Pageoph Topical Volumes

Computational Earthquake Physics: Simulations, Analysis and Infrastructure, Part II

Edited by
Xiang-chu Yin
Peter Mora
Andrea Donnellan
Mitsuhiro Matsu'ura

Birkhäuser

Computational Earthquake Physics: Simulations, Analysis and Infrastructure, Part II

Edited by
Xiang-chu Yin
Peter Mora
Andrea Donnellan
Mitsuhiro Matsu'ura

2007

Birkhäuser Verlag
Basel · Boston · Berlin

Reprint from Pure and Applied Geophysics
(PAGEOPH), Volume 163 (2006) No. 11–12

Editors

Xiang-chu Yin
Institute of Earthquake Science
Chinese Earthquake Administration
Beijing 10036
China

e-mail: xcyin@public.bta.net.cn

Peter Mora
Earth Systems Science
Computational Centre
The University of Queensland
Brisbane, 4072 Qld
Australia
e-mail: director@esscc.uq.edu.au

Andrea Donnellan
Jet Propulsion Laboratory
4800 Oak Grove
Drive-Mail Stop 183-335
Pasadena, CA 91109
USA

e-mail: andrea.donnellan@jpl.nasa.gov

Mitsuhiro Matsu'ura
Dept. of Earth and Planetary Science
The University of Tokyo
7-3-1 Hongo
Bunkyo-Ku
Tokyo 113-0033
Japan
e-mail: matsuura@eps.s.u-tokyo.ac.jp

Library of Congress Control Number: 2006935063

Bibliographic information published by Die Deutsche Bibliothek:
Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliographie; detailed bibliographic data is available in the internet at <http://dnb.ddb.de>

ISBN 3-7643-8130-2 Birkhäuser Verlag, Basel – Boston – Berlin

ISBN 10: 3-7643-8130-2
ISBN 13: 978-3-7643-8130-1

e-ISBN-3-7643-8131-8

9 8 7 6 5 4 3 2 1

Contents

Pure appl. geophys. 163 (2006) 2259–2261
0033–4553/06/122259–3
DOI 10.1007/s00024-006-0138-7

Pure and Applied Geophysics

Computational Earthquake Physics
PART II: Introduction

Xiang-chu Yin,[1,2] Peter Mora,[3] Andrea Donnellan,[4] and Mitsuhiro Matsu'ura[5]

Large earthquakes are catastrophic natural disasters which can potentially cause massive casualties and huge property loss. In the beginning of the new century, large earthquakes violently struck the world, especially in the Asia-Pacific region. Nearly 300,000 people were killed by the magnitude 9.0 Northern Sumatra Earthquake and tsunami, and the magnitude 7.8 Pakistan earthquake of October 8th, 2005, which resulted in 90,000 deaths. In the meantime, there has been great progress in computational earthquake physics. New understanding of earthquake processes, numerous ideas on earthquake dynamics and complexity, next-generation numerical models and methods, higher performance supercomputers, and new data and analysis methods are emerging. These include the SERVO grid and iSERVO, LSM (Lattice Solid particle simulation Model); Australian Computational Earth Systems Simulator (ACcESS); Japan's Earth Simulator; GeoFEM; GeoFEST; QuakeSim; LURR (Load-Unload Response Ratio); earthquake Critical Point Hypothesis, PI (Pattern Informatics), Critical Sensitivity, friction laws and seismicity, episodic tremor, the Virtual California model, interaction between faults and the interactions between earthquakes, ROC (Receiver Operating Characteristic), SMDM (Statistical Mesoscopic Damage Mechanics) and MFEM (Multi-scale Finite-Element Model), among others.

In 2005, the world lost a seminal figure in the study of earthquakes, Professor Keiiti Aki. One of the world's leading seismologists and an important contributor to ACES, Professor Aki had planned to deliver an important lecture at the 2005

[1] Institute of Earthquake Science, Chinese Earthquake Administration, Beijing, 100036. E-mail: xcyin@public.bta.net.cn
[2] Chinese Academy Sciences, State Key Laboratory of Nonlinear Mechanics, Institute of Mechanics, Beijing, 100080. E-mail: yin@LNM.imech.ac.cn
[3] Earth Systems Science Computational Centre, The University of Queensland, Brisbane, 4072 Qld, Australia. E-mail: director@esscc.uq.edu.au
[4] Jet Propulsion Laboratory, 4800 Oak Grove Drive - Mail Stop 183-335, Pasadena 91109, California. E-mail: andrea.donnellan@jpl.nasa.gov
[5] Department of Earth and Planetary Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-0033. E-mail: matsuura@eps.s.u-tokyo.ac.jp

European Geophysical Union meeting titled "Opening of a New Era for Earthquake Prediction Research" shortly before his death. In the abstract to his lecture Professor Aki states: "The Hagiwara symposium ... held at the Sapporo IUGG in 2003...opened a new era for the earthquake prediction research. We now find, however, a growing recognition among earthquake scientists…that modeling is as important as monitoring for a healthy development of earthquake prediction research as a branch of Physical Science." Many of the results to which Professor Aki refers represent the outcome of ACES-related research and activities, and these ideas will be presented in this volume.

The APEC Cooperation for Earthquake Simulation (**ACES**) [1], was endorsed by APEC (Asia-Pacific Economic Cooperation) in 1997. The inaugural workshop, the second and third workshops of ACES were held in 1999, 2000 and 2002, respectively [2–5]. During the week of July 9–14, 2004, China hosted the 4th ACES workshop [6,7] in Beijing. The 4th ACES Workshop was a milestone for ACES as unanimous agreement was reached for the follow-on to ACES, the ACES-iSERVO [8] International Institute (International Solid Earth Research Virtual Observatory Institute). A colloquium on iSERVO was held at the 4th ACES Workshop leading to broad endorsement for establishment of the ACES-iSERVO institute by the international group of over 100 scientists in attendance. The subsequent signing of a formal agreement to establish the institute will initiate a frontier international research institute on simulating the solid earth. The institute's focus will be development of predictive capabilities for solid earth phenomena via simulation and breakthrough science using the computational simulation capabilities aimed at understanding solid earth system complexity.

This special issue is divided into two parts. The first part (Part I) was issued in Pure and Applied Geophysics, Vol. 163, No. 9, 2006, which incorporates Micro-scale Simulation, Macro-scale Simulation and Scaling Physics. The second part (Part II, this issue) incorporates Computational Environment and Algorithms, Data Assimilation and Understanding, Model Applications and iSERVO. Topics covered range from iSERVO and QuakeSim; implementing the international solid earth research virtual observatory by integrating computational grid and geographical information web services; LURR (Load-Unload Response Ratio) described in six papers involving this promising earthquake forecasting model; Pattern Informatics and Phase Dynamics and their applications, which was also a highlighted in the Workshop; computational algorithms, including continuum damage models and visualization and analysis of geophysical data sets; evolution of mantle material; the state vector approach; and assimilation of data such as geodetic data, GPS data, and seismicity and laboratory experimental data.

The 4th ACES workshop (2004) was planned by the ISB (International Science Board) of ACES. Consisting of Peter Mora, Xiang-chu Yin, Mitsuhiro Matsuura, Andrea Donnellan and Jean-Bernard Minster, and was hosted by the Institute of Earthquake Science, China Earthquake Administration and LNM (State Key

Laboratory of Nonlinear Mechanics), Institute of Mechanics, Chinese Academy of Sciences.

We appreciate our sponsors including China Earthquake Administration, Chinese Academy of Sciences, Chinese National Natural Science Foundation, Chinese Ministry of Science and Technology, Chinese Ministry of Finance, Australia-China Fund, Australian Commonwealth Department of Education, Science and Training, Australian Research Council. The University of Queensland and Earth Systems Science Computational Centre, Australian Computational Earth Systems Simulator Major National Research Facility, Japan Society for the Promotion of Science, Research Organization for Information Sciences and Technology (RIST), the National Aeronautics and Space Administration (NASA) and the US National Science Foundation (NSF) and the United States Geological Survey (USGS).

References

[1] http://www.esscc.uq.edu.au/ACES.
[2] Microscopic and Macroscopic Simulation: Towards Predictive Modeling of the Earthquake Process (2000), eds. Mora, P., Matsu'ura, M., Madariaga, R., and Minster, J-B,. Pure Appl. Geophys. *157*, 11/12, 1817–2383.
[3] 2nd ACES Workshop Proceedings, eds. Mitsuhiro Matsuura, Kengo Nakajima, and Peter Mora (2001) (ACES, Brisbane, Australia, ISBN 186499 510 6, GoPRINT).
[4] Computational Earthquake Science, Part I (2004), eds. Andrea Donnellan, Peter Mora, Mitsuhiro Matsu'ura, and Xiang-chu Yin, Pure Appl. Geophys. *161*, 9/10, 1823–2118.
[5] Computational Earthquake Science, Part II, eds. Andrea Donnellan, Peter Mora, Mitsuhiro Matsu'ura, and Xiang-chu Yin, Pure Appl. Geophys. *161*, 11/12, 2119-2460.
[6] http://www.aces-workshop-2004.ac.cn/.
[7] 4th ACES Workshop Proceedings, eds. Xiang-chu Yin, Peter Mora, Andrea Donnellan and Mitsuhiro Matsuura (2006), (ACES, Brisbane, Australia, GoPRINT).
[8] Special issue on *Earth System Modeling* (2005), Guest Editor: John B. Rundle, Paper's authors: Peter Mora, Xiang-chu YIN, Mitsuhiro Matsuura, Lisa B. Grant and Andrea Donnellan *et al.*, Computing in Science and Engineering, vol. *7*, 24–66.

Pure appl. geophys. 163 (2006) 2263–2279
0033–4553/06/122263–17
DOI 10.1007/s00024-006-0126-y

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

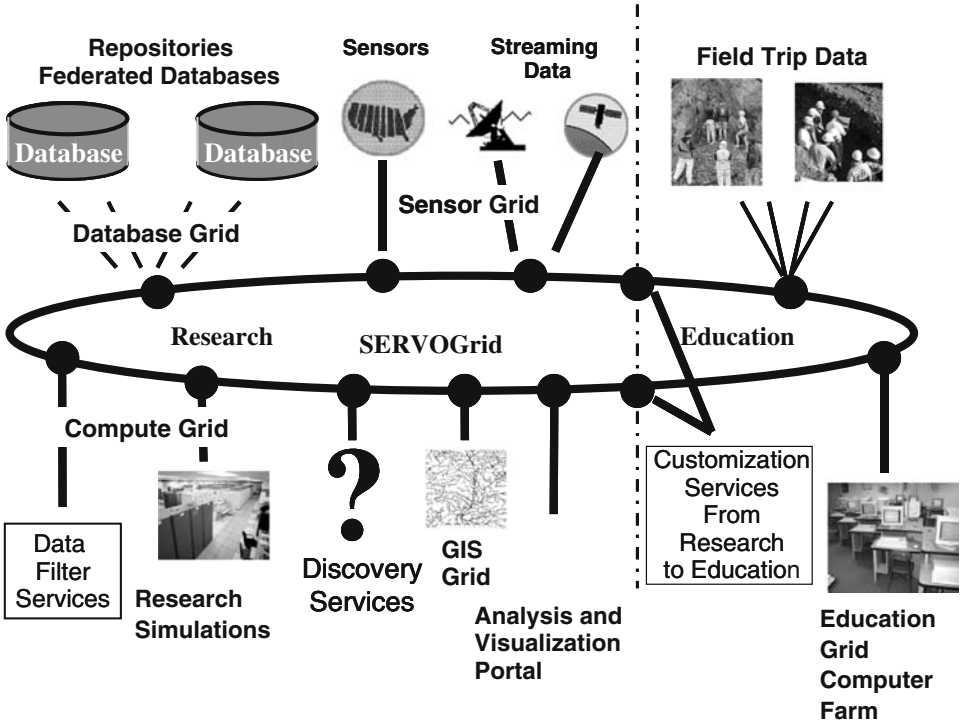


Figure 1
SERVO Grid with typical grid nodes.

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

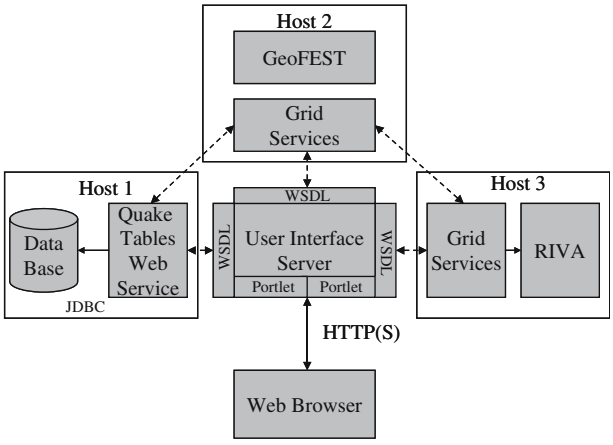


Figure 2
The QuakeSim portal supports distributed web services for interacting with databases (such as the QuakeTables fault database), remote applications (such as GeoFEST), and batch visualization (such as RIVA). The portal creates user interfaces for browser-based users through an aggregation of web 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.

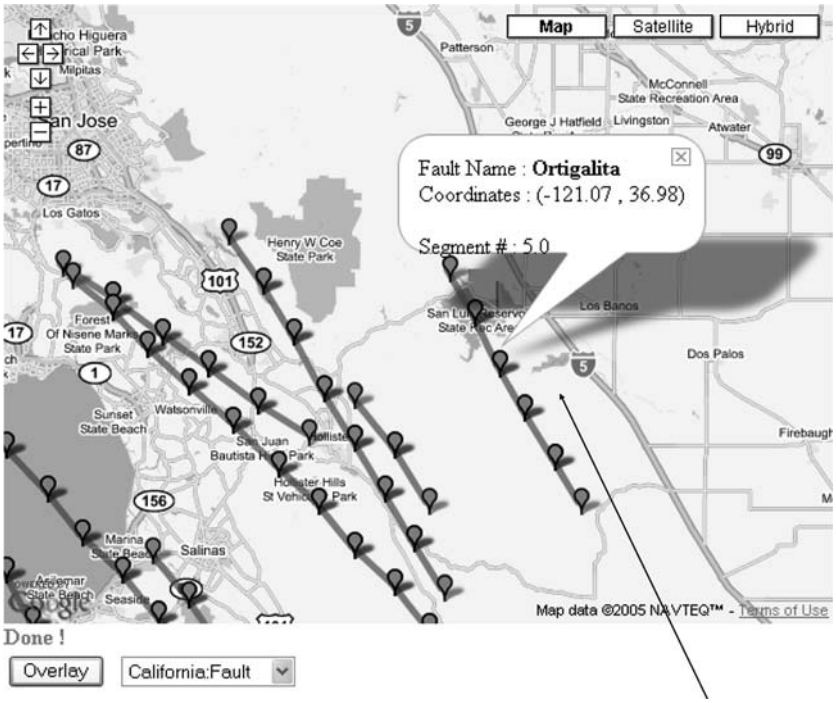


Figure 3
The QuakeTables fault database can be navigated using interactive Google maps as well as through search query HTML forms. Faults are typically stored by segment number, so map interfaces are convenient for both verifying continuity and setting up input files for computing problems.

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

We thank Peggy Li for integrating the visualization tools with QuakeSim and Servo, Charles Norton for adapting GeoFEST to the PYRAMID library, and our project review board led by Brad Hager from MIT. We thank the graduate students, Anne Chen, Sang-Soo Song, and Shan Gao from USC for their work on the database systems, and Miryha Gould, Gabriela Noriega-Carlos and Maggie Ta of UC Irvine for their work on the fault database. Robert Shcherbakov from UC Davis assessed the data assimilation methods. Teresa Baker and Maggi Glasscoe have carried out science models using this system and developed documentation for various components. Michele Judd has proved invaluable in the management of this project. This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, and at the institutions of the co-authors under contract with NASA through their Computational Technologies and Advanced Information Systems Technologies programs.

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.

(Received December 30, 2004, revised November 19, 2005, accepted November 21, 2005)

Pure appl. geophys. 163 (2006) 2281–2296
0033–4553/06/122281–16
DOI 10.1007/s00024-006-0137-8

Pure and Applied Geophysics

iSERVO: Implementing the International Solid Earth Research Virtual Observatory by Integrating Computational Grid and Geographical Information Web Services

Mehmet Aktas,[1] Galip Aydin,[1] Andrea Donnellan,[2] Geoffrey Fox,[3] Robert Granat,[4] Lisa Grant,[5] Greg Lyzenga,[6] Dennis McLeod,[7] Shrideep Pallickara,[1] Jay Parker,[8] Marlon Pierce,[9] John Rundle,[10] Ahmet Sayar,[1] and Terry Tullis[11]

Abstract—We describe the goals and initial implementation of the International Solid Earth Virtual Observatory (iSERVO). This system is built using a Web Services approach to Grid computing infrastructure and is accessed via a component-based Web portal user interface. We describe our implementations of services used by this system, including Geographical Information System (GIS)-based data grid services for accessing remote data repositories and job management services for controlling multiple execution steps. iSERVO is an example of a larger trend to build globally scalable scientific computing infrastructures using the Service Oriented Architecture approach. Adoption of this approach raises a number of research challenges in millisecond-latency message systems suitable for internet-enabled scientific applications. We review our research in these areas.

Key words: Web Services, computing web portals, computational grids, grid computing, earthquake simulation.

[1] Community Grids Lab, Indiana University, Bloomington, 47404-3730, IN.
[2] NASA Jet Propulsion Laboratory, Mail Stop 183-335, 4800 Oak Grove Drive, Pasadena, CA 91109-8099, USA. E-mail: Donnellan@jpl.nasa.gov
[3] Community Grids Laboratory, Departments of Computer Science, Physics, and School of Informatics, Indiana University, Bloomington, Indiana 47404-3730, USA. E-mail: gcf@indiana.edu
[4] NASA JPL, Mail Stop 126-347, 4800 Oak Grove Drive, Pasadena, CA 91109-8099. E-mail: Robert.Granat@jpl.nasa.gov
[5] Environmental Analysis and Design, University of California-Irvine, Irvine, California 92697-7070, USA. E-mail: lgrant@uci.edu
[6] NASA JPL, Mail Stop 126-347, 4800 Oak Grove Drive, Pasadena, CA 91109-8099. E-mail: Gregory.Lyzenga@jpl.nasa.gov
[7] University of Southern Califoria, Mail Code 0781, 3651 Trousdale Parkway, Los Angeles, CA 90089-0742, USA. E-mail: mcleod@pollux.usc.edu
[8] NASA Jet Propulsion Laboratory, Mail Stop 238-600, 4800 Oak Grove Drive, Pasadena, CA 91109-8099, USA. E-mail: jay.w.parker@jpl.nasa.gov
[9] Community Grids Laboratory, Indiana University, Bloomington, Indiana 47404-3730, USA. E-mail: mpierce@cs.indiana.edu
[10] Department of Physics, University of California-Davis, One Shields Avenue, Davis, CA 95616-8677, USA. E-mail: rundle@physics.ucdavis.edu
[11] Department of Geological Sciences, Brown University, Providence, RI 02912-1846, USA. E-mail: Terry_Tullis@brown.edu

Introduction

In this paper we describe the architecture and initial implementation of the International Solid Earth Research Virtual Observatory (iSERVO) (quake-sim.jpl.nasa.gov). We base our design on a globally scalable. distributed computing infrastructure (often termed "cyber-infrastructure" or simply "Grid infrastructure" (FOSTER and KESSELMAN, 2003; BEHRMAN *et al.*, 2003; ATKINS *et al.*, 2003) that enables on-line data repositories, modeling and simulation codes, data mining tools, and visualization applications to be combined into a single cooperating system. We build this infrastructure around Web Services-based approach.

Challenges for Solid Earth Research

The Solid Earth Science Working Group of the United States National Aeronautics and Space Administration (NASA) has identified several challenges for Earth Science research (SOLOMON, 2002). Particularly relevant for iSERVO are the following:

- How can the study of strongly correlated solid earth systems be enabled by space-based data sets?
- What can numerical simulations reveal about the physical processes that characterize these systems?
- How do the interactions in these systems lead to space-time correlations and patterns?
- What are the important feedback loops that mode-lock the system behavior?
- How do processes on a multiplicity of different scales interact to produce the emergent structures that are observed?
- Do the correlations allow for the capability to forecast the system behavior?

In order to investigate these questions, we need to couple numerical simulation codes and data mining tools to observational data sets. These observational data (including crustal fault data from the literature, GPS data, and seismic activity data) are now available on-line in internet-accessible forms, and the quantity of this data is expected to grow explosively over the next decade.

The challenges in solid earth modeling motivate a number of interesting research and development issues in distributed computer science and informatics. Key among these are providing programmatic access to distributed data sources; coupling remote data sources to application codes, including automated searching and filtering; coupling of complementary application codes that are deployed on geographically separated host computers; and providing human level interfaces to these remote services.

We note that the services described in this paper are different from, but complementary with, more traditional code parallelization techniques in high performance computing. As we discuss in more detail below, communication in parallel applications demands microsecond latencies, which can only readily be achieved in tightly coupled systems (such as clusters and "big iron" parallel machines of various flavors). As we describe below, these systems are best thought of as highly specialized services that communicate with each other on Internet timescales (milliseconds or longer).

The iSERVO team possesses a broad range of skills and tools that may be used to investigate solid earth research challenges. Team expertise includes the development of high performance modeling and simulation applications for both the study of large, interacting earthquake systems and the detailed study of individual fault properties; federated database and ontology design; geological characterization of faults; and high performance visualization codes. Welding all of these components into a common distributed computing infrastructure is the subject of this paper.

A Web Service Grid Architecture

Problems in managing distributed computing resources, applications, data and users have been studied for many years. Viewed collectively, when such systems are managed by different organizations, we have what is typically called a computational grid. Typical desired functionality in these systems includes remote command execution, data transfer, security, and high performance messaging. To scale globally, these systems must abandon tight coupling approaches such as distributed object systems and micro-second latency solutions such as MPI. Instead, they should adopt a Service Oriented Architecture (SOA) (BOOTH *et al.*, 2004) that is compatible with millisecond (or longer) communication speeds (FOX *et al.*, 2004). SOAs are implemented around two basic components: Service definition languages (which describe how to invoke the remote service) and message formats for over-the-wire transmissions. In iSERVO, we have adopted the Web Service approach to building an SOA: We use WSDL (CHRISTENSEN *et al.*, 2001) for service description and SOAP (GUDGIN *et al.*, 2003) for message formats. This use of XML for both service description and messaging provides programming language independence: The client does not know or need to know the implementation language of the service.

Web Service systems have an important design feature: Service implementations are decoupled from the user interface components. This enables us to build a number of different clients that can interact with the same remote service, and *vice versa*. Browser-based computing portals are a typical way of managing client user interfaces and have been the subject of research and development work for a number of years (FOX and HEY, 2002). Currently this field is undergoing a revolution as component-based portal systems are being widely adopted, and standard component programming

interfaces have been developed (ABDELNUR *et al.*, 2003; GANNON *et al.*, 2004). This so-called "portlet" approach enables reusability of components: Portals may be built out of standard parts that aggregate content and functionality from many different sources.

SOA and portal standards are not the only relevant standards for building systems such as iSERVO. The Open Geographical Information Systems (GIS) Consortium (OGC) (http://www.opengis.org) defines a number of standards for modeling earth surface feature data and services for interacting with this data. The data models are expressed in the XML-based Geography Markup Language (GML) (COX *et al.*, 2003), and the OGC service framework is being adapted to use the Web Service model. In this paper we describe implementation of GIS services to describe data relevant to the geophysical community (GPS, seismic events, and faults) that we then couple to more typical Grid services for code execution and file management.

Implementing iSERVO

We have implemented an initial set of services and portal components for addressing the problems described in the introduction. We have followed a Web Service-based Grid design described above that uses Web Service standards. The components of the system and their interactions are summarized in Figure 1. Users interact with remote services through a Web browser portal that is run by the User Interface Server (UIS). This portal generates dynamic web pages that collect input

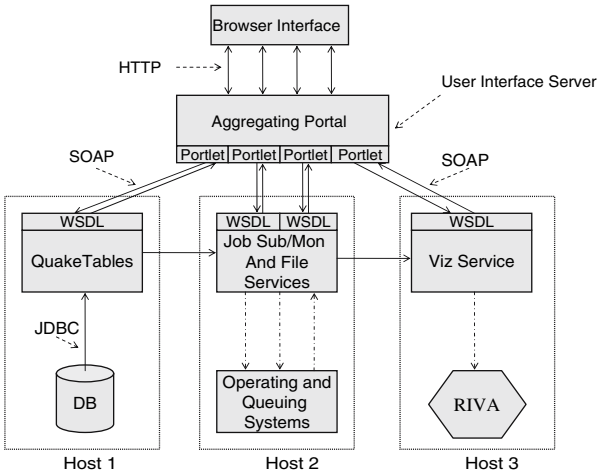


Figure 1
The architecture for the iSERVO portal and services uses Web Service and portal standards. JDBC stands for Java Database Connectivity and DB is an abbreviation for "database". "RIVA" is an example parallel visualization program. Other terms are defined in the text.

information from the user and deliver response messages. The UIS does not directly implement services such as job submission and file transfer. Instead, it maintains client proxies to these remote services. These proxies are responsible for generating the SOAP messages appropriate to the particular services' WSDL descriptions and for receiving the responses from the services. The UIS and most services are implemented in Java using the Apache Axis toolkit (http://ws.apache.org/axis/), but we have also implemented C++ services using gSOAP (VAN ENGELEN and GALLIVAN, 2002) for simple remote visualization.

A typical interaction involves the user selecting a code through the portal, setting up an input file in part through interactions with databases (such as the QuakeTables Fault Database (CHEN *et al.*, 2003; GRANT *et al.*, 2004), invoking the code and monitoring its progress, and having the output visualized through various third party tools of varying sophistication. These interactions are based on a dataflow model: Services communicate by exchanging data files, which must be pulled from one server to another.

In building iSERVO, we have implemented a number of innovations on the standard model components. The portlet component model normally assumes local portlets with content that navigates to other web sites (news portals such as Yahoo and CNN are examples). We have built extensions to this simple model to allow portlet content to be managed remotely, have its display maintained within its component window through a series of navigations, maintain HTTP sessions state with remote content, pass HTTP GET and POST variables, and support SSL security.

Basic iSERVO services include remote command execution, file upload and download, and host-to-host file transfer. We do not directly alter the geophysical applications included in the portal but instead follow a "proxy wrapping" approach (YOUN, *et al.*, 2004). Typically, applications require preprocessing of input files, post processing, and in general require task executions that are distributed across many different hosts. To support this sort of distributed service orchestration, we have developed a simple "workflow" service based on the Apache Ant project (http://ant.apache.org/). This service uses Ant as an engine that may be invoked remotely (as a service on Host 2 in Fig. 1) and may also coordinate service invocations on remote hosts, as needed to complete its task.

More information on the iSERVO, including code downloads, documentation, and information for accessing demonstrations is available from http://quake-sim.jpl.nasa.gov/.

iSERVO couples typical "Execution Grid" services such as described above with "Data Grid" services described in the following section. iSERVO applications work with many different data sources, and we have developed services to automate the coupling of these data to application services. A typical problem is as follows: The iSERVO application RDAHMM (a Hidden Markov Model application) (GRANAT, 2004) needs as input either GPS or seismic activity records. Both data sources are

available online, but there is no programmatic way of working with the remote data archives to, for example, filter data based on a user's criteria, assemble data sets from multiple archive records, or reformat the data based on the application's input formatting. Instead, a researcher typically downloads the data files and edits them by hand. To solve this problem, we have implemented GML-based services for describing these data records, and in the process we have unified several different data formats. These services allow the application user to build search filters on the desired data set (for example, returning seismic events larger than magnitude 5.0 within a particular region of interest since 1990). Additional filters reformat the data into one suitable for RDAHMM, and the data are then shipped to the location of the remote executable, which can then be invoked automatically. We thus replace the process of downloading and hand-editing the entire catalog.

Geographical Information System (GIS) Data Services

iSERVO data service requirements represent an excellent opportunity for further work leveraging open standards for services that will tie iSERVO to this larger community, allowing us to potentially incorporate many additional third-party data sources and tools. The NASA OnEarth project (http://onearth.jpl.nasa.gov/) is an excellent example of a GIS project that may be incorporated with iSERVO in the future. As part of our GIS development work, we are currently re-implementing the OGC standard services Web Feature Service and Web Map Service as iSERVO-compatible Web Services.

We note that the GIS community has other data model and service standards than those defined by the OGC: The commercial vendor ESRI provides another prominent set of data standards along with extensive client tools. Our adoption of OGC standards is intended to take advantage of the significant amount of freely available GIS data that already exists in OGC formats. More importantly, OGC standards define an open architecture that may integrate with Grid/Web service standards for distributed scientific computing discussed in the previous sections. We note further that ESRI and OGC interoperability tools already exist for obvious reasons, so adopting OGC standards does not preclude later integration of our data services with sophisticated ESRI software clients.

Advances in Geographical Information Systems (GIS) introduce several challenges for acquiring, processing and sharing data among interested parties. Different research groups, organizations, and commercial vendors develop their own data models and storage structures. Consequently the data is expressed in various formats and stored in various archives. These archives are often remotely accessible only through simple protocols (like FTP) that do not allow queries and filtering and which are difficult to integrate with geophysical applications. On the other hand the nature of the geographical applications requires seamless integration of spatial data from a

range of providers to produce layers, maps, etc. As a result we see the interoperability between applications and data stores as a significant goal for any GIS.

As an example of how this goal can be accomplished we describe our design of a Service Oriented Architecture for serving a subset of geographical information. We first review the existing data formats in our domain of interest and summarize our initial work for generating a common data format. The next section explains how we employed pure Web Services approach for data conversion, storage and query capabilities. The next section briefly discusses our experience and findings on XML and Relational Databases, and the user interfaces we created for testing the Web Services.

Goals of the Project

We designed a service-based architecture for solving the aforementioned challenges. However, before implementing this system we identified several goals to make the scope of this project clear. These goals are as follows:

- Making GPS and Seismic data easily available for humans and applications alike;
- Providing seamless access to data repositories and computing resources;
- Providing a common data format for each information area;
- Supporting search capabilities on the catalogs for certain properties, filtering the search results, and retrieving the results in various formats; and
- Integrating data with the scientific applications.

Figure 2 illustrates the major components of the system for achieving these goals. Existing public archives maintained by the Southern California Earthquake Center (SCEC) and the Southern California Integrated GPS Network (SCIGN) are accessed through Web Services that download and reformat the data into GML (steps 1 and 2 in Fig. 2). Data sources that we relied upon are more extensively documented at www.crisisgrid.org. We then store the converted data in either native XML or relational databases (step 3).

The above steps summarize administrative services that need to be performed once per external archive for initialization, followed by regular updates. Application users do not need to use these services. They do, however, make use of the search services (right-hand side of Fig. 2). These are also Web Services defined in WSDL and so may be accessed by various client programs.

Data Format Issues and GML

Depending on the data providers use, the catalogs are formatted in various ways. Any application or user who wants to make use of these legacy formats should understand what each column or data segment means and should write scripts to convert them into the target application's own legacy format. This introduces the tedious and resource consuming conversion problem.

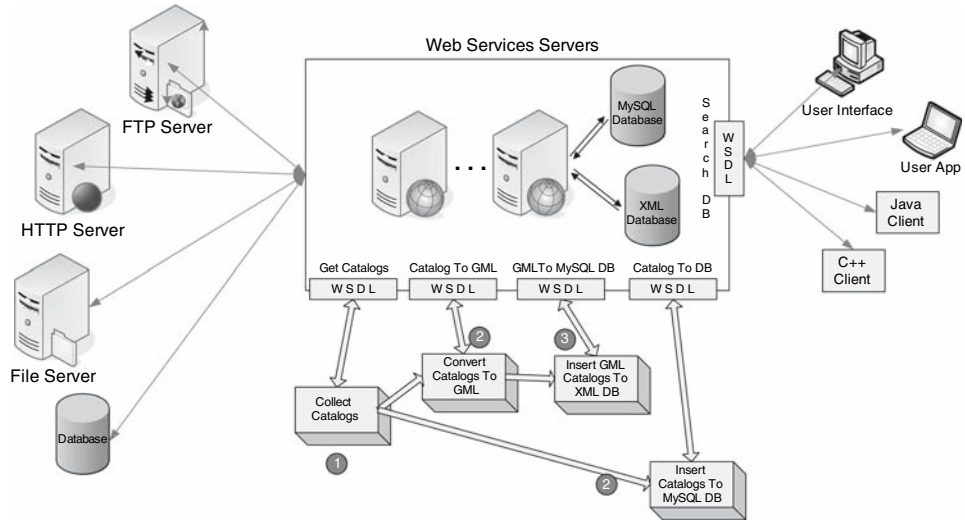


Figure 2
Major parts of the architecture and a sample workflow for processing geo-data using Web Services.

Therefore this was our initial challenge in making the geographic data easily available: *data in this domain come from different sources in different formats*. To solve this problem, we designed and implemented common data formats for each data type (GPS, fault, seismicity) that encapsulates several existing formats.

Our choice for the common data format was GML since it is widely accepted as a common exchange format for spatial information in the GIS community worldwide. GML is actually an extensive collection of integrated XML Schemas that collectively provide data models for describing geographic entities: Features, coordinate reference systems, geometry, topology, time, units of measure and generalized values.

Our GML Schemas support the following formats:

Seismic data formats:

SCSN, SCEDC, Dinger-Shearer, Haukkson

GPS data formats:

JPL, SOPAC, USGS

More details of these formats (both original and our GML compatible versions) are available from http://www.crisisgrid.org/html/servo.html. GML schema for earthquake faults (available in both GML 2 and GML 3) are also available.

The next step after creating a common exchange format for the catalogs was to design a service-based system compatible with the architecture of Figure 1 to process these data. The first thing necessary in the system was to collect data from various sources. GPS and seismicity catalogs are provided by several organizations and they are available to the public via FTP or HTTP servers. We wrote clients for retrieving

the catalogs and saving them in our server for further processing. These retrieval applications are also provided as Web Services so that any workflow scheme can schedule regular catalog updates.

The reference implementation of the user interface to the system allows a user to select one of the two paths to follow after acquiring the catalogs:

Convert the catalogs into GML format and save them as XML files. These files are then inserted into an XML database.

Insert the catalog data into relational database.

The choice in this step is introduced because of the use of both relational and XML databases in the system. We extensively tested and used both types of the databases.

Databases

Since the message exchange format of Web Services is XML and since we convert our data to XML at the beginning of the process, we initially implemented data storage using native XML databases. Implementations were tested with both the Berkeley XML Database system (Sleepycat) and Apache Xindice. However, after extensive testing we found the XML databases performance to be unusable for GML records of the size needed to store existing on-line catalogs such as the Southern California Earthquake Center's seismic activity records. We re-implemented the storage mechanism (without changing the WSDL and so without updating the clients) to use a common, open-source relational database (MySQL), which demonstrated acceptable query time.

Searching the Catalogs

We provide a Web Service for searching the catalog databases. The results are sent to the client application as GML documents. Since there is only one GML Schema for a particular data group (GPS or Seismicity) it is easy to validate and parse the received GML document and extract the information. XPATH (CLARK and DEROSE, 1999) is used as the query language for the XML Databases.

Status

Documentation and software for this work are available from www.crisisgrid.org.

User Interfaces

We have created a set of user interfaces to the Web Services for demonstration purposes. These interfaces also show the steps for making a set of geographic data available to search and retrieve via web services. The search client enables users to create database queries via web page forms. The search results are shown to the user as a simple text page. We use the XML Pull Parser (www.extreme.indiana.edu/xgws/

xsoap/xpp) to quickly extract data from the received GML file and create the web page output.

Our architecture provides several WSDL interfaces to enable both human and application clients to use the services. WSDL does not require the client to be bound to a particular programming language or environment, thus supporting client tools in different programming languages such as Java classes or C++ applications. To demonstrate this feature we have created a simple C++ client using gSOAP (VAN ENGELEN and GALLIVAN, 2002) along with our JSP interfaces and Java clients.

Designing OGC Web Services

As we have described above, GML is a data modeling language that can be used to encode geophysical data. We may then store these data in various archival systems and design Web Services that can query, retrieve, and update the data. These web services are compatible with the "Execution Grid" services illustrated in Figure 1.

These Web Services, because they use a generic data model, may be standardized and generalized. The OpenGIS Consortium defines specifications for several such services, with the Web Feature Service and the Web Map Service as two prominent examples. These services, unfortunately, are not designed to be Web Service compatible (they do not use WSDL or SOAP but rather lower level HTTP GET/POST conventions for messaging). In order to adapt these services to the QuakeSim architecture while taking advantage of existing OGC resources, we have redesigned these OGC services to use Web Service standards.

The Web Feature Service (WFS) (VRETANOS, 2002) describes standards to publish, update, and delete geographic features, such as faults and GPS stations. We designed a Web Service version of OGC WFS that provides WSDL interfaces for the required capabilities. Instead of using HTTP Post, the user or the client application communicates with the WFS using SOAP messaging. The results of the requests are sent to the user as GML documents.

One important property of the WFS is that it can serve multiple feature types. Different features from different data stores are integrated with the WFS and the clients do not realize that the features are retrieved from several sources.

We also are implementing a Web Service version of Web Map Service to generate maps (DE LA BEAUJARDIERE, 2004). The Web Map Service obtains data from various Web Feature Services. We are also building bridging services that allow our Web Service-compatible WMS to interact with non-Web Service versions of WMS. This will allow us to make use of extensive existing mapping resources, such as the NAS OnEarth project (http://onearth.jpl.nasa.gov/). The interaction between these two services is based on SOAP messaging because of the Web Service standards. This work is currently under development.

Status

The initial Web Feature Service implementation work is completed and is being packaged for availability at www.crisisgrid.org. The Web Map Service implementation is in development but links to demonstrations will be available from www.crisisgrid org.

GIS Information Services

Services such as the Web Map and Web Feature service, because they are generic, must provide additional, descriptive metadata in order to be useful. The problem is simple: a client may interact with two different Web Feature Services in exactly the same way (the WSDL is the same), but the Web Feature Services may hold different data. One, for example, may contain GPS data for the Western United States while the other has GPS data for Northern Japan. Clients must be able to query information services that encode (in standard formats) all the necessary information, or metadata, that enables the client to connect to the desired service. This is an example of the very general problem of managing information about Web Services. To address these problems, we are designing a general purpose information system, the Fault Tolerant High Performance Information System (FTHPIS), that we are applying initially to problems in GIS information management.

An approach to solve the problem of locating resources of interests in iSERVO environment was first introduced in (AKTAS *et al.*, 2004). This approach suggests a centralized discovery model by utilizing semantic web technologies to locate iSERVO resources where resources were limited to iSERVO codes and data.

In a FTHPIS, there is a need for registry services to make the information about services available. We use the Universal Description, Discovery, and Integration (UDDI) (BELLWOOD *et al.*, 2003) specifications in our design as a centralized registry. UDDI offers users a unified and systematic way to find service providers through a centralized registry of services. We design an extension to existing UDDI Specifications in order to provide dynamically updated service registry data.

UDDI provides a centralized approach to the Web Services registry research problem. However, one should be able to locate a Web Services satisfying a query in a decentralized, dynamically changing, distributed environment as well. To do this, discovery architecture needs to be defined. As we consider the volatile behavior of Web Services, such a decentralized approach should provide dynamic discovery of services where the temporarily connected Web Service can be discovered. To this end, we design our implementation based on WS-Discovery Specifications which was recently released.

WS-Discovery (BEATTY *et al.*, 2004) defines a multicast protocol to locate services. It allows dynamic discovery of services in *ad hoc* and managed networks.

However, WS-Discovery does not define a metadata model to describe Web Service. To this end, we also define an abstract metadata model for Web Service. This will allow us to pose more complex queries for WS-Discovery.

We classify metadata associated with Web Services as dynamic metadata and static metadata. Dynamic metadata is the session (or state) metadata generated by the individual interactions with Web Services. Such metadata describes the context of the session and has a lifetime. There are different approaches specifying session metadata. For instance, WS-Context (BUNTING *et al.*, 2003) provides an abstract context defining such metadata. Static metadata is the metadata describing a Web Service profile such as its usage cost, availability, bandwidth, computing power, storage capability, etc. We extend existing UDDI and WS-Context specifications in order to associate metadata with Web Service descriptions.

A FTHPIS consists of information services. An information service is a Web Service that provides registry to make the information about services available and also maintains a repository of context information of web services. An information service combines both WS-Context service and WS-Registry service. WS-Registry part of an information service extends existing UDDI technology where one can annotate service descriptions with metadata describing "Quality of Service" aspects of services. The WS-Context part of an information service extends the existing WS-Context service where one can track contexts shared between multiple participants in Web Service interactions. In Figure 3, we show our design where an Information Service provides a gateway to UDDI Registry and WS-Context replica groups. Our

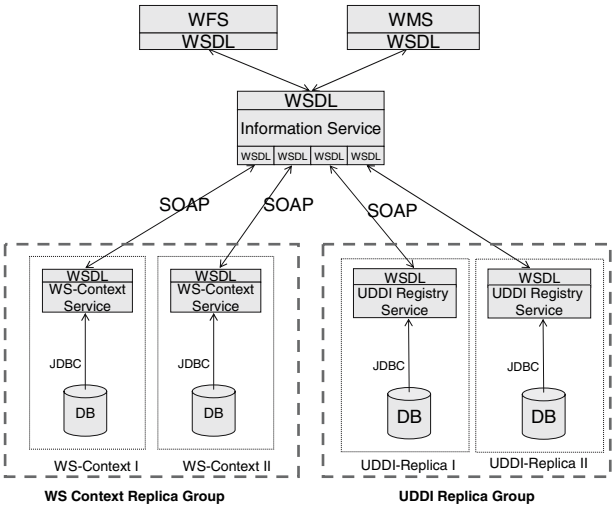


Figure 3
Interaction between an Information Service and OGC Web Services such as Web Map Service and Web Feature Service.

design suggests replication of databases both for registry of services or contextual information replicated with a sequential consistency model and replicated-write consistency protocol in order to provide fault tolerance and avoid single point of failure. We plan on utilizing a caching mechanism to improve performance and reduce the response latency to service discovery queries. In our design for FTHPIS, we suggest each information service peer caches SOAP messages as (query, respond) message pairs. We use read-only caches to improve response time, so updates operation can only be performed on the servers where the origin of the cached data resides. In order to provide cache-coherence we simply suggest propagating the updates to caches.

We illustrate a motivating scenario in which Information Services interacts with Web Map Services and Web Feature Services in Figure 4. In this scenario, Web Feature Services are published into the UDDI-Registry. Each Web Feature Service provides data layers corresponding to geographic entities. An important challenge is that UDDI does not natively support registry of services with a bounding box corresponding to a data layer and representing a location of interest. To overcome this problem, we use a standard capability of UDDI registries which is to classify service entries according to predefined taxonomies. We use geographic taxonomies to classify UDDI service entries based on spatial coverage. This methodology allows us to make coordinate-based spatial queries on the UDDI-Registry. In Figure 4, Web Map Services interact with the Information Service to ascertain available WFS (data services) satisfying the data requirements of a map. As the Information Service responds to a query of WMS with metadata of services satisfying the query, WMS can then start interacting with corresponding WFS to acquire the data layers needed to create maps.

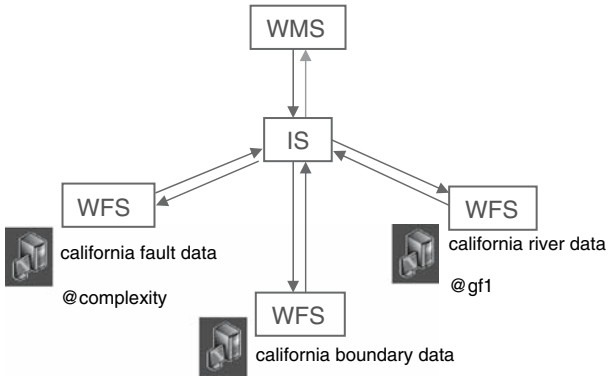


Figure 4
Interaction between an Information Service (IS) and OGC Web Services such as Web Map Service (WMS) and Web Feature Service (WFS).

Status

Our information system is currently under development.

Future Directions for iSERVO

Web Services in the SOA approach communicate with SOAP messages in a loosely coupled fashion. Such systems demand a number of features: Fault tolerance, reliable messaging, message level security (such as authorization and encryption), and message virtualization for firewall tunneling. We term this general class of messages as the Web Service "Internet-On-Internet" (IOI) problem: Many Web Service standards reimplement common TCP/IP features within the SOAP message. We see this as an interesting development, as it allows us to use a messaging system infrastructure (which we do control) to provide quality of service that is independent of the underlying network (which we do not control). We have implemented such a messaging system infrastructure, NaradaBrokering (http://www.naradabrokering.org), and are extending it to support Web Service invocations natively. This is described in more detail in Fox *et al*. (2004).

A common objection to the Web Service approach is that it is too slow. Message speeds across network connections are on the order of milliseconds at best, and so unsuitable for classic metacomputing. For typical iSERVO applications, this is not an issue: Model runs may take several hours or days to complete. The applications themselves may be deployed on clusters or supercomputers and may be parallelized by traditional techniques; we treat such applications as a single service component (Youn *et al*., 2004). However, there are classes of problems, particularly in interactive remote visualization and high performance transfers of large data sets, in which maximum network performance is needed. We are currently researching this within the NaradaBrokering system. The IOI approach for Web Services will allow us to replace TCP/IP with much more efficient UDP transmissions, while retaining desirable TCP/IP features in the SOAP messages.

On top of the IOI infrastructure, we must provide information services: How can one encode machine readably what a particular service in Figure 1 actually does? What data do the service provide or require? How do other components in the system discover it? How can it be classified? How can complicated service interactions be coordinated? We term the higher level information Grid that manages this sort of information as the "Context and Information Environment" (CIE). The information system described in this paper is an implementation of some of the foundation-level services that are needed to build a CIE. All the information requirements that we have enumerated are part of a larger problem in metadata management. In Web and Grid Services, this is an open problem with many competing solutions. iSERVO represents an excellent test case of a real Grid with real information system requirements that can be used to validate the competing solutions.

Acknowledgments

This work was funded by the Computational Technologies Program and the Advanced Information Systems Technology Program, both of NASA's Earth Science Technology Office.

References

Abdelnur, A., Chien, E., and Hepper, S., eds (2003), *Portlet Specification 1.0*. Available from http://www.jcp.org/en/jsr/detail?id = 168.

Aktas, M.S., Pierce, M., Fox, G., and Leake, D. (2004), *A We Based Conversational Case-Based Recommender System for Ontology Aided Metadata Discovery,* 5th IEEE/ACM International Workshop on Grid Computing (GRID 2004).

Atkins, D., Droegemeier, K., Feldman, S., Garcia-Molina, H., Klein, M., Messerschmitt, D., Messina, P., Ostriker, J., and Wright, M. (2003), *Revolutionizing Science and Engineering through Cyberinfrastructure,* Report of the National Science Foundation Blue-Ribbon Advisory Panel on Cyberinfrastructure. Available from http://www.cise.nsf.gov/sci/reports/atkins.pdf.

Beatty, J., Kakivaya, G., Kemp, D., Kuehnel, T., Lovering, B., Roe, B., St. John, C., Schlimmer, J., Simonnet, G., Walter, D., Weast, J., Yarmosh, Y., and Yendluri, P. (2004), *Web Services Dynamic Discovery (WS-Discover),* available from http://msdn.microsoft.com/library/en-us/dnglobspec/html/ws-discovery.pdf.

Bellwood, T., Clement, L., and von Riegen, C., eds (2003), *UDDI Version 3.0.1: UDDI Spec Technical Committee Specification*. Available from http://uddi.org/pubs/uddi-v3.0.1-20031014.htm.

Behrman, F., Fox, G.C., and Hey, T., eds *Grid Computing: Making the Global Infrastructure a Reality* (John Wiley and Sons, 2003).

Booth, D., Haas, H., McCabe, F., Newcomer, E., Champion, M., Ferris, C., and Orchard, D. (2004), *Web Services Architecture,* W3C Working Group Note 11 February 2004. Available from http://www.w3.org/TR/ws-arch/.

Bunting, B., Chapman, M., Hurley, O., Little, M., Mischkinky, J., Newcomer, E., Webber, J., and Swenson, K. (2003), *Web Services Context (WS-Context),* available from http://www.arjuna.com/library/specs/ws_caf_1-0/WS-CTX.pdf.

Chen, A. Y., Chung, S., Gao, S., McLeod, D., Donnellan, A., Parker, J., Fox, G., Pierce, M., Gould, M., Grant, L., and Rundle, J. (2003), *Interoperability and Semantics for Heterogeneous Earthquake Science Data*. In Proceedings of 2003 *Semantic Web Technologies for Searching and Retrieving Scientific Data Conference,* October 20, 2003, Sanibel Island, Florida.

Christensen, E., Curbera, F., Meredith, G., and Weerawarana, S. (2001), *Web Service Description Language (WSDL) 1.1* W3C Note 15 March, 2001.

Cox, S., Daisey, P., Lake, R., Portele, C., and Whiteside, A., eds (2003), *OpenGIS Geography Markup Language (GML) Implementation Specification*. OpenGIS project document reference number OGC 02-023r4, Version 3.0.

Clark, J. and DeRose, S., *XML Path Language (XPath) Version 1.0*, W3C Recommendation, 16 November, 1999.

de La Beaujardiere, J. (2004), *Web Map Service,* OGC project document reference number OGC 04-024.

Foster, I. and Kesselman, C., eds *The Grid 2: Blueprint for a New Computing Infrastructure*. (Morgan Kaufmann, 2003).

Fox, G. and Hey, A., eds (2002), *Concurrency and Computation: Practice and Experience,* Vol. 14, No. 13–15.

Fox, G., Pallickara, S., and Parastatidas, S. (2004), *Towards Flexible Messaging for SOAP Based Services*, In Proceedings of the IEEE/ACM Supercomputing Conference, November 2004, Pittsburgh, PA.

GANNON, D., ALAMEDA, J., CHIPARA, O., CHRISTIE, M., DUKLE, V., FANG, L., FARRELLEE, M., FOX, G., HAMPTON, S., KANDASWAMY, G., KODEBOYINA, D., KRISHNAN, S., MOAD, C., PIERCE, M., PLALE, P., ROSSI, A., SIMMHAN, Y., SARANGI, A., SLOMINSKI, A., SHIRASUNA, S., and THOMAS, T. (2004), *Building Grid Portal Applications from a Web-Service Component Architecture,* to appear in IEEE Dist. Comp. Available from http://grids.ucs.indiana.edu/ptliupages/publications/portal-apps-arch.pdf.

GRANAT, R., (2004), *Regularized Deterministic Annealing EM for Hidden Markov Models*, Doctoral Dissertation, University of California, Los Angeles.

GRANT, L., DONNELLAN, A., MCLEOD, D., PIERCE, M., FOX, G., CHEN, A., GOULD, M., and SUNG, S.-S. (2004), *A Web service based universal approach to heterogeneous fault databases,* submitted to Comp. Sci. Eng.

GUDGIN, M., HADLEY, M., MENDELSOHN, N., MOREAU, J.-J., and NIELSEN, H. (2003), *SOAP Version 1.2 Part 1: Messaging Framework,* W3C Recommendation 24 June 2003. Available from http://www.w3c.org/TR/soap12-part1/.

SOLOMON, S.C. (chair) (2002), *Living on a Restless Planet*, Solid Earth Science Working Group Report. Available from http://solidearth.jpl.nasa.gov/PDF/SESWG_final_combined.pdf.

VAN ENGELEN, R. and GALLIVAN, K. (2002), *The gSOAP Toolkit for Web Services and Peer-to-Peer Computing Networks*. In: *Proceedings of IEEE CCGrid Conferences 2002*. Available from http://www.cs.fsu.edu/~engelen/soappaper.html.

VRETANOS, P., ed (2002), *Web Feature Service Implementation Specification,* OpenGIS project document: OGC 02-058, version 1.0.0.

YOUN, C., PIERCE, M., and FOX, G. (2004), *Building problem solving environments with application Web Service Toolkits,* Future Generation Computing Systems (in press).

(Received December 30, 2004, revised September 16, 2005, accepted September 19, 2005)

Pure appl. geophys. 163 (2006) 2297–2316
0033–4553/06/122297–20
DOI 10.1007/s00024-006-0127-x

Pure and Applied Geophysics

Construction of an Intraplate Fault System Model of South Australia, and Simulation Tool for the iSERVO Institute Seed Project

H.L. XING[1,2] and P. MORA[1,2]

Abstract—To foster ongoing international cooperation beyond ACES (APEC Cooperation for Earthquake Simulation) on the simulation of solid earth phenomena, agreement was reached to work towards establishment of a frontier international research institute for simulating the solid earth: iSERVO = International Solid Earth Research Virtual Observatory institute (http://www.iservo.edu.au). This paper outlines a key Australian contribution towards the iSERVO institute seed project, this is the construction of: (1) a typical intraplate fault system model using practical fault system data of South Australia (i.e., SA interacting fault model), which includes data management and editing, geometrical modeling and mesh generation; and (2) a finite-element based software tool, which is built on our long-term and ongoing effort to develop the R-minimum strategy based finite-element computational algorithm and software tool for modelling three-dimensional nonlinear frictional contact behavior between multiple deformable bodies with the arbitrarily-shaped contact element strategy. A numerical simulation of the SA fault system is carried out using this software tool to demonstrate its capability and our efforts towards seeding the iSERVO Institute.

Key words: Intraplate fault system, finite-element method, South Australian fault model, iSERVO, international cooperation, crustal dynamics.

1. Introduction

During the 3rd ACES (APEC Cooperation for Earthquake Simulation, see http://www.aces.org.au) Working Group Meeting in Brisbane in June 2003, agreement was reached to work towards establishment of a frontier international research institute for simulating the solid earth: iSERVO = International Solid Earth Research Virtual Observatory institute (http://www.iservo.edu.au). iSERVO is the natural follow on to ACES, and aims to foster ongoing international cooperation on simulation of solid earth phenomena. This paper outlines a key Australian contribution towards the iSERVO institute seed project which includes construction of a typical intraplate fault system model and development of a finite-element based

[1] Earth Systems Science Computational Centre (ESSCC), The University of Queensland, St. Lucia, QLD 4072, Australia. E-mail: xing@esscc.uq.edu.au
[2] Australian Computational Earth Systems Simulator (ACcESS), Major National Research Facility.

software tool capable of modelling the dynamics of interacting fault systems. We have chosen the South Australian (SA) fault system because it is representative of an active intraplate fault system and due to availability of sufficient fault data to construct a model. A numerical simulation of the SA fault system is presented to demonstrate the capabilities and usefulness of the software tool. This is carried out as part of our work in constructing the Australian Computational Earth Systems Simulator (ACcESS) Major National Research Facility (see http://www.access.edu.au).

Numerical modelling of crustal fault systems offers an outstanding opportunity to gain an understanding of their dynamics and complex system behavior, and to develop the scientific underpinning for earthquake forecasting. However, there are a number of challenges that must be overcome, such as how to obtain and use geological data to construct the computational model and in particular a suitable unstructured mesh for finite-element analysis. Owen did a survey of unstructured mesh generation technology used in both in-house and commercial software (see http://www.andrew.cmu.edu/user/sowen/survey/index.html); XING and MORA (2003d) did a similar survey but focused on the technologies most relevant to meshing crustal fault systems. Based on these surveys, tetrahedral mesh generation seems to be more easily achieved automatically. Most of the existing tetrahedral mesh generators are based on the so-called Delaunay algorithm and hence, cannot guarantee that mesh points are exactly located at the fault interfaces specified by the input data. The ''advancing front algorithm'' can generate a mesh with the nodes located exactly at fault surfaces, although very few codes are available with this algorithm for the automatic mesh generation. Hexahedral meshes provide an alternative to tetrahedral meshes, while the generation of fully hexahedral meshes in a complex domain is even more difficult. For finite-element analysis of the incompressible or nearly incompressible nonlinear behaviors, it is necessary to use a hexahedral mesh rather than a tetrahedral mesh to obtain the sufficiently accurate results (e.g., TAYLOR, 2000). Fortunately, application of the geometric constraints using the so-called mapped mesh generation algorithm can facilitate hexahedral mesh generation even in some of the complicated cases.

The finite-element method is now widely used for numerical analysis of certain science and engineering problems including geodynamics. Crustal dynamics (i.e., earthquake rupture and interacting fault systems) is a more complicated problem due to the strong nonlinearity related to the stick-slip instability along faults. The static-implicit algorithm is widely applied to this kind of nonlinear problem in current finite-element analysis. Usually this employs an iterative scheme, such as the Newton-Raphson or the modified Newton-Raphson method to treat the governing equations as a system of nonlinear equations and attempt to solve them by applying the unbalanced forces, computing the corresponding displacement increments, and iterating until the drift from the equilibrium state is lower than a prescribed small value. One major disadvantage of the Newton-Raphson family of algorithms is that

the iterations may not converge, particularly when the behavior is strongly nonlinear. Despite numerous efforts to solve this problem of convergence, problems still exist. Thus, the so-called *R*-minimum strategy (YAMADA, *et al*., 1968; XING and MAKINOUCHI, 2002c) which does not involve iterations and hence, does not have non-convergence problems, was extended and employed to construct a finite-element based software tool. Since the *R*-minimum strategy involves no iteration, it belongs to the class of explicit algorithms. One common problem for such explicit algorithms is that the solution trends to drift from the equilibrium state. This effect can be minimized by using the *R*-minimum strategy to limit the size of the time increment or time step. The *R*-minimum strategy has been successfully used in the industry-strength finite element code ITAS3D for simulating sheet metal forming processes (e.g., KAWKA and MAKINOUCHI, 1995; XING and MAKINOUCHI, 2001; SFSRG, 1996).

Frictional contact problems that are required for crustal dynamics simulation are characterized by contact constraints that are imposed on contacting boundaries. Thus, an arbitrarily shaped isoparametric frictional contact element, named as the node-to-point contact element strategy, was proposed and applied here with an efficient contact search algorithm (XING and MAKINOUCHI, 2000, 2002c). Similar research on the contact description has been published in implicit finite-element analysis literature (LAURSEN and SIMO, 1993; PARISCH and LUBBING, 1997). LAURSEN and SIMO (1993) formulates the contact with respect to a convected basis in the reference configuration and in this case, the relative tangential motion is defined by the increment of the convected surface coordinates of the master element. Thus it is problematic when the motion passes across element boundaries during iterations (PARISCH and LUBBING, 1997). To avoid this difficulty, PARISCH and LUBBING (1997) proposed a formulation based on the incremental relative tangential motion measured in space. Unfortunately, problems occur with the iterative calculation of the frictional force. Thus, an explicit time integration algorithm based on the *R*-minimum strategy was introduced to overcome such problems (XING and MAKINOUCHI, 2000, 2002c). The current finite-element based software tool developed for the iSERVO institute seed project uses this strategy and represents our long-term effort to accurately solve frictional contact problems.

This paper will concentrate on our research activity for an iSERVO seed project which aims at construction of a typical intraplate fault system model and a finite-element based tool to simulate the dynamics of such models. It is presented as follows: typical intraplate fault model construction, governing equations applied in the finite-element based software tool, and numerical simulation results.

2. Typical Intraplate Fault Model Construction

With the rapid development of advanced digital image technology and its application in the earth sciences, increasingly more image information of the Earth is

available and may be used to build computational models and even to validate numerical results. The task of model construction, which aims at editing and managing the tectonic data, and constructing 2-D or 3-D unstructured meshes containing faults and plate boundaries, is a very important step for a successful finite-element analysis. It normally includes the following three aspects: (1) data management and editing: conversion of digital images and other data to those data necessary for constructing the fault and plate models, visualization and editing of the above data; (2) geometrical modelling: construction and editing of the lines and curves defining faults and plate boundaries, construction of the parametric surfaces using the above curves, construction and editing of the solid models using the generated surfaces, editing and output of the solid model of the specified region; (3) mesh generation and optimization, specification of loading boundary conditions and material properties, etc. The above process may be achieved by numerous in-house and/or commercial software such as those specified in the surveys mentioned above, but there is so far no software tool available to automatically finalize the above process by generating a high quality mesh for a general fault system, especially for the case of a hexahedral mesh.

The South Australian fault system was chosen as our contribution to the database of the iSERVO seed project as a representative intraplate fault system. This choice was based on South Australia being reasonably representative of an active fault system in Australia and availability of sufficient data and geological expertise for this region to enable a mesh to be constructed. Patran, a commercial graphics software package developed by MSC, was used to construct the model, this being the only such package that is widely accessible in our University and that satisfies our basic requirements. With detailed fault data provided by Professor Mike Sandiford of Melbourne University based on advanced digital images (e.g., Figure 1(a)) and geological knowledge of the region, we constructed a 3-D fault geometry model within a block with dimensions of about $530 \times 350 \times 60$ km^3 (Fig. 1(b)). This involved editing and smoothing the related curves/lines/surfaces defining the faults. In order to easily generate the unstructured mesh and specify the conditions necessary for the finite-element simulation (i.e., boundary conditions and information about faults), the entire geometric model of faults was firstly divided into several different geometrical components representing components of the solid model (e.g., Fig. 2), and these were then used to generate finite-element meshes (e.g., see Fig. 3). Finally, the finite-element meshes generated for the different components were assembled together with 'welding' (node equivalent), or our stick contact algorithm after meshing, and the boundary conditions were specified (see Fig. 4 and the related description in Section 4). As shown in Figures 3 and 4, more regular and fine meshes are used around the faults, while a coarse mesh is used in the other regions. This approach enables more accurate computational results to be obtained with the available and finite computational resources. The discretised model after optimization currently includes 504,471 nodes and 464,620 8-node hexahedron elements. The

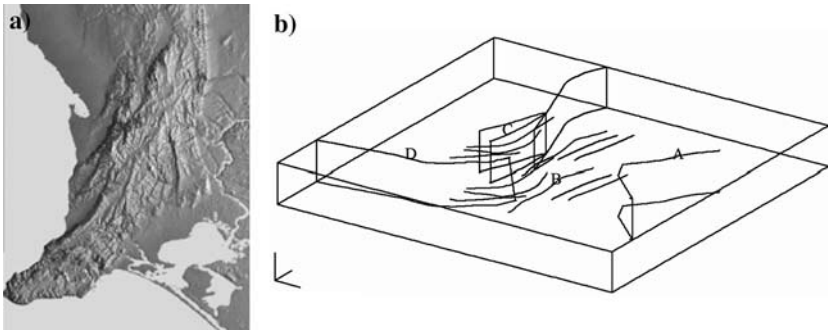


Figure 1
(a) Image of the region of South Australia being considered. (b) 3-D fault model to be analyzed.

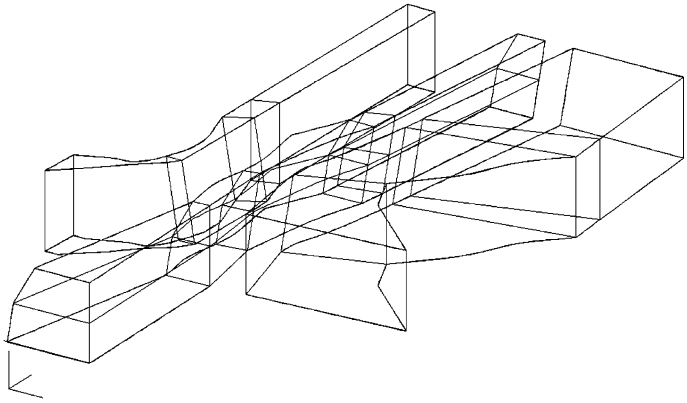


Figure 2
The components of the entire fault system model.

fineness of mesh required for a quasi-static simulation of a complicated fault system is case dependent and relates to the research question being posed and accuracy required to answer this question. Further research is required to gain understanding of mesh fineness considerations in the case of quasi-static simulations. In the case of dynamic simulations (e.g., dynamic rupture and wave propagation problems), one normally requires at least 10 nodes per wavelength to achieve accurate results, although again the exact resolution required is problem dependent.

Difficulties have also been experienced using existing finite-element/graphics packages to construct interacting fault models. Specifically, it is very time-consuming and difficult to use the existing graphics software (i.e., Patran) to construct practical 3-D models including complex fault geometries using the converted fault data from digital images, because the existing software are mostly designed for the mechanical/civil engineering industry (such as the automobile industry). In these industries, there

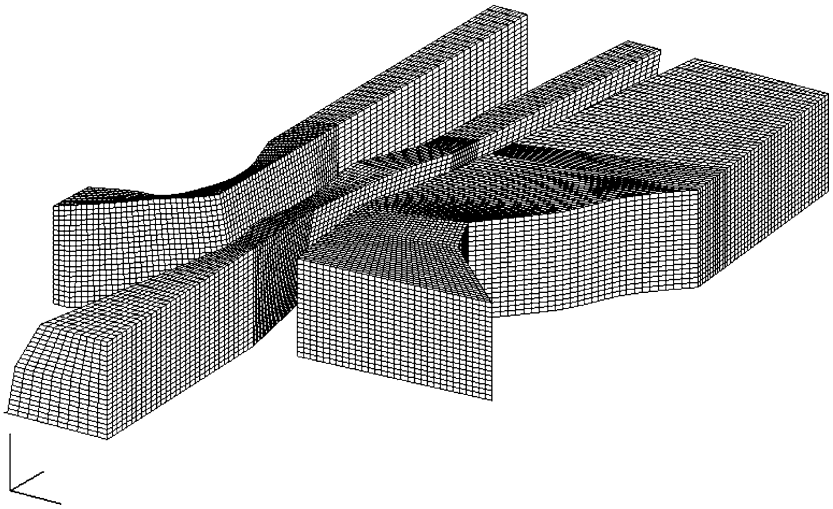


Figure 3
The unstructured mesh generated for the different components of the fault system model.

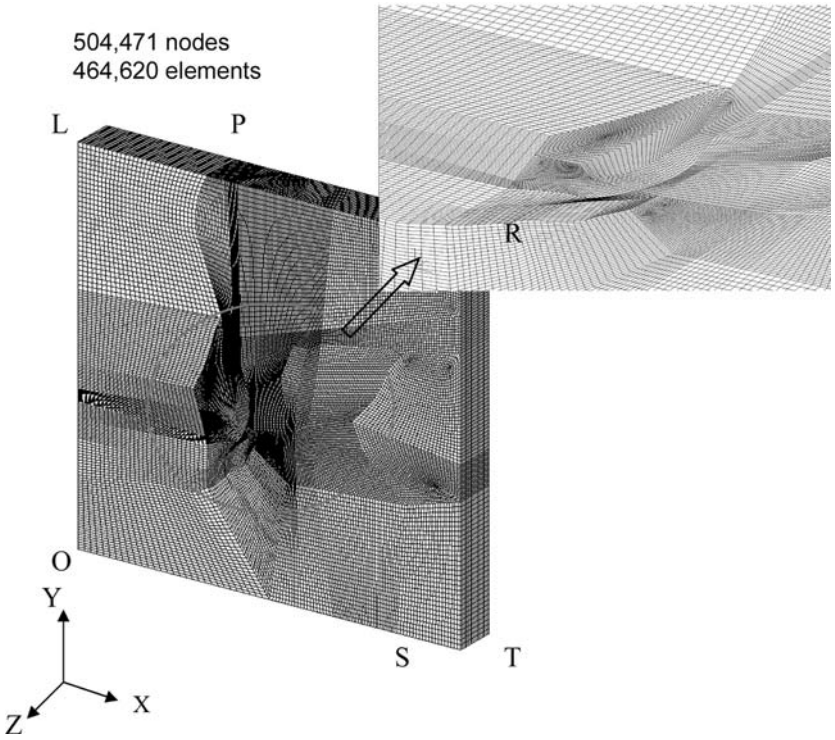


Figure 4
The total mesh after assembling the components together and the boundary conditions applied in the analysis.

are defined shapes and dimensions with reasonable tolerances provided by the designers, whereas in the case of fault data, these are usually specified using a large quantity of point data. An efficient and special purpose module for automatically meshing large complex earth models is still not available, but is extremely important.

3. Governing Algorithm for Finite-element Based Software Development

To simulate interacting fault behaviors efficiently and stably for a long duration (i.e., earthquake cycles), an adaptive time integration algorithm based on a *R*-minimum strategy is used to allow simulation of the stress accumulation and redistribution over long periods of time (i.e., tens/hundreds of years) and dynamic behavior of earthquakes over short periods of time (i.e., seconds/minutes). The *R*-minimum strategy based finite-element computational algorithm for modelling three-dimensional nonlinear frictional contact behavior between multiple deformable bodies with the arbitrarily-shaped contact element strategy was first proposed and developed by the authors, and has been successfully applied in a wide range of fields (e.g., XING *et al.*, 1998–2005). It is chosen as a basis here to construct the finite-element based software tool for the iSERVO institute seed project. The basic theory is briefly summarized as follows.

3.1. Finite-element Formulation

The updated Lagrangian rate formulation is employed to describe the nonlinear contact problem. The rate type equilibrium equation and the boundary at the current configuration are equivalently expressed by a principle of virtual velocity of the form

$$\int_V (\overset{\circ}{\tau}_{ij} - D_{ik}\sigma_{kj} + \sigma_{ik}L_{jk} - \sigma_{ik}D_{kj})\delta L_{ij}\, dV + \int_V \rho \ddot{u}_i \delta \dot{u}_i\, dV = \int_{S_F} \dot{F}_i \delta \dot{u}_i\, dS + \int_{S_C} \dot{f}_i \delta \dot{\tilde{u}}_i\, dS \tag{1}$$

where V and S denote respectively, the domains occupied by the total deformable body B and its boundary at time t; S_F is a part of the boundary of S on which the rate of traction $\dot{F}_i$ is prescribed; $\boldsymbol{\delta \dot{u}}$ is the virtual velocity field which satisfies the condition $\boldsymbol{\delta \dot{u}} = \mathbf{0}$ on the velocity boundary; $\overset{\circ}{\tau}_{ij}$ is the Jaumann rate of Cauchy stress; $\boldsymbol{L}$ is the velocity gradient tensor, $\boldsymbol{L} = \partial \boldsymbol{v}/\partial \mathbf{x}$; $\boldsymbol{D}$ and $\boldsymbol{W}$ are the symmetric and antisymmetric parts of $\boldsymbol{L}$, respectively; $\dot{\boldsymbol{f}}$ is the rate of contact stress on the contact interface S_c of the body.

3.2. Unified Constitutive Equation of Frictional Contact

Most tectonic earthquakes occur as a sudden slip along a pre-existing fault or plate interface. Earthquakes have been recognized as resulting from a stick-slip

frictional instability along faults (BRACE and BYERLEE, 1966). Subsequently, several experiments (e.g., DIETERICH, 1979; OHNAKA *et al.*, 1987; KATO *et al.*, 1999) were carried out in laboratories and frictional constitutive equations were proposed based on the experimental results which may be grouped into two main types: slip dependent (e.g., ANDREWS, 1976; MATSU'URA *et al.*, 1992; OHNAKA and SHEN, 1999) and rate- and state-dependent (e.g., DIETERICH, 1979; RUINA, 1983; MARONE, 1998). Although several other forms were subsequently proposed, they were mainly based on the above two types of relationship, such as the composite law (e.g., KATO, 2001). Researchers have been trying to determine the relationship between the slip-dependent friction law (SDFL) and rate- and state-dependent friction law (RSDFL). Both are able to describe certain phenomena during a rock friction experiment, but focus on different aspects. The former (slip-dependent laws) focuses on the rupture process (i.e., the transition process from stick to slip), but neglects rate effects; while the latter describes the rate and memory effects with a state variable, but cannot describe the real stick state. Researchers are still trying to unify both aspects and explore the relationship between the basic parameters, such as D_c in the slip-dependent case, and L in the rate- and state-dependent friction relation, but no reasonable results have yet been obtained. If we recall the basic conditions and environments of rock experiments used to develop the two types of frictional relations, it is evident that two kinds of basic experiments focused on different aspects, respectively: the SDFL experiment aimed to investigate the initial rupture behavior of rocks along an existing fault (i.e., the transition process from nucleation to high speed rupture propagation), while the RSDFL experiment was defined to investigate the slip frictional behaviors at different velocities (i.e., friction behavior at the slip state corresponding to different velocities, including velocity variations). The results demonstrate that the rate effect is essential for the case of tectonic loading, nucleation and healing. That means the SDFL and RSDFL respectively, focus on different aspects and can be combined together to describe a single whole frictional event: before slip stage (i.e., the static stick state and its transition to slip state) and after slip stage, and the transition process is governed by peak shear strength $\tau_p(=\mu f_n)$, which is dependent on the velocity, contact pressure, and state variable etc. as described by the so-called rate- and state-dependent friction relation. This is similar to elastic-plastic deformation phenomena of materials which are governed by the yield locus.

As for the frictional slip, an increment decomposition of the sticking and the slipping is assumed. A standard Coulomb friction model is applied analogously to the flow plasticity rule to govern the slipping behavior. Thus, the frictional stress can be described as follows (XING and MAKINOUCHI, 2002, 2003a) (a variable with tilde ($\sim$) denotes a relative component between slave and master bodies, and $l, m = 1, 2$; $i, j, k = 1, 2$ in this paper if not otherwise specified):

$$f_l = E_t \tilde{u}_l^e = E_t \Sigma \Delta \tilde{u}_l^e, \quad df_l = E_t d\tilde{u}_l, \tag{2}$$

(in the sticking state)

$$df_l = \frac{\overline{F} E_t}{\sqrt{f_m^e f_m^e}} (\delta_{lm} - \eta_l \eta_m) d\tilde{u}_m + \eta_l \mu \left(df_n + \frac{\partial \mu}{\partial f_n} df_n \right) + \eta_l f_n \left(\frac{\partial \mu}{\partial \dot{\tilde{u}}_{\text{eq}}} d\dot{\tilde{u}}_{\text{eq}} + \frac{\partial \mu}{\partial \varphi} d\varphi \right) \tag{3}$$

(in the slipping state)

Here μ is the friction coefficient, it may depend on the normal contact pressure f_n, the equivalent tangential velocity $\dot{\tilde{u}}_{\text{eq}}$, the state variable φ, i.e., $\mu = \mu(\dot{\tilde{u}}_{\text{eq}}^{sl}, f_n, \varphi)$ (e.g., DIETERICH, 1979; RUINA, 1983; SCHOLZ, 1998).

A state variable φ, used in the rate- and state-dependent friction law, seems to be necessary for theoretical analyses such as the so-called spring-slider model which was applied to investigate the earthquake mechanism (e.g., TSE and RICE, 1986; OKUBO, 1989; DIETERICH, 1992; RICE 1993), because of the assumption of rigid bodies. However, the state variable also introduces a numerical difficulty due to the following reasons as stated by XING and Mora (2006): (1) The friction cannot be described explicitly with the above rate- and state-dependent friction relation, it requires numerical iterations to calculate the state variable effects; (2) for the widely applied rate- and state-dependent frictional equation, it conflicts with the use of the Coulomb friction criteria due to the state variable. It easily falls into a dead cycle and is quite computationally time-consuming to escape from this cycle. Therefore, we proposed using the stick- and slip-decomposition algorithm only and for this to be governed by the Coulomb friction criteria with the rate-dependent strength. This neglects the state variable thereby avoiding related numerical difficulties. Thus, to describe the stick-slip instability phenomena detailed above, Equation (3) can then be rewritten as:

$$df_l = \frac{\overline{F} E_t}{\sqrt{(f_1^e)^2 + (f_2^e)^2}} (\delta_{lm} - \eta_l \eta_m) d\tilde{u}_m + \eta_l \mu df_n + \eta_l f_n \frac{\partial \mu}{\partial \dot{\tilde{u}}_{\text{eq}}^{sl}} d\dot{\tilde{u}}_{\text{eq}}^{sl}. \tag{4}$$

In addition, the penalty parameter method is chosen to satisfy the normal impenetrability condition when contact occurs, thus the normal contact stress can be calculated as

$$f_n = \boldsymbol{f} \cdot \boldsymbol{n} = E_n \mathrm{g}_n \ (\neq 0 \quad \text{only for} \quad \mathrm{g}_n \leq 0). \tag{5}$$

Here E_n is the penalty parameter to penalize the penetration (gap) in the normal direction; g_n is the penetration (gap) in the normal direction.

In summary, from Eqs. (2), (4) and (5), the contact stress acting on a slave node can be described as (denote $\dot{f}_3 = \dot{f}_n$)

$$\dot{f}_i = G_{ij}\dot{\tilde{u}}_j, \tag{6}$$

where $\boldsymbol{G}$ is the frictional contact matrix for both the stick and the slip state.

The simulation results for two typical frictional models demonstrated that our algorithm (i.e., the stick-slip decomposition algorithm together with the unified friction description) can describe frictional instability phenomena, geometrical effects on frictional behavior, and repeated relative motion state changes (i.e., stick-slip instability cycles) that are observed in practical rock experiments without using the so-called state variable. This also avoids numerical difficulties such as additional iterations that are required when using the so-called state variables (see XING and MORA, 2006 for details).

3.3. Arbitrarily-shaped Contact Element Strategy

An arbitrarily-shaped contact element strategy, named as the node-to-point contact element strategy (which needs nothing special to cover the traditional node-to-node, node-to-segment and node-to-edge description), was proposed to handle the frictional contact problems between deformable bodies with the *R*-minimum strategy (XING *et al*., 1998a; XING and MAKINOUCHI, 2000, 2002a, b, and c) and is briefly introduced here. Assume a slave node s has contacted with point c on a surface element (master segment) E', and the surface element E' consists of γ nodes ($\gamma = 4$ in this paper if not otherwise defined), thus the term related with frictional contact in Eq. (1) can be described as

$$\dot{f}_i(\delta\dot{u}_{si} - \delta\dot{u}_{ci}) = \delta\dot{\tilde{u}}_{sci\beta}[\overline{K}_{fik}]_{\beta\alpha}\dot{u}_{sck\alpha}, \tag{7}$$

where

$$\left[\overline{K}_{fik}\right]_{\beta\alpha} = R_\beta \mathbf{e}_i \cdot \left\{G_{hk}R_\alpha \mathbf{e}_h + \left(H_{jm}\hat{\mathbf{e}}_j\left(\left(\overline{C}_{ll}R_{\alpha,m} - \overline{C}_{ml}R_{\alpha,l}\right)\mathbf{e}_k \cdot \tilde{\mathbf{x}} + R_\alpha\left(\overline{C}_{ll}\hat{\mathbf{e}_m} - \overline{C}_{ml}\hat{\mathbf{e}}_l\right) \cdot \mathbf{e}_k\right)\right)\right\}$$
$$(h = 1,3,\ l \neq m \text{ and no sum on } l). \tag{8}$$

Here $\alpha = 1,(\gamma+1), \beta = 1,(\gamma+1)$; $\bar{C}_{ml} = C_{ml} - g_n \boldsymbol{n}\cdot\hat{\boldsymbol{e}}_{m,l}, C_{ml} = \hat{\boldsymbol{e}}_m\cdot\hat{\boldsymbol{e}}_l, \wp = \bar{C}_{11}\bar{C}_{22} - \bar{C}_{12}\bar{C}_{21}$, $\tilde{\boldsymbol{x}} = \boldsymbol{x}_s - \boldsymbol{x}_c$, $E_{ijm} = \hat{\mathbf{e}}_{i,m}\cdot\hat{\mathbf{e}}_j$, $H_{jm} = \hat{f}_i E_{ijm}/\wp$, $\boldsymbol{R} = [1 - N_1 - N_2 \cdots - N_\gamma]^T$, $N_p(p = 1,\gamma)$ is the shape function value of the point c on the surface element E', $\hat{\boldsymbol{e}}_i$ and $\hat{\boldsymbol{e}}_i$ are, respectively the base vectors of the local natural and the local Cartesian coordinate systems on the master segment.

3.4. Time Integration Algorithm

The time integration method is one of key issues to formulate a nonlinear finite-element method. It is well known that the fully implicit method often suffers from bad convergence problems, mostly due to changes of contact and friction states. In order to avoid this, we employ an explicit time integration procedure as follows. It

is assumed that under a sufficiently small time increment all rates in Eq. (1) can be considered constant within the increment from t to as long as there are no drastic changes of state taking place (for example, elastic to plastic at an integration point, contact to discontact or discontact to contact on the contact interface, stick to slide or slide to stick in friction on the contact interface). The R-minimum method is extended and used to limit the step size in order to avoid such a drastic change in state within an incremental step (e.g., XING and MAKINOUCHI, 2002c).

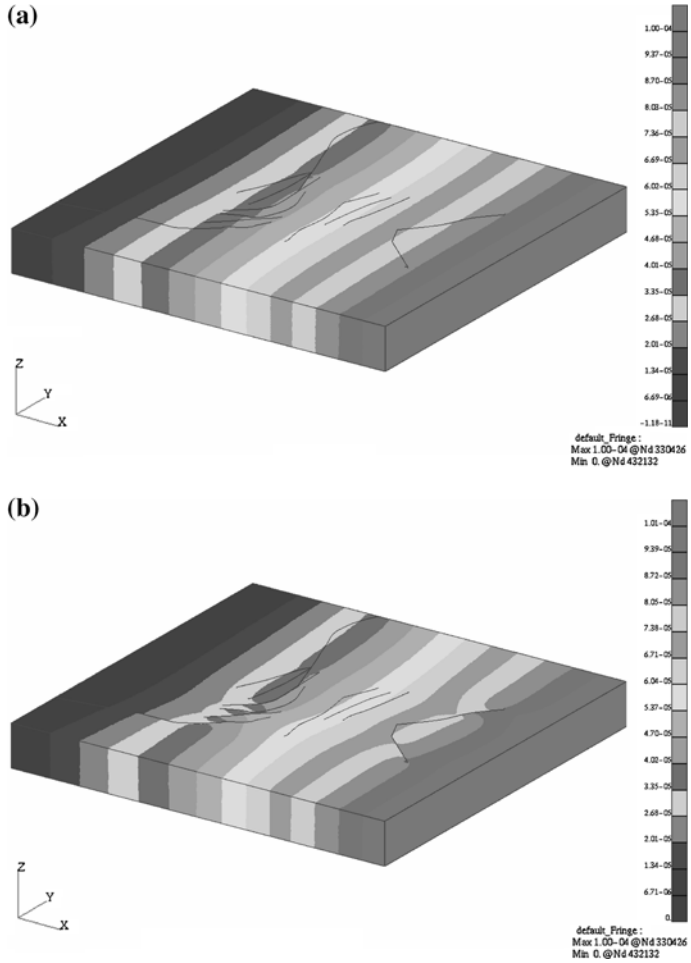


Figure 5
The velocity distribution at the different stages. Loading time = 1000 R/V_{ref}, in which (a) R = 0.002, (b) R = 0.006, (c) R = 0.008, (d) R = 0.00813, (e) R = 0.00829, (f) R = 0.00860, (g) R = 0.00885. The dark blue curves represent the faults.

Thus all the rate quantities used to derive Eq. (1) are simply replaced by incremental quantities as

$$\Delta \mathbf{u} = \mathbf{v}\Delta t = \dot{\mathbf{u}}\Delta t, \quad \Delta \tau = \overset{\circ}{\tau}\Delta t. \tag{9}$$

Finally, in combination with the above equations, Eq. (1) can be rewritten as

$$(\mathbf{K} + \mathbf{K}_f)\Delta \mathbf{u} = \Delta \mathbf{F} + \Delta \mathbf{F}_f, \tag{10}$$

here $\mathbf{K}$ and $\Delta \mathbf{F}$ are, respectively the stiffness matrices and force increment terms corresponding to the various time integration algorithms; $\mathbf{K}_f$ and $\Delta \mathbf{F}_{\mathrm{f}}$ are the stiffness matrices and the force increments of all the contact elements among the

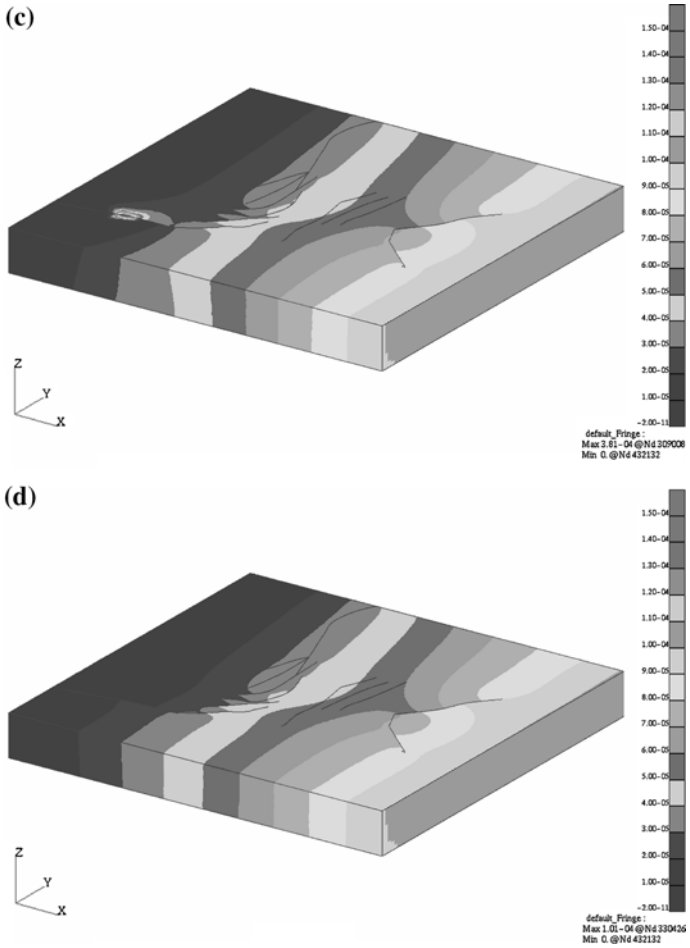


Figure 5
(contd.)

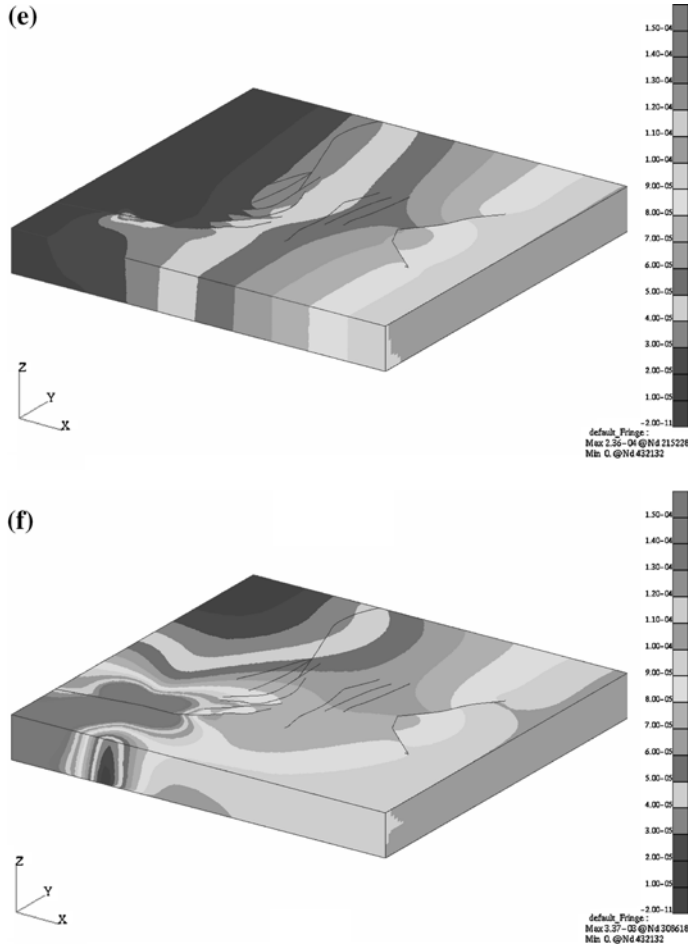


Figure 5
(contd.)

contact pairs. As for the details on how to solve the above equations for large-scale problems, please refer to XING *et al*. (1998) and XING and MAKINOUCHI, 2002b and 2002c).

