) and applications (such as GeoFEST and RIVA of Fig. 2). Web Service Architectures for Grids are analogous to three-tiered architectures (BOOTH *et al.*, 2004).

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:8282 in the testbed). The "Aggregating Portal" is so termed because it collects and manages dynamically generated web pages (in JSP, JavaServer Pages) 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, and 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. For example, a database 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
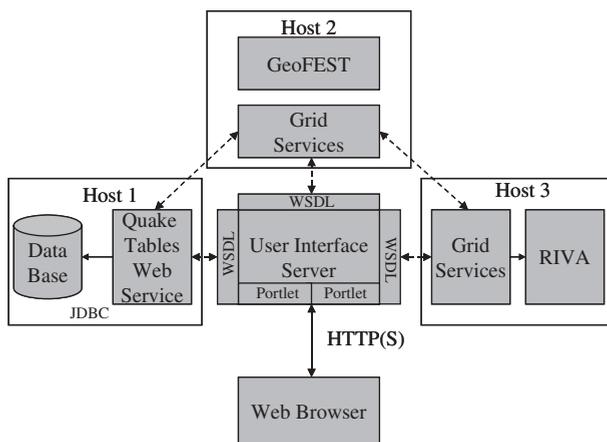

Figure 2
The QuakeSim portal supports distributed web services for interacting with databases (such as the QuakeTables fault database), remote applications (such as GeoFEST), and batch visualization (such as RIVA). The portal creates user interfaces for browser-based users through an aggregation of web components ("portlets"). In the figure, broken arrows represent SOAP connections. JDBC stands for Java Database Connectivity.

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 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. The QuakeSim team collaborates 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. The QuakeSim project benefits from involvement with the OGCE larger community of portal development, providing the option of extending the QuakeSim portal to use capabilities developed by other groups and of sharing the capabilities developed for the QuakeSim portal with the portal-building community.

Earthquake Fault Database

The "database system" for this project manages a variety of types of earthquake science data and information. There are pre-existing collections, with heterogeneous access interfaces; there are also 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. The QuakeSim database is 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 database management system (DBMS), MySQL. Though we

may migrate to a commercially available system in the future, such as Oracle, we are finding that the features of MySQL are adequate and that the advantages of a publicly available DBMS outweigh the more advanced features of a commercial system. 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 schema 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 these 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 populate the fault database. This fault database system has been designed to be scalable to considerably larger amounts of data. Plans are in place to allow the system to respond to more complex requests from both users and program agents via web service and semantic web (semantic grid) technologies.

We have also developed map interfaces to the fault data base. These use both standard Geographical Information System tools like the NASA OnEarth Web Map Server as well as Google Maps. This is depicted in Figure 3.

Application Programs

QuakeSim (http://quakesim.jpl.nasa.gov) is developing three high-end computing simulation tools: GeoFEST, PARK and Virtual California. We have demonstrated parallel scaling and efficiency for the GeoFEST finite element code, the PARK boundary element code for unstable slip, and the Green's functions Virtual California. GeoFEST was enhanced by integration with the Pyramid adaptive meshing library, and PARK with a fast multipole library.
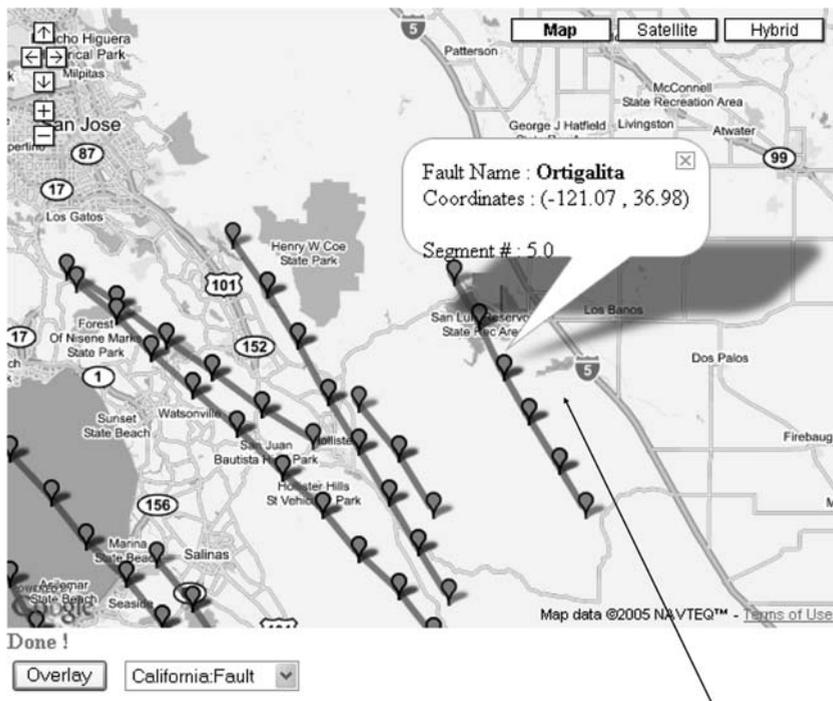
Figure 3