4. Numerical Simulation of South Australia Fault System Model

Finite-element software based on the above theory has been developed and tested with the so-called sandwich fault model (XING and MAKINOUCHI, 2002a),

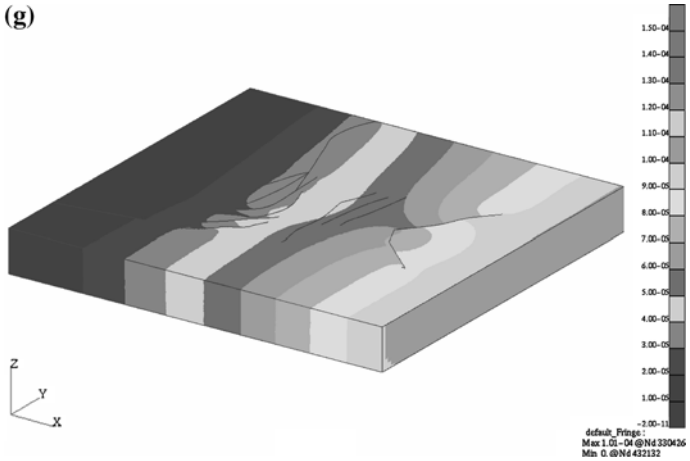


Figure 5
(contd.)

subduction fault model (XING, 2001a; XING, 2002b), single fault bend model (including both the interplate and intraplate cases) (XING *et al.*, 2003a, b, 2004a), multiple fault bends model (XING and MORA, 2003c, 2006) and the Southern California fault model (XING and MORA, 2004b). It is applied here to simulate stress/velocity variations of the South Australian (SA) fault system in a discrete model with 504,471 nodes constructed as above using fault data supplied by Professor Mike Sandiford of Melbourne University. The rate-dependent friction law is used here to describe the frictional behavior along the faults: $\mu = 0.60 + (0.01 - 0.025)\ln(V/V_{\text{ref}})$, where V_{ref} and V are respectively, an arbitrary reference velocity and a sliding velocity. It takes the same form as the so-called rate- and state-dependent frictional law when $d\varphi/dt = 0$ (e.g., RUINA, 1983; SCHOLZ, 1998). All the materials have the same properties: density $\rho = 2.60\ \text{g/cm}^3$, Young's modulus $E = 44.8$ GPa and Poisson's ratio $\gamma = 0.2$. From the above, only the friction law used here is time-dependent, thus we choose V_{ref} as the reference velocity for the entire process here. The following boundary conditions are applied in this analysis (as shown in Fig. 4): at the right hand side surface of the model (TR), a constant relative velocity of $V_{\text{ref}}/10$ is applied along the negative X direction, but nodes are kept fixed along the Z direction; at the left-hand side surface OL, nodes are fixed along both the X and Z direction; while at the surfaces LP and ST, nodes are fixed along the Y direction; all the other surfaces are free. The stress and the frictional contact force of the whole model at the initial stage are assumed as 0. Figures 5 and 6, respectively show a few snapshots of the preliminary results for the velocity and the corresponding equivalent stress-rate distribution at different loading stages. For simplicity in the following description,

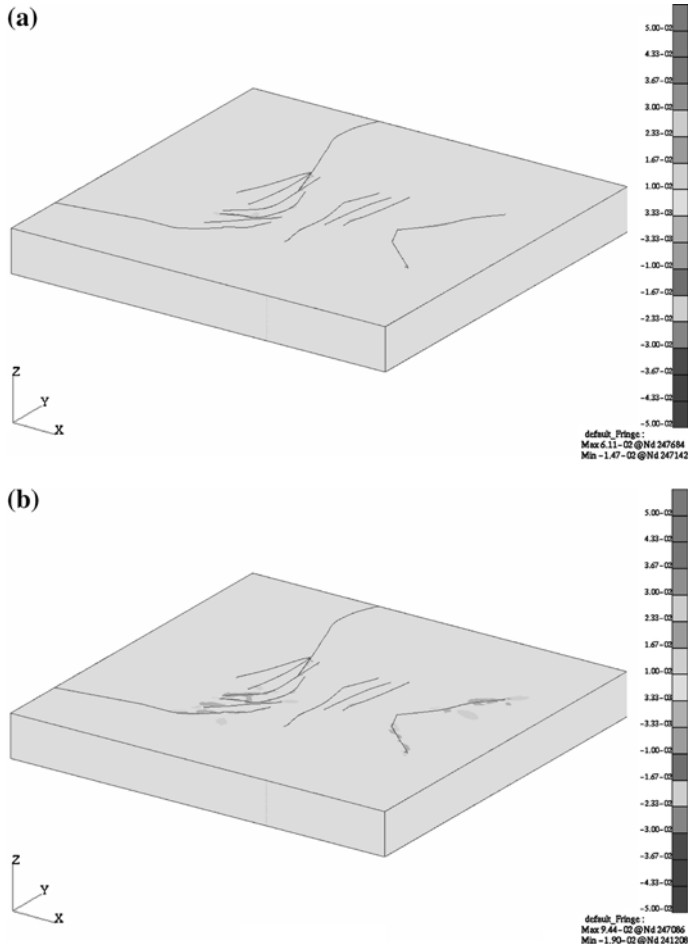


Figure 6
The corresponding equivalent stress rate distribution at the different stages. The positive value represents the stress/energy increase/accumulation, while the negative value is for the stress/energy decrease/release. Loading time = 1000 R/V_{ref}, in which (a) R = 0.002, (b) R = 0.006, (c) R = 0.008, (d) R = 0.00813, (e) R = 0.00829, (f) R = 0.00860, (g) R = 0.00885. The dark blue curves represent the faults.

positions of the faults are denoted as zones A, B, C and D, respectively as shown in Figure 1(b). The velocity distributions show that no relative motions exist along the faults at the beginning loading stages (i.e., all the nodes along the faults are in the stick state, see Fig. 5(a)) accompanying with the stress/energy accumulation (see Fig. 6(a)), but nodes at zones A and C gradually swap into the slip state due to the accumulation of stress coupled with stress/energy release at the related nodes as shown in Fig. 6(b), while nodes at other positions remain in the stick state until the instant of time shown in Figure 5(b). With continued loading, increasingly more

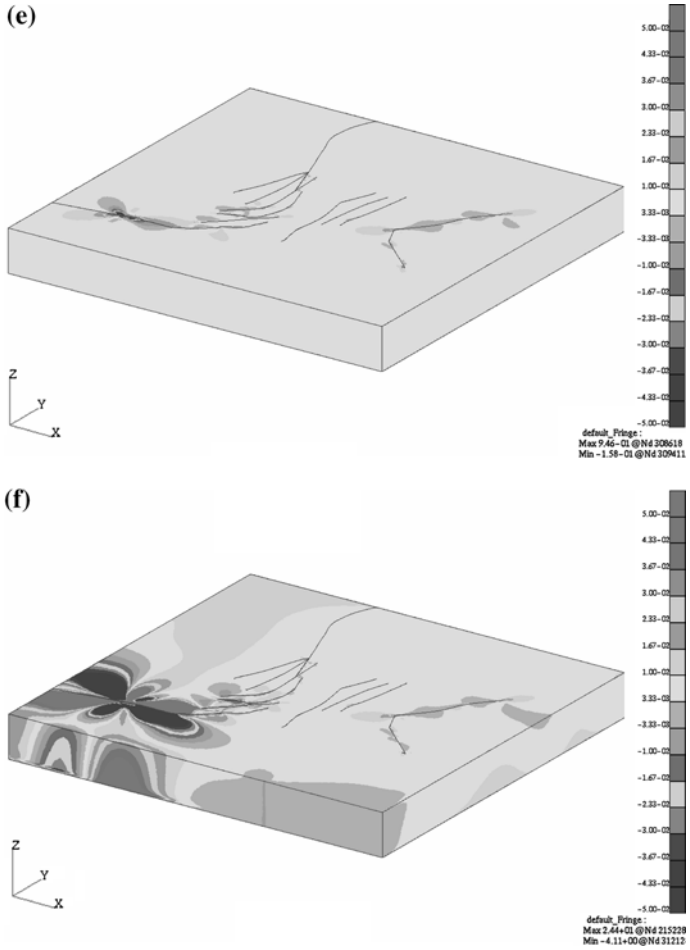


Figure 6
(contd.)

nodes at zone C switch into the slip state and their released stress/energy pushes their neighbor nodes and stress accumulates along the faults (e.g., positive stress rate distribution of the nodes between zones C and D in Fig. 6(b)). Suddenly, the local nodes at zone D jump into the slip state with a larger velocity (see Fig. 5(c) due to the dramatic stress variation along the fault (stress release, see Fig. 6(c)); then these nodes at zone D are returned back to the silent state (see Fig. 5(d)) with their stress being mostly released and accumulated again (see Fig. 6(d)); the above process at zone D repeats again and again (see (c)–(g) of Figs. 5 and 6), and may lead to considerably larger events as show in Figures 5(f) and 6(f). This is similar to practical earthquake cycles observed/recorded in many sites. Of course, the results

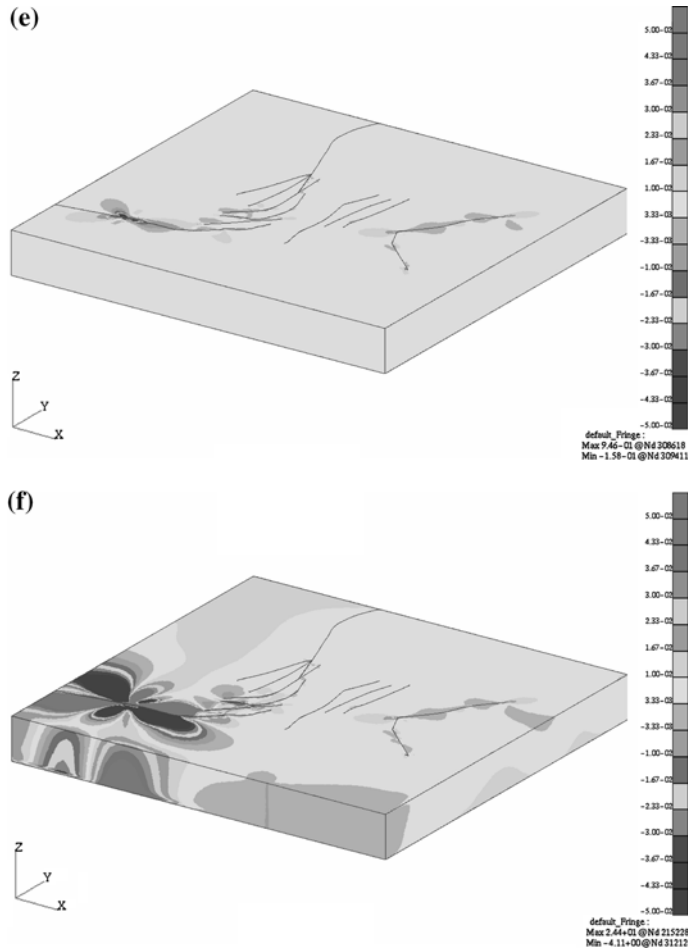


Figure 6
(contd.)

may depend on the initial conditions but not withstanding this, the simulation captures the key earthquake phenomena including the stress variation pattern (e.g., in Figs. 6 (c)–(g)). The numerical results also demonstrate that our work towards the iSERVO seed project is progressing well.

5. *Conclusion*

With the detailed fault data around South Australia (SA), an intraplate fault system model of SA was successfully constructed with 504,471 nodes and 464,620

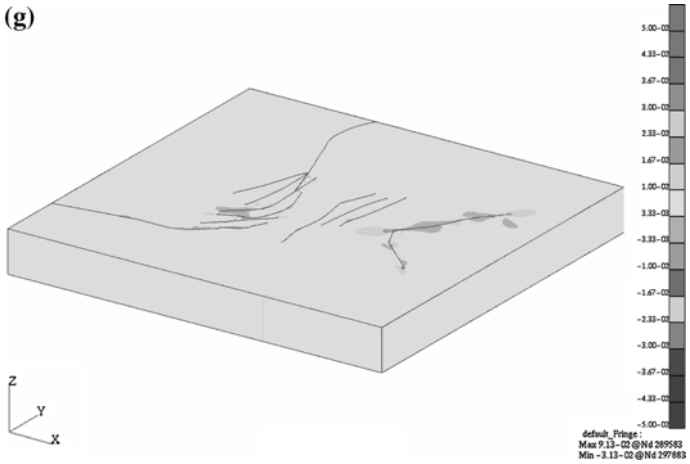


Figure 6
(contd.)

8-node hexahedron elements. Meanwhile, a three-dimensional finite-element based software tool for modelling the nonlinear frictional contact behaviors of the interacting fault system was developed with the arbitrarily-shaped contact element strategy and the unified friction description proposed by the authors and applied here to simulate the SA fault system model. The preliminary numerical results demonstrate that the constructed SA fault model can be used for further finite-element analysis aimed at understanding seismicity of the region and that the developed simulation tool can capture key earthquake phenomena such as earthquake cycles. Thus, our seed work towards establishing the iSERVO institute is progressing well.

Acknowledgments

Support is gratefully acknowledged by the Australian Computational Earth Systems Simulator Major National Research Facility (ACcESS MNRF), the Queensland State Government, The University of Queensland, and SGI. Computations were made using the ACcESS MNRF supercomputer – a 208 processor 1.1 TFlops SGI Altix 3700—which was funded by the Queensland State Government Smart State Research Facility Fund and SGI. The authors are grateful to Professor M. Sandiford from Melbourne University for providing the data for the South Australia Fault system used in this paper, to Dr. Hori of JAMSTEC and another reviewer for their advice/comments that allowed this paper to be improved.

References

ANDREWS, D.J. (1976), *Rupture propagation with finite stress in antiplane strain*, J. Geophs. Res. *81*, 3575–3582.

BRACE, W.F. and BYERLEE, J.D. (1966), Stick-slip as A mechanism for earthquakes, Science *153*, 990–992.

DIETERICH, J.H. (1978), *Time-dependent friction and the mechanics of stick-slip,* Pure Appl. Geophys. *116*, 790–806.

DIETERICH J.H. (1992), *Earthquake nucleation on faults with rate- and state-dependent strength*, Tectonophysics *211*, 115–134.

DIETERICH, J.H. (1979), *Modelling of rock friction 1. Experimental results and constitutive equations*, J. Geophys. Res. *84*, 2161–2168.

KATO, N., SATOH, T., LEI, X., YAMAMOTO, K., and HIRASAWA, T. (1999), *Effect of fault bend on the ruputure propagation process of stick-slip,* Tectonophysics. *310*, 81–99.

KATO, N. and Tullis, T.E. (2001), *A composite rate- and state-dependent law for rock friction*. Geophys. Res. Lett. *28*, 1103–1106.

KAWKA, M. and MAKINOUCHI, A. (1995), *Shell element formulation in the static explicit FEM code for simulation of sheet stamping*. J. Materials Process. Technol. *50*, 105–115.

LAURSEN, T.A. and SIMO, J.C. (1993), *A continuum-based finite element formulation for implicit solution of multi-body, large deformation frictional contact problems*. Int. J. Num. Methods Eng. *36*, 3451–3485.

MARONE, C. (1998), *Laboratory-derived friction laws and their application to seismic faulting*, Annu. Rev. Earth Planet. Sci. *26*, 643–696.

MATSU'URA, M., KATAOKA, H., and SHIBAZAKI, B. (1992), *Slip-dependence friction law and nucleation processes in earthquake rupture,* Tectonophysics *211*, 135–148.

OHNAKA, M., KUWAHARA, Y., and YAMAMOTO, K. (1987), *Constitutive equations between dynamic physical parameters near a tip of the propagating slip zone during stick-slip shear failure*, Tectonophysics *144,* 109–125.

OHNAKA, M. and YAMASHITA, T. (1989), *A cohesive zone model for dynamic shear faulting based on experimentally inferred constitutive relation and strong motion source parameters,* J. Geophys. Res. *94*, B4, 4089–4104.

OHNAKA, M. and KUWAHARA, Y. (1990), *Characteristic features of local breakdown near a crack-tip in the transition zone from nucleation to unstable rupture during stick-slip shear failure*, Tectonophysics *175*, 197–220.

OHNAKA, M. and SHEN, L. (1999), *Scaling of the shear rupture process from nucleation to dynamic propagation: implications of geometric irregularity of the rupture surfaces,* J. Geophys. Res. *104*, B1, 817–844.

OKUBO, P.G. (1989), *Dynamic rupture modeling with laboratory-derived constitutive relations,* J. Geophys. Res. *94*, 12321–12335.

PARISCH, H. and LUBBING, C.H. (1997), *A formulation of arbitrarily shaped surface elements for three-dimensional large deformation contact with friction,* Int. J. Num. Methods Eng. *40,* 3359–3383.

RICE, J.R. (1993), *Spatio-temporal complexity of slip on a fault,* J. Geophys. Res. *98*, 9885–9907.

RUINA, A.L. (1983), Slip *instability and state variable friction laws,* J. Geophys. Res. *88*, 10359–10370.

SCHOLZ, C.H. (1998), *Earthquakes and friction laws,* Nature *391*, 37–42.

SFSRG (Sheet Forming Simulation Research Group, 1996), *ITAS3D User's Manual*, Version 2.0 (RIKEN, Japan).

TAYLOR, R.L. (2000), *A mixed-enhanced formulation for tetrahedral finite elements*, Int. J. Numer. Meth. Eng. *47*, 205–227.

TSE, S.T. and RICE, J.R. (1986), *Crustal earthquake instability in relation to the depth variation of frictional slip properties*, J. Geophys. Res. *91*, 9452–9472.

XING, H.L., FUJIMOTO, T., MAKINOUCHI, A., and NIKISHKOV, G.P. (1998), *Static-explicit FE modeling of 3-D large deformation multibody contact problems on parallel computer*. In: Simulation of Materials Processing: Theory, Methods and Applications (Huetink, J. and Baaijens, F., eds.) (A.A. Balkema, 1998) pp. 207–212.

XING, H.L. and MAKINOUCHI, A. (2000), *A node-to-point contact element strategy and its applications,* RIKEN Review: Focused on High Performance Computing *30*, 35–39.

XING H.L., MIYAMURA, T., MAKINOUCHI, A., HOMMA, T., KANAI, T., and OISHI, Y. (2001a), *Development of high performance finite-element software system for simulation of earthquake nucleation and development*. In Exploration Geodynamics (Moresi, L., Muller, D., and Hobbs, B., eds), Western Australia, 178–188.

XING, H.L. and MAKINOUCHI, A. (2001b), *Numerical analysis and design for tubular hydroforming*. Int. J. Mech. Sci. *43*, 1009–1026.

XING, H.L. and MAKINOUCHI, A. (2002a), *Finite-element analysis of sandwich friction experimental model of rocks*, Pure Appl. Geophys. *159*, 1985–2009.

XING, H.L. and MAKINOUCHI, A. (2002b), *Finite-element modeling of multibody contact and its application to active faults*. Concurrency and Computation: Practice and Experience *14*, 431–450.

XING, H.L. and MAKINOUCHI, A. (2002c), *Three-dimensional finite-element modeling of thermomechanical frictional contact between finite deformation bodies using R-minimum strategy,* Comput. Methods Appl. Mech. Eng. *191*, 4193–4214.

XING, H.L. and MAKINOUCHI, A. (2003a), *Finite-element modeling of frictional instability between deformable rocks*. Int. J. Numer. Anal. Meth. Geomech. *27*, 1005–1025.

XING, H.L., MORA, P., and MAKINOUCHI, A. (2003b), *Finite-element simulation of stress evolution in a frictional contact system,* Lecture Notes in Computational Science LNCS 2659, 798–806.

XING, H.L. and MORA, P. (2003c), *Finite-element analysis of coupling role of multiple fault bends in the stick-slip instability along the faults,* EOS Trans. AGU *84*/(46), Fall Meet. Suppl., Abstract NG41C-78.

XING, H.L. and MORA, P. (2003d), *Mesh generation software survey*, Technical Report of ACcESS, (ESSCC, UQ), pp. 1–20.

XING, H.L., MORA, P., and MAKINOUCHI, A. (2004a), *Finite-element analysis of fault bend influence on stick-slip instability along an intra-plate fault*, Pure Appl. Geophys. *161*, 2091–2102.

XING, H.L. and MORA, P. (2004b), *Finite-element modeling of interacting fault systems,* Proc. 4th ACES Workshop, Beijing, 112–114.

XING, H.L. and MORA, P. (2006), *A unified friction description and its application to the simulation of frictional instability using the finite-element method*, Philos. Magazine *86*, 3453–3475.

YAMADA, Y., YOSHIMURA, N., and SAKURAI, T. (1968), *Plastic stress-strain matrix and its application for the solution of elastic-plastic problems by finite-element method*, Int. J. Mech. Sci. *10*, 343–354.

(Received November 1, 2004, revised December 20, 2005, accepted December 30, 2005)
Published Online First: December 20, 2006

Pure appl. geophys. 163 (2006) 2317–2341
0033–4553/06/122317–25
DOI 10.1007/s00024-006-0135-x

Pure and Applied Geophysics

LURR's Twenty Years and its Perspective

Xiang-Chu Yin,[1,2] Lang-Ping Zhang,[2] Hui-Hui Zhang,[2] Can Yin,[3] Yucang Wang,[2,3] Yongxian Zhang,[2,4] Keyin Peng,[2,4] Haitao Wang,[2,5] Zhiping Song,[2,6] Huaizhong Yu,[2] and Jiancang Zhuang[2,7]

Abstract—Seismogenic process is a nonlinear and irreversible one, so that the response to loading of a seismogenic zone is different from the unloading one. This difference reflects quantitatively the process of an earthquake preparation. A physics-based new parameter-Load/Unload Response Ratio (LURR) was proposed to measure quantitatively the proximity to a strong earthquake and then used to be an earthquake predictor. In the present paper, a brief history of LURR is recalled; inspection of real earthquake cases, numerical simulations and laboratory studies of LURR, prediction efforts in terms of LURR, probability problem of LURR and its prospect are also expatiated.

Key words: Load-Unload Response Ratio (LURR), earthquake prediction, numerical simulation, laboratory study.

1. Introduction

The idea of LURR (Load/Unload Response Ratio) came as a brainstorm in 1984, twenty years ago, when the first author was a visiting professor in Saint Louis University in U.S.A. cooperating with Professor Otto Nuttli, who was extremely interested in LURR and encouraged us to strive for the maximum achievement. He is many years deceased, however we still deeply miss him.

The essence of an earthquake, from the viewpoint of mechanics or physics, is precisely the failure or instability of the focal media accompanied by a rapid release

[1] China Earthquake Administration (CEA), Institute of Earthquake Science, Beijing 100036, China. E-mail: xcyin@public.bta.net.cn

[2] LNM (State Key Laboratory of Nonlinear Mechanics), Institute of Mechanics, Chinese Academy Sciences (CAS), Beijing 100080, China.

[3] ESSCC, University of Queensland, The University of Queensland St. Lucia, Brisbane, QLD 4072, Australia. E-mail: wangyc@esscc.earth.uq.edu.au; canyon@esscc.earth.uq.edu.au

[4] China Earthquake Network Center, CEA, Beijing 100036, China.

[5] Xinjiang Earthquake Bureau, CEA, Urumchi 830011, China.

[6] Shanghai Earthquake Bureau, CEA, Shanghai 200062, China.

[7] Department of Earth and Space Sciences, University of California, Los Angeles, CA 90095-1567. E-mail: zhuang@moho.ess.ucla.edu

of energy. Therefore the preparation process of an earthquake is exactly the damage process of the focal media that is usually a narrow layer of rock.

From the microscopic viewpoint, the damage process for geo-material (rock) has incredible richness in complexity (MEAKIN, 1991; BAI *et al.*, 1994; KRAJCINNOVIC, 1996; XIA *et al.*, 2002). In any rock block there must be a large number of disordered defects (cracks, fissures, joints, faults; caves) with different sizes, shapes and orientations. The damage process includes initiation, growth, interaction, coalition and cascade of cracks. It is an irreversible, far-from -equilibrium and nonlinear process, which has been intensively studied for decades but many fundamental questions remain unsolved. The inherent difficulty of earthquake prediction stems mainly from this complexity. From the macroscopic viewpoint the constitutive curve is a comprehensive description of the mechanical behaviors of the materials. A typical constitutive curve for focal media (rock) is shown in Figure 1. For generality, in Figure 1 the ordinate denotes general load P instead of stress σ and the abscissa is the general response R to load instead of strain ε. If the load acting on the material increases monotonously, the material will experience the regimes of elastic, damage and failure or destabilization. The most essential characteristic of the elastic regime is its reversibility; i.e., the positive process and the contrary process are reversible. In other words, the loading modulus and the unloading one are equal to each other. Contrary to the elastic regime, the damage regime is irreversible and the unloading response is different from the loading response, or the loading modulus is different from the unloading modulus. This difference indicates the deterioration of material due to damage.

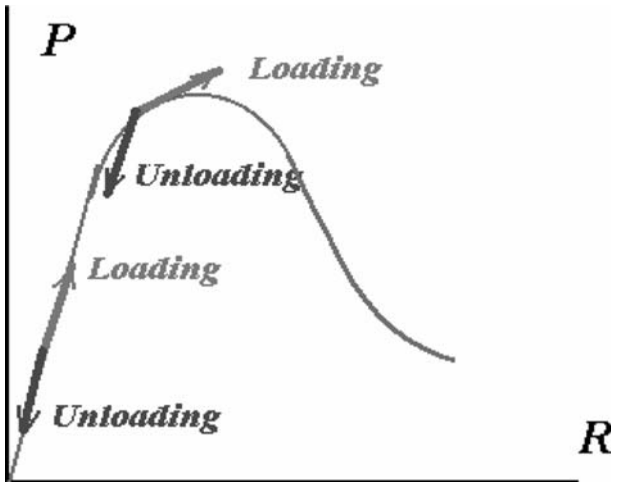


Figure 1
The constitutive curve of focal zone.

In order to measure quantitatively the difference, two parameters are defined as follows:

The first one is the response rate X defined as

$$X = \lim_{\Delta P \to 0} \frac{\Delta R}{\Delta P}, \tag{1}$$

where ΔP and $\Delta R =$ denote the increments of load P and response R, respectively.

The second one is the Load/Unload Response Ratio (LURR) Y defined as

$$Y = \frac{X_+}{X_-}, \tag{2}$$

where X_+ and X_- refer to response rate under loading and unloading condition, respectively.

It is clear that $Y = 1$ for the elastic regime since $X_+ = X_-$ and $Y > 1$ for the damage regime due to $X_+ > X_-$. The more severely the material is damaged, the larger the Y value will be. As the media approach failure the Y value becomes increasingly larger. Therefore the Y value (LURR) could measure the proximity to failure and also act as a precursor for earthquake prediction/forecasting.

In continuum damage mechanics, the damage degree of material is measured by damage variable or damage parameter d. There are many ways to define d -from scalar to high order tensor (KRAJCINNOVIC, 1996). A direct way is to define d as the relative variation of the effective stiffness tensor. But for a simple condition, it can be simplified. According to LEMAITRE's definition (LEMAITRE, 1984), the damage parameter d is defined as:

$$d = (E_0 - E)/E_0, \tag{3}$$

where modulus E_0 denotes the Young's modulus for the undamaged original material and E means the Young's modulus in the damaged state for the uniaxial tension or compression.

In fact, the reduction of modulus is due to the existence of defects (cracks, fissures, joints, faults; caves) in the material. Many scientists have studied the relation between modulus and the contained cracks. According to ODA (1983), the increase of the compliance tensor $\boldsymbol{M}$ (the reciprocal of stiffness and $\varepsilon_{ij} = M_{ijkl}\sigma_{kl}$) could be expressed as

$$M_{ijkl} - M^o_{ijkl} = (\Delta/4)(\delta_{il}F_{jk} + \delta_{jl}F_{ik} + \delta_{jk}F_{il} + \delta_{ik}F_{jl}), \tag{4}$$

where M^o_{ijkl} denotes the compliance tensor of the undamaged medium, M_{ijkl} denotes the compliance tensor of the damaged medium, δ_{ij} is the Kronecker delta, and F_{ij} are the components of the fabric tensor which are defined as

$$F_{ij} = \frac{\pi N}{V} \int_0^{\infty} \int_{\Omega} a^3 D\,(n_k, a)\, n_i\, n_j\, d\Omega\, da, \tag{5}$$

where a is the size of crack, D is a crack distribution density function, n denotes the unit normal vector and $d\Omega$ is the spherical surface element.

Obviously not only the calculation of the fabric tensor is very complicated, but also it is more difficult to obtain enough data for calculating F_{ij} at present. It is known that a crack or seismic fault with size a corresponds to an earthquake with specified magnitude and energy release (KANAMORI and ANDERSON, 1975). Therefore it would be better to define the Y value directly by the released seismic energy in seismology or AE energy in laboratory as follows

$$Y = \frac{\left(\sum_{i=1}^{N^+} E_i^m\right)_+}{\left(\sum_{i=1}^{N^-} E_i^m\right)_-} \tag{6}$$

where E denotes radiated energy which can be calculated according to the Gutenberg-Richter formula (KANAMORI and ANDERSON, 1975), the sign "+" means loading and "–" unloading, $m = 0$ or 1/3 or 1/2 or 2/3 or 1. When $m = 1$, E^m is exactly the energy itself; $m = 1/2$, E^m denotes the Benioff strain; $m = 1/3$, 2/3, E^m represents the linear scale or area scale of the focal zone respectively; $m=0$, Y is equal to N^+/N^-, where N^+ and N^- denote the number of earthquakes which occurred during the loading and unloading duration, respectively. We adopted $m = 1/2$ in this paper.

In order to predict earthquakes in terms of LURR, some problems should be solved. The first one is how to load and unload the crustal blocks with sizes of hundreds and even thousands of kilometers. One of the means to load and unload the crustal blocks is by the earth tide. Tidal forces exerted by the Moon and Sun produce continuously varying stresses in the earth's crust. How to calculate the tide induced stress in the crust has been elucidated in many works (MELCHIER, 1978; YIN, 1987, 1993; YIN and YIN, 1991; YIN *et al.*, 1994a,b; 1995, 2000; MARUYAMA, 1993).

The second problem is the criteria to judge loading and unloading. We adopt the Coulomb failure hypothesis to judge loading or unloading according to the sign of the increment of Coulomb failure stress which was denoted by CFS in recent literatures (e.g., HARRIS, 1998, 2000; REASENBERG and SIMPSON, 1992).

$$CFS = \tau_n + f\sigma_n, \tag{7}$$

where f, τ_n and σ_n stand for inner frictional coefficient, shear stress and normal stress (positive in tension), respectively, $\mathbf{n}$ is the normal of the fault plane on which the CFS

reaches its maximum. ΔCFS denotes the increment of CFS. If $\Delta CFS > 0$, it is referred to as loading; otherwise $\Delta CFS < 0$ is referred to as unloading.

It is well known that the resultant stress σ_{ij} in the crust consists of tectonic stress σ_{ij}^T and the tide induced stress σ_{ij}^t. Since the level of σ_{ij}^T in the crust (in the order of 10^6–10^8 Pa) is considerably higher than the level of σ_{ij}^t (10^3–10^4 Pa), the directions of the principle stress of the crust resultant stress and then also the direction of **n** can be determined by the tectonic stress only. However, the change rate of tidal-induced stress is much larger than the change rate of the tectonic stress (VIDALI *et al.*, 1998) so ΔCFS is mainly due to tidal-induced stress that could be calculated precisely. The calculation of elastic deformation of the earth can be formulated as a system of six differential equations of first order. Following and improving MOLODENSKSY-TAKEUCHI's work, we calculate the tide-induced stress components of any section in the crust in terms of the Runge-Kutta numerical method (MELCHIOR, 1978; YIN and YIN, 1991) . The shear and normal stress on the fault plane with normal **n** can be obtained by stress tensor transform after which the CFS can be calculated easily according to Eq. (7).

2. Inspection of Real Earthquake Cases

The retrospective inspections of hundreds of earthquake cases have validated the LURR method (YIN *et al.*, 1995, 2000). For more than 80% of examined cases the Y value fluctuates around 1 during the early stage of the seismogenic process and it rises when the region approaches the onset of a strong earthquake. Then Y reaches its maximum (significantly larger than 1), but decreases sharply at the eve of the main shock (Figs. 4 and 5). In contrast, we selected seven "stable" regions with low seismicity (no earthquakes with $M \geq 4$ occurred in the examined period) on the Chinese mainland and analyzed the variation in Y for more than two decades (from 1970 to 1992). For all seven regions, the Y value always fluctuated slightly around 1 during the duration (YIN *et al.*, 1995). These results suggest that LURR is available to be a precursor of strong earthquakes.

Certain numerous parameters are necessary such as the strike direction, dip and slip direction of fault, time window and the size and shape of the spatial window, etc. to calculate the variation of LURR. Unfortunately we did not record all the above parameters completely in our early research due to many reasons such as student floating, computer virus attack and lack of experience, so that even our group now cannot reproduce all the previous results. In view of the above-mentioned facts, we have taken two measures to solve this problem:

(1) automatically recording all the parameters on every chart of $Y(t)$.
(2) setting the scaling relation between radius (the circle region), time window and magnitude based on statistical research (YIN *et al.*, 2002 and ZHANG *et al.*, 2005) as below (Fig. 2)

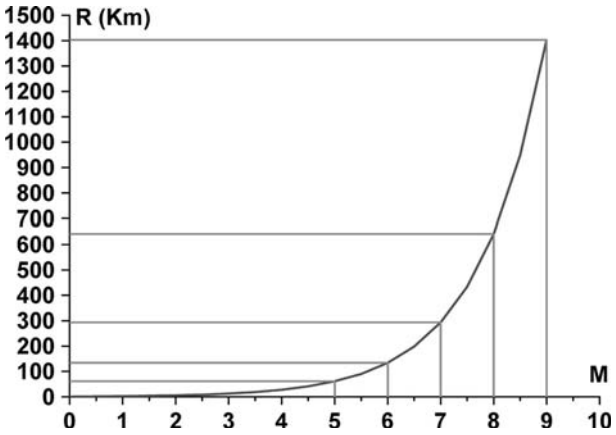


Figure 2
The spatial window size (radius) scale with magnitude.

We often select a circular region as the spatial window in calculating LURR. But in some cases when the epicenters distribute in the forms, which are significantly different from a circular region, we have to select other forms of spatial window as shown in Figure 3.

Figure 4 shows the variation of LURR before some earthquakes with $M \geq 6$ occurred in Southern California, USA and Figure 5 shows the variation of LURR before some earthquakes with $M \geq 7$ occurred on the Chinese Mainland (YIN *et al.*, 1995).

It is well known that the tectonics settings in Southern California, the Chinese Mainland and Japan are significantly different. But the variations of LURR in these regions are similar so that it is suggested that LURR could serve as a precursor of strong earthquakes which occurred in different tectonic settings.

There is a question which was often posed to us: Many scientists have studied the problem of the triggering of earthquakes by the Earth is tide and they found that earthquake occurrence does not correlate well with the Earth's tides (KNOPOFF, 1964; EMTER, 1998; VIDALI *et al.*, 1998). Why were the LURR values in some periods and regions (before a strong earthquake) in our studies considerably larger than 1 (which means in such a situation that earthquake occurrence correlates strongly with the Earth's tides)?

The Earth's crust is not perpetually in a critical state according to the critical point hypothesis (SORNETTE and SORNETTE, 1990; SORNETTE and SAMMIS, 1995; HUANG *et al.*, 1998; JAUME and SYKES, 1999; MORA *et al.*, 2000; RUNDLE *et al.*, 2000). The occurrence of a large earthquake in a region certainly dissipates a sufficient proportion of the accumulated energy and stress and to remove the crust from a critical state so that during the early stage of seismogenic process

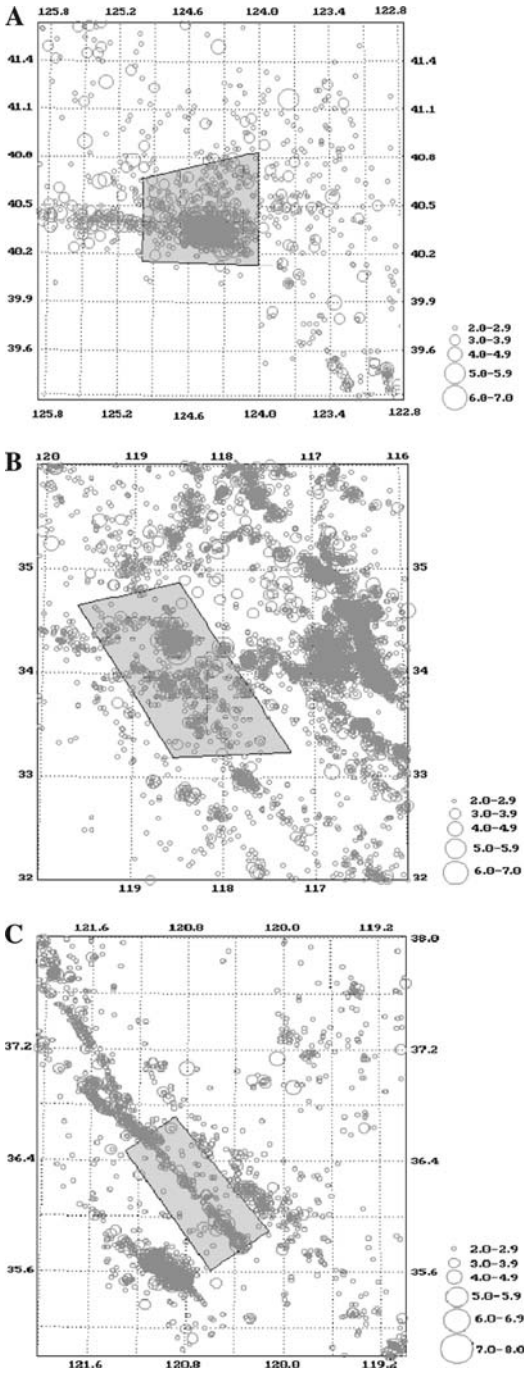


Figure 3
The spatial windows in calculation of LURR for some earthquake cases.

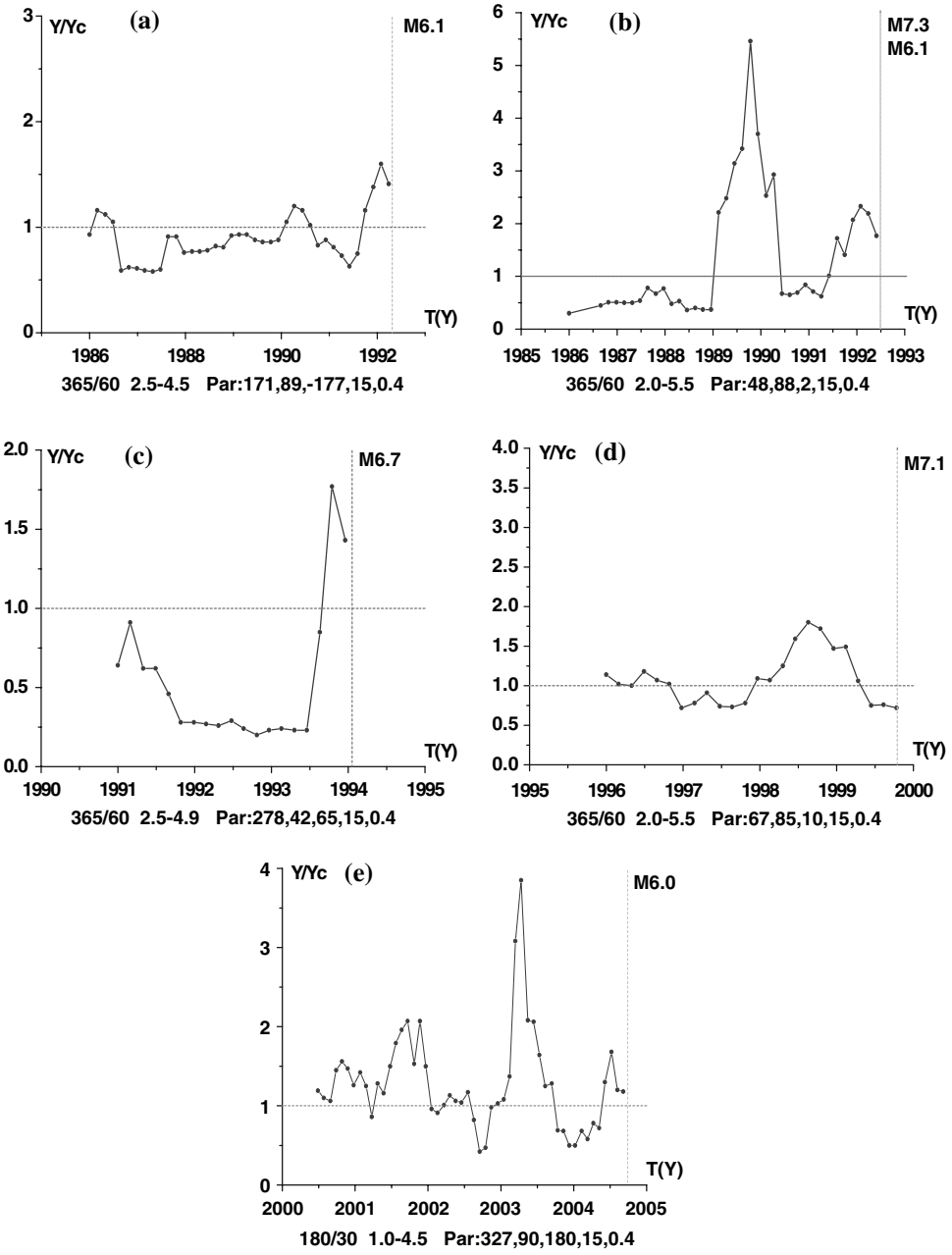


Figure 4

The variation of LURR before earthquakes with $M \geq 6$ occurred in southern California. (a) 1992.4.23, (33.96N,-116.32W), R = 150. (b) 1992.6.28, (34.20N,-116.44W). (c) 1994.1.17, (34.21N,-118.54E). (d) 1999.10.16, (34.59N,-116.27E), R = 320. (e) 2004.9.28, (35.81,-120.365.9).

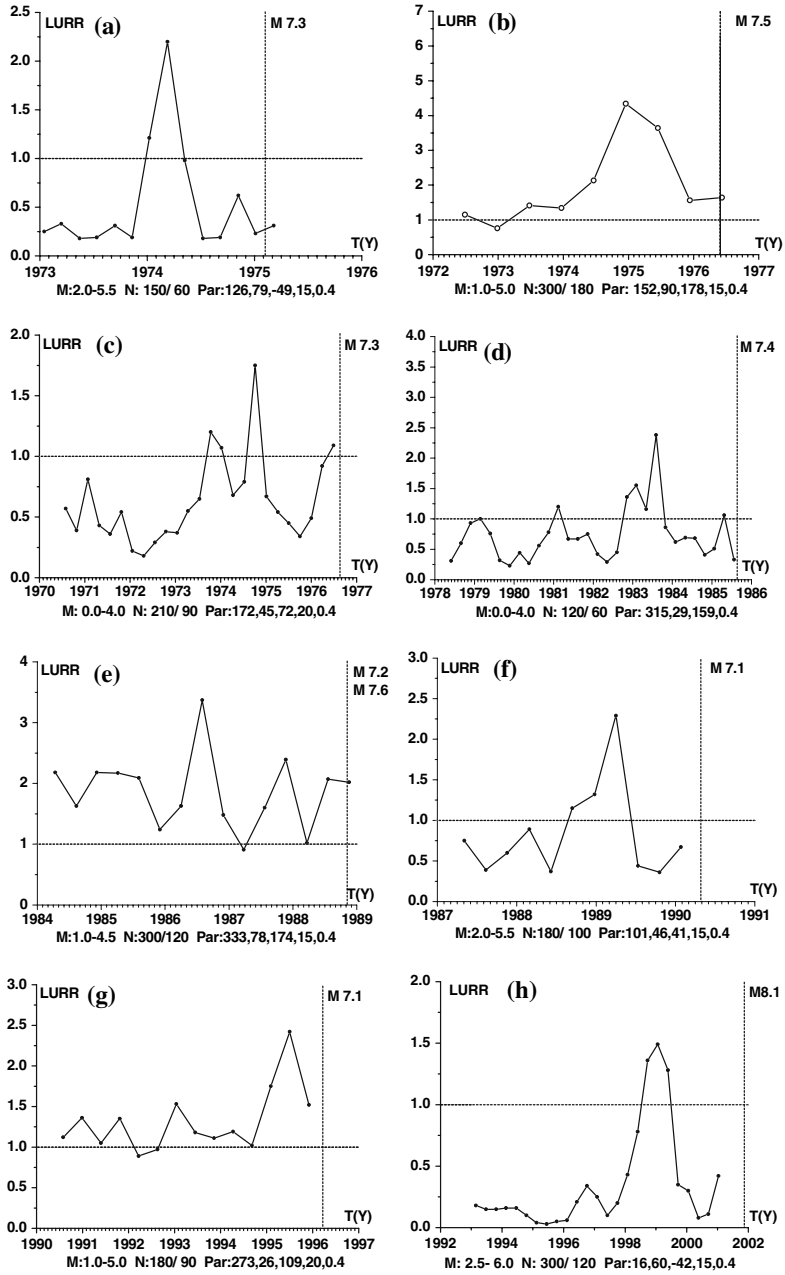


Figure 5
The variation of LURR prior to earthquakes with $M \geq 7$ occurred in the Chinese Mainland. (a) Haicheng Earthquake (1975.2.4). (b) Longling Earthquake (1976.5.29). (c) Songpan Earthquake (1976.8.23). (d) Wuqia Earthquake (1985.8.23). (e) Lancang Earthquake (1988.11.6). (f) Gonghe Earthquake (1990.4.26). (g) Jiashi Earthquake (1996.3.19). (h) Kunlun Mountain (2001.11.14).

(a considerably long period after the occurrence of a large earthquake), the tectonic stress level is low and the Y values are usually close to 1 (which means in such situation that earthquake occurrence correlates weakly with the Earth's tides). Subsequently, tectonic loading drives the crust back towards the critical state. During the establishment of criticality, the crust in this region is damaged severely after which the crust will be sensitive to any tiny external disturbance (such as tidal stress). Just in such a condition could the LURR be larger than 1. Figure 6 is a schematic diagram showing this issue. Since high LURR (significantly larger than unity) only exists in a local region (surrounding the epicenter) and a narrow duration (just a relatively short duration prior to the occurrence of strong earthquakes and this duration is only a very small fraction of the entire reoccurrence time between two strong earthquakes), whereas most of the data set for the above investigations on earthquake tidal triggering **covers a very long time period** which can be comparable to the reoccurrence time **and large region** which is usually considerably larger than the seismogenic region. Naturally no significant correlation exists between the occurrence of the earthquake and the tidal stress. In other words, the sensitive period only occupies a minute fraction of the entire time window which most scientists used to research the triggering earthquake by

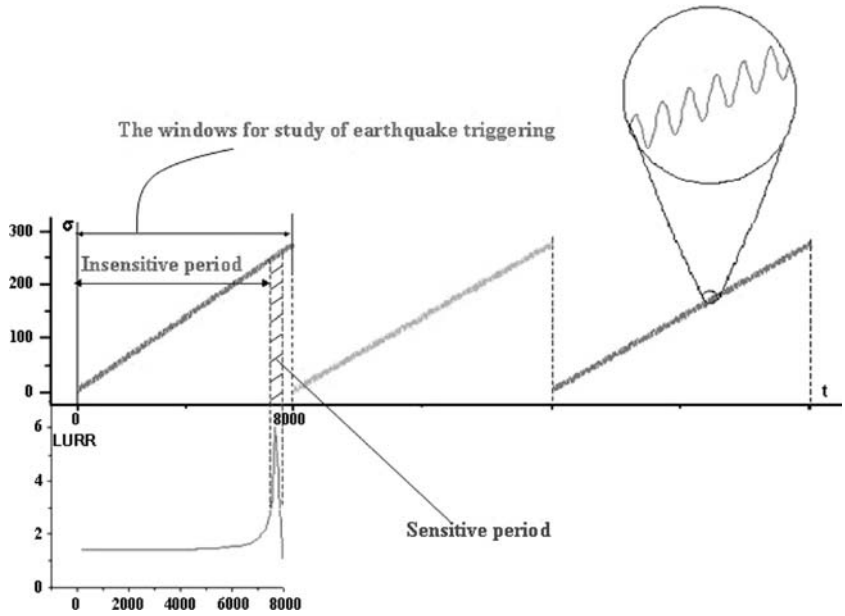


Figure 6
The schematic diagram to illustrate the relation between LURR and the correlation between the occurrence of the earthquake and the tidal stress.

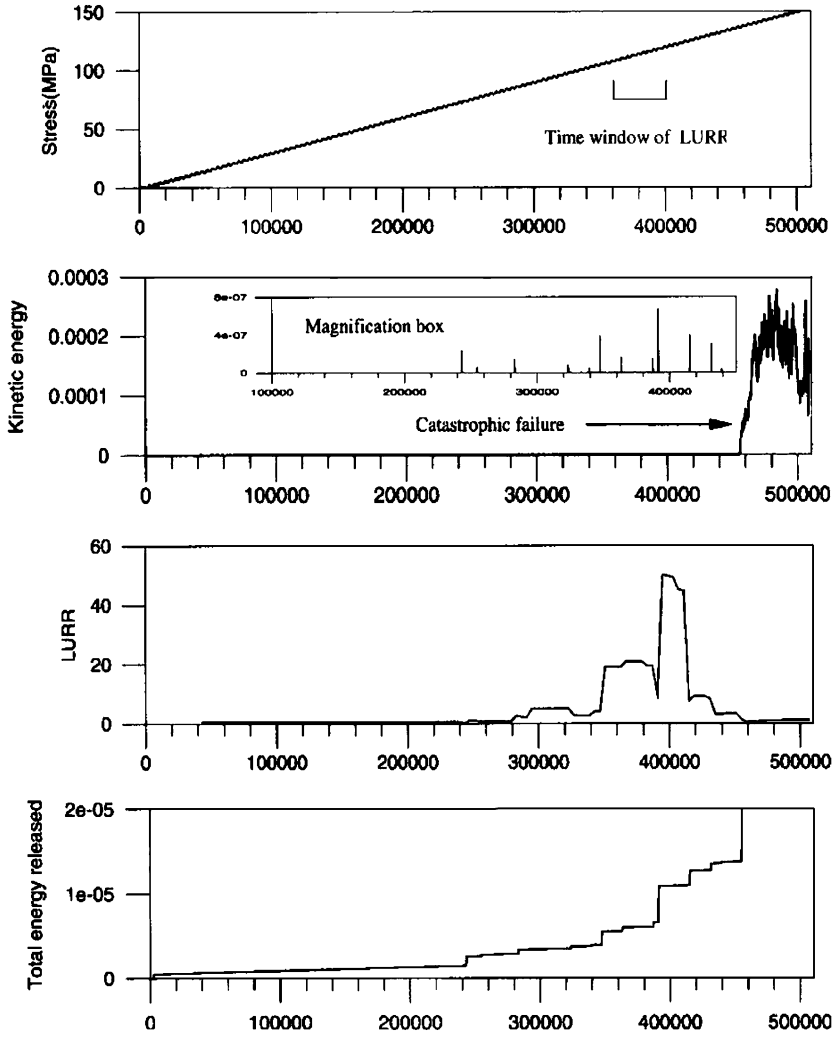


Figure 7
Results of the stress-controlled experiment. From top to bottom, stress measured on the rigid driving plates, kinetic energy within the model, LURR value and cumulative seismic energy release.

tide and the sensitive spatial region is only a slight fraction of the whole region in which most scientists used to research such type of problems.

3. Numerical Simulation

There have been numerous numerical simulations on LURR (MORA *et al.*, 2002; WANG *et al.*, 1998b; 1999; 2000; 2004; WANG, 2000) in recent years. The solid lattice

model, FEM model (RONG, 2006) and chains network model (LIANG *et al.*, 1996) were used to simulate the evolution of damage and the failure of heterogeneous brittle solid and in the meantime to measure the variation of LURR prior to the macro-fracture similar to the strong natural earthquake.

Figure 7 is after MORA *et al.*'s paper (2002). The simulating results show that LURR values become high along with the increasing load and then reach their peak point, then drop rapidly prior to the main event. These results reproduce LURR signals similar to those that have often been observed in earthquake prediction practice, and suggest that LURR is correctly identifying the critically sensitive or unstable regime prior to the catastrophe in the model. The conclusions of this paper said: "The results suggest that LURR provides a good predictor for catastrophic failure and motivates further research to study the underlying physical mechanisms. The results provide encouragement for the prospects of earthquake prediction research and the use of advanced simulation models to probe the physics of earthquakes."

In a recent paper (WANG *et al.*, 2004) the statistical tests of LURR have been conducted and come to the conclusion that: "The statistical studies presented in this paper verified statistical robustness."

4. Laboratory Studies on LURR

There have also been many laboratory studies on LURR (SHI *et al.*, 1994; YIN *et al.*, 2004; YU, 2004; ZHANG *et al.*, this issue). SHI *et al.* (1994) conducted laboratory studies on variations of LURR during rock failure in the early 1990s. The results are similar to theory expectations, that is, Y is close to 1 when the stress is small, Y increases when beyond the elastic phase, and Y grows remarkably just before fracture.

In terms of monitoring acoustic emission (AE) in rock mechanics experiments, it is possible to trace the growth of cracks in rock specimens, to observe the damage evolution of the loaded specimens and then to recognize imminent criticality. Therefore significant similarity or overlap exists between the studies of AE and seismology since both disciplines are concerned with the generation and propagation of stress wave, whereas at different geometry scales and over different frenquency ranges. Based on such considerations, we have carried out AE experiments for large-scale rock specimens by means of international cooperation among Chinese, Russian, Australian and Japanese scientists in 2001 and 2003, respectively. The detailed descriptions of these experiments have been issued in our papers (KYKCEHKO and YIN, 2003; YIN *et al.*, 2004; ZHANG *et al.*, this issue).

The geometry of specimens is rectangular prisms (Fig. 8) with a maximum size reaching 105 cm involving four kinds of rocks (granite, gneiss, sandstone and marble).

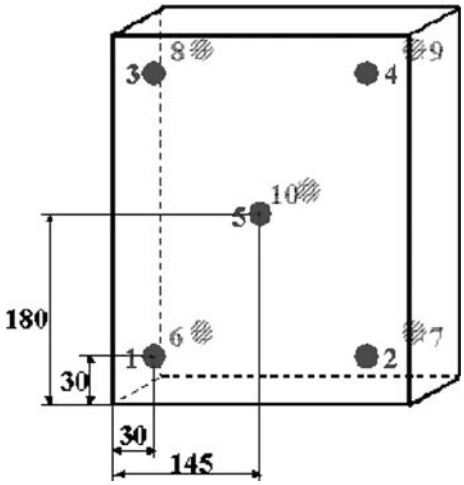


Figure 8
Geometry of the specimens, the loading conditions, and the arrangement of *AE* sensors.

The specimens are loaded in two directions: the axial stress σ_1 and lateral stress σ_2.
Another principal stress σ_3 is zero so that

$$\sigma_1 \neq \sigma_2 \neq \sigma_3.$$

In other words, the stress state is a tri-axial stress state. Under such a stress state most of the specimens undergo shear failure.

In order to simulate the solid Earth tides induced by the Sun and the Moon, the specimens are loaded by the combined action of a constant loading rate, intended to simulate tectonic loading, and a small sinusoidal stress, analogous to the solid Earth tides. The loading history and the typical experimental results are shown in Figure 9. The LURR calculated from the experimental data increases from unity to a high value, and then decreases on the eve of macro-fracture. The experimental results coincide with the seismological observations very well (refer to Figs. 4 and 5).

Here we discuss another question that was also often posed to us: Why the LURR value decreases rapidly on the eve of an earthquake?

According to our understanding, the peak point of LURR indicates the formation of macro-crack in loaded rock specimen (JAEGER and COOK, 1976; RONG, 2006) or the beginning of nucleation of an earthquake (OHNAKA and KUWAHARA, 1990; DIETERICH, 1992). After that the evolution of the system will develop into the so-called self-driving stage. That means it obeys its own law of evolution and will no longer be sensitive to the external disturbance so that the LURR value decreases rapidly on the eve of an earthquake or catastrophic failure of the specimen.

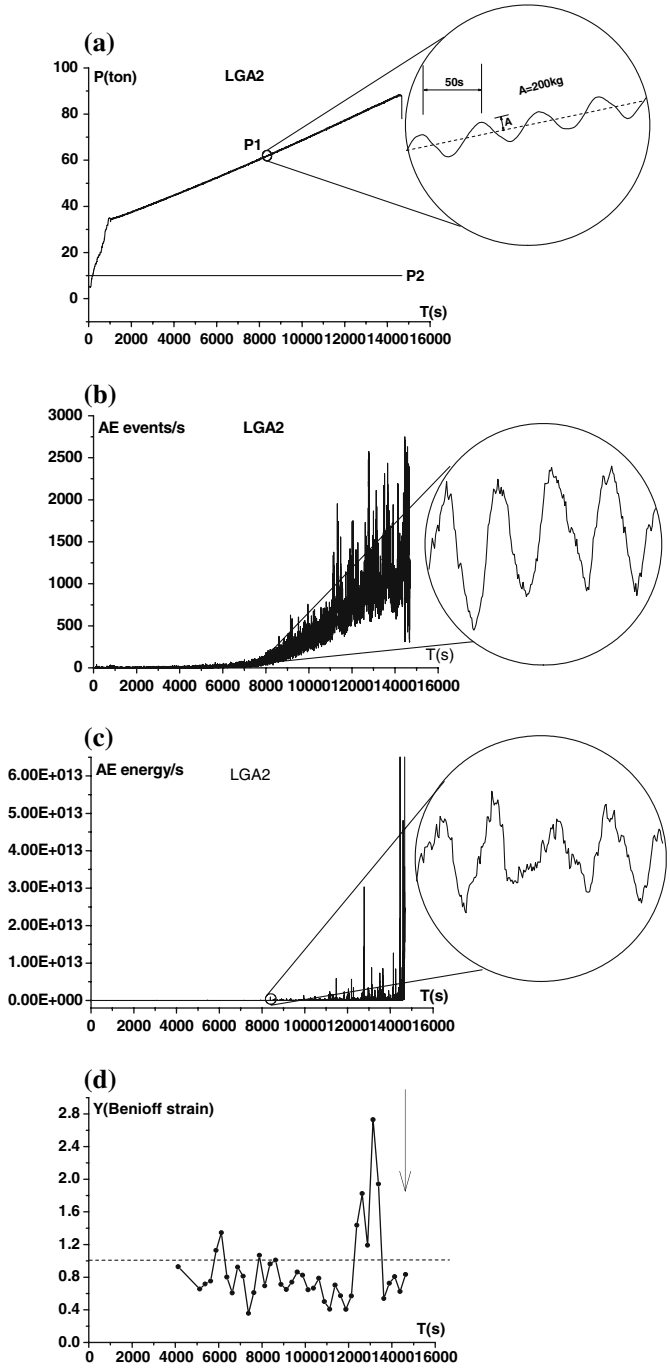


Figure 9
(a) The typical loading history in experiment, (b) the AE rate, (c) the released AE energy per unit time and (d) the variation of LURR during the experimental course.

Some scientists have researched the phenomenon of critical sensitivity which is close relative to LURR (XIA *et al.*, 2002; XU *et al.*, 2004; ZHANG *et al.*, 2004). The numerical simulation of the damage and fracture process for brittle inhomogeneous material also shows a peak of sensitivity prior to the catastrophic rupture; a phenomenon similar to the variation of LURR (refer to Figs. 2c and 3a in ZHANG *et al.*, 2004).

5. Earthquake Prediction in Terms of LURR

LURR is easily accepted theoretically, laboratory modeling and numerical simulation also proved its validity, case inspections are satisfactory too. However, LURR theory must be tested ultimately in real earthquake practices. Bearing this in mind, we applied LURR to real earthquake prediction practice. The statistics of earthquake prediction using LURR during the period 1993–1998 is listed in List 1 of the paper (YIN *et al.*, 2000), among the total predicted 18 cases, 12 cases were correct, 6 cases were false (the predicted earthquakes have not occurred) and no missing case (a strong EQ occurred, but no prediction). The successful cases included the 1994 Northridge earthquake and the Kanto earthquake (1996.09.11, Ms 6.6, 35.5° N, 140.9°E) (YIN *et al.*, 1996).

It is encouraging that the prediction results using LURR have been increasingly better. Recently, especially in 2003–2005, we have made a major breakthrough in intermediate-term earthquake forecasting in terms of LURR. Figure 10 is a map

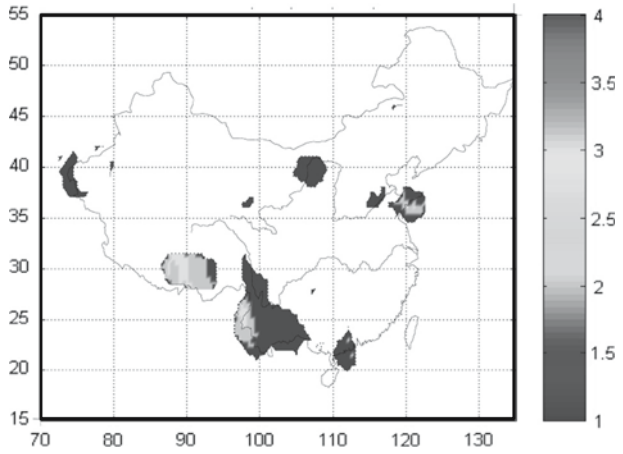


Figure 10
The anomalous LURR regions in the Chinese mainland in 2003 which have a high probability of occurrence for $M \geq 5$ in 2004.

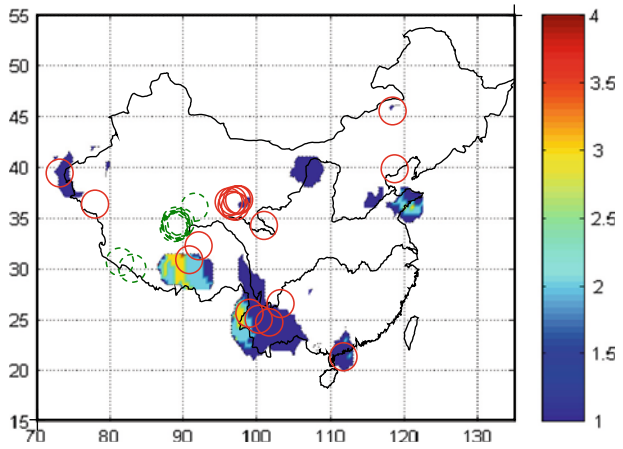


Figure 11
The map of epicenters of earthquakes with magnitude $M \geq 5$ in 2004 in the Chinese mainland. The solid red circles denote the events which occurred in regions in which the observed data were good enough to calculate LURR and the dashed green circles denote the events which occurred in regions where the scarcity of observed data made the LURR calculation impossible.

which shows the anomaly LURR regions in Chinese mainland calculated at the end of 2003. There would be a high probability of occurrence for $M_s \geq 5$ earthquakes in the next year (2004) in these regions. This result was issued in the book *The Research on Seismic Tendency of China in 2004*, edited by the Center for Analysis and Prediction, CSB, pp. 282–285 (YIN *et al.*, 2003) which was issued at the end of 2003 by Seismological Press, Beijing. Figure 11 is a map of epicenters of earthquakes with magnitude $M_s \geq 5$ which occurred in the Chinese mainland in 2004. The red solid circles denote the events $M_s \geq 5$ (total number 17) which occurred in regions, in which the observed data are available to calculate LURR, and the green dashed circles denote the events $M_s \geq 5$ (total number 16) which occurred in regions, where the scarcity of observed data makes the LURR calculation impossible. According to the custom and usage of Chinese seismological circles the earthquakes which occurred in the data scarcity regions should be ignored in the statistics. Therefore there are 15 events among 17 (**a percentage of 88%**) which have fallen into the anomaly LURR regions. In 2005 twelve strong earthquakes with $M \geq 5$ among a total 13 earthquakes (**a percentage of 92%**) have fallen into the LURR anomaly regions (Fig. 11) (YIN *et al.*, 2004). The only missing forecasted one–Jiujiang earthquake with magnitude 5.7 (November 26, 2005) is not in the LURR anomaly regions, rather it is located in the data scarcity regions (ZHANG, 2006).

It is worthy to mention the prediction of the Pakistan earthquake with a magnitude of M 7.8 on October 8, 2005. We began to study the seismic trend of the Euro-Asian seismic belt (30°–50°N, 30°–90°E) in 2003. The outcomes were issued in

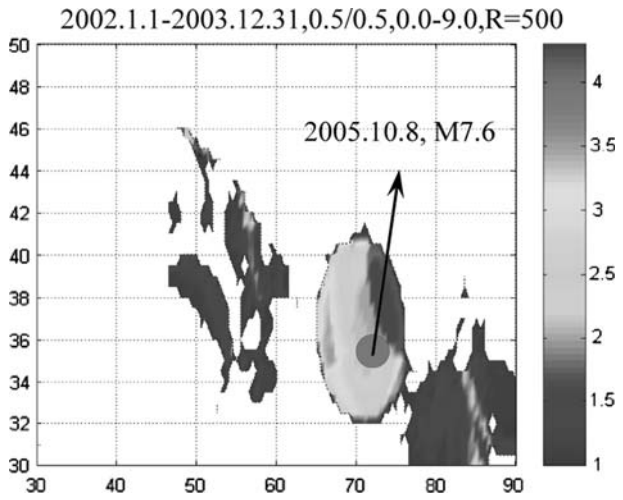


Figure 12
The spatial scanning of LURR in the Europe-Asia earthquake belt in 2002.1.1-2003.12.31.

the Journal *Mechanics in Engineering* (YIN, 2004) and *Supercomputing Newsletter*, sponsored by the Computer Network Information Center, Chinese Academy of Sciences. The Pakistan earthquake on October 8, 2005 exactly fell into the High LURR region (Fig. 12).

Recently the spatial scanning of LURR for the western United States (30°–50°N, 100°–130°W) also has been conducted (Fig. 13), which shows us the anomalous LURR regions in this region during 2003 (from Jan.1, 2003 to Dec. 31, 2003) and the map of epicenters with magnitude also $M \geq 5$ (circles) observed in 2004. All the observed earthquakes with magnitude $M \geq 5$ fell into the LURR anomalous regions.

Figure 14 shows the maps of LURR anomaly regions from the spatial scanning with different special windows (ZHANG *et al.*, 2005). We have obtained the scaling relations of LURR: The size of seismogenic zone (radius R of circle area) should scale with the magnitude of the future earthquake (Fig. 2)(YIN *et al.*, 2002a, b)

$$\mathrm{Log}\, R\,(\mathrm{km}) = 0.087 + 0.34\mathrm{M}, \tag{8}$$

and the duration (months) between the peak of LURR and the occurrence time of the future earthquake should also scale with the earthquake magnitude (ZHANG *et al.*, 2005)

$$\Delta T(\mathrm{month}) = 60(1 - 2.3 * 10^{-0.08M}). \tag{9}$$

According to these scaling relations we predicted the seismic tendency in those regions with high LURR:

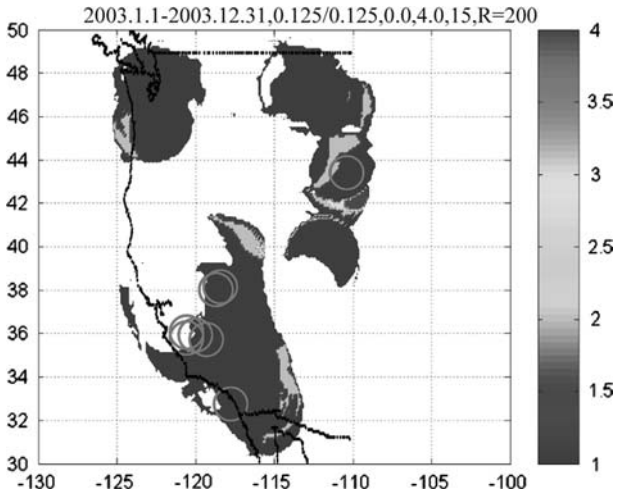


Figure 13
The anomalous LURR regions in the western USA during Jan.1,2003-Dec. 31, 2003 and the map of epicenters with magnitude $M \geq 5$. ◯ denotes an earthquake with magnitude $M \geq 5$.

(1) Earthquakes with magnitude around M5($\pm$0.5) probably occurred in the LURR anomaly regions shown in Figure 14a, especially in the regions before July 31, 2005.
(2) Earthquakes with magnitude around M5.5($\pm$0.5) probably occurred in the LURR anomaly regions shown in Figure 14b, especially in the region before Dec. 31, 2005.
(3) Earthquakes with magnitude around M6.5($\pm$0.5) probably occurred in the LURR anomaly regions shown in Figure 14c, especially in the region before Dec. 31, 2006.
(4) Earthquakes with magnitude near M6($\pm$0.5) probably occurred in the LURR anomaly regions shown in Figure 14d, especially in the region before Dec. 31, 2005.

6. The Random Distribution of LURR

The preparation and occurrence process of earthquakes is controlled not only by deterministic dynamical law but also affected by some stochastic or disorder factors. Therefore we have to study the influence of random factors on LURR in order to judge the height to which the Y value reaches can be considered as a precursor under the specified confidence (e.g., 0.90, 0.95 and 0.99) (ZHUANG and YIN, 1999).

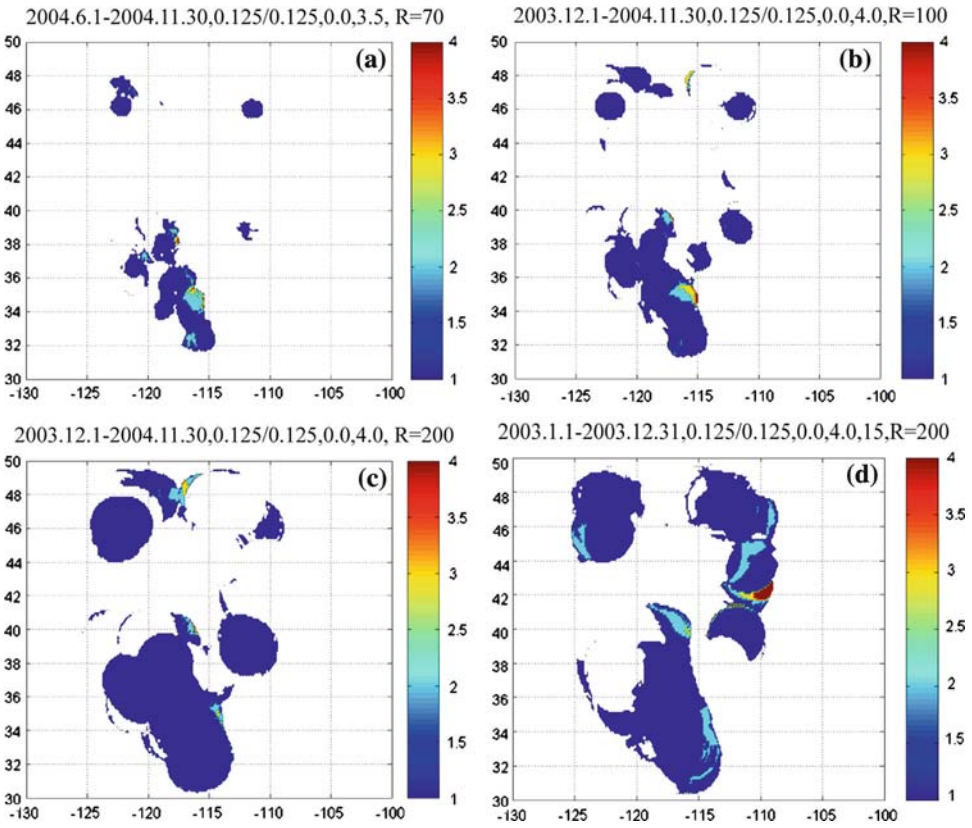


Figure 14
The anomalous LURR regions in the western USA with different special windows. (a) R = 70 km; (b) R = 100 km; (c) R = 200 km; (d) R = 200 km with different time windows.

The Poisson model is used to describe the occurrence times of earthquakes. We assume that earthquakes in a region obey the following basic assumption:

1. The earthquakes occur according to a Poisson process with a constant rate λ. The number of earthquakes occurring in the time interval $[0,T]$ has a Poisson distribution with expectation λT, i.e.,

$$\Pr\{N = n\} = \frac{(\lambda T)^n}{n!} e^{-\lambda T}. \tag{10}$$

2. The distribution of the magnitudes obeys the Gutenberg-Richter law, i.e., an exponential distribution with the probability density function.
3. The probabilities of an earthquake falling in a loading period and unloading period are equal, both 1/2.

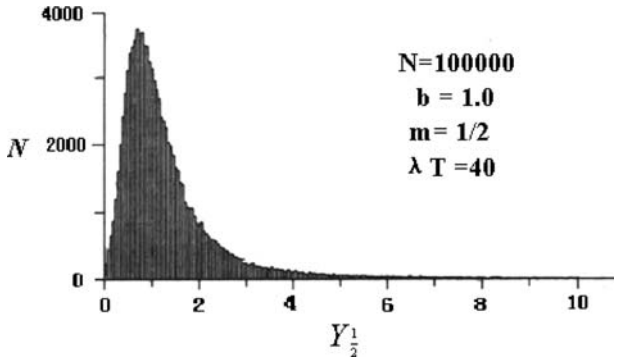


Figure 15
The random distribution of Y under the specified condition.

Based on the above assumptions, a simulation algorithm for computing the distribution and the confidence bands of the LURR is outlined below:

1. For each time interval (assumed as unit time interval), simulate two random variables P, Q belonging to a Poisson distribution of rate $\lambda/2$, where λ is the occurrence rate of earthquakes. P, Q can be regarded as the number of loading earthquakes and unloading earthquakes occurring in the time interval, respectively.
2. According to the given b value, simulate P magnitudes for the loading earthquake and Q magnitudes for the unloading earthquakes.
3. Calculate the Y value.
4. Repeat step 1 to 3 for one million times, and draw the histogram of Y values, which could be regarded as the probability density function (p.d.f) of LURR. Figure 15 shows an example for $b = 1$, $\lambda T = 40$ and $m = 1/2$.
5. Find the 0.90, 0.95 and 0.99 confidence bands from the p.d.f of LURR, respectively. Figure 16 is an example of such kind of simulated results. For example, for the condition: Occurrence rate $= 50$, $b = 1$, $m = 1/2$, confidence 95%, Y value should be equal or greater than 2.4 which denoted as Y_c.

The results show that the variation of LURR is controlled by the occurrence rate for the Poisson model and the parameter $S(S = b/m$, where b is the constant in the Gutenberg-Richter law and m is the power in (6)). The larger the occurrence rate and the S value, the more stable or more concentrated around 1 the Y value is.

Instead of Y the parameter Y/Y_c will be used to measure the anomaly of LURR hereafter. It is considered that LURR is an anomaly when $Y/Y_c > 1$. The larger the value of Y/Y_c, the greater the possibility of strong earthquakes in the specified region and time window will be.

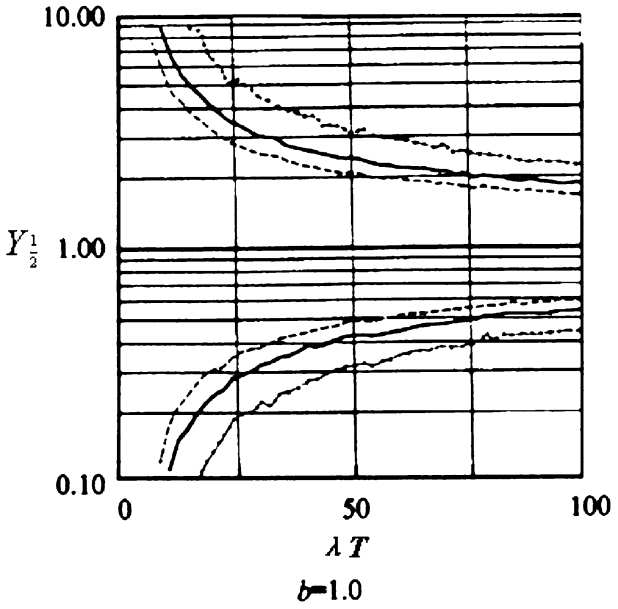


Figure 16
Variation of confidence bands of Y with occurrence rates under the Poisson model.

7. Perspective

The LURR has been become a new promising approach to earthquake prediction after twenty years effort and there exits considerable room for improvement and development in it (YIN, 2005; ZHANG, 2006).

As mentioned above, the LURR is a parameter which mirrors the closeness to strong earthquakes and serves as an approach to natural earthquake prediction. It has also been applied to the prediction of reservoir induced earthquakes (CHEN and YIN *et al.*, 1995), mining rock bursts (YIN *et al.*, 2004b) and landslides (XU and HUANG, 1995; JIANG, 2004). It could potentially be used for the prediction of other critical phenomena and the damage evaluation or "healthy-diagnoses"of major engineering structures.

Acknowledgements

The authors gratefully acknowledge the support from Special Funds for Major State Basic Research Project under Grant Nos.2004 CB418406, NSFC under grant Nos. 10232050, 10572140, the Informatization Construction of Knowledge Innovation Projects of the Chinese Academy of Sciences "Supercomputing Environment Construction and Application" (INF105-SCE-2-02).

The authors are grateful to many of our teachers, friends, colleagues, leaders and students for their support, encouragement and assistance. They are Qian Xue Sen, Zheng zhe-min, Qin Xin Ling, Bai Yi Long, Mei Shi Rong, Chen Zhang Li, Xia Mengfen, Liang Nai Gang, Ke Fujiu, Chi Xuebin, Li Jie, Keiiti Aki, Leon Knopoff, Peter Mora, Bernard Minster, Dave Jacksen, Takuo Maruyama and Zheng-kang Shen.

REFERENCES

BAI, Y.L., LU, C.S., KE, F.J, *et al*. (1994), *Evolution induced catastrophe,* Phys. Lett. *A185*, 196–201.

BOWMAN, D.D., OUILLEN, G., SAMMIS, C.G., SORNETTE, A., and SORNETTE, D. (1998), *An observational test of the critical earthquake concept,* J. Geophys. Res. *103*, 24,359–24,372.

BAKUN, W.H., and LINDH, A.G. (1985), *The Parkfield, California, earthquake prediction experiment,* Science *229*, 619–624.

CHEN, J.M., ZHANG, Z.D., and YANG, L.Z. (1994), *Seismic abnormalities in LURR of ground water lever*, Earthquake *1*, 73–78.

CHEN, X.Z. and YIN, X.C. (1995), *Application of LURR theory to the earthquake prediction for reservoir-induced earthquakes*, Earthquake Research in China *11*, 361–367.

COTTON, L. A. (1922), *Earthquake frequency, with special reference to tidal stresses in the lithosphere*, Bull. Seismol. Soc. Am. *12*, 47–198.

CUNDALL, P.A. and STRACK, O.D.L. (1979), *A discrete element model for granular assemblies*, Geotechnique *29*, 47–65.

DIETERICH J.H. (1992), *Earthquake nucleation on faults with rate- and state-dependent strength*, Tectonophysics *211*, 115–134.

EMTER, D. (1997), *Tidal triggering of earthquakes and volcanic events*. In: (WILHELM, H., *et al*., ed.), *Tidal Phenomena* (Springer-Verlag 1997), pp. 293–309.

HARRIS, R.A. (1998), *Introduction to special section: stress triggers, stress shadows, and implication for seismic hazard*, J. Geophys. Res. *103*, 24347–24358.

HARRIS, R.A. (2000), *Earthquake stress triggers, stress shadows, and seismic hazard,* Current Science *79*, 1215–1225.

HUANG, Y., SALEUR, H., SAMMIS, C.G. and SORNETTE, D. (1998), *Precursor, aftershocks, criticality and self-organization criticality*, Europhys. Lett. *41*, 43–48.

HUANG, R. and XU, Q. (1997), *Scientific Analysis Principle and Its Applications to Generalized System in Engineering Geology* (Geological Press, Beijing 1997) (in Chinese).

JAEGER, J.C. and COOK, N.G.W., *Fundamentals of Rock Mechanics* (Chapman and Hall, London, 1976).

JAUME, S.C. and SYKES, L.R. (1999), *Evolving towards a critical point: A review of accelerating seismic moment/energy release prior to large and great earthquakes*, Pure Appl. Geophys. *155*, 279–306.

JIANG, T. (2004), The Analysis of landslide in terms of Load/Unload Response Ratio, Ph.D. Thesis, Institute of Geology, Chinese Seismological Bureau (in Chinese).

KANAMORI, H. and ANDERSON, D.L. (1975), *Theoretical basis of some empirical relation in seismology,* Bull. Seismol. Soc. Am. *65*, 1073–1096.

KNOPOFF, L. (1964), *Earth tide as a triggering mechanism for earthquakes*, Bull Seismol. Soc. Am. *54*, 1865–1870.

KRAJCINNOVIC, D., *Damage Mechanics* (Elsevier, Amsterdam, 1996).

KYKCEHKO, B.C. and YIN, X.C (2003), *Влияние слабых механических воздействий наповедение очага разрушения*. Известия Академии Наук. Серия Физическая, *том 67*, No. 6c. 877–881 (in Russian).

LEMAITRE, J., *Formulation and identification of damage kinetic constitutive equations*. In *Continuum Damage Mechanics* (ed. Krajcinnovic, D.) (Springer-Verlag, Wien, New York 1987).

LIANG, N., LIU, Q., LI, J., and SONG, H., *A chains net-work model simulating meso- mechanics behavior and micro-damage evolution of in situ reinforced ceramics*. In *Advanced in Engineering Plasticity and its Applications* (eds. Abe, T. and Tsuta, T.) (Pergamon, Amsterdam, 1996), pp. 141–146.

MARUYAMA, T. (1993), *Earthquake Prediction in China*, Zisin *19*(5), 68–76 (in Japanese).

MELCHIER, P. The *Tide of the Planet Earth,* (Pergamon Press, New York 1978).

MEAKIN, P. (1991), *Model for material failure and deformation*, Science *252*, 226–234.

MORA, P. and PLACE, D. (1993), *A lattice solid model for the nonlinear dynamics of earthquakes*, Int. J. Mod. Phys. *C4*, 1059–1074.

MORA, P., PLACE, D., ABE, S., and JAUME, S. (2000), *Lattice solid simulation of the physics of faultzones and earthquakes: The model, results and directions*. In *Geocomplexity and the Physics of Earthquakes* (eds. Rundle, J.B., Turcotte, D.L., and Klein, W.) (AGU, Washington 2000), pp. 105–125.

MORA, P., WANG, Y., YIN, C., PLACE, D., and YIN, X.C. (2002), *Simulation of Load-Unload Response Ratio and critical sensitivity in the lattice solid model*, Pure Appl. Geophys. *159*, 2525—2536.

ODA, M. (1983), *A method for evaluating the effect of crack geometry on the mechanical behavior of cracked rock mass,* Mech. Mater. *2*, 163–171.

OHNAKA, M. and KUWAHARA, Y. (1990), *Characteristic features of local breakdown near a crack-tip in the transition zone from nucleation to unstable rupture during stick-slip shear failure*, Tectonophysics *175*, 197–220.

QIANG XU and RUN-QIU HUANG (1995), *Investigation on precursor of slope instability in term of LURR*, Chinese J. Geolog. Disasters *6*, 2–16, (in Chinese).

REASENBERG, P.A. and SIMPSON, R.W. (1992), *Response of regional seismicity to the static stress change produced by the Loma prieta earthquake*, Science *255*, 1687–1690.

ROELOFFS, E.A. and LANGBEIN, J. (1994), *The earthquake prediction experiment at parkfield, California,* Rev. Geophys. *32*, 315–336.

RONG, F. (2006), *Multiscale Simulation of damage Evolution in Heterogeneous Brittle Media,* Ph.D. Thesis of the Graduate School, China Academy Science.

RUNDLE, J.B. *et al*. (2000), Precursor seimic activation and critical-point phynomena, Pure Appl. Geophys. *157*, 2165–2182.

SHI, X., XU, H., WAN, Y., LU, Z., and CHEN, X. (1994), *The rock fracture under simulated tide force—Laboratory study on the Loading and Unloading Response Ratio (LURR) theory*, Acta Geophysica Sinica *37*, 631–636.

SORNETTE, A. and SORNETTE, D., (1990), *Earthquake rupture as a critical point: Consequences for telluric precursors,* Tectonophysics *179*, 327–334.

SORNETTE, D. and SAMMIS, C. G. (1995), *Complex critical exponents from renormalization group theory of earthquake prediction*, J. Phys. I. France *5*, 607–619.

VIDALI, J.E., AGNEW, D.C., JOHNSTON, M.J.S., and OPPENHEIMER, D.H. (1998), *Absence of earthquake correlation with earth tides: An indication of high preseismic fault stress rate*, J. Geophys. Res. *103*, 24567–24572.

WANG, H.T., PENG, K.Y., ZHANG, Y.X., WANG, Y.C., and YIN, X.C. (1998a), *Characters of variation of LURR during the earthquake sequence of Xinjiang*, Chinese Sciences Bulletin *43*, 1752–1755.

WANG, T.W. (1995), *The application of Load/Unload Response Ratio in earthquake prediction by magnetic method*, Observation and Research of Seismic Geomagnetism *16*, 26–29 (in Chinese).

WANG, Y. C., YIN, X.C., and WANG, H.T. (1998b), *The simulation of rock experiment on Load/Unload response for earthquake prediction*, Earthquake Research in China *14*(2), 126–130.

WANG, Y.C., YIN, X.C., and WANG, H.T. (1999), *Numerical simulation on Load/Unload Response Ratio (LURR) theory*, Acta Gephysica Sinica *42*, 669–676.

WANG, Y.C., YIN, X.C., KE, F.J., XIA, M.F., and PENG, K.Y. (2000), *Simulation of rock failure and earthquake process on mesoscopic scale*, Pure Appl. Geophys. *157*, 1905–1928.

WANG, Y.C. (2000), *Simulation of earthquake process and studies on predictability,* Post-Doctor Report, Institute of Mechanics, CAS (in Chinese).

WANG, Y.C., MORA, P., YIN, C., and PLACE, D. (2004), *Statistical tests of Load-Unload Response Ratio signals by lattice solid model: Implication to tidal triggering and earthquake prediction*. Pure Appl. Geophys. *161*, 1829–1839.

XIA, M.F, WEI, Y.J., KE, F.J., and BAI, Y.L. (2002), *Critical sensitivity and transscale fluctuations in catastrophe rupture*, Pure Appl. Geophys. *159*(10), 2491–2509.

XU, Q. and HUANG, R.Q. (1995), *Investigation on precursor of slope instability in term of LURR*, Chinese J. Geological Disasters *6*, 2, 1995 (in Chinese).

YANG, L.Z., HE, S.H., and XI, Q.W. (1994), *Study on the variation of the property of rock elasticity by Load/Unload Response Ratio of tidal volume strain,* Earthquake Research in China *10*, 90–94.

YIN, C. (2005), Exploring the Underlying Mechanism of Load/Unload Response Ratio Theory and its Application to Earthquake Prediction, Ph.D. Thesis, The University of Queensland, Australia.

YIN, X.C. (1987), *A new approach to earthquake prediction,* Earthquake Research in China *3*, 1–7 (in Chinese with English abstract).

YIN, X.C. and YIN, C. (1991), *The precursor of instability for nonlinear system and its application to earthquake prediction*, Science in China *34*, 977–986.

YIN , X.C. (1993), *A new approach to earthquake prediction* (Russia's *Nature*), *1*, 21–27 (in Russian).

YIN, X.C., YIN, C., and CHEN, X.Z. (1994a), *The precursor of instability for nonlinear system and its application to earthquake prediction—The Load- Unload Response Ratio theory*. In *Nonlinear Dynamics and Predictability of Geophysical Phenomena* (eds. Newman, W.I., Gabrelov, A. and Turcotte, D.L.) Geophysical Monograph *83*, IUGG Volume *18*, 55–60.

YIN, X.C., CHEN, X.Z., SONG, Z.P., and YIN, C. (1994), *The Load-Unload Response Ratio theory and its application to earthquake prediction*, J. Earthquake Prediction Research *3*, 325–333.

YIN, X.C., CHEN, X.Z., SONG, Z.P., and YIN, C. (1995), *A new approach to earthquake prediction: The Load/Unload Response Ratio (LURR) theory*, Pure Appl. Geophys. *145,* 701–715.

YIN X.C., SONG Z.P., and WANG, Y.C. (1996), *The temporal variation of LURR in Kanto and other regions of Japan and its application to earthquake prediction*, Earthq. Res. in China *10*, 381–385.

YIN X.C., WANG Y.C., PENG K.Y., BAI Y.L., WANG H.T., and YIN X.F., (2000), *Development of a new approach to earthquake prediction—Load/unload Response Ratio (LURR) theory*, Pure Appl. Geophys. *157*, 2365–2383.

YIN, X.C., MORA, P., PENG, K.Y., WANG, Y.C., and WEATHERLY, D. (2002a), *Load-Unload Response Ratio and accelerating moment/energy release, critical region scaling and earthquake prediction*, Pure Appl. Geophys. *159*, 2511–2524.

YIN, X.C., PENG, K.Y., YU, H.Z., WANG, Y.C., YIN, C., and ZHANG, Y.X. (2002b), *Scaling of LURR Critical Region*, 3rd ACES Workshop, May 5–10, 2002, Island of Maui, Hawaii, U.S.A.

YIN, X.C., YU, H.Z., KUKSHENKO, V., XU, Z.Y., WU, Z.S., LI, M., PENG, K.Y. ELIZAROV, S., and LI, Q. (2004b), *Load-Unload Response ratio (LURR), Accelerating Energy release (AER) and State Vector evolution as precursors to failure of rock specimens*, *Pure Appl. Geophys 161*, 11–12, 2405–2416.

YIN, X.C. (2005), *A new approach to earthquake prediction-Load/Unload Response Ratio (LURR) theory*, Mechanics in Engineering *26*, 1–7, (in Chinese).

YU, H.Z. (2004), *Experimental research on precursors of brittle heterogeneous media and earthquake prediction methods,* Ph.D. Thesis, Institute of Mechanics, CAS (in Chinese).

ZHANG, H.H., YIN, XIANGCHU, and LIANG NAIGANG (2005), *The spatial variation of LURR and seismic tendency in western United States*, Earthquake Research in China *19*(3), 338–345.

ZHANG H.H. (2006), *Prediction of Catastrophic failure in Heterogeneous Brittle Media—Study and Practice of Load/Unload Response Ratio (LURR)*, Ph.D. Thesis, Graduate School, CAS.

ZHANG, J.H. (1995), *The analysis of the geomagnetic abnormalities in Load/Unload Response Ratio method*, Observation and Research of Seismic Geomagnetism *16*, 61–63 (in Chinese).

ZHANG, X.H., XU, X.H., XIA, M.F., and BAI, Y.L. (2004), *Critical sensitivity in driven nonlinear threshold systems*, Pure Appl. Geophys. *161*, 1931–1944.

ZHANG YONGXIAN and XIANGCHU YIN (2004), *LURR and the San Simeon M6.5 earthquake in 2003 and the seismic tendency in California*, The 4th ACES Workshop, July 9–14, 2004, Beijing China.

ZHEN, X. P. (1996), *Earthquake and Load/Unload Response Ratio of terrestrial magnetic field to solar wind,* Observation and Research of Seismic Geomagnetism *17,* 49–53 (in Chinese).

ZHUANG, J.C. and YIN, X.C. (1999), *Random distribution of the Load/Unload Response Ratio (LURR) under assumptions of poisson model*, Earthquake Research in China *15*, 128–138.

(Received November 11, 2004, revised October 10, 2005, accepted December 1, 2005)

Pure appl. geophys. 163 (2006) 2343–2351
0033–4553/06/122343–9
DOI 10.1007/s00024-006-0125-z

Pure and Applied Geophysics

LURR and the San Simeon M 6.5 Earthquake in 2003 and the Seismic Tendency in CA

Yongxian Zhang,[1,2] Xiangchu Yin,[1,2] Keyin Peng,[1,2] Haitao Wang,[3] Jianchang Zheng,[4] Yongjia Wu,[1] and Lang-ping Zhang[2]

Abstract—The spatial and temporal variation of LURR (Load/Unload Respond Ratio) in California during April 2002 to June 2004 was studied in this paper. The result shows that before the San Simeon earthquake (35.7 N, 121.1 W) on Dec. 22, 2003, Y/Y_c anomalous region occurred successively near the epicenter from April 2002 to June 2002, and the maximum anomaly of Y/Y_c occurred in May, 2002. The published research work pointed out that the Y/Y_c anomaly near the San Simeon earthquake appeared from March, 2002. Compared with the five earthquake cases out of the six with M ≥ 6.5 in California during the period from 1980 to 2001, the maximum Y/Y_c and duration of Y/Y_c anomaly before this earthquake are among the normal ranges, but the time delay from the maximum anomaly time to the occurrence time of this earthquake is the longest one. The result also shows that two areas with Y/Y_c anomalies occurred from Oct. 2002 and Dec. 2002, respectively. According to statistical characteristics of the relationship between Y/Y_c anomalies and the coming earthquakes, the seismic tendency in California was discussed in this paper.

Key words: LURR (Load/Unload Respond Ratio) anomaly, San Simeon earthquake, seismic tendency, California.

1. Introduction

The physical essence of an earthquake is failure or instability of the focal media. When a seismogenic system is in a stable state, its response to loading is nearly the same as its response to unloading, whereas when the system is in an unstable state, the response to loading is more sensitive than that to unloading (Yin, 1987; Yin and Yin, 1991; Yin *et al.*, 1995, 2000). Consequently a parameter LURR (Load/Unload

[1] China Earthquake Networks Center, China Earthquake Administration, Beijing 100036, China. E-mail: zhang.yongxian@263.net; yixc@btmail.net.cn; peng_keyin@163.com; wuyongj@mails.gucas.ac.cn
[2] LNM, Institute of Mechanics, Chinese Academy of Sciences, Beijing 100080, China. E-mail: yin@lnm.imech.ac.cn
[3] Seismological Bureau of Xinjiang Uygur Autonomous Region, Urumqi 830011, China. E-mail: whtxj@mail.xj.cninfo.net
[4] Shandong Earthquake Administration, Jinan 250014, China. E-mail: zjcmail@yeah.net

Response Ratio) was put forward by YIN (1987) to describe if a system is in stable or unstable state, which is defined as

$$Y = X^+/X^-, \tag{1}$$

where X^+ and X^- are the response rates during loading and unloading measured by some method. According to the LURR idea, when a seismogenic system is in a stable or linear state, $Y \sim 1$, whereas when the system lies outside of the linear state, $Y > 1$.

In earthquake prediction practice with LURR, the earth's tide is taken as a load and unload source, and the response parameter of X is chosen as Benioff strain calculated from magnitudes of earthquakes in this region. The periods of loading and unloading are determined by calculating perturbations in the Coulomb Failure Stress induced by the earth's tides (YIN *et al.*, 1995, as illustrated in the next part). Experimental and numerical simulation have validated LURR (MORA *et al.*, 1999, 2000; WANG *et al.*, 1999; WANG *et al.*, 2000; WANG *et al.*, 1999). In retrospective studies, high Y values have been observed months to years prior to most significant events and some successful intermediate-term earthquake predictions have been made (YIN *et al.*, 2000).

Based on the theory of LURR and its recent development, spatial and temporal variation of Y/Y_c (value of LURR/critical value of LURR) in Southern California and its adjacent area (32° N to 40° N, 114° W to 125° W) during the period from 1980 through March, 2002 has been studied (ZHANG *et al.*, 2004). The scanning results show that obvious Y/Y_c anomalies occurred before five out of the total six earthquakes with $M \geq 6.5$. The areas with Y/Y_c anomaly are near the epicenters of the strong earthquakes and the Y/Y_c anomalies occur months to years prior to the earthquakes. According to the contour of LURR in March 2002, a forecast of "There might be a moderate earthquake near 5 around (36° N, 119° W) and (36° N, 121.6° W) within about 1 year. If the Y/Y_c anomalous region grows larger and persists for one half to one year or more, there might a stronger earthquake in these regions." (Presented at the 3rd ACES Workshop in May, 2002 in Hawaii, ZHANG *et al.*, 2003, 2004).

What happened from April 2002 to Dec. 2003 in California? Two earthquakes with $M \geq 5.0$ occurred in the studied area. One was an M 5.4 (34.31° N, 116.85° W) earthquake on Feb. 22, 2003, and the other was the San Simeon M 6.5 (35.7° N, 121.1° W) earthquake on Dec. 22, 2003. The San Simeon M 6.5 earthquake occurred very near one of the LURR anomalous regions mentioned above (36° N, 121.6° W). In order to obtain the evolution process of LURR before this quake, an earthquake catalogue from April 2002 to June 2004 has been downloaded from ANSS (the former CNSS). Under the same calculation parameters, snapshots of Y/Y_c contours for each month from April 2002 to March 2004 are obtained. From these snapshots, the feature of anomalous Y/Y_c before the San Simeon M 6.5 earthquake was abstracted, and a comparison of this earthquake with the former five cases was given in this paper.

2. Method to Calculate LURR

2.1. Determination of Loading and Unloading Periods

YIN *et al.* (1995) resorted to the Coulomb failure criterion (JAEGER and COOK, 1976) to determine the loading and unloading periods. We calculate perturbations in the Coulomb Failure Stress (CFS) (e.g., HARRIS, 1998; RESERNBERG and SIMPSON, 1992) induced by the earth's tides.

$$\mathrm{CFS} = \tau_n + f\sigma_n, \tag{2}$$

where σ_n stands for normal stress, τ_n denotes shear stress, f represents the coefficient of internal friction (taken as 0.6 in this paper), and n is the normal direction of the fault plane on which CFS reaches its maximum. When the increment of Coulomb Failure Stress (ΔCFS) is positive, it is in a loading state; otherwise, when ΔCFS is negative, it is in an unloading state.

As we know, stress in the crust σ_{ij} consists of tectonic stress σ_{ij}^T and the stress induced by the earth's tide σ_{ij}^t. Since the level of σ_{ij}^T (on the order of 10^6–10^8 Pa) far exceeds the level of σ_{ij}^t (10^3–10^4 Pa), directions of the principle stress in the crust and the direction of n can be determined from the tectonic stress only. However, the rate of change of the tidal stress is considerably larger than that of the tectonic stress (VIDALI *et al.*, 1998), thus ΔCFS is mainly due to stress induced by the tide, which can be calculated precisely, and could be resolved along the dominant fracture mechanism determined in each region.

2.2. Tectonic Stress Field in Southern California

An outline of the stress field in Southern California can be obtained from the world stress map (ZOBACK, 1992). The stress field is supplemented by the fault system in Southern California which is provided by SCECDC (Southern California Earthquake Data Center). With these two sets of information, we divided the Southern California region into eleven parts, in each of which the fault property stress field is almost uniform, as shown in Table 1.

With the fault property and basic stress field in Southern California, we can calculate ΔCFS along the dominant fracture mechanism with time, and determine the loading and unloading periods.

2.3. Calculation of LURR

As mentioned in the reference (ZHANG, 2004), Y is determined by the ratio of Benioff strain during the loading period over the unloading period.

Since the preparation and occurrence process of earthquakes is controlled not only by deterministic dynamical law but also affected by stochastic or disorder factors, ZHUANG and YIN (1999) studied the influence of random factors on LURR

Table 1

Fault property and basic stress field in Southern California

Subdivision	Fault Property	P-axis direction/(°)	T-axis direction/(°)
Sc01	right-lateral strike-slip	−15	75
Sc02	right-lateral strike-slip	−5	85
Sc03	reverse, right-lateral	45	0
Sc04	right-lateral strike-slip	−15	75
Sc05	normal, right-lateral	0	−20
Sc06	right-lateral strike-slip	−10	80
Sc07	normal, right-lateral	0	80
Sc08	right-lateral strike-slip	−5	85
Sc09	reverse right-lateral	45	0
Sc10	right-lateral strike-slip	0	90
Sc11	right-lateral strike-slip	10	100

in order to estimate the threshold Y value which can be regarded as an earthquake precursor within a specified confidence level. They gave the critical value of LURR Y_c that depends on the number of earthquakes under different specified confidence levels. For instance, at the confidence level of 90%, Y_c is equal to 3.18 if the number of earthquakes in the time and space window is 20, which means that Y should be equal to or greater than 3.18 for the medium to be considered in an unstable state when the number of earthquakes is 20. For the confidence level of 99%, Y_c is 7.69 if the number of earthquakes in the specific time and space window is 20. The greater the earthquake number is, the lower the Y_c (critical value of LURR) is.

In this paper, we give critical space-time regions of LURR by Y/Y_c instead of Y under a confidence level of 99%. When $Y/Y_c \geq 1.0$, the seismogenic system lies outside of the linear state.

3. Data and Scanning Parameters

The earthquake catalogue we use in this paper is from ANSS (Advanced National Seismic System).

In order to speed up the calculations and avoid disturbance from outstanding earthquakes, we chose magnitude thresholds according to the Gutenberg-Richter relation in each unit area. The threshold of minimum magnitude for each unit area is M 2.0, and the threshold of maximum magnitude for each unit area is M_{max} ($M_{max} = 3.5$ when the magnitude intercept of line $\lg N = a - bM$ is among M 4.0 and M 5.9, and $M_{max} = 4.0$ when the intercept of line $\lg N = a - bM$ is equal or larger than M 6.0).

A circle region with a radius of 100 km was selected as the spatial window within which a value of Y/Y_c (LURR/critical LURR) was calculated for a specific time

window (1 year), then the circle center was moved step by step in both latitude and longitude by increments of 0.25 degrees, and the contour of Y/Y_c in each month could be obtained.

4. The Feature of Y/Y_c Anomaly before the San Simeon Earthquake and Tendency of Earthquake Occurrence in Southern California

4.1. The Feature of Y/Y_c Anomaly before the San Simeon Earthquake

The results show that Y/Y_c anomalous region (with $Y/Y_c \geq 1.0$) occurred successively near the epicenter of the San Simeon earthquake from March 2002 to June 2002, and the maximum anomaly of Y/Y_c occurred in May 2002 (Fig. 1).

The maximum Y/Y_c and duration of Y/Y_c anomaly before the San Simeon earthquake are 1.4 and 15 months, respectively. The interval between the occurrence time of maximum Y/Y_c and the occurrence time of the earthquake is 19 months. Compared with the 5 earthquake cases out of the 6 with $M \geq 6.5$ during the period from 1980 to 2001, the maximum Y/Y_c and duration of Y/Y_c anomaly before this earthquake are among the normal ranges, however the interval between the occurrence time of maximum anomaly and the occurrence time of this earthquake is the longest one (Table 2).

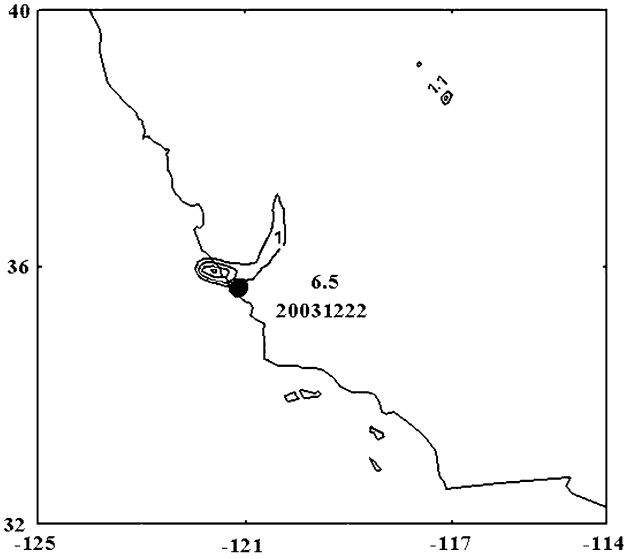


Figure 1
Contour of Y/Y_c in May, 2002. (The solid black circle represents the epicenter of the San Simeon M 6.5 earthquake on Dec. 22, 2003; Y/Y_c is among 1.0 to 1.4, and the increment of Y/Y_c is 0.1 between each two neighboring isolines.)

Table 2

Y/Y_c anomalies before seven strong earthquakes in Southern California during the period from 1980 through 2003

Date	Epicenter	Magnitude/ Δ(km)	Max Y/Y_c	Lasting time of anomalous Y/Y_c (month)	$\Delta t(m)$
1983.5.2	(36.23° N, 120.32° W) Coalinga	6.7/?	?	?	?
1987.11.24	(33.01° N, 115.85° W) Superstition Hills	6.6/0	1.4	21	11
1989.10.18	(37.04° N, 121.88° W) Loma Prieta	7.0/100	1.2	24	10
1992.6.28	(34.20° N, 116.44° W) Landers	7.3/100	1.0	18	8
1994.1.17	(34.21° N, 118.54° W) Northridge	6.6/200	1.2	25	13
1999.10.16	(34.59° N, 116.27° W) Hector Mine	7.1/100	1.4	15	1
2003.12.22	(35.7° N, 121.1° W) San Simeon	6.5/50	1.4	15	19

Note: Δ(km) is the distance between the earthquake epicenter and the maximum Y/Y_c point.
Δt is the duration between the occurrence time of maximum anomaly and the occurrence time of the earthquake.

4.2. Tendency of Earthquake Occurrence in Southern California

The results of this study also show that two areas with Y/Y_c anomalies occurred from Oct. 2002 and Dec. 2002, respectively. One is in the southern part of Southern California during Oct. 2002 to May. 2003, and the other is in the northern part of Northern California during Dec. 2002 to Aug. 2003. Figure 2 shows the contour of Y/Y_c in March 2003. This figure implies that two earthquakes with $M \geq 6.5$ might

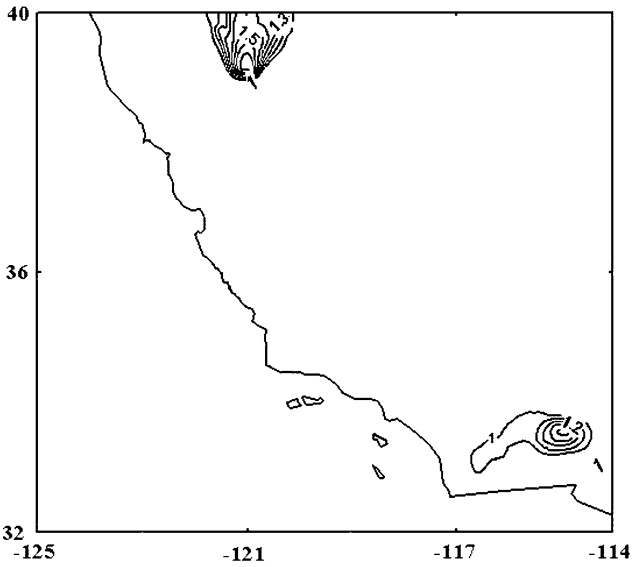


Figure 2
Contour of Y/Y_c in March, 2003. (The increment of Y/Y_c is 0.1 between each two neighboring isolines.)

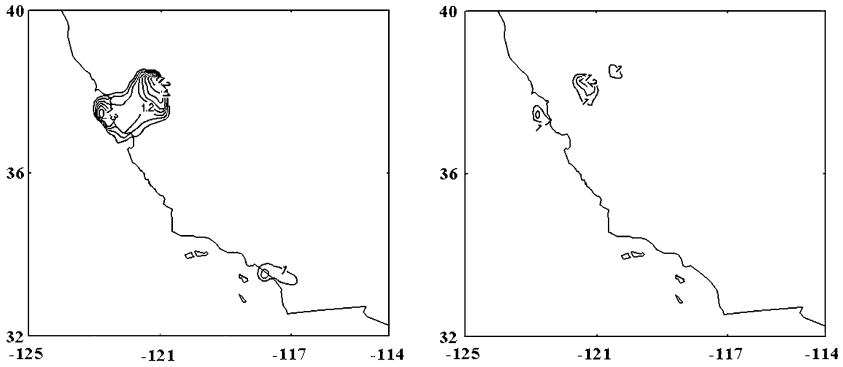


Figure 3
Contours of Y/Y_c in Feb. 2004 (left hand) and June 2004 (right hand). (The increment of Y/Y_c is 0.1 between each two neighboring isolines.)

occur in California in the future; one is in the circle region with the center of (33.7 N, 115 W) and the radius of 200 km (100 km with the possibility of 80%), and the other is in the circle region with the center of (39.1 N, 121.1 W) and the radius of 200 km (100 km with the ratio of 5/7). The former is more reliable than the latter because we obtain more detailed data about the stress field and fault system in Southern California than those in Northern California. According to statistical characteristics of the relationship between Y/Y_c anomalies and the coming earthquakes (Table 2), the most probable time for the coming earthquake is before Oct., 2004.

The above forecast was reported on the 4th ACES Workshop in Beijing in July, 2004. In fact, a new region which is involved with anomalous Y/Y_c occurred from Feb. 2004 (left contour in Fig. 3) and blossomed in June 2004 (right contour in Fig. 3). Since we had not seen the shrinkage of the area, we could not judge when there would be an earthquake near this region. The Parkfield M_w 6.0 Earthquake* on Sept. 28, 2004 is about 200 km from the point with maximum Y/Y_c value. According to Table 2, an earthquake with magnitude larger than 6.5 might occur near this region during the coming 19 months. Although the Parkfield M_w 6.0 earthquake occurred near this region, it is not the expected one because it is smaller than the lower threshold of the predicting events. In other words, there would be an earthquake with magnitude larger than 6.5 near this region during the next 19 months from June, 2004.

* Parkfield M_w6.0 Earthquake (11 km SSE of Parkfield, CA, 35.815 N, 120.374 W) occurred on Sept. 28, 2004, which is about 200 km from the northern Y/Y_c anomalous region.

5. Conclusions and Discussion

The original anomaly of LURR near the epicenter of the San Simeon earthquake was given beforehand (ZHANG *et al.*, 2004), and the feature of LURR anomaly before the San Simeon earthquake in this retrospective study coincides with those before the five strong earthquakes in the former study, which proves that LURR is a promising approach to intermediate-term earthquake prediction prior to strong earthquakes with M 6.5 or greater in Southern California.

The accuracy of earthquake prediction might be improved if detailed knowledge of the stress field in Southern California and higher quality of earthquake catalogues could be accessed. Meanwhile, suitable scanning parameters might improve the results.

Acknowledgements

Support is gratefully acknowledged by the Key Program of NSFC (Grant Nos. 10232050, 10572140) and MOST under grant's (2004CB418406, 2002CB412706). We also express our thanks to ANSS for earthquake data.

References

BULLEN, K.E. and BOLT, B.B., *An Introduction to the Theory of Seismology* (Cambridge University Press, Cambridge Press 1985).

HARRIS, R.A. (1998), *Introduction to special section: Stress triggers, stress shadows, and implication for seismic hazard*, J. Geophys. Res. *103*, 24,347–24,358.

JAEGER, J.C. and COOK, N.G.W., *Fundamentals of Rock Mechanics* (Chapman and Hall, London 1976) 78–79.

KANAMORI, H. and ANDERSON, D.L. (1975), *Theoretical basis of some empirical relation in seismology*, Bull. Seismol. Soc. Am. *65*, 1073–1096.

MORA, P., PLACE, D., WANG, Y.C., YIN, X.C., PENG, K.Y., and WEATHERLEY, D. (1999), *Earthquake Forecasting: Retrospective Studies in Australia – the Newcastle and Burra Earthquakes and Numerical Simulation of the Physical Process*, AEES (Australian Earthquake Engineering Society) Annual Meeting, November, 2000, Hobart, Australia.

MORA, P., PLACE, D., WANG, Y.C., and YIN, X.C. (2000), *Simulation of Load-Unload Response Ratio in the Lattice Solid Model*, AGU (American Geophysical Union) 2000 Fall Meeting, San Francisco, pp. 10–17, December 2000.

RESERNBERG, P.A. and SIMPSON, R.W. (1992), *Response of regional seismicity to the static stress change produced by the Loma Preita earthquake*, Science *255*, 1687–1690.

VIDALI, J.E., AGNEW, D.C., JOHNSTON, M.J.S., and OPPENHEIMER, D.H. (1998), *Absence of earthquake correlation with earth tides: An indication of high preseismic fault stress rate*, J. Geophys. Res. *103*, 24,567–24,572.

YIN, X.C. and YIN, C. (1991), *The precursor of instability for nonlinear systems and its application to earthquake prediction*, Science in China *34*, 977–986.

YIN, X.C. (1987), *A new approach to earthquake prediction*, Earthq. Res. China *3*, 1–7 (in Chinese with English abstract).

YIN, X.C., CHEN, X., SONG, Z.P., and YIN, C. (1995), *A new approach to earthquake prediction: the Load\Unload Response Ratio (LURR) theory*, Pure Appl. Geophys. *145*, 701–715.

YIN, X.C., WANG, Y.C., PENG, K.Y., and BAI, Y.L. (2000), *Development of a new approach to earthquake prediction: Load/Unload Response Ratio (LURR) theory*, Pure Appl. Geophys. *157*, 2365–2383.

WANG, Y.C., YIN, X.C., and PENG, K.Y. (1999), *Numerical simulation on Load/Unload Response Ratio theory*, Chinese J. Geophys. *42*(4), 527–536.

ZHANG, Y.X., YIN, X.C., and PENG, K.Y.(2004), *Spatial and temporal variation of LURR and its implication for the tendency of earthquake occurrence in Southern California*, Pure Appl. Geophys. *161*(11–12th), 2359–2367.

ZHANG, Y.X., YIN, X.C., and PENG, K.Y. (2003), *Study on Spatial and Temporal Variation of LURR in Southern California and Its Implication to Earthquake Tendency in this Region*, APEC Cooperation for Earthquake Simulation (ACES) 3rd ACES Workshop Proceedings, pp. 365–371 (Chief editor: Andrea Donnellan, co-editor: Peter Mora). (Published by: APEC Cooperation for Earthquake Simulation (ACES) , QUAKES, Department of Earth Science, the University of Queensland, Brisban).

ZHUANG, J.C. and YIN X.C. (1999), *Random distribution of the Load/Unload Response Ration (LURR) under assumptions of Poisson model,* Earthq. Res. China *15*, 128–138.

ZOBACK, M.L. (1992), *First- and second-order patterns of stress in the lithosphere: The world stress map project*, J. Geophys. Res. *97*, 11,703–11,728.

(Received December 1, 2004, revised September 10, 2005, accepted September 12, 2005)
Published Online First: December 20, 2006

Pure appl. geophys. 163 (2006) 2353–2362
0033–4553/06/122353–10
DOI 10.1007/s00024-006-0124-0

Pure and Applied Geophysics

A Statistical Investigation of the Earthquake Predictions Using LURR

KEYIN PENG,[1] XIANGCHU YIN,[1,2,3] and LANG-PING ZHANG[2]

Abstract—In terms of the spatial scanning of LURR (Load-Unload Response Ratio), we have been predicting the seismic tendency within the next year for the mainland of China from 1995 to 2003. In order to make the quantitative retrospective assessment of LURR method, we compare the results with Poisson null hypothesis. The results show that the prediction by LURR method is much better than Poisson hypothesis.

Key words: Load-Unload Response Ratio (LURR), earthquake prediction, statistical assessment.

1. Introduction

The LURR (Load-Unload Response Ratio) method is an intermediate-term earthquake prediction approach. It was first proposed by YIN (1987) about twenty years ago and since then has undergone many improvements (YIN *et al*., 1991, 1995, 1996, 2000, 2004). The basic idea of LURR is: when the seismogenic system is in a stable state, its response to loading is nearly the same as its response to unloading, whereas when the system is approaching the unstable state, the response to loading is more sensitive than that to unloading. When a seismogenic system is in a stable or linear state the LURR value equals unity, whereas when the system lies outside of the linear state the LURR value will be larger than unity. The tide provides the loading and unloading in the current implementation.

During the first ten years, the retrospective inspections were performed by YIN *et al*. (1995) with hundreds of earthquake cases to validate the LURR method. For more than 80% of cases the LURR value (denoted by Y) fluctuates around 1 during the early stage of the seismogenic process and rises along with the approaching occurrence of a strong earthquake. Then the Y values decrease after they reach their

[1] China Earthquake Networks Center, CEA (China Earthquake Administration), Beijing 100036, China. E-mail: pengkeyin@seis.ac.cn

[2] LNM (Laboratory for Non-linear Mechanics), Institute of Mechanics, CAS, Beijing 100080, China. E-mail: yin@LNM.imech.ac.cn; zhanglp@LNM.imech.ac.cn

[3] CSB, Institute of Geophysics, Beijing 100084. E-mail: xcyin@public.bta.net.cn

maximum (significantly larger than 1). In contrast, for seven "stable" regions with low seismicity (no earthquake with $M \geq 4$ occurred in the examined period) on the Chinese mainland, the Y value always fluctuates slightly around 1 during the entire duration without any significant deviations (YIN *et al.*, 1995). These results suggest that LURR is available to be a precursor of strong earthquakes.

The Lattice Solid Model (MORA *et al.*, 2002; WANG *et al.*, 1998–2000, 2004), FEM model and chains network model (LIANG *et al.*, 1996) were used to simulate the evolution of damage and failure of heterogeneous brittle solid and in the meantime to measure the variation of LURR prior to the macro-fracture similar to the strong natural earthquake. The simulating results show that LURR values become high along with the increasing load and reach their peak point, then drop rapidly prior to the main event.

There also have been numerous laboratory studies on LURR (SHI *et al.,* 1994; YIN *et al.*, 2004). XING-JUE SHI *et al.* (1994) conducted laboratory studies on variations of LURR during rock failure in the early 1990s. The result is similar to the theory expectation, that is, Y is close to 1 when the stress is small, increases when beyond the elastic phase, and grows remarkably just before fracture.

Although LURR theory is easily accepted theoretically and has been proved valid by laboratory modeling, numerical simulation and case inspections as well, it must be tested ultimately in real earthquake practices. YIN *et al.* (2000) applied LURR to real earthquake prediction practice in the period of 1993–1998. Among the 18 total predicted cases, 12 cases were correct, 6 cases false. The successful cases included the 1994 Northridge, USA earthquake and the 1995 Kanto, Japan earthquake (YIN *et al.,* 1996).

2. Earthquake Prediction in Terms of LURR

After retrospective study with hundreds of earthquake cases and some tentative practice of earthquake predictions in selected regions, we attempted to predict the seismic tendency of China from 1995. We used the catalogue provided by the network of China. For each quake we calculated its increment of Coulomb Failure Stress caused by tides, to judge loading or unloading, and then performed spatial scanning of LURR with the formula:

$$Y = \frac{\left(\sum_{i=1}^{N^+} E_i^m\right)_+}{\left(\sum_{i=1}^{N^-} E_i^m\right)_-}, \tag{1}$$

where E denotes seismic energy which can be calculated according to the Gutenberg-Richter formula (KANAMORI and ANDERSON, 1975), the sign "+" means loading and "−" unloading, $m = 1/2$ (thus E^m denotes the Benioff strain).

Since the preparation and occurrence process of earthquakes are controlled not only by deterministic dynamical law but also affected by stochastic or disorder factors, ZHUANG and YIN (1999) studied the influence of random factors on LURR in order to estimate the threshold above which Y value can be regarded as an earthquake precursor within a specified confidence level. They gave the critical value of LURR Y_c that depends on the number of earthquakes under different specified confidence levels. The greater the earthquake number is, the lower the Y_c (critical value of LURR). The earthquake number is usually different. Consequently we use the ratio Y/Y_c instead of Y where Y_c is the threshold at a confidence level of 95%.

In the end of every year we obtained a distribution map which shows the anomalous LURR regions in the Chinese mainland from 1995. According to the maps we predicted that earthquakes with magnitude $\geqq 5$ would occur in the anomalous LURR regions in the next year.

From 1995 to 2003 we have been predicting the seismic tendency within the next year for the mainland of China, according to the spatial scanning of LURR. The results were issued in the books pertaining to the research on the seismic tendency of China between 1996–2004 (YIN *et al.,* 1995–2003). Now it is time to evaluate the prediction effect of LURR.

3. Statistical Tests

How to evaluate the prediction effect is a controversial topic. In China the R score method has been the most popular one. An R score is approximately 0 for completely random guesses, and approximately 1 for completely successful predictions. SHI *et al.* (2001) evaluated the annual earthquake predictions of the China Seismological Bureau (CSB) by means of an R score. They concluded ''the average R score of the annual predictions in China in the period 1990–1998 is about 0.184, and concluded that earthquake prediction in China is still in a very preliminary stage, barely above a pure chance level'' and YAOLIN *et al.* (2001) believed also that it is fact to use a single index to cover all important aspects in evaluating earthquake predictions. Previous tests of earthquake forecasts in the USA have emphasized the likelihood test (KAGAN and JACKSON, 2000; RUNDLE *et al.*, 2002; HOLLIDAY *et al.*, 2005). ''These tests have the significant disadvantage that they are overly sensitive to the least probable events'' (HOLLIDAY *et al.*, 2005). In the present paper we adopt a method suggested originally by D. Jackson.

We used the catalogues published by the China Seismograph Network (CSN) in our test. Using a strict interpretation, only earthquakes with magnitude $M \geq 5.0$ would be counted.

The Y value contour maps were subdivided into spaces by $0.5^\circ \times 0.5^\circ$. For each anomaly LURR region, called 'dangerous' region, we count the number of earthquakes of $M \geq 5.0$ which occurred in the next year and acquire Poisson

earthquake probability directly from the data file of probability density of earthquakes for $M \geq 5.0$, provided by Kagan who used Poisson null hypothesis which calculated the probability density of earthquakes worldwide based on Harvard CMT data catalogues. The probabilities mean the rates of earthquakes occurred intermediately in the predicted regions every year.

According to the data of CSN more than 20 earthquakes with magnitude $M \geq 5$ occurred generally on the chinese mainland every year. A part of such events occurred in regions where the scarce data are unavailable for LURR calculation. Therefore the earthquakes which occurred in the data scarcity regions were ignored in this paper.

Tabulate the number of earthquakes of $M \geq 5.0$ which occurred in the next year and the accumulative Poisson earthquake probability for each prediction region (Table 1 is an example in 1997), and plot the accumulated earthquake number against cumulative Poisson earthquake probability.

Figure 1 is the test results of the forecasts according to LURR spatial scanning from 1995 to 2003. The figure in the box in Figure 1 is the year when the prediction was made for the next year. For example, "2003" means the prediction for seismic tendency in 2004 made at the end of 2003. If the actual earthquake strictly obeys Poisson hypothesis it is a bisector line in the figure (see the red lines in Fig. 1). If the predictions are better than the Poisson hypothesis, the accumulated earthquake

Table 1

Statistics of EQ prediction using LURR and Poisson hypothesis in the example of Year 1997

Region No.	Probability	Σ Probability	Earthquake	Σ Earthquake
1	0.8649	0.8649	1	1
2	0.0454	0.9103	0	1
3	0.5351	1.4454	1	2
4	0.2087	1.6541	1	3
5	0.6365	2.2906	0	3
6	0.0552	2.3458	0	3
7	0.1281	2.4739	1	4
8	1.0618	3.0618	2	6
9	0.0884	3.6241	0	6
10	0.0288	3.6529	0	6
11	0.036	3.6889	0	6
12	0.2938	3.9827	3	9
13	0.5278	4.5105	0	9
14	0.1278	4.6383	1	10
15	0.0096	4.6479	0	10
16	0.1387	4.7866	0	10
17	0.0487	4.8353	0	10
18	0.2238	5.0591	1	11
19	0.0094	5.0685	0	11
20	0.4657	5.5342	1	12
21	0.6621	6.1963	1	13

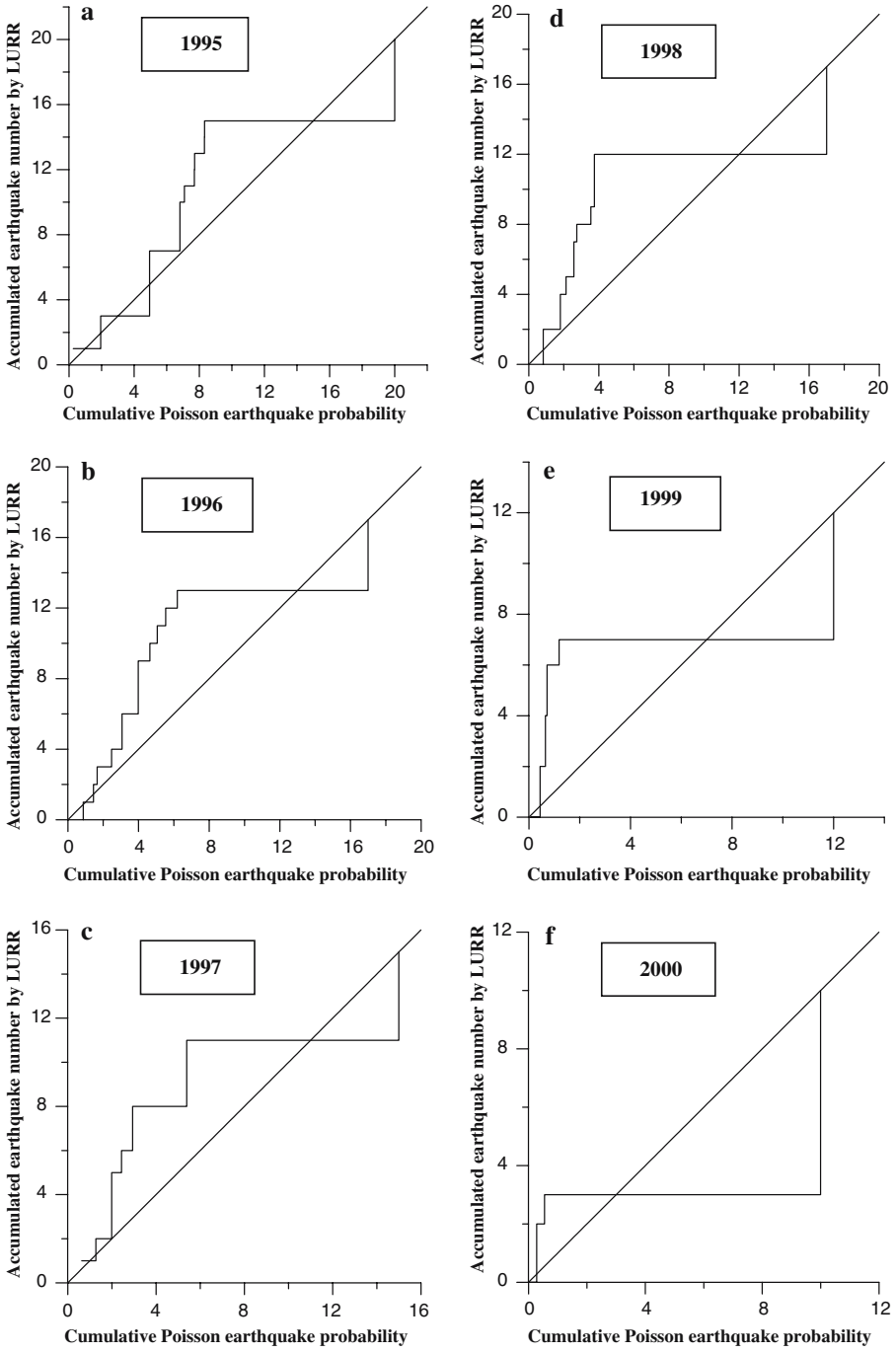


Figure 1
Accumulated earthquake number predicted by LURR vs. cumulative Poisson earthquake probability from 1995 to 2003 on the Chinese mainland.

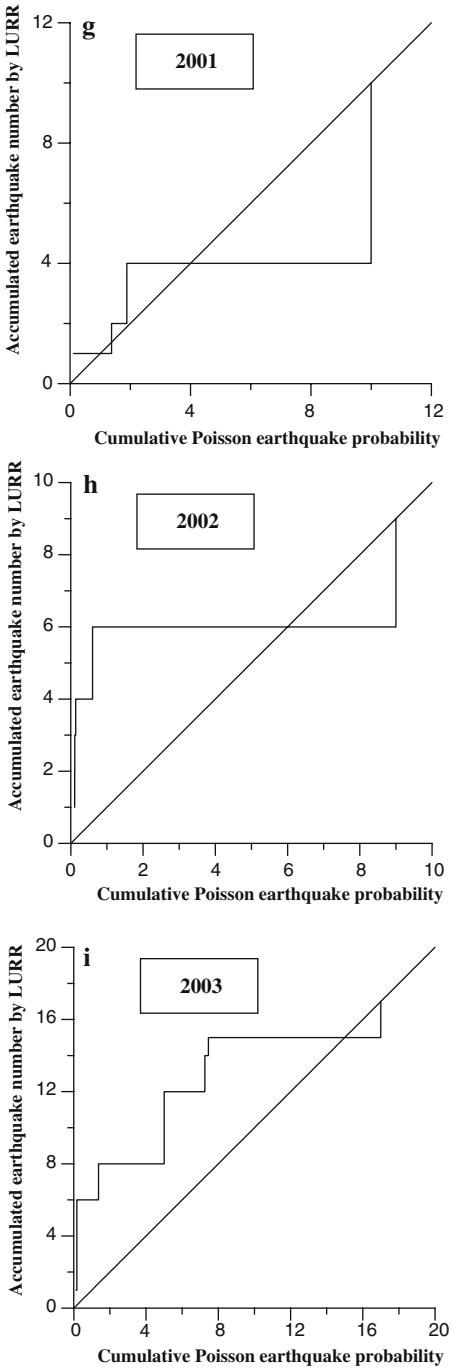


Figure 1
(contd.)

numbers will be greater than the cumulative Poisson earthquake probability and the curves will be above the bisector lines.

It is true that some regions predicted by LURR are worse than Poisson hypothesis (see the 2nd, 5th, 10th and 19th regions in Table 1), however those predicted successfully (see the 1st, 4th, 7th and 14th regions in Table 1) outperform Poisson hypothesis to such a degree that the overall effect still shows the better performance of the LURR method (see the left part to the diagonal of the plots). Regarding the missing cases (strong EQ occurred without prediction, corresponding to the vertical line right to the diagonal of the plots), a majority of results shows that the number of predicted cases is obviously greater than the number of missing cases except in 2000 and 2001.

4. Critical Region

Usually the scan radius we used is 100 to 150 km. However, before the Kunlun Earthquake (M 8.1, Nov. 14, 2001) we did not find an anomaly in the LURR contour map near the epicenter. Further study shows that anomalous high value LURR appears prior to the Kunlun main shock while the radii are from 400 to 1100 km.

When we calculate the values of Y for a region and specified time window, the Y values are a function of the size of the region. The size which maximizes the peak value of LURR is called critical region size. In Figure 2 the curve is plotted for LURR against the radius R of the region for Kunlun earthquake. The critical region size for Kunlun earthquake is $R = 800$ km.

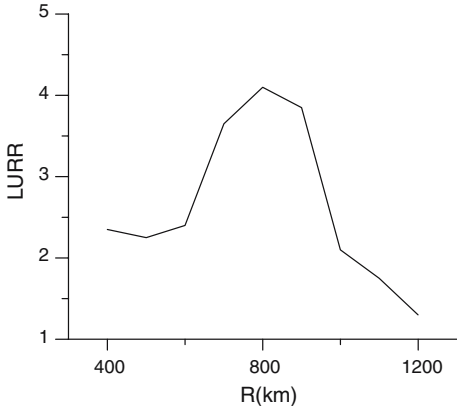


Figure 2
The peak values of LURR against the radius R of the region for Kunlun earthquake.

Fourteen earthquakes that occurred in China have been studied with magnitudes ranging from 5.1 to 8.1. It is found that the critical region size scales with the magnitude of earthquake (Fig. 3).

Fitting the data in Figure 3, the function between critical region size and the magnitude is

$$\log R(\mathrm{km}) = 0.087 + 0.34\,\mathrm{M}. \tag{2}$$

JAUME and SYKES (1999) have found that the critical regional size for accelerating energy release (AER/AMR) is

$$\log R \propto 0.36\,\mathrm{M} \tag{3}$$

and BOWMAN *et al.* (1998) found

$$\log R \propto 0.44\,\mathrm{M}. \tag{4}$$

Obviously they agree with each other quite well.

In our previous work (YIN *et al.*, 2002), we have discovered that a strong correlation is evident between the AMR/AER and LURR critical regional sizes that suggest these two observations maybe have a common physical mechanism. Furthermore, based on the results presented in Figure 3, the critical region size-magnitude scaling relation for LURR provides a means to estimate the magnitude of an upcoming earthquake.

5. Discussion and Conclusion

The statistical results show that the prediction by LURR method is considerably better than the Poisson hypothesis. Considering the complexity of the earthquake,

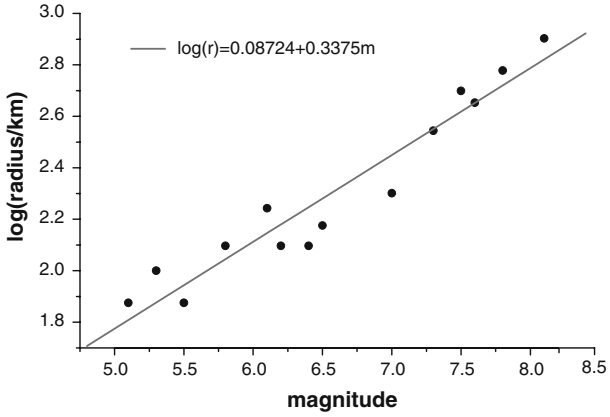


Figure 3
Critical region radius r as a function of the magnitude of 14 earthquakes which occurred in China.

some factors, such as critical size, should be considered. As discussed above, to predict earthquakes with different magnitude, we should select different radii.

It is exhilarating that the prediction results using LURR have been increasingly better. Especially in 2003 (see the last plot in Fig. 1), for the total of 16 earthquakes with magnitude M 5 or larger which occurred in regions in which the observed data are good enough to calculate LURR in 2004, the number of predicted cases is 15, whereas the number of missing cases is only 1.

Acknowledgements

This research is supported by the Natural Science Foundation of China (Grant Nos. 10232050, 10572140, and 2005CB724800), the key project of the ninth five-year plan of MOST (2004CB418406, 2002CB412706) and CEA (10604). We are grateful to Professors Davi Jackson, Zhengkang Shen for their highly-valued suggestions and Professor Y. Kagan for his Poisson earthquake probability data files for $M \geq 5.0$ based on PDE (1968–1995) data catalogues.

References

Bowman, D.D., Ouillon, G., Sammis, C.G., Sornette, A., and Sornette, D. (1998), *An observational test of the critical earthquake concept*, J. Geophys. Res. *103,* 359–372.

Holliday, J.R., Rundle, J. B., Tiampo, K. F., Klein, W., and Donnellan, A. (2005), *Systematic procedural and sensitivity analysis of the pattern informatics method for forecasting large (*$M \geq 5$*) earthquake events in southern California*, Pure Appl. Geophys., in press (2005).

Jaume, S.C. and Sykes, L.R. (1999), *Evolving towards a critical point: A review of accelerating seismic moment/energy release prior to large and great earthquakes*, Pure Appl. Geophys. *155,* 279–306.

Kagan, Y. Y. and Jackson, D. D. (1995), *New seismic gap hypothesis: Five year's after*, J. Geophys. Res. *100(B3)*, 3943–3959.

Kagan, Y. Y. and Jackson, D. D. (2000), *Probabilistic forecasting of earthquakes*, Geophys. J. Int. *143*, 438–453.

Kanamori, H. and Anderson, D.L. (1975), *Theoretical basis of some empirical relation in seismology*, Bull. Seismol. Soc. Am. *65*, 1073–1096.

Liang, N., Liu, Q., Li, J., and Song, H. (1996), *A chains network model simulating meso-mechanics behavior and micro- damage evolution of in situ reinforced ceramics*, *Advanced in Engineering Plasticity and Its Applications* (eds. Abe, T. and Tsuta, T.) (Pergamon, Amsterdan, 1996) pp. 141–146.

Mora, P., Wang, Y., Yin, C., Place, D., and Yin, X.C. (2002), *Simulation of load-unload response ratio and critical sensitivity in the lattice solid model*, Pure Appl. Geophys. *159*, 2525–2536.

Rundle, J. B., Tiampo, K. F., Klein, W., and Martins, J. S. S., *Sel-forganization in leaky threshold systems: The influence of near mean field dynamics and its implications for earthquakes, neurobiology, and forecasting*, Proc. Natl. Acad. Sci. USA *99*, 2514–2521, Suppl. 1 (2002).

Shi, X., Xu, H., Wan, Y., Lu, Z., and Chen, X. (1994), *The rock fracture under simulated tide force—Laboratory study on the loading and unloading response ratio (LURR) theory*, Acta Geophysica Sinica *37*, 631–636.

Shi, Y. L., Liu, J., and Zhang, G. M. (2001), *An evaluation of Chinese annual earthquake predictions, 1990–1998*, J. Appl. Prob. *38a*, 222–231.

WANG, Y.C., YIN, X.C., and WANG, H.T. (1998), *The simulation of rock experiment on load/unload response for earthquake prediction*, Earthq. Res. China *14(2)*, 126–130.

WANG, Y.C., YIN, X.C., and WANG, H.T. (1999), *Numerical simulation on load/unload response ratio (LURR) theory*, Chinese J. Geophys. *42(4)*, 527–536.

WANG, Y.C., YIN, X.C., KE, F.J., XIA, M.F., and PENG, K.Y. (2000), *Simulation of rock failure and earthquake process on mesoscopic scale*, Pure Appl. Geophys. *157*, 1905–1928.

WANG Y.C., MORA, P., YIN, C., and PLACE, D. (2004), *Statistical tests of load-unload response ratio signals by lattice solid model: Implication to tidal triggering and earthquake prediction*, Pure Appl. Geophys. *12*, 1829–1839.

YIN, X.C. (1987), *A new approach to earthquake prediction*, Earthq. Res. China *3*, 1–7 (in Chinese with English abstract).

YIN X.C. and YIN, C. (1991) *The precursor of instability for nonlinear systems and its application to earthquake prediction*, Science in China, *34*, 977–986.

YIN, X.C., CHEN, X.Z., SONG, Z.P., and YIN, C. (1995), *A new approach to earthquake prediction: The load/unload response ratio (LURR) theory*, Pure Appl. Geophys. *145*, 701–715.

YIN, X.C., SONG, Z.P., and WANG, Y.C. (1996), *Temporal variation of LURR in kanto and other regions in Japan and its application to earthquake prediction*, Earthq. Res. China *10*, 381–385.

YIN, X.C., WANG, Y.C., PANG, K.Y., BAI, Y.L., WANG, H.T., and YIN, X.F. (2000), *Development of a new approach to earthquake prediction—load/unload response ratio (LURR) theory*, Pure Appl. Geophys. *157*, 2365–2383.

YIN, X.C., MORA, P., PENG, K.Y., WANG, Y.C., and WEATHERLY, D. (2002), *Load-unload response ratio and accelerating moment/energy release, critical region scaling and earthquake prediction*, Pure Appl. Geophys. *159*, 2511–2524.

YIN, X.C., ZHANG, H.H., and YU, H.Z. (2004), *Acoustic emission experiment of rock failure under cycling load*, The 4th ACES Workshop, July 9–14, Beijing China (2004).

YIN, X.C., ZHANG, H.H., YU, H.Z., ZHANG, Y.X., and PENG, K.Y. (2003), *Prediction of seismic tendency of Chinese mainland in 2004 in terms of LURR*, in book *The Research on Seismic Tendency of China in 2004,* (ed. by Center for Analysis and Prediction), CSB, pp. 282–285, (Seismological Press, Beijing, 2003) (in Chinese).

ZHUANG, J.C. and YIN, X.C. (1999), *Random distribution of the load/unload response ratio (LURR) under assumptions of poisson model*, Earthq. Res. China *15*, 128–138.

(Received December 20, 2004, reveised September 12, 2005, accepted September 14, 2005)

Pure appl. geophys. 163 (2006) 2363–2373
0033–4553/06/122363–11
DOI 10.1007/s00024-006-0133-z

Pure and Applied Geophysics

Stress Reorientation and LURR: Implication for Earthquake Prediction Using LURR

CAN YIN[1] and PETER MORA[1,2]

Abstract—After reviewing the problems associated with the current implementation of the LURR theory, we suggest that taking account of stress field complexity and stress reorientation may resolve these problems. By introducing the concept of Maximum Faulting Orientation (MFO), we propose a new approach for calculating LURR. Results presented for the case of the Northridge earthquake provide encouragement for the stress-reorientation explanation and the new approach.

Key words: Tectonic stress, reorientation, LURR, earthquake prediction, tidal triggering.

Introduction

The LURR (Load-Unload Response Ratio) theory of earthquake prediction states that just prior to a major event, the response of the Earth's crust under loading compared to its response to unloading will change remarkably (YIN and YIN, 1991; YIN *et al.*, 1995, 2000). The physical basis of LURR theory is thought to be related to critical sensitivity (XIA *et al.*, 2002; ZHANG *et al.*, 2004), which can be easily understood in the light of a famous proverb: *the last straw broke the camel's back*. Here the tidal stress acts as the last straw to break the crust. In the current LURR implementation, the loading/unloading condition is defined as the contribution of tidal stress to tectonic Coulomb Failure Stress (CFS) on a predetermined fault plane, while the seismic energy release is taken as the response. This implementation assumes that more seismic energy is released in the loading period than in the unloading period when the crust is close to instability, thus implying that statistically positive tidal CFS would trigger earthquakes when the crust is near instability.

[1] QUAKES, Earth Systems Science Computational Centre, The University of Queensland, St. Lucia, Brisbane, QLD 4072, Australia. E-mail: yincan@yahoo.com

[2] Australian Computational Earth Systems Simulator (ACcESS), MNRF, The University of Queensland, St. Lucia, Brisbane, QLD 4072, Australia.

A triggering mechanism exists in many situations in nature. A more relevant example is that quakes on the Moon are triggered by the tide due to the Earth's gravitation (MELCHIOR, 1978; NASA website).

In rock experiments (YIN *et al.*, 2004) and numerical simulations (WANG *et al.*, 2003), similar phenomena have been observed: When a small sinusoidal perturbation is superimposed on an otherwise linear loading, most events occur when the perturbation is increasing the loading. However, the relationship between earthquakes and tides is more difficult to determine. By examining numerous major earthquakes of known fault planes, we found that at the instant when main shocks occur, the chance that the tidal CFS is positive (promoting failure) is almost equal to the chance that it is negative (inhibiting failure). This fact means the timing of main shock is independent of tidal stress but has no other direct implication to the LURR results because the overwhelmingly majority of earthquakes used in LURR calculation are smaller events. The reason why no strong correlation has been found between major earthquakes and tides may be because a major event involves more processes or more complex interactions than smaller events, and the tidal stress is not the only process contributing to breaking the crust and may be offset by other factors. At the time a main shock occurs, the crust may have become critically sensitive such that the earth is primed for the large event, consequently events of all sizes start to occur at any time independently of the tidal loading effect, with stress redistribution from these events acting as the trigger for one another.

Problems Perplexing the LURR Theory

When LURR was first introduced as an approach for earthquake prediction (YIN *et al.*, 1995), it was expected to be a robust and quantitative indicator of instability, both retrospectively and prospectively. But this objective is yet to be realized in the current implementation of the LURR theory.

To analyze this issue, two well-known earthquakes which occurred in California, USA in the 1990s are taken as examples. Shown in Figure 1 are LURR vs. time curves for (a) the 1992 Landers earthquake and (b) the 1994 Northridge earthquake. In both cases, tidal CFS is calculated on the main-shock fault plane. An outstanding peak appears two years prior to the Northridge main shock, while the peak prior to the Landers main shock is not impressive.

The sharp contrast between the two patterns raises a number of questions:

First, why LURR anomaly is obvious in Northridge but not in the temporally and spatially adjacent Landers earthquake?

Secondly, even though the random effect is taken into account and LURR will not exceed unity except when there is statistical significance (see ZHUANG and YIN 2000), one may ask why do the LURR values still deviate far away from the unity

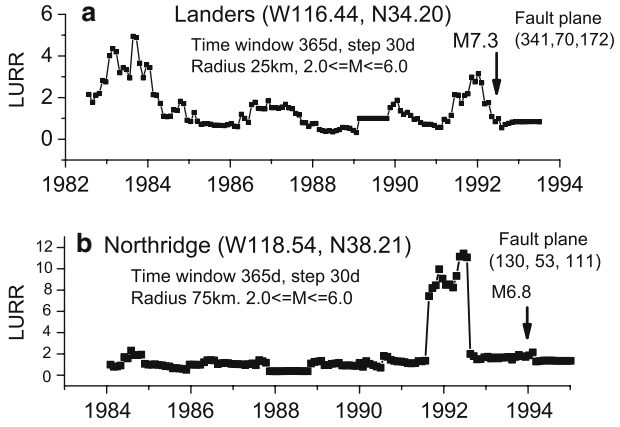


Figure 1
LURR as a function of time for (a) the 1992 Landers earthquake; and (b) 1994 Northridge earthquake (b). In both cases, time window is 1 year, time step 30 days and earthquake magnitude ranges from 2.0 to 6.0. The only difference is the radius is 75 km for Northridge and 25 km for Landers. An obvious peak appears 2 years before the Northridge main shock but not in Landers.

when the crust is far from a major earthquake? Could time variable stress field complexity induced by earthquake events be the reason?

Thirdly, if critical sensitivity is the underlying mechanism of the LURR phenomenon, the time period immediately prior to the main shock must be more critical than any other time. If LURR value represents the degree of criticality, one would expect to observe the highest peak at the instant immediately before the main shock. Then why does the highest peak appear two years before in Northridge instead? Could it be that after the crust becomes critically sensitive, a nucleation process of stress release and stress redistribution starts to occur and takes over as the dominant triggering mechanism in the leadup to the large event?

Anatomy of the Problems

Since tidal CFS is directly related to these questions, we try to solve the problem by analyzing CFS in LURR.

If you want to break a camel's back with a straw, you have to put the straw on its back rather than under its feet. In the case of an earthquake, this means tidal CFS must be superimposed upon coseismic tectonic CFS to break the crust. At a large scale, the regional (first-order) tectonic stress is quite uniform and may remain so for decades, even centuries. This is manifested by the world stress map (ZOBACK, 1992), which would be useless if regional stress changed frequently. However, an earthquake is generated directly by local (second-order) *in situ* stress

which might be variable. For example, it has been found that sometimes the tectonic stress rotates before and after major earthquakes (ZHAO and KANAMORI, 1998; HARDEBECK and HAUKSSEN 1999). The local stress regime could rotate from the first-order stress by up to 90°, and thus totally change the faulting pattern. We usually do not have the comprehensive knowledge of the development of *in situ* tectonic stress at local scale. This is one of the difficulties in geophysics.

On the other hand, in the rock experiments or numerical simulations mentioned above, the perturbation is always in alignment with the main loading, therefore it is always known when the perturbation is increasing the loading and when it is decreasing. In the real crust, however, we do not have an unequivocal knowledge of the main loading—tectonic stress—at any time, particularly at local scale which controls the initiation and development of faulting. Because an earthquake can occur on a fault of any orientation, it becomes very difficult if not impossible to determine whether a given fault is being loaded or unloaded in terms of Coulomb Failure Criterion which is the prerequisite of LURR calculation.

The dilemma is illustrated in Figure 2 which shows the focal mechanisms of all the larger earthquakes ($M \geq 4.0$) which occurred in Southern California between 01/01/1980 and 31/12/1999 (according to the Harvard CMT Project). Obviously these earthquakes occurred on various orientations of faulting, even though the tectonic movement direction of San Andreas Fault system is quite constant at large scale. Although the focal mechanisms of smaller events, the main body of LURR calculation, are hard to obtain, it is almost certain that they must have occurred on various orientations either. Therefore, the implementation of LURR by defining

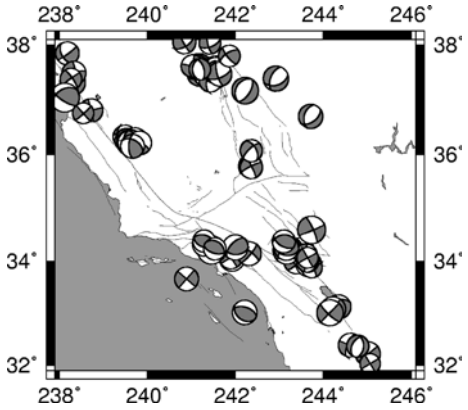


Figure 2
The focal mechanisms of larger earthquakes in Southern California between 01/01/1980 and 31/12/1999, according to Harvard CMT project.

CFS for all events preceding the main shock on the main shock fault plane might be wrong.

Because the main-shock fault plane is unknown for prediction purpose, researchers in China have tried to use an optimal fault plane determined by knowledge of regional tectonics in place of the main shock fault plane (YIN *et al.*, 1995). This method has generated more satisfactory results in a retrospective examination of LURR and brought successful real-time forecasts. However, because the choice of the optimal plane is relatively subjective and partly depends on experience, this method is arguable. Is there any better way to find the CFS-defining plane? What is the justification of doing so?

A Possible Solution

To solve the dilemma, the critical point hypothesis (CPH) is considered. According to CPH theory (SORNETTE and SORNETTE, 1990; SORNETTE and SAMMIS, 1995; BOWMAN *et al.*, 1998; JAUMÉ and SYKES, 1999), a near-critical system is characterized by both susceptibility to external perturbation and strong correlation between its parts. In the case of tidal triggering, tidal stress has the potential to facilitate the propagation of small cracks which may then coalesce and cascade into a large earthquake. However, not any perturbation will push the system towards failure, only those in accordance with the driving tectonic loading do so. When the system is far from criticality, the local tectonic stress around these microcracks evolves in random orientations, thus the possibility of tidal stress facilitating the propagation of the majority of them would be very low. However when the crust is approaching a large failure, the number of fracture orientations is decreasing, which may be manifested by the decreasing b-value that represents the fractal dimension of active faults before the large earthquake. Lower dimension not only means that cracks are coalescing but also indicates that the occurrence of fractures is becoming concentrated in fewer orientations before the major event, or more precisely, only microcracks of certain orientations become activated. The scenario can be considered a special long-range correlation. The triggering mechanism will more likely become obvious at this stage.

Because of the existence of tidal stress, the failure of the crust would be either facilitated or inhibited. Suppose failure of the crust obeys the Coulomb Failure Criterion, i.e., the failure only occurs on the plane where tectonic CFS reaches its maximum. Facilitation by the triggering mechanism will also be maximized on this plane. Let us call it the maximum faulting orientation (MFO). LURR should be maximum when tidal CFS is defined on MFO. In such a particular orientation, the tidal triggering effect should be more prominent than in others. By varying the faulting orientation among ALL possible orientations, i.e. combinations of strike, dip and rake and calculating the corresponding LURR, a particular orientation by

which LURR reaches its maximum will be identified. This orientation should be identical with MFO. After a major earthquake, the tectonic stress at local scale will normally change dramatically. Consequently the MFO will follow the change and a new MFO may emerge during another earthquake cycle — a process of reorientation. Therefore LURR values defined by the old MFO may not show peaks after the major event but undulate around unity. In a certain way, LURR defined by a particular focal mechanism represents the polarization degree of vulnerable small faults in this particular orientation at that moment, or the maturity of long-range correlation. Accordingly, reorientation of the driving tectonic stress due to smaller events will cause the undulation of LURR values.

MFO may or may not agree with the main shock fault plane because of the complexity of initiation process of faulting. However, MFO must not contradict the regional tectonic stress because it is controlled by tectonic stress. In return, the agreement between MFO and regional tectonic stress, once proven, will be a strong support for the above mechanism.

Before the MFO at local scales are examined, it is helpful to understand the general nature of MFO at a larger scale. As an example, MFO of Southern California within a $8° \times 8°$ area delimited by longitude W122° to W114°, latitude N32° to N40° was calculated using all the $M \geq 2.0$ earthquakes from 1970.01.01 to 2000.01.01. The MFO obtained in this way should represent the long term consequence of tidal stress exerting on this large geological structure.

MFO calculated from this dataset is (130°, 50°, 150°), representing an oblique right-lateral strike slip faulting with a little thrust. Though different from the focal mechanisms of most major earthquakes since 1980, it agrees with the general movement of San Andreas Fault system and the tectonic stress regime according to the stress map.

Nonetheless, regional MFO cannot replace local MFO because MFO is after all calculated using earthquakes which are controlled by tectonic stress at local scale. The tectonic stress field at the local scale is very complicated, varying from time to time and place to place. Therefore MFO at local scale should also vary from time to time (Figure 3) and place to place Figure 4 as illustrated below.

In the case of Northridge (Fig. 3), several years prior to the main shock, MFO changes from other orientations to those very similar to the main shock fault plane and changes again after the main shock. The calculation used a one-year time window and 100 km of radius. The size of the beach balls (focal mechanism) is proportional to the LURR value associated with such a mechanism.

The spatial variation of local MFO for the one-year time span prior to the Northridge main shock is shown in Figure 4. The dots are locations of earthquakes with $M \geq 3$ that occurred within the selected time window. The size of the dot is scaled with its magnitude. The future epicenter is almost on the highest anomaly.

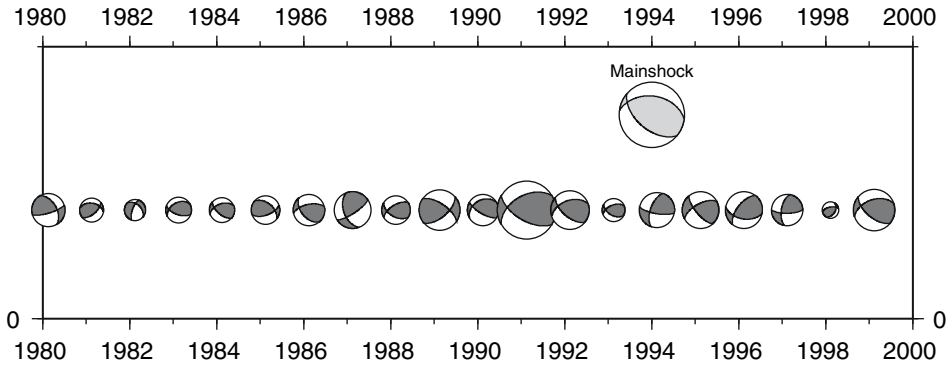


Figure 3
MFO as a function of time for Northridge in the past ten years before the main shock. As a reference, the actual focal mechanism of the main shock is shown in light gray. The size of the beach ball (focal mechanism) is proportional to the LURR value associated with such a mechanism. The time window is one year (backdated). Window radius = 100 km. Event magnitude ranges from M 2.0 to M 6.0.

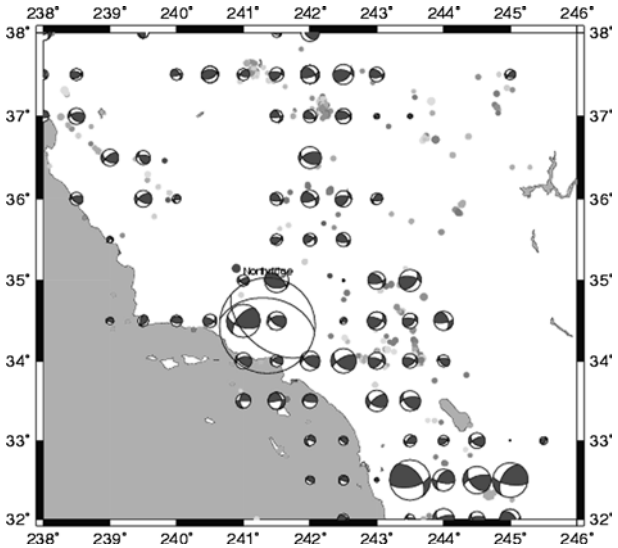


Figure 4
Spatial distribution of MFO at local scale in Southern California for the year before January 16, 1994 when the Northridge earthquake occurred. (The actual focal mechanism of the main shock is shown in transparency). The size of the beach ball (focal mechanism) is proportional to the LURR value associated with such a mechanism. The time window is one year (backdated). Window radius = 100 km. Event magnitude ranges from M 2.0 to M 6.0.

Because the MFO of entire Southern California represents the long-term effect of tidal stress over this whole region, we would expect the tidal CFS defined on this MFO to be more correlated with earthquakes than on other focal mechanisms. The

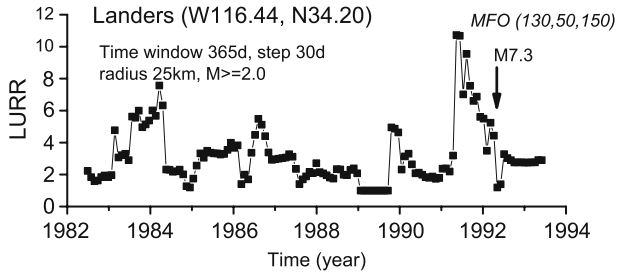


Figure 5
LURR vs. time in Landers using regional MFO for whole Southern California. The time window is 1 year, step 30 days, radius 25 km, $2.0 \leq M \leq 6.0$. An obvious peak appears immediately before the main shock.

variation of LURR with time for Landers earthquake using MFO is shown in Figure 5. Compared with the pattern of LURR defined on actual main shock fault planes (Fig. 1), the LURR pattern in Figure 5 shows more prominent anomaly prior to the main shock. If it is proven that regional MFO can effectively apply to calculating LURR at local scales, MFO will be an ideal approach to prospective application of LURR. It provides the potential of practically forecasting earthquakes because no knowledge about the main shock is needed any more.

The calculation of MFO is conducted as follows:

The maximization is achieved by varying strike, dip and rake independently among all possible combinations. If not specifically indicated, strike will vary in the range from 0 to 360°, dip from 20° to 90° and rake from −180° to 180°, with a 10° step (smaller steps are time-consuming and not necessary). Some physical restrictions, such as no normal fault should dip less than 45°, are set to exclude unrealistic combinations. After such a preprocessing, there are still more than 5000 possible combinations remaining in total. With such a large pool of possibilities, the calculated MFO is considered to be statistically reliable.

At last, to explain the possible procedure of earthquake forecast using LURR, we spatially scan the region from longitude W114° to W122 ° and from latitude N32° to N38° for the one-year time span prior to Landers earthquake. The result is shown in Figure 6. The time window is one year back dated from the time the main shock occurred and radius is 100 km. Magnitude $6.0 \geq M \geq 2.0$. The future epicenter is close to the highest peak. By advancing the time window forward step by step, we should be able to make forecast in real time.

Conclusions

Based on the above results we suggest that: If the plane on which LURR is defined happens to be close to the MFO, LURR curve will have a high peak immediately prior to the main shock. Otherwise, the LURR anomaly may not

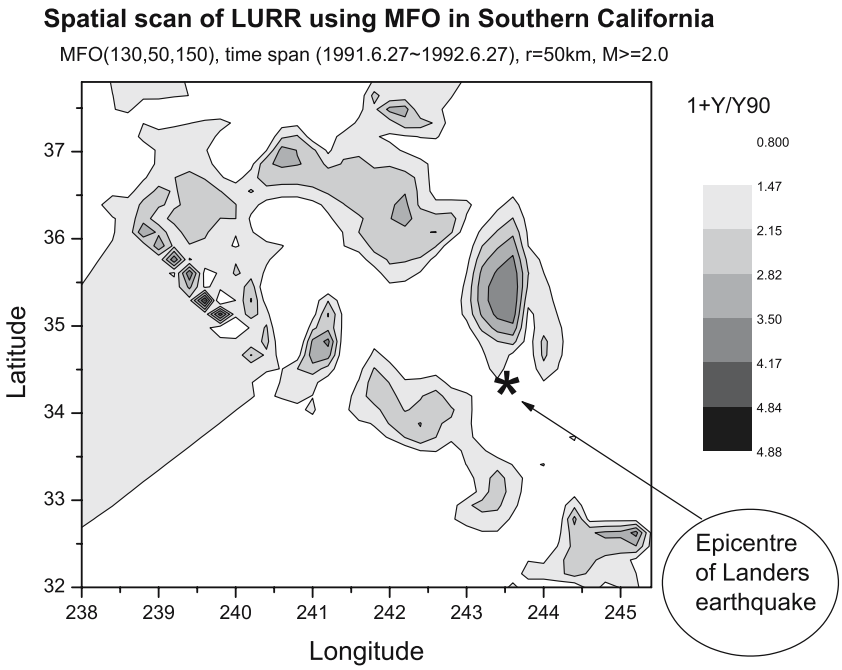


Figure 6
Spatial scanning of LURR defined on the regional MFO for one year prior to the main shock of the Landers earthquake. LURR value is set as $1 + Y/Y_{90}$ when $Y > Y_{90}$ and unity otherwise. Time window is 1 year (backdated) and radius 50 km. The future epicentre is close to but not on the highest LURR anomalies.

appear, even if the plane is the main shock fault plane. Since tectonic stress may change from time to time both regionally and locally, LURR values calculated on a fixed plane will inevitably undulate. The LURR peak timing is dependent on the timing of MFO change, which may follow the rotation of the tectonic stress. Considering that the way in which stress rotation is related to the upcoming major event is not yet clear, one should remain cautious regarding the predictive ability of the LURR method.

Whether the LURR method is able to predict earthquakes or not is subject to a clear demonstration of the existence of systematic large scale reorientation of the stress field due to earthquakes on smaller faults prior to the main shock. If the evidence is found in more cases in various geological settings, the predictive ability of LURR theory will be justified. Even though it is yet to be proven, its application to identifying the tectonic stress orientation in a very convenient way, and even forecasting the focal mechanism of future earthquakes, seems promising.

Acknowledgments

Funding support is gratefully acknowledged by Earth Systems Science Computational Centre and The University of Queensland. Computations were made using the ACcESS MNRF supercomputer – a 208 processor 1.1 TFlops SGI Altix 3700 – which was funded by the Queensland State Government Smart State Research Facility Fund and SGI.

The earthquake catalogue used in this paper came from ANSS (Advanced National Seismic System, previously known as the Council of National Seismic System) earthquake catalogue at Berkeley (website http://www.anss.org). Ocean tidal models used in tidal CFS calculation courtesy of D. C. Agnew (AGNEW, 1996).

REFERENCES

AGNEW, D.C. (1996), SPOTL: *Some programs for ocean-tide loading*, SIO Ref. Ser. 96–8, 35 pp., Scripps Inst. of Oceanogr., La Jolla, Calif., 1996.

BOWMAN, D.D., OUILLON, G., SAMMIS, C.G., SORNETTE, A., and SORNETTE, D. (1998), *An observational test of the critical earthquake hypothesis*, J. Geophys. Res. *103*, 24, 359–24, 372.

HARDEBECK, J.L. and HAUKSSON, E. (1999), *Role of fluids in faulting inferred from stress field signatures*, Science *285*, 236–239.

JAUMÉ, S.C. and SYKES, L.R. (1999), *Evolving towards a critical point: a review of accelerating seismic moment/energy release prior to large and great earthquakes*, Pure Appl. Geophys. *155*, 279–306.

MELCHIOR, P., *The Tides of the Planet Earth* (Pergamon Press 1978).

NASA website, http://spacekids.hq.nasa.gov/osskids/animate/moon.html.

SORNETTE, A. and SORNETTE, D. (1990), *Earthquake rupture as a critical point: Consequences for telluric precursors* Tectonophysics *179*, 327–334.

SORNETTE, D. and SAMMIS, C.G. (1995), *Complex critical exponents from renormalization group theory of earthquakes: Implications for earthquake predictions*, J. Phys. I France *5*, 607–619.

WANG, Y.C., MORA, P., YIN, C., and PLACE, D. (2003), *Statistical test of the Load-Unload Response Ratio (LURR) signals using the Lattice Solid Model (LSM): Implication to tidal triggering and earthquake prediction*, Pure Appl. Geophys. *161*, 1829–1839.

XIA, M.F, WEI, Y.J., KE, F.J., and BAI, Y.L. (2002), *Critical sensitivity and trans-scale fluctuations in catastrophe rupture,* Pure Appl. Geophys**.** *159* (10), 2491–2509.

YIN, X.C. and YIN, C. (1991), *The precursor of instability for nonlinear systems and its application to earthquake prediction,* Science in China *34*, 977–986.

YIN, X.C, WANG, Y.C., PENG, K.Y., and BAI, Y.L. (2000), *Development of a new approach to earthquake prediction: Load/Unload Response Ratio (LURR) Theory*, Pure Appl. Geophys. *157*, 2365–2383.

YIN, X.C., YU, H., KUKSHENKO, V., PENG, K., and XU, Z. (2004), *Load-Unload Response ratio (LURR), Accelerating Energy release (AER) and State Vector evolution as precursors to failure of rock specimens*, Pure Appl. Geophys. *161*, 11/12.

YIN, X.C., CHEN, X.Z., SONG, Z.P., and YIN, C. (1995), *A new approach to earthquake prediction: The Load/Unload Response Ratio (LURR) theory*, Pure Appl. Geophys. *145*, 701–715.

ZHANG, X.H., XU, X.H., XIA, M.F., and BAI, Y.L. (2004)*, Critical sensitivity in driven nonlinear threshold systems,* Pure Appl. Geophys. *161*, 11/12.

ZHAO, D. and KANAMORI, H. (1998), *The 1994 Northridge earthquake: Tomography of the source area and temporal stress rotation*, 1998 Fall Meeting of the Seismological Society of Japan, October 26–28, 1998, Fukui, Japan, Programme and Abstracts, B38.

ZHUANG, J.C. and YIN, X.C. (2000), *Random distributions of the loading and unloading response ratio under the assumptions of the Poisson model*, Earthquake Research in China *14*, 37–48.

ZOBACK, M.L. (1992), *First- and Second-Order Patterns of Stress in the Lithosphere: The World Stress Map Project,* J. Geophys. Res.-Solid Earth *97,* 11703–11728.

(Received December 30, 2004, revised December 23, 2005, accepted December 30, 2005)

Pure appl. geophys. 163 (2006) 2375–2387
0033–4553/06/122375–13
DOI 10.1007/s00024-006-0128-9

Pure and Applied Geophysics

An Independent Assessment of the Load/Unload Response Ratio (LURR) Proposed Method of Earthquake Prediction

J.E. Trotta and T.E. Tullis

Abstract—The Load/Unload Response Ratio (LURR) method is a proposed technique to predict earthquakes that was first put forward by Yin (1987). LURR is based on the idea that when an area enters the damage regime, the rate of seismic activity during loading of the tidal cycle increases relative to the rate of seismic activity during unloading in the months to one year preceding a large earthquake. Since earth tides generally contribute the largest temporal variations in crustal stress, it seems plausible that earth tides would trigger earthquakes in areas that are close to failure (e.g., Vidale *et al.*, 1998). However, the vast majority of studies have shown that earth tides do not trigger earthquakes (e.g., Vidale *et al.*, 1998; Heaton, 1982; Rydelek *et al.*, 1992). In this study, we conduct an independent test of the LURR method, since there would be important scientific and social implications if it were proven to be a robust method of earthquake prediction. Smith and Sammis (2004) undertook a similar study and found no evidence that there was predictive significance to the LURR method. We have repeated calculations of LURR for the Northridge earthquake in California, following both the parameters of X.C. Yin (personal communication) and the somewhat different ones of Smith and Sammis (2004). Though we have followed both sets of parameters closely, we have been unable to reproduce either set of results. Our examinations have shown that the LURR method is very sensitive to certain parameters. Thus it seems likely that the discrepancies between our results and those of previous studies are due to unaccounted for differences in the calculation parameters. A general agreement was made at the 2004 ACES Workshop in China between research groups studying LURR to work cooperatively to resolve the differences in methods and results, and thus permit more definitive conclusions on the potential usefulness of the LURR method in earthquake prediction.

Key words: Earthquake prediction, tides, LURR, triggering, seismicity.

Introduction

Scientists have been researching earthquake prediction for over 100 years and yet there is no reliable short-term method of earthquake prediction (Geller, 1997). Loading caused by earth tides is one of many different potential earthquake triggers that have been suggested over the years. Regions near failure have tectonic stresses nearly high enough for a large earthquake to occur. Thus it seems more likely that smaller earthquakes in the region would be triggered when the tidal stresses add to

Department of Geological Sciences, Brown University, 324 Brook Street, Providence, Rhode Island, 02912, USA. E-mail: Julie_Trotta@alumni.brown.edu

the tectonic ones. However, until recently even the most careful studies suggested that the effect of tidal stresses on earthquake occurrence is very small and difficult to detect. New studies have shown that there is a tidal triggering effect on shallow thrust faults in areas with strong tides from ocean loading (TANAKA *et al.*, 2002; COCHRAN *et al.*, 2004).

The Load/Unload Response Ratio (LURR) method is a new approach to earthquake prediction first proposed by YIN (1987), and further developed by YIN *et al.* (2000). The LURR method is based on the idea that when a region is near failure, there is an increase in the rate of seismic activity during loading of the tidal cycle relative to the rate of seismic activity during unloading. Typically the numerator of the LURR ratio is the number, or the sum of some measure of the size (e.g., Benioff strain), of small earthquakes that occur during loading of the tidal cycle, whereas the denominator is the same as the numerator except it is calculated during unloading. The LURR method suggests this ratio should increase in the months to year preceding a large earthquake.

We conducted an independent assessment of the LURR method, since there would be important scientific and social implications if the LURR method were proven to be a useful means of earthquake prediction. SMITH and SAMMIS (2004) undertook a similar study, and concluded there was no evidence that the LURR method had any predictive significance. We have repeated calculations of LURR for the Northridge earthquake (M_0 = 6.7, 34.44°N, −118.64°E) that occurred in California on January 17, 1994, following both the parameters of X.C. YIN (personal communication) and the somewhat different ones of SMITH and SAMMIS (2004). We have been unable to reproduce either set of results, though we have had some success validating our calculations for a test case with C. YIN. Through the published works and discussions with X.C. YIN, we have attempted to duplicate the parameters used by YIN *et al.* (2000) and X.C. YIN (personal communication). However, our examinations have shown that the LURR method is very sensitive to certain parameters, such as the size and shape of the search area. Thus, it is plausible that some unknown parameter discrepancy could cause the disagreement between our results and those of previous studies. A general agreement was reached at the 2004 ACES Workshop in China between research groups studying LURR, to work cooperatively to discover the causes of these differing in results. The goal of this cooperative effort is to resolve the differences methods and results to allow for more conclusive evidence on the potential usefulness of LURR in earthquake prediction.

Data Sets and Parameters of Previous Studies

X.C. YIN (personal communication) and SMITH and SAMMIS (2004) used different data sets for their analyses. SMITH and SAMMIS (2004) used estimates of the

parameters determined from published works. However, their estimates were somewhat different from the actual parameters used by X.C. YIN (personal communication) and the two calculations did not lead to the same conclusions. Both groups used data sets that were composed of earthquakes occurring several years before the Northridge earthquake. However the data sets varied in several significant ways. The size and shape of the search areas (the geographic region in which earthquake data was collected) were different. The search area used by SMITH and SAMMIS (2004) was much larger than that of X.C. YIN (personal communication) and the data were collected over a longer time period, resulting in a data set that contained many more earthquakes. X.C. YIN (personal communication) ended their data sets with the last earthquake before Northridge occurred. The data set used by SMITH and SAMMIS (2004) does not end until a year after the Northridge earthquake and includes Northridge and all of its aftershocks. In addition, the upper and lower magnitude cutoffs for earthquakes included in the data sets were different in the two studies.

There are also several calculation parameters that vary in significant ways in the two studies. Both groups used overlapping time windows with a two-month time step between windows, but the length of the averaging time window used and the approximate number of earthquakes in each data point were different. X.C. YIN (personal communication) requires a minimum of 20 earthquakes in each data point. SMITH and SAMMIS (2004) need no such requirement, as there are several dozen earthquakes in each of their data points due to the large volume of their data set. The two groups also plot the data points in different locations within the time windows.

SMITH and SAMMIS (2004) considered the effects of ocean loading, whereas X.C. YIN (personal communication) did not. Thus, in following the methods of X.C. YIN (personal communication), we did not consider ocean loading, though we did when following the techniques of SMITH and SAMMIS (2004). X.C. YIN (personal communication) considered the depth-dependence of the solid earth tidal loading whereas SMITH and SAMMIS (2004) did not. We follow SMITH and SAMMIS (2004), and assume that deformation and stress will only vary slightly over the depth range relevant to crustal earthquakes, since tidal effects are a body force exerted throughout the earth from remote sources and the < 20 km depth range of interest is a small fraction of the radius of the earth. Nevertheless, there are small variations in solid earth tides that are included in the calculations of X.C. YIN (personal communication) and neglected by SMITH and SAMMIS (2004) and us. Though the depth variations can be significant for ocean loading, we did not have access to programs capable of calculating ocean loading stresses at depth. Thus, like SMITH and SAMMIS (2004), we did not consider depth-dependence in any of our analyses. The main differences between the parameters of X.C. YIN (personal communication) and SMITH and SAMMIS (2004) are summarized in Table 1.

Table 1

Parameters used in previous LURR studies. In this study, we followed the parameters presented in this table, except where noted

Parameter	YIN	SMITH and SAMMIS
Search area	Polygonal	120 km radial search area; Northridge epicenter at center
Magnitude earthquakes included in data set	2.5 – 4.9	2.0 and up
Length of time window	1 year	6 months
Northridge and its aftershocks included?	No	Yes
Tides included	Solid earth only	Solid earth + ocean
Depth variations of solid earth tides included?	Yes*	No
Depth variations of ocean loading stresses included?	No	No
Data points plotted	End of time window	Middle of time window
Δ Time used to determine change in Coulomb Failure Stress (ΔCFS)	10 minutes	1 hour
Earthquake catalog used	Council of the National Seismic System (CNSS)	Northern California Earthquake Data Center (NCEDC)

* We neglect any depth-dependence of solid earth tides in our calculations.

Tidal Calculations

Solid earth tidal strains were calculated using code provided by AGNEW (personal communication), which we modified to calculate shear strain as well as longitudinal strain. Shear and normal strains on the fault plane and normal strain on the plane perpendicular to the slip vector were calculated for a N-S strike-slip fault with a coordinate system in which x_1 corresponds to east, x_2 to north, and x_3 to up. We follow the convention that tensile stress is positive. For the comparisons with SMITH and SAMMIS (2004), we also calculated the strain due to ocean loading using programs provided by AGNEW (1996) and summed the solid and ocean-loading strains to obtain the total effect of the tides. We included local ocean-tide effects in the Gulf of California. Stresses were calculated from the strains according to:

$$\sigma_{12} = 2G\varepsilon_{12},$$

$$\sigma_{11} = \frac{2G}{1-\nu}(\varepsilon_{11} + \nu\varepsilon_{22}),$$

$$\sigma_{22} = \frac{2G}{1-\nu}(\varepsilon_{22} + \nu\varepsilon_{11}),$$

where $\boldsymbol{\sigma}$ is the stress, ε is the strain, υ is Poisson's ratio, and G is the shear modulus. We used $\upsilon = 0.25$, and $G = 3 \times 10^4$ MPa. These stresses are then transformed into the Northridge fault plane (strike = 130°; dip = 53°; slip/rake = 111°) according to:

$$\sigma'_{12} = a_{11}a_{21}\sigma_{11} + a_{11}a_{22}\sigma_{12} + a_{12}a_{21}\sigma_{21} + a_{12}a_{22}\sigma_{22},$$
$$\sigma'_{11} = a_{11}a_{11}\sigma_{11} + a_{11}a_{12}\sigma_{12} + a_{12}a_{11}\sigma_{21} + a_{12}a_{12}\sigma_{22},$$
$$\sigma'_{22} = a_{21}a_{21}\sigma_{11} + a_{21}a_{22}\sigma_{12} + a_{22}a_{21}\sigma_{21} + a_{22}a_{22}\sigma_{22},$$

where $\boldsymbol{\sigma}'$ is the stress in the new coordinate system with x'_1 normal to the fault plane and $+x'_2$ parallel to the direction of slip of the $+x'_1$ side of the fault, and $\mathbf{A}$ is the 3×3 coordinate transformation matrix. Since we assume there is no depth variation, $\sigma_{3i} = 0$.

Loading or unloading is determined by the Coulomb Failure Stress (*CFS*), which is calculated according to:

$$CFS = \sigma'_{12} + f\sigma'_{11},$$

where σ'_{12} is the shear stress on the fault plane (positive in the sense of slip), f is the coefficient of friction, and σ'_{11} is the normal stress on the fault plane. We used a standard value of $f = 0.6$ for the friction coefficient. If $\Delta CFS > 0$, the fault plane is undergoing loading, if $\Delta CFS < 0$, the fault plane is unloading. Once loading or unloading has been determined, the LURR value (Y) can be calculated according to:

$$Y_m = \frac{\left(\sum_{i=1}^{N^+}\left(E_i^m\right)\right)_+}{\left(\sum_{i=1}^{N^-}\left(E_i^m\right)\right)_-},$$

Where E is the seismic energy, + indicates loading, and – indicates unloading. We calculate E following Bullen (1947): $E \approx E_0 10^{1.8M}$, where M is the earthquake magnitude and E_0 is the energy of a zero magnitude earthquake (2×10^{11}ergs). When $m = 1$, E^m is the seismic energy itself; when, $m = 1/2$, E^m is the Benioff strain ($\sqrt{E}$); when $m = 0$, $Y = \mathrm{N}^+/\mathrm{N}^-$; where N^+ is the number of earthquakes that occurred during loading and N^- is the number that occurred during unloading (Yin *et al.,* 2000).

Results

We examined the Northridge earthquake following both the methods of X.C. Yin (personal communication) and Smith and Sammis (2004). While we followed both sets of parameters closely, we were unable to reproduce either set of results, even though we were able to nearly reproduce the results of C. Yin (personal communication) for the tidal stresses at the earth's surface for an example day. We have observed in our examinations that the LURR method is very sensitive to various parameters, such as the size and shape of the search area and the length of

the averaging window. We feel that slight differences may cause large changes in the resulting Y values, so any unaccounted for difference may explain why we have not yet been able to duplicate the results of the previous studies. LURR values from this study obtained using the parameters of X.C. YIN (personal communication) and SMITH and SAMMIS (2004) are shown in Figures 1 and 2. Results from this study are shown along with results of the previous studies in Figure 3.

The inclusion of ocean loading effects only had a major impact on the calculation of LURR when adding its effects changed the sign of a moderate earthquake. This is illustrated in Figure 2, which shows the LURR values calculated with and without including the effects of ocean loading. The only point at which the results in Figure 2 strongly deviate is for $Y_{1/2}$ values (LURR values weighted by the earthquake magnitude) between 1991 and 1992. This deviation is caused by a sign change of a magnitude 5.8 earthquake occurring in that interval. This moderate earthquake occurred during unloading if including only solid earth tides in the calculations, but occurred during loading once the ocean loading effects were added. The effect of

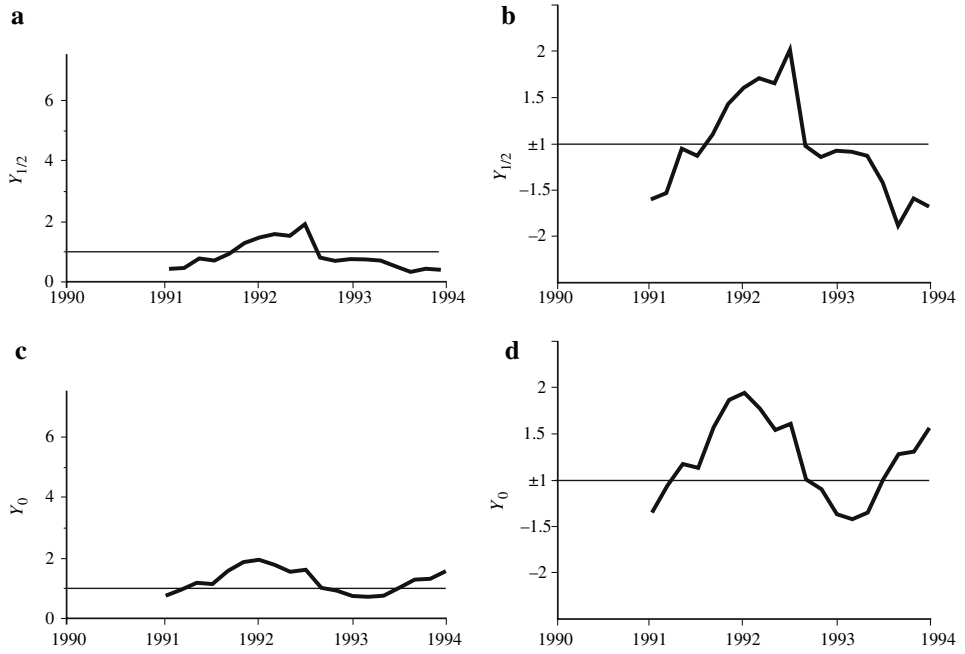


Figure 1
LURR results from this study, calculated following the methods and parameters of X.C. YIN (personal communication). The Northridge earthquake occurred on January 17, 1994. No increase in seismic activity during loading of the tidal cycle relative to unloading is observed. **a)** $Y_{1/2}$ values from this study shown on a standard linear scale. **b)** $Y_{1/2}$ values from this study shown on an improved linear scale in which the minus inverse of the Y ratio for $0 < Y < 1$ is plotted below the $Y = \pm 1$ line to give equal weight to small LURR values. **c)** Y_0 values from this study shown on a standard linear scale. **d)** Y_0 values from this study shown on an improved linear scale.

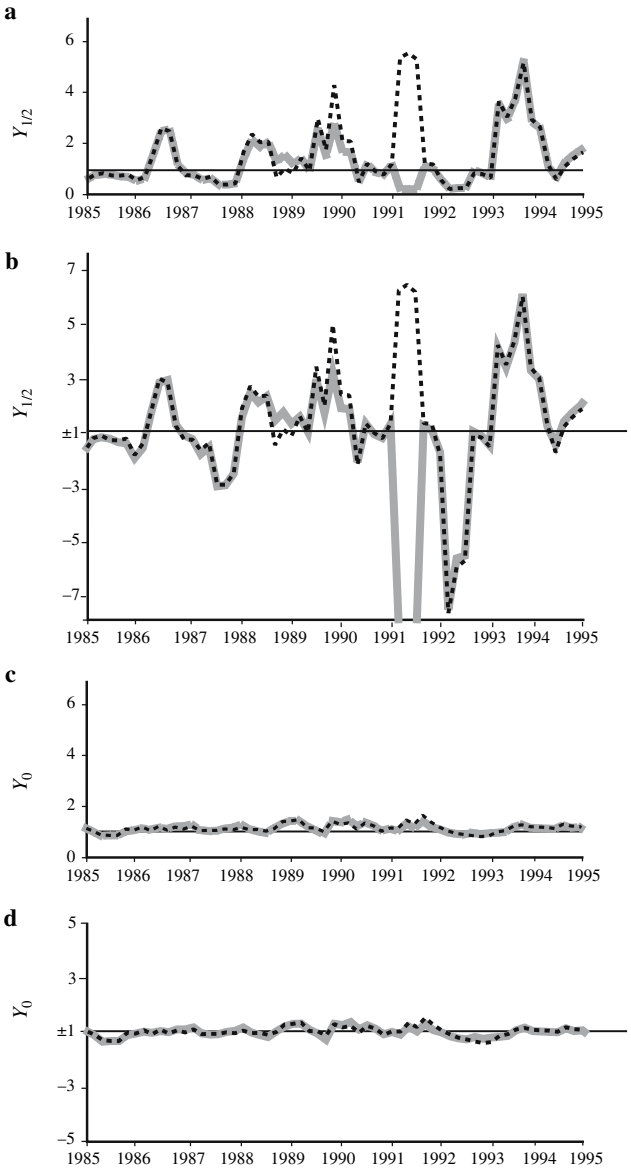


Figure 2

LURR results from this study, calculated following the methods and parameters of SMITH and SAMMIS (2004). Grey lines indicate the values were obtained using solid loading only and the dashed line is for values calculated using solid and ocean loading. The Northridge earthquake occurred on January 17, 1994. $Y_{1/2}$ plots indicate there may be a precursory effect, however Y_0 plots show there is no increase in seismic activity during loading compared to unloading, indicating that the spike in $Y_{1/2}$ values is more likely due to Northridge and its aftershocks. **a)** $Y_{1/2}$ values from this study shown on the standard linear scale. **b)** $Y_{1/2}$ values from this study shown on the previously described new style of linear plot. **c)** Y_0 values from this study shown on a standard linear scale. **d)** Y_0 values from this study shown on an improved linear scale.

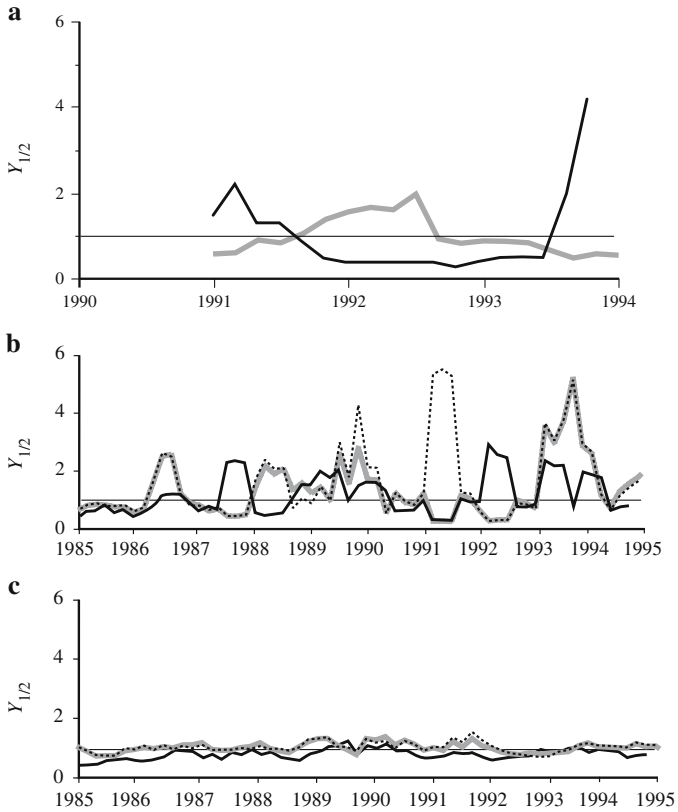


Figure 3
LURR results from this study, shown along with results from previous studies. **a)** $Y_{1/2}$ values from this study are shown in grey and results from X.C. YIN (personal communication) are shown in black. There is no correlation in results and there is a precursory effect seen in the results of X.C. YIN (personal communication) that is not observed in the results of this study. **b)** $Y_{1/2}$ values from this study calculated using solid loading only are shown in grey, $Y_{1/2}$ values calculated using solid and ocean loading are shown in the dashed line. Results of SMITH and SAMMIS (2004) are shown in the solid black line. There are some similarities between the results, but it is clear they are not well correlated. **c)** Y_0 values from this study calculated using solid loading only are shown in grey, Y_0 values calculated using solid and ocean loading are shown in the dashed line. Results of SMITH and SAMMIS (2004) are shown in the solid black line. Results from both studies remain roughly 1 over the observed interval, but the results still do not appear to be well correlated.

adding the ocean loading is much less pronounced when looking at LURR values calculated using only event count (Y_0), as can be seen in Figure 2. It is interesting to note that SMITH and SAMMIS (2004) did not see any changes in sign of moderate earthquakes due to addition of the ocean load.

If there was an increase in seismic activity during loading relative to the seismic activity during unloading, that increase should be apparent in the LURR ratio regardless of whether or not the earthquakes are weighted by their magnitude. None

of our results show a noticeable increase in Y_0 values leading up to the Northridge earthquake. In fact, the ratio stays near 1 in all of our results, indicating there is an approximately equal amount of seismic activity during loading and unloading over the length of the data sets. SMITH and SAMMIS (2004) point out that showing the LURR ratio on a standard linear scale biases the display of the results, because values between zero and one are confined to a small part of the plot. We also show our results on a new style of linear plot in which the minus inverse of the Y ratio for $0 < Y < 1$ is plotted below the $Y = \pm 1$ line to give equal weight to small LURR values. Displaying the results on this improved linear scale helps to illustrate that there are approximately equal amounts of data plotted above and below 1, meaning that seismic activity is approximately equal during loading and unloading.

Synthetic Data Sets

Since such close detail was paid to data set compilation and calculation parameters, we were surprised to see the large variations between our results and the results of the previous studies. This problem also intrigued X.C. YIN, C. YIN, and SAMMIS when our results were presented and discussed at the ACES Workshop, 2004. An agreement was made to investigate the source of these discrepancies using synthetic data sets to test our results.

To begin such tests, two synthetic data sets were developed. We started with one artificial earthquake occurring every hour at the same location. All of the earthquakes are located at the Northridge epicenter (34.44°N, −118.64°E) and we arbitrarily did calculations for the 5 days beginning on January 1, 2000 and ending on January 5, 2000. We took this synthetic set of 120 earthquakes and calculated *CFS* to determine loading or unloading, using one hour between calculations. We then picked 10 earthquakes that occurred during loading and 10 that occurred during unloading from a plot of *CFS* vs. time to comprise our synthetic data set. This procedure was followed for a N-S vertical strike-slip fault assumed to be slipping in a left lateral sense and for the Northridge fault plane. All of the earthquakes that occurred during loading were assigned a magnitude of 4.0, and the earthquakes that occurred during unloading were assigned a magnitude of 2.5. These magnitude assignments have no physical meaning, however they did allow us to calculate an expected LURR value by simple equations to determine whether or not our calculations using our standard computer programs produced the expected result. In both cases all calculations were done at the earth's surface and the programs used in this study did produce the expected LURR values.

We have compared stress calculations for the first day of the vertical strike-slip fault data set with results provided by C. YIN (personal communication). Figure 4 illustrates that our results are in good agreement with those of C. YIN. We have not yet compared results for the entire synthetic data set or for the Northridge fault

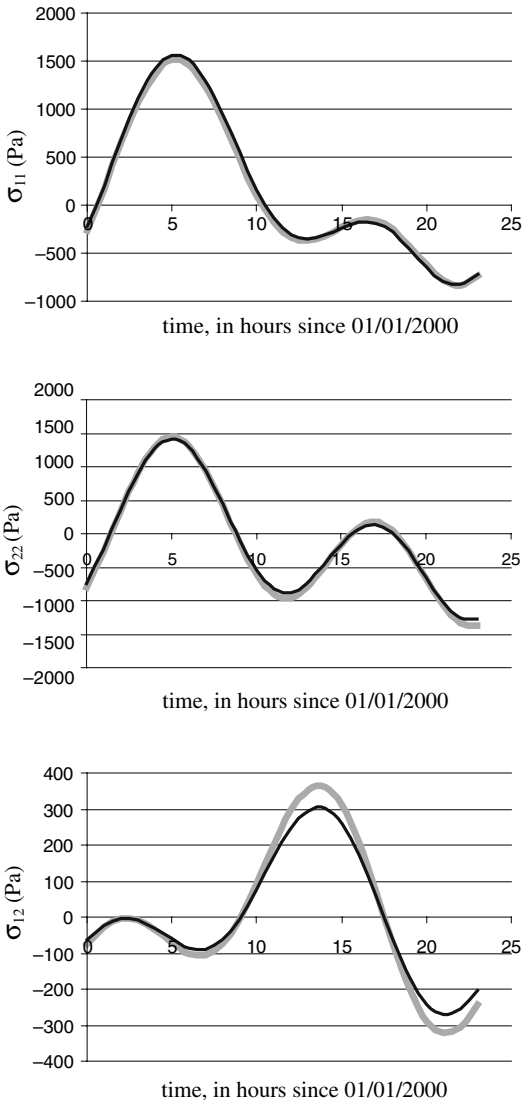


Figure 4
Stress comparisons for a N-S vertical strike-slip fault assumed to be slipping in a left lateral sense with results of C. YIN (personal communication). Stresses were calculated at the earth's surface. Results from this study are shown in grey and results from C. YIN are shown in black. Generally, there is very good agreement between the two results.

plane. It is clear that our calculations or data sets must differ somehow from those of X.C. YIN (personal communication), as our LURR calculations leading up to the Northridge earthquake do not agree. Our goal is to use these synthetic data sets to discover where the calculations deviate and to resolve why those deviations have such

a significant effect on the resultant LURR values. Since we calculate solid earth tides at the surface and X.C. YIN (personal communication) and C. YIN (personal communication) calculate the tides at depth, the differences in results could be due to this discrepancy, although the depth variation is so small that we think this unlikely. Further testing should resolve this issue. We have sent our results for these synthetic data sets to the other participants, who have not yet done the same calculations using their programs. We anticipate a comparison of results by all parties will be available in the future.