The QuakeTables fault database can be navigated using interactive Google maps as well as through search query HTML forms. Faults are typically stored by segment number, so map interfaces are convenient for both verifying continuity and setting up input files for computing problems.

The software applications are implemented in QuakeSim for individual use, or for interaction with other codes in the system. As the different programs are developed they will be added to the QuakeSim/SERVO portal. From the Web services point of view, an application Web service proxy that encapsulates both the internal and external services needed by an application component wraps each of the applications. 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. Basic code descriptions in the system are as follows:

disloc

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

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.

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.

VC (VirtualCalifornia)

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

park

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

DAHMM

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

PDPC

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

RIVA

RIVA (Remote Interactive Visualization and Analysis) System is used for visualizing the vertical and horizontal surface deformation generated by GeoFEST and Virtual California. The vertical deformation is represented as displacement on top of a digital elevation model and the horizontal deformation is represented as a semi-opaque overlay on top of LandSat images.

PARVOX

The 3-D stress and strain data generated by GeoFEST is visualized with ParVox (Parallel Voxel Renderer). ParVox is a parallel 3-D volume rendering system capable of visualizing large time-varying, multiple variable 3-D data sets.

Data Assimilation and Mining Infrastructure

Solid earth science models must define an evolving, high-dimensional nonlinear dynamical system, and the problems are large enough that they must be executed on high performance computers. Data from multiple sources must be ingested into the

models to provide constraints, and methods must be developed to increase throughput and efficiency in order for the constrained models to be run on high-end computers.

Data Assimilation

Data assimilation is the process by which observational data are incorporated into models to set these parameters, and to "steer" or "tune" them in real time as new data become available. The result of the data assimilation process is a model that is maximally consistent with the observed data and is useful in ensemble forecasting. Data assimilation methods must be used in conjunction with the dynamical models as a means of developing an ensemble forecast capability.

In order to automate the modeling process we are developing approaches to data assimilation based initially on the most general, and currently the most used, method of data assimilation, the use of *T*angent linear and *A*djoint *M*odel *C*ompilers (TAMC; GIERING and KAMINSKI, 1997). Such models are based on the idea that the system state follows an evolutionary path through state space as time progresses, and that observations can be used to periodically adjust model parameters, so that the model path is as close as possible to the path represented by the observed system.

Our research in the areas of ensemble forecasting and data assimilation is in three areas: 1) Static assimilation with event search for optimal history using cost function; 2) assessment of applicability of existing TAMC methods; and 3) use of the Genetic Algorithm (GA) approach (RUNDLE *et al.*, 2004; GIERING and KAMINSKI, 1997; RAWLINS, 1991a,b; BHATTACHARYYA *et al.*, 1999). All three of these tasks require that a cost function be defined quantifying the fit of the simulation data to the historic data through time (RUNDLE *et al.*, 2006). These data types include earthquake occurrence time, location, and moment release, as well as surface deformation data obtained from InSAR, GPS, and other methods. The data will ultimately be assimilated into viscoelastic finite element and fault interaction models and into a Potts model (CHAIKIN and LUBENSKY, 1995) in which data must be assimilated to determine the universal parameters that defined the model. Once these parameters are determined, the Potts model represents the general form of a predictive equation of system evolution.

Datamining: Pattern Analysis as a General Course Graining

In many solid earth systems, one begins by partitioning (tesselating) the region of interest with a regular lattice of boxes or tiles. Physical simulations demonstrate that the activity in the various boxes is *correlated* over a length scale ξ, which represents the appropriate length scale for the fluctuations in activity. The *analysis* (as opposed to the simple *identification*) of space-time patterns of earthquakes and other earth science phenomena is a relatively new field. Several techniques of pattern analysis