Random Data Sets

We performed a series of calculations in which we randomly assigned loading or unloading to each earthquake in the data set, effectively removing any actual effect of the tides. If large Y values were indicative of an increase in the ratio of seismic activity during loading of the tidal cycle relative to unloading, one would expect that randomizing the data sets would result in Y values that remained near 1 for the observed interval. Representative results of five such random data sets (out of the 100

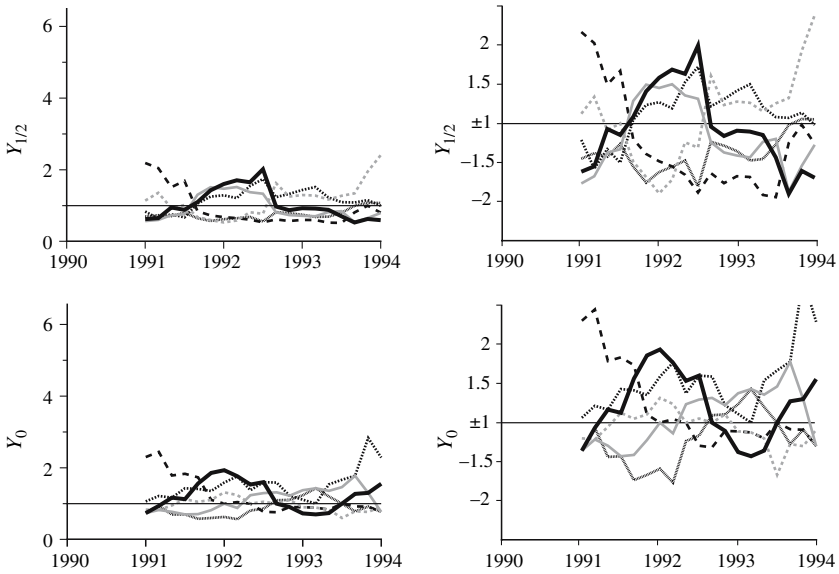


Figure 5

Results of randomized data sets. The heavy black line is the actual LURR values from this study, calculated following the methods and parameters of X.C. YIN (personal communication). The other 5 lines represent 5 data sets in which loading or unloading has been randomly determined. There appears to be an equal amount of variation in the randomized data sets as in the actual data set. **a)** $Y_{1/2}$ values for the actual data and 5 randomized data sets. **b)** $Y_{1/2}$ values for the actual data and 5 randomized data sets shown on an improved linear scale. **c)** Y_0 values for the actual data and 5 randomized data sets. **d)** Y_0 values for the actual data and 5 randomized data sets shown on an improved linear scale.

that we examined) are shown in Figure 5. While we have not done rigorous testing of these results, it appears that randomly assigning loading or unloading causes an equal amount of variation in the Y value as does calculating the real effect of the tides when viewed on the usual linear scale. When viewed on the improved linear scale, it is clear that there is no increase in seismic activity during loading relative to that during unloading for any of the random data sets, or for the actual data set. Figure 5 illustrates how showing the results on the standard linear scale may bias the display. The results of the time randomizations imply that it may not be the tides causing the elevated Y values seen leading up to the Northridge earthquake in the results of YIN *et al.* (2000) and X.C. YIN (personal communication), because random fluctuations appear to have the same effect on LURR values that YIN *et al.* (2000) report is controlled by increased seismic activity during loading relative to unloading.

Conclusions

Though the majority of studies have shown that there is no statistically significant correlation between earth tides and earthquakes except for under very special conditions, the LURR method is a novel approach to earthquake prediction that appears to have some merit. Retrospective examinations presented by YIN *et al.* (2000) and X.C. YIN (personal communication) indicate that $Y_{1/2}$ values typically do increase beginning several months to one year before the predicted earthquake. However, the results of this study and the results of SMITH and SAMMIS (2004) do not find the same precursory activity. In fact, using event count as an indicator, this study and the study of SMITH and SAMMIS (2004) found no increase in seismic activity during loading of the tidal cycle relative to unloading of the tidal cycle leading up to the Northridge earthquake. In addition, data sets in which the time of each event was randomly assigned as loading or unloading showed an equal amount of variation in the LURR values leading up to an event as did the actual data set, indicating that the apparent precursory effect presented by YIN *et al.* (2000) and X.C. YIN (personal communication) may be influenced by random fluctuations. Since LURR calculations are so sensitive to parameters used, the process by which parameters are chosen must be made clear in order to use the LURR method as an earthquake prediction tool. We are hopeful that testing of the synthetic data sets will resolve the differences between the studies and facilitate determination of the potential usefulness of the LURR method as a means of earthquake prediction.

Acknowledgements

Earthquake catalogs used in this study were obtained on-line from the Northern California Earthquake Data Center (NCEDC) and the member networks of the Council of the National Seismic System (CNSS). Earthquake mechanisms were

provided by the on-line Harvard CMT Project. Earth tide programs were made available by D. C. Agnew. Particular gratitude to X.C. Yin, C. Yin, S. Smith, and C. Sammis for providing details regarding their data sets and calculations as well as useful conversations on LURR method. Appreciation also to D.C. Agnew for providing the tidal calculation codes and useful discussions on their implementation and to John Vidale for his suggestions regarding the random data sets.

References

Agnew, D.C. (1996), *SPOTL: Some Programs for ocean-tide loading,* SIO Ref. Ser. 96–8, 35 pp., Scripps Institution of Oceanography, La Jolla, CA.

Bullen, K.E., *Introduction to the Theory of Seismology* (University Press, Cambridge 1947).

Cochran, E.S., Vidale, J.E., and Tanaka, S. (2004), *Earth tides can trigger shallow thrust fault earthquakes,* Science *306,* 1164–1166.

Geller, R.J. (1997), *Earthquake prediction: A critical review,* Geophys. J. Int. *131,* 425–450.

Heaton, T.H. (1982), *Tidal triggering of earthquakes,* Bull. Seismol. Soc. Am. *72,* 2181–2200.

Rydelek, P.A., Sacks, I.S., and Scarpa, R. (1992), *On tidal triggering of earthquakes at Campi Flegrei, Italy,* Geophys. J. Int. *109,* 125–137.

Smith, S.W. and Sammis, C. G. (2004), *Revisiting the tidal activation of seismicity with a damage mechanics and friction point of view,* Pure Appl. Geophys. *161*, 2393–2404.

Tanaka, S., Ohtake, M., and Sato, H. (2002), *Evidence for tidal triggering of earthquakes as revealed from statistical analysis of global data*, J. of Geophys. Res. *107*(B10), 2211, doi:10.1029/2001JB001577.

Vidale, J.E., Agnew, D.C., Johnston, M.J.S., and Oppenheimer, D.H. (1998), *Absence of earthquake correlation with earth tides: An indication of high preseismic fault stress rate,* J. Geophys. Res. *103,* 24,567–24,572.

Yin, X.C. (1987), *A new approach to earthquake prediction,* Earthq. Res. China *3,* 1–7 (in Chinese with English abstract).

Yin, X.C., Chen, X.Z., Song, Z.P., and Yin, C. (1995), *A new approach to earthquake prediction: The Load/Unload Response Ratio (LURR) theory,* Pure Appl. Geophys. *145,* 701–715.

Yin, X.C., Wang, Y., Peng, K., Bai, Y., Wang, H., and Yin, X. (2000), *Development of a new approach to earthquake prediction: Load/Unload Response Ratio (LURR) theory,* Pure Appl. Geophys. *157,* 2365–2383.

(Received January 6, 2005, revised May 15, 2005, accepted June 9, 2005)
Published Online First: December 20, 2006

Pure appl. geophys. 163 (2006) 2389–2406
0033–4553/06/122389–18
DOI 10.1007/s00024-006-0129-8

Pure and Applied Geophysics

Acoustic Emission Experiments of Rock Failure Under Load Simulating the Hypocenter Condition

Hui-hui Zhang,[1,6] Xiang-chu Yin,[1,2,3] Nai-gang Liang,[1] Huai-zhong Yu,[1] Shi-yu Li,[3] Y.C. Wang,[4] C. Yin,[4] Victor Kukshenko,[5] Nikita Tomiline,[5] and Surguei Elizarov[6]

Abstract—A series of acoustic emission (AE) experiments of rock failure have been conducted under cyclic load in tri-axial stress tests. To simulate the hypocenter condition the specimens are loaded by the combined action of a constant stress, intended to simulate the tectonic loading, and a small sinusoidal disturbance stress, analogous to the Earth tide induced by the Sun and the Moon. Each acoustic emission signal can indicate the occurrence time, location and relative magnitude of the damage (micro-crack) in the specimen. The experimental results verified some precursors such as LURR (Load/Unload Response Ratio) and AER (Accelerating Energy Release) before macro-fracture of the samples. A new parameter, the correlation between the AE and the load, has been proposed to describe the loading history. On the eve of some strong earthquakes the correlation between the Benioff strain and the Coulomb failure stress (CFS) decreases, similar to the variation of LURR prior to strong earthquakes.

Key words: Acoustic emission (AE), cyclic load, load/unload response ratio (LURR), critical point hypothesis (CPH), accelerating energy release (AER), correlation (Corr).

Introduction

In recent years many investigators have attempted to use the methods of statistical physics to understand regional seismicity (Rundle, 1988a,b; 1989a,b). One approach has been put forward to model the earthquake process as a critical phenomenon (Sornette and Sornette, 1990; Sornette and Sammis, 1995; Bowman *et al.*, 1998; Rundle *et al.*, 1999; Huang, 1998; Ito, 1990). According to

[1] LNM (State Key Laboratory of Nonlinear Mechanics), Institute of Mechanics, CAS (Chinese Academy of Sciences), Beijing, 100080. E-mail: zhanghh@lnm.imech.ac.cn
[2] Institute of Earthquake Prediction, China Seismological Bureau, Beijing 100036. E-mail: xcyin@public.bta.net.cn
[3] Institute of Geophysics, China Seismological Bureau, Beijing, 100081, China.
[4] QUAKES, Department of Earth Sciences, The University of Queensland, Brisbane, 4072, Australia. E-mail: wangyc@esscc.uq.edu.au
[5] Ioffe Physical Technique Institute, Russian Academy of Sciences, Russia. E-mail: Victor.Kuksenko@pop.ioffe.ssi.ru
[6] School of Civil Engineering and Mechanics, Yanshan University, China, 066004.

the Critical Point Hypothesis (CPH), the earth's crust does not remain perpetually in or near a critical state, but repeatedly approaches and retreats from a critical state. The critical point hypothesis for large earthquakes predicts two different precursory phenomena in space and time, an accelerating moment release and the growth of the spatial correlation length between its different parts. XIA *et al.* (2002) presented critical sensitivity and trans-scale fluctuations associated with catastrophic rupture in a system. Critical sensitivity means that the system becomes significantly sensitive near the catastrophe transition. This could lead to triggering of earthquakes by tidal stress (GRASSO and SORNETTE, 1998) and consequently anomalously high values of LURR (YIN and YIN, 1991; YIN, 1993; Yin *et al.*, 1994–2000, 2002) are observed prior to strong earthquakes. The establishment of long-range correlations in the regional stress field (SAMMIS and SMITH, 1999; MORA and PLACE, 2000, 2002; WEATHERLEY, 2002) and accelerating seismic activity of moderate-sized earthquakes (ELLSWORTH *et al.*, 1981; KEILIS-BOROK, 1990; SORNETTE and SAMMIS, 1995; KNOPOFF *et al.*, 1996 and BOWMAN *et al.*, 1998) and equivalently, the accelerating seismic release prior to large earthquakes follows a power-law time-to-failure formula. LURR and AER reflect the changing critical sensitivity and growing correlation length before large earthquakes, respectively.

It is recognized by many scientists that, from the viewpoint of physics, earthquake is the failure or instability of the focal media accompanied by a rapid release of energy. Similarly, the acoustic emissions in rock experiment are elastic waves generated in conjunction with energy release during crack onset and propagation and internal deformations in the rock body. Laboratory experiments of rock deformation are considered as a tool for understanding the occurrence of natural earthquakes. By recording AE during a laboratory test of rock samples, significant additional information of the failure process can be obtained. Information about the onset and propagation of micro-cracking and fracture in rock samples, subjected to different stress regimes can be determined by recording the time and location of AE during the test. Using AE to map out fault nucleation and propagation may also be useful in understanding earthquake mechanisms and may contribute to solving the problem of earthquake prediction (LOCKNER *et al.*, 1991).

The main motivation of our work is to investigate the fracture precursors of heterogeneous brittle material. We are going to monitor the acoustic emission (AE) from micro-fractures that occur before the final fracture, and test LURR and AER using acoustic emission data recorded during rock fracture experiments.

Experimental Setup and Procedures

The experiment designed for this research program utilizes large rectangular prisms, rock samples. The samples include sandstone from Chengdu City, Sichuan Province, and gneiss from Jinzhou City, Liaoning Province. The dimensions

of Chengdu sandstone and Jinzhou gneiss are $300 \times 360 \times 20\,\mathrm{mm}^3$ and $300 \times 360 \times 25\,\mathrm{mm}^3$, respectively. The other parameters for Jinzhou gneiss are: Young's modulus $E = 27$ Gpa, Possion ratio $v = 0.26$, density $\rho = 2.6 \times 10^3\,\mathrm{kg/m}^3$, and longitudinal sound velocity $v = 2900\,\mathrm{m/s}$; and for Chengdu sandstone: Young's modulus of $E = 19$ GPa, Possion ratio $v = 0.26$, density $\rho = 2.5 \times 10^3\,\mathrm{kg/m}^3$ and longitudinal sound velocity $v = 1500\,\mathrm{m/s}$.

The experiments are conducted using MTS-100 servo-control experimental equipment in Institute of Geophysics, China Seismological Bureau. The maximum load for this facility is 100 ton in axial direction and 30 ton in lateral direction. Boundary-displacement control is used to load the system until final failure. Samples are subjected to both axial stress σ_1 and lateral load σ_2 simultaneously and another principle stress σ_3 is zero. The greatest, intermediate and least principle stress are $\sigma_1, \sigma_2, \sigma_3$. So that:

$$\sigma_1 \neq \sigma_2 \neq \sigma_3.$$

Therefore, the stress state is a tri-axial stress state (Fig. 1). Under such a stress state all of the specimens undergo shear failure at the last stage.

In our experiments the lateral stress σ_2 is symmetrically applied to samples with steel plates on the two smaller sides of the specimens and it keeps constant until the samples fracture. Then the axial stress σ_1 is applied.

σ_1 consists of two parts: The constant loading rate of tectonic stress build-up and a sinusoidal stress perturbation which simulates the periodic loading and unloading

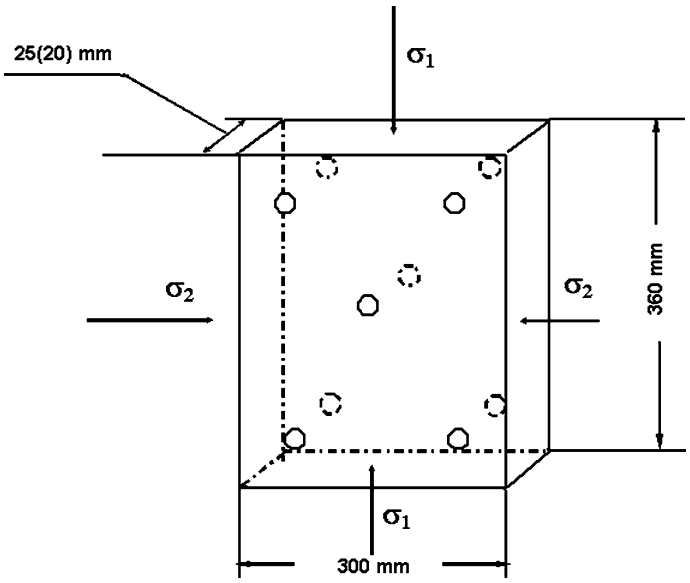


Figure 1
The Geometry of the specimens, the loading conditions, and the arrangement of AE sensors (circles).

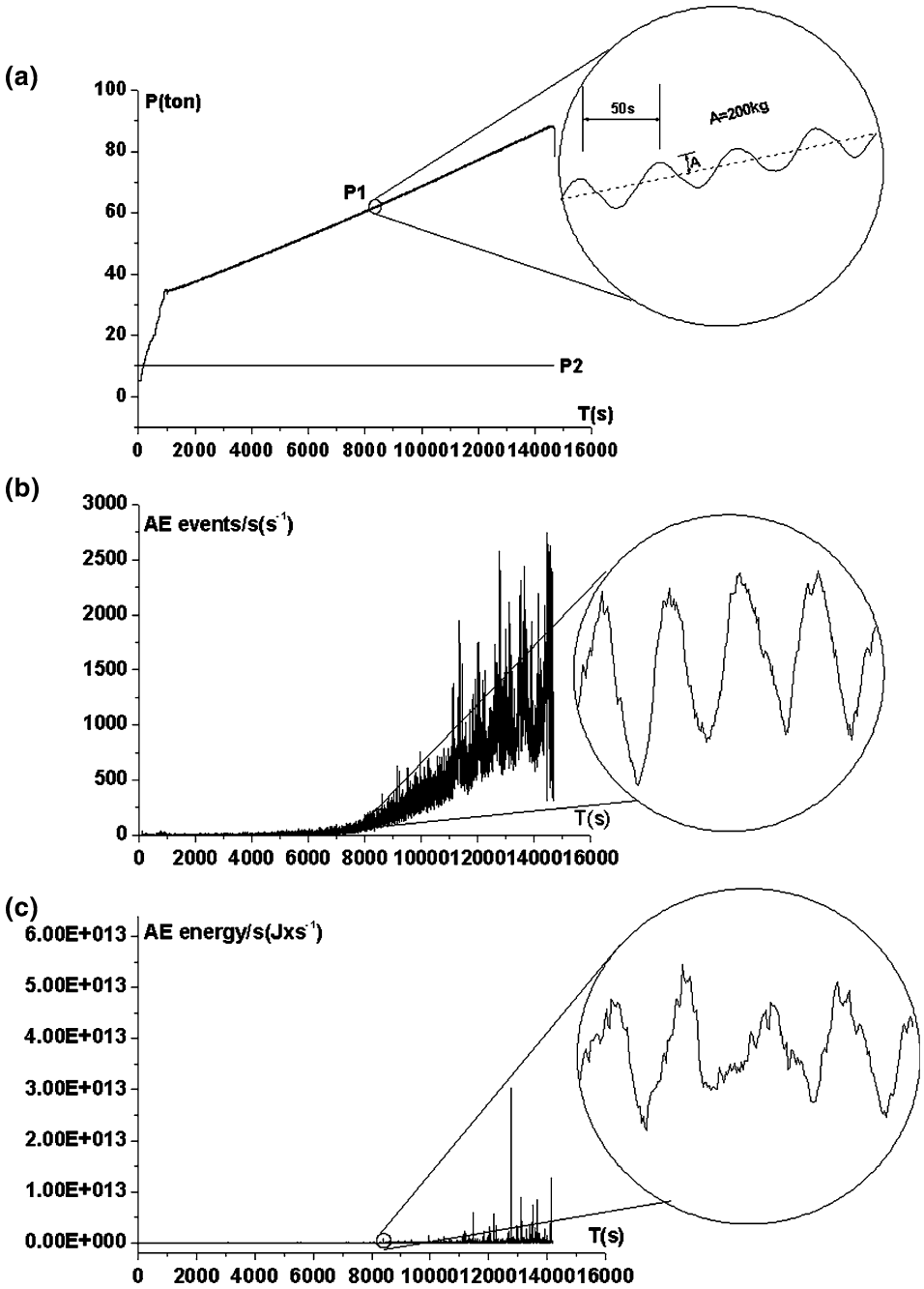
Figure 2 ▶

The loading history and the corresponding experimental results for gneiss specimen G2. The average loading velocity is 0.4 μm/s. (a) The loading history in the experiment for gneiss specimen G2. P1 is the axial load. The amplitude and the periodicity of the cyclic period loading cycles is 200 kg and 50 s, respectively. P2 is the lateral load and it keeps 10 tons during the entire course. (b) The AE event rate versus time for gneiss specimen G2. (c) The AE energy rate versus time for gneiss specimen G2.

cycles induced by tidal forces. During the early stage of seismogenic process, which is a prolonged period before the occurrence of a large earthquake and maybe several or dozens of years, the tectonic stress level is low. Subsequently, tectonic loading drives the crust towards the critical state. During the establishment of criticality, the crust in this region is damaged severely after which the crust will be sensitive to any tiny external disturbance, such as tidal stress whose minimum period is only a dozen hours. Thus the tidal stress period is only a very small fraction of the entire reoccurrence time between two strong earthquakes. On the other hand, it is well known that the resultant stress σ_{ij} in the crust consists of tectonic stress σ_{ij}^T and the tide induced stress σ_{ij}^t. Although the level of σ_{ij}^T in the crust (in the order of 10^6–10^8 Pa) is considerably higher than the level of σ_{ij}^t (10^3–10^4 Pa), the change rate of tidal-induced stress is considerably larger than the change rate of the tectonic stress (VIDALI *et al.*, 1998). Since the tidal stress period is small compared to the whole reoccurrence time between two strong earthquakes.

The loading history of gneiss sample G2 is shown in Figure 2(a). P_2 is the lateral load which keeps constant (10 ton) during the loading until the samples fracture. The axial load, P_1, increases to a certain value (33 ton) and then cyclic loading is added until the final fracture. The amplitude (200 kg) and the periodicity (50 sec) of the cyclic loading cycles are enlarged in the circle for more details. Figure 3(a) is the loading history of sandstone sample S1.

The acoustic emission equipment (12 digits, 32 channels) of Ioffe Physical Technique Institute, Russian Academy of Science is used to keep a continuous log (recording) of the time, location and intensity of micro-cracks within rock samples. Each channel consists of an AE sensor, a preamplifier and an AECB (Acoustic Emission Channel Board). The AE sensors pick up the sound waves from the specimens and convert them into an electrical signal that is then amplified by a preamplifier and converted into a digital data stream in an AECB. AE features such as arrival times, rise-times, duration, peak amplitude, energy and counts are extracted by a FPGA (Field Programmable Gate Array). Since an AE event is generally corresponding to a micro-cracking, AE hypocenter distribution reflects directly the distribution of micro-cracking. The rapid AE monitoring system can record AE waveforms without major loss of events, even for AE event rate on the order of several thousand events per second. Ten piezoelectric transducers (resonance from 50 to 250 kHz) are attached directly to the rock sample surface ('o' represents the location of transducers in Fig. 1), five on each side. They are used to monitor the high-frequency acoustic emissions generated in the sample as it is stressed. The


(a)
P(ton)
100
80
60
40
20
0
P1
P2
T(s)
0 2000 4000 6000 8000 10000 12000 14000 16000
50s
A=200kg
TA
(b)
AE events/s(s-1)
3000
2500
2000
1500
1000
500
0
T(s)
0 2000 4000 6000 8000 10000 12000 14000 16000
(c)
AE energy/s(Jxs-1)
6.00E+013
5.00E+013
4.00E+013
3.00E+013
2.00E+013
1.00E+013
0.00E+000
T(s)
0 2000 4000 6000 8000 10000 12000 14000 16000

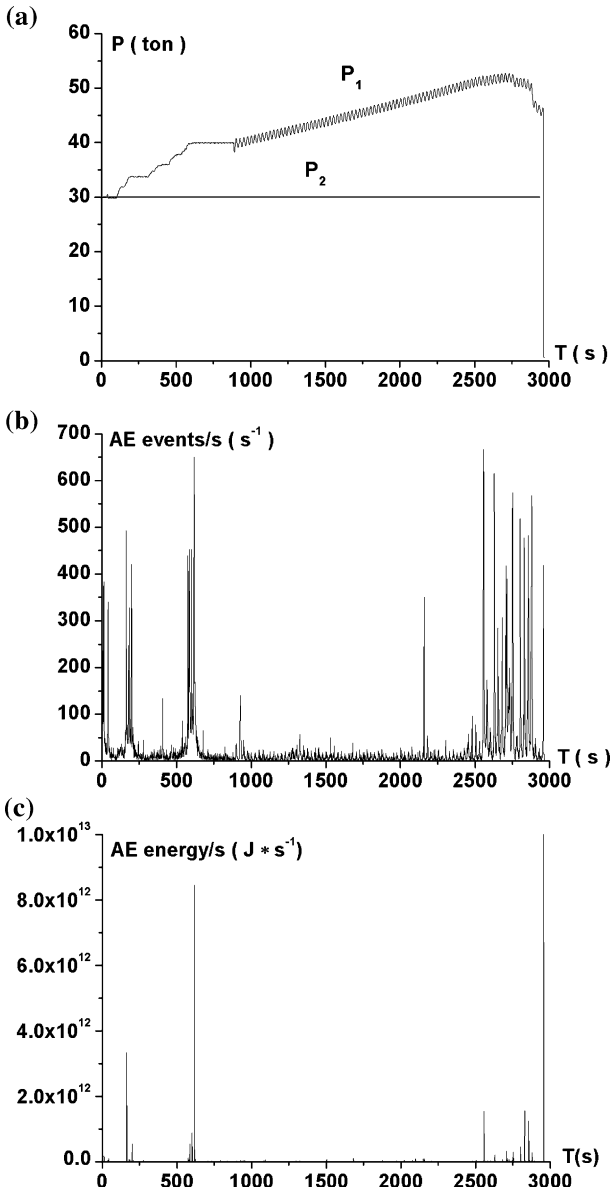


Figure 3

The loading history and the corresponding experimental results for sandstone specimen S1. The average loading velocity is 0.3 μm/s. (a) The loading history in the experiment for sandstone specimen S1. P1 is the axial load. The amplitude and the periodicity of the cyclic period loading cycles is 300 kg and 25 s, respectively. P2 is the lateral load and it is 30 tons during the whole course. (b) The AE event rate versus time for sandstone specimen S1. (c) The AE energy rate versus time for sandstone specimen S1.

nucleation and growth of micro-cracks are represented by AE event. Thus we are able to investigate the nucleation and growth of fracture by acquiring and analyzing AE signals (LOCKNER *et al.*, 1991, 1993; TOMILIN *et al.*, 1994).

Experimental Results

The experiments have been conducted by 20 scientists from three countries (China, Russia and Australia) during 40 days. For each specimen millions of AE events have been recorded, among which thousands of events can be located. Figures 2 (b,c) and Figures 3 (b,c) are the AE records versus time for gneiss specimen G_2 and sandstone S_1, respectively. At the early stage of the loading there are few micro-fractures in the samples so the AE event rate and energy rate are relatively low. With increasing stress, more and larger micro-cracks appear and some micro-cracks coalesce into larger cracks which lead to increased AE event rate and energy rate. Thereafter additional and larger AE events concentrate before the macro-fracture of the samples.

The acquired AE data allow us to examine the details concerning the appearance and localization of micro-cracks. Figure 4 shows the plots of the spatial distribution of AE hypocenters during different stages. In the beginning micro-fractures are roughly uniformly distributed. Afterwards they begin to concentrate and to nucleate, eventually producing a major fault with the increasing pressure, followed by the final catastrophic failure. This phenomenon has also been observed experimentally by many people such as Lockner *et al.* in granite samples undergoing monotonic loading (LOCKNER *et al.*, 1991; LOCKNER 1993). They distinguish three stages in the process: randomly distributed micro-fractures, nucleation and growth of the nucleation sites. Our experiment supports this feature. Besides, other quantities such as LURR and AER have also been obtained.

YIN and YIN (1991), YIN (1993), YIN *et al.* (1994, 1995, 2000, 2002) proposed an approach called the Load-Unload Response Ratio (LURR) to test for the crustal criticality. It is defined as

$$Y = \frac{X_+}{X_-}, \tag{1}$$

where X_+ and X_- are the response rates during loading and unloading according to some measure.

The idea that motivated the LURR earthquake prediction approach is that when a system is in elastic regime, its response to loading is nearly the same as its response to unloading, whereas the response to loading and unloading becomes quite different when the system is in the damage regime. The ratio of the response to loading and unloading can measure quantitatively the damage extent.

It is clear that $Y = 1$ for the elastic regime since $X_+ = X_-$ and $Y > 1$ for the damage regime due to $X_+ > X_-$. The more severely damaged the material, the larger

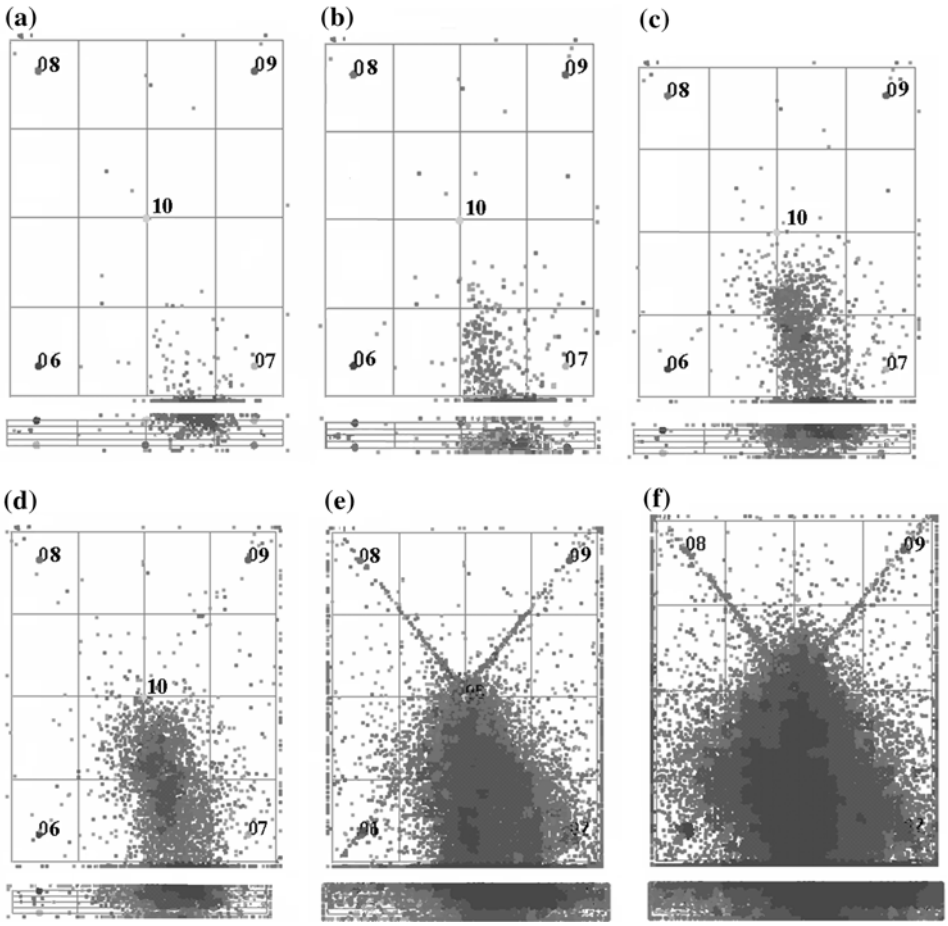


Figure 4
Evolution of damage in terms of the AE hypocenters location with the increasing load. Micro-fractures detected in each interval are represented in plots (a)–(e) (increasing stress). In (f) all the micro-fractures are plotted. The upper picture is seen from the front side and the lower one is seen from the top.

the Y value will become. As the media approach failure the Y value becomes increasingly larger so that the Y value could measure the proximity to failure and also acts as a precursor for earthquake prediction.

In order to predict earthquakes by means of the parameter LURR several main problems need to be solved. One is how to load and unload a block of crust and how to distinguish loading from unloading. The linear dimension of a seismogenic zone may reach hundreds even thousands of kilometers. One of the means to load and unload is by the earth's tide. Tidal force varies periodically, so the induced stresses in the crust continue to loading and unloading it periodically. To distinguish loading from unloading for rock materials in the three-dimensional stress state, we resort to

the Coulomb failure criterion (RESERNBERG and SIMPSON, 1992; HARRIS, 1998). It can be expressed as the following:

$$CFS = \tau_n + f\sigma_n, \tag{2}$$

where τ_n and σ_n are shear and normal stress resolved on the fracture plane. f is the coefficient of inner friction. n is the normal of the fault plane on which the CFS reaches its maximum. ΔCFS is the increment of CFS. If the increment of Coulomb failure stress $\Delta CFS > 0$, it is referred to as loading; otherwise $\Delta CFS < 0$ is referred to as unloading.

When calculating CFS the total stress tensor at every point in the earth's crust, including tectonic stress σ_T and tidal stress σ_t, is used. As to the tectonic stress field, we mainly use the results of ZHONGHUAI XU *et al.* for Chinese Mainland (ZHONGHUAI XU *et al.,* 1995), and Mary Lou Zoback's results for other regions (http://www.world-stress-map.org/). On the basis of Molodensksy-Takeuchi's work, we independently wrote the code of calculating stress tensor at any point in the earth's crust. The shear and normal stresses on the fault plane with normal n can be obtained by stress tensor transform and subsequently the CFS can be calculated easily according to (2).

High LURR values indicate that a region is prepared for a strong earthquake. In previous years, a series of successful intermediate-term predictions have been made for strong earthquakes in China and other countries. In LURR theory, Y is defined directly by the released seismic energy as follows:

$$Y_m = \frac{\left(\sum_{i=1}^{N^+} E_i^m\right)_+}{\left(\sum_{i=1}^{N^-} E_i^m\right)_-}, \tag{3}$$

where E denotes released seismic energy, the sign "+" means loading and "–" means unloading, $m = 0$ or $1/3$ or $1/2$ or $2/3$ or 1. For $m = 1$, E^m is exactly the energy itself; for $m = 1/2$, E^m denotes the Benioff strain; for $m = 1/3, 2/3$, E^m represents the linear scale and area scale of the focal zone, respectively; for $m = 0$, Y is equal to N^+/N^-, where N^+ and N^- denote the number of earthquakes occurring during the loading and unloading periods, respectively. In general m is chosen as $1/2$, which means that Y is determined by Benioff strain during the loading and unloading period. While the LURR reaches a high value several months or years prior to the occurrence of strong earthquakes, on the eve of strong earthquakes the LURR decreases to a low level and then the large event occurs.

According to our understanding, the peak-point of the LURR curve suggests the formation of macro-crack in loaded rock specimens (JAEGER, and COOK, 1979) or the beginning of nucleation of an earthquake. After that the system will evolve into the so called self-driving stage, which means that it will obey its own evolution law

(fracture dynamics) and the system is no longer sensitive to the external disturbance. Therefore the LURR decreases to a low level before the large event occurs. XIA *et al.,* (2002) presented critical sensitivity to describe the damage evolved in materials. The numerical simulation of the damage and fracture process for brittle inhomogeneous material also shows a peak of sensitivity prior to the catastrophic rupture which is similar to the variation of LURR (ZHANG *et al.,* 2004). According to Beeler and Lockner's result the minimum typical duration of earthquake nucleation on the San Andreas fault system is about 1 year (BEELER and LOCKNER, 2003) which is in a good agreement with our results.

In our experiment AE energy is used to calculate LURR. The true AE energy is directly related to the area under the acoustic emission waveform which in turn can be measured by digitizing and integrating the waveform signal. As a simplification, the AE event energy can be approximated as the square of the peak amplitude (LOCKNER *et al*., 1991). The resulting values are actually more representative of the intensity of the event but are commonly referred to as energy calculations in the AE literature. This is due to their approximately linear relationship with energy (the units of this term are given in decibels or dB, which can be defined as 10 times the logarithm, to the base 10, of the ratio of two mean square values of voltage). The main reason to perform this type of "energy" analysis is to accentuate events with abnormally large amplitudes. This type of energy measurement is not an absolute energy, but a relative quantity proportional to the true energy.

The results of LURR in this experiment are shown in Figure 5. The arrows indicate the failure time of the specimens. From Figure 5 we can see that at the beginning of the experiment LURR is low and fluctuates around 1. But prior to the final failure of the specimens, the LURR reaches a high value, then the LURR decreases prior to the occurrence of macro-fracture. The experimental results very favorably coincide with the seismological observations (YIN *et al.* 1994, 1995, 2000, 2002, *2004*)?

AER (Accelerating Energy Release)

Observational studies indicate that large earthquakes are sometimes preceded by phases of accelerated seismic release characterized by cumulative Benioff strain following a power-law time-to-failure relation. BUFE and VARNES (1993) and BUFE *et al.* (1994) found that the clustering of intermediate events before a large shock produces a regional increase in cumulative Benioff strain, $\varepsilon(t)$, which can be fit by a power-law time-to-failure relation of the form

$$\varepsilon(t) = A + B(t_c - t)^z \tag{4}$$

where t_c is the time of the large event, the constant A and B are fit parameters. The exponent z is set to 0.3 according to numerous AER studies, e.g., BUFE and VARNES

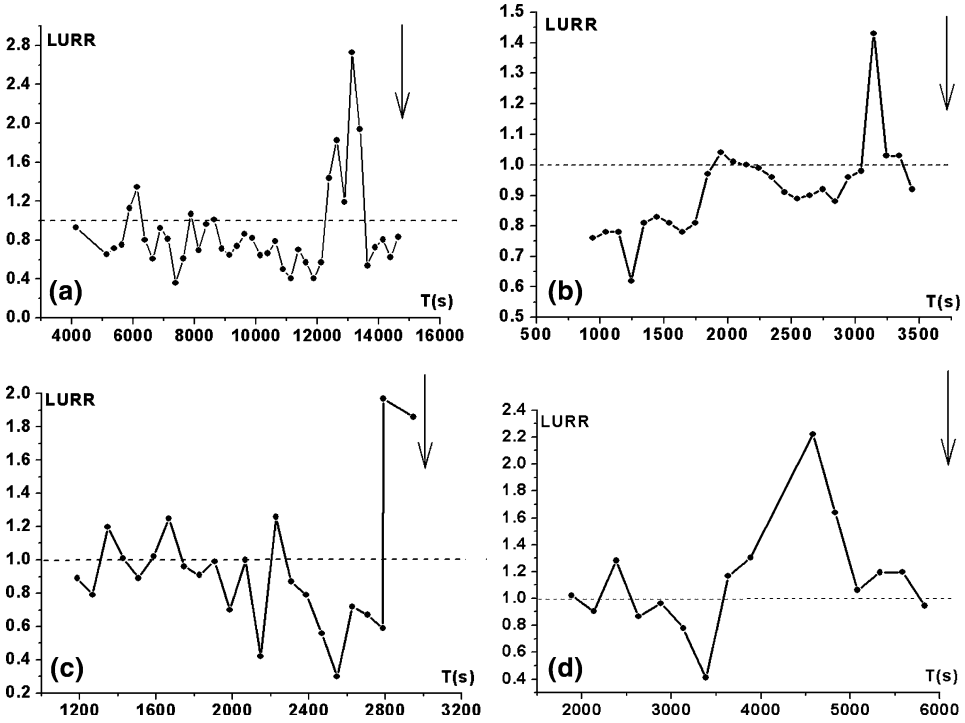


Figure 5
LURR anomaly during rock fracture experiments involving specimens: (a) gneiss G2, (b) gneiss G3, (c) sandstone S1, (d) sandstone S5.

(1993) and it is an exponent that defines the curvature of the power-law acceleration. A is the value of $\varepsilon(t)$ when $t = t_c$, i.e., the final Benioff strain up to and including the largest event. The cumulative Benioff strain at time t is defined as

$$\varepsilon(t) = \sum_{i=1}^{N(t)} E_i(t)^{1/2} \tag{5}$$

where E_i is the energy of the i-th event and $N(t)$ is the number of events at time t.

In order to quantify the degree of acceleration in the seismicity, a curvature parameter C is defined, where

$$C = \frac{\text{power-law fit root-mean-square error}}{\text{linear fit root-mean-square error}}. \tag{6}$$

Therefore when the data are best characterized by a power-law curve, the root-mean-square (RMS) error for the power-law fit will be small compared to the RMS error of the linear fit, C will be small. Conversely, if the seismicity linearly increases then the power-law fit will be statistically indistinguishable from a linear fit, the parameter C will be at or near unity.

The energy of AE event recorded in the experiment represented the elastic energy released by rock specimen. Thus we can investigate the evolution of elastic energy release of brittle materials by analyzing the energy of AE event. Figure 6 shows the experimental results for the specimens and the best fit to equation (4) for each specimen.

In this context the exponent z is in focus of the problem. SORNETTE (1992) found that the mean field value of the exponent z associated with a critical phase transition is $z = 1/2$. RUNDLE *et al.* (2000) used scaling arguments to show that power-law time-to-failure buildup of cumulative Benioff strain may represent the scaling regime of a spinodal phase transition, with an exponent $z = 1/4$. BEN-ZION and LYAKHOVSKY (2002) concluded that $z = 1/3$ for the damage rheology model of LYAKHOVSKY *et al.* (1997). TURCOTTE *et al.* (2002) used the discrete fiber-bundle model to investigate time-dependent failure of chipboard and fiberglass and obtained similar results. BEN-ZION and LYAKHOVSKY (2002) also listed the z value of seismological observations from various authors. The exponents fell in the range of 0.1 to 0.55 with the peak of the distribution at $z = 0.29$. In our experiment the measured z values

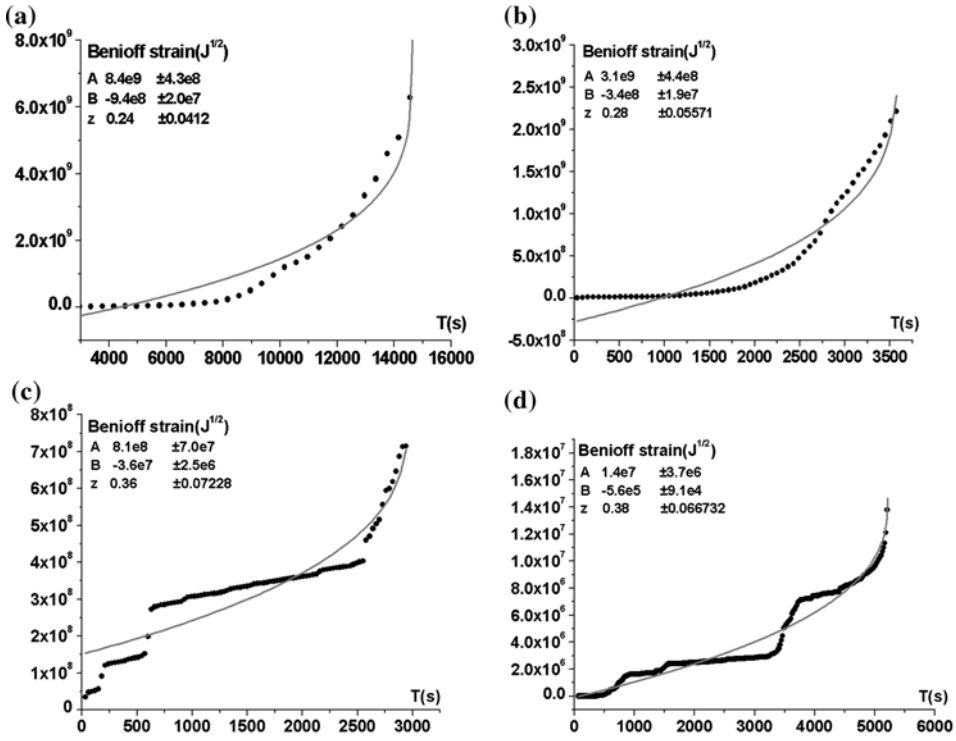


Figure 6
The AER curve prior to rock fracture for specimens: (a) gneiss G2, (b) gneiss G3, (c) sandstone S1, (d) sandstone S5.

are 0.24, 0.28, 0.36 and 0.38 for specimens gneiss G2, G3 and sandstone S1, S5, respectively, which is shown in Figure 6. These values fluctuate around 1/3. Our results are consistent with the results mentioned above.

On the other hand, the curvature parameter C is another focus of the problem. BOWMAN *et al*. (1998) tested the cumulative seismic strain release which increased as a power-law time to failure before large earthquakes in terms of the statistic physics of a critical phase transition. They found C fell in the range of 0.4 to 0.6 within the critical region. YUCANG WANG *et al*. (2003) analyzed the cumulative Benioff strain release preceding seven earthquakes with magnitudes above 5.0 in Australia since 1980. They found $0.45 \leq C \leq 0.65$ with z is about 0.3. In our experiment the curvature parameter C is 0.39, 0.53, 0.49 and 0.56 for specimens gneiss G2, G3 and sandstone S1, S5, respectively. Our results are consistent with the real earthquake results mentioned above.

The scatter in z and C values of our experiments is similar to that seen in observational earthquake data AER studies (BOWMAN *et al*., 1998, YUCANG WANG *et al*., 2003). The results confirm that energy release accelerates prior to macro-failure of the specimens and provide further support for the CPH. With the tidal stress the accelerating energy release is more significant than without that. Nonetheless until now no contrast tests have been conducted and further study is necessary to compare the difference of the behavior of the AER with and without tidal stress.

Correlation between the AE and Load

In our experiment the loading stress follows the formula below:

$$P = kt + A \sin \frac{2\pi}{T} t, \tag{7}$$

where k is rate of tectonic stress σ_{ij}^{T}, A and T are amplitude and period of the sinusoidal stress perturbation to simulating tidal stress.

During the experiment the obvious intense correlation between the AE and load is observed. At the beginning the AE event rate increases gradually with the increase of loading. At the intermediate and later loading stage the AE event rate in the specimen increases with the load increasing and decreases with the load decreasing. Many larger events appear and focus temporarily on the specimen. Figure 7 is the load and AE event rate versus time during the experiment. It is seen that the AE event rate increases when the load increases and decreases when the load decreases. A new parameter called the correlation between the AE and load has been put forward to describe the loading stage.

The correlation coefficient between the AE activity and load is

$$\mathrm{Corr} = \frac{\sum (P_i - \overline{P})(R_i - \overline{R})}{\sqrt{\sum (P_i - \overline{P})^2 \sum (R_i - \overline{R})^2}}, \tag{8}$$

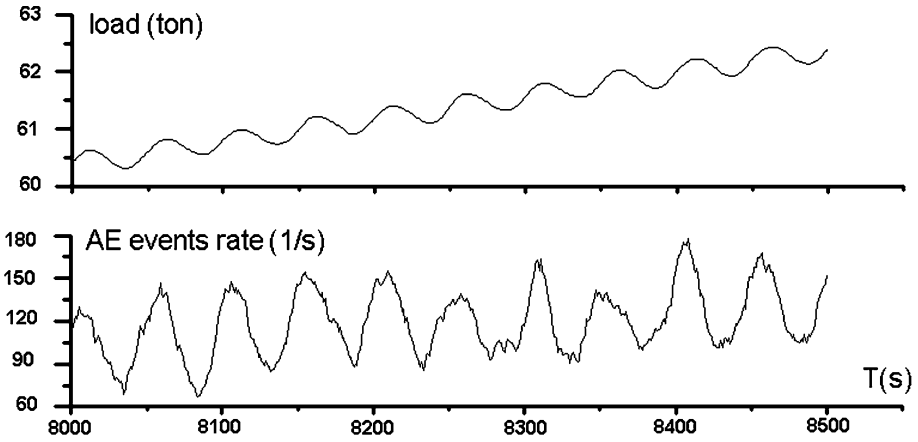


Figure 7
The load and AE events rate versus time during the loading of specimen gneiss G2.

where Corr is the correlation coefficient between P and R, P is the load and R is the AE event rate, $\overline{P}$, $\overline{R}$ is the average of P and R, respectively in the entire time window.

Using formula (8) we can get the correlation coefficient between the AE event rate and load of the specimens at different times. Figure 8 is the correlation between AE event rate and load versus time for specimen gneiss G2, G3 and sandstone S1, S5 respectively. At the outset of loading the Corr is very low. With the increase of the load the Corr also increases and reaches its peak value. However, before the fracture of the specimen the Corr decreases, which is similar to the variation of LURR with time.

Based on this phenomenon we have calculated the correlation between the Coulomb failure stress and the Benioff strain before two strong earthquakes; the results of which are show in Figure 9. The arrows denote the earthquake occurrence. Here the Coulomb failure stress is obtained as in LURR calculation. The results show that on the eve of the earthquakes the Corr falls down, exactly similar with LURR. This could be a new precursor of rock fracture and earthquake prediction.

Conclusions

In our experiment a series of rectangular prisms rock samples of sandstone and gneiss are stressed to failure under a cyclic load in tri-axial compression tests using acoustic emission (AE) monitoring. The experimental results show that high LURR (Load/Unload Response Ratio) and AER (Accelerating Energy Release) occur before the macro fracture of the specimens. These results are interpreted as evidence that brittle failure of heterogeneous media is a CP phenomenon. A new parameter Corr (the correlation between the AE and load) is first proposed to describe the

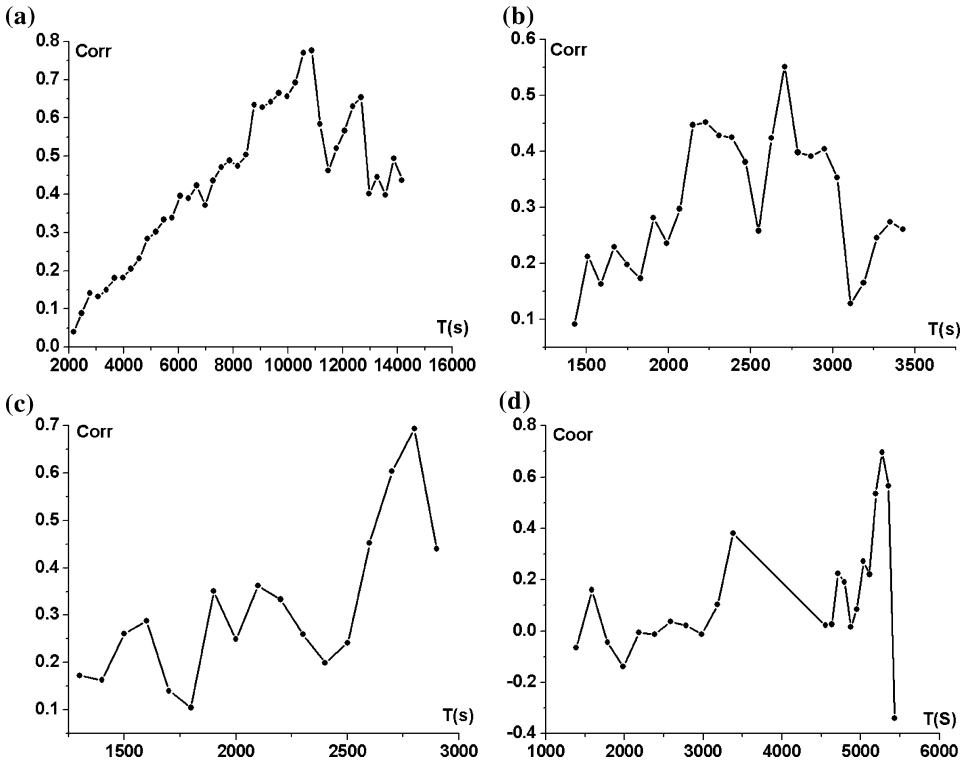


Figure 8
The correlation between the AE and load in rock fracture experiment: (a) gneiss G2, (b) gneiss G3, (c) sandstone S1, (d) sandstone S5.

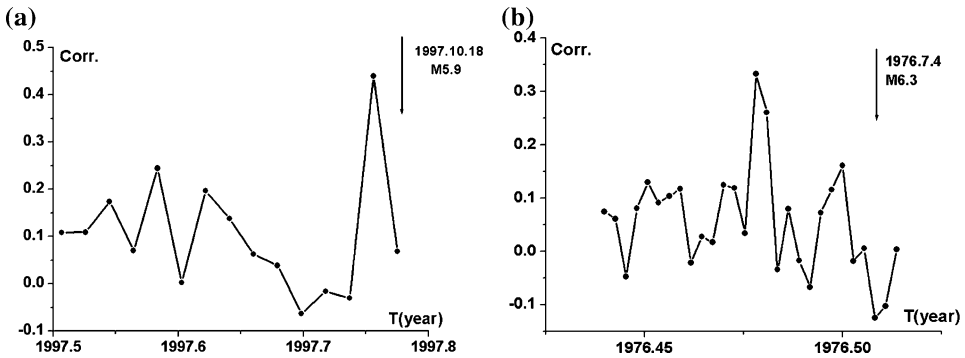


Figure 9
The correlation between the CFS and Benioff strain versus time before two strong earthquakes. (a) 1997.10.18, M 5.9, (39.60 N, 76.90 E), Jiashi Earthquake, Xinjiang Province. (b) 1976.7.4, M 6.3, (24.37 N, 98.50 E), Longling Earthquake, Yunan Province.

evolution of damage for the rock samples loading stage. The correlation between the CFS (Coulomb failure stress) and the Benioff strain before two strong earthquakes investigated reaches its peak point significantly is similar with LURR. This study brings new hopes to rock fracture and earthquake prediction.

Acknowledgments

The authors benefited greatly from anonymous reviewers in revising the manuscript. This research is supported by Natural Sciences Foundation of China (Grant No. 10232050, 10572140), MOST (Ministry of Science and Technology, China, Grant No.2004CB418406 and NO.2002CB412706), the Informatization Construction of Knowledge Innovation Projects of the Chinese Academy of Sciences "Supercomputing Environment Construction and Application" (INF105-SCE-2-02).

References

Beeler, N.M. and Lockner, D.A. (2003), *Why earthquakes correlate weakly with the solid Earth tides: Effects of periodic stress on the rate and probability of earthquake occurrence*, J. Geophys. Res. (in press).

Ben-Zion, Y. and Lyakhovsky, V. (2002), *Accelerated seismic release and related aspects of seismicity patterns on earthquake faults*, Pure Appl. Geophys. *159*, 2385–2412.

Bowman, D.D., Ouillon, G., Sammis, C.G., Sornette, A., and Sornette, D. (1998), *An observational test of the critical earthquake concept*, J. Geophys. Res. *103*, 24359–24372.

Bufe, C.G. and Varnes, D.J. (1993), *Predictive modeling of the seismic cycle of the greater San Francisco Bay region*, J. Geophys. Res. *98, B6*, 9871–9883.

Bufe, C.G., Nishenko, S.P., and Varnes, D.J. (1994), *Seismicity trends and potential for large earthquakes in the Alaska-Aleutian region*, Pure Appl. Geophys. *142*, 83–99.

Ellsworth, W.L., Lindh, A.G., Prescott, W.H., and Head, D.G. (1981), *The 1906 San Francisco earthquake and the seismic cycle*. In *Earthquake Prediction: An International Review (Maurice Ewing Series 4),* Washington, Amer. Geophys. Union, 126–140.

Grasso, J. and Sornette, D. (1998), *Testing self-organized criticality by induced seismicity,* J. Geophys. Res. *103, B12*, 29965–29987.

Harris, R.A. (1998), *Introduction to Special Section: Stress triggering, stress shadows, and implication for seismic hazard,* J. Geophys. Res. *103*, 24347–24358.

Huang, Y., Saleur, H., Sammis, C.G., and Sornette, D. (1998), *Precursors, aftershocks, criticality and self-organized criticality*, Europhys. Lett. *41,* 43–48.

Jaeger, J.C. and Cook, N.G.W. *Fundamentals of Rock Mechanics, 3rd Edition* (Chapman and Hall, 1979).

Keilis-Borok, V. *The lithosphere of the earth as a large nonlinear system*. In *Quo Vadimus: Geophysics for the Next Generation*, Geophys. Monogr. Ser. (eds. G. D. Garland and J. R. Apel) (AGU, Washington, D. C. (1990)) *60*, 81–84.

Knopoff, L., Levshina, T., Keilis-Borok, V.I., and Mattoni, C. (1996), *Increased long-range intermediate-magnitude earthquake activity prior to strong earthquakes in California,* J. Geophys. Res. *101*, 5779–5796.

Lockner, D.A., Byerlee, J.D., Kuksenko, V. *et al.* (1991), *Quasi-static fault growth and shear fracture energy in granite,* Nature *350* (7 MARCH) 39–42.

Lockner, D.A. (1993), *The role of emission in the study of rock failure,* Int. J. Rock Mech. Min. Sci. and Geomech. Asstr. *30* (7), 883–899.

LYAKHOVSKY, V., BEN-ZION, Y., and AGNON, A. (1997), *Distributed damage. faulting and friction*, J. Geophys. Res. *102*, 27,635–27,649.

MORA, P., PLACE, D., ABE, S., and JAUME, S., *Lattice solid simulation of the physics of fault zones and earthquakes: The model, results and directions*. In *Geocomplexity and the Physics of Earthquakes* (eds. Rundle, J.B., Turcotte, D.L. and Klein, W.) (AGU, Washington, DC, 2000), 105–125.

MORA, P. and PLACE, D. (2002), *Stress correlation function evolution in lattice solid elasto-dynamic models of shear and fracture zones and earthquake prediction,* Pure Appl. Geophys. *159*, 2413–2428.

RESERNBERG, P.A. and SIMPSON, R.W. (1992), *Response of regional seismicity to the static stress change produced by the Loma Prieta Earthquake,* Science *255*, 1687–1690.

RUNDLE, J.B. (1988a), *A physical model for earthquakes, 1. Fluctuations and interactions*, J. Geophys. Res. *93*, 6237–6254.

RUNDLE, J.B. (1988b), *A physical model for earthquakes, 2. Application to Southern California,* J. Geophys. Res. *93*, 6255–6274.

RUNDLE, J.B. (1989a), *Derivation of the complete Gutenberg-Richter magnitude frequency relation using the principle of scale invariance*, J. Geophys. Res. *94*, 12,337–12,342.

RUNDLE, J.B. (1989b), *A physical model for earthquakes: 3. Thermodynamical approach and its relation to nonclassical theories of nucleation*, J. Geophys. Res. *94*, 2839–2855.

RUNDLE, J.B., KLEIN, W., and GROSS, S. (1999), *A physical basis for statistical patterns in complex earthquake populations: models, predictions and tests,* Pure Appl. Geophys. *155*, 575–607.

RUNDLE, J.B., KLEIN, W., TIAMPO, K., and GROSS, S. (2000), *Precursory seismic activation and critical-point phenomena,* Pure Appl. Geophys. *157*, 2165–2182.

SAMMIS, C.G. and SMITH, S.W. (1999), *Seismic cycles and the evolution of stress correlation in cellular automation models of finite fault networks*, Pure Appl. Geophys. *155*, 307–334.

SORNETTE, A. and SORNETTE, D. (1990), *Earthquake rupture as a critical point: Consequences for telluric precursors*, Tectonophysics *179*, 327–334.

SORNETTE, D. (1992), *Mean-field solution of a Block-spring Model of Earthquake*, J. Phys. I, France *2*, 2089–2096.

SORNETTE, D. and SAMMIS, C.G. (1995), *Complex critical exponents from renormalization group theory of earthquake prediction*, J. Phys. I, France *5*, 607–619.

TOMILIN, N.G., DAMASKINSKAYA, E.E., and KUKSENKO, V.S. (1994), *Formation of a fracture focus during the deformation of heterogeneous materials (Granite),* Phys. Solid State *36* (10), 1649–1653.

TURCOTTE, D.L., NEWMAN, W.I., and SHCHERBAKOV, R. (2002), *Micro- and macroscopic models of rock fracture*, Geophys. J. Int. *152*, 718–728.

VIDALI, J.E., AGNEW, D.C., JOHNSTON, M.J.S., and OPPENHEIMER, D.H. (1998), *Absence of earthquake correlation with earth tides: An indication of high preseismic fault stress rate*, J. Geophys. Res. *103*, 24567–24572.

WANG, Y.C., YIN, C., MORA, P. YIN, X.C. and PENG, K.Y., *Spatiotemporal scanning and statistical test of the accelerating moment release (AMR) model using Australian earthquake data,* 3rd APEC Cooperation for Earthquake Simulation Workshop Proceedings (eds. A. Donnellan and Peter Mora Goprint, Brisbane, 2003), 265–270.

WEATHERLEY, D., MORA, P., and XIA, M. (2002), *Long-range automaton models of earthquakes: Power-law accelerations, correlation evolution, and mode switching*, Pure Appl. Geophys. *159,* 2469–2490.

XIA, M.F., WEI, Y.J., KE, F.J., and BAI, Y.L. (2002), *Critical sensitivity and transscale fluctuations in catastrophic rupture*, Pure Appl. Geophys. *159*, 2491–2510.

YIN, X.C. and YIN, C. (1991), *The precursor of instability for nonlinear systems and its application to earthquake prediction*, Science in China *34*, 977–986.

YIN, X.C. (1993), *New approach to earthquake prediction*, Preroda (Russia's "Nature"). *1*, 21–27 (in Russian).

YIN, X.C., YIN, C., and CHEN, X.Z. (1994), *The Precursor of instability for nonlinear system and its application to earthquake prediction-the Load-Unload Response Ratio theory*. In *Non-linear Dynamics and Predictability of Geophysical Phenomena* (eds. Neuman, W.I., Gabrelov, A. M., and Turcotte, D.L.) (AGU Geophysical Monograph *83*, 1994), 55–60.

YIN, X.C., CHEN, X.Z., SONG, Z.P., and YIN, C. (1995), *A new approach to earthquake prediction – The Load-Unload Response Ratio (LURR) theory*, Pure Appl. Geophys. *145*, 3/4, 701–715.

YIN, X.C., CHEN, X.Z., SONG, Z.P., and WANG, Y.C. (1996), *The temporal variation of LURR in Kanto and other regions in Japan and its application to earthquake prediction*, Earthq. Res. in China *10*, 381–385.

YIN, X.C., WANG, Y.C., PENG, K.Y., BAI, Y.L., WANG, H., and YIN, X.F. (2000), *Development of a new approach to earthquake prediction: Load/Unload Response Ratio (LURR) theory*, Pure Appl. Geophys. *157*, 1923–1941.

YIN, X.C., MORA, P., PENG, K.Y., WANG, Y.C., and WEATHERLY, D. (2002), *Load-Unload Response Ratio and Accelerating Moment/Energy Release, critical region scaling and earthquake prediction,* Pure Appl. Geophys. *159*, 2511–2524.

ZHONG, HUAI, XU, SU, YUN, WANG, YANXIANG, YU, and AJIA, GAO (1995), *The basic features of the tectonic stress filed in China mainland, The developments of scientific and observational technology in earth and space* (eds. Yuntai Chen) (Earthquake Press), 312–318 (in Chinese).

ZHANG, X.H., XU, X.H., XIA, M.F. and BAI, Y.L. (2004), *Critical sensitivity in driven nonlinear threshold systems*, Pure Appl. Geophys. *161*, 1931–1944.

(Received December 13, 2004, revised August 10, 2005, accepted August 12, 2005)
Published Online First: December 20, 2006

Pure appl. geophys. 163 (2006) 2407–2416
0033–4553/06/122407–10
DOI 10.1007/s00024-006-0134-y

Pure and Applied Geophysics

Stress Shadows Determined from a Phase Dynamical Measure of Historic Seismicity

K.F. TIAMPO,[1] J.B. RUNDLE,[2] and W. KLEIN[3]

Abstract—The Pattern Informatics (PI) technique (TIAMPO *et al*., 2002) is founded on the premise that changes in the seismicity rate are a proxy for changes in the underlying stress. These stress changes are correlated over large spatial regions, and can be quantified using a phase dynamical analysis of the changes in historic seismicity rates. This new approach to the study of seismicity quantifies its local and regional space-time patterns and identifies regions of local quiescence or activation. Here we study those local changes in an attempt to objectively quantify short-term stress shadow regions. We determine that, while persistent stress shadows are detectable with this method, they do not occur uniformly throughout the spatio-temporal region.

Key words: Seismicity, stress shadows, earthquake physics, seismic hazard.

Introduction

Recent work by RUNDLE *et al*. (2000) and TIAMPO *et al*. (2002) produced a new approach to the study of seismicity, the pattern informatics (PI) technique, a methodology for quantifying the spatiotemporal seismicity rate changes in historic seismicity. Recognizing that the seismicity rate is proportional to the stress change rate, and thus systematic space-time variations in seismicity occur in response to changes in the underlying stress field (DIETERICH, 1994; DIETERICH *et al*., 2002; TODA *et al*., 2002), the PI methodology therefore indirectly quantifies the underlying stress change in terms of the change in seismicity rate relative to the background seismicity (TIAMPO *et al*., 2002). Here we employ the PI technique to study the temporal evolution of the seismicity, in particular the growth of stress shadows.

Stress shadows are those regions, on various scales, where a reduced rate of larger earthquakes is hypothesized to occur as a result of the occurrence of a previous large

[1] Department of Earth Sciences, University of Western Ontario, London, ON, Canada. E-mail: ktiampo@uwo.ca
[2] Center for Computational Science and Engineering and Department of Physics, University of California, Davis, CA, U.S.A. E-mail: jbrundle@ucdavis.edu
[3] Department of Physics, Boston University, Boston, MA, U.S.A. E-mail: klein@bphyc.edu

earthquake. Every large earthquake produces a change in the stress field, increasing the stress in certain regions and decreasing the stress in other regions (JAUMÉ and SYKES, 1992; HARRIS and SIMPSON, 1992, 1998; HAUKSSON, 1994; DENG and SYKES, 1997; JONES and HAUKSSON, 1997; KING and BOWMAN, 2003). Those areas where a stress decrease occurs are thought to be regions where fewer events will occur for some time period after the event (HARRIS and SIMPSON, 1992, 1998). The location, size, and duration of these regions, or stress shadows, are a function of the type and size of the main shock.

The location of stress shadows, and their inverse, regions of increased activity due to the stress increase in certain regions adjacent to these large events, has been used to calculate the probability of future events in this adjacent region (STEIN *et al.*, 1997; PARSONS *et al.*, 2000; WIEMER, 2000; TODA and STEIN, 2002; WIENER *et al.*, 2002). In recent years, researchers have searched for confirmation of these stress shadows in the seismicity data using a variety of techniques, with mixed results (STEIN *et al.*, 1992; HARDEBECK *et al.*, 1998; STEIN, 1999; TODA *et al.*, 2002; MARSAN *et al.*, 2003, FELZER *et al.*, 2004). Here we use the PI index to quantify the same phenomena, stress shadows, in California seismicity.

Methodology

In phase dynamical systems, the space-time evolution of the system can be described by changes in the amplitude and phase angle of a state vector, or phase function (MORI and KURAMOTO, 1998). Computer simulations strongly suggest that earthquake seismicity can be described by phase dynamics, in which the important changes in seismicity are associated only with rotations, or changes in phase angle, of the vector function that describes the seismicity in a high-dimensional correlation space (RUNDLE *et al.*, 2000; TIAMPO *et al.*, 2002). Here, space-time pattern structures are non-local in character, a consequence of the strong correlations in the underlying dynamics, where non-local means that events widely separated in space and time may be correlated with each other.

The PI procedure we apply here is based upon the idea that seismic activity is characterized by a vector or phase function, $\hat{\mathbf{S}}$. Changes in the state of seismicity in space and time, or the change in stress rate, are characterized by changes in $\hat{\mathbf{S}}$. It should be noted that we have performed the same calculation for seismic moment, and the results are similar, although the assignment of moment to the events adds another source of error to the calculation. As a result, we have confined our recent work to quantifying the numbers of events.

Define the seismic activity rate $\psi_{\text{obs}}(x_i, t)$ as the number of earthquakes per unit time, of any size, within the box centered at x_i at time t. We define the time-averaged seismicity function $\mathbf{S}(x_i, t_0, t)$ over the interval $(t - t_0)$:

$$\mathbf{S}(x_i, t_0, t) = \frac{1}{(t - t_0)} \int_{t_0}^{t} \psi_{\text{obs}}(x_i, t) dt. \tag{1}$$

Since there are N locations for $\mathbf{S}(x_i, t_0, t)$, and we assume t_0 to be a fixed time, $\mathbf{S}(x_i, t_0, t)$ is the ith component of a general, time-dependent vector evolving in an N-dimensional space. Denoting spatial averages over the N boxes by $\langle\rangle$, the phase function $\hat{\mathbf{S}}(x_i, t_0, t)$ is then defined to be the mean-zero, unit-norm function obtained from $\mathbf{S}(x_i, t_0, t)$:

$$\hat{\mathbf{S}}(x_i, t_0, t) = \frac{\mathbf{S}(x_i, t_0, t) - \langle \mathbf{S}(x_i, t_0, t) \rangle}{\|\mathbf{S}(x_i, t_0, t)\|}, \tag{2}$$

where $\|\mathbf{S}(x_i, t_0, t)\|$ is the square root of the variance (the standard deviation) over all spatial boxes.

Under the assumption of phase dynamics, the important changes in seismicity will be given by $\Delta\hat{\mathbf{S}}(x_i, t_1, t_2) = \Delta\hat{\mathbf{S}}(x_i, t_0, t_2) - \Delta\hat{\mathbf{S}}(x_i, t_0, t_1)$, a rotation of the N-dimensional unit vector $\hat{\mathbf{S}}(x_i, t_0, t)$ in time. The change in probability at x_i, $\Delta\mathbf{P}(x_i, t_1, t_2)$, from time t_1 to t_2, is then the PI index,

$$\Delta P(x_i, t_1, t_2) = \{\Delta\hat{\mathbf{S}}(x_i, t_1, t_2)\}^2 - \mu_P, \tag{3}$$

where μ_P is the spatial mean of $\{\Delta\hat{\mathbf{S}}(x_i, t_1, t_2)\}^2$. Note that rotations in $\hat{\mathbf{S}}(x_i, t_0, t)$ correspond to a change in the local rate of seismicity, or number of events, normalized to a regional background rate, resulting in an objective measure of the changing seismicity rate and, as a result, the changing stress state (DIETERICH, 1994; TIAMPO *et al.*, 2002).

Figure 1a shows a plot of $\Delta\hat{\mathbf{S}}$ for the years 1978 to 1991. The primary seismicity data set used to calculate $\Delta\hat{\mathbf{S}}$ in this figure is a compilation catalog from 1932 through the end of 1991, obtained from the Southern California Earthquake Center (SCEC) database (available at http://www.scecdc.scec.org.) and the Northern California Seismic Network (NCSN, http://quake.geo.berkeley.edu/ncsn/), with all blast events removed. A positive (red) value represents anomalous seismic activity, and a negative (blue) value represents anomalous quiescence. Figure 1b shows the corresponding PI index, $\Delta\mathbf{P}(x_i, 1978, 1991)$, where the color scale is now logarithmic, with the number denoting the exponent to the base 10. Note the correspondence between hotspot areas and large events that occurs subsequent to 1991 (shown as black circles). In the following section, we compute $\Delta\hat{\mathbf{S}}$, a measure of the rate of seismicity change relative to the regional background rate, to look for evidence of stress shadows.

Results

While the occurrence of stress shadows subsequent to large events has long been hypothesized, their detection has been difficult (FELZER, 2004). While a number of

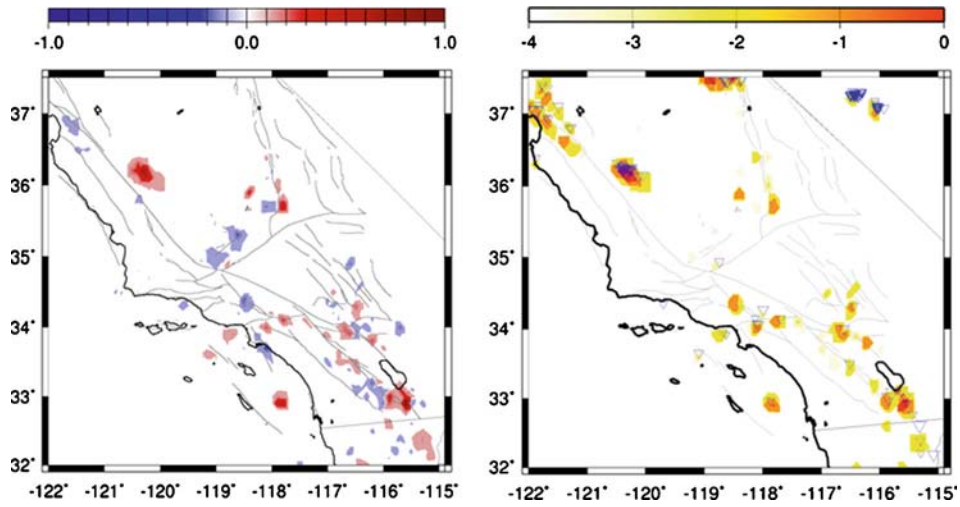


Figure 1

a) Map of $\Delta\mathbf{S}$, normalized to the maximum absolute value, for the time period 1991–1978. The color scale is linear, blue to white to red; and b) color contour plot of PI index, $\Delta\mathbf{P}$, for locations at which $\Delta\mathbf{P} > 0$, 1991–1978. Color scale is logarithmic, where the number represents the exponent. Inverted triangles are events that occurred from 1978–1991 for events of $5 < m$ (triangle size scales with magnitude). Circles are events that occurred from 1992–present, again for $5 < m$ (circle size also scales with magnitude).

researchers have detected reduced rates of triggered events (STEIN *et al.*, 1992; STEIN, 1999; TODA *et al.*, 2002), others have found little or no evidence for such a phenomena (MARSAN *et al.*, 2003; HARDEBECK, 2004). This discrepancy is generally due in part to the slight variation in rate changes after an earthquake. As noted in HARDEBECK *et al.* (1998), sixty percent of aftershocks occur in regions where there is a stress increase related to a large event, and therefore forty percent of all aftershocks occur in areas designated as stress shadows. This results in only a small net change in the number of events between a region of stress increase and one of stress decrease. In addition, the various algorithms for declustering catalogs revise the statistics in such a way that it is difficult to quantify rate changes in an objective way (FELZER, 2004). And finally, the particular choice of the region size and location results in widely varying values for the background rate to test against in order to detect rate changes.

On the other hand, the PI index, and the quantity $\Delta\hat{\mathbf{S}}$ that is used to calculate this index, is an *objective* measure of seismicity rate changes. No declustering algorithm is employed on the catalog, and the data are normalized by the regional spatial variance in seismicity rate; a constant that does not depend on region size or location.

The primary seismicity data set used to calculate $\Delta\hat{\mathbf{S}}$ for central and southern California employed in this analysis is, as in Figure 1, a compilation catalog obtained from SCEC and NCSN, in this case through the end of 1994. Relevant data include location, in latitude and longitude, and the time the event occurred. Seismic events between $-115°$ and $-125°$ longitude and 32° and 40° latitude were selected, and

events of all quality were acquired. The seismicity was binned into squares of 0.1° latitude and 0.1° longitude to a side, approximately 11 km, and a magnitude cut of 3.0 was applied to the data. $\Delta\hat{\mathbf{S}}$ was then calculated for periods of only two months, in order to evaluate the change in seismicity rate for short periods of time subsequent to large events. Note also that, for each event, a subset of the catalog was used that ended with the time period shown on each plot, so that the value of $\Delta\hat{\mathbf{S}}$ is calculated only through that month.

Figure 2a shows $\Delta\hat{\mathbf{S}}$ for the region surrounding the Superstition Hills event of November 1987, a predominantly strike-slip event. Here the blue region denotes anomalous quiescence, or a stress shadow region. Note the consistent spatial pattern that starts in March of the following year, and continues through December of 1988, outlining a region of stress rate decrease, or a stress shadow.

Figure 2b shows $\Delta\hat{\mathbf{S}}$ for a region that displays the region surrounding the Northridge earthquake of 1994, a blind thrust event. In this case, no stress shadow region is detected using the PI method.

Three additional events are shown in Figure 3. At the top is shown $\Delta\hat{\mathbf{S}}$ for the San Fernando event of 1971. Here the stress shadow to the southeast of the rupture zone shows a consistent spatial pattern for ten months after the main shock. The stress shadow associated with the central region of the Loma Prieta earthquake that occurred in October, 1989, on the other hand, does not occur immediately, fades in and out over a six-month period, and has no consistent spatial pattern. Finally, the Landers sequence has a consistent stress shadow to the south, in one of the regions identified by STEIN (1999).

One interesting result is that the PI index does not, in any of these cases, identify stress shadow/stress triggering regions that exactly conform to the stress change lobes predicted from a standard Coulomb stress change analysis (STEIN *et al.*, 1992; STEIN, 1999; KING and BOWMAN, 2003). This result leads to the conclusion that the local stress field is substantially more complicated than this simple model, and is modified by such effects as the heterogeneous local fault regime and the interaction of long-and short-range effects such as pore water pressure or dynamic triggering. This observation is supported again by the fact that forty percent of aftershocks occur in the stress shadow lobes (HARDEBECK *et al.*, 1998), and supports the conclusion of others that it is difficult to objectively observe stress shadows (MARSAN *et al.*, 2003; FELZER *et al.*, 2004).

Using the PI index (equation (3), above), it is possible to calculate the relative change in probability for each time period. Again, the time period for this calculation is two months, so that the change in probability calculated from the PI index is for the subsequent two months. In this case, we calculate the PI index for the Superstition Hills series of events, the Loma Prieta earthquake, and the Landers sequence. Again, the scale logarithmic, as in Figure 1, but the time periods over which the PI index is calculated correspond to those in Figures 2 and 3 above, two months. Individual snapshots throughout the aftershock period are shown and labeled in Figure 5.

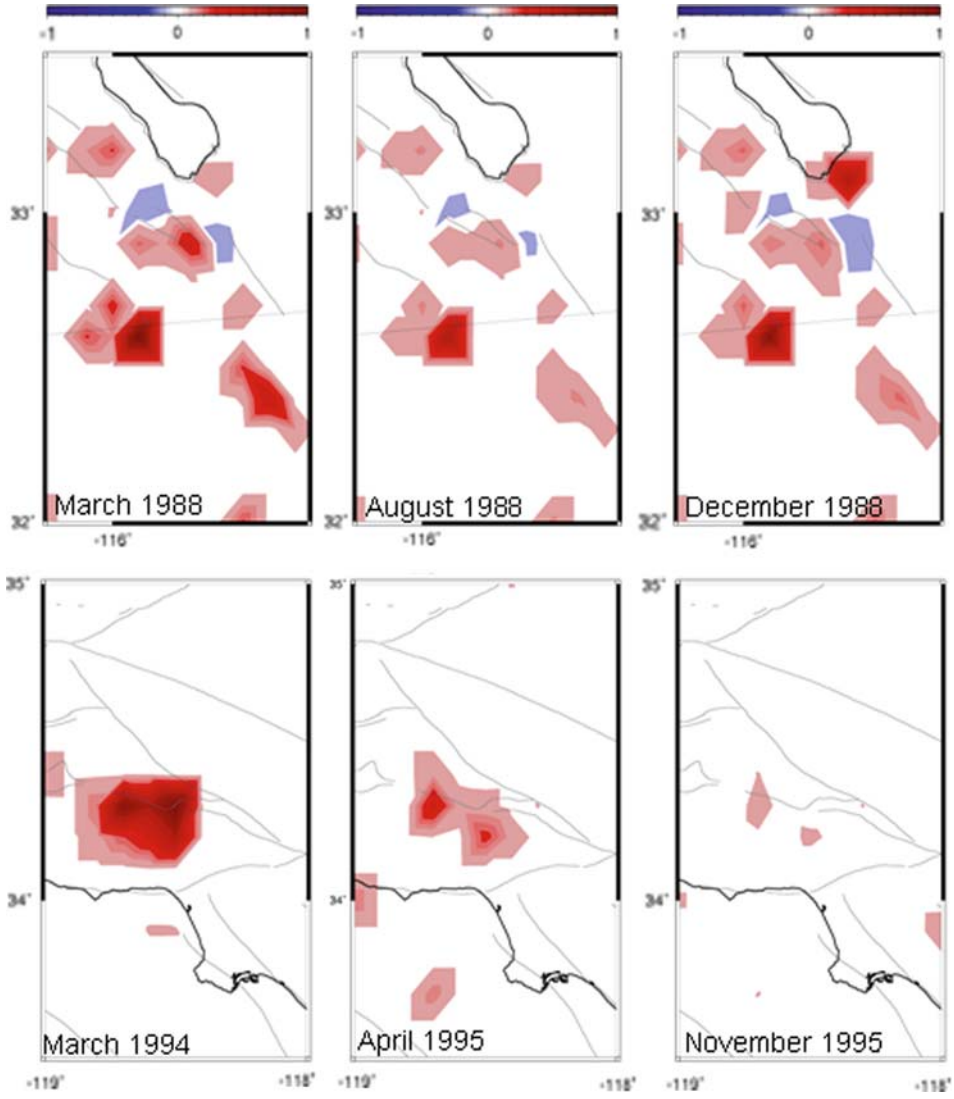
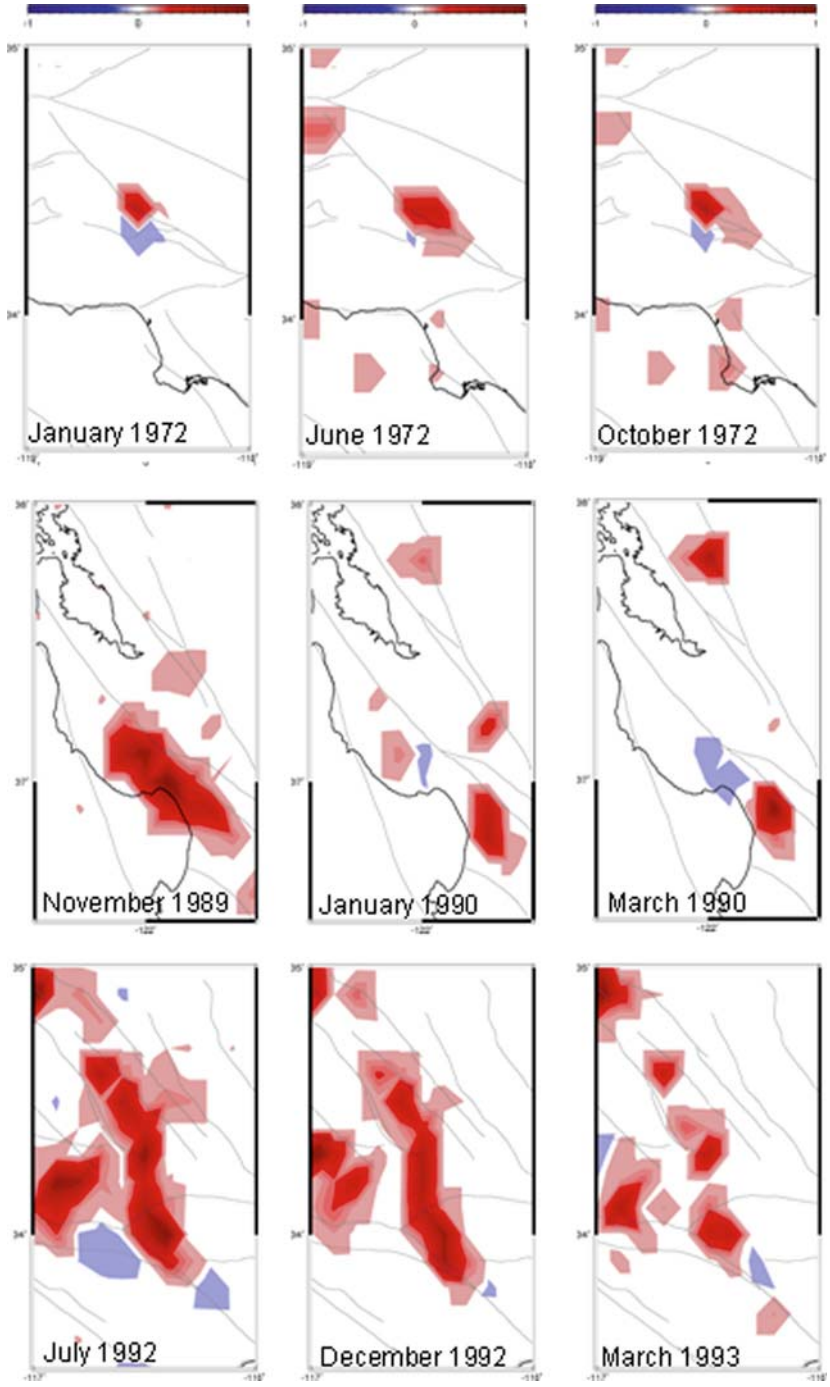


Figure 2
a) $\Delta\hat{\mathbf{S}}$ in the region of the Superstition Hills earthquake that occurred in November of 1987. Shown is the relatively continuous spatial pattern that is first detected in the spring of 1988 and ends in December of that year and; b) $\Delta\hat{\mathbf{S}}$ in the region of the Northridge earthquake of January, 1994. Color scale is as shown in Figure 1a.

Figure 3
At the top is shown $\Delta\hat{\mathbf{S}}$ for in the region of the San Fernando earthquake of February, 1971. In the middle is shown the same calculation, but for the Loma Prieta earthquake of October, 1989. The lower plots show the same calculation for the June, 1992 Landers sequence. Again, the color scale is as in Figure 1a.


January 1972
June 1972
October 1972
November 1989
January 1990
March 1990
July 1992
December 1992
March 1993

Note that, in every case, there is a persistent increase in probability of a large aftershock in most of the region surrounding the main shock, as would be expected in the aftershock regions to large events. However, those smaller regions which show persistent stress shadows do not have as great a probability of a large aftershock in the next few months, although there is still a positive change in probability, consistent with the stress shadow hypothesis. This suggests that short-term probability forecasts are possible using the PI method.

Conclusions

The PI index is a method for determining regions of anomalous seismic moment release, and $\Delta\hat{\mathbf{S}}$, a measure of relative changes in seismicity rate and the relevant quantity used to calculate the PI index, quantifies significant seismic rate changes relative to the region background rate. Here we have demonstrated that this

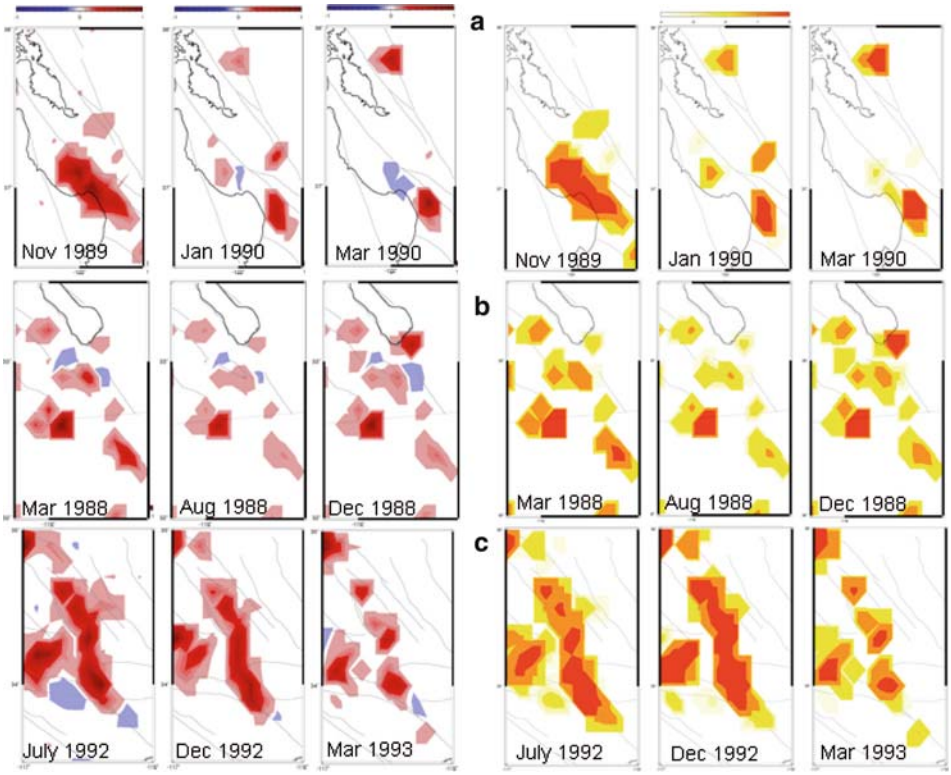


Figure 4

$\Delta\hat{\mathbf{S}}$ (left) and the PI index (right) calculated for a) the Loma Prieta earthquake of 1989; b) the Superstition Hills events of 1987, and c) the Landers sequence of 1992. Color scales are the same as that shown in Figure 1, but the time periods over which the PI index is calculated correspond to those in Figures 2 and 3 above, two months. Snapshots for individual months throughout the aftershock period are shown here.

technique can be used to locate regions of stress decrease, or stress shadows, and that these regions can be detected following certain large events in California, including the San Fernando, Superstition Hills, Loma Prieta, and Landers events. In addition, while stress shadows do occur subsequent to certain earthquakes, not all large events produce detectable stress shadows, most notably in the case of thrust events such as the 1994 Northridge earthquake (Fig. 3), and the Coalinga event of 1983, not shown. Further work should be performed to determine if this lack of stress shadow is a persistent effect, or if it is the result of a depth component not included in this analysis.

In addition, the PI index can be used to identify regions with varying degrees of increased probability of large aftershocks for the purposes of short-term hazard analysis, as shown in Figure 5.

Finally, it appears that while, in general, those regions of persistent seismic quiescence, or stress shadows, tend to be smaller than those predicted by the stress change calculations, areas of seismic activation tend to be larger than those predicted from the same calculations. Future work will include the quantification of both regions of stress shadow and stress increase in an attempt to provide insight into both the stress state of the crust, and a better measure of time-dependent seismic hazard. This work will necessitate the calculation of the specific Coulomb stress change for each of these large events, and testing of the statistical significance of these regions against those changes.

Acknowledgments

The authors would like to thank both three anonymous reviewers and Dr. Dion Weatherley for helpful comments that improved this manuscript significantly. This research was supported by the Southern California Earthquake Center. SCEC is funded by NSF Cooperative Agreement EAR-8920136 and USGS Cooperative Agreements 14-08-0001-A0899 and 1434-HQ-97AG01718. The SCEC contribution number for this paper is 812. Research by JBR was funded by grant DE-FG02-04ER15568 from the US Department of Energy, Office of Basic Energy Sciences to the University of California, Davis. Research by WK was supported by USDOE/OBES grant DE-FG02-95ER14498 and W-7405-ENG-6 at LANL. WK would also like to acknowledge the hospitality and support of CNLS at LANL. Images were plotted with the help of GMT software developed by P. Wessel and W.H.F. Smith.

References

DENG, J. and SYKES, L.R. (1997), *Evolution of the stress field in Southern California and triggering of moderate-size earthquakes: A 200-year perspective*, J. Geophys. Res. *102*, 9859–9886.

DIETERICH, J. (1994), *A constitutive law for rate of earthquake production and its application to earthquake clustering*, J. Geophys. Res. *99*, 2601–2618.

DIETERICH, J.H., CAYOL, V., and OKUBO, P. (2002), *The use of earthquake rate changes as a stress meter at Kilauea volcano*, Nature *408*, 457–460.

FELZER, K., BRODSKY, E., and ABERCROMBIE, R. (2004), *The absence of stress shadows*, Spring Meeting, Seismol. Soc. of America.

HARDEBECK, J. L., NAZARETH, J. J., and HAUKSSON, E. (1998), *The static stress change triggering model: Constraints from two Southern California aftershock sequences*, J. Geophys. Res. *103*, 24,427–24,438.

HARDEBECK, J.L. (2004), *Stress triggering and earthquake probability estimates,* J. Geophys. Res. *109*, 2354, doi: 10.1029/2003JB002437.

HARRIS, R. A. and SIMPSON, R.W. (1992), *Changes in static stress on Southern California faults after the 1992 Landers earthquake*, Nature *360*, 251–254.

HARRIS, R. A. and SIMPSON, R.W. (1998), *Suppression of large earthquakes by stress shadows: A comparison of Coulomb and rate-and-state failure,* J. Geophys. Res. *103*, 24,439–24,451.

HAUKSSON, E. (1994), *State of stress from focal mechanisms before and after the 1992 Landers earthquake sequence,* Bull. Seismol. Soc. Am. *84*, 917–934.

JAUMÉ, S.C. and SYKES, L.R. (1992), *Evolving towards a critical point: A review of accelerating seismic moment/energy release prior to large and great earthquakes,* Pure Appl. Geophys. *155*, 279–306.