have been recently developed and are being applied: *Eigenpattern Method* ("*Karhunen-Loeve*" or "*Principal Components*") (RUNDLE *et al.*, 2000, 2001, 2002; TIAMPO *et al.*, 2002) uses catalogs of seismic activity to define matrix operators on a coarse-grained lattice of sites within a spatial region. *Hidden Markov Models (HMM)* (RABINER, 1989; UEDA and NAKANO, 1998; SMYTH, 2000) provide a framework in which, given the observations and a few selected model parameters, it is possible to objectively calculate a model for the system that generated the data, as well as to interpret the observed measurements in terms of that model. *Wavelet-Based Methods* (DAUBECHIES, 1992; CHUI, 1992) emphasize the use of wavelets to 1) search for scale-dependent patterns in coarse-grained space-time data, and 2) analyze the scale-dependence of the governing dynamical equations, with a view towards developing a reduced set of equations that are more computationally tractable yet retain the basic important physics. One goal of QuakeSim is to automate *Ensemble Forecasting* (RUNDLE *et al.*, 1995, 1997, 2004; FISCHER *et al.*, 1997; EGOLF, 2000) using data assimilation and pattern recognition.

While pattern dynamics offers system characterization modeling on an empirical level, and fast multipoles and multigrid methods automatically calculate useful levels of coarse graining, we do not neglect physics-based techniques that enable abstractions at various scales through direct modeling. The advantages of such methods are numerous: clues to direct physical relationships, insights into potential coarse-graining and coupling, and meaningful interpolation and visualization. For fault systems, boundary elements, finite elements and hybrid coupled systems are the most promising. Implementation of these technologies is required to enable realistic simulations using space-based and other data.

On Demand Science

The QuakeSim/SERVO 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 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 (OKADA, 1985). 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 following facts were widely available within two days of the earthquake (see e.g., HAUKSSON *et al.*, 2004). 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 within the first days. Additionally, no surface rupture had been identified. Seismic inversion indicated 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.

In another application of the portal, GPS data show 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. 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 (e.g., DOLAN et al., 2003).

The portal was also used to calculate displacements from the 2004, December 26 magnitude 9.0 earthquake off the coast of Sumatra. For this study the module disloc was used to calculate and map surface displacements from the earthquake mainshock. The results were calculated for a grid, but can be used to compare to the regional GPS observations. This type of analysis can yield information about the slip distribution of the earthquake as well as the mechanical properties of the crust and mantle.

Acknowledgments

REFERENCES

ARBIB, M. and GRETHE, J. eds., *Federating Neuroscience Databases, Computing the Brain: A Guide to Neuroinformatics* (Academic Press, 2001).

ATKAS, M., AYDIN, G., DONNELLAN, A., FOX, G., GRANAT, R., GRANT, L., LYZENGA, G., MCLEOD, D., PALLICKARA, S., PARKER, J., PIERCE, M., RUNDLE, J., SAYAR, A., and TULLIS, T. *iSERVO: Implementing the International Solid Earth Research Virtual Observatory by integrating computational grid and geographical information Web Services*, Pure appl. geophys., *163*, 2281–2296.

BHATTACHARYYA, J., SHEEHAN, A.F., TIAMPO, K., and RUNDLE, J.B. (1999), *Using genetic algorithms to model broadband regional waveforms for crustal structure in the western United States*, Bull. Seismol. Soc. Am. *89*, 202–214.

BOOTH, D., HAAS, H., MCCABE, F., NEWCOMER, E., CHAMPION, M., FERRIS, C., and ORCHARD, D. *Web Service Architecture*, W3C Working Group Note 11, February 2004. Available from http://www.w3.org/TR/ws-arch/.

CHAIKIN, P.M. and LUBENSKY, T.C., *Principles of Condensed Matter Physics* (Cambridge University Press, Cambridge UK, 1995).

CHUI, C.K. (Academic Press, Boston, 1992).

DAUBECHIES, I., *Ten Lectures on Wavelets* (SIAM, Philadelphia, 1992).

DOLAN, J.F., CHRISTOFFERSON, S.A., and SHAW, J.H. (2003), *Recognition of paleoearthquakes on the Puente Hills Blind Thurst Fault, California*, Sci. *300*, 115–118.

EGOLF, D.A. (2000), *Equilibrium regained: From nonequilibrium chaos to statistical mechanics,* Sci. *287*, 101–104.

FISCHER, D.S., DAHMEN, K., RAMANATHAN, S., and BEN-ZION, Y. (1997), *Statistics of earthquakes in simple models of heterogeneous faults*, Phys. Rev. Lett. *78*, 4885.

FOX, G. and HEY, T. *Grid Computing: Making the Global Infrastructure a Reality* (ed. Fran Berman) (John Wiley & Sons, Chichester, England, 2003) ISBN 0-470-85319-0.