JONES, L.M. and HAUKSSON, E. (1997), *The seismic cycle in Southern California: Precursor or response?* Geophys. Res. Lett. *24*, 469–472.

KING, G.C.P. and BOWMAN, D.D. (2003), *The evolution of regional seismicity between large earthquakes,* J. Geophys. Res. *108*, doi:10.1029/2001JB000783.

MARSAN, D. and BEAN, C.J. (2003), *Seismicity response to stress perturbations, analysed for a world-wide catalogue*, Geophys. J. Int. *158*, 179–195.

MORI, H. and KURAMOTO, Y., *Dissipative Structures and Chaos* (Springer-Verlag, Berlin 1998).

PARSONS, T., TODA, S., and STEIN, R.S. (2000), *Heightened odds of large earthquakes near Istanbul: An interaction-based probability calculation*, Science *288*, 661–665.

RUNDLE, J.B., KLEIN, W., TIAMPO, K.F., and GROSS, S. (2000), *Linear pattern dynamics of nonlinear threshold systems*, Phys. Rev. E *61*, 2418–2432.

STEIN, R.S. (1999), *The role of stress transfer in earthquake occurrence*, Nature *402*, 605–609.

STEIN, R.S., BARKA, A.A., and DIETERICH, J.H. (1997), *Progressive failure on the North Anatolian fault since 1939 by earthquake stress triggering*, Geophys. J. Int. *128*, 594–604.

STEIN, R.S., KING, G.C.P., and LIN, J. (1992), *Change in failure stress on the southern San Andreas fault system caused by the magnitude = 7.4 Landers earthquake*, Science *258*, 1328–1332.

TIAMPO, K.F., RUNDLE, J.B., MCGINNIS, S., GROSS, S., and KLEIN, W. (2002), *Mean-field threshold systems and phase dynamics: An application to earthquake fault systems*, Europhys. Lett. *60*, 481–487.

TODA, S., STEIN, R.S., and SAGIYA, T. (2002), *Evidence from the AD 2000 Izu Islands earthquake swarm that stressing rate governs seismicity*, Nature *419*, 58–61.

WIEMER, S. (2000), *Introducing probabilistic aftershock hazard mapping*, Geophys. Res. Lett. *27*, 3405–3408.

WIEMER, S., GERSTENBERGER, M., and HAUKSSON, E. (2002), *Properties of the aftershock sequence of the 1999 M_w 7.1 Hector Mine earthquake: Implications for aftershock hazard*, Bull. Seismol. Soc. Am. *92*, 1227–1240.

(Received December 30, 2004, revised August 26, 2005, accepted August 31, 2005)
Published Online First: December 20, 2006

Pure appl. geophys. 163 (2006) 2417–2432
0033–4553/06/122417–16
DOI 10.1007/s00024-006-0130-2

Pure and Applied Geophysics

Pattern Informatics and its Application for Optimal Forecasting of Large Earthquakes in Japan

K.Z. NANJO,[1,2,4] J.B. RUNDLE,[2] J.R. HOLLIDAY,[2] and D.L. TURCOTTE[3]

Abstract—Pattern Informatics (PI) technique can be used to detect precursory seismic activation or quiescence and make an earthquake forecast. Here we apply the PI method for optimal forecasting of large earthquakes in Japan, using the data catalogue maintained by the Japan Meteorological Agency. The PI method is tested to forecast large (magnitude $m \geq 5$) earthquakes spanning the time period 1995–2004 in the Kobe region. Visual inspection and statistical testing show that the optimized PI method has forecasting skill, relative to the seismic intensity data often used as a standard null hypothesis. Moreover, we find in a retrospective forecast that the 1995 Kobe earthquake ($m = 7.2$) falls in a seismically anomalous area. Another approach to test the forecasting algorithm is to create a future potential map for large ($m \geq 5$) earthquake events. This is illustrated using the Kobe and Tokyo regions for the forecast period 2000–2009. Based on the resulting Kobe map we point out several forecasted areas: The epicentral area of the 1995 Kobe earthquake, the Wakayama area, the Mie area, and the Aichi area. The Tokyo forecast map was created prior to the occurrence of the Oct. 23, 2004 Niigata earthquake ($m = 6.8$) and the principal aftershocks with $5.0 \leq m$. We find that these events were close to in a forecasted area on the Tokyo map. The PI technique for regional seismicity observation substantiates an example showing considerable promise as an intermediate-term earthquake forecasting in Japan.

Key words: Pattern informatics, earthquake, forecast, seismicity, 1995 Kobe earthquake, 2004 Niigata earthquake.

1. Introduction

Earthquakes have great scientific, societal, and economic significance. The 17 Jan. 1995 Kobe, Japan, earthquake (hereinafter referred as the 1995 Kobe earthquake), which was only a magnitude $m = 7.2$, event, killed nearly 6,000 persons and produced an estimated $200 billion loss. Similar scenarios are possible at any time in

[1] The Institute of Statistical Mathematics, Minato-ku, Tokyo 106-8569, Japan.
[2] Center for Computational Science and Engineering, c/o Department of Physics, University of California at Davis, One Shields Avenue, Davis, CA 95616, USA.
E-mail: nanjo@cse.ucdavis.edu; jbrundle@ucdavis.edu; holliday@physics.ucdavis.edu
[3] Department of Geology, University of California at Davis, One Shields Avenue, Davis, CA 95616, USA. E-mail: turcotte@geology.ucdavis.edu
[4] Swiss Seismological Service, Institute of Geophysics, ETH Hoenggerberg, HPP P, 8093, Zürich, Switzerland. E-mail: kazuyoshi.nanjo@sed.ethz.ch

San Francisco, Seattle, and other U.S. urban centers along the Pacific plate boundary. The magnitude of potential loss of life and property is so great that reliable earthquake forecasting should be at the forefront of research goals, especially in Japan.

Millions of dollars and thousands of work years have been spent on observational programs searching for reliable precursory phenomena. Possible precursory phenomena include changes in seismicity, changes in seismic velocity, tilt and strain precursors, electromagnetic signals, hydrologic phenomena, and chemical emissions (TURCOTTE, 1991; SCHOLZ, 2002). For example, TSUNOGAI and WAKITA (1995, 1996) found that the ion concentrations of ground water issuing from deep wells located near the epicenter of the 1995 Kobe earthquake showed precursory phenomena (see also JOHANSEN *et al.,* 1996). A few successes have been reported, but, to date, no precursors to large earthquake have been detected that would provide reliable forecasts (Nature Debates, denoted on earthquake forecasting, http://www.nature.-com/nature/debates/earthquake/, 1999).

The earth's crust is clearly extremely complex and it is generally accepted that earthquakes are a chaotic phenomenon. Thus, as in the case of weather forecasting, earthquake forecasting must be considered on a statistical basis (RUNDLE *et al.*, 2003). A fundamental question is whether the statistical properties of seismicity patterns can be used to forecast future earthquakes. Premonitory seismicity patterns were found for some strong earthquakes in California and Nevada using algorithm "CN" and for $m > 8$ worldwide using algorithm "M8" (e.g., KEILIS-BOROK, 1990; KEILIS-BOROK and ROTWAIN, 1990; KEILIS-BOROK and KOSSOBOKOV, 1990; KEILIS-BOROK and SOLOVIEV, 2003).

Alternatively, a new approach to earthquake forecasting, the pattern informatics (PI) approach, has been proposed by RUNDLE *et al.* (2002), TIAMPO *et al.* (2002a,b,c), and HOLLIDAY *et al.* (2006). This approach is based on the strong space-time correlations that are responsible for the cooperative behavior of driven threshold systems and arises both from threshold dynamics as well as from the mean field (long range) nature of the interactions. The PI technique can be used to detect precursory seismic activation or quiescence and make earthquake forecasts. Applications to earthquake data from Southern California show that the PI method is a powerful technique for forecasting large events, but no one has attempted to apply this method to Japanese earthquakes. Moreover, it will be interesting to test whether the 1995 Kobe earthquake could have been forecasted using the method.

The purpose of this paper is to study the applicability of the pattern informatics (PI) algorithm for forecasting large earthquakes in Japan. As an example, we will present a forecast of large ($m \geq 5$) earthquakes during the time period 1995–2004 in the Kobe region: The region that includes the epicenter of the 1995 Kobe earthquake. First, we will briefly introduce the PI method. Next, we will describe the earthquake catalogues used in this paper. Carrying out visual inspections and statistical testing,

we will find that the method can have more precision in forecasting large earthquakes than a simple look at where earthquakes have occurred in the past. Then, we will find in a retrospective forecast that the 1995 Kobe earthquake falls into a seismically anomalous area.

The true test of the forecasting algorithm is to make a future forecast that proves correct. TIAMPO *et al.* (2002b), RUNDLE *et al.* (2002, 2003), and HOLLIDAY *et al.* (2006) did this for the period of 2000–2009 in Southern California. In this paper we will also make a future forecast extending from 2000 to 2009 for two regions: The Kobe region and the Tokyo region in which the Tokyo Metropolitan area is included. This paper will show that the PI method has considerable promise as an intermediate-term event forecasting tool for Japanese earthquakes.

2. Pattern Informatics

The PI approach is a six-step process that creates a time-dependent system state vector in a real valued Hilbert space and uses the phase angle to predict future states (RUNDLE *et al.*, 2003). The method is based on the idea that the future time evolution of sesmicity can be described by pure phase dynamics (MORI and KURAMOTO, 1998; RUNDLE *et al.*, 2000a,b). This PI method was originally developed by TIAMPO *et al.* (2002b). HOLLIDAY *et al.* (2006) showed that the modified PI method recently proposed by KLEIN (2006) has better forecasting skills than the original method. Here we use the modified PI method for forecasting large earthquakes in Japan.

First the study area is divided into N square boxes. The center of the i-th box is denoted by $\vec{x}_i$ and each box has an edge length Δx . The seismic intensity in box i is defined to be the total number of earthquakes $n\left(\vec{x}_i, t_b, t\right)$ in the box during the period from the base time t_b to time t ($> t_b$) with magnitude greater than m_c. For each box an activity rate function $s\left(\vec{x}_i, t_b, t\right)$ is defined to be the average rate of the occurrence of earthquakes in box i during the period t_b to t. That is,

$$s\left(\vec{x}_i, t_b, t\right) = \frac{n\left(\vec{x}_i, t_b, t\right)}{t - t_b}. \tag{1}$$

Values of the average rate of occurrence of earthquakes in box i are obtained by taking the t_b values at daily intervals from t_0 to $t-1$. These values are averaged to give its mean value

$$s^*\left(\vec{x}_i, t_0, t\right) = \frac{1}{t - 1 - t_0} \sum_{t_b=t_0}^{t-1} s\left(\vec{x}_i, t_b, t\right). \tag{2}$$

The normalized mean value $\hat{s}^*\left(\vec{x}_i, t_b, t\right)$ is found by subtracting the spatial mean for all boxes and dividing by the spatial standard deviation

$$\hat{s}^*\left(\vec{x}_i, t_b, t\right) = \frac{s^*\left(\vec{x}_i, t_0, t\right) - \frac{1}{N}\sum_{j=1}^{N} s^*\left(\vec{x}_j, t_0, t\right)}{\sqrt{\frac{1}{N}\sum_{j=1}^{N}\left[s^*\left(\vec{x}_j, t_0, t\right) - \frac{1}{N}\sum_{k=1}^{N} s^*\left(\vec{x}_k, t_0, t\right)\right]^2}}. \tag{3}$$

The change in the normalized mean value is found by subtracting the normalized mean value for the time period t_0 to t_1 ($> t_0$) from the normalized mean value for the time period t_0 to t_2 ($> t_1$)

$$\Delta\hat{s}^*\left(\vec{x}_i, t_0, t_1, t_2\right) = \hat{s}^*\left(\vec{x}_i, t_0, t_2\right) - \hat{s}^*\left(\vec{x}_i, t_0, t_1\right). \tag{4}$$

Finally we introduce a probability of change of activity $P\left(\vec{x}_i, t_0, t_1, t_2\right)$ in a box. This is related to the square of $\Delta\hat{s}^*\left(\vec{x}_i, t_0, t_1, t_2\right)$ by

$$P\left(\vec{x}_i, t_0, t_1, t_2\right) = \left[\Delta\hat{s}^*\left(\vec{x}_i, t_0, t_1, t_2\right)\right]^2. \tag{5}$$

Because the $\Delta\hat{s}^*\left(\vec{x}_i, t_0, t_1, t_2\right)$ is squared, the probability is a measure of both seismic activation and seismic quiescence.

Schematically, this whole process can be represented by

$$n \rightarrow s \rightarrow s^* \rightarrow \hat{s}^* \rightarrow \Delta\hat{s}^* \rightarrow P, \tag{6}$$

where the symbol $\hat{}$ means "apply normalization", the symbol Δ means "calculate the change in rate", and the symbol * means "average over base times".

We are interested in seismic activation and seismic quiescence relative to the background; the new probability function $P'\left(\vec{x}_i, t_0, t_1, t_2\right)$ is defined by the difference between $P\left(\vec{x}_i, t_0, t_1, t_2\right)$ and its spatial mean

$$P'\left(\vec{x}_i, t_0, t_1, t_2\right) = P\left(\vec{x}_i, t_0, t_1, t_2\right) - \frac{1}{N}\sum_{j=1}^{N} P\left(\vec{x}_j, t_0, t_1, t_2\right). \tag{7}$$

The use of $P'\left(\vec{x}_i, t_0, t_1, t_2\right)$ to forecast earthquakes is referred to as pattern informatics. Forecasts should convey information for time t in the range: $t_3 > t > t_2$. We call the time interval $t_2 - t_1$ the "change interval" and the interval $t_3 - t_2$ the "forecast interval". In this paper, to improve performance, the PI method is optimized by adjusting the length of the change interval ($t_2 - t_1$) and the initial time t_0.

3. Application of PI Method

We use a seismic catalog maintained by the Japan Meteorological Agency (JMA). This catalog includes the data of earthquakes with $m \geq 0$ for the time period since 1923 in and around Japan. The relevant data consist of time, magnitude, and

location given by east longitude, north latitude, and depth. To ensure the completeness of the earthquake catalog, events of magnitude equal to or larger than the lower cut-off magnitude $m_c = 3$ with depth shallower than 20 km dated from Jan. 1, 1955 to the present (as of Mar. 14, 2004) are selected.

To apply the PI method to Japanese earthquakes we need to determine Δx. Following TIAMPO *et al.* (2002b) we use $\Delta x = 0.1^\circ \approx 11$ km. Boxes of this size correspond roughly to the linear scale size of $m \approx 5$–6 earthquakes. We try to forecast earthquakes of $m \geq 5$. The idea is to use information on small events having spatial scales $\lambda < \Delta x$ to forecast the occurrence of large events having scales $\lambda > \Delta x$.

We are also interested in a retrospective forecast of the 1995 Kobe earthquake ($m = 7.2$). The epicenter was 135.03° east longitude and 34.58° north latitude with the depth of 16 km. This earthquake produced a surface rupture with a length of about 9 km appearing along the pre-existing right lateral Nojima fault. Six aftershocks registering magnitude $m \geq 5$ followed the main shock along this fault.

To examine whether our forecasting capability is changed if the size of our study regions is changed, we consider 5 study regions: The extent of these regions is changed but they have the same center which is near the epicenter of the 1995 Kobe earthquake. Table 1 summarizes the study regions. Regions 4 and 5 include the epicenter of the Oct. 6, 2000 Tottori earthquake of $m = 7.3$ (hereinafter referred to as the 2000 Tottori earthquake). The epicenter was 133.35° east longitude and 35.27° north latitude with a depth of 9 km.

Before discussing the PI method forecast for the Kobe region during the period 1995–2004 we describe earthquake data from the Tokyo region (No. 6 in Table 1). The analyzed earthquake data are from the region between 136.0–142.0° east longitude and between 33.0–38.0° north latitude (depth shallower than 20 km). The

Table 1

Shown are the study regions used for forecasting large earthquakes in the Kobe and Tokyo regions and the statistical results. The regions are numbered. 0.1° corresponds roughly to 11 km. The number of 0.1° × 0.1° × 20 km boxes is given in the column "N". The log-likelihoods computed using the global Gaussian model and local Poissonian model for the PI method forecast are given in the columns $\log_{10}(L_G)$ and $\log_{10}(L_P)$, respectively. Similarly, the log-likelihoods computed using the global Gaussian model and local Poissonian model for the seismic intensity data (null hypothesis) are given in the columns $\log_{10}(L_{GN})$ and $\log_{10}(L_{PN})$, respectively

Region	No.	North lat. (deg.)	East long. (deg.)	Depth (km)	N	$\mathrm{Log}_{10}(L_G)$	$\mathrm{Log}_{10}(L_P)$	$\mathrm{Log}_{10}(L_{GN})$	$\mathrm{Log}_{10}(L_{PN})$
	1	34.1–35.2	134.2–135.8	20	176	−17.0	−32.4	−21.2	−49.6
Kobe	2	33.7–35.6	133.8–136.2	20	456	−27.2	−40.0	−31.7	−71.5
	3	33.3–36.0	133.4–136.6	20	864	−34.3	−47.4	−39.8	−81.2
	4	32.9–36.4	133.0–137.0	20	1500	−47.0	−64.7	−58.9	−108.7
	5	32.5–36.8	132.6–137.4	20	2064	−48.4	−67.0	−63.2	−114.2
Tokyo	6	33.0–38.0	136.0–142.0	20	3000	–	–	–	–

Tokyo Metropolitan area is located at the center of region 6. Eighty-two earthquakes measuring $m \leq 5$ occurred from Jan. 1, 2000 to the present (Mar. 14, 2004). An earthquake swarm associated with Miyake volcano started on Jun. 26, 2000 (hereinafter referred to as the 2000 Miyake earthquake swarm). Since then there have been 74 earthquakes with $5.0 \leq m < 6.0$ and six earthquakes with $6.0 \leq m$. This swarm is located at about 33.8–34.3° north latitude and 139.0–139.5° east longitude.

4. Forecasting Large Earthquakes in the Kobe Region during the Period 1995–2004

4.1. Visual Inspection

The PI approach to forecasting large earthquakes in the Kobe region during the period Jan. 1, 1995 to Mar. 14, 2004 is best illustrated using a specific example. A

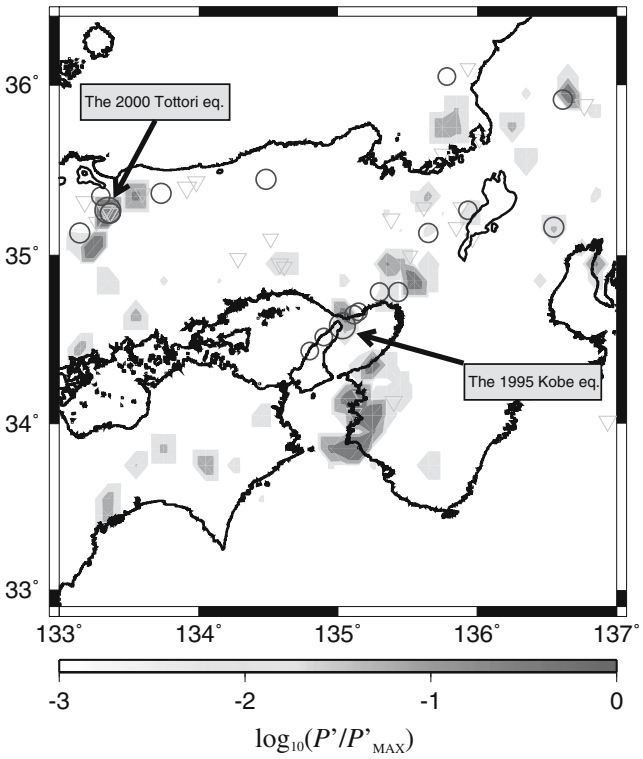


Figure 1
The optimized PI method forecast for the Kobe region (No. 4 in Table 1) for the period 1995–2004. Relative probabilities $\log_{10}\left\{P'\left(\vec{x}_i, t_0, t_1, t_2\right)/P'_{\mathrm{MAX}}\right\}$ are given using the color code. The times used are t_0 = Jan. 1, 1960, t_1 = Jan. 1, 1968, and t_2 = Dec. 31, 1994. Earthquakes with $5.0 \leq m$ that took place during 1968–1994 are shown as inverted triangles. Circles represent events with $5.0 \leq m$ during the time period 1995–2004.

PI-method optimal forecast of earthquake occurrence in region 4 is given in Figure 1. In applying the method to the $N = 1500$ $0.1° \times 0.1° \times 20$ km boxes with $m_c = 3$, the times used are $t_0 =$ Jan. 1, 1960, $t_1 =$ Jan. 1, 1968, $t_2 =$ Dec. 31, 1994, and $t_3 =$ Mar. 14, 2004. The t_2 value is just before the occurrence of the 1995 Kobe earthquake. In region 4 eighteen earthquakes with $m \leq 5$ occurred from 1995–2004.

Relative values of the probability of activity are given in the form $\log_{10}\left\{P'\left(\vec{x}_i, t_0, t_1, t_2\right) / P'_{\mathrm{MAX}}\right\}$ where P'_{MAX} is the maximum in the probabilities $P'\left(\vec{x}_i, t_0, t_1, t_2\right)$. The colour coded anomalies are shown in Figure 1. Note that only positive values of $\log_{10}\left\{P'\left(\vec{x}_i, t_0, t_1, t_2\right)/P'_{\mathrm{MAX}}\right\}$ are given. Thus the color-coded regions represent regions of anomalously high seismic activation or high seismic quiescence. The colour-coded anomalies are associated with large ($m \leq 5$) earthquakes for both current (triangles, $t_1 < t < t_2$) and future (circles, $t_2 < t < t_3$) time periods in this figure. Fifteen of eighteen future earthquakes occur either on areas of forecasted anomalous activity or within the margin of error of 11 km (the coarse-grained box size). Note that the 1995 Kobe earthquake (123.0° east longitude, 34.6° north latitude) and the 2000 Tottori earthquake (133.4° east longitude, 35.3° north latitude) fall into warmer colored anomalies.

Figure 2 shows seismic intensity $n(\vec{x}_i, t_0, t_2)/n_{\mathrm{MAX}}$ using the data from t_0 to t_2, where n_{MAX} is the largest value of $n(\vec{x}_i, t_0, t_2)$. The large ($m \leq 5$) earthquakes in the change interval (triangles, $t_1 < t < t_2$) and forecast interval (circles, $t_2 < t < t_3$) are included in this figure. Comparing Figures 1 and 2 shows that the optimized PI method narrows the possible locations where large earthquakes are expected.

We also carry out visual inspections for the other regions (Nos. 1, 2, 3, and 5 in Table 1). We take $t_0 =$ Jan. 1, 1960, $t_1 =$ Jan. 1, 1968, $t_2 =$ Dec. 31, 1994, and $t_3 =$ Mar. 14, 2004 as done for region 4. We find that the 1995 Kobe earthquake again falls into one of the color-coded areas. Moreover, for region 5 the 2000 Tottori earthquake also falls in a color-coded area. These visual inspections substantiate that the 1995 Kobe earthquake falls in the forecast area.

4.2. Statistical Testing

Visual inspection of Figure 1 shows that the retrospective forecast is reasonably successful, nonetheless rigorous statistical testing is needed. For a null hypothesis we use the actual seismic intensity data in the time period from t_0 to t_2 $\left\{n\left(\vec{x}_i, t_0, t_2\right)/n_{\mathrm{MAX}}\right\}$ as a probability density. The use of seismic intensity data has been proposed for the standard null hypothesis (KAGAN and JACKSON, 2000). This hypothesis was used for testing the PI forecast for Southern California earthquakes (TIAMPO *et al.*, 2002b).

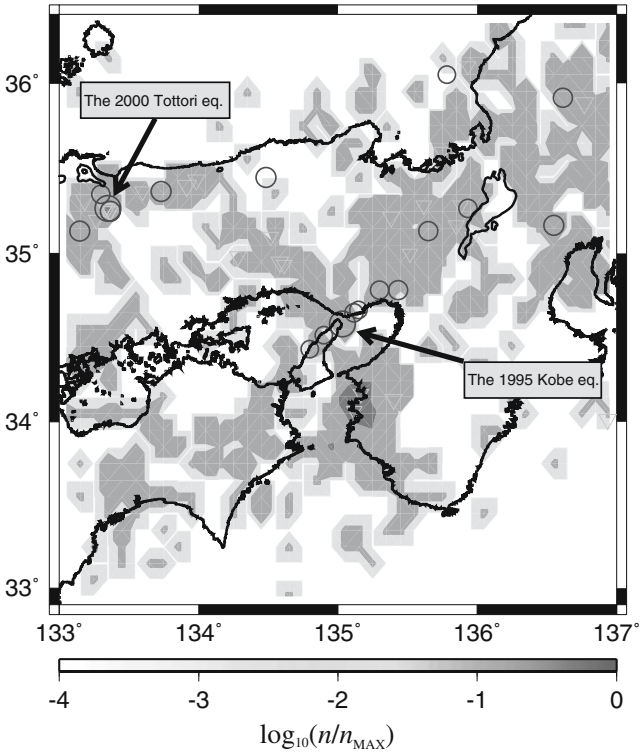


Figure 2

Relative seismic intensities $\log_{10}\left\{n\left(\vec{x}_i, t_0, t_2\right)/n_{\mathrm{MAX}}\right\}$ for the Kobe region (No. 4 in Table 1) for the period from t_0 = Jan. 1, 1960 to t_2 = Dec. 31, 1994 (triangles: earthquakes with $5.0 \leq m$ in 1968–1994; Circles: earthquakes with $5.0 \leq m$ in 1995–2004).

To test the PI method forecast we utilize the maximum likelihood test. This is accepted as the standard approach of testing earthquake forecasts (e.g., BEVINGTON and ROBINSON, 1992; GROSS and RUNDLE, 1998; KAGAN and JACKSON, 2000; SCHORLEMMER *et al.*, 2006). This test is used to evaluate the accuracy with which probability measure $P\left(\vec{x}_i, t_0, t_1, t_2\right)$ can forecast future ($t_2 < t < t_3$) large ($m \leq 5$) events, relative to a forecast from the null hypothesis. The likelihood L is a probability measure that can be used to assess the utility of one forecast measure over another. Typically, one computes the logarithm of the likelihood $\log_{10}(L)$ for the proposed measure L and compares that to the likelihood measure L_N for a representative null hypothesis. The ratio of these two values then yields information about which measure is more accurate in forecasting future events.

In the likelihood test, a probability density function is required. Following HOLLIDAY *et al.* (2006), we use a global Gaussian model and a local Poissonian

model in our study. The use of the global Gaussian model was proposed for testing the PI method forecast (TIAMPO *et al.*, 2002b). The second model used is based on work performed by the Regional Earthquake Likelihood Methods (RELM) group (SCHORLEMMER *et al.*, 2006). The likelihood values for the PI method forecast are defined as L_G for the global Gaussian model and L_P for the local Poissonian model. Similarly, the likelihoods for the null hypothesis are defined as L_{GN} for the global Gaussian model and L_{PN} for the local Poissonian model.

We first take the global Gaussian model. Then we compute the log-likelihood $\log_{10}(L_G)$ for the forecast of Figure 1 and the log-likelihood $\log_{10}(L_{GN})$ for the seismic intensity map in Figure 2. The results are summarized in the row of No. 4 in Table 1. The computed log-likelihood $\log_{10}(L_G) = -47.0$ is larger than the log-likelihood $\log_{10}(L_{GN}) = -58.9$. Next we take the local Poissonian model. As done for the global Gaussian model, the log-likelihoods are computed for the forecast of Figure 1, $\log_{10}(L_P) = -64.7$, and for the seismic intensity map in Figure 2, $\log_{10}(L_{PN}) = -108.7$ (Table 1). The value of $\log_{10}(L_P)$ is larger than that of $\log_{10}(L_{PN})$. Since larger values of the log-likelihoods indicate a more successful hypothesis, the logical conclusion is that the optimized PI method has better forecast skill than the actual seismic intensity data.

We also carry out statistical testing for the other regions (Nos. 1, 2, 3, and 5 in Table 1). We take t_0 = Jan. 1, 1960, t_1 = Jan. 1, 1968, t_2 = Dec. 31, 1994, and t_3 = Mar. 14, 2004 as done for the visual inspections. We first take the global Gaussian model. For all regions we find that the log-likelihood $\log_{10}(L_G)$ is larger than that $\log_{10}(L_{GN})$. We next take the local Poissonian model. We again find that the log-likelihood $\log_{10}(L_P)$ is larger than that $\log_{10}(L_{PN})$ for all regions. These results support our conclusion that the method has forecasting skill.

5. *Forecasting Large Earthquakes for the Period 2000–2009 in the Kobe and Tokyo Regions*

TIAMPO *et al.* (2002b) and RUNDLE *et al.* (2002, 2003) proposed that the true test of any forecasting algorithm is to make a future forecast that proves correct. The diffusive, mean field nature of the dynamics (FISHER *et al.*, 1997; KLEIN *et al.*, 1997; RUNDLE *et al.*, 2000b) leads to a hypothesis that forecasts should convey information for time t approximately in the range: $t_3\{= t_2 + (t_2 - t_1)\} > t > t_2$. According to this hypothesis, previous studies (RUNDLE *et al.*, 2003; TIAMPO *et al.*, 2002b; HOLLIDAY *et al.*, 2006) took the length of the change interval to be equal to the length of the forecast interval for future large event forecast during the period of 2000–2009 in Southern California. Here we also do this for a future forecast in the period from 2000 to 2009 ($t_3 - t_2$ = 10 years) for the Tokyo and Kobe regions. For both regions we assume that $t_3 - t_2 = t_2 - t_1$ = 10 years. That is, the times used are t_1 = Jan. 1, 1990 and t_2 = Dec. 31, 1999. We optimize the initial time t_0 for

forecasting large earthquakes from Jan. 1, 2000 to Mar. 14, 2004 by using statistical likelihood testing.

5.1. Kobe Region

We first consider the Kobe region (No. 4 in Table 1) where 6 earthquakes with $m \leq 5$ occurred in the period 2000–2004. Statistical testing using local Gaussian and local Poissonian models indicates that the time t_0 is optimized to be $t_0 = 1980$ (years) for region 4. Our forecast map is shown in Figure 3. We use times $t_1 =$ Jan. 1, 1990 and $t_2 =$ Dec. 31, 1999 with the optimized initial time $t_0 =$ Jan. 1, 1980. Values of $\log_{10}\left\{P'\left(\vec{x}_i, t_0, t_1, t_2\right)/P'_{\mathrm{MAX}}\right\}$ are given using the same color code as in Figure 1. Inverted triangles are events, $m \geq 5$, during 1989–1999. Events of $m \geq 5$ that occurred between Jan. 1 2000 to Mar. 14, 2004 are plotted with circles. The 2000 Tottori

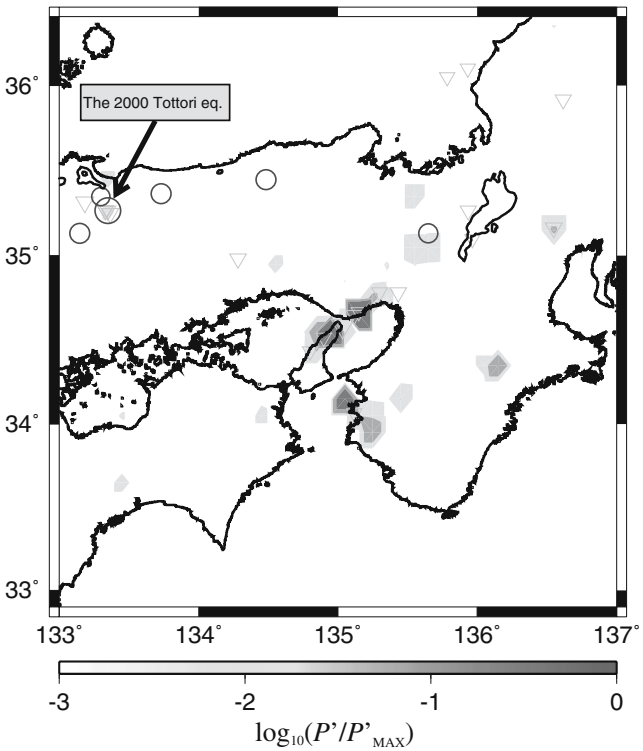


Figure 3

Pattern informatics method forecast for the Kobe region (No. 4 in Table 1) during the period 2000–2009. Relative probabilities $\log_{10}\left\{P'\left(\vec{x}_i, t_0, t_1, t_2\right)/P'_{\mathrm{MAX}}\right\}$ are given using color code. The times used are also $t_0 =$ Jan. 1, 1980, $t_1 =$ Jan. 1, 1990, and $t_2 =$ Dec. 31, 1999. Circles are events that occurred from 2000 to 2004 for $5.0 \leq m$. Inverted triangles are events that occurred from 1990 to 1999 for again $5.0 \leq m$.

earthquake (133.4° east longitude, 35.3° north latitude) again falls into one of the forecasted areas. This figure calls attention to regions that seem to be at risk for larger earthquakes having $m \geq 5$ during 2000–2009. The regions most at risk (orange and red colored) include the epicentral area of the 1995 Kobe earthquake approximately located at 135.0° east longitude and 34.5° north latitude, the Wakayama area approximately located at 135.1° east longitude and 34.0° north latitude, the Mie area approximately located at 136.1° east longitude and 34.3° north latitude, and the Aichi area approximately located at 136.5° east longitude and 35.2° north latitude.

5.2. Tokyo Region

Next the PI method is applied for forecasting future large earthquakes in the Tokyo region. Since the Sep. 1, 1923 Kanto earthquake ($m = 7.9$) that killed more

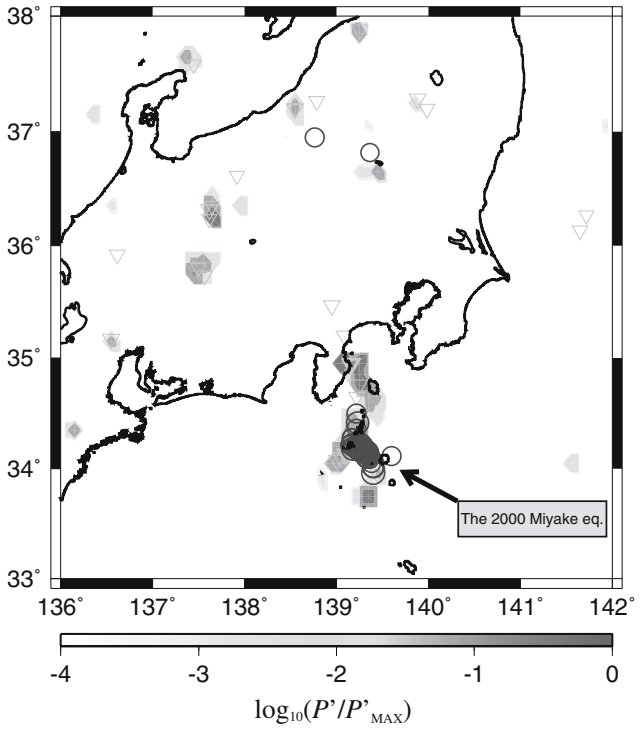


Figure 4
Pattern informatics method forecast for the Tokyo region (No. 6 in Table 1) for the period 2000–2009. Relative probabilities $\log_{10}\left\{P'\left(\vec{x}_i, t_0, t_1, t_2\right)/P'_{\mathrm{MAX}}\right\}$ are given using color code. The times used are t_0 = Jan. 1, 1965, t_1 = Jan. 1, 1990, and t_2 = Dec. 31, 1999. Inverted triangles and circles represent events with $5.0 \leq m$ that have occurred during the time periods 1990–1999 and 2000–2004, respectively.

than 140,000 people, the Tokyo region has experienced a seismically quiet period (e.g., RIKITAKE, 1990). However, earthquakes that could inflict disaster on the Tokyo region were pointed out (e.g., MOGI, 1985). Moreover, using the exponential probability distribution model for recurrence times, FERRAES (2003) estimated that a damaging earthquake ($m \leq 6.4$) may occur before the year June 2009. In these circumstances, it will be interesting to make the PI hazard map for forecasting large ($m \leq 5$) future (2000–2009) earthquakes for the Tokyo region.

As was done for the Kobe region, we assume times t_1 = Jan. 1, 1990 and t_2 = Dec. 31, 1999 and utilize initial time t_0 that is optimized for the forecast of large ($m \leq 5$) large quakes spanning Jan. 1, 2000 to the present (Mar. 14, 2004) by the statistical tests. The optimized value is t_0 = Jan. 1, 1965. The forecast map is shown in Figure 4. Values of $\log_{10}\left\{P'\left(\vec{x}_i, t_0, t_1, t_2\right)/P'_{\mathrm{MAX}}\right\}$ are given using the color code as in Figure 2. Earthquakes with $m \geq 5$ in the change and forecast intervals are represented by inverted triangles and circles, respectively. The 2000 Miyake

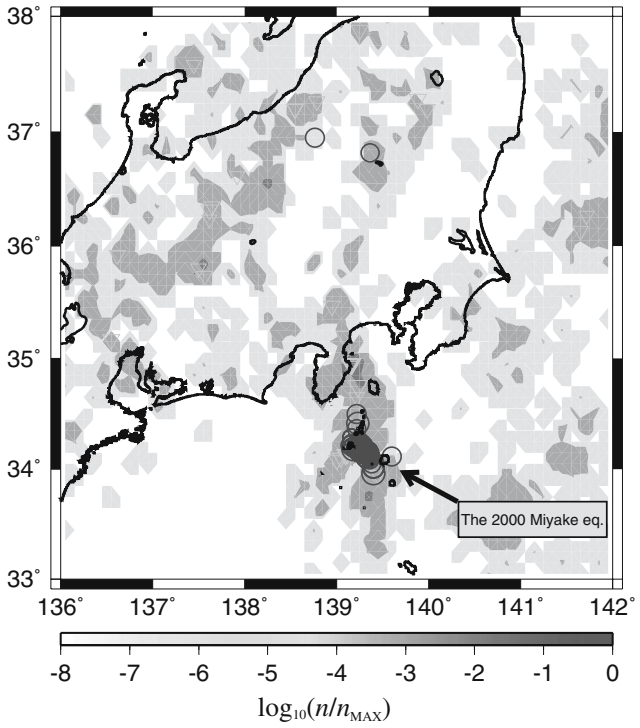


Figure 5

Relative seismic intensities $\log_{10}\left\{n\left(\vec{x}_i, t_0, t_2\right)/n_{\mathrm{MAX}}\right\}$ for the Tokyo region (No. 6 in Table 1) for t_0 = Jan. 1, 1965 to t_2 = Dec. 31, 1999. Earthquakes with $5.0 \leq m$ that took place during 1990–1999 are shown as inverted triangles. Circles represent events with $5.0 \leq m$ that have occurred during the period 2000–2004.

earthquake swarm falls in color-coded anomalies in Figure 4. For comparison, the spatial distribution of relative seismic intensities for region 6 for the time period 1965 to 1999 is given in Figure 5. The relative intensity is defined as the ratio $n\left(\vec{x}_i, t_0, t_2\right)/n_{\mathrm{MAX}}$ where t_0 = Jan. 1, 1965 and t_2 = Dec. 31, 1999.

The forecast map in Figure 4 was first presented by one of the authors (JBR) at a lecture on Oct. 13, 2004 at Kyoto University, Japan (Organizer: Professor James Mori, Kyoto University) and at the International Conference on Geodynamics held on Oct. 14–16, 2004 at the University of Tokyo, Japan (Organizers: Professor Mitsuhiro Matsu'ura and professor Teruyuki kato, University of Tokyo). After this appearance, the Oct. 23, 2004 Niigata earthquake of $m = 6.8$ (hereinafter referred to as the 2004 Niigata earthquake) occurred. The epicenter was 138.7° east longitude and 37.3° north latitude and the depth was 16 km. Large aftershocks with $5.0 \le m$ occurred around the hypocenter of this earthquake. Figure 6 is the same as Figure 4 but these events are added in Figure 6. Note that the 2004 Niigata earthquake is very close to a color-coded anomaly and some aftershocks fall into the anomaly,

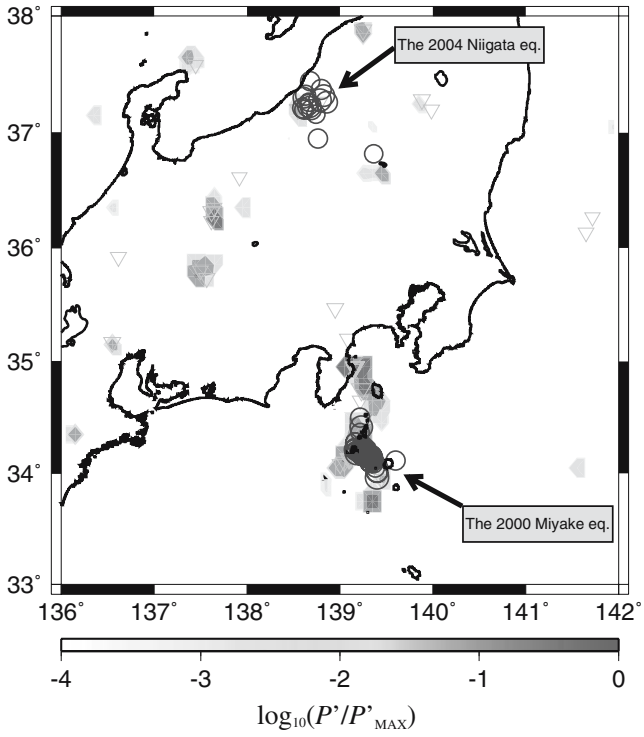


Figure 6
Recreated forecast map for the Tokyo region (No. 6 in Table 1) for the period 2000–2009. This figure is the same as Figure 4, but the 2004 Niigata earthquake ($m = 6.8$) and its large aftershocks ($5.0 \le m$) are added. See also the caption of Figure 4.

indicating that the method may have considerable promise for forecasting future large earthquakes in Japan. Future monitoring of the Tokyo and Kobe regions will test the accuracy and reliability of our PI studies.

6. Conclusion

We have utilized the pattern informatics (PI) technique to obtain the seismic hazard in the Kobe and Tokyo regions of Japan. First, we introduced the PI method. Next we briefly described the JMA catalog with the locations, magnitudes, and times of earthquakes in and around Japan. Earthquakes of magnitude equal to or larger than $m_c = 3$ with depths shallower than 20 km in the time period from 1955 to 2004 were selected. We applied the method for optimal forecasting of large ($m \leq 5$) earthquakes in Kobe region during the time period Jan. 1, 1995 – Mar. 14, 2004 ($t_3 - t_2 \sim 9$ years). This was best illustrated using a specific example. The created forecast map demonstrated that retrospective forecast is reasonably successful. We noted that the 1995 Kobe and 2000 Tottori earthquakes fall in anomalous areas. For comparison, the seismic intensity distribution (null hypothesis) was shown. We found that the optimized PI method narrows the possible locations of the occurrence of future large earthquakes. Statistical testing for the PI method forecast was carried out. For a null hypothesis, the actual seismic intensity data were used. We used maximum likelihood tests with the probability density functions, global Gaussian model and local Poissonian model. The results of the statistical tests showed that the log-likelihood for the PI forecast was larger than that for the null hypothesis. We concluded that the optimized PI method has better forecast skill than the actual seismic intensity data. For further tests of the forecasting algorithm, we make a future forecast to prove correct. The PI method was applied for creating potential maps for large earthquake events with $m \geq 5$ for the forecast time period from Jan. 1, 2000 to Dec. 31, 2009 in the Kobe and Tokyo regions. Several forecast areas on the Kobe map were pointed out: The epicentral area of the 1995 Kobe earthquake, the Wakayama area, the Mie area, and the Aichi area. The Tokyo forecast map was obtained prior to the Oct. 23, 2004 Niigata earthquake ($m = 6.8$) and the principal aftershocks with $5.0 \leq m$ occurred. Some of these earthquakes occurred in an anomalous area in the Tokyo hazard map. This is an example that the PI technique has considerable promise as an intermediate-term earthquake forecasting instruments in Japan.

Acknowledgements

DLT acknowledges the National Science Foundation (USA) under grant NSF ATM-03–27571. JBR and JRH are supported by a grant from US Department of Energy, Office of Basic Energy Sciences to the University of California, Davis DE-

FG03–95ER14499, and through additional funding from the National Aeronautics and Space Administration under grant through the Jet Propulsion Laboratory to the University of California, Davis. JRH is supported also by NASA Headquarters under the Earth System Science Fellowship Grant NGT5. KZN thanks W. Klein for his constructive review, JSPS Research Fellowship for Young Scientists for financial support, JMA for providing Japanese earthquake data, and H. Tsuruoka for discussion of the JMA earthquake data. Figures 1–6 are generated using the GMT software (WESSEL and SMITH, 1998).

REFERENCES

BEVINGTON, P.R. and ROBINSON, D.K. *Data Reduction and Error Analysis for the Physics Sciences* (McGraw-Hill, New York 1992).

FERRAES, S.G. (2003), *The conditional probability of earthquake occurrence and the next large earthquake in Tokyo, Japan*, J. Seismol. *7* (2), 145–153.

FISHER, D.S., DAHMEN, K., RAMANATHAN, S., and BEN-ZION, Y. (1997), *Statistics of earthquakes in simple models of heterogeneous faults*, Phys. Rev. Lett. *78* (25), 4885–4888.

GROSS, S. and RUNDLE, J.B. (1998), *A systematic test of time-to-failure analysis*, Geophys. J. Int. *133*, 57–64.

HOLLIDAY, J.R., RUNDLE, J.B., TIAMPO, K.F., KLEIN, W., and DONNELLAN, A. *Systematic procedural and sensitivity analysis of the pattern informatics method for forecasting large ($m \leq 5$)* earthquake events in Southern California, Pure Appl. Geophys. *163*, 2433–2454.

JOHANSEN, A., SORNETTE, D., WAKITA, H., TSUNOGAI, U., NEWMAN, W.I., and SALEUR, H. (1996), *Discrete scaling in earthquake precursory phenomena: Evidence in the Kobe earthquake*, Japan, J. Phys. I France *6* (10), 1391–1402.

KAGAN, Y.Y. and JACKSON, D.D. (2000), *Probabilistic forecasting of earthquakes*, Geophys. J. Int. *143*, 438–453.

KEILIS-BOROK, V.I. (1990), *The lithosphere of the Earth as a nonlinear system with implications for earthquake prediction*, Rev. Geophys. *28*, 19–34.

KEILIS-BOROK, V.I. and KOSSOBOKOV, V.G. (1990), *Premonitory activation of earthquake flow: Algorithm M8*, Phys. Earth Planet. Inter. *61*, 73–83.

KEILIS-BOROK, V.I. and ROTWAIN, I.M. (1990), *Diagnosis of time of increased probability of strong earthquakes in different regions of the world*, Phys. Earth Planet. Inter. *61*, 57–72.

KEILIS-BOROK, V.I. and SOLOVIEV, A.A. eds., *Nonlinear Dynamics of the Lithosphere and Earthquake Prediction* (Springer-Verlag, New York 2003).

KLEIN, W. (2006, unpublished), *Stress field evolution near the spinodal*.

KLEIN, W., RUNDLE, J.B., and FERGUSON, C.D. (1997), *Scaling and nucleation in models of earthquake faults*, Phys. Rev. Lett. *78*, 3793–3796.

MORI, H. and KURAMOTO, Y., *Dissipative Structures and Chaos* (Springer-Verlag, Berlin 1998).

MOGI, K., *Earthquake Prediction* (Academic Press, Tokyo 1985).

RIKITAKE, T. (1990), *Threat of an earthquake right under the capital in Japan*, Earthq. Volcanol. *22*, 209–210.

RUNDLE, J.B., KLEIN, W., TIAMPO, K., and GROSS, S.J. *Dynamics of seismicity patterns in systems of earthquake faults*. In *GeoComplexity and the Physics of Earthquakes, Geopghys. Monogr. Ser., vol. 120* (eds. RUNDLE, J.B., D.L. TURCOTTE, D.L., and KLEIN, W.) (American Geophysical Union, Wasgington D. C. 2000a) pp. 127–146.

RUNDLE, J.B., KLEIN, W., TIAMPO, K., and GROSS, S. (2000b), *Linear pattern dynamics in nonlinear threshold systems,* Phys. Rev. E *61* (3), 2418–2431.

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. Natl. Acad. Sci. U.S.A. *99*, 2514–2521.

RUNDLE, J.B., TURCOTTE, D.L., SHCHERBAKOV, R. , KLEIN, W., and SAMMIS, C. (2003), *Statistical physics approach to understanding the multiscale dynamics of earthquake fault systems,* Rev. Geophys. *41* (4), 1019, doi:10.1029/2003RG000135.

SCHOLZ, C.H. *The Mechanics of Earthquakes and Faulting,* 2nd ed. (Cambridge University Press, New York 2002).

SCHORLEMMER, D., GERSTENBERGER, M., WIEMER, S., and JACKSON, D.D. (2006, submitted), *Earthquake likelihood-model testing,* Seismol. Res. Lett. http://www.earthquake.ethz.ch/docs/drafts/draft-relmtest.pdf.

TIAMPO, K.F., RUNDLE, J.B., MCGINNIS, S., GROSS, S.J., and KLEIN, W. (2002a), *Eigenpatterns in southern California seismicity,* J. Geophys. Res. *107* (B12), 2354, doi: 10.1029/2001JB000562.

TIAMPO, K.F., RUNDLE, J.B., MCGINNIS, S., GROSS, S.J., and KLEIN, W. (2002b), *Mean field threshold systems and phase dynamics: an application to earthquake fault systems,* Europhys. Lett. *60*, 481–487.

TIAMPO, K.F., RUNDLE, J.B., MCGINNIS, S., and KLEIN, W. (2002c), *Pattern dynamics and forecast methods in seismically active regions,* Pure Appl. Geophys. *159*, 2429–2467.

TSUNOGAI, U. and WAKITA, H. (1995), *Precursory chemical-changes in-ground water: Kobe earthquake, Japan,* Science *269*, 61–63.

TSUNOGAI, U. and WAKITA, H. (1996), *Anomalous changes in groundwater chemistry - possible precursors of the 1995 Hyogo-Ken Nanbu earthquake, Japan,* J. Phys. Earth *44* (4), 381–390.

TURCOTTE, D.L. (1991), *Earthquake prediction,* Annu. Rev. Earth Planet. Sci. *19*, 263–281.

WESSEL, P. and SMITH, W.H.F. (1998), *New improved version of the generic mapping tool released,* EOS Trans. AGU *79*, 579.

(Received December 1, 2004, revised September 10, 2005, accepted September 12, 2005)
Published Online First: December 20, 2006

Pure appl. geophys. 163 (2006) 2433–2454
0033–4553/06/122433–22
DOI 10.1007/s00024-006-0131-1

Pure and Applied Geophysics

Systematic Procedural and Sensitivity Analysis of the Pattern Informatics Method for Forecasting Large (M > 5) Earthquake Events in Southern California

J.R. Holliday,[1,2] J.B. Rundle,[1,2] K.F. Tiampo,[3] W. Klein,[4] and A. Donnellan[5]

Abstract—Recent studies in the literature have introduced a new approach to earthquake forecasting based on representing the space-time patterns of localized seismicity by a time-dependent system state vector in a real-valued Hilbert space and deducing information about future space-time fluctuations from the phase angle of the state vector. While the success rate of this Pattern Informatics (PI) method has been encouraging, the method is still in its infancy. Procedural analysis, statistical testing, parameter sensitivity investigation and optimization all still need to be performed. In this paper, we attempt to optimize the PI approach by developing quantitative values for "predictive goodness" and analyzing possible variations in the proposed procedure. In addition, we attempt to quantify the systematic dependence on the quality of the input catalog of historic data and develop methods for combining catalogs from regions of different seismic rates.

Key words: Pattern Informatics, earthquake forecasting.

1. Introduction

Large magnitude earthquakes are devastating events which can have great social, scientific, and economic impact. The 26 December 2003 magnitude 6.7 Iran earthquake killed nearly 30,000 persons. The 16 January 1995 Japan magnitude 6.9 earthquake produced an estimated $200 billion loss. Similar scenarios are possible at any time in San Francisco, Seattle, and other U.S. urban centers along the Pacific plate boundary, especially in Southern California. The gravity of potential

[1] Center for Computational Science and Engineering, University of California, One Shields Avenue, Davis, CA 95616-8677, U.S.A. E-mail: holliday@cse.ucdavis.edu
[2] Department of Physics, University of California, One Shields Avenue, Davis, CA 95616-8677, U.S.A. E-mail: jbrundle@ucdavis.edu
[3] Department of Earth Sciences, University of Western Ontario, Biology and Geological Sciences Bldg., London, Ontario, Canada, N6A 5B7. E-mail: ktiampo@uwo.ca
[4] Department of Physics, Boston University, 590 Commonwealth Avenue, Boston, MA 02215, U.S.A. E-mail: klein@buphyc.bu.edu
[5] Earth and Space Sciences Division, Jet Propulsion Laboratory, Mail Stop 183-335, 4800 Oak Grove Drive, Pasadena, CA 91109-8099, U.S.A. E-mail: donnellan@jpl.nasa.gov

loss of life and property is so great that reliable earthquake forecasting should be at the forefront of research goals.

While millions of dollars and thousands of work years have been spent on observational programs searching for reliable precursory phenomena, to date few successes have been reported and no precursors to large earthquake events have been detected that provide reliable forecasts. Indeed, many wonder if earthquake forecasting is even possible (see, for example, the online debate hosted at http://www.nature.com/nature/debates/earthquake).

A new approach to earthquake forecasting, the pattern informatics (PI) approach, has been proposed by RUNDLE *et al.* (2000a, b, 2002, 2003) and TIAMPO *et al.* (2002a, b, c). This approach is based on the strong space-time correlations that are responsible for the cooperative behavior of driven threshold systems and arise both from threshold dynamics as well as from the mean field (long range) nature of the interactions.

Using both simulations and observed earthquake data, they have shown that the space-time patterns of threshold events (earthquakes) can be represented by a time-dependent system state vector in a Hilbert space. The length of the state vector represents the average temporal frequency of events throughout the region and is closely related to the rate at which stress is dissipated. It can be deduced that the information about space-time fluctuations in the system state is represented solely by the phase angle of the state vector. Changes in the norm of the state vector represent only random fluctuations and can for the most part be removed by requiring the system state vector to have a constant norm. A more detailed summary of the method is given in section 4.

2. Background

Earthquake fault systems are now believed to be a complex example of a highly nonlinear system (BAK and TANG, 1989; RUNDLE and KLEIN, 1995). Interactions among a spatial network of fault segments are mediated by means of a potential that allows stresses to be redistributed to other segments following slip on any particular segment. For faults embedded in a linear elastic host, this potential is a stress Green's function whose exact form can be calculated from the equations of linear elasticity, once the current geometry of the fault system is specified. A persistent driving force, arising from plate tectonic motions, increases stress on the fault segments. Once the stresses reach a threshold characterizing the limit of stability of the fault, a sudden slip event results. The slipping segment can also trigger slip at other locations on the fault surface whose stress levels are near the failure threshold as the event begins. In this manner, earthquakes occur that result from the interactions and nonlinear nature of the stress thresholds.

The Karhunen-Loeve method (FUKUNAGA, 1970; HOLMES *et al.*, 1996), a linear decomposition technique in which a dynamical system is decomposed into a complete set of orthonormal subspaces, has been applied to a number of other complex nonlinear systems over the last fifty years, including the ocean-atmosphere interface, turbulence, meteorology, biometrics, statistics, and even solid earth geophysics (HOTELLING, 1993; FUKUNAGA, 1970; AUBREY and EMERY, 1983; PREISENDORFER, 1988; SAVAGE, 1988; PENLAND, 1989; VAUTARD and GHIL, 1989; GARCIA and PENLAND, 1991; PENLAND and MAGORIAN, 1993; PENLAND and SARDESHMUKH, 1995; HOLMES *et al.*, 1996; MOGHADDAM *et al.*, 1998). The notable success of this method in analyzing the ocean-atmosphere interface and such features as the El Niño Southern Oscillation (ENSO), a nonlinear system whose underlying physics is governed by the Navier-Stokes equation, suggested its application to the analysis of the earthquake fault system (NORTH, 1984; PREISENDORFER, 1988; PENLAND and MAGORIAN, 1993; PENLAND and SARDESHMUKH, 1995). Building on these methods for analyzing nonlinear threshold systems, space-time seismicity patterns can be identified in both observed phenomena and numerical simulations using realistic earthquake models for southern California (BUFE and VARNES, 1993; BOWMAN *et al.*, 1998; GROSS and RUNDLE, 1998; BREHM and BRAILE, 1999; JAUME and SYKES, 1999; TIAMPO *et al.*, 1999, 2000; RUNDLE *et al.*, 2000b).

The PI method is an adaptation of the Karhunen-Loeve expansion technique to the analysis of observed seismicity data from Southern California in order to identify basis patterns for all possible space-time seismicity configurations. These basis states represent a complete, orthonormal set of eigenvectors and associated eigenvalues, obtained from the diagonalization of the correlation operators computed for the regional historic seismicity data, and, as such, can be used to reconstitute the data for various subset time periods of the entire dataset.

3. Data

The primary dataset employed in this analysis is the entire historic seismic catalog from 1 January 1932 through 31 December 1999, obtained from the Southern California Earthquake Data Center (SCEDC) online searchable database[1] , with all non-local and blast events specifically removed. The relevant data consists of location, in East longitude and North latitude, and the date the event occurred. Seismic events between -122° and -115° longitude and between 32° and 37° latitude (any depth and quality) and with magnitude greater than or equal to $M_{min} = 3.0$ were selected.

While the SCEDC catalog is among the best available, both in completeness and historic depth, there are a number of known deficiencies[2] that undoubtedly affect the

[1] http://www.data.scec.org/catalog_search/index.html

[2] http://www.data.scec.org/catalog_search/known_issues.html

quality of our constructed forecast hot-spot maps. The most notable of these issues is that the four-year span of data from 1977–1980 is currently not available to web searching. Fortunately, data for these missing years are available from the older Southern California Seismic Network (SCSN) archives[3] and was hand inserted for this analysis. Unless otherwise indicated, all analysis was performed using SCEDC data with the additional SCSN data.

A second source of data employed in this analysis was acquired from the Northern California Earthquake Data Center (NCEDC) online searchable database,[4] with all non-local and blast events again specifically removed. When incorporating this catalog, seismic events between -122° and -115° longitude and between 35° and 37° latitude (any depth and quality) and with magnitude greater than or equal to $M_{min} = 3.0$ were selected. The necessity for utilizing an additional catalog in some of our analysis arises from various earthquake events in the vicinity of 35° North latitude missing from the SCEDC catalog but present in the NCEDC collection.

4. Basic Method

Here we summarize the current PI method as described by RUNDLE *et al.* (2003) and TIAMPO *et al.* (2002c). The PI approach is a six step process that creates a time-dependent system state vector in a real valued Hilbert space and uses the phase angle to predict future states (RUNDLE *et al.*, 2003). The method is based on the idea that the future time evolution of seismicity can be described by pure phase dynamics (MORI and KURAMOTO, 1998; RUNDLE *et al.*, 2000a, b). Hence, a real-valued seismic phase function $\hat{S}(\mathbf{x}_i, t_b, t)$ is constructed and allowed to rotate in its Hilbert space. Since seismicity in active regions is a noisy function (KANAMORI, 1981), only temporal averages of seismic activity are utilized in the method. The geographic area of interest is partitioned into N square bins centered on a point $\mathbf{x}_i$ and with an edge length dx determined by the nature of the physical system. For our analysis we chose $dx = 0.1^\circ \sim 11$ km, corresponding to the linear size of a magnitude M$\sim$6 earthquake. Within each box, a time series $\psi_{obs}(\mathbf{x}_i, t)$ is defined by counting how many earthquakes with magnitude greater than M_{min} occurred during the time period t to $t + dt$. Next, the activity rate function $S(\mathbf{x}_i, t_b, T)$ is defined as the average rate of occurrence of earthquakes in box i over the period t_b to T:

$$S(\mathbf{x}_i, t_b, T) = \frac{\sum_{t=t_b}^{T} \psi(\mathbf{x}_i, t)}{T - t_b}. \quad (1)$$

[3] http://www.data.scec.org/ftp/catalogs/SCSN/

[4] http://quake.geo.berkeley.edu/ncedc/catalog-search.html

If t_b is held to be a fixed time, $S(\mathbf{x}_i, t_b, T)$ can be interpreted as the i-th component of a general, time-dependent vector evolving in an N-dimensional space (TIAMPO *et al.*, 2002c). Furthermore, it can be shown that this N-dimensional correlation space is defined by the eigenvectors of an $N \times N$ correlation matrix (RUNDLE *et al.*, 2000a, b). The activity rate function is then normalized by subtracting the spatial mean over all boxes and scaling to give a unit-norm:

$$\hat{S}(\mathbf{x}_i, t_b, T) = \frac{S(\mathbf{x}_i, t_b, T) - \frac{1}{N}\sum_{j=1}^{N} S(\mathbf{x}_j, t_b, T)}{\sqrt{\sum_{j=1}^{N}\left[S(\mathbf{x}_j, t_b, T) - \frac{1}{N}\sum_{k=1}^{N} S(\mathbf{x}_k, t_b, T)\right]^2}}. \tag{2}$$

The requirement that the rate functions have a constant norm helps remove random fluctuations from the system. Following the assumption of pure phase dynamics (RUNDLE *et al.*, 2000a, b), the important changes in seismicity will be given by the change in the normalized activity rate function for the time period t_1 to t_2:

$$\Delta\hat{S}(\mathbf{x}_i, t_b, t_1, t_2) \;=\; \hat{S}(\mathbf{x}_i, t_b, t_2) - \hat{S}(\mathbf{x}_i, t_b, t_1). \tag{3}$$

This is simply a pure rotation of the N-dimensional unit vector $\hat{S}(\mathbf{x}_i, t_b, T)$ through time. In order to remove the last free parameter in the system, the choice of base year, and to further reduce random noise components, changes in the normalized activity rate function are averaged over all possible base-time periods:

$$\Delta\underline{\hat{S}}(\mathbf{x}_i, t_0, t_1, t_2) \;=\; \frac{\sum_{t_b=t_0}^{t_1} \Delta\hat{S}(\mathbf{x}_i, t_b, t_1, t_2)}{t_1 - t_0}. \tag{4}$$

Finally, the probability of change of activity in a given box is deduced from the square of its base averaged, mean normalized change in activity rate:

$$P(\mathbf{x}_i, t_0, t_1, t_2) \;=\; [\Delta\underline{\hat{S}}(\mathbf{x}_i, t_0, t_1, t_2)]^2. \tag{5}$$

In phase dynamical systems, probabilities are related to the square of the associated vector phase function (MORI and KURAMOTO, 1998; RUNDLE *et al.*, 2000b). This probability function is often given relative to the background by subtracting off its spatial mean:

$$P'(\mathbf{x}_i, t_0, t_1, t_2) \;=\; P(\mathbf{x}_i, t_0, t_1, t_2) - \frac{1}{N}\sum_{j=1}^{N} P(\mathbf{x}_j, t_0, t_1, t_2), \tag{6}$$

where P' indicates the probability of change in activity and is measured relative to the background.

Schematically, this whole process can be represented by

$$N \;\rightarrow\; S \;\rightarrow\; \hat{S} \;\rightarrow\; \Delta\hat{S} \;\rightarrow\; \Delta\underline{\hat{S}} \;\rightarrow P,$$

where the *hat* symbol is understood to mean "calculate normalization in space", the capital Delta means "calculate the change in rate", and the underscore symbol means

"average over base times". Note that this method implicitly assumes earthquake fault systems are in an unstable equilibrium state and can be treated linearly about their equilibrium points.

4.1. Variations in Order

To determine the optimal application of the PI method, we identified and analyzed all physically meaningful variations of the described procedure. While we have outlined above a six step process, there are considerably fewer than 6! = 720 variations that need to be investigated. A forecast analysis must always begin with binning the available data and end with a calculation of probability change. Also, base-time averaging and calculation of changes in the activity rate functions can only be performed after creating the activity rate vectors. With these constraints imposed, there are only eight possible variations in the order to which each step is performed. Table 1 lists these eight variations with the original method denoted Method **I**.

On the basis of theoretical arguments and assumptions of linearity within the system, we expect that Methods **I** through **VI** should perform qualitatively similar to each other. This is due largely to the fact that the operations being permuted are all linear and commute with each other. Qualitatively it is unclear which variation should yield the best correlation with actual future events other than to expect Methods **II** and **III** might perform better than Method **I** due to the movement of when the change in activity rate is calculated to after the normalization and base-time averaging steps. This essentially places all of the activity rate vectors on equal footing and legitimizes the vector rotation. We also expect that Methods **VII** and **VIII** will yield both qualitatively and quantitatively inferior forecast hot-spot maps. This is due to the direct normalization of the binned data. Such a step destroys correlations between different spatial locations by independently scaling the relative historic intensity rates. Each of these expectations are verified in the results section below.

Table 1

Possible variations in the procedure ordering. The analysis must always begin with data binning and end with probability calculation. Recall N is binned data, S is the activity rate, P is a probability calculation, the $\hat{S}$ symbol represents normalization in space, the Δ symbol represents calculation of change in rate, and the underscore symbol represents averaging over base times

Method	Procedure										
I	N	$\rightarrow$	S	$\rightarrow$	$\hat{S}$	$\rightarrow$	$\Delta\hat{S}$	$\rightarrow$	$\Delta\underline{\hat{S}}$	$\rightarrow$	P
II	N	$\rightarrow$	S	$\rightarrow$	$\hat{S}$	$\rightarrow$	$\underline{\hat{S}}$	$\rightarrow$	$\Delta\underline{\hat{S}}$	$\rightarrow$	P
III	N	$\rightarrow$	S	$\rightarrow$	$\underline{S}$	$\rightarrow$	$\underline{\hat{S}}$	$\rightarrow$	$\Delta\underline{\hat{S}}$	$\rightarrow$	P
IV	N	$\rightarrow$	S	$\rightarrow$	ΔS	$\rightarrow$	$\Delta\hat{S}$	$\rightarrow$	$\Delta\underline{\hat{S}}$	$\rightarrow$	P
V	N	$\rightarrow$	S	$\rightarrow$	ΔS	$\rightarrow$	$\Delta\underline{\hat{S}}$	$\rightarrow$	$\Delta\underline{\hat{S}}$	$\rightarrow$	P
VI	N	$\rightarrow$	S	$\rightarrow$	$\underline{S}$	$\rightarrow$	$\Delta\underline{\hat{S}}$	$\rightarrow$	$\Delta\underline{\hat{S}}$	$\rightarrow$	P
VII	N	$\rightarrow$	$\hat{N}$	$\rightarrow$	S	$\rightarrow$	$\Delta\hat{S}$	$\rightarrow$	$\Delta\underline{\hat{S}}$	$\rightarrow$	P
VIII	N	$\rightarrow$	$\hat{N}$	$\rightarrow$	S	$\rightarrow$	$\underline{\hat{S}}$	$\rightarrow$	$\Delta\underline{\hat{S}}$	$\rightarrow$	P

4.2. Variations in Binning

In addition to the original binning method, we also analyzed time-centered, cumulative, and detrended binning. For time-centered binning, we took each time series and removed the temporal mean:

$$\psi_{\text{obs}}(\mathbf{x}_i, t) \Rightarrow \psi_{\text{obs}}(\mathbf{x}_i, t) - \frac{\sum_{t=t_0}^{t_2} \psi_{\text{obs}}(\mathbf{x}_i, t)}{t_2 - t_0}. \tag{7}$$

For cumulative binning we allowed each time series to build on its past events:

$$\psi_{\text{obs}}(\mathbf{x}_i, t) \;\Rightarrow\; \sum_{T=t_0}^{t} \psi_{\text{obs}}(\mathbf{x}_i, T) \tag{8}$$

For detrended binning, we took each cumulative time series, fit it to a first order polynomial, and subtracted the fitted line:

$$\psi_{\text{obs}}(x_i, t) \;\Rightarrow\; \sum_{T=t_0}^{t} \psi_{\text{obs}}(x_i, T)[A + Bt], \tag{9}$$

where A and B are the parameters of the regression fit. Figure 1 shows the effect of each binning procedure on a synthetic data sample. We will denote the four different binning methods with the labels **A**, **B**, **C**, and **D**, respectively, with **A** denoting the

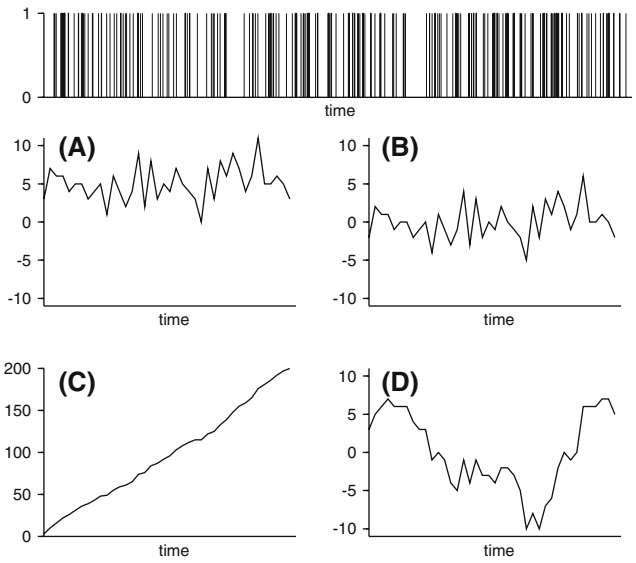


Figure 1
The topmost plot represents random earthquake events over an arbitrary time scale. The four lower plots show the results of the different binning methods: A) normal, B) time-centered, C) cumulative, and D) detrended.

unmodified method. Methods **B** and **D** are significant in that they remove the mean for each time series from the data. Thus, anomalous activity away from background seismicity is expected to be emphasized. Method **C** is reminiscent of an unbiased estimator in the cumulative distribution Kolmogorov-Smirnov Test (PRESS *et al.*, 2002) and could in theory allow more accurate comparisons among the different time series.

We also investigated magnitude- and energy-weighted binning in which the value at each time step is proportional to either the total magnitude M_{tot} of all the events in the time period or to the total energy ($\sim 10^{Mtot}$) of all the events. These weighting factors, however, had the effect of selecting out time periods surrounding only the largest events and were thus unsuitable for the analysis. We did not investigate *Boolean* binning in which each time step is given an initial value of either *1* if one or more events occur in that time period or *0* otherwise due to the realization that this effect can be achieved by sufficiently reducing the time step *dt*. Also, we desired the method to scale appropriately as *dt* is increased.

4.3. Variations in Projection

In addition to calculating the change in the activity rate function through the vector rotation during the time period t_1 to t_2, we also investigated the effect of linear projection of change into future times:

$$\Delta S(\mathbf{x}_i, t_b, t_1, t_2) \rightarrow S(\mathbf{x}_i, t_b, t_2) + \Delta S(\mathbf{x}_i, t_b, t_1, t_2). \tag{10}$$

The motivation behind this investigation was that for regions with a near constant rate of seismicity (or with frequencies higher than an inverse time step), $\Delta\hat{S}(\mathbf{x}_i, t_b, t_1, t_2) \approx 0$. By linear projection, we mean that the future seismic activity for this type of situation would be approximately equal to the present seismic activity with a small correction added. For notational purposes, we will denote the unmodified approach of calculating the change in the activity rate function with the label **1** after the method specification. We will denote the linear projection approach with the label **2**.

4.4. Variations in dt

While the spatial width of the boxes, *dx,* is determined by the nature of the physical system, the temporal binning width *dt* is arbitrary. Larger values of *dt* result in greater bin statistics and faster execution time of the algorithm while lower values may potentially yield greater sensitivity to high frequency periodicity.

To investigate the effect, we performed the analysis with representative values for *dt* ranging from one day to one year. If the catalog is uniform in its completeness and not missing bands of data at quasi-periodic intervals, we would expect to find a smooth transition through the varying choices of *dt* with perhaps some optimal selection. On the other hand, large fluctuations in the forecast as *dt* is slowly modified

may indicate underlying chaotic phenomena and would bring into question the assumptions and treatment of linearity within the system.

5. Statistical Tests

To test the hypothesis that the probability measure P_i can forecast future ($t > t_2$) large ($M > 5$) events, we performed a set of maximum likelihood tests (BEVINGTON and ROBINSON, 1992; GROSS and RUNDLE, 1998; KAGAN and JACKSON, 2000; TIAMPO *et al.*, 2002b; SCHORLEMMER *et al.*, 2003). The likelihood **L** is a probability measure that can be used to assess the quality of one forecast measure over another. Typically, one computes $\boldsymbol{L} = \log(\mathbf{L})$ for the proposed forecast measure **L** and compares that to the likelihood measure $\boldsymbol{L^0} = \log(\mathbf{L^0})$ for a representative null hypothesis. The ratio of these two values then yields information about which measure is more accurate in forecasting future events. In the likelihood ratio test, a probability density function (PDF) is required. Two different PDFs were used in this analysis: A global, Gaussian model and a local, Poissonian model. These distributions differ significantly in that the Gaussian model assumes purely random, normal statistics while the Poissonian model assumes independent statistics over small time intervals with no temporal clustering (WALPOLE and MYERS, 1993).

5.1. Global Gaussian Model

In their original analysis, TIAMPO *et al.* (2002b) calculated likelihood values by defining $P_i = P[\mathbf{x}_i]$ to be the union of a set of N Gaussian density functions $p_G(|\mathbf{x}-\mathbf{x}_i|)$ (BEVINGTON and ROBINSON, 1992) centered at each location $\mathbf{x}_i$. Each individual Gaussian density has a standard deviation equal to the box width dx and a peak value equal to the calculated probability of change in activity P_i divided by the standard deviation squared. $P[\mathbf{x}(e_j)]$ is therefore a probability measure that a future large event e_j occurs at location $\mathbf{x}(e_j)$:

$$P[\mathbf{x}(e_j)] = \sum_i \frac{P_i}{\sigma^2} e^{-\frac{|\mathbf{x}(e_j)-\mathbf{x}_i|^2}{\sigma^2}}. \tag{11}$$

If there are J future events, the normalized likelihood **L** that all J events are forecast is:

$$L = \prod_j \frac{P[e(\mathbf{x}_j)]}{\sum_i P[\mathbf{x}_i]}. \tag{12}$$

Furthermore, the log-likelihood value $\boldsymbol{L}$ for a given calculation can be calculated and used in ratio comparison tests:

$$\log(L) = \sum_j \log \frac{P[e(\mathbf{x}_j)]}{\sum_i P[\mathbf{x}_i]}. \tag{13}$$

Before performing the statistical analysis, the change in activity values P_i were first truncated by scaling all the probabilities equally up-wards and performing a *histogram cut* to enforce the restriction $\Delta P \leq 1$. This was used to eliminate the exponential tail on the high end of the PDF and ensure that events that occurred during the forecasting time period had a probability $\Delta P = 1$ of occurring (which, in fact, they did).

5.2. Local Poissonian Model

The second model used is based on work performed by the Regional Earthquake Likelihood Models (RELM) group (SCHORLEMMER *et al.*, 2003). For each bin i an expectation value λ_i is calculated by scaling the local probability P_i by the number of earthquakes that occurred over all space during the forecast time period:

$$\lambda_i = nP_i, \tag{14}$$

where n is the number of post-t_2 events. Note that for any future time interval (t_2, t_3), n could in principle be estimated by using the Gutenberg-Richter relation. For each bin an observation value w_i is also calculated such that w_i contains the number of post-t_2 earthquakes that actually occurred in bin i. For the RELM model, it is assumed that earthquakes are independent of each other. Thus, the probability of observing w_i events in bin i with expectation λ_i is the Poissonian probability

$$p_i(w_i|\lambda_i) = \frac{\lambda_i^{w_i}}{w_i!} e^{-\lambda_i}. \tag{15}$$

The log-likelihood for observing w earthquakes at a given expectation λ is defined as the logarithm of the probability $p_i(w_i, \lambda_i)$, thus

$$\log(L(w|\lambda)) = \log p(w|\lambda) = -\lambda + w \log \lambda - \log(w!). \tag{16}$$

Since the joint probability is the product of the individual bin probabilities, the log-likelihood value for a given calculation is the sum of $\log(L(w, \lambda))$ over all bins i.

When using this PDF function, we preprocess the change in activity values P_i by performing the same *histogram cut* as with the Gaussian model.

6. Results

Results for the procedural analysis with variations in binning and calculation of activity rate are presented in Tables 2 and 3. All values of $\boldsymbol{L}$ are given relative to $\boldsymbol{L^0}$ defined to be the value supplied by our original, unaltered Method **I-A1**. Since these are ratio tests, greater values indicate better predictive ability.

As statistical evaluations of earthquake forecasts are still under development, it is instructive to weigh the quantitative (“predictive goodness” values) against the

Table 2

*Relative likelihood values $L_G - L^0$ using a global Gaussian model over the time period $t = 1984 \rightarrow 1994$ for the various variations in order, binning, and calculation of change in activity rate. Recall that **A–D** denote normal, time-centered, cumulative, and detrended binning, respectively, while **1** and **2** denote normal and projected calculations of change in activity rate. For our null hypothesis, L^0, we took the value from Method* **I-A1**. *Larger (more positive) values are better correlated with actual events*

Method	A1	B1	C1	D1	A2	B2	C2	D2
I	0.00	−13.06	−11.27	−18.80	−36.47	−32.23	−19.43	−24.62
II	3.33	−8.65	−21.91	−17.96	−36.14	−30.92	−14.17	−23.27
III	2.70	−1.04	−32.58	−19.89	−15.28	−15.28	−14.74	−21.99
IV	−2.89	−2.08	−16.10	−13.87	−31.20	−16.43	−15.94	−12.57
V	−7.99	−4.75	−14.35	−19.70	−34.48	−12.94	−14.67	−21.51
VI	−2.76	−2.92	−17.63	−19.92	−33.23	−10.88	−14.54	−21.05
VII	−20.32	−17.41	−14.87	−32.44	−48.93	−10.90	−16.03	−33.38
VIII	−16.65	−21.57	−37.77	−32.02	−47.32	−10.99	−15.05	−33.42

Table 3

*Relative likelihood values $L_P - L^0$ using a local Poissonian model over the time period $t = 1984 \rightarrow 1994$ for the various variations in order, binning, and calculation of change in activity rate. Recall that **A – D** denote normal, time-centered, cumulative, and detrended binning, respectively, while **1** and **2** denote normal and projected calculations of change in activity rate. For our null hypothesis, L^0, we took the value from Method* ***I – A1***. *Larger (more positive) values are better correlated with actual events*

Method	A1	B1	C1	D1	A2	B2	C2	D2
I	−0.00	1.29	−38.14	−30.87	−57.74	−44.65	−5.77	−74.67
II	4.93	5.58	−60.65	−28.60	−18.05	−29.54	−2.09	−48.88
III	2.94	14.74	−59.22	−26.22	5.04	5.04	−2.01	−35.93
IV	7.75	6.77	−7.27	−12.30	−32.11	−14.98	−3.15	−11.46
V	0.43	−0.52	−7.38	−43.10	−45.47	−5.94	−2.12	−45.89
VI	0.84	0.63	−9.89	−40.51	−21.67	−3.99	−2.04	−55.60
VII	−59.34	−51.33	−61.89	−85.76	−81.90	−47.86	−44.11	−81.12
VIII	−45.73	−57.16	−76.22	−87.33	−83.09	−48.66	−44.12	−81.55

qualitative (pictorial representation of the forecast hot-spot maps). Thus, representative maps for each procedural variation are given in Figures 2 and 3.

Only Methods **II** and **III**, using normal binning and change of activity calculation, performed better than the original method under the two statistical tests. Naively, this result is expected as both methods wait until after normalization and base year averaging to calculate the change in activity rate, thus giving the calculations in each box equal statistical weight. For all other investigated variations, no method performed better on both likelihood tests and qualitative analysis.

While a few of the binning and change of activity variations fared well on one or the other likelihood tests (for example, **III-B1**), most performed poorly qualitatively. Probability calculations gave predictions of activity that spread well into areas with no recorded activity. These results can be understood by considering their

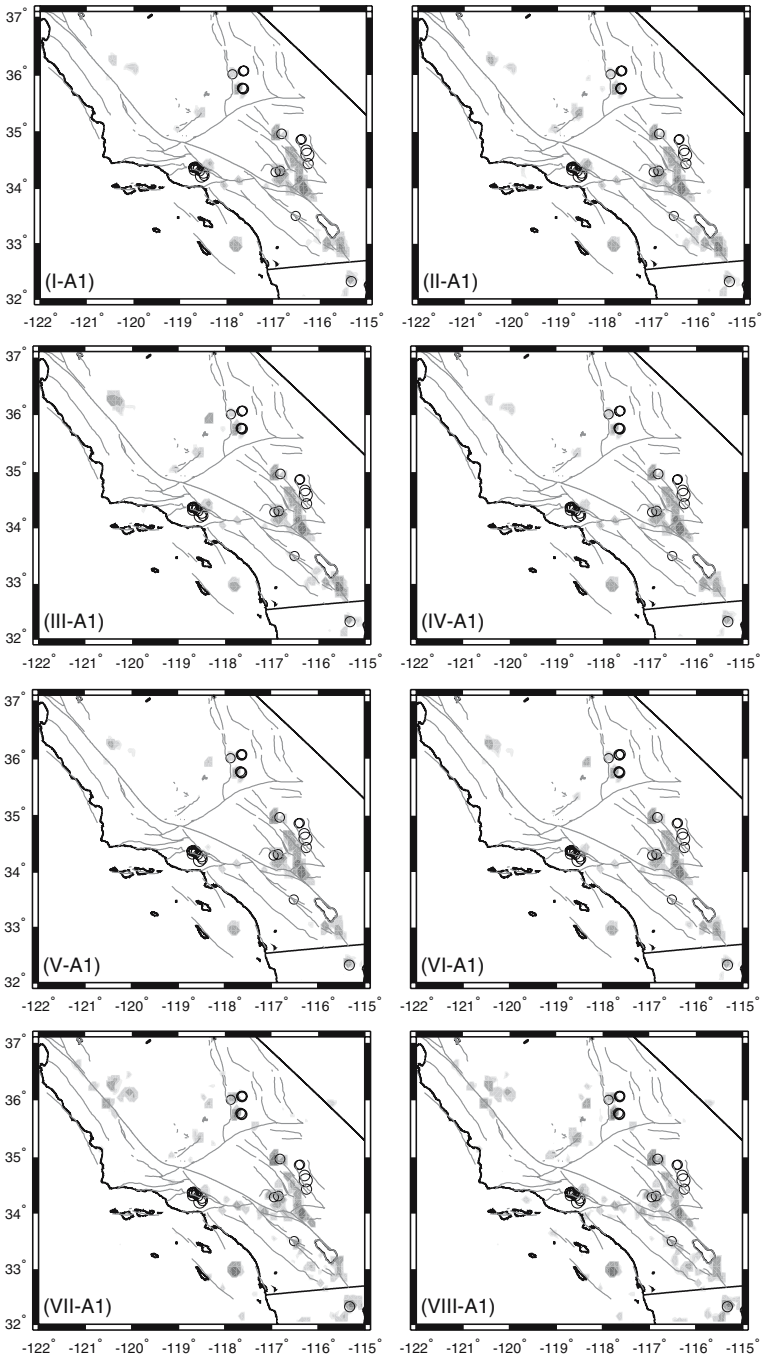


Figure 2
Representative forecast hot-spot maps created using each of the order variations with normal binning and calculation of change in activity rate for the time period t = 1984 to 1994. Note the increase in apparent noise for Methods VII and VIII.

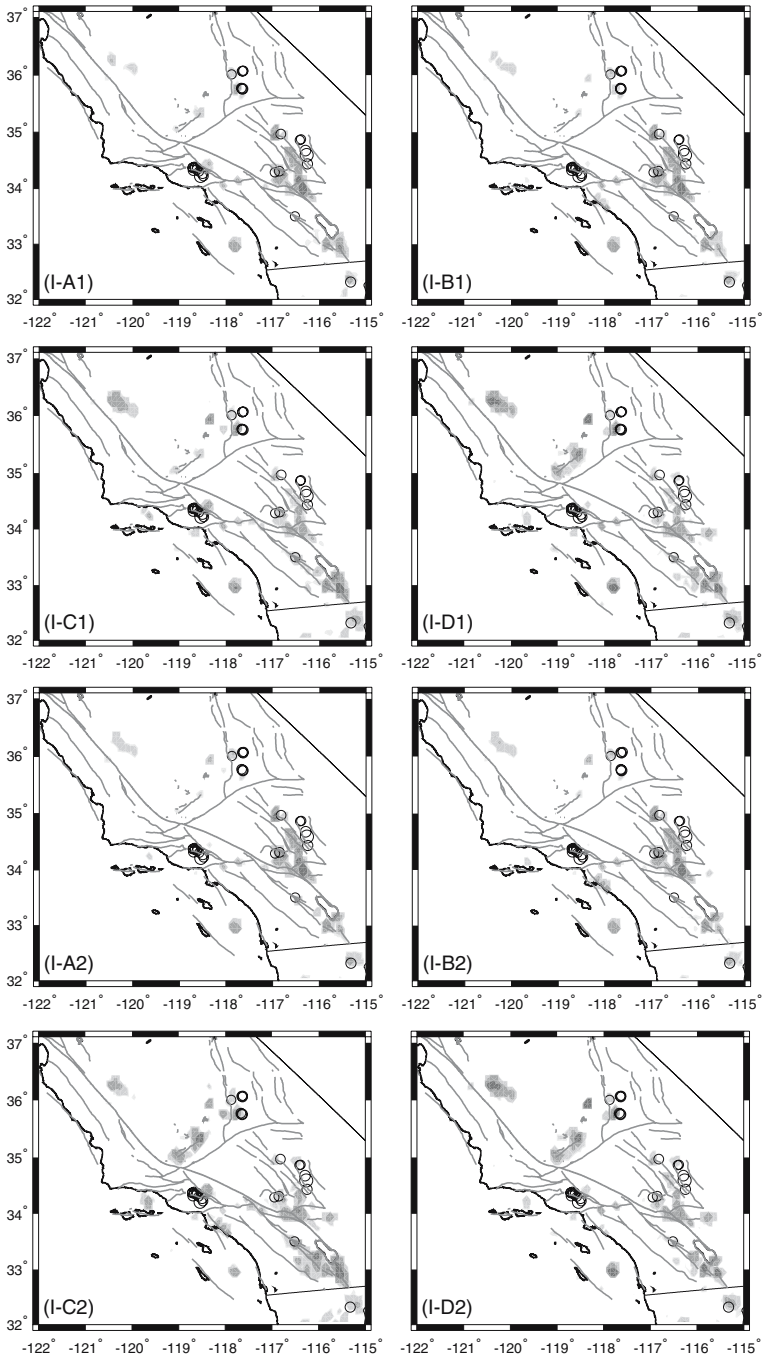


Figure 3
Representative forecast hot-spot maps created using each of the variations in binning and calculation of change in activity rate for Method I over the time period t = 1984 to 1994.

Table 4

Relative likelihood values using Method ***III-A1*** *with varying time steps (in days) over the time period t = 1984 → 1994. For our null hypothesis, we took the value at dt = 1 day. Larger (more positive) values are better correlated with actual events*

dt	=	1	3	5	7	15	30	60	90	180	365
$L_G - L^0$	=	0.00	−0.07	−0.16	−0.13	−0.63	−1.33	−2.69	−17.00	−34.06	−20.17
$L_P - L^0$	=	0.00	−0.55	−0.70	−1.43	−4.52	−7.31	−9.66	−24.43	−85.22	−33.66

mathematical operations. By linearly projecting the change in activity rate, heavy weight is placed on the most recent seismic history. For the procedure to identify anomalous changes in the seismicity, however, the entire history must be considered equally. Also, the cumulative and detrended variations in the binning method create time series that are significantly altered from those apparent in nature.

While only Methods **II-A1** and **III-A1** performed better than the original PI procedure on both statistical tests, it should be stressed that at this time none of the methods can be claimed to be superior. There is still a subjective element over which forecast hot-spot map to prefer. Based on theoretical and mathematical considerations, Method **III-A1** is the authors' preferred choice. This method creates a unique state vector at every time step and allows the purest interpretation of a vector rotation.

Table 4 shows the results of varying the time step in the analysis (note that Method **III-A1** was used). Likelihood values for this investigation were referenced against a choice of $dt = 1$ day. Note that the accuracy of the calculated forecast decreases with increasing time step, slowly decreasing up to around $dt = 1$ week and then rapidly decreasing. While larger choices of dt decrease time of computation for the PI algorithm, they do so at the cost of accuracy. Evaluating the data from Table 4, along with the corresponding forecast hot-spot maps, the authors believe $dt = 7$ days to be a suitable compromise. This choice of time step is low enough to probe the seismic periodicity at all scales with reasonable accuracy while being large enough to significantly speed up the computation.

7. Catalog Sensitivity

To gauge the sensitivity of the PI method on the quality of the input catalog, we decimated the available data by systematically increasing both the starting date of catalog information (and thus affecting t_0) and the minimum magnitude threshold. Figures 4 and 5 show the effect on the relative likelihood values of varying either parameter individually. Both probability density functions—Poissonian and Gaussian—were used to calculate log likelihood indexes.

In Figure 4 we see the surprising result that the forecast is relatively stable as t_0 is increased, up to around 1965. This would indicate that accurate forecast hot-spot

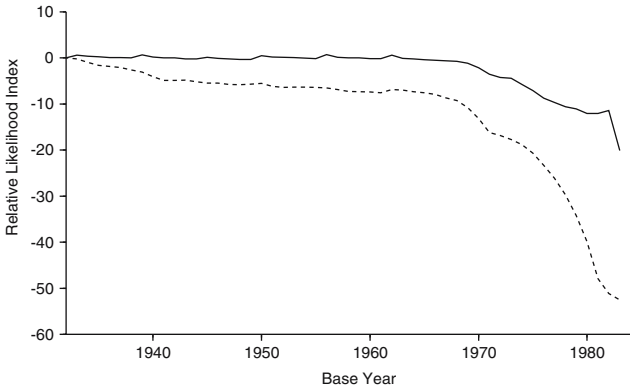


Figure 4
Relative likelihood values for two different probability density functions, Gaussian (solid) and Poissonian (dashed), as a function of t_0. Larger (more positive) values are better correlated with actual events. The plateau in the data before $t_0 = 1965$ indicates that only ~40 years of historic data is necessary for the analysis.

maps can be created using only approximately 40 years of historic data. When the normalized activity rate functions are averaged over all possible base-time periods, more recent data gets weighted heavier than more historic data. The threshold at which historic data no longer influences the forecast appears to be approximately 40 years before the onset of the forecast, i.e., t_2. With less than 40 years of historic data, however, the likelihood values drop sharply.

The Poissonian analysis in Figure 5 seems to indicate that higher accuracy in the forecast can be obtained by raising the minimum magnitude cut-off threshold of the analysis from $M_{\min} = 3.0$ to ~3.7. This may have the effect of removing low

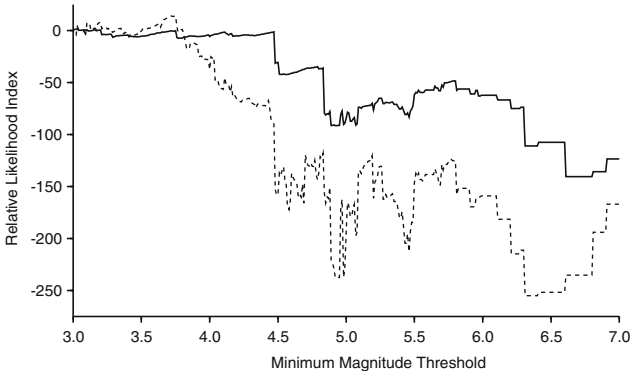


Figure 5
Relative likelihood values for two different probability density functions, Gaussian (solid) and Poissonian (dashed), as a function of the minimum magnitude cut-off threshold. Larger (more positive) values are better correlated with actual events. Using the Poissonian PDF, more probable forecasts appear possible by increasing the magnitude threshold slightly.

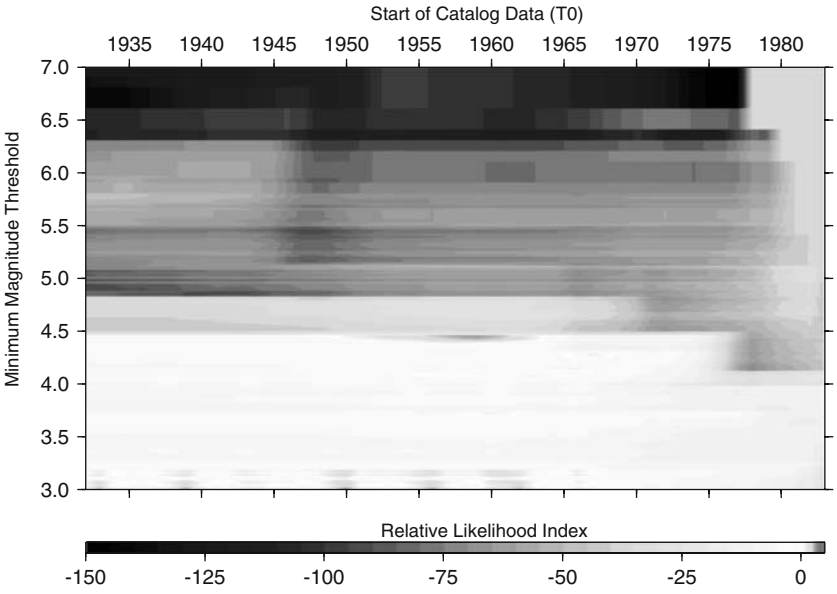


Figure 6
Relative likelihood index calculated using a Gaussian density function as a function of both t_0 and minimum magnitude cut-off threshold. Warmer colors are better correlated with actual events.

magnitude events that are uncorrelated with future large magnitude events and thereby eliminate background noise from the analysis. Care must be taken, however, as the likelihood values drop quickly as the magnitude threshold is raised too high. It is interesting to note the sudden drop in likelihood values as the magnitude threshold reaches 4.5 (and again near 4.8, 5.5, and 6.3). While statistics may be playing a role in the latter three drops, the discontinuity at $M_{\min} = 4.5$ appears to identify an unknown deficiency in the catalog.

Figures 6 and 7 show the effect on the relative likelihood values of varying both parameters simultaneously. For these two-dimensional plots, warmer colors indicate better correlation between the forecast and actual events. All of the features mentioned above are again evident as well as the surprising observation that increasing $M_{\min}$ allows accurate forecasts with less historic data (as indicated by the positive slope of the high-likelihood-edge surrounding $M_{\min} = 3.6$ and $t_0 = 1967$).

8. Application of the Method

To test the optimization on the PI method, we recreated the forecast seismic hot-spot map originally presented by RUNDLE *et al.* (2002) for the time period 1 January 2000 to 31 December 2009 using Method **III-A1**. The result is shown in Figure 8. The original forecast was made using only data from the SCEDC

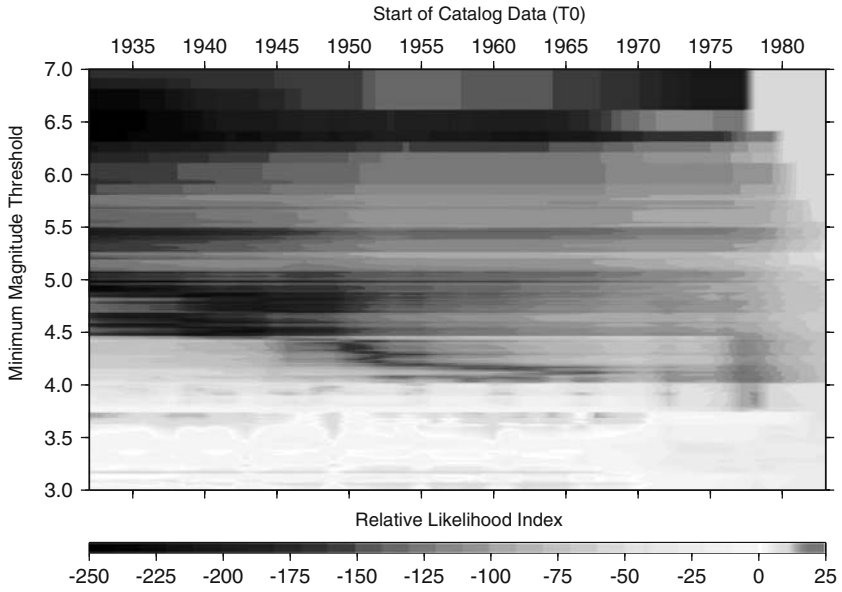


Figure 7
Relative likelihood index calculated using a Poissonian density function as a function of both t_0 and minimum magnitude cut-off threshold. Warmer colors are better correlated with actual events.

catalog, which does not contain earthquakes from the San Simeon region (location of the M = 6.5, 2003 event; label #7 in Fig. 8). Our revised forecast was made using data from both the NCEDC catalog (for latitude above 35°) and the SCEDC catalog (for latitude below 35°).

Since the cut-off date for the forecast of 31 December 1999, eight large earthquake events with M > 5 have occurred in Central or Southern California. The first seven events all occurred either on areas of forecasted anomalous activity or within the margin of error of ± 11 km. While this hot-spot map was made after each of these events occurred, it was done so using only data prior to 31 December 1999 and could have in principle predicted these events. Scorecards using the original method and the current optimized method can be found at the JPL QuakeSim website[5] .

9. Combining Catalogs

The issue of how to combine historic catalogs in order to create forecast hot-spot maps for large regions is a difficult one. Problems arise from the fact that different

[5] http://www-aig.jpl.nasa.gov/public/dus/quakesim/scorecard.html

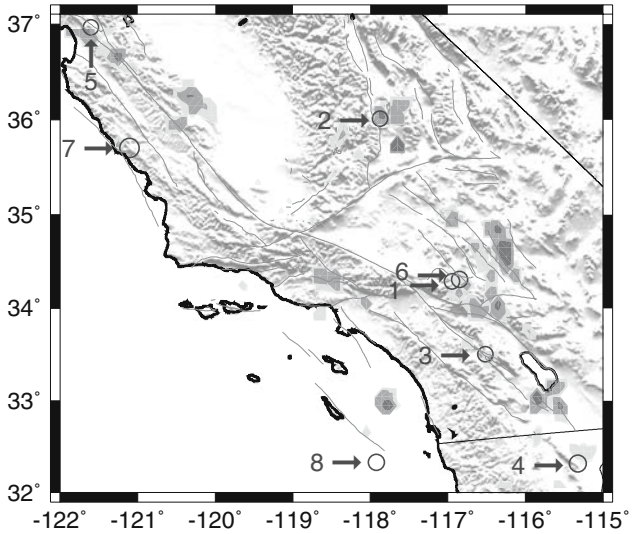


Figure 8
Seismic hot-spot map for large earthquake events with M > 5 for the forecast time period 1 January 2000 to 31 December 2009. Since the cut-off date for the forecast, eight large earthquake events with M > 5 have occurred in Central or Southern California. Seven of the eight events occurred either on areas of forecasted anomalous activity or within the margin of error of ±11 km. Data from the SCEDC catalog were used below 35° North latitude, and from the NCEDC catalog above 35° North latitude.

areas will normally have widely different seismic rates, and these differences become smoothed out when we normalize our state vectors.

One way to try and account for these differences is to apply a weighting factor to the different catalogs as they are merged into an aggregate catalog. This method, however, tends to emphasize near threshold-level anomalous activity in the catalog with the highest weighted activity rate. In Figure 9 we created a forecast hot-spot map by combining data from the NCEDC and SCEDC catalogs with two different weighting ratios. With equal weighting between the two catalogs (Fig. 9A), event #3 (Anza) occurs near a threshold-level anomalous region. Event #7 (San Simeon), however, is missed completely. As the relative weighting for the northern catalog is increased to account for its lower total seismic rate (Fig. 9B), anomalous activity begins to appear under event #7, but disappears from event #3.

Another way to attempt to demonstrate the differences is to apply a weighting factor to each individual time series based on its own statistics. This method, unfortunately, also has failings. By weighing each time series individually, correlations between local events are destroyed. In practice, this approach has effects similar to the earlier proposed modifications **VII** and **VIII** to the PI procedure and simply

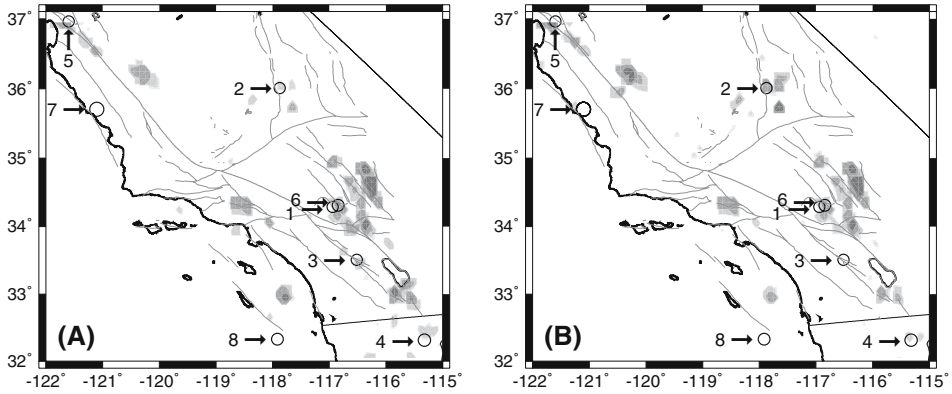


Figure 9
Equal weight for both catalogs (A) vs. higher weighting for northern catalog (B). With the equally weight map, event #3 occurs near a threshold-level anomalous region while event #7 does not. The opposite is true with the unequally weight map.

results in more apparent noise in the forecast and less correlation with actual future events.

Currently, the best approach (at least for this time period and these catalogs) appears to be to treat all catalogs and regions separately, combining only at the end of the analysis and normalizing over all spatial bins to allow for correlations across the catalog seams.

10. Conclusion

We have analyzed the current PI procedure and developed a more optimized approach for creating accurate forecast hot-spot maps. First, historic seismic data are binned by counting the number of earthquakes per unit time, of any size greater than or equal to $M_{\min}$, within a geographic box centered at x_i at some time t. The geographic region defined by dx is taken large enough so that seismic activity can be considered an incoherent superposition of phase functions. Second, an activity rate function is defined as the average rate of occurrence of earthquakes in box i over the period t_b to T. Third, the activity rate function is averaged over all possible base-time periods. Fourth, the base-year averaged activity rate function is normalized by subtracting the spatial mean over all boxes and scaling to give a unit-norm. Fifth, changes in the base-year averaged, mean-normalized activity rate function are calculated by allowing the vector to rotate over time. Finally, the probability of change of activity in a given box–calculated relative to the background–is deduced from the square of its base-year averaged, mean-normalized change in activity rate.

We also showed that the choice of *dt* is relatively unimportant to the calculation if it is taken low enough, that only approximately 40 years of complete historic data is necessary for accurate forecasts, and that the assumptions of linearity and near-equilibrium appear valid for Southern California seismic fault systems. Applying our new procedure, we recalculated and updated the Southern California forecast hot-spot map presented by RUNDLE *et al.* (2002) and showed that the 22 December 2003 San Simeon event could have been foreseen. Finally, we identified pitfalls associated with combining seismic catalogs from different regions in an attempt to create a composite forecast hot-spot map.

There is movement in the forecast verification community to part with likelihood calculations, which lightly reward successes and heavily penalize failures, and embrace ROC verification diagrams (JOLIFFEE and STEPHENSON, 2003). Additional analyses that utilize these verification techniques are currently underway.

Acknowledgments

The authors are grateful to the anonymous reviewers for their helpful criticisms and suggestions. This work has been supported by a grant from US Department of Energy, Office of Basic Energy Sciences to the University of California, Davis DE-FG03-95ER14499 (JRH and JBR), by a NASA Earth Science Fellowship NN-6046Q98H (JRH), and through additional funding from the National Aeronautics and Space Administration under grants through the Jet Propulsion Laboratory to the University of California, Davis.

REFERENCES

AUBREY, D.G. and EMERY, K.O. (1983), *Eigenanalysis of recent United States sea levels*, Continental Shelf Res. *2*, 21–33.

BAK, P. and TANG, C. (1989), *Earthquakes as self-organized critical phenomena*, J. Geophys. Res. *94*, 15,635–15,637.

BEVINGTON, P.R. and ROBINSON, D.K. (1992), *Data Reduction and Error Analysis for the Physical Sciences*, McGraw-Hill.

BOWMAN, D.D., OUILLON, G. SAMMIS, C.G., SORNETTE, A., and SORNETTE, D. (1998), *An observational test of the critical earthquake concept,* J. Geophys. Res. *103*, 24,359–24,372.

BREHM, D.J. and BRAILE, L.W. (1999), Intermediate-term earthquake prediction using the modified time-to-failure method in Southern California, *BSSA 89*, 275–293.

BUFE, C.G. and VARNES, D.J. (1993), *Predictive modeling of the seismic cycle of the greater San Francisco bay region,* J. Geophys. Res. *98*, 9871–9883.

FUKUNAGA, K., *Introduction to Statistical Pattern Recognition*, (Academic Press, New York (1970)).

GARCIA, A. and PENLAND, C. (1991), *Fluctuating hydrodynamics and principal oscillation pattern analysis,* J. Stat. Phys. *64*, 1121–1132.

GROSS, S. and RUNDLE, J. B. (1998), *A systematic test of time-to-failure analysis,* Geophys. J. Int. *133*, 57–64.

HOLMES, P. LUMLEY, J.L., and BERKOOZ, G., *Turbulence, Coherent Structures, Dynamical Systems and Symmetry*, (Cambridge University Press, Cambridge, U.K 1996).

HOTELLING, H. (1993), *Analysis of a complex of statistical variables into principal components,* J. Educ. Psych. *24*, 417–520.

JAUMÉ, S.C. and SYKES, L.R. (1999), *Evolving towards a critical point: A review of accelerating seismic moment/energy release prior to large and great earthquakes,* Pure Appl. Geophys. *155*, 279–306.

JOLIFFEE, I.T. and STEPHENSON, D.B., *Forecast Verification* (John Wiley. (2003)).

KAGAN, Y.Y. and JACKSON, D.D. (2000), *Probabilistic forecasting of earthquakes,* Geophys. J. Int. *143*, 438–453.

KANAMORI, H., The nature of seismicity patterns before large earthquakes. In *Earthquake Prediction: An International Review*, Geophys. Monogr. Ser., pp. 1–19, AGU (Washington, D. C. (1981)).

MOGHADDAM, B., WAHID, W., and PENTLAND, A. (1998) Beyond eigenfaces: Probabilistic matching for face recognition. In *Third IEEE Intl. Conf. on Automatic Face and Gesture Recognition*, pp. 1–6.

MORI, H. and KURAMOTO, Y., *Dissipative Structures and Chaos*, (Springer-Verlag, Berlin. (1998)).

NORTH, G.R. (1984), *Empirical orthogonal functions and normal modes,* J. Atm. Sci. *41* (5), 879–887.

PENLAND, C. (1989), *Random forcing and forecasting using principal oscillation pattern analysis,* Monthly Weather Rev. *117*, 2165–2185.

PENLAND, C. and MAGORIAN, T. (1993), *Prediction of Niño 3 sea-surface temperatures using linear inverse modeling,* J. Climate *6*, 1067–1076.

PENLAND, C. and SARDESHMUKH, P.D. (1995), *The optimal growth of tropical sea surface temperature anomalies,* J. Climate *8*, 1999–2024.

PREISENDORFER, R.W., *Principle Component Analysis in Meteorology and Oceanography* (Elsevier, Amsterdam. (1988)).

PRESS, W.H., TEUKOLSKY, S.A., VETTERLING, W.T., and FLANNERY, B.P., *Numerical Recipes in C* (Cambridge University Press, Cambridge, MA. (2002)).

RUNDLE, J.B. and KLEIN, W. (1995), *New ideas about the physics of earthquakes,* Rev. Geophys. Space Phys. Suppl. (July) *283*, 283–286.

RUNDLE, J.B., KLEIN, W., GROSS, S.J., and TIAMPO, K.F., Dynamics of seismicity patterns in systems of earthquake faults. In *Geocomplexity and the Physics of Earthquakes*, Geophys. Monogr. Ser., vol. 120 (eds. by J.B. Rundle, D.L. Turcotte, and W. Klein,) pp. 127–146 (AGU, Washington, D. C. 2000a).

RUNDLE, J.B., KLEIN, W., TIAMPO, K.F., and GROSS, S.J. (2000b), *Linear pattern dynamics in nonlinear threshold systems,* Phys. Rev. E. *61*, 2418–2432.

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. Natl. Acad. Sci. U. S. A. *99*, 2514–2521: Suppl. 1.

RUNDLE, J.B., TURCOTTE, D.L., SHCHERBAKOV, R., KLEIN, R., and SAMMIS, C. (2003), *Statistical physics approach to understanding the multiscale dynamics of earthquake fault systems*. Rev. Geophys. *41*(4), 1019, doi:10.1029/2003RG000135.

SAVAGE, J.C. (1988), *Principal component analysis of geodetically measured deformation in long valley caldera, eastern California*. 1983–1987, J. Geophys. Res. *93*, 13,297–13,305.

SCHORLEMMER, D., JACKSON, D.D., and GERSTENBERGER, M. (2003), *Earthquake likelihood model testing*. http://moho.ess.ucla.edu/~kagan/sjg.pdf.

TIAMPO, K.F., RUNDLE, J.B., KLEIN, W., and GROSS, S.J. (1999), *Systematic evolution of nonlocal space-time earthquake patterns in Southern California*. EOS Trans. AGU *80*, 1013.

TIAMPO, K.F., RUNDLE, J.B., MCGINNIS, S., GROSS, S.J., and KLEIN, W., *Observation of systematic variations in non-local seismicity patterns from southern California*. In *Geocomplexity and the Physics of Earthquakes*, Geophys. Monogr. Ser. vol. 120 (eds. J.B. Rundle, D.L. Turcotte, and W. Klein,) pp. 211–218 (AGU, Washington, D. C. 2000).

TIAMPO, K.F., RUNDLE, J.B., MCGINNIS, S., GROSS, S.J., and KLEIN, W. (2002a), *Eigenpatterns in Southern California seismicity*. J. Geophys. Res. *107* (B12), 2354, doi:10.1029/2001JB000562.

TIAMPO, K.F., RUNDLE, J.B., MCGINNIS, S., GROSS, S.J., and KLEIN, W. (2002b), *Mean field threshold systems and earthquakes: An application to earthquake fault systems*. Europhys. Lett. *60* (3), 481–487.

TIAMPO, K.F., RUNDLE, J.B., MCGINNIS, S., and KLEIN, W. (2002c), *Pattern dynamics and forecast methods in seismically active regions*. Pure Appl. Geophys *159*, 2429–2467.

VAUTARD, R. and GHIL, M. (1989), *Singular spectrum analysis in nonlinear dynamics, with applications to paleodynamic time series*. Physica *D 35*, 395–424.

WALPOLE, R.E. and MYERS, R.H., Probability and Statistics for Engineers and Scientists (Prentice Hall. 1993).

(Received December 28, 2004, revised July 17, 2005, accepted August 29, 2005)
Published Online First: December 20, 2006

Pure appl. geophys. 163 (2006) 2455–2465
0033–4553/06/122455–11
DOI 10.1007/s00024-006-0132-0

Pure and Applied Geophysics

Visualization and Analysis of Multi-terabyte Geophysical Datasets in an Interactive Setting with Remote Webcam Capabilities

Benjamin J. Kadlec,[1] David A. Yuen,[1] and Gordon Erlebacher[2]

Abstract—Visualizing and analyzing datasets in the geosciences is becoming increasingly complicated as their volumes are growing explosively. This poses a challenging problem for researchers who must sift through terabytes of data to discover useful relationships inside the information. There is a great need for geophysicists to interactively explore their data sets. Conventional visualization systems lack adequate bandwidth and rendering capabilities necessary for the largest data sets. CAVE and Powerwall display devices are necessary for researchers to explore their data sets in an immersive setting. We describe a utilitarian system targeted specifically at the cost-effective interactive exploration of data sets tens of terabytes in size and harness this system for visualization and analysis of geophysical simulations. Webcams can be used as a steering device to track a local region of interest, which is useful for remote visualization of large data sets. This system will be employed as a web-service under the auspices of Narada-Brokering, while using webcam technologies to enable remote visualization for collaborating researchers. Webcams can be incorporated in a point-to-point network for rapid exchange of information and quickly announcing natural disasters, such as tsunamis, landslides and earthquakes.

Key words: Remote visualization, interactive visualization, webcam, grid computing.

1. Introduction

Multi-terabyte data sets are becoming commonplace in the geosciences due to the proliferation of global observatories and large-scale numerical simulations. Projects such as EarthScope [http://www.earthscope.org] currently provide a wealth of observations on physical processes controlling earthquakes, volcanic eruptions, and tsunamis. Satellite missions, a part of EarthScope, will contribute 200GB/day of data to ongoing projects. The present progress and future needs in computational geosciences have been summarized in a recent report (Cohen, 2005). As

[1] Minnesota Supercomputing Institute and Virtual Laboratory for Earth and Planetary Materials, University of Minnesota, 413 Walter Library, 117 Pleasant St. SE, Minneapolis, MN 55455-0219, U.S.A. E-mail: kadlec@msi.umn.edu

[2] School of Computational Sciences and Information Technology, Florida State University, Tallahassee, FL 32306-4120, U.S.A. E-mail: davey@krissy.geo.umn.edu; erlebach@csit.fsu.edu

described in this report, computational geoscientists continue to run simulations resulting in increasingly large data sets. For example, numerical simulations associated with earthquakes and the predictions of 3-D tsunami wave propagation (WARD, 2004) can result in hundreds of gigabytes of data. On a larger scale, simulations of geophysical phenomena such as 3-D mantle convection are already producing terabyte-size data sets. This multitude of data is easily realized when considering a simulation on a 1000^3 grid with 5 output variables and 1000 time steps results in 20 Tbytes of data ($1000^3 \times 5 \times 1000 \times 4$ bytes $= 2 \times 10^{13}$ bytes). Computing multiple runs of the same simulation will then result in hundreds of terabytes of data. The results of simulations, in particular those relating to natural disasters, must be available in a timely fashion by using fast and efficient techniques for storing and visualizing these data sets. The International Solid Earth Research Virtual Observatory (iSERVO) (http://www.iservo.edu.au, AKTAS *et al.*, 2005) is a globally scalable grid-infrastructure that can provide remote access to data sets and visualization applications through a web-portal interface. We describe a cost-effective data exploration system using remote webcam technologies that can be integrated into the iSERVO network as a web-based component. The remote webcam service will be based on the framework of our Web-based Data Interrogation System (WEB-IS) (GARBOW *et al.*, 2003; WANG *et al.*, 2004). This system will allow visualizations and the analysis to be quickly transmitted to collaborating researchers, decision makers, and the broader community. The aims of this paper are as follows:

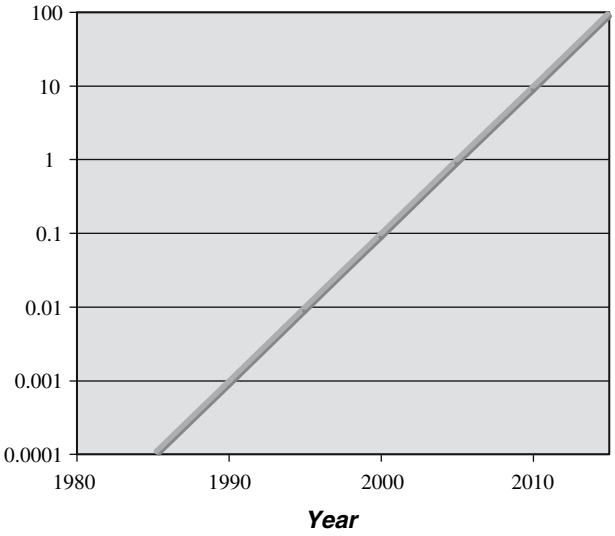


Figure 1
Exponential growth of data sets for 35 years starting from 1980.

- Point out the cost-effectiveness of webcam technology for remote visualization.
- Describe how the use of webcam will fit into the greater framework of WEB-IS and iSERVO.

2. Visualizing Multi-terabyte Datasets

The limiting process in obtaining scientific understanding from computer simulation is not the time required to perform the calculation, but instead the considerably longer time required to analyze and visualize the resulting data. Therefore the main objective of our work is to decrease the time needed to understand scientific data, which typically sets the rate of scientific progress, in the hopes of increasing scientific productivity. This urgency is particularly important when considering the promulgation of information related to impending natural disasters, such as tsunamic wave propagation in realistic ocean basins. Simulation and visualization are tightly coupled and visualizations must closely model the underlying computation (2-D or 3-D). We draw examples from techniques used in 2-D and 3-D simulations of slab dynamics and mantle convection, but other areas in computational geosciences, such as floods, volcanic eruptions and tsunami wave generation, also will produce voluminous data sets.

We have studied the two-dimensional multiscale dynamics of hydrous cold plumes at subduction zones (RUDOLPH *et al.*, 2004) using techniques for extremely high-resolution 2-D visualization, involving on the order of a billion unknowns. Commercial software, such as Matlab, Origin, and Tecplot, was found to be extremely hard-pressed for visualizing this type of large-scale data sets, involving more than 10^7 unknowns, and therefore new tools have been developed for addressing both local and remote visualization solutions. A web-based, zoomable image service (WEB-IS) has been developed to allow a user to explore high-resolution 2-D subduction zones through time, across many thermo-physical properties, and through different spatial scales (RUDOLPH *et al.*, 2004). We use a 13 million-pixel Powerwall display device for comprehending the relationships between multiple physical and chemical properties. The Powerwall enables parallel visualization of multiple fields in an immersive environment (Fig. 2). Recently available large (30-inch) LCD displays with resolutions of over four million pixels have opened the doors for extremely high-resolution powerwalls. The Electronic Visualization Laboratory [http://www.evl.uic.edu] at the University of Illinois at Chicago is developing a large-scale display with a wall of 30-inch LCD screens and a capacity of 4 Megapixels on each panel, essentially creating a powerwall with around one hundred million pixels! These new display walls will be enthusiastically received as simulations are already surpassing resolutions resulting from computations of one billion tracers.

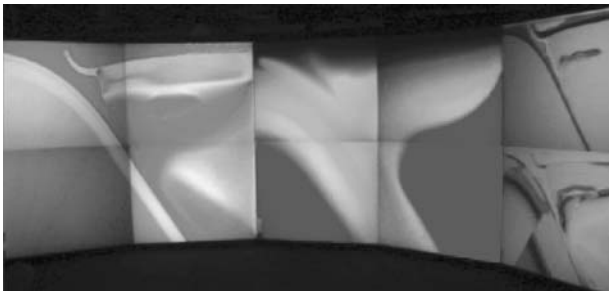


Figure 2
Multi-physics visualization of slab dynamics: Here we show four fields simultaneously on the Powerwall display device. The larger images from left to right are represented by five million pixels and show the viscosity and temperature fields, respectively. For further details see RUDOLPH *et al.* (2004).

We now consider visualization techniques for 3-D mantle convection simulations (Fig. 3) carried out on a large grid space (ERLEBACHER *et al.*, 2002). We have used a large database of 3-D simulations computed at the Earth Simulator Center (KAMEYAMA *et al.*, 2004). For modest data sets, the fast and interactive 3-D environment of Amira [www.amiraviz.com] works excellent when visualizing mantle convection data sets on workstations. Unfortunately, large time-evolving data sets are too big when using workstations with Amira and therefore more sophisticated software and hardware are required for creating large-scale time-evolving 3-D visualizations. Until now the solution has been to work with a down-sampled or smaller subset of the original data set, but this approach does not allow researchers to see their simulation at the highest resolution possible and can result in missed information. We will show that an interactive visualization system coupled with a large-scale display device can solve this problem.

We see promising results in using a CAVE (CRUZ-NEIRA, 1993) display device for 3-D simulations. Its immersive design allows for a natural exploration of 3-D data

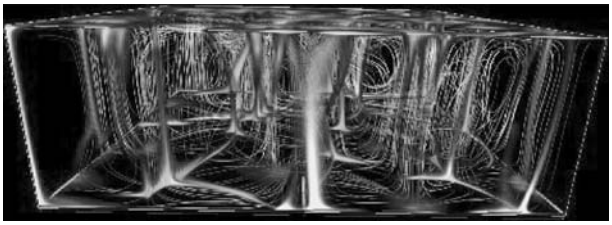


Figure 3
Visualization of 3-D mantle convection at a Rayleigh number of 10^8, using 400^3 evenly spaced points (ERLEBACHER *et al.*, 2002). Larger data sets from runs taken from the Earth Simulator Center (KAMEYAMA *et al.*, 2004) would further aggravate this data-extraction and visualization problem, as there are close to one billion unknowns to be sorted through at each time step. The cold descending flows and the hot rising plumes are depicted by the blue and yellow colors, respectively; streamlines show the directions of the instantaneous flow motion.

sets, such as those from mantle convection or geodynamo simulations. A CAVE is effectively being used for looking interactively at spherical shell convection at the Earth Simulator Center under the direction of Dr. Akira Kageyama [http://www.es.jamstec.go.jp/esc/research/Solid/members/kage/index.en.html]. The visualization centers supporting large displays such as the CAVE or Powerwall have systems tailored to storing and analyzing large data sets. We use the interactive visualization system at the Laboratory for Computational Science and Engineering (LCSE) [http://www.lcse.umn.edu] at the University of Minnesota for exploring our data sets.

The LCSE system uses a Hierarchical Volume Rendering (HVR) software for storing large geophysical data sets at multiple time steps and in full resolution. HVR (Fig. 4) is designed for simulations that require visualizing volumetric data sets, such as those from mantle convection simulations. The software works by saving a simulation's variables at every grid cell as compressed graphics files that contain all the information needed to make a snapshot of any variable over the problem domain. The resulting HVR-files (Figs. 5 and 6) are used for loading data sets into the interactive visualization system. Using a single byte per value, this hierarchical format is able to produce volume rendered images very quickly (PORTER *et al*., 2004) and enables interactive rotation and zoom of large data sets. We have adapted our data sets to fit into this hierarchical data framework by using a Block of Bytes (BOB) (CHIN-PURCELL, 1995) intermediate file format as a way to create a hierarchical layout of bytes. The HVR software runs on a cluster of 10 rendering machines at the LCSE, each of which is connected to one of the Powerwall's ten display panels. Each node is equipped with 16 Gigabytes of RAM, a high-performance graphics card, and

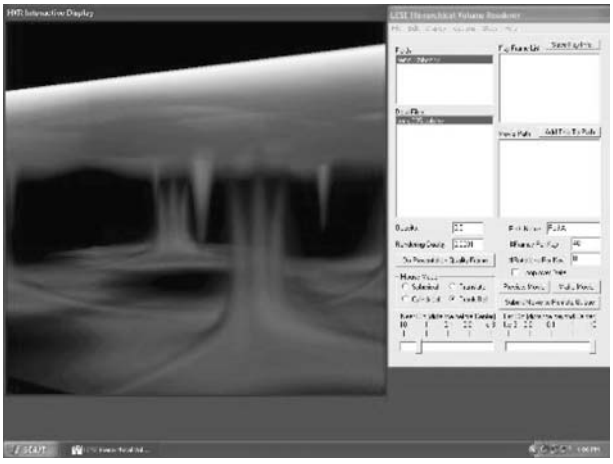


Figure 4
Using the HVR software for visualizing mantle convection data in the LCSE system. The situation depicted is layered convection with a Rayleigh number of 10 billion.

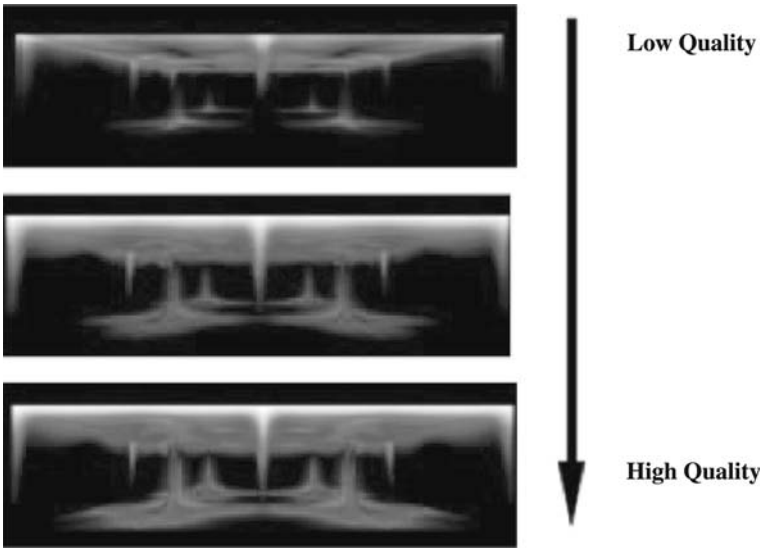


Figure 5
Mantle Convection Data Viewed in a Hierarchical Rendered Format. Rayleigh number of 10^{10} and 500^3 evenly spaced points have been employed (ERLEBACHER *et al.*, 2002).

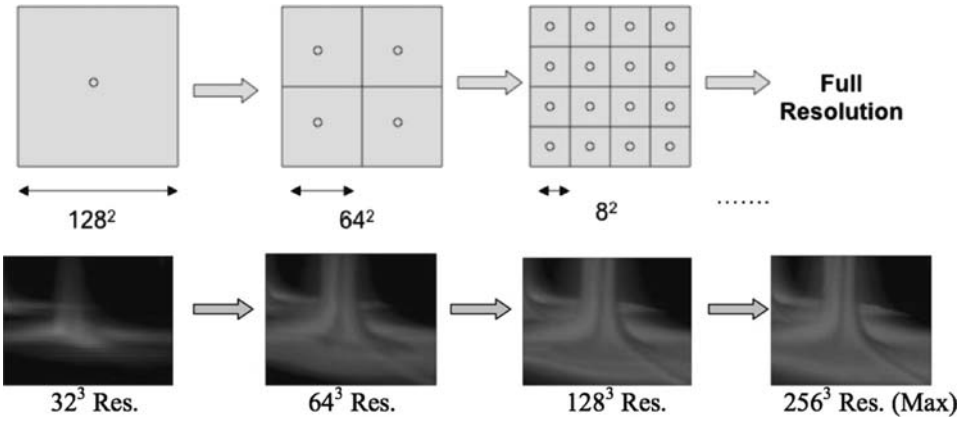


Figure 6
Storing data sets in a hierarchical data format for fast volumetric rendering (top, left to right) (PORTER *et al.*, 2004). The base of a 3-D mantle plume (ERLEBACHER *et al.*, 2002) displayed at increasing resolution (bottom, left to right).

connected to a hundred-terabyte RAID system for staging the largest data sets. This cluster of visualization machines works in-sync to allow interactive visualization of time-evolving data on the Powerwall using HVR.

3. Remote Visualization Using Webcam

As methods for large-scale interactive visualization are quickly being developed, there is increasing motivation for remotely visualizing data at interactive rates, using protocols such as virtual network computing [http://www.realvnc.com]. This need is especially dire in the geosciences, where global collaboration has quickly become commonplace with organizations arising such as ACES [http://www.aces-workshop-2004.ac.cn] and the iSERVO [http://www.iservo.edu.au, ATKAS *et al.*, 2005] group in earthquake physics. We focus now on enabling collaborative visualization by sharing with remote users. Data storage, graphic rendering, and high-resolution display hardware commonplace at regional visualization centers could provide a great benefit to distant researchers with limited access to advanced hardware and software. Webcam technology can be employed as a cost-effective solution for remote collaborative visualization by sharing these resources with remote users, who are looking at different portions of a Powerwall display (Fig. 7).

We have devised a solution for remotely viewing data sets located at distant computing centers and displayed on a large display device, such as the Powerwall, by using webcam technology. Using a method called remote scanning, visualizations are captured from a display device using a webcam equipped with a controllable TrackerCam base [http://www.trackercam.com] and a stream of normal-sized images is broadcasted at a rate of around 10 frames per second to remote users through a java applet (Fig. 8). A demonstration of this visual broadcasting technology was presented during the 2004 ACES Meeting in Beijing, China [http://www.aces-workshop-2004.ac.cn, KADLEC *et al.*, 2004]. Webcams achieve significant data

Figure 7
Schematic representation of collaborative visualization using remote webcam capabilities. Webcam scans geophysical data from the Powerwall and broadcasts it individually to remote users (laptops) through the Naradabrokering middleware. The viewers can look at different portions of the grand image.

compression by using photons to transmit visual information rather than bulky raw data files. The data compression inherent to webcams allows fast transmission of large visualizations, although some information of fine structures in the data set may be lost. Proprietary webcams, such as the Apple iSight®, are cost-effective solutions preferred over more expensive video conferencing hardware, since they are cheap enough to make widely available to the general public. The Apple iSight® can capture video with resolutions up to 1024×768 pixels. This is the maximum number of pixels that can be captured by a webcam when scanning a visualization at its highest resolution, which amounts to an area of 3 ft by 2 ft on the Powerwall. Larger areas of a visualization can be captured when the webcam reverse-zooms and scans a down-sampled version of the larger visualization area. This approach is effective for fast viewing of simulation results when high-resolution is not as necessary as understanding the greater structure of what is being visualized. Computational down-sampling is not necessary with webcam visualizations, and therefore a significant amount of time and resources are saved using this cost-effective approach. The remote java client (Fig. 8) captures a stream of images directly from the webcam and currently only supports a resolution of 640×480 pixels. The effective resolution on the client side is highly dependent on the bandwidth available to remote users, since greater video frame rates and resolutions correspond to a greater load on the network. Latency also needs to be considered when discussing network issues since as bandwidth increases, network latency decreases to under milli-seconds and allows for still greater webcam resolutions to be transmitted.

Recent events following the Sumatran Earthquake on December 26, 2004 have greatly increased the need for a system to quickly broadcast information on pending and ongoing natural disasters. The consequences that follow are well known when these systems are not in place. Therefore, the importance of using webcams for

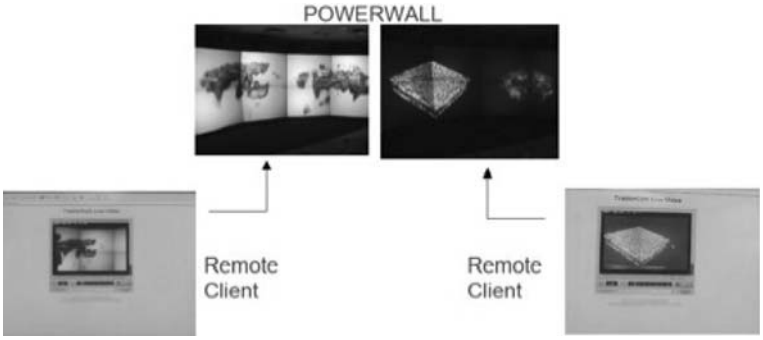


Figure 8
Remote scanning of the Powerwall using webcam. The upper middle images show visualizations of 3-D seismic tomography (top left) and 3-D mantle convection (top right) being scanned by two webcams while displayed on the Powerwall. The lower images are taken from a remote client viewing the webcam visualizations through an interactive java applet at 640×480 pixels.

remote collaboration is relevant now more than ever. Since webcams can be cheaply implemented in remote regions of the world, they remain an effective tool for transmitting visual information to sparsely located researchers and decision makers. By deploying webcams as a web-service in the iSERVO framework, they can be used for interactive remote visualization and broadcasting visual information on disaster events. Integrating additional controls with a web-based remote visualization service has the potential to facilitate additional remote interaction features with this service. The remote webcam service can be modeled after the Web-Based Data Interrogation System (WEB-IS) (GARBOW *et al.*, 2003 and WANG *et al.*, 2004) in the hopes to apply remote webcam scanning to iSERVO using the web-services middleware Narada Brokering [see http://www.naradabokering.org].

This approach follows popular methods currently being used for distance audio/video communication between researchers (KADLEC *et al.*, 2004). Recent advances in the Apple ® Macintosh operating system 10.4 has made it possible for the iChat [http://www.apple.com/ichat] webcam software to be used for multi-user conferencing with both Macintosh and PC users. Additional updates by other webcam softwares (MSN Messenger ®, AOL Instant Messenger ®) over the last year, have primed webcam for greater utility in collaborative movements by allowing for easy webcam communication in environments consisting of diverse computing systems (Windows ® PC, Mac®, Linux, etc.). Online virtual laboratories, such as the iSERVO and the Virtual Laboratory for Earth and Planetary Materials [http://www.vlab.msi.umn.edu], are now in a position to embrace webcam in order to fulfill goals aimed at educational outreach and bridging gaps between disciplines in the geosciences. By allowing different webcam systems to communicate with each other, less commonly used operating systems, such as Macintosh and Linux, are now able to take part in collaboration and outreach where they had previously been neglected.

4. Conclusion

Without any doubt visualization will assume an ever more prominent role in computational endeavors in the foreseeable future. As discussed already in the Cohen report (COHEN, 2005), many young researchers in the geosciences are already enthusiastic to use advanced visualization, but they lack the funding for expensive visualization equipment. They require the use of expensive interactive visualization systems for their simulation results, but unfortunately these resources are only available to a few privileged and well endowed Earth science departments and institutes. Therefore, regional visualization centers would satisfy this need by providing via remote visualization and webcam technologies the means for accessing advanced visualization tools and storing large data sets for many researchers at less well-supported universities. The Cohen report also highlights the need for conveying rapidly visual information for interpreting simulations and transmitting essential

results. The jump to greater bandwidth and Internet-III will allow these aspirations to be satisfied.

In light of recent tragedy from the Sumatran Earthquake, it was shown that the means were not available to disseminate visual information warning of an impending tsunami disaster. In preparing for future geophysical events, simulations like those of tsunami wave propagation will need to be computed and visualized quickly so that results can be broadcasted to decision makers located in remote regions. To this end, we stress the necessity of the webcam as a fast and cost-effective implementation for making visual information available to a broader worldwide community. Through integration with visualization centers and the web-services devoted iSERVO framework, the webcam becomes an invaluable tool for researchers and globally wide communities.

Acknowledgements

We would like to thank Geoffrey Fox, Maxwell Rudolph and Charley Kameyama for stimulating discussions. This research has been supported by the Virtual Laboratory for Earth and Planetary Materials, Math-Geo, and ITR grants from the National Science Foundation.

References

Aktas, M., Aydin, G., Donnellan, A., Fox, G., Granat, R., Grant, L., Lyzenga, G., McLeod, D., Pallickara, S., and Parker, J., Marlon Pierce iSERVO: Implementing the International Solid Earth Research Virtual Observatory by Integrating Computational Grid and Geographical Information Web Serv Pure Appl. Geophys. *163*, 2281–2296.

Bollig, E.F., Kadlec, B.J., Erlebacher, G., Yuen, D.A., and Palchuk, Y.M. (2004), *Interactive collaborative visualization in the geosciences*, Eos Trans. AGU *85*(46), Fall Meet. Suppl., Abstract NG43A-0443.

Cohen, R.E. (2005), *High performance computing requirements for the computational solid earth sciences*, Workshop on *Computational Geoinformatics*, http://www.geo-prose.com/computational_SES.html, May 2–4, 2004.

Cave (2005), *Cave Automatic Virtual Environment*, Electronic Visualization Laboratory at University of Illinois, Chicago, http://www.evl.uic.edu/pape/CAVE/.

Chin-Purcell, K. (1995), *BOB-Brick Of Bytes*, University of Minnesota Army High Performance Computing Research Center.

Cruz-Neira, C., Sandin, D., and DeFanti, T. (1993), *Virtual Reality: The design and implementation of the CAVE*, Proceedings of SIGGRAPH 93 Computer Graphics Conference 08/01/1993 - 08/01/1993, ACM SIGGRAPH, pp. 135–142.

Erlebacher, G., Yuen, D.A., and Dubuffet, F.W. (2002), *Visualization and analysis of high Rayleigh number 3D convection in the Earth's mantle*, Proceedings of IEEE Visualization Conference, Oct.–Nov. pp. 493–496, IEEE Publishers.

Erlebacher, G. and Yuen, D.A. (2004), *A wavelet toolkit for visualization and analysis of large data sets. In Earthquake Research*, Pure Appl. Geophys. *161*, 2215–2229.

GARBOW, Z.A., YUEN, D.A., ERLEBACHER, G., BOLLIG, E.F., and KADLEC, B.J. (2003), *Remote visualization and cluster analysis of 3-D geophysical Data over the internet using off-screen rendering*, Visual Geosciences.

KADLEC, B.J., YUEN, D.A., and SEVRE, E.O.D. (2004), *Webcam for remote visualization and scanning of Giga- to Tera-byte size data sets on large-scale display devices*, APEC Cooperation for Earthquake Simulation (ACES), July 9–14, 2004.

KADLEC, B.J., SEVRE, O.D., YUEN, D.A., LILLI, X.L., BOLLIG, E.F., WANG, Y., ERLEBACHER, G., and RUDOLPH, M. (2004), *Web cam's potential for collaborative activities in the earth sciences*, Visual Geosciences.

KADLEC, B.J., SEVRE, O.D., YUEN, D.A., LILLI, X.L., BOLLIG, E.F., WANG, Y., ERLEBACHER, G., and RUDOLPH, M. (2003), *Web cam as an easy-to-use way for distant collaborative activities*, University of Minnesota Supercomputing Institute Research Report UMSI 2003/194.

KAMEYAMA, M., YUEN, D.A., and KAGEYAMA, A. (2004), *High-resolution 3-D numerical studies on the interplay between variable thermal conductivity and post-Perovskite phase transition*, Eos Trans. AGU *85*(47), Fall Meet. Suppl., Abstract MR23A-0194.

PORTER, D.H., WOODWARD, P.R., and IYER, A. (2004), *Initial experiences with grid-based volume visualization of fluid flow simulations on PC clusters*, University of Minnesota Supercomputing Institute Research Report UMSI 2004/114.

RUDOLPH, M.L., GERYA, T.V., YUEN, D.A., and DEROSIER, S. (2003), *Visualization of complex multiscale phenomena at subduction zones*, Eos. Trans. AGU *84*(46), Fall Meet. Suppl., Abstract NG11A-0167.

RUDOLPH, M.L., GERYA, T.V., YUEN, D.A., and DEROSIER, S. (2004), *Visualization of multiscale dynamics of hydrous cold plumes at subduction zones*, Visual Geosciences.

RUDOLPH, M.L., GERYA, T.V., and YUEN, D.A. (2004), *Vizualization challenges of a subduction simulation using one billion markers*, Eos Trans. AGU *85*(47), Fall Meet. Suppl., Abstract NG43A-0449.

WANG, Y., BOLLIG, E.F., KADLEC, B.J., GARBOW, Z.A., ERLEBACHER, G., YUEN, D.A., RUDOLPH, M., YANG, X.L., and SEVRE, E.O.D. (2004), *WEB-IS (Integrated System): An overall view*, University of Minnesota Supercomputing Institute Research Report UMSI 2004/66.

(Received December 30, 2004, revised August 26, 2005, accepted August 30, 2005)
Published Online First: December 20, 2006

Pure appl. geophys. 163 (2006) 2467–2483
0033–4553/06/122467–17
DOI 10.1007/s00024-006-0146-7

Pure and Applied Geophysics

A Grid Framework for Visualization Services in the Earth Sciences

G. Erlebacher,[1] D. A. Yuen,[2] Z. Lu,[1] E. F. Bollig,[2]
M. Pierce,[3] and S. Pallickara[3]

Abstract—Scientific visualization is an ingredient essential to understanding the large amounts of data generated from large-scale numerical simulations, laboratory experiments and geological surveys. Visualization forms an integral component of any complete framework, together with services to handle mathematical and statistical analysis, storage, feature extraction, and other functions. To support rapid and seamless collaborations and communication between researchers across geographically disparate regions necessitates a distributed infrastructure that supports redundancy, fault tolerance, and most importantly, ease of use. We describe herein an architecture based on Naradabrokering, a publish/subscribe framework that supports the above requirements. We have implemented an initial version of this architecture and describe some initial experiments.

Key words: Middleware, collaboration, visualization, web-services, grids.

1. Introduction

In the past fifteen years, scientific visualization (SV) has evolved from a tool used by a few researchers to understand their data into an indispensible approach towards improving understanding of the extremely large data sets currently being produced. SV is used in many areas of applied research, ranging from meteorology and oceanography, hydrology, material sciences, earth physics, etc., with various degrees of sophistication. Techniques used to enhance data understanding, beyond looking at rows of numbers on printouts, range from simplistic line plots to time-dependent volumetric renderings of data sets at resolutions of 1000^3 and beyond. There is also a need for higher-dimensional representations such as complex higher-dimensional phase diagrams.

[1] School of Computational Science and Information Technology, Department of Mathematics/Computer Science, Florida State University, Tallahassee, FL.
E-mail: zhenyulu@cs.fsu.edu; erlebach@csit.fsu.edu
[2] Minnesota Supercomputer Institute, Department of Geophysics, University of Minnesota, Minneapolis, MN. E-mail: davey@msi.umn.edu; bollig@scs.fsu.edu
[3] Community Grids Lab, Indiana University, Bloomington, IN.
E-mail: mpierce@cs.indiana.edu; spallick@indiana.edu

Comparing the current hardware requirements to those of the early 1990s, we find much in common. In 1990, high-end supercomputers cost in the range of $20–30 million while desktop workstations cost upwards of $30,000. Today the discrepancy is even more severe. Our largest supercomputers cost $100,000,000 while desktop computers with high-end graphics processors cost in the range of $3,000. While the ratio between high-end and low-end computer costs and performance have increased over the years, the network bandwidth has only increased moderately fast, although storage requirements are increasing exponentially fast. Over the past 15 years, there has been an increasing emphasis on collaborative efforts within a given institution, across institutions, even across continents. For example, the ACES program (http://www.uq.edu.au/ACES) studies earthquake dynamics, and involves researchers in Australia, China, and the United States. There is a pressing need to not only share data, but to facilitate sharing of the visual experience, both static and animated. The fast-paced evolution of technology makes it all the more difficult to achieve real collaboration in the midst of changing standards, evolving tools, and the lack of a common software infrastructure (COHEN, 2005).

Large numbers of researchers wish to collaborate with one-another, but are restricted to data visualization on local hardware resources (workstations, large displays, etc.). There is very little support for *easily* sharing their visualizations with one another. Researchers are increasingly mobile, own a wide variety of end display devices, ranging from laptops to hand-held devices, and have access to commodity (i.e., cheap) graphics cards and graphics processors of incredible power (ERLEBACHER and YUEN, 2004). Yet there is very little support to provide a seamless integration between these various technologies. For example, if a researcher develops a tool to visualize earthquake clusters, environments should exist to simplify the integration of this tool within themselves, making it available to all. This vision is addressed by various researchers working on grid technology (FOSTER and KESSELMAN, 1999; FOSTER *et al.*, 2002; UYAR *et al.*, 2003; FOX *et al.*, 2003), a loose collection of hardware and software resources made available to anybody around the world with adequate permissions. Catastrophic events, such as the recent tsunami that produced wide-spread devastation across multiple countries, calls for collaborative systems capable of reaching wide audiences across multiple channels of communication (GADGIL *et al.*, 2005).

2. Dataset Glut

In all areas of the earth sciences, datasets are growing in size at what appears to be exponential rates. For example, when a 3D dataset expands by a factor of two in each coordinate direction, the size of one snapshot expands by a factor of eight, or approximately an order of magnitude.

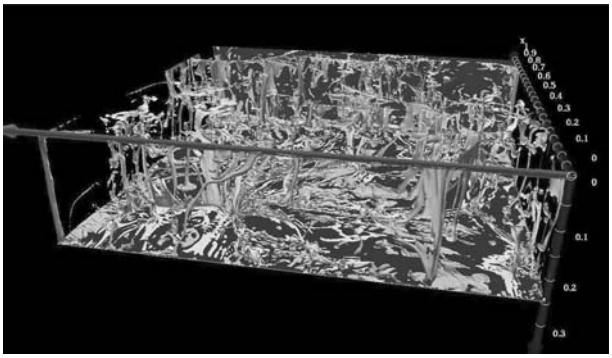


Figure 1
An example of strongly time-dependent high Rayleigh number convection in the Earth's mantle. A Rayleigh number of 10^8 is employed here for a purely basally-heated configuration with an aspect-ratio of $4 \times 4 \times 1$. A grid configuration of $400 \times 400 \times 400$ points has been employed. The temperature fields cast in volumetric rendering are displayed here for both the upwelling (pink isosurface)and downwelling (blue isosurface). Data courtesy Fabien W. Dubuffet.

A 512^3 numerical simulation that stores five variables at the finest resolution in single precision requires $2^{29} \approx 540$ Mbytes of disk space. Storing 200 time steps for future analysis, for example to track the evolution of vortices and other features, requires 100 Gbytes of disk space. The same data set at a resolution of 1024^3 would require 800 Gbytes of storage. As computer power continues to follow the predictions of Moore's law, storage requirements will become increasingly severe.

Satellite systems such as Erbes [NASA] currently collect a Terabyte of data daily. The resolving power of field equipment is also increasing. For example, high resolution spectral channel equipment can scan terrain over 100 simultaneous channels. Resolution is expected to increase further in the immediate future.

3. New Science

Mantle convection, currently recognized as the driving mechanism of plate tectonics, governs the thermal and chemical evolution of the earth. Mantle convection involves both thermal and compositional transport and is a phenomenon with a high range of temporal and spatial scales due to the strong nonlinearities in the governing equations, ranging from rheology to sharp compositional gradients. Figure 1 shows temperature isosurfaces in the earth's mantle using numerical simulation data at $Ra = 10^8$. The complexity of the isosurfaces and the richness of the plume structures suggest the development of plume extraction techniques and interactive navigation (i.e., exploration) capabilities that could be shared between multiple users.

Based on the results of 2-D simulations, convective transitions are expected to occur at Rayleigh numbers within the range $Ra = 10^{10}$ and $Ra = 10^{12}$ (VINCENT and YUEN, 2000). Such high Rayleigh numbers would require a grid of at least 1250^3 points spaced uniformly. It would be very impractical to process and visualize the huge time-dependent data sets from these simulations at the desktop. Rather, a large number of tools will be available to the community, some open source, some commercial, to help datamine the data, extract relevant spatial-temporal subsets (themselves very large) and display them in a manner that clarifies the physics. To this end, users will leverage existing parallel graphics APIs (not useful at the desktop), and other existing (or yet to exist) analysis tools. We have used visualization techniques to examine the interrelationship between the temperature and velocity fields at $Ra = 10^{10}$ at a grid resolution of 512^3. We have found that at this *Ra*, the convective cells do not extend fully between upper and lower boundaries, but rather, break up into one set close to the cold slab, and one near the heated slab, thus cutting off the heat transport between the two boundaries (Fig. 2). This finding required the computation of a large number of streamlines (Fig. 3). Although we downloaded one full data set to the desktop to accomplish this task, it was an inefficient procedure due to lack of memory and insufficient computing resources. It would be more efficient to extract the streamlines on a supercomputer using parallel algorithms and display them either locally (if the number of streamlines is moderate) or remotely. Given the difficulties encountered with a single data set, it is clear that the analysis of a collection of hundreds of such files at the desktop is all but impossible, requiring multiple Terabytes of disk space, and upwards of a Terabyte of local fast memory.

Recently, the role of water in geodynamics is receiving increased attention, driven by laboratory experiments, seismic analysis and numerical modeling.

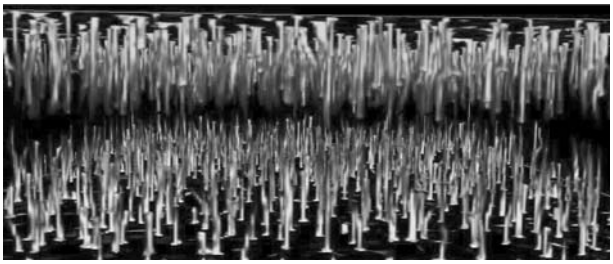


Figure 2
An example of extremely high Rayleigh number convection at $Ra = 10^{10}$. A grid configuration of $600 \times 600 \times 600$ points has been employed. We notice that the system is partially layered from the increasingly large Rayleigh number, which is in agreement with earlier results in 2-D convection by (VINCENT and YUEN, 2000). The upwellings and downwellings are denoted by the yellow and blue colors, respectively. The morphologies are bounded by a certain magnitude of the dimensionless temperature.

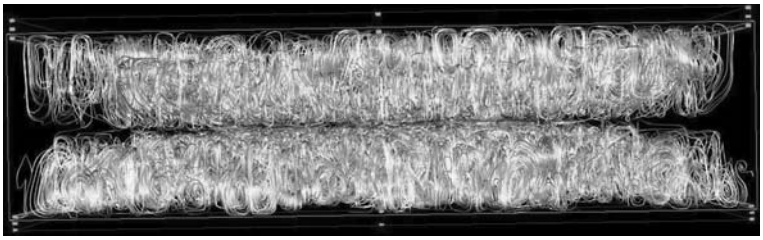


Figure 3
Streamlines corresponding to the data in Figure 2 at $Ra = 10^{10}$. They clearly indicate a layered configuration.

Introducing water into the dynamical models changes the nature of the equations to a complex system characterized by multiple components with many different intrinsic scales. We have modeled the dynamics of hydrous plume development above subducting slabs. Our 2-D model incorporates the effects of thermal-chemical buoyancy as well as the combined effects of both hydration and melting. We have used up to *one billion* tracers to delineate the evolution of the lithological, rheological and deformation structure of the subduction zone with a resolution between 10 and 50 meters (GERYA and YUEN, 2003b). Tracers are integrated simultaneously with a Lagrangian energy equation and the solution is interpolated back to an Eulerian system for the solution of the thermal-chemical momentum equation. The rheologies of the different components are explicitly included. Using visualization, we have found that, contrary to prevailing wisdom, hot rising mantle flows prevail in the mantle wedge above the subducting slab, while the partially molten hydrated upwellings (wet cold plumes) in the mantle wedge are characterized by a colder thermal anomaly of 300 to 400 degrees (GERYA and YUEN, 2003b).

Visualizing one billion tracers is indeed a daunting task and a grand challenge problem. Currently there exists no single display device available for unveiling all of the minute details simultaneously. In order to address this serious obstacle in visualization, we promote the use of remote visualization. Our web-based remote solution is an image service with zooming capability that requires minimal bandwidth, allows the user to explore the data across the temporal dimension, across several thermo-physical properties, and at different spatial scales.

A 3-D configuration requires at least 100 billion tracers to achieve a resolution of several hundred meters. (Tackling this problem will also help address other problems in volcanology.) It should be noted that while it is possible to display the 2-D data in its entirety or in planar subsets without ambiguity (Fig. 4), the same cannot be said for 3-D. New techniques will be developed to create a multiscale representation of the 3-D data to automatically detect and extract relevant flow features, and to display them in a meaningful form. Unfortunately, it is highly unlikely that the networks will

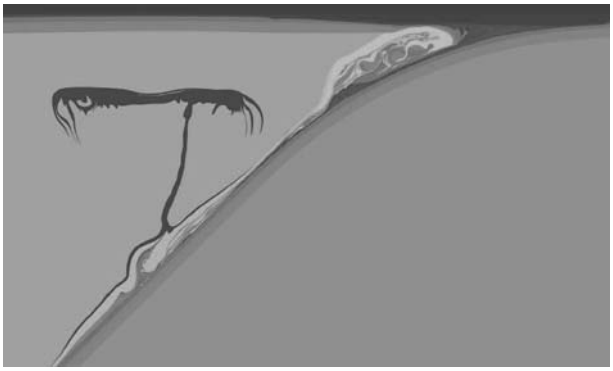


Figure 4
Compositional fields developed from subduction dynamics modeled with the tracer method (GERYA and YUEN, 2003a). The different components of the oceanic crust and lithosphere are portrayed by the various colors. Twelve components have been employed and one billion tracers have been used (RUDOLPH *et al.*, 2004). Further information can be found in (GERYA and YUEN, 2003a) and a web-interface can be found on the **WebIS** site at http://webis.msi.umn.edu.

ever become fast enough to transmit the raw data, or meaningful subsets thereof, to local sites. Rather, it will be more expedient to maintain the data at one or more storage archives and use local computational resources to reduce and display the data, returning only the 2-D bitmaps to the desktop.

4. Middleware

For the foreseeable future, there is an urgent need for *easy to use* tools to help earth scientists cull information from the onslaught of data accumulating at daily rates in the range of hundreds of Gigabytes to Terabytes. For example, the Earthscope project (http://www.earthscope.org), in particular, recognizes the value of data to facilitate better understanding of our planet, judging from the range of new experimental and computational techniques the project supports to expand data collection in the geosciences. Unfortunately, with the exception of a few sites such as the Southern California Earthquake Center, JPL and the Scripps Institution of Oceanography, geoscientists generally lack the necessary technical knowledge to tackle the challenges of remote visualization and analysis over the Internet, instead capitalizing on the burgeoning developments in grid technology. Moreover, easy to use software and transparent middleware able to handle large data sets in a *collaborative mode are still not available*. Although Fox and his group (PIERCE *et al.*, 2003; LEE *et al.*, 2003) have been actively involved in the architectural design of (SERVO 2002) through the creation of sophisticated portlet technology and XML schemas describing a variety of geoscience processes, their

systems remain point-to-point, noncollaborative, and visualization remains at best a painful process of generating single images of two-dimensional data. While many projects promise visualization tools, they remain low on their priority lists behind code access, code and metadata management, and security. Recent work has begun to investigate the use of Narada messaging for certain problems within SERVO (GADGIL *et al.*, 2005) with the objective, among several, of removing the point-to-point restriction.

Substantial analysis of the huge data sets is impeded due to insufficient network bandwidth. The promise of higher speed networks (Internet III) only postpones these problems temporarily since the rate of increase in network bandwidth appears to be substantially slower than the rate of increase of the average data set. Currently, transmitting a high-resolution microscopy image (at a resolution of $20,000 \times 20,000$, which is as dense as a high-resolution Schmidt plate image used in astronomy, over the current Internet would take at least one to two hours. It would be more efficient to break up the data into multiple files and at several resolutions, and access the subsets as needed (BEYNON *et al.*, 2000). Very often the scientists must travel to another site to optimize their productivity.

Visualization is but one of several components necessary to promote effective collaboration between distributed teams of researchers. As these teams keep growing in size and diversity, the grand challenge problems that can be addressed become increasingly complex and difficult to solve. Grids were developed to address this concern, although they are still difficult to use and often do not present the user with an easy to use, extensible interface. There are several efforts underway to develop web portals (PIERCE *et al.*, 2002) to present users with a ''simple'' mechanism to access backend resources (supercomputers, storage areas, visualization and analysis servers).

When a user accesses a grid resource, its identity must often be known in advance, following the client/server model. There is an urgent need for a more general mechanism to link client devices (hand held devices, laptop and desktop computers, and large screen displays) with servers that support a variety of services (or functionalities) necessary to the researcher. Figure 5 illustrates the basic difference between point-to-point communication and a publish/subscribe paradigm. In the latter, a client *publishes* a message, including a topic tag, to the middleware fabric, in our case Naradabrokering (http://www.naradabrokering.org). A destination point (a server, service, or other generic utility) only processes those messages tagged with a topic it has subscribed to.

We propose a middleware architecture that enables users to enter requests into the system without consideration for how the request will be executed or by whom. One or more schedulers are responsible for analyzing the completed requests and submitting them to appropriate resources for execution. The results of the completed request are eventually (faster is better) returned to the user. A successful middleware framework should support fault-tolerance, service redundancy, journaling systems,

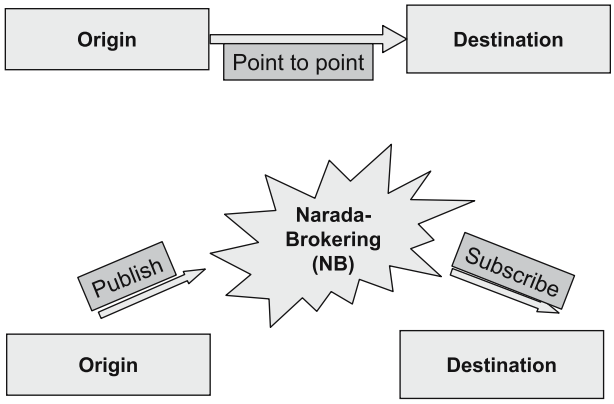


Figure 5
Top: The client (origin) must explicity define the IP address or hostname of the destination server. Bottom: The origin and destination machines are decoupled through the publish/subscribe middleware (see text).

interactivity, and most importantly, collaboration. In the area of visualization, tools are required for users to share displays. Although bitmap sharing can already be accomplished through the Virtual Network Computing API (http://www.realvnc.com), there are no generally accepted tools for two-way interaction via direct manipulation of these bitmaps. Alternative approaches are desirable, such as permitting multiple users to share an exploration session through a spatio-temporal data set. The underpinning of our architecture is the NaradaBrokering system described next.

4.1. NaradaBrokering

NaradaBrokering (NB) (http://www.naradabrokering.org) is a distributed messaging infrastructure that can be used to intelligently route data between the originators (publishers) and registered consumers (FOX and PALLICKARA, 2003; FOX and PALLICKARA, 2002a,b, 2001). Its strength lies in the complete decoupling between message source and destination. Rather than specify the precise origin and destination of a message with an IP address or any other tag that uniquely identifies the actual hardware, messages are tagged by topic. Thus, NB is a ''black box'': publishers send (i.e., publish) input messages, with an associated topic, while subscribers receive (i.e., consume) messages that correspond to any topic they subscribed to. NB manages messages, tracks topics, and routes message content to the appropriate subscribers. As a result, a given message can be recevied by several subscribers, opening the door to effective collaborative systems. NaradaBrokering has already been used to build an audio/video conferencing system (BULUT *et al.*,

2003), and is being investigated for certain aspects of the iSERVO program (GADGIL *et al.*, 2005).

The smallest unit of the messaging infrastructure that would provide a backbone for routing these messages must be able to intelligently process and route messages while working with multiple underlying network communication protocols. We refer to this unit as a *broker* to distinguish it clearly from the application *servers* that would be among the sources/links to messages processed within the system. Entities within the system use the broker network to effectively communicate and exchange data with each other. These interacting entities could be any combination of users, resources, services and proxies. They are also referred to as clients.

NaradaBrokering has a cluster-based architecture that allows the system to scale to support large client concentrations by adding new broker nodes, thus sustaining a high volume of messages between interacting entities. NaradaBrokering manages all communication complexities and optimizations pertaining to the deployment of specific network transports or to the communication across domains separated by firewall boundaries. The main challenge is to develop a topic naming scheme and standardized approaches to connect/disconnect clients, schedulers and other types of entities.

4.2. Our Solution: publish/subscribe

Our architecture will be fully web-service compliant when completed. Any task, such as file storage/retrieval, file filtering, visualization, video production, etc. can be recast as a web service with its own resources, such as a single cluster to visualize large data sets.

Our ultimate goal is to provide an infrastructure aimed at providing visualization-related services to the geophysics community with the following attributes:

1. Complete decoupling between the users who *request* services and the resources that *execute* these services.
2. Availability of multiple machines to perform visualization and computational services.
3. Support for data files stored on one of several computational platforms.
4. No single point of failure.
5. Automatic resource discovery transparent to the users.
6. User interfaces built into the leading web browsers.
7. Support of several versions of client software to accommodate users without the latest java plug-in technology.
8. Automatic scalability, a result of the built-in scalability of the NaradaBrokering middleware.
9. Language independence for clients and services.

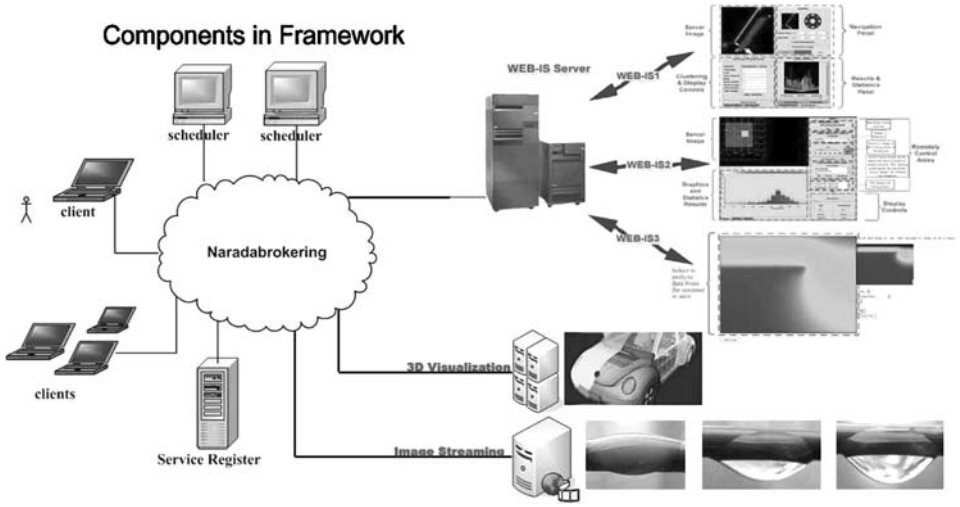


Figure 6
Example web-services connected to our proposed middleware system.

We have developed the framework for such an architecture, described in the next section.

4.3. System Components

Our system is constructed from five principal components (Fig. 6): the broker network, the clients, the services, the project managers, and the service registry/topic managers. In what follows, *entity* refers to any client/service/manager connected to NB.

1. *The Broker Network* is set up by a set of cooperating broker nodes (they define the middleware fabric) responsible for effectively organizing entity interests and routing the right message to the right set of entities.
2. *Client* entities make task requests and often display the task results through graphical user interfaces (GUIs). The client publishes a high level task description to NB along with any additional metadata necessary to complete the task (i.e., data set descriptions, commands). One consequence of the client proxy is that a SOAP message can be received by multiple subscribers, which will form the basis of future collaborative services.
3. *Resources*/*Services* within the system execute the tasks (or subtasks) requested by the clients. Resources subscribe to themselves, simplifying resource discovery. Services are typically accessed through a set of procedures described within a WSDL file, stored in a public database along with descriptive metadata. This will enable existing geoinformatics projects, such as GEON (http://www.geongrid.org), CHRONOS (http://www.chronos.org), CIG (http://www.geodynamics.org),

iSERVO (http://www.iservo.edu.au/), and ACCESS (http://www.access.edu.au), to develop their own interfaces to our services.

4. *Project Managers/Task Schedulers* offload the clients and are responsible for successful task completion. A task is a unit of work specified by a client. Sometimes, a specified task can be decomposed into several subtasks. For the most part, the client is solely interested in the correct, complete, and expedient execution of the task. For example, a volume rendering task could be executed on multiple servers and the various images composited before returning the final image to the client. The task scheduler is responsible for taking a task and splitting it up into a set of smaller subtasks. The task scheduler then constructs a dependency chain for these subtasks and decides the sequence in which these tasks need to be performed, taking concurrency into account where possible.
5. *Topic Managers and Service Registries* are fundamental to the user transparency of our architecture. Clients must discover a scheduler to manage requests on their behalf; project managers must discover resource entities to accomplish particular tasks. In a future version, the resource entities may have their own discovery requirements. We are investigating two complementary approaches, the first for predefined topics, the second for dynamically generated topics.

Entities willing to share resources subscribe to predefined topics, one for each resource type. Entities interested in locating these resources simply publish to these topics, together with credentials and other relevant information. The broker network routes the discovery request to all entities that expressed their interest through subscription. Depending on the policy maintained, on the receiving entities, and on the credentials (or the lack thereof) in the request, the sharing entity may choose to respond with the resource or information pertaining to the resource, request additional credentials from the requesting entity or not respond at all.

4.4. Flow of Information through NaradaBrokering

Figure 7 illustrates the flow of information within NB given a client who wishes to compute the factorial of some integer. This client, having no knowledge of available services, solicits the help of a scheduler that identifies one or more services capable of accomplishing the requested task, and chooses one among them based on computational load.

1. The client publishes a message to the topic ''Sch'', subscribed to by all schedulers.
2. All schedulers publish a response including their own ID (used as a topic name) and information about their current load.

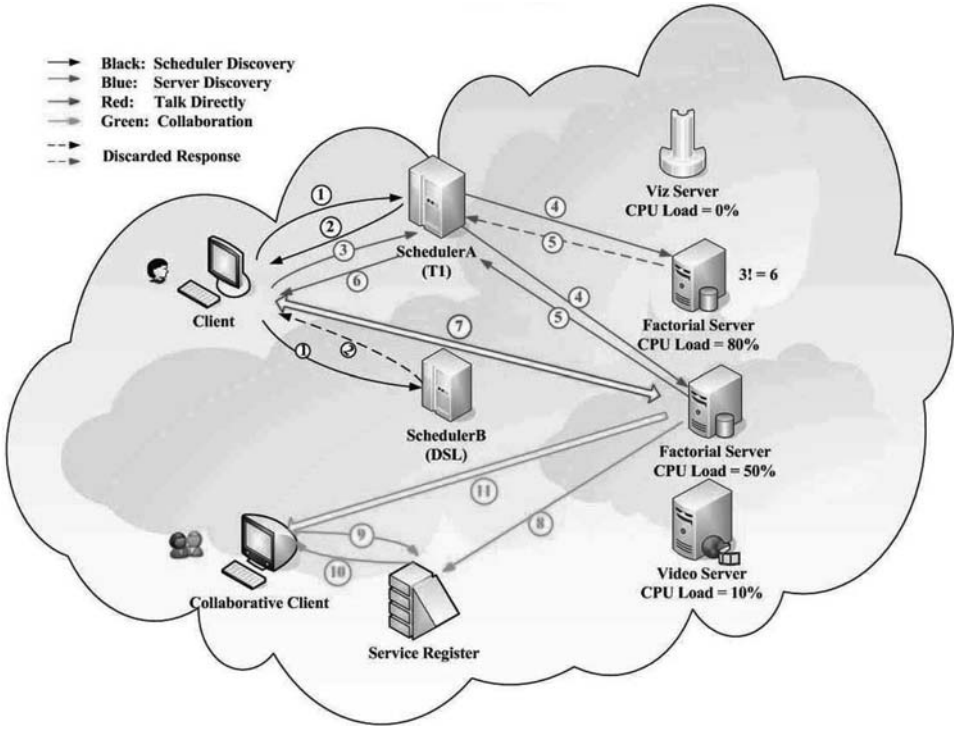


Figure 7
Flow of information between clients and web services through the control of a scheduler.

3. The client chooses one scheduler (SchedulerA) and publishes the service name (Factorial) to its ID.
4. SchedulerA publishes a message to topic "Factorial" subscribed to by all service providers that can provide the Factorial service.
5. Each service provider that provides a factorial service publishes a message that includes a unique topic ID and load information.
6. The scheduler chooses one service provider and sends its topic ID to the client.
7. The client retrieves the topic ID of the factorial service, should it be necessary for client and services to establish a direct link for reasons of efficiency (e.g., for the transfer of bitmaps). (The following steps pertain to collaborative clients):
8. The chosen service provider will register the current session to a service register and unregister itself at the end of the session.
9. The collaborative client sends a request to the service register to get the IDs of current running sessions.
10. The service register returns the list of current running sessions to the collaborative client.
11. New clients can now join the session using the server and client IDs. All clients that collaborate will, under the current setup, see identical displays.

The architecture of Figure 7 has already been implemented. We are currently working on strategies to fold web services into the framework.

4.5. Advantages of the Proposed Scheme

A system like NaradaBrokering, based on publish/subscribe has many advantages once it is coupled with the naming and discovery scheme just described:

1. No constraints are imposed on the number or location of clients, resources, task schedulers, or topic managers connected to NB.
2. The system allows for asynchronous discovery of task schedulers, available resources, resource metadata, file locations, etc.
3. Both task schedulers and resources can accept tasks based on an estimate of load.
4. The system has the ability to maintain audit trails to track system usage, task execution, and task execution failures.
5. In the event of failures, it will be possible to return partial results and the status of the computation to the client.
6. The middleware supports built-in collaboration mechanisms: Task results can be viewed by multiple users sharing one or more task IDs. Additional filtering may be required to allow for different client devices (PDA, laptop, etc.) (LEE *et al.*, 2003).
7. Task updates can be published to a "Task update" topic, perhaps with an associated task ID, that clients subscribe to separately. This would allow clients to keep track of task execution, similar to monitoring a FedEx package through a tracking number.
8. NB provides the option to cache results to improve performance in the presence of repetitive tasks.

5. Portals, Visualization Web Services

Recent efforts have focused on the creation of a web portal to enable users to interrogate their data sets stored on a remote server (GARBOW *et al.*, 2001; 2005a,b). These efforts have led to the **WebIS** system (Fig. 8). The existing framework promotes user interactivity and portability across wired and wireless networks. The challenge was to simultaneously provide users with the flexibility to explore their data while minimizing the effects of network latency. Currently, the users choose one of several precomputed slices through a data set that is downloaded to the client. After an initial delay, the user then has the ability to zoom, translate, perrform statistical analysis over a subregion and display histograms. Some magic lens technology is also implemented. A magic lens is a localized spatial region through which alternate information can be viewed. For example, thermal conductivity can be viewed through a lens that overlays the temperature field. These concepts are extended in YUEN *et al.* (2004) to include the remote visualization of earthquake clusters over the

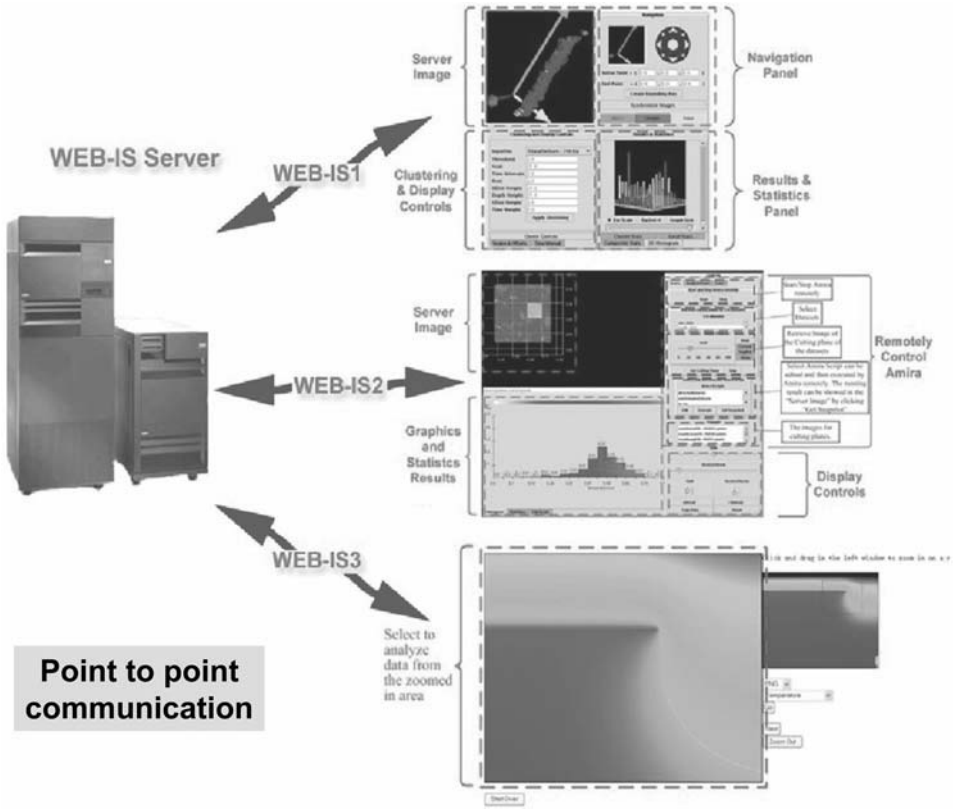


Figure 8
WebIS is a software system that allows the visualization and analysis of large data sets that reside on a server or a client. At this time, the communication between client and server is point-to-point.

network combined with off-screen rendering. This suggests the use of server farms dedicated to visualization via Web services.

Service nodes on the Naradabrokering framework described above will be populated with the server side of the **WebIS** system, along with several other visualization-related web services, such as image/video processing and movie creation. On the client side, user interfaces will be crafted and adapted to the variety of devices in use. These interfaces will be decoupled from the server side and communicate with them through web service specifications (WSDL files).

One of the most important challenges faced by these types of portals is compatibility across a range of devices. The largest displays have on the order of 10^6–10^7 pixels, while current (IPAQ, Jornada, OQO) handheld devices have much smaller resolutions, on the order of 10^4–10^5 pixels. This extreme range makes it virtually impossible to design a consistent user interface. An interface designed for a large screen will be very tedious to interact with on a small screen. On the other hand,

a user interface crafted for a handheld device wastes a lot of space on the larger screen of a workstation.

The range of networking speeds is also an issue. Our middleware will use the built-in protocols of NaradaBrokering, supplemented by compression algorithms suitable to the transmission of encoded imagery and thresholded wavelet transforms. Furthermore, while command messages will travel through the middleware system, be logged, discover available services and execute commands on them, resulting images *may* be transmitted directly to the client for reasons of efficiency.

6. Future Trends and Perspectives

There has been insufficient attention paid to the efficient analysis and facile means of visualization and display of large-scale data sets in the earth sciences. Initiatives in computing, such as iServo (AKTAS *et al*., 2005) (http://quakesim.jpl. nasa.gov), CIG (http://www.geodynamics.org) in geodynamics, and the Vlab in computational mineral physics (http://www.vlab.msi.umn.edu) will have to take a leadership role in promoting remote visualization and the practical use of allied services, such as **WebIS**.

Education is an important aspect of this program and workshops are sorely needed in areas that span the disciplines of visualization and standards-based web technologies (Java, XML, web services). By hosting several meetings, ACES has already played an important role in laying the foundation in these areas. Inroads at the national level have also been visible through several special symposiums held at the Fall American Geophysical Union meeting since 2001.

In this paper we have described a workable framework suitable for carrying out remote visualization by geographically distributed researchers. To this framework, we will connect a variety of visualization-related web services and address the issue of efficiency: Minimize communication and increase user interactivity. To move ahead, we must develop regional visualization centers, such as present at NCSA at the University of Illinois, which has recently installed large shared-memory (on the order of Terabytes) visualization facilities to enable entirely new classes of data analysis and visualization. We hope to use such powerful machines to host web services made available to the Earth Science community. These services would go a long way towards enhancing multi-way communication and exchange of data, which would be of great utility in times of disaster, such as the recent earthquake and tsunami in Asia in late 2004.

Acknowledgments

This research has been supported by the Geophysics, Advanced Computational Research, Math-Geo, Information Technology Research (VLab project: ITR-0428774) and the Advanced Computer and Infrastructure and Research programs

of the National Science Foundation, the Dept. of Energy and the Environmental Protection Agency. We would also like to thank Geoffrey Fox for useful discussions on the project.

References

Aktas, 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. (2005), *iSERVO: Implementing the International Solid Earth Research Virtual Observatory by integrating computational grid and geographical information web services*, Pure Appl. Geophys. *163*, 2281–2296.

Beynon, M., Ferreira, R., Kurc, T. M., Sussman, A., and Saltz, J. H. (2000), *Datacutter: Middleware for filtering very large scientific data sets on archival storage systems*. In *IEEE Symposium on Mass Storage Systems*, 119–134, http://citeseer.nj.nec.com/beynon00datacutter.html.

Bulut, H., Fox, G., Pallickara, S., Uyar, A., and Wu, W. (2003), *Integration of Naradabrokering and audio/video conferencing as a web service*. In *IASTED International Conference on Communications, Internet, and Information Technology 2002*.

Cohen, Ronald, E. E. (2005), *High performance computing requirements for the computational earth sciences*, Tech. Rep., Carnegie Institution of Washington, http://www.geo-prose.com/computational_SES.html.

Erlebacher, G. and Yuen, D. A. (2004), *A wavelet toolkit for visualization and analysis of large data sets in earthquake research*, Pure Appl. Geophys. 2215–2229.

Foster, I. and Kesselman, eds., *The Grid. Blueprint for a New Computing Infrastructure* (Morgan Kaufmann Publishers, Inc. (1999)).

Foster, I., Kesselman, C., Nick, J. M., and Tuecke, S. (2002), *The physiology of the grid, an open grid architecture for distributed systems integration*, Tech. Rep., Global Grid Forum, http://www.globus.org/research/papers.html.

Fox, G. and Pallickara, S. (2001), *An approach to high performance distributed web brokering*, Ubiquity *2*(38).

Fox, G. and Pallickara, S. (2002a), *An event service to support grid computational environments*, Concurrency and Computation: Practice and Experience, *1*, 1–34.

Fox, G. and Pallickara, S., *The Narada event brokering system: Overview and extensions*. In *Proceedings of the 2002 International Conference on Parallel and Distributed Processing Techniques and Applications* (PDPTA'02) (ed. H. Arabnia, *I*), pp. 353–359 (CSREA Press 2002b).

Fox, G. and Pallickara, S., *Naradabrokering: An event based infrastructure for building scaleable durable peer-to-peer grids*. In *Grid Computing: Making the Global Infrastructure a Reality* chap. 22, 579–600 (John Wiley, West Sussex, England (2003)).

Fox, G., Gannon, D., Ko, S.-H., Lee, S., Pallickara, S., Pierce, M., Qiu, X., Rao, X., Uyar, A., Wang, M., and Wu, W. (2003), *Peer-to-peer grids*. In *Making the Global Infrastructure a Reality* (ed. F. Berman, G. C. Fox, and A. J. G. Hey), pp. 471–489 (John Wiley, West Sussex, England 2003).

Gadgil, H., Fox, G., Pallickara, S., Pierce, M., and Granat, R. (2005), *A scripting based architecture for management of streams and services in real-time grid applications*. In *Proceedings of the IEEE/ACM Cluster Computing and Grid 2005 Conference (CCGrid 2005)*, to appear, http://grids.ucs.indiana.edu/ptliupages/publications/HPSearch-mgmtArch. pdf.

Garbow, Z. A., Olson, N. R., Yuen, D. A., and Boggs, J. M. (2001), *Interactive web-based map: Applications to large data sets in the geosciences*, Visual Geosciences *6*, 3, 1–14.

Garbow, Z. A., Erlebacher, G., Yuen, D. A., Boggs, J. M., and Dubuffet, F. W. (2003), *Web-based interrogation of large-Scale geophysical data sets from handheld devices*, Visual Computing. *8*, 1, 1–20.

Garbow, Z. A., Erlebacher, G., Yuen, D. A., Sevre, E. O. D., Nagle, A. R., and Kaneko, J. Y. (2002), *Web-based interrogation of large-scale geophysical data sets and clustering analysis of many earthquake events from desktop and handheld computers*, in *EOS, TRANS, AGU abstract* .

Gerya, T. V. and Yuen, D. A. (2003a), *Characteristics-based marker-in-cell method with conservative finite-differences schemes for modelling geological flows with strongly variable transport properties*, Phys. Earth Planet. Int. *140*, 295–320.

GERYA, T. V. and YUEN, D. A. (2003b), *Rayleigh Taylor instabilities from hydration and melting propel 'cold plumes' at subduction zones*, Earth and Planetary Sci. lett. *212*, 47–62.

LEE, S., KO, S., FOX, G., KIM, K., and OH, S. (2003), *A web service approach to universal accessibility in collaboration services*. In *Proceedings of 1st International Conference on Web Services, Las Vegas*.

PIERCE, M., FOX, G., YOUN, C., MOCK, S., MUELLER, K., and BALSOY, O. (2002), *Interoperable web services for computational portals*. In *Proceedings of High Performance Networking and Computing SC'02*.

PIERCE, M., YOUN, C., and Fox, G. (2003), *Interacting data services for distributed earthquake modeling*. In *International Conference on Computational Science*, pp. 863–872.

RUDOLPH, M., GERYA, T. V., YUEN, D. A., and DE ROSIER S. (2004), *Visualization of complex multiscale phenomena at subduction zones*, Visual Geosciences. SERVO (2002), Solid Earth Research Virtual Observatory, http://www.servogrid.org.

UYAR, A., PALLICKARA, S., and FOX, G. (2003), *Towards an architecture for audio video conferencing in distributed brokering systems*, Communications in Computing , 17–23.

VINCENT, A. P. and YUEN, D. A. (2000), *Transition to turbulent thermal convection beyond* $Ra = 10^{10}$ *detected in numerical simulations*, Phys. Rev. E *61*(5), 5241–5247.

YUEN, D. A., GARBOW, Z. A., and ERLEBACHER, G. (2004), *Remote Data Analysis, Visualization and Problem Solving Environment (PSE) Based on Wavelets in the Geosciences*, Visual Geosciences, Online First.

(Received December 30, 2004, revised August 10, 2005, accepted August 12, 2005)

Pure appl. geophys. 163 (2006) 2485–2495
0033–4553/06/122485–11
DOI 10.1007/s00024-006-0143-x

Pure and Applied Geophysics

Thermal Effects in the Evolution of Initially Layered Mantle Material

M. Davies,[1] H. Mühlhaus,[1,2] and L. Gross[1]

Abstract—A simplified model for anisotropic mantle convection based on a novel class of rheologies, originally developed for folding instabilities in multilayered rock (Mühlhaus *et al.*, 2002), is extended through the introduction of a thermal anisotropy dependent on the local layering. To examine the effect of the thermal anisotropy on the evolution of mantle material, a parallel implementation of this model was undertaken using the *Escript* modelling toolkit and the *Finley* finite-element computational kernel (Davies *et al.*, 2004). For the cases studied, there appears too little if any effect. For comparative purposes, the effects of anisotropic shear viscosity and the introduced thermal anisotropy are also presented. These results contribute to the characterization of viscous anisotropic mantle convection subject to variation in thermal conductivities and shear viscosities.

Key words: Anisotropy, folding, alignment, mantle.

1. Introduction

The convection of the Earth's mantle is the underlying dynamic process that drives the motion of its lithospheric plates. The convection process, in turn, is a result of the gravitational instability of material within the mantle and its temperature-dependent density. As warmer buoyant material rises, it displaces colder, denser material that sinks. Conversely, as the warmer material rises, it cools, and, as the colder material sinks, it warms and increases its buoyancy. It is through this continuous pseudo-cyclic mechanism that currents are established within the Earth that carry warmer material toward the surface, displacing cooler material to the interior.

The simulation of anisotropic mantle convection has been a specialized area where it is often assumed that the instantaneous flow patterns predicted by convection simulations could be directly related to seismic anisotropy. However, recent studies indicate that recorded observations diverge from this assumption and

[1] Earth Systems Science and Computational Centre (ESSCC), The University of Queensland, St Lucia, QLD 4072, Australia. E-mail: matt@esscc.uq.edu.au; gross@esscc.uq.edu.au

[2] CSIRO Division of Exploration and Mining, 26 Dick Perry Ave, Kensington, WA 6051, Australia. E-mail: muhlhaus@esscc.uq.edu.au

that the effect of anisotropic shearing is of a complex nature (SIMONS *et al.*, 2002). In response to these observations, a novel class of rheologies for folding instabilities in multilayered rock have been developed (MÜHLHAUS *et al.*, 2002) and then later extended (MÜHLHAUS *et al.*, 2004) to address feedback processes for large-scale flow where the effects of layer obstruction, misalignment and realignment are evident. The approach is unique in that the local material layering is treated as an independently evolving state of the convection process. This paper adopts this approach in a simple model to investigate the evolution of initially layered mantle material. The simple model is formulated here to include an anisotropic thermal conductivity.

2. The Model

The physics of mantle convection processes are typically modeled under the standard assumption that mantle material under large deformation can be compared to an incompressible, highly-viscous fluid. Based on this assumption, a viscous anisotropic mantle convection model has been developed for the study of deformation processes in multilayered geological formations. In terms of a nondimensionalized system, the general governing equations of mantle convection are given by the stress equilibrium equation and the heat conservation equation, respectively:

$$\sigma'_{ij,j} - p^{th}_{,i} + Ra^c T z_i = 0, \tag{1}$$

$$\dot{T} = \left(K_{ij} T_{,j}\right)_{,i} + \frac{Di^c}{Ra^c} \sigma'_{ij} D_{ij}, \tag{2}$$

where σ' is the deviatoric stress, p^{th} is the pressure due to thermal effects, T is the temperature, $\dot{T}$ denotes the material time derivative of the temperature, z_i is a unit vector in the opposite direction to gravity, K is the thermal conductivity tensor, Di^c is the computational dissipation number, Ra^c is the computational Rayleigh number and D is the stretching (deformation rate) tensor. As the mantle material is incompressible ($v_{i,i} = 0$), the Boussinesq approximation is adopted in the last term of (1). The Boussinesq approximation requires that the fluid bouyancy is proportional to the fluid temperature in the convective system and is implemented as a linear density variation with temperature as given by

$$\rho = \rho_0 (1 - \alpha (T - T_c)), \tag{3}$$

where α is the coefficient of thermal expansion relative to T_c, the temperature at the cold boundary. In equation (1) we assume $T_c = 0$.

2.1. Anisotropic Model Refinements

An initially layered, transverse-isotropic, viscous mantle material which is heated from below is now considered. Such a material can represent an alternating sequence

of composite heterogeneous layers of variable rheology or a weakly bonded sequence of homogeneous layers. After the onset of convection, the initially anisotropic layered mantle material mixes as it evolves. As a result, the anisotropic effects arising from the layer evolution change dynamically. In this paper we study the effects of viscous shearing with an introduced formulation for explicit thermal conductivity variations along a layer.

The anisotropic viscous shearing effect is characterized by two effective shear viscosities, denoted as η and η_S, where η_S is introduced as the shear viscosity for flow along a layer. The shear stress arising from the deviatoric stretching is consequently corrected by the 4th order anisotropy tensor Λ (see (MÜHLHAUS *et al.*, 2002) for details on the derivation of Λ) such that

$$\sigma'_{ij} = 2\eta D'_{ij} + 2(\eta_S - \eta)\Lambda_{ijkl} D'_{kl}, \tag{4}$$

where $\Lambda_{ijkl} = \frac{1}{2}\left(n_i n_k \delta_{lj} + n_j n_k \delta_{il} + n_i n_l \delta_{kj} + n_j n_l \delta_{ik}\right) - 2n_i n_j n_k n_l$ and n is a normal surface vector, or *"director"*, defining the orientation of the material layering known.