GANNON, D., CHIU, K., GOVINDARAJU, M., and SLOMINSKI , A. (2002), *A revised analysis of the Open Grid Services Infrastructure*, Comput. Inform. *21*, 321–332.

GIERING, R. and KAMINSKI, T. (1997), *Recipes for adjoint code construction*, ACM Trans. Math. Software *23*, 437.

HAUKSSON, E., OPPENHEIMER, D., and BROCHER, T.M. (2004), *Imaging the source region of the 2003 San Simeon earthquake within the weak Franciscan subduction complex, central California*, Geophys. Res. Lett. *31*, L20607.

OKADA, Y. (1985), *Surface deformation due to shear and tensile faults in a half space*, Bull. Seismol. Soc. Am. *75*, 1135–1154.

RABINER, L.R. (1989), *A tutorial on hidden markov models and selected applications in speech recognition*, P. IEEE *77*, 257.

RAWLINS, G.J.E, *Foundations of Genetic Algorithms*, 1 (Morgan Kaufman, San Mateo, CA, 1991a).

RAWLINS, G.J.E, *Foundations of Genetic Algorithms*, 2 (Morgan Kaufman, San Mateo, CA, 1991b).

RUNDLE, J.B., KLEIN, W., GROSS, S., and TURCOTTE, D.L. (1995), *Boltzmann fluctuations in numerical simulations of nonequilibrium threshold systems*, Phys. Rev. Lett. *75*, 1658.

RUNDLE, J.B., KLEIN, W., GROSS, S., and FERGUSON, C.D. (1997), *The traveling density wave model for earthquakes and driven threshold systems*, Phys. Rev. E *56*, 293.

RUNDLE, J.B., KLEIN, W., and K. TIAMPO (2000), *Linear pattern dynamics in nonlinear threshold systems*, Phys. Rev. E *61*, 2418–2431.

RUNDLE, P.B., RUNDLE, J.B., TIAMPO, K.F., MARTINS, J.S.S., MCGINNIS, S., and KLEIN, W. (2001), *Nonlinear network dynamics on earthquake fault systems*, Phys. Rev. Lett. *87*, 148–501.

RUNDLE, J.B., TIAMPO, K.F., KLEIN, W., and MARTINS, J.S.S. (2002), *Self-organization in leaky threshold systems: The influence of near mean field dynamics and its implications for earthquakes, neurobiology and forecasting*, Proc. Nat. Acad. Sci. USA *99*, Supplement 1, 2514–2521.

RUNDLE, J.B., TURCOTTE, D.L., SAMMIS, C., KLEIN, W., and SHCHERBAKOV, R. (2003), *Statistical physics approach to understanding the multiscale dynamics of earthquake fault systems (invited)*, Rev. Geophys. Space Phys. *41*, DOI 10.1029/2003RG000135.

RUNDLE, J.B., RUNDLE, P.B., DONNELLAN, A., and FOX, G.C. (2004), *Gutenberg-Richter statistics in topologically realistic system-level earthquake stress-evolution simulations*, Earth, Planets and Space *56*, 761–771.

RUNDLE, P.B, RUNDLE, J.B., TIAMPO, K.F., DONNELLAN, A., and TURCOTTE, D.L. *Virtual California: Fault Model, Frictional Parameters, Applications,* Pure appl. geophys. *163*, 1819–1846.

SHETH, A.P. and LARSON, J.A. (1990), *Federated database systems for managing distributed, heterogeneous, and autonomous databases*, Comput. Sur. *22*, 183–236.

SMYTH, P. (2000), *Model selection for probabilistic clustering using cross-validated likelihood*, Statistics and Computing *10*, 63–72.

TALIA, D. (2002), *The Open Grid Services Architecture: Where the grid meets the Web*, IEEE Internet Computing *6*, 67–71.

TIAMPO, K.F., RUNDLE, J.B., MCGINNIS, S. GROSS, S., and KLEIN, W. (2002), *The phase dynamics of earthquakes, Implications for forecasting in southern California*, Europhys. Lett. *60*, 481–487.

UEDA, N. and NAKANO, R. (1998), *Deterministic annealing em algorithm*, Neural Networks *11*, 271.

WIEDERHOLD, G. (1994), *Interoperation, Mediation, and Ontologies*, Proceedings of the International Symposium on Fifth Generation Computer Systems, Tokyo, Japan, pages 33–84.

To access this journal online:
http://www.birkhauser.ch