In a similar manner, we introduce a thermal anisotropic effect, characterized by two thermal conductivities κ and κ_L, where κ_L is thermal conductivity along a layer. The thermal conductivity tensor K is then composed of an isotropic part and an anisotropic correction such that

$$K_{ij} = \kappa_L \delta_{ij} + (\kappa - \kappa_L) n_i n_j. \tag{5}$$

Effects such as anisotropic flow misalignment can be captured through the use of the director as a distinct material property. As the director transforms with a material surface element in a continuum, Nanson's relation (BELYTSCHKO *et al.*, 2001) can be used to describe the evolution of the director with time as a both a translation and rotation with the velocity field. In turn, this relation can be simplified conditionally where the direction of the director is the only unknown

$$\dot{n} = -v_{j,i} n_j. \tag{6}$$

The use of this simplification requires that care is taken to ensure that the director retains unit length during computation. It is shown in (MÜHLHAUS *et al.*, 2004) that viscous slip planes tend to align with velocity vectors in steady states, and subsequently that a solution to (6) is indeed flow-aligned at steady state.

3. Computational Strategy

The domain is a square two-dimensional domain with the length of a side being equal to 1 (nondimensionalized), representative of an isolated region of the Earth's mantle. The base of the region is maintained at a constant temperature as it is heated from below. Similarly, the top surface of the region is a cold boundary that is also of

constant temperature. There is no temperature change across the vertical surfaces. The velocity, v, director, n, and temperature, T, are independent variables that are solved sequentially at each time step of the simulation using the finite-element method (FEM). As mentioned earlier, care is taken to ensure that the magnitude of the director is reset to unit length as soon as it is solved.

Substitution of (4) into (1) yields a second order linear equation for which the incompressible velocity, v, and the pressure due to thermal expansion, p^{th}, are unknown. Adopting a simple penalty method, a discretization is derived for solution of v and p^{th}at time $t + \delta t$:

$$p^{th(\alpha+1)} = p^{th(\alpha)} - Pv_{j,j}^{(t+\delta t)(\alpha)} \tag{7}$$

$$\left[(\eta(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}) + 2(\eta_S - \eta)\Lambda_{ijkl}^{(t)} + P\delta_{ij}\delta_{kl})v_{k,l}^{(t+\delta t)(\alpha+1)}\right]_{,j} - \left[p^{th(\alpha)}\delta_{ji}\right]_{,j} + Ra^c T^{(t)} z_i = 0. \tag{8}$$

Here, $P \gg \eta$ is introduced as a penalty factor.

In a similar manner, the heat equation (2) and the director evolution equation (6) are discretized through the introduction of both a backward difference approximation for the time derivative and additional terms for upwinding (upwinding is discussed in section 3.2).

For unstructured meshes, a forward-centred time step δt^+ is determined from a first-order implicit extrapolation based on the Courant condition. The scheme computes δt^+ from the average value of the maximum velocity predicted implicitly for the ensuing computation

$$\delta t^+ = \frac{\sqrt{\delta t^-(\delta t^- v_{\max}^{(t)2} + 2(v_{\max}^{(t)} - v_{\max}^{(t-\delta t)})Ch)} - \delta t^- v_{\max}^{(t)}}{v_{\max}^{(t)} - v_{\max}^{(t-\delta t)}}, \tag{9}$$

where C is an appropriate Courant number, h is the mesh discretization scale, and $v_{\max}$ is the maximum magnitude of the nodal point velocities. Due to the adaptive nature of this extrapolation scheme, larger values of C can be trialed.

3.1. Implementation

The authors[1] have designed and developed a *Python* scripting language toolkit "*Escript*" to facilitate the simulation of solid Earth processes (Davies *et al.*, 2004). *Escript* provides a scripting interface to parallelised computational kernels in order to shield a modeller from developing low-level parallelism constructs. For the results of this paper, the "*Finley*" FEM computational kernel is used to implement the finite-

[1] The *Escript* and *Finley* projects at ESSCC are led by Lutz Gross.

element method to solve the velocity, director and temperature equations. A brief overview of *Escript*, *Finley* and the FEM implementation is presented in this section.

3.1.1. Escript. At the core of *Escript* is the *Data* module. *Data* is semantically associated with an arbitrary computational domain or distribution which is implemented or managed through a computational kernel interface. Through the provision of a standard interface, the *Data* module provides the means to exchange data between various kernels while maintaining full contextual information in regard to its source and intended use. *Data* objects may be scalars, vectors, or tensors of up to 4th order.

The association of domain information with *Data* objects enables an implicit form of domain-based parallelism. As an example, the *Data* object expression `C = A + B`, where `C`, `A` and `B` are associated with the nodes of a FEM mesh, executes an addition in parallel. Composite numerical methods such as specialized nonlinear solvers and time-differencing schemes can be rapidly developed using a compact high-level domain-wide syntax with a simple implicit parallelism.

3.1.2. Finley. To use *Finley*, *Data* expressions are used to transform an initial boundary value problem (IBVP) into a sequence of linear boundary value problems (BVPs) to be solved at each time step. Each linear BVP can then be provided to *Finley* to assemble a stiffness matrix using a discretization based on the standard variational formulation (ZIENKIEWICZ and TAYLOR, 2000). The discretisation is appropriate for the mesh and element type.

For an unknown vector function u, the partial differential equations (PDEs) of a BVP (eg. (8), (6)) are provided to *Finley* through the specification of the coefficients of the following templated form in tensorial notation:

$$-(A_{ijkl}u_{k,l})_{,j} - (B_{ijk}u_k)_{,j} + C_{ikl}u_{k,l} + D_{ik}u_k = -X_{ij,j} + Y_i. \tag{10}$$

The tensorial coefficients A, B, C, D, X and Y are functions of their location in the physical domain and are supplied to *Finley* as *Data* objects or expressions. For example, to solve the velocity equation (8), a simple rearrangement into a form comparable with (10) determines that it is necessary only to provide *Finley* with A, X, and Y as *Data* objects or expressions. On the other hand, it is necessary to provide *Finley* with all of the coefficients[2] to solve the director evolution equation (6) after introducing the required upwinding terms.

To complete the BVP, *Finley* also provides forms for both Neumann and Dirichlet boundary conditions as given by the respective systems:

$$n_j(A_{ijkl}u_{k,l} + B_{ijk}u_k) + d_{ik}u_k = n_jX_{ij} + y_i \text{ on } \Gamma^N \tag{11}$$

[2] As per section 2, the coefficients (being like terms) of a modified equation are provided to *Finley* in this case.

$$u_i = r_i \text{ on } \Gamma^D \tag{12}$$

where n denotes the outer boundary normal, and A, B and X are as previously defined. Here, d and y are coefficients defined on the natural boundary Γ^N while r is a function defined on the Dirichlet boundary Γ^D. The complete linear BVP for systems of equations is then defined by (10)–(12). *Finley* also defines a steady BVP for a single equation in a scalar unknown.

3.2. Upwinding

Both the hyperbolic director evolution equation (6) and the parabolic heat equation (1000) are advection dominated at high Rayleigh numbers. To avoid the spurious oscillations that would otherwise arise during the computation of their solution, it is necessary to introduce an upwinding scheme to their discretisation. The *de facto* standard upwinding scheme is the streamline upwind Petrov-Galerkin (SUPG) method (ZIENKIEWICZ and TAYLOR, 2000), which implements a modification to the weight functions of the FEM formulation. However, as *Finley* does not permit the direct manipulation of the FEM assembly process, it is not possible to use this scheme. To address this issue, a modified PDE is derived in-place. This modified PDE is based on the original PDE and includes additional terms that eliminate the most significant advective truncation error terms for an explicit discretization[3]. For a general first-order differential expression $Lu = au + bu_{,i}v_i + c$, the modified operator is given by:

$$L \longleftarrow L - (L\beta_i)_{,i}, \tag{13}$$

where $\beta_i = \frac{h}{2}\frac{v_i}{\sqrt{v_j v_j}}$. For pure advection problems with regular quadrilateral mesh elements, this approach is equivalent to the SUPG method. The strategy of this approach can be compared to that of the Taylor-Galerkin upwinding scheme described in (ZIENKIEWICZ and TAYLOR, 2000).

4. Results and Discussion

Simulations have been run for a range of thermal and viscous anisotropies. The simulations are based on 50×50 square meshes of bilinear quadrilateral finite elements. The system is at rest initially with an initial layering perpendicular to the direction of depth and an initial temperature perturbation is given by

$$T = 1 + \frac{1}{10}\cos(\pi x_1)\sin(\pi x_2) - x_2. \tag{14}$$

[3] This approach is commonly applied in the finite-difference method.

A parameter space of $\eta_S/\eta = \{0.1, 0.5, 1.0\}$ against $\kappa_L/\kappa = \{0.9, 1.0, 1.1\}$ for $Ra^c = 10^5$ and $Di^c = 0$ was studied. A comparison of Nusselt number plots and maximum shear stresses is shown in Figure 1. The Nusselt number derived for the

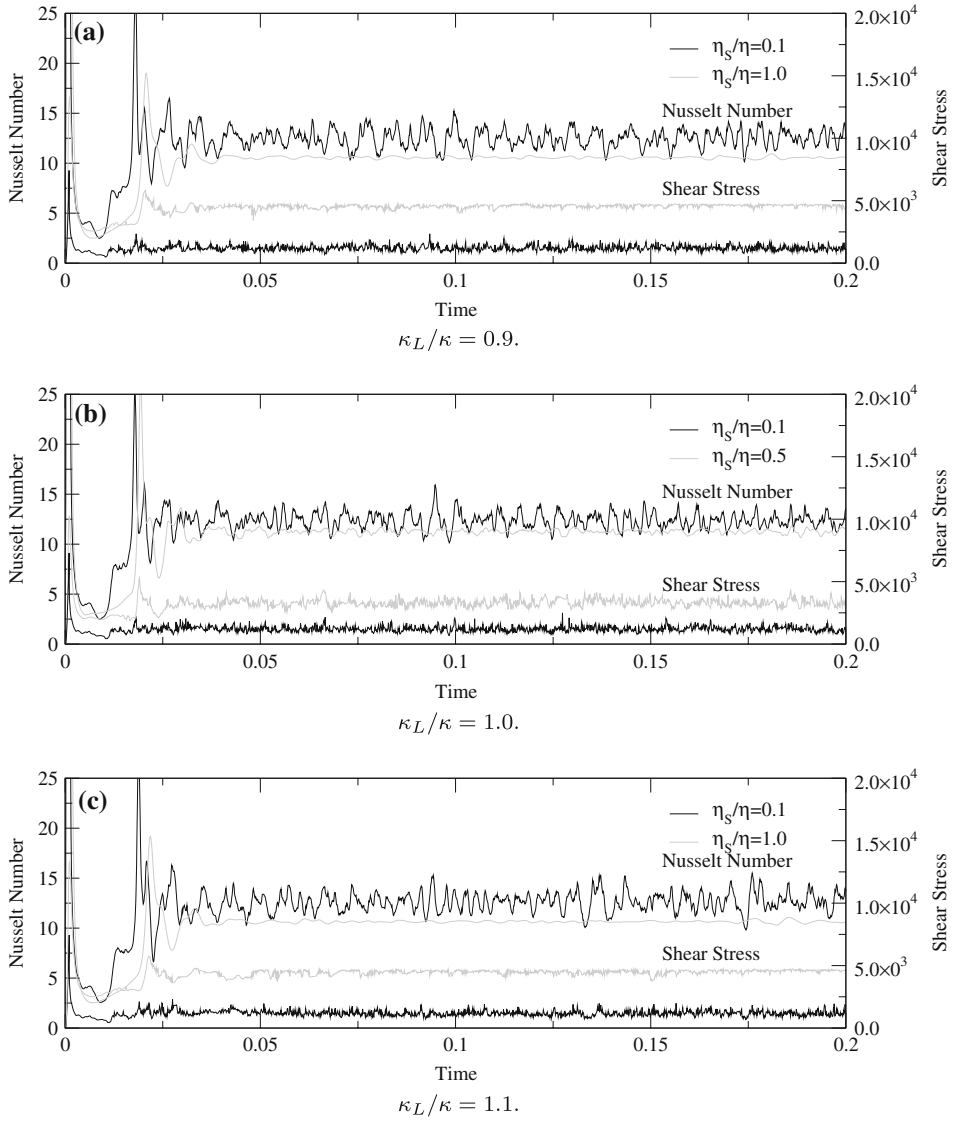


Figure 1
A comparison of Nusselt numbers for $Ra^c = 10^5$ and maximum shear stresses ($\max(\sqrt{\sigma_{ij}\Lambda_{ijkl}\sigma_{kl}})$) along the layers.

purpose of characterizing the resultant convection profiles is expressed in terms of the mechanical power

$$Nu = 1 - \frac{1}{Ra^c} \frac{\int_V v_{i,j}\sigma_{ij} dV}{\int_V K_{ij} T_{,j} z_i dV}. \tag{15}$$

A comparison of these plots illustrates that the thermal conductivity effects add an additional wavelength to the Nusselt number which appears to dampen with time. The effects also appear to have little effect on the maximum shear profile and despite the additional damping wavelength, the average Nusselt number is unchanged. It has been proposed by CHRISTENSEN (1987), that the Nusselt number for an anisotropic convection process can be compared to the Nusselt number of an isotropic convection process with average viscosity $(\eta + \eta_S)/2$. Such a comparison has been made in Figure 2. While the Nusselt number of the isotropic process exceeds marginally the running average of the original anisotropic process, it does maintain a comparable trajectory.

A comparison of the temperature field and isotherms is shown in Figure 3. at $t = 0.2$ for $Ra^c = 10^5$. It is shown that there are but subtle differences in the anisotropic temperature profiles (b), (c) and (d). A comparison of sub-figures (b) and (c) demonstrate that for $\kappa_L/\kappa = 1.1$ there is a subtle increase in the thermal gradient at the cell boundaries. Conversely, a comparison of sub-figures (b) and (d) demonstrate that for $\kappa_L/\kappa = 0.9$ there is a subtle decrease in the thermal gradient at the cell boundaries.

A comparison of the streamlines, velocity and the nonalignment field is shown in Figure 4 at $t = 0.2$ for $Ra^c = 10^5$. The nonalignment field is measured by the function $N = abs(n_i v_i/\sqrt{v_k v_k})$. While both isotropic and anisotropic convection has been reported to produce cells that encompass a mechanically-stagnant core (MÜHLHAUS *et al.*, 2004), the results obtained for this study evidenced subtle periodic adjustments

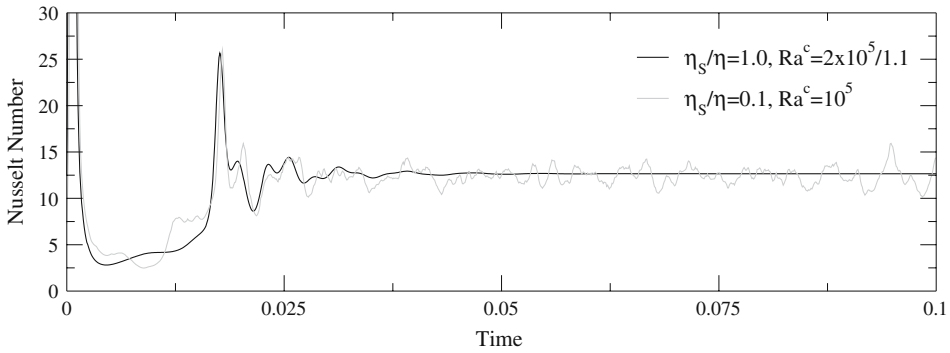


Figure 2
A comparison between the Nusselt plots of an anisotropic convection process and a similar but isotropic $(\eta_S/\eta = 1)$ convection process.

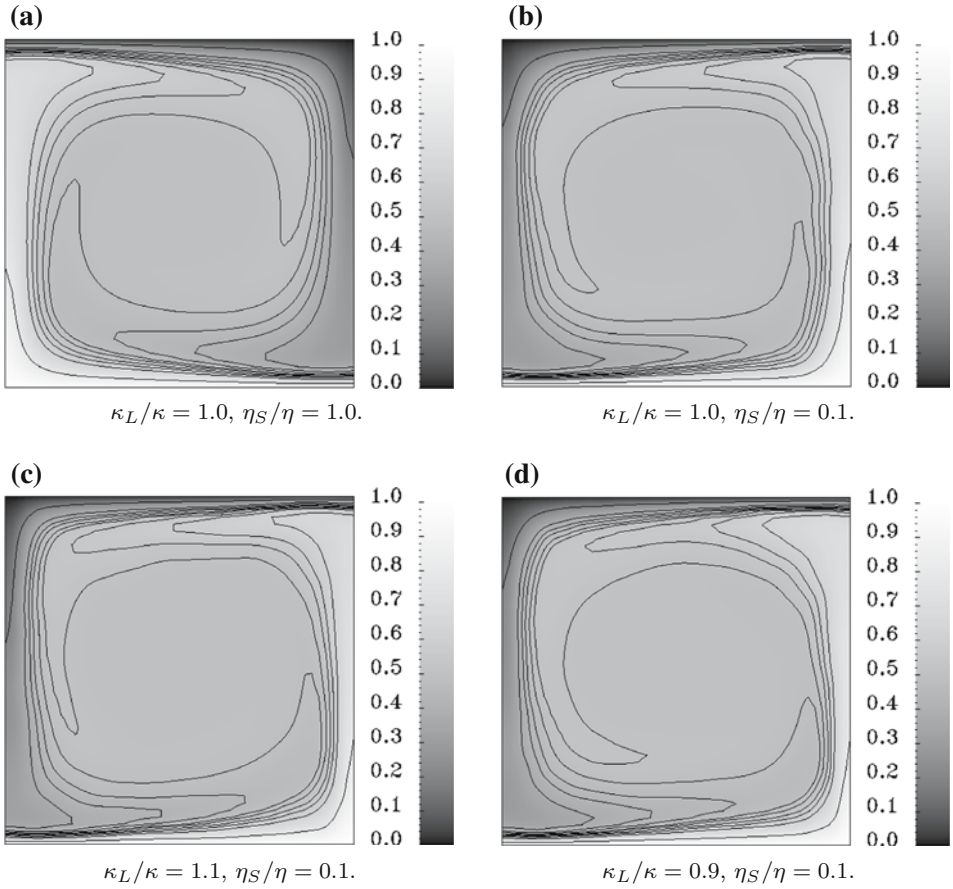


Figure 3
A comparison of temperature fields and isotherms at $t = 0.2$ for $Ra^c = 10^5$. Subfigure (a) depicts the isotropic case.

in the core position. Feedback processes of anisotropic systems appear to result in a higher degree of mixing and episodic periods of misalignment in comparison with the corresponding isotropic system. The range of simulations performed evidenced that large fluctuations in the degree of alignment were spasmodic and perhaps oscillatory. The degree of oscillation evident in the convection process for long running simulations with anisotropic shear viscosity questions the existence of a true steady state in this computational scheme.

It can be argued that the irregular oscillatory nature of the convection system as a result of anisotropic shear viscosity is both a result of feedback processes and a potential artifact of computation involving spurious discontinuities arising in director interpolation. There are two observations in support of this: (1) it has been demonstrated that the degree of oscillation decreases with mesh element size

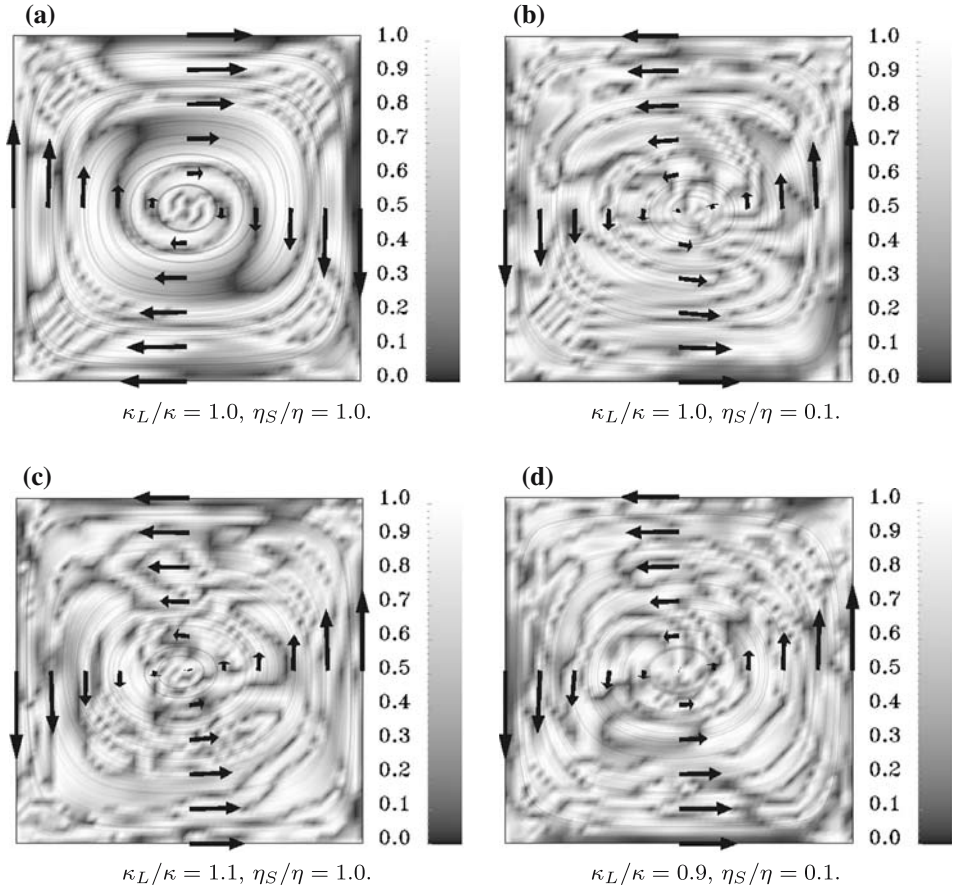


Figure 4
A comparison of streamlines, velocity and nonalignment fields at $t = 0.2$ for $Ra^c = 10^5$. Subfigure (a) depicts the isotropic case.

(particularly observable for $Ra^c = 10^6$); (2) the interaction of opposing directors for similarly-aligned folding layers at a displacement less than or equal to the mesh element size is left untreated. The first observation reflects the fact that the interpolation of directors in regions of high velocity gradient is inaccurate. The second observation reflects the need for further modeling and a more effective computational treatment. A potential solution to this problem in-part would be to evolve the directors' rotational orientation instead of their vector representation.

5. *Conclusion*

A simple model has been presented to investigate thermal conductivity effects in the evolution of initially layered material. For the presented case studies, it has been

demonstrated that the evolution of anisotropic thermal conductivity effects on once-layered mantle material appears only to have a subtle influence on the resultant temperature gradients and velocity streamlines as steady state is approached.

Further simulation in the study of both larger two-dimensional and three-dimensional domains is likely to provide a better illustration of the variation in thermal convection patterns. Finally it is noted that the computational treatment of the evolution of internal fluid surfaces in geodynamics remains a significant challenge at this time.

Acknowledgments

This work has been supported by the Australian Computational Earth Systems Simulator Major National Research Facility (ACcESS MNRF), the Queensland State Government, the University of Queensland and SGI. The Australian Commonwealth Government, participating institutions and both the Victorian and Western Australian State Governments fund the ACcESS MNRF.

References

Belytschko, T., Liu, W. K., and Moran, B., *Nonlinear finite elements for continua and structures* (John Wiley and Sons, 2001).

Christensen, U.R. (1987), *Some geodynamical effects of anisotropic viscosity*, Geophys. J. R. Astr. Soc. *91*, 711–736.

Davies, M., Gross, L., and Mühlhaus, H.-B. (2004), *Scripting high performance earth systems simulations on the sgi altix 3700*. In: Proceedings of the 7th International Conference on High Performance Computing and Grid in the Asia Pacific region, pp. 244–251.

Mühlhaus, H.-B., Moresi, L., Hobbs, B., and Dufour, F. (2002), *Large amplitude folding in finely layered viscoelastic rock structures*, Pure Appl. Geophys. *159*, 2311–2333.

Mühlhaus, H.-B., Moresi, L., and Čada, M. (2004), *Emergent anisotropy and flow alignment in viscous rock*, Pure Appl. Geophys. *161*, 2451–2463.

Simons, F.J., van der Hilst, R.D., Montagner, J.P., and Zielhuis, A. (2002), *Multimode rayleigh wave inversion for heterogeneity and azimuthal anisotropy of the Australian upper mantle*, Geophy. J. Int. *151*, 738–754.

Zienkiewicz, O.C., and Taylor, R.L., *The finite-element method* (Butterworth Heinemann, 2000), 5th ed., vol. 3.

(Received December 1, 2004, revised September 23, 2005, accepted September 25, 2005)

Pure appl. geophys. 163 (2006) 2497–2512
0033–4553/06/122497–16
DOI 10.1007/s00024-006-0142-y

Pure and Applied Geophysics

Detecting Regional Events via Statistical Analysis of Geodetic Networks

Robert Granat

Abstract—We present an application of hidden Markov models (HMMs) to analysis of geodetic time series in Southern California. Our model-fitting method uses a regularized version of the deterministic annealing expectation-maximization algorithm to ensure that model solutions are both robust and of high quality. Using the fitted models, we segment the daily displacement time series collected by 127 stations of the Southern California Integrated Geodetic Network (SCIGN) over a two-year period. Segmentations of the series are based on statistical changes as identified by the trained HMMs. We look for correlations in state changes across multiple stations that indicate region-wide activity. We find that although in one case a strong seismic event was associated with a spike in station correlations, in all other cases in the study, time period strong correlations were not associated with any seismic event. This indicates that the method was able to identify more subtle signals associated with aseismic events or long-range interactions between smaller events.

Key words: Hidden Markov models, annealing, geodesy, segmentation.

1. Introduction

In this work, we apply hidden Markov models (HMMs) to analysis of geodetic time series data. Hidden Markov models are a well-known tool that have been successfully applied to a number of problems. The HMM models observations as being generated by a discrete sequence of underlying (hidden) states. Each state depends only the previous state, and thus the sequence forms a first-order Markov chain. Changes in the statistics of the observation sequence are indicative of changes in the underlying state. Fitting a model to the observation data results in an estimate of the underlying state sequence, allowing classification of observations according to associated state, as well as an estimate of the model statistics.

Hidden Markov models have been used most prominently in the fields of speech synthesis and recognition (continuous output HMMs), and protein matching and

Jet Propulsion Laboratory, California Institute of Technology, 4800 Oak Grove Dr., Pasadena, 91109 CA. E-mail: granat@jpl.nasa.gov

analysis (discrete output HMMs). In these domains the difficult nonlinear optimization problem of fitting the model to the observation data has primarily been addressed by the addition of explicit and implicit constraints that act to reduce the number of free model parameters. These methods include restrictions on the form of the state-to-state transition probability matrix (JUANG and RABINER, 1985; FARAGO and LUGOSI, 1989; MCGUIRE *et al.*, 2000), restrictions on the form of the output distributions (EPHRAIM *et al.*, 1989), and parameter tying (BELLEGARDA and NAHAMOO, 1990; YOUNG and WOODLAND, 1994; BOCCHIERI and MAK, 2001). These constraints are supported by extensive knowledge about the underlying system being modeled. For instance, in speech analysis, we know not only the rules of language that govern the ordering of sounds and words (LEE and HON, 1989; LEE, 1990) but also the details of the actual physical process which generates sound waves (JUANG and RABINER, 1991).

In the analysis of geodetic time series such constraints are not as readily available. To address this difficulty, we use an alternative optimization approach that uses deterministic annealing and statistics-based regularization. The usual approach to HMM optimization is to use the expectation-maximization (EM) algorithm (DEMPSTER *et al.*, 1977). The EM algorithm has many good properties and works well in practice for many problems, but only guarantees convergence to a local maxima (the local maxima found are dependent on the initial conditions). When there are a large number of local maxima, repeated applications of the method can result in very different answers. The annealing portion of our method addresses this problem by guiding the solution towards strong local maxima that emerge first from the optimization surface as the problem is "cooled." The regularization portion of our work acts in a complimentary fashion, by pushing the solution away from solutions with redundant states. The method can be demonstrated to work by empirically measuring the number of local maxima encountered for different random initializations. A comparison with the standard EM algorithm reveals that the number of local maxima found is dramatically reduced with no loss in solution quality.

We apply this method to time series of displacement measurements collected by the SCIGN (Southern California Integrated Geodetic Network) array of GPS stations. By training an HMM on an individual time series, we can classify the displacement observations in the series according to the estimate of the underlying state. This allows us to objectively segment the series into different behavioral modes. By repeating this process for all stations in the network, we can gain perspective of the overall activity in the region. In particular, we can detect regional events by looking for correlations between state changes at different stations at a given point in time. We present the results of this method used in a two-year study of GPS measurements and compare it against the seismic record during that same period.

2. Hidden Markov Models

We begin with a short review of hidden Markov models (HMMs). A hidden Markov model is a statistical model for ordered data. The observed data are assumed to have been generated by an unobservable statistical process of a particular form. This process is such that each observation is coincident with the system being in a particular state. Furthermore it is a first-order Markov process: The next state is dependent only on the current state. The model is completely described by the initial state probabilities, the first-order Markov chain state-to-state transition probabilities, and the probability distributions of observable outputs associated with each state.

A. Notation

Our notation is similar to that employed by RABINER (1989) and is as follows: A hidden Markov model λ with N states is composed of a vector of initial state probabilities $\pi = (\pi_1, \ldots, \pi_N)$, a matrix of state-to-state transition probabilities $A = (a_{11}, \ldots, a_{ij}, \ldots, a_{NN})$, and the observable output probability distributions $B = (b_1, \ldots, b_N)$. The observable outputs can be either discrete or continuous. In the discrete case, the output probability distributions are denoted by $b_i(m)$, where m is one of M discrete output symbols. In the continuous case, the output probability distributions are denoted by $b_i(y, \theta_{i1}, \ldots, \theta_{ij}, \ldots, \theta_{iM})$ where y is the real-valued observable output (scalar or vector) and the θ_{ij} are the parameters describing the output probability distribution. For the normal distribution we have $b_i(y, \mu_i, \Sigma_i)$. An observation sequence O of length T is denoted $O_1 O_2 \cdots O_T$ and a state sequence Q of the model is denoted $q_1 q_2 \cdots q_T$.

B. Model Optimization Problem

In this section we concentrate on maximizing the likelihood of the observation sequence given the model, $P(O|\lambda)$; this is the *maximum likelihood* objective function. Many other objective functions have also been proposed for hidden Markov models, including the state-optimized joint likelihood for the observations and underlying state sequence (JUANG and RABINER, 1990), maximum mutual information (MMI) (BAHL *et al.*, 1986), minimum discrimination information (MDI) (EPHRAIM *et al.*, 1989), and maximum classification error (MCE) (CHOU *et al.*, 1994). Of these, all but the first require labeled training examples on which to train the models, making them inappropriate for our target application domain. The first, used as the basis for the so-called "segmental K-means" algorithm, suffers from similar initialization-dependent local maxima issues as does the more common maximum likelihood criteria, and so we skip an independent analysis of it in this work.

For the series of observations $O = O_1 O_2 \cdots O_T$, we consider the possible model state sequences $Q = q_1 q_2 \cdots q_T$ to which this series of observations could be assigned.

For a given fixed state sequence Q, the probability of the observation sequence O is given by

$$P(O|Q,\lambda) = \prod_{t=1}^{T} P(O_t|q_t,\lambda). \tag{1}$$

Assuming statistical independence of observations,

$$P(O|Q,\lambda) = b_{q_1}(O_1)b_{q_2}(O_2)\cdots b_{q_T}(O_T). \tag{2}$$

The probability of the given state sequence Q is

$$P(Q|\lambda) = \pi_{q_1}a_{q_1q_2}a_{q_2q_3}\cdots a_{q_{T-1}q_T}. \tag{3}$$

The joint probability of O and Q is the product of the above, so that

$$P(O,Q|\lambda) = P(O|Q,\lambda)P(Q|\lambda), \tag{4}$$

and the probability of O given the model is obtained by summing this joint probability over all possible state sequences Q:

$$P(O|\lambda) = \sum_{\text{all } Q=q_1q_2\cdots q_T} \pi_{q_1}b_{q_1}(O_1)a_{q_1q_2}b_{q_2}(O_2)\cdots a_{q_{T-1}q_T}b_{q_T}(O_T). \tag{5}$$

We can pose the optimization of $P(O|\lambda)$ as a nonconvex optimization problem, often presented in terms of the equivalent problem of maximizing the log *likelihood* $\log P(O|\lambda)$. The most common method for solving this problem is the EM algorithm, although alternative approaches exist, such as those employing genetic algorithms (KWONG *et al.*, 2001), recursive predictive error techniques (COLLINGS *et al.*, 1994), or gradient projection (HUO and CHAN, 1993).

3. Expectation-Maximization

We can pose the EM algorithm generally as follows: We wish to maximize a likelihood $P(\lambda)$ where λ is a set of model parameters. Given $p(x,\lambda)$, a positive real-valued function on $x \times \Lambda$ measurable in x for fixed λ with measure μ, we define

$$P(\lambda) = E[p(x,\lambda)|\lambda] = \int_{\mathcal{X}} p(x,\lambda)\, d\mu(x) \tag{6}$$

and

$$Q(\lambda,\lambda') = E[\log p(x,\lambda')|\lambda] = \int_{\mathcal{X}} p(x,\lambda)\log p(x,\lambda')\, d\mu(x), \tag{7}$$

where λ' is also a set of model parameters on Λ. Here x is the so-called *hidden variable*, while $p(x,\lambda)$ is often referred to as the *complete data likelihood*. The function

Q is often referred to as the Q *function*. Note that the function p may be a function of observable outputs y as well as the parameters of the model λ, so we have $p(x, y, \lambda)$. In this case, the integrals are over $\mathcal{X} \to \mathcal{Y}(\mathcal{X})$.

Assume $Q(\lambda, \overline{\lambda}) \geq Q(\lambda, \lambda)$ for some set of model parameters $\overline{\lambda}$. Then $P(\overline{\lambda}) \geq P(\lambda)$.

Proof:

$$\begin{aligned}
\log P(\overline{\lambda})/P(\lambda) &= \log \int_{\mathcal{X}} p(x, \overline{\lambda})\, d\mu(x)/P(\lambda) \\
&= \log \int_{\mathcal{X}} [p(x, \lambda)\, d\mu(x)/P(\lambda)] p(x, \overline{\lambda})/p(x, \lambda) \\
&\geq \int_{\mathcal{X}} [p(x, \lambda)\, d\mu(x)/P(\lambda)] \log[p(x, \overline{\lambda})/p(x, \lambda)] \\
&= (P(\lambda))^{-1}[Q(\lambda, \overline{\lambda}) - Q(\lambda, \lambda)] \geq 0.
\end{aligned}$$

From this we can show that for a transformation $\mathcal{F}$ that if $\mathcal{F}(\lambda)$ is a critical point of $Q(\lambda, \lambda')$ as a function of λ', then the fixed points of $\mathcal{F}$ are critical points of P. This gives us the EM algorithm:

1. Start with $k = 0$ and pick a starting $\lambda^{(k)}$.
2. Calculate $Q(\lambda^{(k)}, \lambda)$ (expectation step).
3. Maximize $Q(\lambda^{(k)}, \lambda)$ over λ (maximization step). This gives us the transformation $\mathcal{F}$.
4. Set $\lambda^{(k+1)} = \mathcal{F}(\lambda^{(k)})$. If $Q(\lambda^{(k+1)}, \lambda) - Q(\lambda^{(k)}, \lambda)$ is below some threshold, stop. Otherwise, go to step 2.

Note that this method is inherently sensitive to the initial conditions $\lambda^{(0)}$, and only guarantees eventual convergence to a local maxima of the objective function, not the global maximum. Nevertheless, it is widely used in practice and often achieves good results.

For the hidden Markov model, we have the complete data likelihood

$$p(Q, O, \lambda) = \pi_{q_1} b_{q_1}(O_1) a_{q_1 q_2} b_{q_2}(O_2) \cdots a_{q_{T-1} q_T} b_{q_T}(O_T), \tag{8}$$

with $P(\lambda) = E[p(q, O, \lambda)|\lambda]$ defined as in (5). The forward-backward method suggested by BAUM and COLLEAGUES (BAUM, 1972; BAUM and EGON, 1967; BAUM and PETRIE, 1966; BAUM *et al.*, 1970; BAUM and SELL, 1968) can be used to efficiently calculate each step in the EM algorithm.

4. Deterministic Annealing

Deterministic annealing is a technique based on the principles of statistical mechanics that can be used to modify the EM method to mitigate its inherent

sensitivity to initial conditions. Deterministic annealing uses the principle of maximum entropy to specify an alternative posterior probability density for the hidden variables; this allows us to define a new effective cost function depending on the temperature that is analogous to the thermodynamic free energy. Maximization of the likelihood at a given temperature is achieved via minimization of this cost function. Deterministic annealing differs from simulated annealing (KIRKPATRICK *et al.*, 1983), in which a stochastic search is performed on the energy surface, in that the cost function is deterministically optimized at each temperature.

Use of deterministic annealing has been proposed for vector quantization (ROSE *et al.*, 1992) and for clustering problems (BUHMANN and KUHNEL, 1993; WONG, 1993). Yuille and colleagues (YUILLE *et al.*, 1994) showed that the EM algorithm can be used in conjunction with deterministic annealing. Recently the deterministic annealing technique has been applied to a variety of problems (ROSE, 1998). The particular framework we present here was first applied by Ueda and Nakano to mixture density estimation problems (UEDA and NAKANO, 1994) and then extended to the general case (UEDA and NAKANO, 1998), and involves a reformulation of the EM algorithm so that it incorporates deterministic annealing.

In effect, the method involves optimizing over a series of smoothed approximations to the original objective function. By slowly increasing the computational temperature parameter γ (inverse temperature), the effect of each observation is gradually localized. At $\gamma = 1$, the parameterized Q function is equivalent to the original Q function for the problem. We start the algorithm at a $\gamma_{\min}$ such that the modified objective function has a single maximum in λ. We thereafter assume that at each new γ, the global maximum of the new objective function is close to that of the previous, so that the method tracks the global maximum as γ increases. In cases where this assumption does not hold true, the method will fail to track the global optimum.

Our application of the deterministic annealing method to HMM optimization is similar to that presented by ROSE and RAO (2001) but differs from it in some important respects. First, it is not a supervised training method, and optimizes the likelihood rather than the minimum classification error. Second, it employs EM rather than gradient descent at each temperature. Our method is described in full in GRANAT and DONNELLAN (2001) but can be summarized as follows: On the kth iteration at each temperature we optimize over the function

$$U(\gamma, \lambda|\lambda^{(k)}) = \sum_{i=1}^{N} \tau_{i1}^{(k)}(\gamma) \log \pi_i + \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T-1} \tau_{ijt}^{(k)}(\gamma) \log a_{ij} + \sum_{i=1}^{N} \sum_{t=1}^{T} \tau_{it}^{(k)}(\gamma) \log b_i(O_t), \quad (9)$$

where $\tau_{i1}^{(k)}(\gamma)$ and $\tau_{ijt}^{(k)}(\gamma)$ are estimates of the state assignment and the transition probabilities calculated via a modified forward-backward algorithm in which all probabilities are raised to the power γ.

5. Regularization

As was observed by WHILEY and TITTERINGTON (2002), many local maxima of the deterministic annealing EM method are located where the states are under-utilized, in other words where $b_i = b_j$ for some $i \neq j$. Our approach is design regularization terms that act to push the optimization procedure away from these parts of the parameter space. The approach we present here was first applied to geodetic time series data in previous work (GRANAT, 2003), but without the use of deterministic annealing.

In general when applying regularization terms it is convenient to work directly with the so-called Q function for the HMM which is maximized during each EM iteration:

$$Q(\lambda, \lambda^{(k)}) = \sum_{i=1}^{N} \tau_{i1}^{(k)} \log \pi_i + \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{t=1}^{T-1} \tau_{ijt}^{(k)} \log a_{ij} + \sum_{i=1}^{N} \sum_{t=1}^{T} \tau_{it}^{(k)} \log b_i(O_t). \quad (10)$$

Since this is separable in π, A, and B, we can divide this into the sum of three functions, $Q_1(\pi)$, $Q_2(A)$, and $Q_3(B)$ that can be maximized individually. Since we are interested in the output distributions we concentrate our attention on the last of these.

No general regularization term exists to assist in avoiding the condition where $b_i = b_j$. However, for particular forms of the output distribution regularization terms can be devised. For example, for Gaussian output distributions, we can add a regularization term based on the squared Euclidean distance:

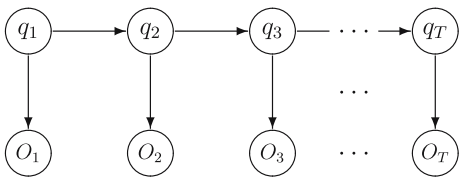


Figure 1
A representation of the hidden Markov model, with hidden nodes in underlying system states q, and observable variables O.

$$Q_3' = \sum_{i=1}^{N}\sum_{t=1}^{T}\tau_{it}^{(k)}\left(\log n - \frac{1}{2}\log\det(\Sigma_i) - \frac{1}{2}(m_i-\mu_i)^T\Sigma_i^{-1}(m_i-\mu_i)\right.$$
$$\left. -\frac{1}{2}(O_t-m_i)^T\Sigma_i^{-1}(O_t-m_i) + \frac{\omega_{Q_3}}{2}\sum_{j=1}^{N}(\mu_i-\mu_j)^T(\mu_i-\mu_j)\right). \quad (11)$$

To find the means and covariances we solve the simultaneous equations

$$\left(N\omega_{Q_3}I_{N^2\times N^2} + \omega_{Q_3}\begin{bmatrix} I_{N\times N} & \cdots & I_{N\times N} \\ \vdots & \ddots & \vdots \\ I_{N\times N} & \cdots & I_{N\times N}\end{bmatrix}\right)U = \begin{bmatrix}\Sigma_1^{-1} & & \\ & \ddots & \\ & & \Sigma_N^{-1}\end{bmatrix}M, \quad (12)$$

and

$$\Sigma_i = \frac{\sum_{t=1}^{T}\tau_{it}^{(k)}(O_t-\mu_i)(O_t-\mu_i)^T}{\sum_{t=1}^{T}\tau_{it}^{(k)}}. \quad (13)$$

We note that for sufficiently large values of ω_{Q_3} equation (11) is no longer concave in the means and covariances, thereby invalidating the M-step of the EM algorithm. To address this, we require that $\omega_{Q_3} \leq ||\Sigma_i^{-1}||/2N$ at each iteration; this guarantees concavity of the modified Q function. This is the method we use to obtain the results in this work.

6. Single Station Results

In this section, we present some results of using the combined deterministic annealing and regularization techniques to train hidden Markov models on a GPS time series collected by a station in Claremont, California.

This data set, which we designate *clar*, consists of relative displacement measurements in three dimensions (north-south, east-west, and vertical) collected daily over about two years spanning 1998–1999. We choose this particular data set because it contains certain clear signals of deformation processes, thereby providing some measure of ground truth against which we can evaluate models fit to this data. Figure 2 shows this data set; note the slow, recovering displacement around days 100–200 and the sudden east-west jump on day 626. The former is the result of ground water pumping and subsequent refilling of a local aquifer, the latter is an effect of the 1999 Hector Mine earthquake (magnitude 7.1).

Our regularization scheme was based on the squared Euclidean distance as described in the preceding section. Instead of choosing a particular set value for ω_{Q_3} we set the value to the upper bound throughout the optimization procedure. That is, at each iteration $\omega_{Q_3} = \min_i ||\Sigma_i^{-1}||/2N$. We note that because of this recalculation of

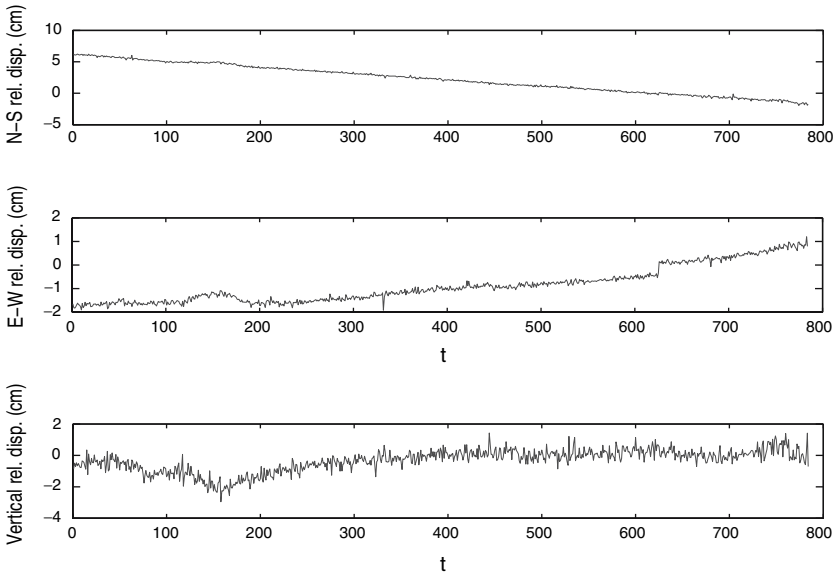


Figure 2
The data set *clar*, collected by a SCIGN GPS station in Claremont, California.

the regularization weight, our procedure is in fact not a true EM optimization method. However, our implementation does require that the log likelihood function decrease at every iteration and so our procedure satisfies the requirements of a generalized expectation-maximization (GEM) method, guaranteeing convergence to a local maxima.

To evaluate the effectiveness of our method in avoiding local maxima, we need to determine which solutions correspond to different local maxima. Our approach is to compare the individually most likely state assignments for the observation sequence (i.e., the classification results):

$$q_t = \arg\max_{1 \le i \le N}(\tau_{it}),\ t = 1, \ldots, T, \tag{14}$$

where τ_{it} is an estimate of the probability of being in state i at time t, given the observation sequence and the model. We use a linear assignment method based on bipartite graph matching (FORD and FULKERSON, 1956) to resolve equivalent state permutations. Solutions with different state sequence assignments are considered to be different maxima. This means that models that produce identical classification sequences are considered to be the same local maxima, even if the model parameters are not identical. To determine the number of maxima found by an algorithm when applied to a particular data set, we can run repeated trials with uniform random initializations of the model parameters and count the number of different solutions based on this criterion. While this method does not guarantee identification of all

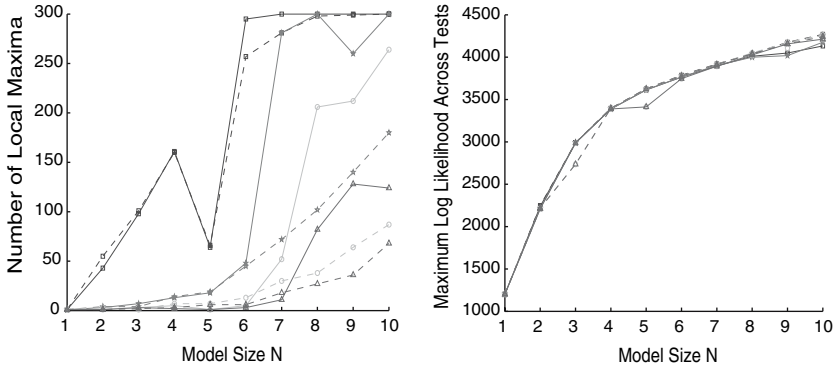


Figure 3
Left: Number of experimentally determined local maxima for HMMs with varying numbers of hidden states applied to the data set *clar*. *Right*: Maximum log likelihood among all experiments for HMMs with varying numbers of hidden states applied to the data set *clar*. Squares show results for the baseline HMM with standard EM optimization; Circles results with schedule $\Delta\beta = 0.01$; Triangles results with schedule $\Delta\beta = 0.001$. Dashed lines are the results with deterministic annealing only, solid lines are the results of the combined annealing and regularization technique.

local maxima, we can have confidence in the results if after some number of tests the number of identified local maxima fails to increase.

In Figure 3 we present results of the method as applied to the data set *clar*. We note that the combined method has fewer local maxima than both the standard EM approach and the deterministic annealing alone for all three annealing schedules. In fact, we observe that for the combined method at the slowest annealing schedule there are three or fewer maxima for $N = 1, \ldots, 6$ and only eleven solutions for $N = 7$, three of which comprise 92% of the experimental results. However, after this point there is an abrupt rise in the number of experimentally determined local maxima. We propose that this rise is due to the fact that we have exceeded the true number of classes in the data set: Since the combined method acts to reduce the number of redundant maxima, if we exceed the true number of maxima in the data set, then we expect radically worse results as the method forces the existence of additional, distinct classes. Figure 5 displays the majority classification result of the combined method for $N = 7$. We see that the method has identified all the major modes of the system including not only the before and after Hector Mine earthquake states and the water pumping signal, but also a number of more subtle signals.

7. Multiple Station Results

In the preceding section we presented results of our method applied to a single SCIGN GPS station in Claremont, California. Here we are interested in detecting

geophysical events with geographically disperse signatures and therefor wish to use the entire network. As background to our study we note that while earthquake events are of course of considerable interest, recently the geophysics community has become interested in aseismic events linked to crustal block motion or stress transfer between earthquake faults. These types of events have been observed in a few instances (MELBOURNE and WEBB, 2002, 2003; ROGERS and DRAGERT, 2003; MELBOURNE *et al.*, 2002; MILLER *et al.*, 2002; HIROSE *et al.*, 1999; HEKI *et al.*, 1997), but detections remain rare due to the subtlety of the signals. We hope to observe evidence of not only seismic but also aseismic events in the SCIGN data.

To do this, we extract GPS signals from all 127 available stations in a 820-day window. When GPS displacement values for a given station are not available on a particular day due to signal dropout or incomplete installation, we assume a zero displacement measurement for that day at that station. We note that since actual measurements are almost never of zero displacement, this in effect adds an additional "dropout" class to the data. Our next step is to train N-state hidden Markov models on each of these GPS signals. Since the GPS signals have similar statistics to one another, we can use the results of our experiments on the data set *clar* to estimate the model size. We see that there were very few maxima for $N < 7$ and only eleven local maxima for $N = 7$, rising rapidly after that. So we can guess that a good number of states to use would be in the range of 5–7, with an additional state added to account for the dropout class. Once all models of a particular size have been trained on each of the GPS time series, we can use the models to perform state assignments of each

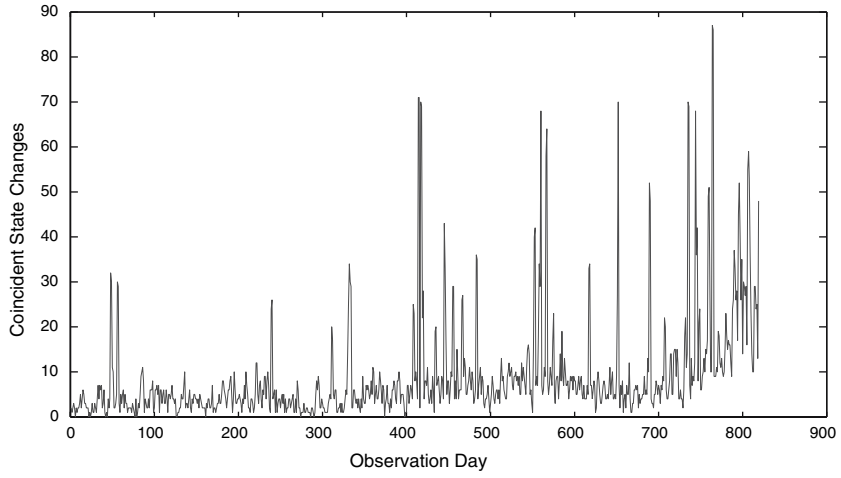


Figure 4
Coincident state changes for six-state HMMs trained on signals from each of 127 SCIGN GPS stations.

observation. We suspect that interesting geophysical events will manifest themselves as changes in the signals across multiple GPS stations, so we look for correlations in state changes across the network.

Figure 4 shows the number of coincident state changes across all observation days with training done with six-state models. We see that there are a number of strong peaks indicating correlated state changes. Of note is the strong peak on day 652, which corresponds to the Hector Mine earthquake (visible as an E-W displacement jump in the *clar* data). We also observe that there is an increasing trend in the average number of coincident state transitions; this is because of the increasing number of stations installed and activated during the observation period.

In Figure 6 we compare the results of using the baseline EM algorithm (blue) for training the HMMs used in this study against the results of using the regularized deterministic annealing EM training (red). We see that the noise level in the coincident state transition signal is significantly reduced by employing the latter method. We compare the coincident state changes against the earthquake record during the same time period in Figure 7. We see that correlations across the GPS network (blue) are only strongly correlated with an earthquake event (red) in the case of the aforementioned Hector Mine earthquake. There are no other strong earthquakes in the time window studied. The implication of this is that the regional activity indicated by the state transition correlations is either an aseismic effect or the result of subtle long-range interactions between small (magnitude $<= 4.0$) events.

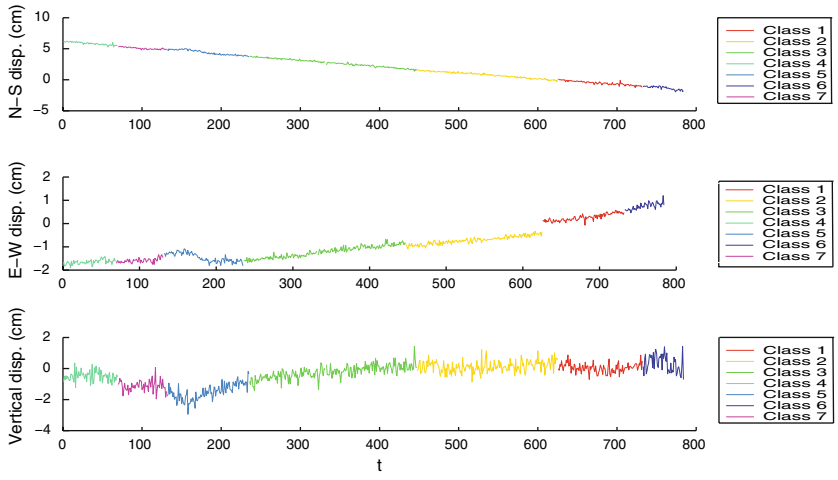


Figure 5
Classification results for a seven-state HMM applied to the data set *clar*.

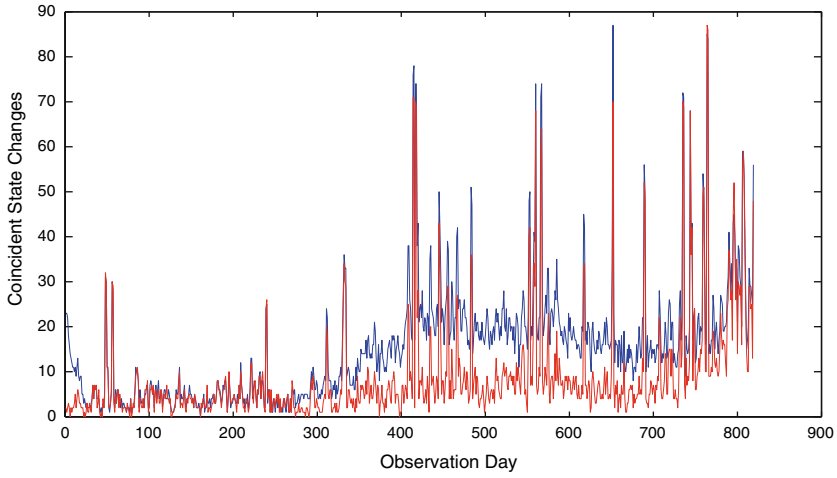


Figure 6
Coincident state changes for six-state HMMs trained using standard EM (blue) and regularized deterministic annealing EM (red) on signals from each of 127 SCIGN GPS stations.

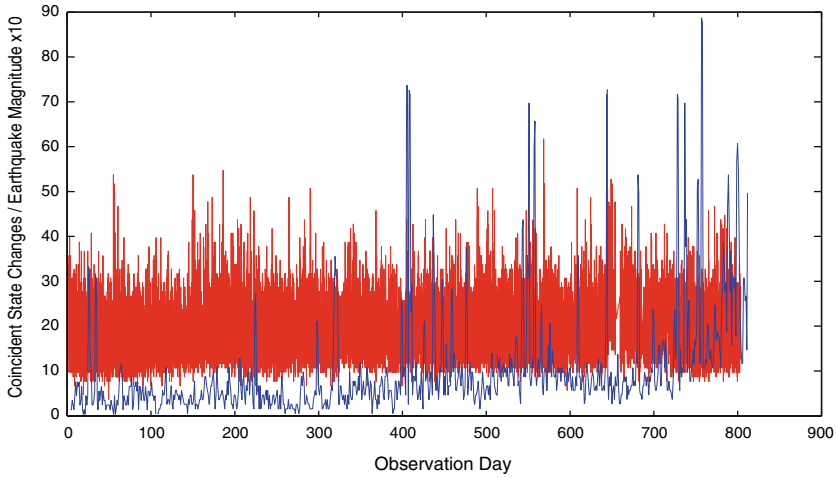


Figure 7
Comparison of coincident state changes for six-state HMMs trained using the regularized deterministic annealing EM (blue) and the Southern California earthquake record (red). Earthquake magnitudes, exaggerated by a factor of 10 for visibility, are presented on the vertical axis.

8. *Conclusions*

We have presented a method that uses deterministic annealing and regularization to modify the standard expectation-maximization (EM) method for fitting hidden Markov models (HMMs). We show that for typical geodetic time series the method

greatly improves the robustness of the solution, as measured by the propensity to converge to different local maxima given random initial conditions. At the same time, it still preserves solution quality as indicated by the solution log likelihood measure and by comparison with ground truth as identified by domain experts.

Application of this method to training HMMs on displacement time series data collected by the Southern California Integrated Geodetic Network (SCIGN) enabled the reliable statistical segmentation of 127 of those time series over approximately a two-year span. Comparing the timing of state changes across all of the stations, we discovered that large correlations between multiple stations were found at particular points in time. In only one case, that of the Hector Mine earthquake, were these correlations found to be connected with a large seismic event. We can be confident that these correlations are not the result of algorithm artifacts because each station is evaluated independently. Furthermore, in many cases we can eliminate transient effects, such as weather or network glitches, because transient effects should result in a characteristic double spike. This is because under such circumstances the effected stations switch into the anomalous state, but then switch quickly back. The implication, therefore, is that in the remaining cases these jumps in station correlations are indicative of aseismic activity or of more subtle interactions between smaller scale events.

Further study will involve extending the analysis to longer time series and inclusion of data collected by all of the more than 250 SCIGN GPS stations. In addition, correlation spikes will be analyzed in more detail to determine if the correlated stations can be associated with any particular crustal block motion or particular fault interactions. Extensions of the method include use of techniques such as generalized conjugate-gradient acceleration (JAMSHIDIAN and JENNRICH, 1993) to speed up solution convergence, particularly in flat portions of the objective function, as well as combination with Kalman filter type approaches to estimate continuous state trajectory dynamics.

Acknowledgements

This research was carried out at the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration.

REFERENCES

BAHL, R., BROWN, P., SOUZA, P.D., and MERCER, R. (1986), *Maximum mutual information estimation of hidden Markov model parameters for speech recognition*, Proc. 1986 IEEE Int. Conf. Acoustics, Speech, and Signal Processing pp. 49–52, Tokyo.

BAUM, L.E. (1972), *An inequality and associated maximization technique in statistical estimation for probabilistic functions of Markov processes*, Inequalities *3*, 1–8.

BAUM, L.E. and EGON, J.A. (1967), *An inequality with applications to statistical estimation for probabilistic functions of a Markov process and to a model for ecology*, Bull. Amer. Math. Soc. *73*, 360–363.

BAUM, L.E. and PETRIE, T. (1966), *Statistical inference for probabilistic functions of finite state Markov chains*, Ann. Math. Stat. *37*, 1554–1563.

BAUM, L.E., PETRIE, T., SOULES, G., and WEISS, H. (1970), *A maximization technique occuring in the statistical analysis of probabilistic functions of Markov chains*, Ann. Math. Stat. *41*, 164–171.

BAUM, L.E. and SELL, G.R. (1968), *Growth functions for transformations on manifolds*, Pac. J. Math. *27*, 211–227.

BELLEGARDA, J. and NAHAMOO, L. (1990), *Tied mixture continuous parameter modeling for speech recognition*, IEEE Trans. on Acoustics, Speech, and Signal Proc. *38*, 2033–2045.

BOCCHIERI, E. and MAK, B. (2001), *Subspace distribution clustering hidden markov model*, IEEE Trans. on Speech and Audio Proc. *9*, 264–275.

BUHMANN, J. and KUHNEL, H. (1993), *Complexity optimized data clustering by competitive neural networks*, Neural Computation *5*, 75–88.

CHOU, W., LEE, C., JUANG, B., and SOONG, F. (1994), *A minimum error rate pattern recognition approach to speech recognition*, Int. J. Pattern Recogn. Artificial Intelligence *8*, 5–31.

COLLINGS, I., KRISHNAMURTHY, V., and MOORE, J. (1994), *Online identification of hidden Markov models via recursive prediction error techniques*, IEEE Transactions on Signal Processing *42*, 3535–3539.

DEMPSTER, A.D., LAIRD, N.M., and RUBIN, D.B. (1977), *Maximum likelihood from incomplete data via the EM algorithm*, J. Royal Stat. Soc. *B-39*, 1–37.

EPHRAIM, Y., DEMBO, A., and RABINER, L. (1989), *A minimum discrimination information approach for hidden Markov modeling*, IEEE Transactions on Information Theory *35*, 1001–1013.

FARAGO, A. and LUGOSI, G. (1989), *An algorithm to find the global optimum of left-to-right hidden Markov model parameters*, Problems of Control and Information Theory-Problemy Upravleniya *I*. Teorii Informatsii *18*, 435–444.

FORD, L.R. and FULKERSON, D.R. (1956), *Maximal flow through a network*, Canadian J. Math. *8*, 399–404.

GRANAT, R. (2003), *A method of hidden markov model optimization for use with geophysical data sets*, Comput. Science - ICCS 2003, PT III, Proceedings *2659*, 892–901.

GRANAT, R.A. and DONNELLAN, A. (2001), *Deterministic annealing hidden markov models for geophysical data exploration*, Proc. 3rd Workshop Sci. Datamining.

HEKI, K., MIYAZAKI, S., and TSUJI, H. (1997), *Silent fault slip following an interplate thrust earthquake at the Japan trench*, Nature *386*, 595–598.

HIROSE, H., HIRAHARA, K., KIMATA, F., FUJII, N., and MIYAZAKI, S. (1999), *A slow thrust slip event following the two 1996 Hyuganada earthquakes beneath the Bungo channel, southwest Japan*, Geophys. Res. Lett. *26*, 3237–3240.

HUO, Q. and CHAN, C. (1993), *The gradient projection method for the training of hidden Markov models*, Speech Communication *13*, 307–313.

JAMSHIDIAN, M. and JENNRICH, R.I. (1993), *Conjugate gradient acceleration of the EM algorithm*, J. Am. Stat. Assoc. *88*, 221–228.

JUANG, B. and RABINER, L. (1985), *Mixture autoregressive hidden Markov models for speech signals*, IEEE Transactions on acoustics speech and Signal Processing *33*, 1404–1413.

JUANG, B. and RABINER, L. (1990), *The segmental k-means algorithm for estimating parameters of hidden Markov models*, IEEE Transactions on Acoustics Speech and Signal Processing *38*, 1639–1641.

JUANG, B. and RABINER, L. (1991), *Hidden Markov models for speech recognition*, Technometrics *33*, 251–272.

KIRKPATRICK, S., GELATT, C., and VECCHI, M. (1983), *Optimization by simulated annealing*, Science *220*, 671–680.

KWONG, S., CHAU, C., MAN, K., and TANG, K. (2001), *Optimisation of hmm topology and its model parameters by genetic algorithms*, Pattern Recognition *34*, 509–522.

LEE, K. (1990), *Context-dependent phonetic hidden Markov models for speaker-independent continuous speech recognition*, IEEE Transactions on Acoustics Speech and Signal Processing *38*, 599–609.

LEE, K. and HON, H. (1989), *Speaker-independent phone recognition using hidden Markov models*, IEEE Transactions on Acoustics Speech and Signal Processing *37*, 1641–1648.

MCGUIRE, G., WRIGHT, F., and PRENTICE, M. (2000), *A bayesian model for detecting past recombination events in dna multiple alignments*, J. Comput. Bio. *7*, 159–170.

MELBOURNE, T. and WEBB, F. (2002), *Precursory transient slip during the 2001 $m_w = 8.4$ peru earthquake sequence from continuous gps*, Geophys. Res. Lett. *29*, art. no.–2032.

MELBOURNE, T. and WEBB, F. (2003), *Slow but not quite silent*, Science *300*, 1886–1887.

MELBOURNE, T., WEBB, F., STOCK, J., and REIGBER, C. (2002), *Rapid postseismic transients in subduction zones from continuous gps*, J. Geophys. Res. – Solid Earth *107*, art. no.–2241.

MILLER, M., MELBOURNE, T., JOHNSON, D., and SUMNER, W. (2002), *Periodic slow earthquakes from the cascadia subduction zone*, Science *295*, 2423–2423.

RABINER, L.R. (1989), *A tutorial on hidden Markov models and selected applications in speech recognition*, Proc. IEEE *77*, 257–286.

ROGERS, G. and DRAGERT, H. (2003), *Episodic tremor and slip on the Cascadian subduction zone: The chatter of silent slip*, Science *300*, 1942–1943.

ROSE, K. (1998), *Deterministic annealing for clustering, compression, classification, regression, and related optimization problems*, Proc. IEEE *86*, 2210–2239.

ROSE, K., GUREWITZ, E., and FOX, G.C. (1992), *Vector quantization by deterministic daing*, IEEE Trans. Inf. Theory *38*, 1249–1257.

ROSE, K. and RAO, A.V. (2001), *Deterministically annealed design of hidden Markov model speech recognizers*, IEEE Trans. on Speech and Audio Processing, *9*, 111–126.

UEDA, N. and NAKANO, R. (1994), *Mixture density estimate via EM algorithm with deterministic annealing*, Proc. IEEE Neural Networks for Signal Processing, 69–77.

UEDA, N. and NAKANO, R. (1998), *Deterministic annealing EM algorithm*, Neural Networks *11*, 271–282.

WHILEY, M. and TITTERINGTON, D.M. (2002), *Applying the deterministic annealing expectation maximisation algorithm to naive Bayesian networks*, Univ. Glasgow Tech. Report.

WONG, Y. (1993), *Clustering data by melting*, Neural Computation *5*, 89–104.

YOUNG, S. and WOODLAND, P. (1994), *State clustering in hidden Markov model-based continuous speech recognition*, Computer Speech and Language *8*, 369–383.

YUILLE, A.L., STOLORZ, P., and UTANS, J. (1994), *Statistical physics, mixtures of distributions and the EM algorithm*, Neural Computation *6*, 334–340.

(Received December 3, 2004, revised December 12, 2005, accepted December 29, 2005)

Pure appl. geophys. 163 (2006) 2513–2528
0033–4553/06/122513–16
DOI 10.1007/s00024-006-0140-0

Pure and Applied Geophysics

A Recent Application of the ETAS Model and a Proposed Method for Prediction of Strong Aftershocks

W. B. Liu[1,2] and L. Ma[1]

Abstract—In this paper, 28 aftershock sequences are selected, which are distributed in different areas including north China, southwest of China, northwest of China, Taiwan area, Turkey and Greece. In order to investigate the characteristics of these sequences along with different temporal and spatial coordinates, each sequence has been divided into dozens of segments called "sub-sequences". The ETAS (Epidemic Type Aftershock Sequences) model is applied to each "sub-sequence", and therefore the vectors of parameters of ETAS could be evaluated. Another model named LR (Logistic Regression) model is used to seek the correlate relation between the parameters of ETAS applied to every earthquake "sub-sequence" and seismicity. All the analyses and estimations imply that the characteristic of decay of aftershock sequences in different temporal and spatial domains seems to be characterized by the parameters of the ETAS model applied to some aftershock sequences or "sub-sequences", and there are some proportional correlate relations between the evaluation of LR model and the occurrence probability of the succeeding strong seismic energy release.

Key words: ETAS model, LR model, "sub-sequences", MLE, aftershock sequences, success probability.

1. Introduction

As a result of stress or strain field change somewhere in the lithosphere layer strong earthquakes usually trigger their own aftershock sequences. To the phenomena there are many explanations based on both the observation of the test of the rock mechanism in the laboratory and some theories for the mechanism of the shocks. The foregone research often focused on the qualitative analysis, however it is significant in hazard quantitative estimation and prediction to some degree if tectonics, and dynamic condition, medium property, and other interrelated factors are considered. Therefore the quantitative analysis in some way might be considerably more necessary to explore the characteristic of profiles and detailed information of aftershock sequences.

[1] Institute of Earthquake Science, CEA (China Earthquake Administration), Beijing 100036, China. E-mail: liuwenbing@seis.ac.cn; mali@seis.ac.cn

[2] Earthquake Administration of Tianjin Municipality, Tianjin 300201, China.

The study of aftershock sequences has become an important branch of seismology, and there are a few models that have been adopted to describe the character of aftershock sequences, such as Omori formula, UTSU (1961, 1970, 1977), and LIU (1984) have been conducting systemic analysis of aftershock, and put forward the approach of how to determine the type of earthquake sequences, estimate how many times the aftershock will occur in a certain time cycle, the magnitude of the strongest aftershock, and how long these events will last. The line of sight might be drawn for the history of his effective prediction through this method. In addition, the detailed rules of his method have been constituted to predict the strong aftershocks in China.

In fact, this branch of seismology research exists for considerably more than a hundred years, and Japanese researcher OMORI (1894) put forward his law, which shows the inverse power relation between the cumulative number of events and the time after the main shocks, and the exponent is usually -1 in his formula; UTSU (1975, 1987) modified the exponent in Omori's relation after analysis of more aftershock sequences in Japan; OGATA *et al.* (1982, 1999) built up such a model named Epidemic-type Aftershock-sequence (ETAS) model, which combines the modified Omori law with Gutenberg-Richter frequency magnitude relation for a history-dependent occurrence rate of a point process, and applied this theory to the analysis of many aftershock sequences worldwide; ZHUANG *et al.* (2002) goes further in a thinning theory applied to declustering of aftershock sequences; MA and VERE-JONES (1997) and MA and ZHUANG (2001) used this model to determine precursory quiescence and active seismicity and predicts the next strong events in Jiashi region of Xinjiang in China.

However, describing the subtle characters of sequences in detail is more necessary than the cursory presentation of the entire profile. Given such an assumption, there are approximately similar characteristics in seismic tectonics, geodynamic environment, and stress distribution in the region where aftershocks occur, and in terms of the fractal and self-alike phenomena in both the seismic ruptures and time series of the aftershocks sequences, the local characteristics of aftershocks sequences could be discovered in such a way that any whole sequence might be divided into dozens of segments (named "sub-sequences") in advisable approaches to make some quantitative analysis to these segments based on combination ETAS model with the LR model. Two advantages regarding this method could be posed at least. One of the feasibilities is that all of the "sub-sequences" are abstracted from the whole sequence and among them there should be some conjunct and common characteristics; The other point to be noted is that the estimated parameters of the ETAS model applied to dozens of "sub-sequences" could act as the samples of Logistic Regression (LR) model, as we want to perform tests to provide the precursory quiescence before strong aftershocks (MINQI *et al.,* 1995).

2. Models

2.1. ETAS Model

Earthquakes always occur stochastically, and it is very necessary for statistical seismologists to consider and adopt the statistical methods to explore the time series characteristics (HAWKES, 1971; VERE-JONES, 1976, 1978, 1989, 2001; KAGAN and KNOPOFF, 1980). Based on the analysis of aftershock sequences and background events, OGATA (1988, 1989) posed interrelated hypotheses: (1) The background of an aftershock sequence is always constant; (2) all the events (including aftershocks) follow the Omori law, and can trigger their own aftershocks; (3) the number of aftershocks are only related to their parents; (4) all the events in the same generation of one aftershock sequence are independent. Referring to the relation between the magnitude of main event (M) and the scale size of aftershock region (S) posed by UTSU (1970).

$$\log_{10} S = 1.0M - 3.9 \tag{1}$$

Based on the self-alike and stochastic point process, OGATA (1988) put forward temporal Epidemic-type Aftershock-sequence model (ETAS) in the form of the intensity function

$$\lambda(t, M) = \mu + K \sum_{t_i < t} \frac{e^{\alpha(M_i - M_0)}}{(t - t_i + c)^p} \quad M_i \geq M_c \tag{2}$$

which represents the seismicity of aftershocks, or the mathematic occurrence probability of such an event with magnitude $[M, M + \Delta M]$ time interval $[t, t + \Delta t]$, where μ is the occurrence probability of background events, and K is the effect ability of the main shock to aftershocks, and p is the attenuation rate of aftershock sequences, and c means the time intervals from major event to maximum rate of seismicity, and α the capability of events triggering their own aftershock, and M_0 is magnitude of major event, M_c is the threshold magnitude.

In furtherance, similar pattern could be found in a spatial coordinate, thus the temporal-spatial ETAS model is defined as (OGATA *et al.,* 2002)

$$\lambda(t, x, y | H_t) = \mu + K_0 \sum_{t_i < t} (t - t_i + c)^p \times \left\{ \frac{r(x - x_i, y - y_i)}{e^{2\alpha(M_i - M_c)}} + d \right\}^{-q} \tag{3}$$

where

$$r(x - x_i, y - y_i) \equiv (x, y)\Sigma(x, y)^t = \frac{1}{\sqrt{1 - \rho^2}} \left(\frac{\sigma_2}{\sigma_1} x^2 - 2\rho xy + \frac{\sigma_1}{\sigma_2} y^2 \right) \tag{4}$$

in which σ_1, σ_2, ρ are positive constants, and Σ is a 2×2 positive definite matrix when the anisotropic medium is considered.

2.2. LR Model

We assume there is theoretically a continuous variable y^*, which denotes the probability of an event occurring, varying in $(-\infty, +\infty)$. When y^* beyond one certain critical value, we think the event occurs.

Given that the correlation relation between random 0–1 variable Y and determined variable $X_1, X_2, \ldots, X_p$ exists, and the observation sample $(x_{i1}, x_{i2}, x_{i3}, \ldots x_{ip}, y_i)$, $1 \leq i \leq n$ with sample space n, and assuming the relation between y_i and $(x_{i1}, x_{i2}, x_{i3}, \ldots x_{ip})$ can be given as

$$E(y_i) = p_i = f(\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}) = f\left(\beta_0 + \sum_{k=1}^{p} \beta_k x_{ik}\right) \tag{5}$$

where $f(x) = \frac{e^x}{1+e^x}$ is the logistic regression function, resulting in the probability intensity of variable y_i

$$p_i = \frac{\exp\left(\beta_0 + \sum_{k=1}^{p} \beta_k x_{ik}\right)}{1 + \exp\left(\beta_0 + \sum_{k=1}^{p} \beta_k x_{ik}\right)} \tag{6}$$

can be obtained, and then the occurrence probability of y_i (MINQI *et al.*, 1995),

$$P(y_i) = p_i^{y_i}(1 - p_i)^{1-y_i}, \tag{7}$$

where

$$y_i = \begin{cases} 1 \\ \text{or } 0 \end{cases}. \tag{8}$$

3. Data and Elementary Analysis

All of the aftershock sequences, catalogues are provided by personnel serving in local earthquake administrations in China or downloaded from some limited websites from the Internet (see Table 1). Total data include 58 sequences, although merely 30 of them are ineligible, and the main causes are: (1) Enough precision is not available in some regions; (2) the form of magnitude of events is not very uniform (some are M_L, however some are M_S, or M_W), while it is quite difficult for us to unify them in the same format; (3) the number of events is insufficient to satisfy dividing them into more segments and the requirements of simulation.

In any case, it is indeed necessary to provide explanations about these data. All of them scatter in different places in China, Turkey, and Greece. However, most of them are located in China Mainland, and others fall in Taiwan (TW) Island and the countries of Turkey and Greece. Therefore, these sequences can be classified into two

Table 1

Summary of the 28 aftershock sequences, which includes columns of regions, location of aftershocks, the magnitude of major shock, time span of each sequence, number of aftershocks in every sequence, the threshold magnitude of each sequence, and the number of segments divided, and all the sequences are sprinkled in North China (NC) and northeast of China (NE), southwest of China (SW), northwest of China (NW), Taiwan (TW) Island, and others fall into the countries of Turkey and Greece (just for reference in this paper)

Region		Location	Magnitude	Time interval	Number of	Threshold	Segments
NC&NE	1	Datong	5.7	19891018∼20010225	4832	1.7	157
	2	Baotou	6.8	19960503∼19971228	954	1.4	10
	3	Xiuyan	5.9	19991129∼20001111	1589	0.6	24
	4	Wuhai	3.8	20030121∼20030620	303	0.8	16
	5	Ninglang 1	5.7	19880110∼19880310	1014	1.2	12
	6	Lancang	7.6	19881106∼19890210	3998	2.0	10
	7	Ninglang 2	5.4	19910412∼19910711	150	1.1	9
	8	Shidian1	5.0	19910701∼20011022	469	2.5	9
	9	Yongsheng	5.6	19921218∼1993331	94	1.0	5
	10	Puer	6.3	19930127∼19930410	304	1.2	14
	11	Dayao	5.5	19930201∼19930414	154	1.1	9
	12	Zhongdian	5.8	19930717∼19931002	1171	1.6	16
	13	Jinggu	4.6	19940101∼19941227	56	2.1	6
	14	Menglian	6.2	19950431∼19950630	760	2.7	6
SW	15	Wuding	6.5	1995124∼19960101	2823	2.2	14
	16	Lijiang1	7.0	19960203∼19960810	3998	2.0	19
	17	Lijiang2	4.8	19960925∼19961231	241	2.1	4
	18	Lijiang3	4.0	19970101∼19971231	615	1.2	12
	19	Ninglang3	5.3	19981002∼19981215	1074	1.2	10
	20	Chengjiang	5.2	19991125∼19991231	124	0.8	7
	21	Yaoan	5.9	20000115∼20000331	1299	1.6	11
	22	Shidian2	5.2	20010410∼20010722	396	3.0	17
	23	Ninglang3	5.8	20010524∼20010717	56	2.6	3
	24	Yongsheng	5.9	20011028∼20020128	60	2.0	4
NW	25	Kunlun	8.1	0011114∼20020530	2682	0.9	51
TW	26	Chi-chi	7.3	19990920∼19991128	7397	1.3	14
Turkey	27	Düzce	6.6	19991112∼19991128	597	1.0	35
Greece	28	Patrai	5.3	20021202∼20030831	146	2.8	6

types: One is the interpolate events (mainly in strike-slip, or dip-slip, or reverse-slip style, i.e., those occurring in Mainland China, Turkey, and Greece (PAPADOPOULOS *et al.*, 2002), and the other one is the plate-boundary type (such as the underthrusting faulting in Taiwan Island) (WANG and SHIN, 1998; HUANG *et al.*, 2001), due to different patterns of earthquake occurring.

To every sequence we have selected, the threshold magnitude must be determined to ensure reliability of data based on the Gutenberg-Richter relation which is one of the greatest successes in empirical studies. The linear model is adopted to search the appropriate threshold magnitude M_c (see Fig. 2), which must exist in the magnitude span from the minimum magnitude $M_{\min}$ to the maximum magnitude $M_{\max}$. This operation can be implemented as step A follows.

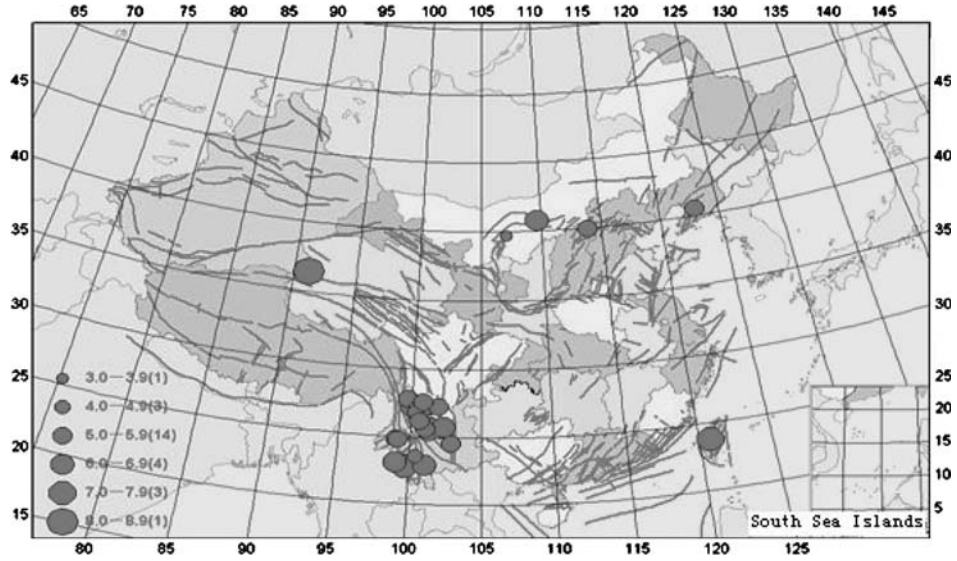


Figure 1
Geographical distribution of the epicenters of the main shock of the aftershock sequences occurring in China Mainland and Taiwan considered in this paper.

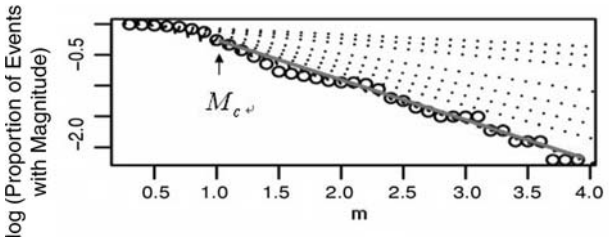


Figure 2
The sketch map registering the appropriate threshold magnitude M_c.

Step A1. Assume the value $M_t = M_{\min}$ as the first potential threshold value, and apply the general linear model to the data both in $[M_{\min}, M_t]$ and $[M_t, M_{\max}]$, respectively, and the vectors of residual values ε_{10} and ε_{20} of the two simulations can generate;

Step A2. Make the value $M_t = M_{\min} + \Delta M \times n$, and $n \in (1, (M_{\max} - M_{\min})/\Delta M)$, and do the implementation in the step 1 repeatedly, and the vectors of residual values ε_{1n}, and ε_{2n} can be got.

Step A3. Compare all of the residual values ε_{2n}, $n \in (0, (M_{\max} - M_{\min})/\Delta M)$ to look for the minimum one among them, and the corresponding M_t will be regarded as the threshold magnitude of this sequence.

4. Simulation and Results

Epidemic-type Aftershock-sequences (ETAS) model has been regarded as a successful point process model generated from the empirical laws (G-R law and modified Omori law), and a more validated version than Gaussian distribution from the aspect of simulation of seismicity of aftershock sequences (OGATA *et al.*, 1982; OGATA, 1988, 1989, 1992, 1998). Having assumed the information of tectonics and dynamic conditions, and the earthquake sources could be different among different seismic tectonic regions, and reflected through analysis of the parameters of ETAS model, in order to describe the more detailed and intricate characteristics in every aftershock sequence in different regions, we continue the process as step B follows.

Step B1. Apply the temporal ETAS model to 28 whole aftershock sequences.

Step B2. Divide each sequence into more or less (about ten or more) segments, or sub-sequences, in which there are the same numbers of events or same span of time.

Step B3. Apply the ETAS model to all the segments and estimate corresponding parameters $\hat{\theta} = (\hat{\mu}, \hat{K}, \hat{\alpha}, \hat{c}, \hat{p})$ by maximum likelihood estimates (MLEs), and Akaike's information criterion (AIC) is used to indicate a goodness of simulation.

Step B4. Conceive the LR (Logistic Regression) Model, and critical value could be set as the mean seismicity, thus the corresponding response variable could be set "1", when the seismicity of the segment is active enough, or set "0" for weak enough seismicity.

Some essential points in the process of operations and simulations have to be brought forward as follows.

(1) Enough events in each segment are necessary in simulation of ETAS model, though from another angle the number of segments in each sequence is also very important for LR simulation. Considering the actual sequence data we have collected, the A step must be satisfied in the first place to assure the ETAS model fitting well, whereas enough segments or sub-sequence are expected.

(2) The methods of maximum likelihood estimates (MLEs) and iterative algorithm are used to calculate the parameters of ETAS model. Yet the convergence is not always good in simulating all of the data, and the way of setting the initial parameter repeatedly is adopted. For parameter μ almost constant, so we replace the initial μ with $0.9 \times u_0$ convergence is not good enough.

The results are illustrated in Table 2, Fig. 3 and Fig. 4.

From Figure 3, parameter estimate of ETAS model $K \approx 0.05$ in north China, α and c values are located in the range (0.001,0.6), p values in the range (0.9,1.4); in the southwest region of China these parameters fluctuate more: K values range from 0.3 to 0.5, α and c values with the range (0.01,0.8), p values in (0.6,1.6); in northwest of China, K values are in (0.4,0,5), α and c are situated in (0.02,07), and the other parameter p in (1.0,1.4). Parameters μ and K smoother than the other three parameters (see Table 3 for details).

Table 2

The parameter estimations of the 28 whole sequences

	Location	$\hat{\mu}$	$\hat{K}$	$\hat{\alpha}$	$\hat{c}$c	$\hat{p}$	*AIC*
1	Datong	0.0000	0.0000	0.0074	0.2819	1.3917	−3889.10
2	Baotou	0.0101	0.0532	0.0728	0.3984	1.2368	−8597.32
3	Xiuyan	0.0157	0.0512	0.1329	0.5764	1.2568	−2563.28
4	Wuhai	0.0002	0.4965	0.0524	0.4638	1.2369	−1536.47
5	Ninglang1	0.0000	0.0000	0.0233	0.4702	1.3210	−5292.36
6	Lancang	0.0000	0.0000	0.1958	0.5380	1.3591	−6787.04
7	Ninglang2	0.0120	0.0490	0.1256	0.4857	1.2654	−897.03
8	Shidian1	0.0002	0.0000	0.0684	0.4863	1.2576	−2874.36
9	Yongsheng1	0.0000	0.0000	0.0832	0.4512	1.2847	−2584.55
10	Puer	0.0000	0.0000	0.1039	0.4935	1.1600	−3797.33
11	Dayao	0.0100	0.0490	0.0800	0.4500	1.2563	−500.65
12	Zhongdian	0.0000	0.0000	0.1041	0.4622	1.1743	−6013.54
13	Jinggu	0.0001	0.0000	0.0607	0.4802	1.0492	−919.22
14	Menglian	0.0000	0.0000	0.1083	0.2816	1.8209	−701.24
15	Wuding	0.0001	0.0001	0.1236	0.6027	0.9586	−4388.18
16	Lijiang1	0.0000	0.0000	0.0488	0.6066	1.0797	−7821.78
17	Lijiang2	0.0000	0.0000	0.0847	0.6258	1.1632	−824.62
18	Lijiang3	0.0001	0.0002	0.1632	0.2835	1.3623	−354.74
19	Ninglang3	0.0000	0.0000	0.0635	0.4855	1.2254	−4326.79
20	Chengjiang	0.0000	.0000	0.0520	0.2440	1.3867	−9954.92
21	Yaoan	0.0000	0.0000	0.0823	0.6341	1.0536	−7958.55
22	Shidian2	0.0001	0.0000	0.1238	0.4783	1.2864	−2369.51
23	Ninglang3	0.0000	0.0000	0.1034	0.5583	1.1259	−4726.45
24	Yongsheng2	0.0001	0.0000	0.0476	0.0512	1.2478	−300.56
25	Kunlun Mont.	0.0000	0.0000	0.0372	0.3718	1.3111	−27489.51
26	Chi-chi	0.2500	0.1200	0.4000	0.1500	1.5320	−4263.25
27	Düzce	0.0100	0.1200	0.0800	0.5320	1.2300	−1235.28
28	Patrai	0.0000	0.1350	0.5136	0.4260	1.2460	−2145.21

The characteristics of aftershock sequences in different regions are distinguished in all: (a) From the facet of seismicity, the Chi-Chi sequence in Taiwan is the strongest in all the sequences in China, then the sequences in the southwest and northwest regions of China are much stronger than those in the north China region, which is approximately equal to Patrai aftershock sequence in Greece. However, the Düzce aftershock sequences in Turkey are the strongest of all the earthquake sequences we study in this paper; (b) from the rate of decay of aftershocks, the order from bigger to small is: Southwest of China, northwest of China, north of China, and the other three earthquakes are almost the same as north of China; (c) from the ability of triggering aftershocks, the order from strong to weak is: southwest of China, northwest of China, north China, and the other three sequences; (d) about the goodness of fitting using the ETAS model, according to Akaike's information criterion (*AIC*), the best are these sequences in the southwest region of China, Chi-Chi sequence in Taiwan Island, Düzce sequence in Turkey, Patrai in Greece, and the sequences in the northwest region of China and north China are not fitted perfectly.

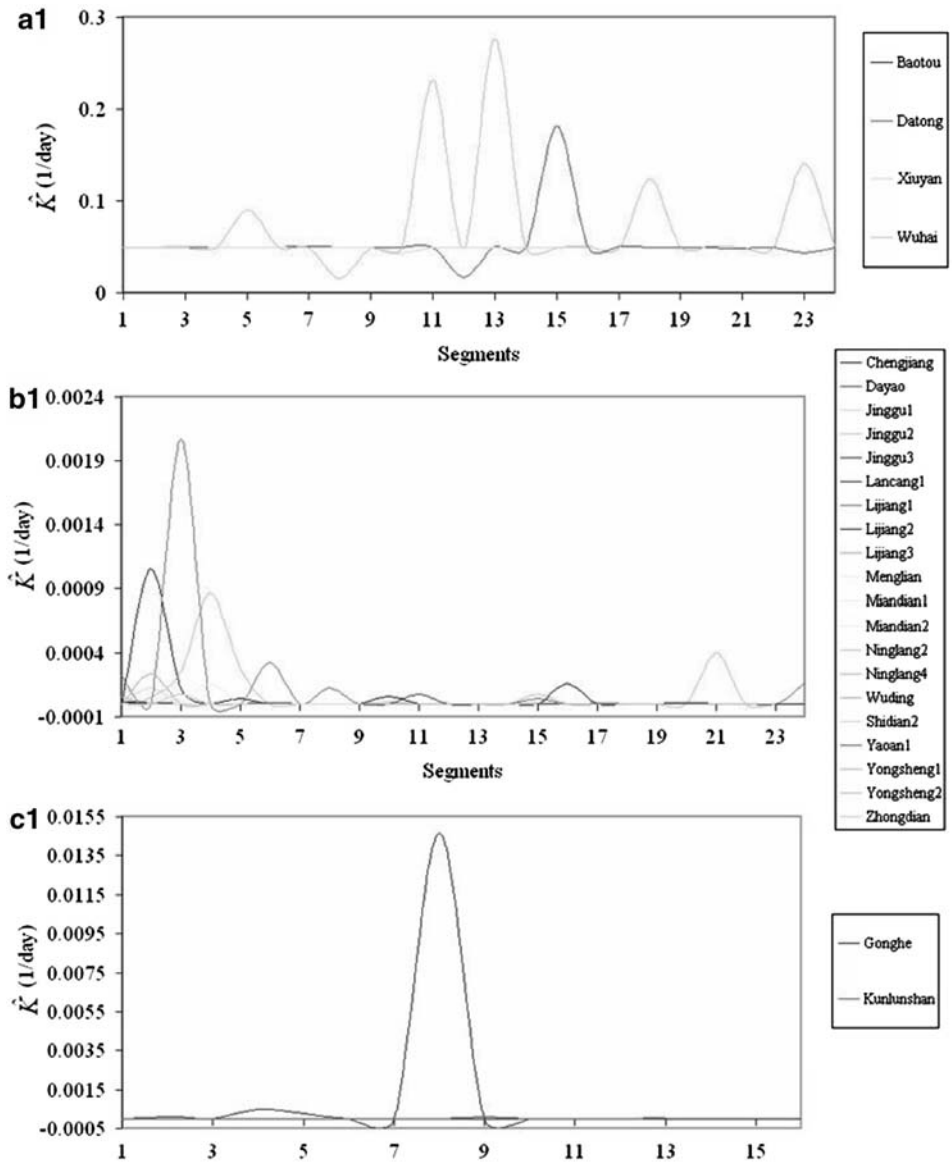


Figure 3

The superposition plots of curves of varieties of ETAS parameter estimates $(\hat{K}, \hat{\alpha}, \hat{c}, \hat{p})$aftershock sequences in different regions of China on the temporal coordinate: (a) The sequences in the region of north China; (b) The sequences in the region of southwest China; (c) the sequences in the region northwest China. And the numbers 1, 2, 3, and 4 following the letter a, b, and c denote different ETAS parameters: K, α, c, and $p$order, respectively.

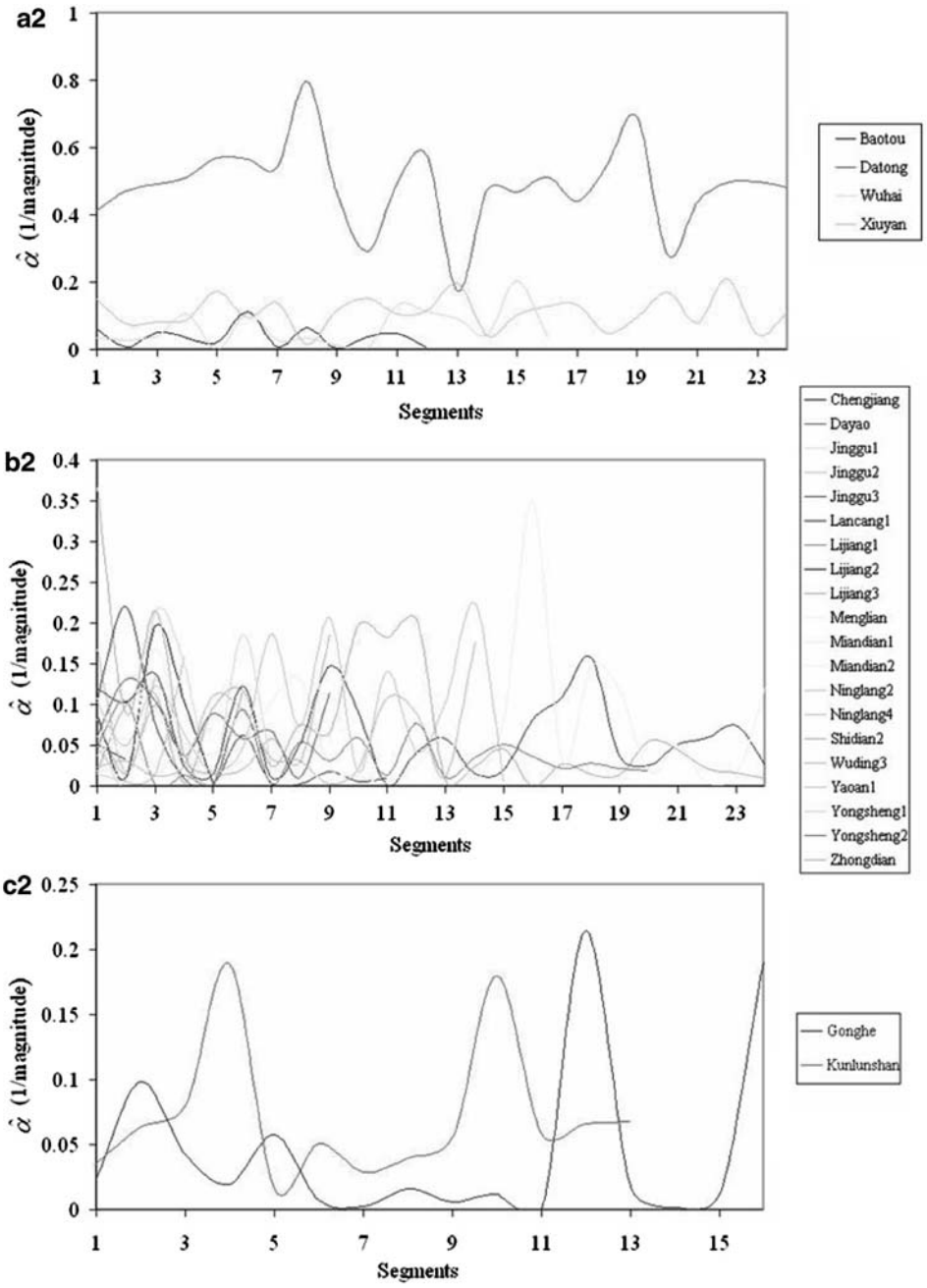


Figure 3
(Contd.)

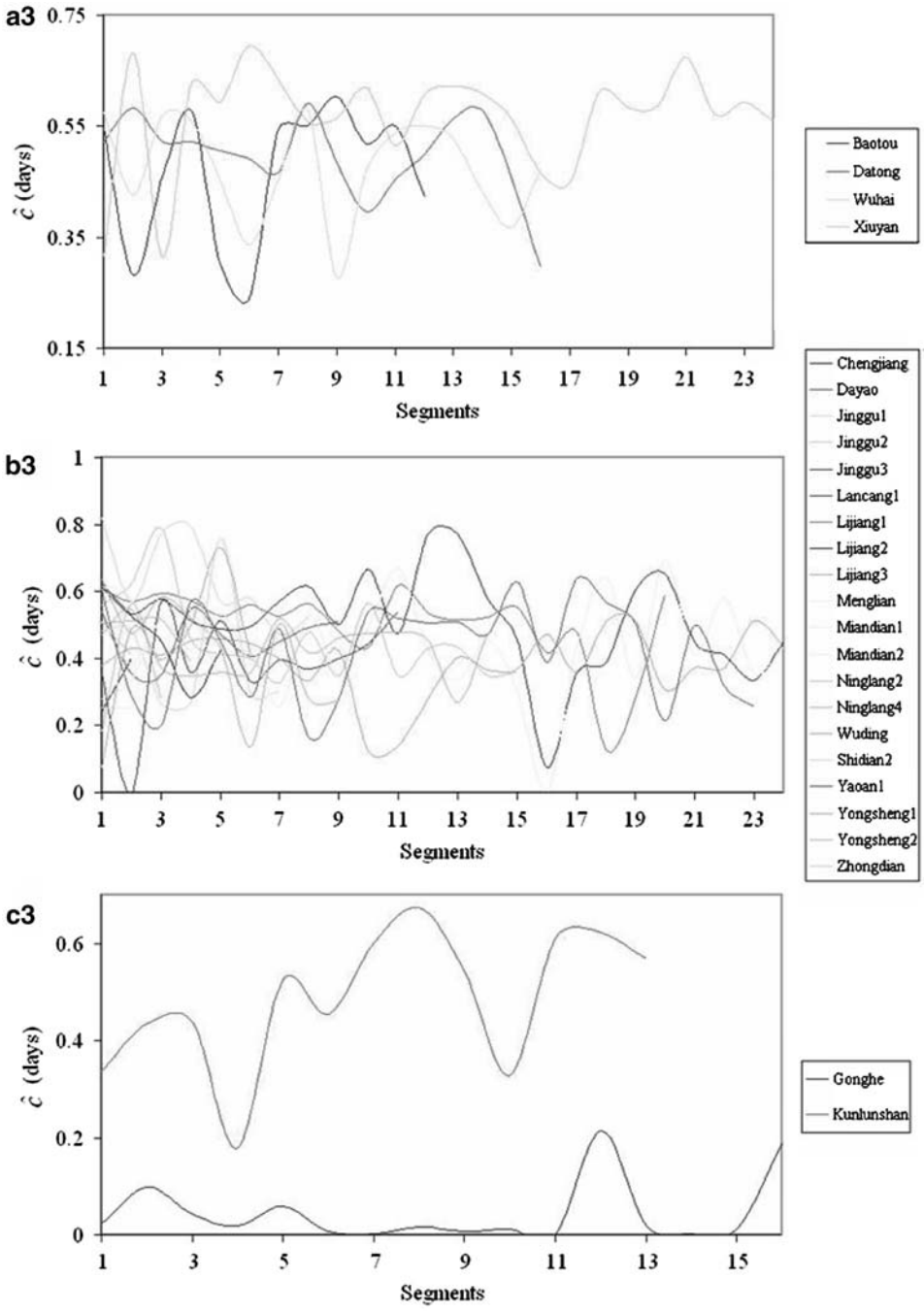


Figure 3
(Contd.)

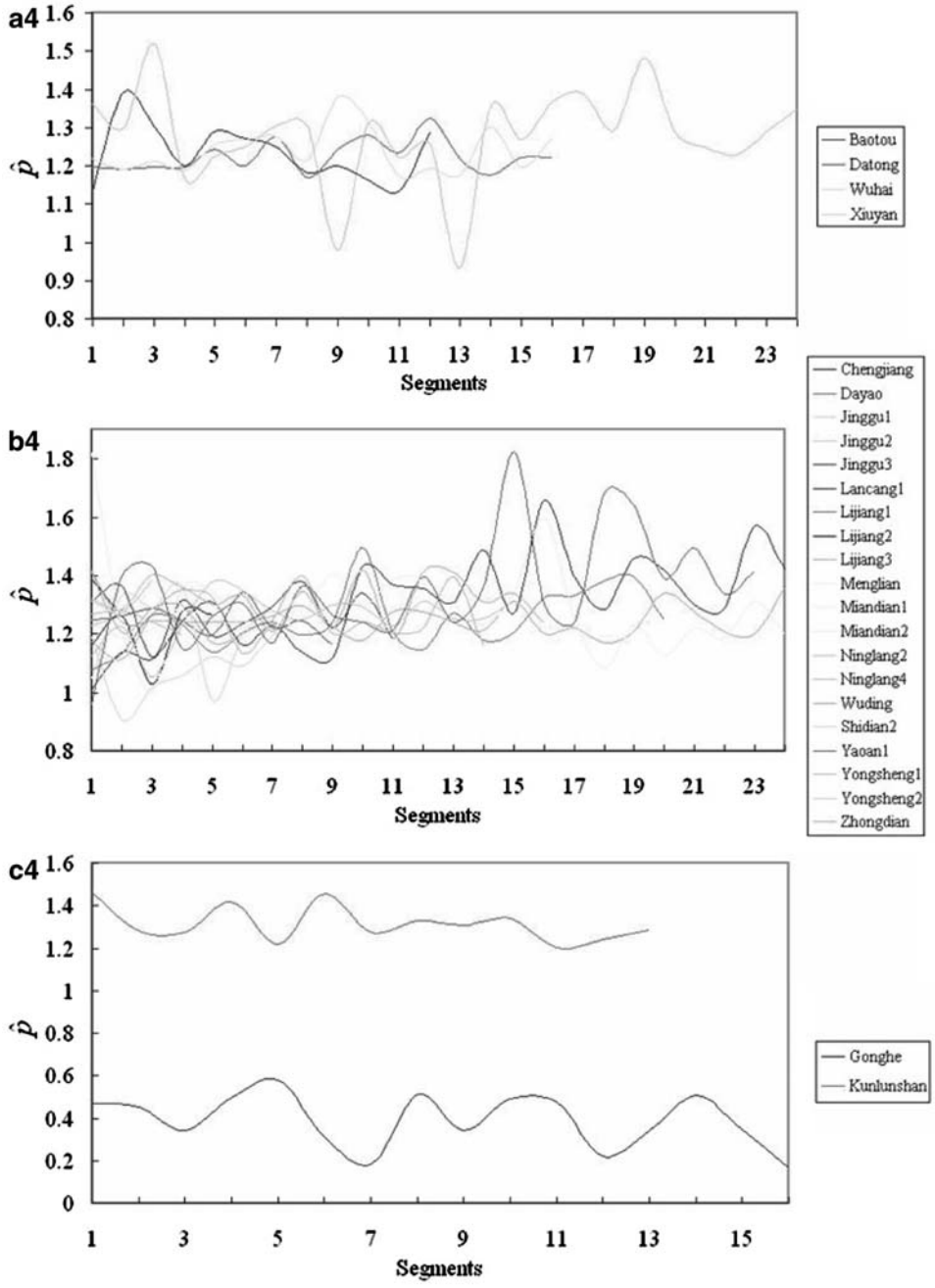


Figure 3
(Contd.)

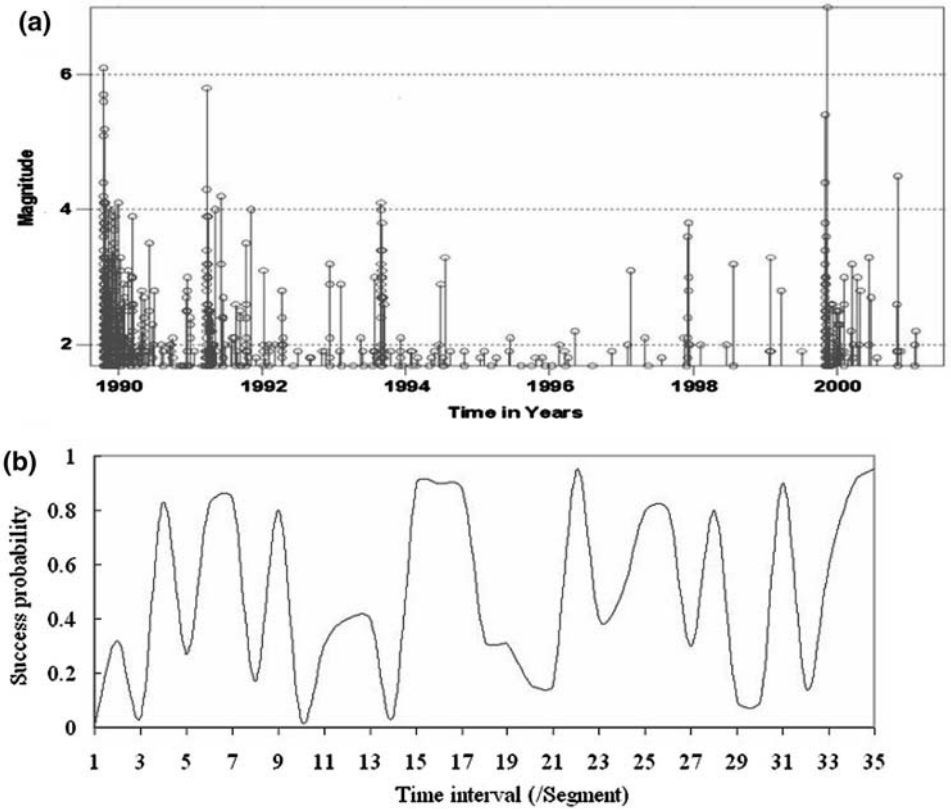


Figure 4
An example of applying the Logistic Regression model to aftershock sequences. (a) The probability of strong shocks occurring through simulation of segments ofETAS model; (b) magnitude- time plot of Datong sequences in the north China region.

Table 3

The estimates of parameters of the ETAS Model applied to different regions. NC&NE mean north China and northeast China; SW means southwest region of China; NW denotes northwest region of China, and the other two sequences which occurred in Turkey and Patrai, respectively

Region	$\hat{\mu}$	$\hat{K}$	$\hat{\alpha}$	$\hat{c}$	$\hat{p}$	*AIC*
NC&NE	$0 \sim 0.001$	$0 \sim 0.05$	$0.001 \sim 0.6$	$0.001 \sim 0.6$	$0.9 \sim 1.4$	$-1500 \sim -8600$
SW	$0 \sim 0.001$	$0.3 \sim 0.5$	$0.01 \sim 0.8$	$0.01 \sim 0.8$	$0.6 \sim 1.6$	$-500 \sim -15000$
NW	$0 \sim 0.001$	$0.4 \sim 0.5$	$0.02 \sim 0.7$	$0.02 \sim 0.7$	$1.0 \sim 1.4$	$-100 \sim -59000$
Chi-Chi	$0 \sim 0.001$	$0.1 \sim 2.2$	$0 \sim 12$	$0 \sim 3$	$0.6 \sim 1.8$	$-200 \sim -6000$
Düzce	$0 \sim 0.001$	$0.5 \sim 0.8$	$0.04 \sim 0.14$	$0.5 \sim 0.75$	$1.0 \sim 1.35$	$-150 \sim -6000$
Patrai	$0 \sim 0.001$	$0.05 \sim 0.12$	$0.03 \sim 0.12$	$0.4 \sim 0.5$	$1.1 \sim 1.3$	$-100 \sim -1700$

From Figure 4, we can factually find that there is always precursory quiescence of seismicity where the success probability value of Logistic Regression (LR) model is high, and the next active seismicity will appear soon. And in the opposite way the success probability value is low, and the seismic quiescence will come forward. Perhaps the phenomena could be explained by the theory of precursory quiescence before strong earthquakes.

5. *Conclusions*

This paper describes one recent application of ETAS (Epidemic-type Aftershock-sequences) model and LR (Logistic Regression) model to 28 aftershock sequences occurring in China Mainland, Taiwan, Turkey and Greece. The innovative work is to divide all the 28 sequences into dozens of segments (called sub-sequences), respectively, which is regarded as the sequences we calculate and analyze. We wish the characteristics of varieties of all the ETAS parameter estimates could indicate the seismicity of every sequence on temporal coordinate. It is significant to apply the LR model to these parameter estimates in order to provide the precursory quiescence before the next strong aftershocks.

There are conclusions which can be drawn according to the simulation and analysis of 28 aftershock sequences as follows.

(1) There are different characteristics in the aftershock sequences in different regions, which can be incarnated through the variety of parameter estimates of ETAS model.
(2) The size of the main shock is very important in this paper, and the aftershocks sequences with more sizable main shocks would decay with time, and characteristics of attenuation are more obvious; otherwise, the aftershock sequences would decay in a shorter time and the characteristics of attenuation are not clear, too.
(3) With regard to the goodness of fitting on the earthquake sequences with ETAS, good *AIC*(considerable low value) can be gained in the simulation of the former segments of most of the whole sequences investigated in this paper, and those later parts of the sequences are not fitted very well using the ETAS model. The most imaginable cause is the attenuation characteristics of those later parts of aftershock which sequences are not seen obviously.
(4) The Logistic Regression Model could be used to identify the quiet seismicty and predict the strong aftershocks. However, the goodness of this kind of operation is related to the number of samples, which is the result of fitting on each segment with the ETAS model.

Acknowledgements

The authors gratefully acknowledge the financial support of grant-in-aid 40134010, 40574020, 40234042 and 10371012 objects from NSFC, China. The theme of this paper is inspired by the discussion in the statistical seismology workshop by Professor Ma L. from Institute of Earthquake Sciences, China Earthquake Administration (CEA) in Beijing. The completion of this paper is facilitated by considerably generous assistance of Ma W. J. from Qinghai Earthquake Administration, Yang L. M. from Gansu Earthquake Administration, Chuo Y. Q. from Shanxi Earthquake Administration, Zhang C. J. from Graduate School of Chinese Academy Sciences for their provision of abundant data, as well as beneficial discussion regarding statistical simulation and software package of R and SSLib with Wang L. F. from Institute of Earthquake Sciences, China Earthquake Administration (CEA) in Beijing, David and Brownrigg L. F. from Victoria University of Wellington, New Zealand.

References

Hawkes, A.G. (1971), *Point spectra of some mutually exciting point processes*, J. R. Statist. Soc. *B33*, 438–443.

Huang, W.G., Wang, J.H., and Huang, B.S. *et al.* (2001), *Estimates of source parameters for the 1999 Chi-Chi, Taiwan, earthquake based on Brune's source model*, Bull. Seismol. Soc. Am. *91(5)*, 1190–1198.

Kagan, Y.Y. and Knopoff L. (1980), *Spatial distribution of earthquakes: The two-point correlation function*, Geophysics. J. R. Astron. Soc. *62*, 303–320.

Liu, Z.Y. (1984), *Earthquake frequency and prediction*, Bull. Seismol. Soc. Am. *74(1)*, 255–265.

Ma, L. and Vere-Jones, D. (1997), *Application of M 8 and Lin-Lin Algorithms to New Zealand earthquake data*, New Zealand J. Geol. and Geophys. *40*, 77–89.

Ma, L. and Zhuang, C. (2001), *Relative quiescence within Jiashi Swarm in Xinjiang, China: An application of the ETAS point process model*, J. Appl. Probab. *38A*, 213–221.

Minqi, W., Eddy, J., and Fitzhugh, E. (1995), *Application of odds ratio and logistic models in epidemology and health research*, Health Values *19*, 59–62.

Ogata, Y., Akaike, H., and Katsura, K. (1982), *The application of linear intensity models to the investigation of causal relations between a point process and another point process*, Ann. Inst. Statist. Math. *34*, 373–387.

Ogata, Y. (1988), *Statistical models for earthquake occurrence and residual analysis for point process*, J. Am. Stat. Assoc. *83*, 9–27.

Ogata, Y. (1989), *Statistical models for standard seismicity and detection of anomalies by residual analysis for point processes*, Tectonophysics *83*, 9–27.

Ogata, Y. (1992), *Detection of relative quiescence before great earthquake occurrence, through a statistical model*, J. Geophys. Res. *97*, 19845–19871.

Ogata, Y.(1998), *Space-time point process models for earthquake occurrence*, Ann. Inst. Stat. Math. *50*, 379–402.

Ogata, Y. (1999), *Real time discrimination of forshocks (in Japanese)* Chikyu Monthly *24*, 167–173.

Omori (1894), *On the after-shocks of earthquakes*, J. College. Sci. Imp. Univ. Tokyo *7*, 111–200.

Papadopoulos, G., Ganas, A., and Plessa, A. (2002), *The Skyros (M_w6.5) of 26 July 2001 and Precursory seismicity patterns in the North Aegean Sea*, BSSA. *92*, 1141–1145.

Utsu, T. (1961), *Statistical study on the occurrence of aftershocks*, Geophys. Mag. *30*, 521–605.

UTSU, T. (1970), *The aftershock and earthquake statistic (II): Further Investigation of aftershocks and other earthquakes sequence based on a new classification of earthquake sequences*, J. Faculty Sci., Hokkaido Univ. Ser. VII(geophysics) 3, 198–266.

UTSU, T. (1977), *Probabilities in earthquake prediction*, Zisin Ser. *30(2)*, 179–185.

VERE-JONES, D. (1976), *Branching model for crack propagation*, Pure Appl. Geophys. *114*, 711–726.

VERE-JONES, D. (1978), *Earthquake prediction — statistics view*, J. Phys. Earth. *26,* 129–146.

VERE-JONES, D. (1989), *Predicting earthquakes by analyzing accelerating precursory seismic activity*, Pure Appl. Geophys. *130*, 661–686.

VERE-JONES, D. (2001), *The marriage of statistics and seismology*, J. Appl. Probab. *38A*, 1–5.

WANG, C.Y. and SHIN, T.C. (1998), *illustration 100 Years of Taiwan seismicity*, TAO *9(4)*, 589–614.

ZHUANG, J.C. (2000), *Statistical modeling of seismicity patterns before and after the Cape Palliser earthquake*, New Zealand J. of Geology and Geophys. *43*, 447–460.

ZHUANG, J.C., OGATA, Y., and VERE-JONES, D. (2002), *Stochastic declustering of space-time earthquake occurrences*, J. Amer. Stat. Assoc. *97*, 369–380.

(Received: December 3, 2004, revised December 26, 2005, accepted December 30, 2005)

Pure appl. geophys. 163 (2006) 2529–2543
0033–4553/06/122529–15
DOI 10.1007/s00024-006-0136-9

Pure and Applied Geophysics

Orientation-based Continuum Damage Models for Rocks

Xiaoyu Liu,[1] Xiangchu Yin,[2,3] and Naigang Liang[2]

Abstract—A general formulation of the Helmholtz free energy used in thermodynamics of damage process of rocks is derived within a multi-scale framework. Such a physically-based thermodynamic state potential has a hybrid, discrete/continuum, nature in the sense that it adopts a continuum description but subsumes the statistical ensemble average of the action of the entirety of microscopic degrees of freedom. The choice of the relevant damage variables results therefore directly from the breaking of contact cohesive bonds, and it naturally obeys the Clausius-Duhem inequality. Furthermore, motivated by the fact that the free energy is formulated by the integral of potentials independently defined on different orientations over the upper hemisphere, the damage evolution equation is formulated on a generic orientation. Consequently, the mechanical behavior of a rock material generally becomes anisotropic characteristics in the inelastic regime even if the material is initially isotropic, thus introducing dissipation-induced anisotropy in a very natural and simple way. Finally, the development of the lattice solid model can be cast into the framework of the orientation based continuum constitutive model.

Key words: Multi-scale modeling, thermodynamics, constitutive relation, damage, lattice solid model.

1. Introduction

As a naturally discontinuous, anisotropic and inhomogeneous geological material, the rocks are assemblages of blocks formed by connected fractures, and the interactions between the rock blocks and fractures (interfaces between blocks) are the main factors affecting the mechanical behavior of the fractured rock masses. Therefore the discrete approach, represented mostly at present by the distinct element method (DEM) and discontinuous deformation analysis (DDA) (Jing, 1998), has become prominent as a tool for providing more realistic representation of problem geometry of fractured rocks since fractures are explicitly represented, both geometrically and mechanically. Nevertheless, this advantage entails sacrificing the problem size since representations of individual blocks and fractures put large demands on computer memory and speed. Therefore, despite the popularity of the

[1] Division of Engineering Sciences, Institute of Mechanics, Chinese Academy Sciences, Beijing 100080, China. E-mail: liuxy@imech.ac.cn
[2] LNM, Institute of Mechanics, Chinese Academy Sciences, Beijing 100080, China.
[3] Center of Analysis and Prediction, China Earthquake Administration, Beijing 100036, China.

'discrete approach', its applicability to solve practical problems remains a challenging task.

Alternatively, the ultimate in reduction of fractured rock masses is represented by continuum field theory where all discrete degrees of freedom are replaced by continuous field variables obeying macroscopic differential equations. Because details of blocks and fracture systems are coarse grained out of the formulation, continuum models are not subject to length and time scale limitations, and the material parameters entering the models such as stresses and displacements are suited directly to measurements and senses. The constitutive relations, however, are often empirically determined, making their reliability over a wide range of deformations dubious.

The purpose of the current paper is, on one hand, to focus on multi-scale constitutive modeling for rocks. In this context, a multi-scale modeling methodology is presented naturally to combine the interactions of discrete micro-structures of a rock material on the small scale with the usual framework of continuum mechanics on the very large scale. By using such micromechanics theory, we develop an orientation-based continuum damage model having the following virtues: (1) the Helmholtz free energy is calculated exclusively from the microscopic contact constitutive relationship, the discrete micro-structures and the macroscopic kinematical variables without empirical phenomenological assumption; (2) the microscopic formation of damage is directly the result of the breaking of cohesive bonds, and it naturally obeys the laws of thermodynamics; (3) damage-induced anisotropy is incorporated into the constitutive model in a natural and simple way.

On the other hand, based on the fact that the free energy is expressed as the integral of potentials defined on different orientation over the upper hemisphere, and with the help of the damage evolution equations formulated on a generic orientation, the lattice solid model (NAPIER and DEDE, 1997; MORA and PLACE, 1999; LI *et al.*, 2000), which has been used to simulate fracture initiation and propagation in rocks as an attractive alternative to finite-element analysis, can be derived from the directional discretization of the continuum constitutive model. Consequently, the orientation-based constitutive model not only can be numerically implemented through the lattice solid model, but also offers a unified framework for the development of the lattice solid model.

Symbolic and index notations are used in the present paper; for instance, the strain tensor is denoted by either $\boldsymbol{E}$ or E_{ij}. In the symbolic notation, '·' and ':' stand for the first- and second-order contractions, '$\otimes$' denotes the tensor product.

2. Multi-scale Modeling of Rocks

Consider a rock mass consisting of abstract material points on the macroscopic scale indicated in Figure 1(a). On the microscopic scale, an infinitesimal neighbor-

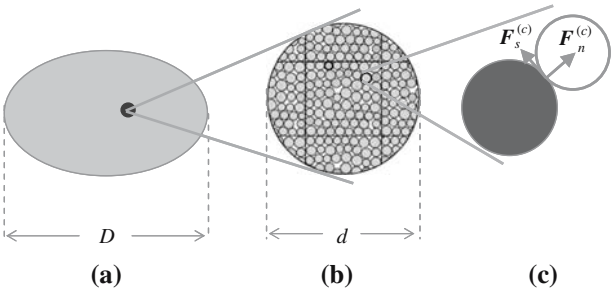


Figure 1
A rock body consisting of discrete block ensembles (RVEs): (a) Macrostructure, (b) microstructure and (c) contact blocks.

hood of a material point, namely the so-called representative volume element, is associated with an ensemble of discrete blocks formed by connected fractures, cf. Figure 1(b). Those blocks with sufficient large numbers are idealized as rigid elements in this paper. Denote the material configuration of the solid by A and the configuration at the present time τ by B. The distance vectors between a pair of adjacent elements a and b in configuration A and configuration B, $\boldsymbol{X}^{(a,b)}$ and $\boldsymbol{x}^{(a,b)}$, are written as

$$\boldsymbol{X}^{(a,b)} = \boldsymbol{X}^{(a)} - \boldsymbol{X}^{(b)}, \quad \boldsymbol{x}^{(a,b)} = \boldsymbol{x}^{(a)} - \boldsymbol{x}^{(b)}, \quad \text{with} \quad \boldsymbol{X}^{(a)} - \boldsymbol{X}^{(b)} = R^{(a,b)}\boldsymbol{N}^{(a,b)}, \tag{1}$$

where $R^{(a,b)}$ and $\boldsymbol{N}^{(a,b)}$ are the length and unit vector of $\boldsymbol{X}^{(a,b)}$, respectively.

As shown in Figure 1(c), the contact actions among adjacent elements are assumed to be concentrated at the contact points. The local stress in the contact region between elements a and b thus is replaced by concentrated force at the contact point c

$$\boldsymbol{F}^{(c)} = \boldsymbol{F}_n^{(c)} + \boldsymbol{F}_s^{(c)}, \tag{2}$$

where $\boldsymbol{F}_n^{(c)}$ is the normal force, and $\boldsymbol{F}_s^{(c)}$ is the shear force, while the local strain is replaced by relative displacement

$$\boldsymbol{u}^{(c)} = \boldsymbol{u}^{(a)} - \boldsymbol{u}^{(b)} = \boldsymbol{u}_n^{(c)} + \boldsymbol{u}_s^{(c)}, \tag{3}$$

where the normal relative displacement, $\boldsymbol{u}_n^{(c)}$, and the shear relative displacement, $\boldsymbol{u}_s^{(c)}$, are given by, respectively,

$$\boldsymbol{u}_n^{(c)} = (\boldsymbol{u}^{(a)} - \boldsymbol{u}^{(b)}) \cdot (\boldsymbol{N}^{(a,b)} \otimes \boldsymbol{N}^{(a,b)}), \tag{4a}$$

$$\boldsymbol{u}_s^{(c)} = (\boldsymbol{u}^{(a)} - \boldsymbol{u}^{(b)}) \cdot (\boldsymbol{I} - \boldsymbol{N}^{(a,b)} \otimes \boldsymbol{N}^{(a,b)}). \tag{4b}$$

As response functions for the contact actions, the following two types of linear elastic contact constitutive relations

$$\boldsymbol{F}_n^{(c)} = K^{(n)}\boldsymbol{u}_n^{(c)}, \qquad \boldsymbol{F}_s^{(c)} = K^{(s)}\boldsymbol{u}_s^{(c)} \tag{5}$$

are considered, in which $K^{(n)}$, $K^{(s)}$ are the respective elastic constants in directions normal and tangential to the contact. Correspondingly, the mean energy of the contact actions within the RVE in configuration B takes the form

$$\begin{aligned} e &= \frac{1}{V}\sum_{<a,b>} W(\boldsymbol{u}^{(a)} - \boldsymbol{u}^{(b)}) \\ &= \frac{1}{2V}\sum_{<a,b>}\left[\left(K^{(n)} - K^{(s)}\right)\left((\boldsymbol{u}^{(a)} - \boldsymbol{u}^{(b)})\cdot \boldsymbol{N}^{(a,b)}\right)^2 + K^{(s)}(\boldsymbol{u}^{(a)} - \boldsymbol{u}^{(b)})\cdot(\boldsymbol{u}^{(a)} - \boldsymbol{u}^{(b)})\right]. \end{aligned} \tag{6}$$

Here V is the volume of the RVE, $\langle a, b\rangle$ runs over all the contacts within the RVE.

In order to construct a continuum model on the macroscopic scale based on the discrete model on the microscopic scale, the crucial step lies in establishing a correspondence between the discrete and the continuum kinematical descriptors. For that, the deformation of element a within the RVE is split into two parts. In the first part the centroid $\boldsymbol{X}$ of the RVE moves to a spatial position $\boldsymbol{x}$ in a vector of macro-displacements $\boldsymbol{u}$, while in the second, element a having a relative position $\Xi^{(a)}$ with respect to the centroid moves to a new relative spatial position $\xi^{(a)}$ in a vector of micro-displacements $v^{(a)}$, cf. Figure 2. Following the assumption that the size of the RVE is assumed to be conceptually infinitesimal on the macroscopic scale, but is much larger than the size of discrete blocks on the microscopic scale, the relative position $\Xi^{(a)}$ of the element a to the centroid of the ''RVE'', which is not necessarily small on the microscopic scale, must be small from the macroscopic view of point. As a result, the displacement of element a can be approximated by a polynomial expansion, i.e.,

$$\begin{aligned} \boldsymbol{u}^{(a)}(\boldsymbol{x}^{(a)}) &= \boldsymbol{u}(\boldsymbol{x} + \Xi^{(a)}) \\ &= \boldsymbol{u}(\boldsymbol{x}) + (\boldsymbol{u}(\boldsymbol{x})\otimes\nabla_X)\cdot\Xi^{(a)} + \frac{1}{2}(\boldsymbol{u}(\boldsymbol{x})\otimes\nabla_X\otimes\nabla_X) : (\Xi^{(a)}\otimes\Xi^{(a)}) + \cdots, \end{aligned} \tag{7}$$

where ∇_X denotes the gradient operator with respect to $\boldsymbol{X}$. When constructing a continuum model accounting for short-range interactions, it suffices to consider homogeneous deformations, which implies the second and higher order derivatives of the displacement field can be neglected. In this case, the above equation becomes

$$\begin{aligned} \boldsymbol{u}^{(a)}(\boldsymbol{x}^{(a)}) &\approx \boldsymbol{u}(\boldsymbol{x}) + (\boldsymbol{u}(\boldsymbol{x})\otimes\nabla_X)\cdot\Xi^{(a)} \\ &= \boldsymbol{u}(\boldsymbol{x}) + \left[\frac{1}{2}(\boldsymbol{u}(\boldsymbol{x})\otimes\nabla_X + \nabla_X\otimes\boldsymbol{u}(\boldsymbol{x})) + \frac{1}{2}(\boldsymbol{u}(\boldsymbol{x})\otimes\nabla_X - \nabla_X\otimes\boldsymbol{u}(\boldsymbol{x}))\right]\cdot\Xi^{(a)}, \end{aligned} \tag{8}$$

where the displacement gradient u_{ij} is decomposed into a symmetric part and an antisymmetric part. In many homogenization studies on granular assemblies the

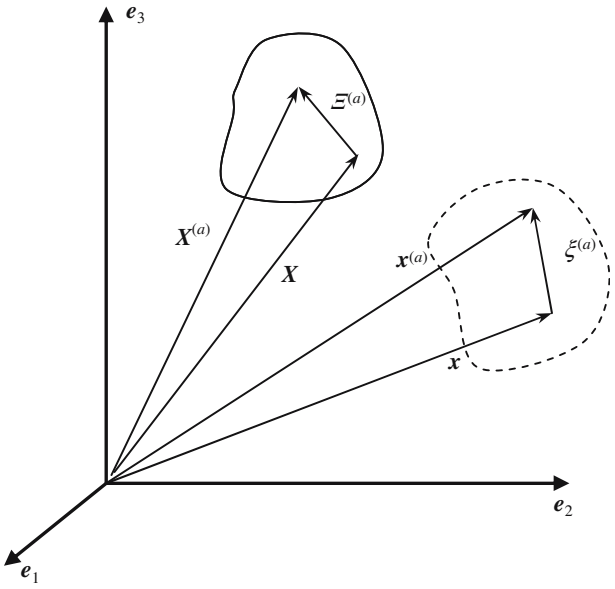


Figure 2
Schematic of the deformation of element a within the RVE.

contribution of the antisymmetric part is neglected (CHRISTOFFERSON *et al.*, 1981; WALTON, 1987; BATHURST and ROTHENBURG, 1988). Accordingly, the relative displacement vector between elements a and b is rewritten as

$$\boldsymbol{u}^{(a)} - \boldsymbol{u}^{(b)} \approx E(\boldsymbol{x}) \cdot (\boldsymbol{X}^{(a)} - \boldsymbol{X}^{(b)}), \quad \boldsymbol{E}(\boldsymbol{x}) = \frac{1}{2}(\boldsymbol{u}(\boldsymbol{x}) \otimes \nabla_{\boldsymbol{X}} + \nabla_{\boldsymbol{X}} \otimes \boldsymbol{u}(\boldsymbol{x})), \tag{9}$$

where $\boldsymbol{E}$ is the Lagrangean strain tensor. By substituting Eq. (9) into Eq. (6) and making use of Eq. (1), the energy density of the discrete system, expressed in index notation, is as follows

$$\begin{aligned} e &= \frac{1}{V} \sum_{<a,b>} W(R^{(a,b)} \boldsymbol{E} \cdot \boldsymbol{N}^{(a,b)}) \\ &= \frac{(K^{(n)} - K^{(s)}) E_{ij} E_{st}}{2V} \sum_{<a,b>} (R^{(a,b)})^2 N_i^{(a,b)} N_j^{(a,b)} N_s^{(a,b)} N_t^{(a,b)} \\ &\quad + \frac{K^{(s)} E_{ij} E_{st}}{8V} \sum_{<a,b>} (R^{(a,b)})^2 (\delta_{is} N_j^{(a,b)} N_t^{(a,b)} + \delta_{it} N_j^{(a,b)} N_s^{(a,b)} \\ &\quad + \delta_{js} N_i^{(a,b)} N_t^{(a,b)} + \delta_{jt} N_i^{(a,b)} N_s^{(a,b)}). \end{aligned} \tag{10}$$

Furthermore, we require that the power stored within the RVE is equal to the total energy spent over the material neighborhood occupying the same region. As

consequences of the energy equivalence between micro and macro scale, the Helmholtz free energy without energy dissipation is calculated exclusively from the microscopic contact constitutive relationship, the discrete micro-structures and the macroscopic kinematical variables, and does not involve an empirical phenomenological assumption.

3. Orientation-based Continuum Constitutive Model

3.1. Elastic Behavior

Before moving on to the more complex irreversible nonlinear behavior of rocks, it is worth describing the material response without energy dissipation. In terms of Fung's (FUNG, 1969) classification of the different forms of elastic model, all nondissipative materials are required to be hyperelastic since this is the only means by which the material is guaranteed to obey the First Law of Thermodynamics. Because of the fact that the hyperelastic behavior is entirely defined by a Helmholtz free energy function, a physical and general constitutive law can be obtained by virtue of Eq. (10). The stresses are given by

$$\sigma_{ij} = \frac{\partial e}{\partial E_{ij}} \tag{11}$$

and the tangent stiffness tensor obtained by further differentiation

$$C_{ijst} = \frac{\partial^2 e}{\partial E_{ij} \partial E_{st}}. \tag{12}$$

Since the energy density in Eq. (10) is formulated by a summation form, the calculation of the constitutive laws thus need to extend over all the relative distance vectors within the RVE. For a material with randomly distributed geometries of discrete microstructures, the calculation is often unfeasible and unnecessary. It becomes necessary to replace the summation setting in Eq. (10) by its continuum surrogates. The latter removes the discrete degrees of freedom but subsumes the statistical ensemble average of the action of the entirety of microscopic degrees of freedom. For that, select a spherical coordinate system (R, θ, ϕ) such that a distance vector is characterized by a corresponding position in the upper hemispherical coordinate system.[1] As a result, the discrete summation expression of Eq. (10) can be translated into an integral form by

[1] Because the potential energy density is independent of the direction of the distance vector, we herein require the distance vector to be oriented in the upper half spherical coordinate system.

$$
\begin{aligned}
e &= \int_0^{2\pi}\int_0^{\pi/2}\int_{R_1}^{R_2} D(R,\theta,\phi)W(R\boldsymbol{E}\cdot\boldsymbol{N})\sin\phi dR d\theta d\phi \\
&= \frac{(K^{(n)}-K^{(s)})E_{ij}E_{st}}{2}\int_0^{2\pi}\int_0^{\pi/2}\int_{R_1}^{R_2} D(R,\theta,\phi)R^2 N_i N_j N_s N_t \sin\phi dR d\theta d\phi \\
&\quad + \frac{K^{(s)}E_{ij}E_{st}}{8}\int_0^{2\pi}\int_0^{\pi/2}\int_{R_1}^{R_2} D(R,\theta,\phi)R^2(\delta_{is}N_jN_t + \delta_{it}N_jN_s \\
&\quad + \delta_{js}N_iN_t + \delta_{jt}N_iN_s)\sin\phi dR d\theta d\phi. \qquad (13)
\end{aligned}
$$

Here $D(R,\theta,\phi)$ is the distribution function of the positions in the upper hemispherical coordinate system which is located by the relative distance vectors, and reads

$$
D(R,\theta,\phi) = \sum_I\sum_J\sum_K \frac{M(R^{(I)},\theta^{(J)},\phi^{(K)})}{2V\sin\phi}\delta(R-R^{(I)})\delta(\theta-\theta^{(J)})\delta(\phi-\phi^{(K)}), \qquad (14)
$$

where I, J and K extend over all the positions located by the distance vectors in the spherical coordinate system, respectively. δ is the Dirac delta function, and $M(R^{(I)},\theta^{(J)},\phi^{(K)})$ is the number of the distance vectors locating at site $(R^{(I)},\theta^{(J)},\phi^{(K)})$. In the spherical coordinates N_i is represented as

$$
N_1 = \sin\phi\cos\theta, \quad N_2 = \sin\phi\sin\theta, \quad N_3 = \cos\phi. \qquad (15)
$$

Particularly, for isotropic assemblages of blocks with fully randomized relative distance vectors, the distribution function D only depends on R. Making use of the following three integration formulae

$$
\begin{aligned}
&\alpha_0 = \int_{R_1}^{R_2} D(R)R^2 dR \\
&\int_0^{2\pi}\int_0^{\pi/2} N_iN_j\sin\phi d\theta d\phi = \frac{2\pi}{3}\delta_{ij} \\
&\int_0^{2\pi}\int_0^{\pi/2} N_iN_jN_sN_t\sin\phi d\theta d\phi = \frac{2\pi}{15}(\delta_{ij}\delta_{st}+\delta_{is}\delta_{jt}+\delta_{it}\delta_{js}) \qquad (16)
\end{aligned}
$$

an analytical evaluation of Eq. (13) is given by

$$
e = \alpha_0\Big(\frac{\pi}{15}(K^{(n)}-K^{(s)})(\delta_{ij}\delta_{st}+\delta_{is}\delta_{jt}+\delta_{it}\delta_{js}) + \frac{\pi}{6}K^{(s)}(\delta_{is}\delta_{jt}+\delta_{it}\delta_{js})\Big)E_{ij}E_{st}. \qquad (17)
$$

After substituting Eq. (17) into Eq. (12), we can obtain the tangent stiffness tensor

$$
C_{ijst} = \frac{\pi}{15}\alpha_0\Big(2(K^{(n)}-K^{(s)})(\delta_{ij}\delta_{st}+\delta_{is}\delta_{jt}+\delta_{it}\delta_{js}) + 5K^{(s)}(\delta_{is}\delta_{jt}+\delta_{it}\delta_{js})\Big). \qquad (18)
$$

Comparison of Eq. (18) and the standard isotropic linear elastic constitutive equation (LANDAU and LIFSHITZ, 1970)

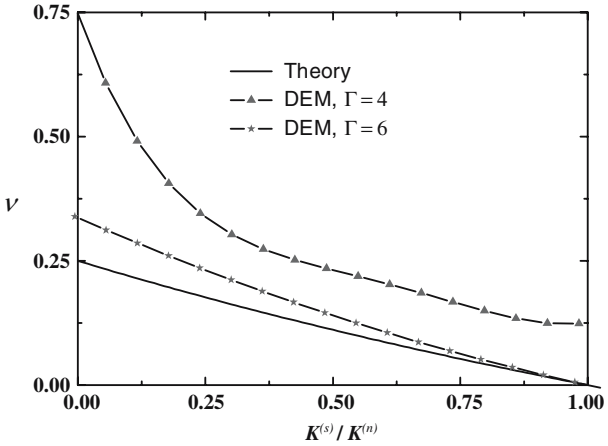


Figure 3
Showing the analytical results for the effect of contact stiffness ratio $K^{(s)}/K^{(n)}$ on effective Poisson ratio v compared to the discrete element results.

$$C_{ijst} = \lambda\delta_{ij}\delta_{st} + \mu(\delta_{is}\delta_{jt} + \delta_{it}\delta_{js}) \tag{19}$$

leads to the following equivalent relationships between the elastic constants $K^{(n)}, K^{(s)}$ and the macroscopic Yang's modulus E and Poisson ratio v, respectively,

$$K^{(n)} = \frac{3}{2\pi\alpha_0}\frac{E}{1-2v}, \quad K^{(s)} = \frac{3}{2\pi\alpha_0}\frac{E(1-4v)}{(1+v)(1-2v)}. \tag{20}$$

According to Eq. (20), the influence of the contact stiffness ratio $K^{(s)}/K^{(n)}$ on the macroscopic Poisson ratio v is plotted in Figure 3. It shows that v is zero when $K^{(s)}$ is equal to $K^{(n)}$. This is in agreement with the discrete element simulations for dense assemblies (the average number of contacts per particle $\Gamma = 6$) (KRUYT and ROTHENBURG, 2004). The analytical data in Figure 3 also show that v is not larger than 0.25 for positive shear stiffness. Although this is not very close to the numerical data, it should be noted that for most granular materials, Poisson ratio is in the range 0.2–0.25. Therefore, we argue that for a statistical homogeneous rigid system, in which the interactions are concentrated at the contact points, the influence of the parameters $K^{(n)}, K^{(s)}$ on the macroscopic behaviour follows Eq. (20).

3.2. Dissipation and Internal State Variables

Irreversible nonlinear behavior of rock masses is the consequence of energy dissipation which is activated during the loading process, and can be traced largely to the contact breakage. In this paper, we assume that all damage dissipation in rocks results from a continuous process of the breakage of contact cohesive bonds between

the blocks, and an internal variable characterizing the state of the bond between elements a and b at the present time τ

$$q^{(a,b)} = \begin{cases} 0 & \text{The failure criterion has not been satisfied during loading process} \\ 1 & \text{The failure criterion has been satisfied during loading process} \end{cases} \tag{21}$$

is introduced such that the breaking of this contact bond is modeled by irreversibly modifying the contact potential to zero when a certain breaking rule has been satisfied, i.e.,

$$W^{(a,b)} = \left(1 - q^{(a,b)}\right) W(R^{(a,b)} \boldsymbol{E} \cdot \boldsymbol{N}^{(a,b)}). \tag{22}$$

Consequently, the discrete summation form of the free energy is represented as

$$e = \frac{1}{V} \sum_{<a,b>} (1 - q^{(a,b)}) W(R^{(a,b)} \boldsymbol{E} \cdot \boldsymbol{N}^{(a,b)}) \tag{23}$$

and its integral form is given by

$$e = \int_0^{2\pi} \int_0^{\pi/2} \int_{R_1}^{R_2} (D(R, \theta, \phi) - D^q(R, \theta, \phi)) W(R\boldsymbol{E} \cdot \boldsymbol{N}) \sin \phi dR d\theta d\phi, \tag{24}$$

where $D^q(R, \theta, \phi)$ is the distribution function of the positions in the upper hemispherical coordinate system which are located by all broken contact bonds at the current time.

Defining the energy functions on the orientation (θ, ϕ) as

$$\Phi(\theta, \phi, E \cdot N) = \int_{R_1}^{R_2} D(R, \theta, \phi) W(R\boldsymbol{E} \cdot \boldsymbol{N}) dR \tag{25}$$

and extracting the damage variable on that orientation,

$$d(\theta, \phi)\Phi(\theta, \phi, \boldsymbol{E} \cdot \boldsymbol{N}) = \int_{R_1}^{R_2} D^q(R, \theta, \phi) W(R\boldsymbol{E} \cdot \boldsymbol{N}) dR. \tag{26}$$

We introduce the general form of the Helmholtz free energy which is sufficient to describe all reversible processes in rock masses. The total free energy per unit volume depends on the observable variable $\boldsymbol{E}$ as well as the damage variable $d(\theta, \phi)$, and takes the following compact form

$$e = \int_\Omega (1 - d^\Omega) \Phi^\Omega (\boldsymbol{E} \cdot \boldsymbol{N}) d\Omega. \tag{27}$$

Applying the Clausius-Duhem inequality to the derived free energy

$$\sigma_{ij} \dot{E}_{ij} - \dot{e} \geq 0 \tag{28}$$

we arrive according to Eq. (27) at

$$\dot{E}_{ij}\left(\sigma_{ij} - \int_{\Omega}(1-d^{\Omega})\frac{\partial\Phi^{\Omega}(\boldsymbol{E}\cdot\boldsymbol{N})}{\partial E_{ij}}d\Omega\right) + \int_{\Omega}\dot{d}^{\Omega}\Phi^{\Omega}(\boldsymbol{E}\cdot\boldsymbol{N})d\Omega \geq 0. \tag{29}$$

This inequality must hold for arbitrary strain-rates. Therefore we can conclude that Eq. (29) and the following statements are equivalent:

- The stress strain relationship

$$\sigma_{ij} = \int_{\Omega}(1-d^{\Omega})\frac{\partial\Phi^{\Omega}(\boldsymbol{E}\cdot\boldsymbol{N})}{\partial E_{ij}}d\Omega \tag{30}$$

from which the expression for the strain-rates results again by time differentiation

$$\dot{\sigma}_{ij} = \dot{E}_{st}\int_{\Omega}(1-d^{\Omega})\frac{\partial^2\Phi^{\Omega}(\boldsymbol{E}\cdot\boldsymbol{N})}{\partial E_{ij}\partial E_{st}}d\Omega - \int_{\Omega}\dot{d}^{\Omega}\frac{\partial\Phi^{\Omega}(\boldsymbol{E}\cdot\boldsymbol{N})}{\partial E_{ij}}d\Omega. \tag{31}$$

- The damage dissipation inequality

$$\int_{\Omega}\dot{d}^{\Omega}\Phi^{\Omega}(\boldsymbol{E}\cdot\boldsymbol{N})d\Omega \geq 0. \tag{32}$$

Based on the fact that the energy stored in the contact cohesive bonds is positive and also, that the dissipative energy gradually increases during the damage processes, the definitions as shown in Eq. (26) apparently ensure that the damage dissipation inequality always holds on each orientation. In other words, the Clausius-Duhem inequality (28) is naturally fulfilled in the course of the physical choice of the damage variables.

Because accurate description of the local effects of the defects are not available for the material under investigation, at the present time information on d^{Ω} and its evolution has to depend on the development of recovery experiments. On the other hand, the orientation-independent nature of the free energy function Φ^{Ω} in Eq. (32) motivates us to assume that the damage dissipation also evolves independently on each orientation, thus identifying Φ^{Ω} as the thermodynamic force conjugate to d^{Ω}. It should be noted that this basic assumption overlaps with certain concepts in which multiple-plane representations of inelasticity are derived (Seaman and Dein, 1983; Bažant and Gambarova, 1984; Ju and Lee, 1991a,b; Espinosa and Brar, 1995; Pensée *et al.*, 2002).

In the context of the irreversible thermodynamics theory (Lemaitre, 1992; Maugin, 1992), a general form of the damage evolution equations on a generic orientation can consequently be formulated through a dissipation potential function in the space of Φ^{Ω} in combination with the maximum dissipation principle (Simo and Hughes, 1998). The detailed expressions are given as follows:

- Define a damage potential function $\pi^{\Omega}(\Phi^{\Omega}, d^{\Omega})$ on a generic orientation such that this damage potential function coincides with the damage surface $\pi^{\Omega}(\Phi^{\Omega}, d^{\Omega}) = 0$ on that orientation. To ensure that the damage evolution follows the path in the damage variable space that maximizes the dissipation on that orientation, we gain the evolution law

$$\dot{d}^{\Omega} = \dot{\lambda}^{\Omega} \frac{\partial \pi^{\Omega}(\Phi^{\Omega}, d^{\Omega})}{\partial \Phi^{\Omega}}, \tag{33}$$

where parameter $\dot{\lambda}_{\Omega}$ is determined from the consistency condition $\dot{\pi}_{\Omega} = 0$:

$$\dot{\lambda}^{\Omega} = -\dot{\Phi}^{\Omega} / \frac{\partial \pi^{\Omega}(\Phi^{\Omega}, d^{\Omega})}{\partial d^{\Omega}}. \tag{34}$$

- The loading/unloading conditions are defined by the condition that each of the following Kuhn-Tucker relations always should be satisfied (Ju, 1990)

$$\dot{\lambda}^{\Omega} \geq 0, \quad \pi^{\Omega}(\Phi^{\Omega}, d^{\Omega}) \leq 0, \quad \dot{\lambda}^{\Omega} \pi^{\Omega}(\Phi^{\Omega}, d^{\Omega}) = 0. \tag{35}$$

Given the evolution equations of d^{Ω} on a generic orientation, the overall responses of brittle materials result from the cumulative effect of processes taking place on all possible orientations in the upper hemisphere. Therefore, Eqs. (27), (30), (31), (33), (34) and (35) furnish a closed constitutive formulation for rocks, which we identify as the orientation-based continuum damage model in this paper. As a consequence of the independent evolution of dissipation processes on different orientations, the mechanical behavior of materials generally becomes anisotropic characteristics in the inelastic regime even if the material is initially isotropic, thus introducing damage-induced anisotropy in a very natural and simple way.

4. Lattice Solid Model

Subdivide the upper hemispherical envelop into M discrete contiguous solid angles, $\Delta\Omega^{(m)}$, each associated with a corresponding discrete orientation $N_i^{(m)}$. The total energy of the infinitesimal neighborhood of the material point can thus be estimated approximately by the sum of a finite number of discrete components defined on M orientations, i.e.,

$$Ve = V \int_{\Omega} (1 - d^{\Omega}) \Phi^{\Omega} (\boldsymbol{E} \cdot \boldsymbol{N}) d\Omega \approx \sum_{m=1}^{M} (1 - d^{(m)}) \Phi^{(m)} (\boldsymbol{E} \cdot \boldsymbol{N}^{(m)}) V \Delta\Omega^{(m)}. \tag{36}$$

Because both the evolution of damage variables $d^{(m)}$ and the calculation for the free energy functions $\Phi^{(m)}$ are independent of the orientations, we can analogously construct a lattice cell with the total of M rod elements, in which the m-th rod element with orientation $N_i^{(m)}$, cross section $A^{(m)}$ and length $l^{(m)}$ makes the

deformation compatible with $\boldsymbol{E}\cdot\boldsymbol{N}^{(m)}$, obeys the evolution laws of $d^{(m)}$, and has the energy

$$\left(1-d^{(m)}\right)\Phi^{(m)}(\boldsymbol{E}\cdot\boldsymbol{N}^{(m)})A^{(m)}l^{(m)}. \tag{37}$$

In this way, once the following equation

$$V\Delta\Omega^{(m)}=A^{(m)}l^{(m)} \tag{38}$$

holds, the infinitesimal neighborhood of the material point can be replaced with the energy equivalent lattice cell. This counterpart is considerably coarser than the characteristic length of the block, but can approximately perform the macroscopic response of the assemblage of blocks with the enormously large number of degrees of freedom under the macroscopic strain $\boldsymbol{E}$.

For example, we estimate the energy using 13 integration points oriented in $\{1,0,0\}+\{1,1,0\}+\{1,1,1\}$ directions on the envelope of the upper hemisphere, and the values of the corresponding solid angles are given by

$$\begin{aligned}
&\Delta\Omega^{[1,0,0]}=\Delta\Omega^{[0,1,0]}=\Delta\Omega^{[0,0,1]}=\frac{10}{51}\pi\\
&\Delta\Omega^{[1,1,0]}=\Delta\Omega^{[1,0,1]}=\Delta\Omega^{[0,1,1]}=\Delta\Omega^{[1,\bar{1},0]}=\Delta\Omega^{[\bar{1},0,1]}=\Delta\Omega^{[0,\bar{1},1]}=\frac{12}{85}\pi\\
&\Delta\Omega^{[1,1,1]}=\Delta\Omega^{[\bar{1},1,1]}=\Delta\Omega^{[1,\bar{1},1]}=\Delta\Omega^{[\bar{1},\bar{1},1]}=\frac{12}{85}\pi
\end{aligned} \tag{39}$$

It can be proved that this 13-point formula can satisfy the zeroth moment constraint, i.e.,

$$\begin{aligned}
&\int_{\Omega}d\Omega=\sum_{m=1}^{13}\Delta\Omega^{(m)}\\
&\int_{\Omega}N_iN_j\,d\Omega=\sum_{m=1}^{13}N_i^{(m)}N_j^{(m)}\Delta\Omega^{(m)}\\
&\int_{\Omega}N_iN_jN_sN_t\,d\Omega=\sum_{m=1}^{13}N_i^{(m)}N_j^{(m)}N_s^{(m)}N_t^{(m)}\Delta\Omega^{(m)}
\end{aligned} \tag{40}$$

Correspondingly, a cubic lattice cell is constructed, which consists of 13 rod elements oriented in$\{1,0,0\}+\{1,1,0\}+\{1,1,1\}$directions, as shown in Figure 4(a). Let the side length of the cubic be a. The lengths of the rod elements are $a, a\sqrt{2}$ and $a\sqrt{3}$, respectively.

In particular, for isotropic assemblages of blocks, the energy of the m-th rod element is derived from Eqs. (13), (16), (20), (37) and (38) as

$$\frac{a^3\Delta\Omega^{(m)}}{2(l^{(m)})^2}\left[\frac{3}{2\pi}\frac{E}{1-2v}\delta u_i^{(m)}\delta u_i^{(m)}+\frac{3}{2\pi}\frac{E(1-4v)}{(1+v)(1-2v)}\delta v_i^{(m)}\delta v_i^{(m)}\right], \tag{41}$$

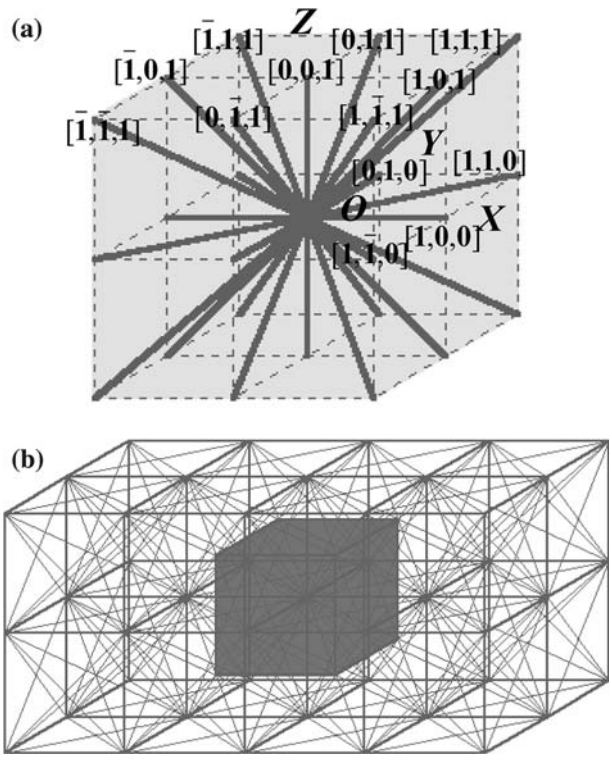


Figure 4
(a) Three-dimensional regular lattice unit cell; (b) the periodic lattices made up of the unit cell.

where

$$\delta u_i^{(m)} = l^{(m)} E_{st} N_i^{(m)} N_s^{(m)} N_t^{(m)}, \quad \delta v_i^{(m)} = l^{(m)} E_{ij} N_j^{(m)} - \delta u_i^{(m)}. \tag{42}$$

As a result, the elastic normal stiffness $K_n^{(m)}$ and the shear stiffness $K_s^{(m)}$ of the rod element with orientation $N_i^{(m)}$ have the following format

$$k_n^{(m)} = \frac{3Ea^3 \Delta\Omega^{(m)}}{2\pi (l^{(m)})^2 (1-2v)}, \quad k_s^{(m)} = \frac{3Ea^3 \Delta\Omega^{(m)} (1-4v)}{2\pi (l^{(m)})^2 (1+v)(1-2v)}. \tag{43}$$

Taking this three-dimensional lattice cell as the unit cell of a periodic discrete lattice shown in Figure 4(b), the constructed lattice solid model can be conceived as an alternative manner of the numerical implementation of the orientation-based constitutive model, instead of the finite-element method.

5. Conclusions

We present a multiscale modeling methodology to describe the transition from micro to macro domain of rock, and formulate the Helmholtz free energy for a non-dissipative case.

By physically introducing the damage variables into the Helmholtz free energy, we construct the orientation-based continuum damage model, which is highly promising for the solution of engineering problems involving a damage process with a greatly simplified mathematical description.

The discretization of the orientation-based constitutive model furnishes a unified framework for the development of the lattice solid model.

Acknowledgements

We gratefully acknowledge the support from National Natural Science Foundation of China under grant Nos. 10232050 and 10572140, Special Funds for Major state Basic Research Project under Grant Nos. 2002CB412703 and 2002CB412706.

References

Bathurst, R.J. and Rothenburg, L. (1988), *Micromechanical aspects of isotropic granular assemblies with linear contact interactions*, J. Appl. Mech. *55*, 17–23.

Bažant, Z. and Gambarova. P. (1984), *Crack shear in concrete: Crack band microplane model*, ASCE J. Eng. Mech. *110*, 2015–2035.

Christofferson, J., Mehrabadi, M.M., and Nemat-Nasser, S. (1981), *A micromechanical description of granular material behavior*, J. Appl. Mech. *48*, 339–344.

Espinosa, H.D. and Brar, N.S. (1995), *Dynamic failure mechanisms of ceramic bars: experiments and numerical simulations*, J. Mech. Phys. Solids. *43*, 1615–1638.

Fung, Y.C., *Foundations of Solid Mechanics* (Prentice, New Jersey 1969).

Jing, L. (1998), *Formulations of discontinuous deformation analysis for block systems*, Eng. Geology. *49*, 371–381.

Ju, J.W. (1990), *Isotropic and anisotropic damage variable in continuum damage mechanics*, ASCE J. Eng. Mech. *116*, 2764–2770.

Ju, J.W. and Lee, X. (1991a), *On three-dimensional self-consistent micromechanical damage models for brittle solids, Part I: Tensile loadings*, ASCE J. Eng. Mech. *117*, 1495–1514.

Ju, J.W. and Lee, X. (1991b), *On three-dimensional self-consistent micromechanical damage models for brittle solids, Part II: Compressive loadings*, ASCE J. Eng. Mech. *117*, 1515–1536.

Kruyt, N.P. and Rothenburg, L. (2004), *Kinematic and static assumptions for homogenization in micromechanics of granular materials*, Mech. Mater. *36*, 1157–1173.

Landau, L.D. and Lifshitz, E.M., *Theory of Elasticity* (Pergamon, Oxford 1970).

Lemaitre, J., *A Course on Damage Mechanics* (Springer, Berlin 1992).

Li, L.H., Bai, Y.L., Xia, M.F., Ke, F.J., and Yin, X.C. (2000), *Damage localization as a possible precursor of earthquake rupture,* Pure Appl. Geophys. *157*, 1929–1943.

Maugin, G.A., *The Thermomechanics of Plasticity and Fracture* (Cambridge University Press, Cambridge 1992).

Mora, P. and Place D. (1999), *The weakness of earthquake faults*, Geophys. Res. Lett. *26*, 123–126.

NAPIER, J.A.L. and DEDE, T.A. (1997), *Comparison between random mesh schemes and explicit growth rules for rock fracture simulation*, Int. J. Rock Mech. Min. Sci. *34*, 356.

PENSÉE, V., KONDO, D., and DORMIEUX, L. (2002), *A micromechanical analysis of anisotropic damage in brittle materials*, ASCE J. Eng. Mech. *128*, 889–897.

SEAMAN, L. and DEIN, J.L., *Representing shear band damage at high strain rates*. In *Proc. IUTAM Symposium on Nonlinear Deformation Waves* (ed. Tallin, Estonia) (Springer-Verlag, Berlin 1983).

SIMO, J.C. and HUGHES, T.J.R., *Computational inelasticity*. In *Interdisciplinary Applied Mathematics* (Springer, New York 1998).

WALTON, K. (1987), *The effective elastic moduli of a random packing of spheres*, J. Mech. Phys. Solids. *35*, 213–226.

(Received November 17, 2004, revised December 23, 2005, accepted December 28, 2005)

Pure appl. geophys. 163 (2006) 2545–2560
0033–4553/06/122545–16
DOI 10.1007/s00024-006-0139-6

Pure and Applied Geophysics

Experimental Study of the Earthquake Recurrence Period and the Trend of Post-seismic Development

Bin Wang,[1,2] Zhaoyong Xu,[2] Xiangchu Yin,[3] Yili Hu,[4] Runhai Yang,[2] Jingguan Cai,[2] and Shunyun Chen[2]

Abstract—In order to study the earthquake recurrence and the characteristics of earthquake series, rupture tests of rock samples and plexiglass samples were made. On rock samples, a number of acoustic emission (AE) and strain measuring points were deployed; the load was one side direct shear. The variation characteristics of AE and strain at different detecting points around the extra large fracture were observed and studied. On plexiglass samples, a series of inclined cracks were prefabricated by a small-scale compressive testing machine. The samples were then loaded on a shockproof platen, when the samples were loaded, the stress intensity factor (SIF) was determined by the laser interferometric technique and shadow optical method of caustics. The fracture conditions such as material toughness around the extra large fracture were also studied. From those experimental results and the theory of fracture mechanics, the earthquake recurrence period and the trend of post-seismic development were studied.

Key words: Earthquake period, post-seismic trend, experimental study, stress intensity factor, strain energy, low-frequency wave.

Introduction

After an earthquake has occurred, will still greater shocks occur? This problem is very important, both for the victims and for the government. Many experts have studied the types of earthquakes' series. The most basic types are the fore-main-aftershock type, fore-mainshock type, main-aftershock type and swarm type. They have also proposed methods to judge the earthquake type. For example, the well-known Mogi's experiments (Mogi, 1962, 1963) simulated the foreshock-mainshock-aftershock type by the fracture of homogeneous materials. Although

[1] School of Earth and Space Sciences, University of Science and Technology of China, Hefei 230026, PRC. E-mail: wbin8@163.com
[2] Disaster Prevention Research Institute of Yunnan Province, Kunming 650041, PRC. E-mail: zhaoyongx@126.com; renhaiyang@163.com
[3] Institute of Mechanics, Chinese Academy of Science, LNM, Beijing 100080, PRC. E-mail: xcyinlurr@vip.163.com; xcyin@public.bta.net.cn
[4] Department of Geophysics, Yunnan University, Kunming 650091, PRC. E-mail: ylhu929@163.com

there are some drawbacks in his experiments examined from the present point of view, he has set a precedent for studying the types of earthquakes by experiments. The principle used for seismic zonation is that earthquakes are likely to occur repeatedly at the same place or in its vicinity. In earthquake prediction, the concept of long-term (such as 20∼50 a) alternation of earthquake active and quiet periods is frequently used. There are even some results of special studies of the recurrence period of earthquakes (MEI SHIRONG, 1960; SHI ZHENLIANG *et al.*, 1984; XU SHAOXIE, 1982; ZHANG GUOMIN, 1987; ZHANG GUOMIN *et al.*, 1993; CAI JINGGUAN and LIU ZHENGRONG, 1990). However, being limited by the conditions, experimental and theoretical confirmation of these results has not been done fully. In view of this, we improved the experimental conditions and designed some new experiments. First, rupture tests of rock samples, on which a number of AE and strain measuring points were deployed, were made by one side direct shear; the purpose was to induce ruptures of the samples producing various types of earthquakes. Second, plexiglass samples that contain a number of prefabricated inclined cracks were loaded by a small-scale compressive testing machine which was placed on a shockproof platen; the SIFs were determined by the laser interferometric technique and shadow optical method of caustics and the signals of microcracking were detected. At the same time, some indicative physical quantities in those experiments were recorded. These results were taken as an experimental basis, consequently a theoretical study was made.

1. Experimental Methods

1.1. One Side Direct Shear Test of Rock Samples

Direct shear tests were made on rectangular parallelepiped rock samples. The sample materials were sandstone and granite, which were pieced together into rectangular parallelepiped 80∼120 mm by 40∼60 mm by 60∼80 mm in size. The strength of sandstone is lower and that of granite is higher. On the free side surfaces of the sample, 12 strain measuring points and 8 AE measuring points were deployed (see Fig. 1b). The sample was placed on a true triaxial compressive testing machine, its lower half was clamped firmly from the two free side surfaces. In the mid-height, a gap of 1∼3 mm wide was arranged in advance for shear fracture. A push was applied on one of the free side surfaces of the upper half of the sample, with the opposite surface left free. On the top surface, there was a combination of two steel blocks, with 2∼4 (depending on the size of sample) rollers in between; a pressure transducer was placed above it. A normal stress of about 4∼8 MPa in the vertical direction was applied first. The normal stress was kept stable after it had reached the prescribed value, and a push was applied in the horizontal direction to do the direct shear test up to fracture (see Fig. 1a). In the beginning, the loading rate was a bit slower, about 4 MPa/min. After the shear stress had reached about

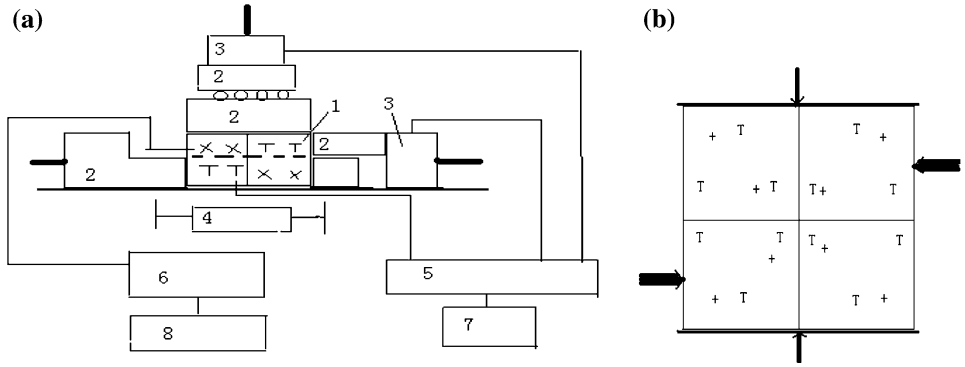


Figure 1
A block diagram showing the setup of direct shear test up to rupture (a), and a sketch of the distribution of measuring points (b). (a) 1. Rock sample (sandstone and granite pieced together); 2. Steel block; 3. Pressure transducer; 4. Displacement transducer; 5. Strainmeter; 6. Transient wave memory; 7, 8. Computer. (b) T: strain gage; + AE transducer; Vertical arrow: normal pressure; Horizontal thick arrow: shear load.

50% of the shear strength, the loading rate was raised. The strain gages were parallel to the horizontal push and were connected to the Wheatstone bridge circuit as the active gage in the half bridge. The LB-2 digital strainmeter designed by the Institute of Geology of China Seismological Administration was used to record the pressure, displacement and strain. The LB-2 digital strainmeter is capable of continuous sampling; it has 16 channels and each channel has a sampling rate that can be adjusted between 1~1,000 Hz and a resolution of 16 bit. In the meantime, the AE information was recorded synchronously by the microcrack-recording system that consists of detector, amplifier, transient wave memory, and so on (XU *et al.*, 2002a,b).

1.2. Laser Interferometric Experiment of Plexiglass Samples

The sample was made of plexiglass and was 200 mm by 200 mm by 15 mm in size. On the sample, a number of inclined penetrating cracks were cut by laser. The cracks make an angle of about 30° with the direction of the major principal stress σ_1 (σ_y) and are 35 mm long and 0.1~0.3 mm wide. The cracks (faults) were arranged in an *en echelon* array to form *Y*-shaped block boundaries. A small-scale compressive testing machine, capable of applying vertical and horizontal pressures respectively up to 300 KN, was installed on a shockproof platen. At the beginning of the experiment, horizontal and vertical pressures (in x and y directions, respectively) were synchronously applied to the sample. When the pressures had reached a certain degree (letting $\sigma_x = \sigma_y = 5$ MPa), σ_x was kept approximately unchanged. Then, σ_y was increased continuously until the final rupture of the sample. In another case, the

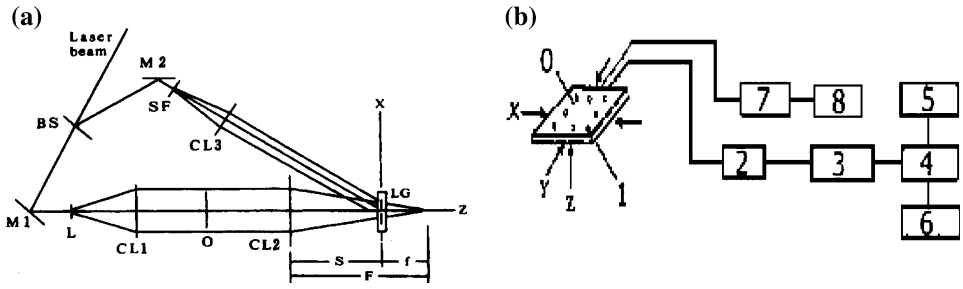


Figure 2
(a) Experimental setup of photography by real-time holographic interferometry BS: beam splitter; M1, M2: mirror; SF: spatial filter; L, CL1, CL2, CL3: lens; LG: liquid gate; O: transparent plexiglass sample. (b) Diagram showing the method of loading and the arrangement of transducers 0 Prefabricated crack; 1 PZT transducer; 2 amplifier; 3 transient wave memory; 4 and 8 computer; 5 printer; 6 plotter; 7 strainmeter.

sample was unloaded when σ_y had been increased to about 70% of the rupture stress. From the beginning of loading, the strain patterns were recorded photographically by the real-time holographic interferometry (Fig. 2a) (XIONG *et al.*, 2002); at the same time, the AE activity was recorded by the system for microcracking information storage and analysis (Fig. 2b) (XU *et al.*, 2002a,b).

2. Test Results

2.1. Characteristics of Stress-strain Variations in the Rock Sample During the Loading Process

Figure 3a shows the variations of stress and strain with time in the rock sample during the loading process. Among the curves, one shows the horizontal shear stress (the dark line in Fig. 3), and those remaining show the strains of measuring points on the rock sample. Figure 3b gives the stress-strain curve of horizontal shear. In Figure 3, except the tendentious variations, there are three large and obvious sudden changes, among which the latter two are even more remarkable; they respectively correspond to extra large fractures. When viewed from the figure, the general trend is that the shear force and strains at various points increase gradually with time. The modes of strain variation around the three extra fractures are not the same. After the first fracture, the strains of many points increased drastically; in contrast, after the latter two fractures, the strains of points all decreased drastically. In addition, when the three especially clear extra large fractures were impending, the various strain measuring points showed characteristic variations: the strain of some points increased or decreased following the original trend; the strain of other points changed the trend, from increasing to decreasing or the opposite; whereas there are

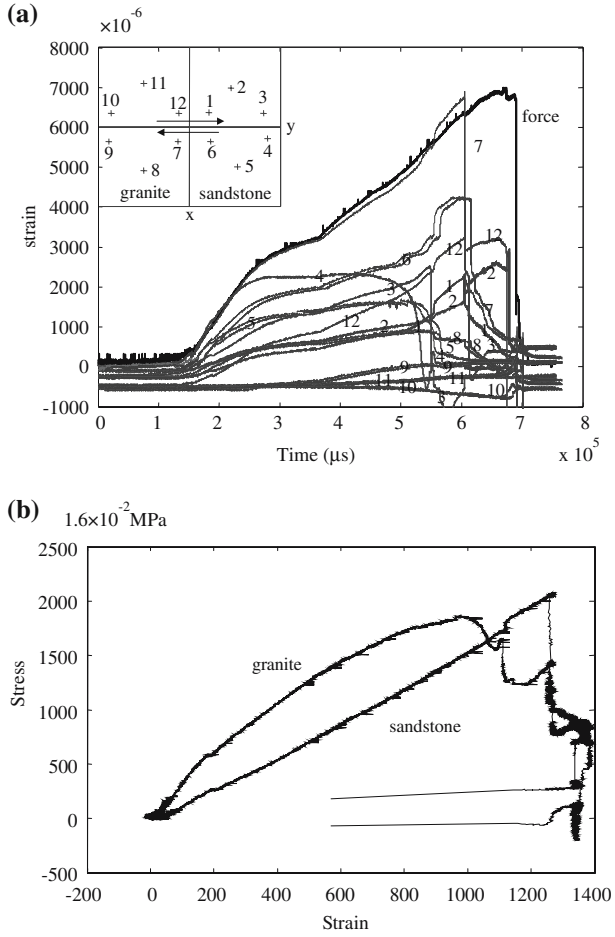


Figure 3
The stress-strain variations of rock sample during the loading process (a), and the stress-strain curve of horizontal shear (b).

still others the strain of which remained basically unchanged. XU *et al.* (1989; 2001a,b) and other authors (MOGI, 1984; HU *et al.*, 1997) have pointed out these characteristics of short-term and imminent strain precursors frequently. These characteristics appeared again in the present study; this has formed an important and feasible experimental basis for us to predict the time of earthquake occurrence.

2.2. Characteristics of the Variation of AE Frequency and Intensity with Time During the Loading Process of Rock Sample

The statistical averages of the number of AE events per 10 s were determined, with which the variation of AE rate with time during the loading process of rock

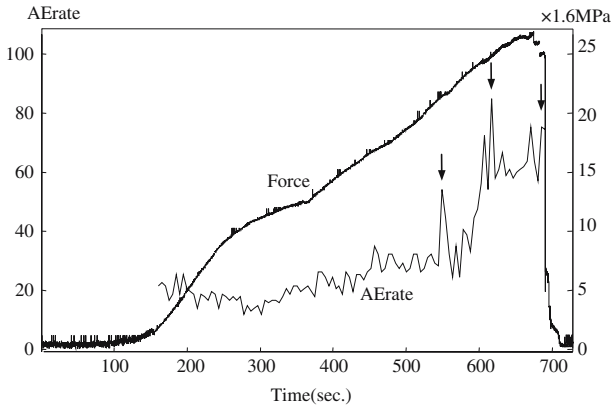


Figure 4
The variation of AE frequency with time during the loading process of rock sample.

samples was plotted, as shown in Figure 4. Arrows in the figure indicate the time position of occurrence of each large fracture. It can be seen from the figure that the number of microcracking events (AE rate) fluctuated more frequently (largely) when the extra large fracture was impending. There are some relatively longer time periods of AE active and quiet alternation; among them, three are most remarkable. Before each extra large fracture, the AE rate always increased (clustered) first, became relatively quiet next, then reactivated (clustered), and finally the extra large fracture occurred.

The intensity of extra event of AE is considerably higher than that accompanying other events. When viewed from the waveform record (as shown in the following), they were all amplitude-limited and lasted a very long time. In this paper, the size (intensity) of an event is determined according to the length of its time duration. The formula is

$$M = c + d\log(T_c - T_0), \tag{1}$$

where T_0 is the time of fracture occurrence; T_c is the ending time of the coda, i.e., the corresponding time when the amplitude becomes smaller than or equal to a certain specified value. In case that the record was amplitude-limited or incomplete, T_c can be determined by the formula given below. Assuming that the wave train amplitude of events obeys the rule

$$A = A_0 t^{-k}, \tag{2}$$

T_c can be determined once A_c is given.

Figure 5 shows the variation of AE intensity with time before and after the first extra large fracture during the loading process of rock sample. It is seen from the figure that in a period of time before the extra large fracture (the highest vertical line

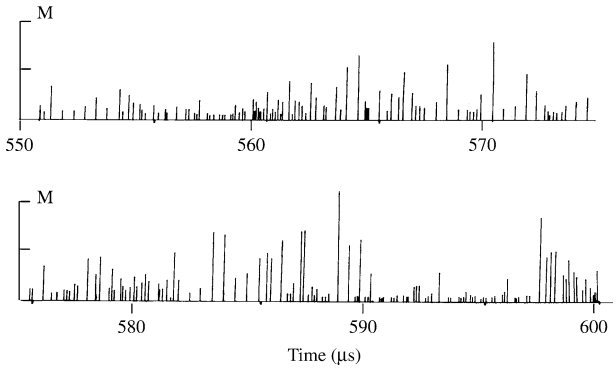


Figure 5
The magnitude variation of AE events with time before and after the first extra large fracture during the loading process of rock sample.

in the lower part of the figure), AE events were of a large number but low intensity; later on, they became less in number but were of higher intensity; finally, the extra large fracture occurred when the frequency of AE events increased again. Before the preceding large fracture, AE events also showed characteristics similar to those mentioned above. If there are very few large AE events but a few good small ones to occur after the large fracture, there might be another large fracture to follow; otherwise, no large fracture would reoccur. The status for the second and third extra fractures is similar to this, and hence the figures are omitted here.

2.3. Characteristics of the Variation of Microcracking Waveform During the Loading Process of Rock Sample

Before and after the three extra large fractures of rock sample, the waveforms of a few fractures are of basically similar characteristics. In the early stage, most of the fractures were of an isolated type. With the increase of stress and the impending extra large fracture, the types of fractures tended to be complete. There were various types of fractures to occur, such as the types of fore-main-after-fracture fore-main-fracture, main-after-fracture, and swarmed fracture. Figure 6a is the waveform record of a typical fore-main-after-fracture type. Other types can be regarded as part of it. In the swarmed-fracture type, the amplitudes (sizes) of events are close to one another. This type is thoroughly discussed elsewhere. Moreover, with the pending extra large fracture, there would be events that contain obviously more low-frequency components in the waveform to occur. Figure 6b is a record of this type. It is obvious that the low-frequency components in Figure 6b are more than those in Figure 6a. Figure 6c is the first extra large fracture and Figure 6d is the after-fracture nearest this extra large fracture. The waveform in Figure 6c is seriously amplitude-limited, showing that this event is of a much higher intensity than other events.

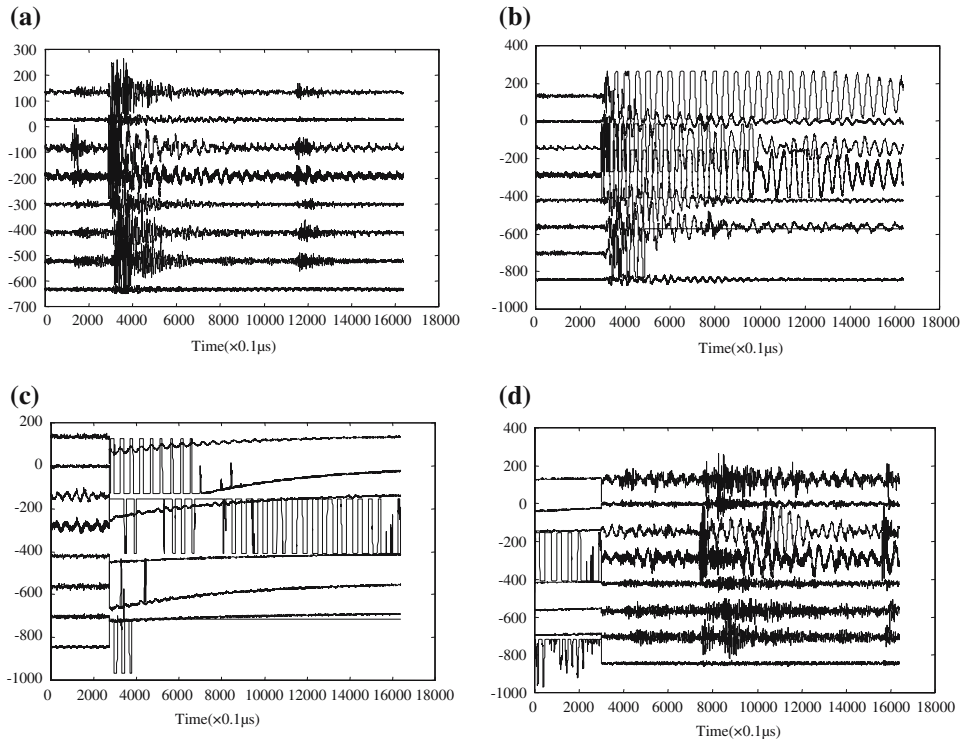


Figure 6
Waveform records of some fractures before and after the extra large fracture of rock sample (a) High-frequency (wave) fore-fracture; (b) Low-frequency (wave) fore-fracture; (c) The first extra large fracture; (d) After-fracture nearest the extra large fracture.

Figure 6d shows that there were not only high-frequency (wave) events, but also low-frequency (wave) events after the extra large fracture. It is noteworthy that, to the extent low-frequency (wave) events exist, there remains the possibility for large or even larger fractures to occur.

2.4. Characteristics of the Variation of Shadow Areas of Caustics During the Fracture Preparation Process of the Plexiglass Sample

In the experiment, the shadow areas of caustics were clear and obvious, as shown in Figure 7. At the crack tip, there is an obvious dark shadow area which has an obvious demarcation line with the high-density fringe area that neighbors it. The bright demarcation line at the crack tip is the caustic curve, and the dark shadow area is the shadow area of caustics (for the local characteristics, see Figure 8c). They grew with increasing load and reduced with decreasing load. Figure 7 shows that the shadow grew with the increase of pressure. Untill the late stage, shadows occurred

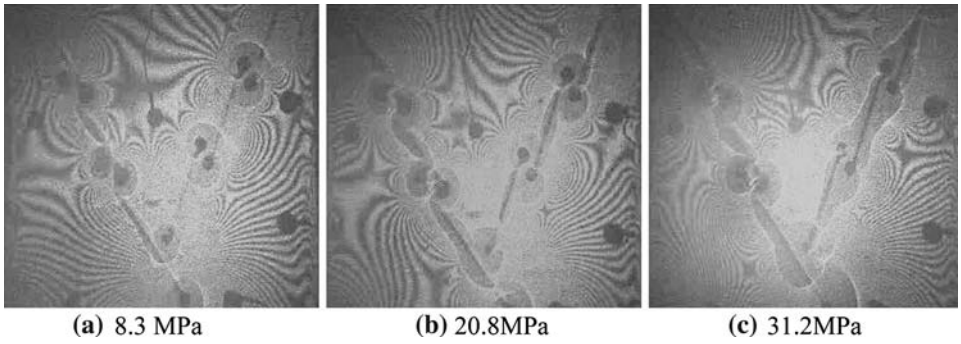


Figure 7
The variation of shadow areas of caustics with loading.

not only at the crack tip but also along the crack surface. The growth of these shadow areas means that the frictional action between crack surfaces had become fairly strong. In due time, the shadows between crack surfaces overlapped with those at the crack tip. When the shadow area of caustics grew from small to large to a certain degree, (micro) cracks occurred, which were just in the shadow area of caustics (for more details, see XU *et al.*, 2002a,b). After this, the shadow area of caustics no longer grew and kept approximately unchanged.

2.5. Characteristics of the Variation of Stress Intensity Factors During the Fracture Preparation Process of the Plexiglass Sample

Experimental mechanics shows that, for cracks of pure mode II, the typical caustic curves under different conditions are as shown in Figures 8a and 8b. Figure 8a gives the negative real image of transmitting light, negative virtual image of reflecting light, and so on, while Figure 8b gives the positive real image of

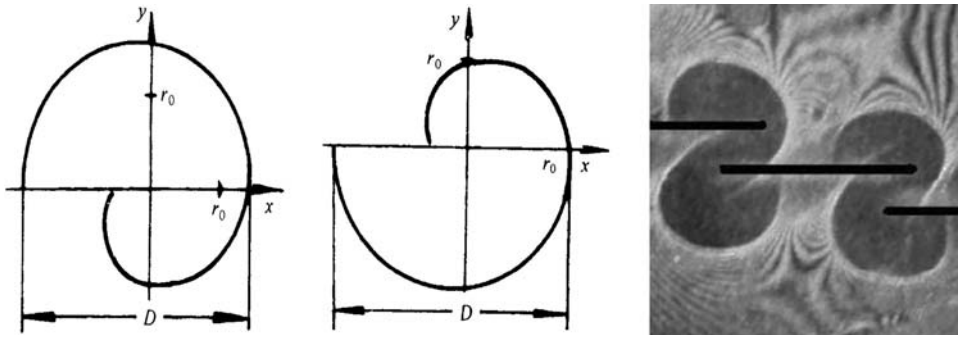


Figure 8
Schematic diagrams of typical caustic curves (a and b) and patterns recorded in this study (c).

transmitting light, positive virtual image of reflecting light, and so on (SU and ZHIHUI, 1987; SU and ZHENG, 1997). Figure 8c gives the typical record of the present experiment.

The empirical formula for calculating SIFs by use of caustic curves is (SU *et al.*, 1987, 1997)

$$K_{\mathrm{II}} = 2(2\pi)^{1/2} D^{5/2} / [3(3.02)^{5/2} |z_0||c| d \lambda_m^{\frac{3}{2}}], \tag{3}$$

where d is the sample thickness; z_0 the distance from the image plane of caustics to object (sample) plane; and D is called the characteristic length of caustic curve. For elastic materials, D is the only variable on the right-hand side of Eq. (3) when the experimental setup is given. In the present experiment, the cracks are inclined. When a single inclined crack is under tension, the SIFs are

$$K_{\mathrm{I}} = \sigma(\pi a)^{1/2} \sin\beta \sin\beta = K_{\mathrm{Is}} \sin^2\beta, \tag{4}$$

$$K_{\mathrm{II}} = \sigma(\pi a)^{1/2} \sin\beta \cos\beta = K_{\mathrm{IIs}} \cos\beta. \tag{5}$$

This means that the calculation result of cracks of pure mode I or II differs from that of the single inclined crack in the tension state only by a constant related to the angle between the crack and loading direction. In relevant literature (YIN, 1985), the result is considered suitable for compression too. In fact, the friction between crack surfaces under compression may increase; this has also been confirmed in the present experiment.

There are a total of six macroscopic cracks in the sample. The SIFs for the tips of the six cracks are determined by Eq. (3). The six cracks gave a total of nine relatively clearer measuring points. Three cracks had two measuring points each, the remaining three had one measuring point each, hence their SIFs are average results. The tips of three cracks were close to the sample boundary, which may cause errors in the results.

The results are as shown in Figure 9. During the entire loading process, the rate of variation of the SIF was not the same. With the increase of pressure, there were mainly three stages of variation of the SIF. In the first stage, the SIFs increased very rapidly; however, the rate and time of increase were not the same because the states of macroscopic cracks were not all the same. In the second stage, the SIFs had an obvious change, and the rate of variation reduced evidently. In the third stage, the SIFs tended to be close or equal to the fracture toughness and the cracks began to propagate. The fracture toughness values obtained from the six cracks are approximately the same, except those from cracks Nos. 1 and 4 which are slightly on the high side. At that time, the SIF at the crack tip had already reached the fracture toughness and hence the external load mainly played the role of supplying the energy for fracture propagation. When the energy accumulated to a certain degree, there would be microcrack to occur so that the energy could be released. For

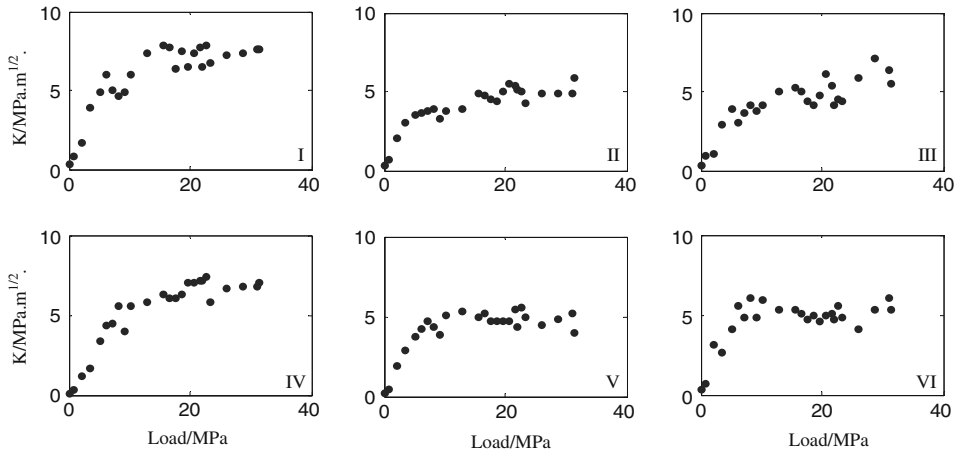


Figure 9
The variation of SIFs of the six cracks with time.

the production of the next fracture, a new turn of energy accumulation is required. Such a process manifests itself on the SIF diagram as a fluctuation around the fracture toughness.

2.6. Characteristics of the Variation of Microcracking Waveforms During the Fracture Preparation Process of the Plexiglass Sample

Like that in the rock sample, most of the microcracking events during the fracture preparation process were high-frequency (wave) events; the variation of their frequency and intensity was also similar to that in the rock sample. However, before an extra large event or a large event, often low-frequency (wave) events would occur, and they were considerably nearer the extra large fracture. Typical records are as shown in Figure 10. In the figure, the extra large fracture was seriously amplitude-limited, but the fore-fracture just before it was a low-frequency (wave) event; the frequency of that event was much lower than that of fractures before.

3. Discussion and Conclusions

3.1. On the Earthquake Recurrence Period

There are two kinds of concepts pertaining to about the earthquake recurrence period: One is the time separation for earthquakes to recur at the same place; the other is the interval for the active-quiet alternation of earthquakes in a larger region. However, the implication obtained from the experiment is that, regardless of which kind the recurrence period is, the most basic condition is that enough energy must be

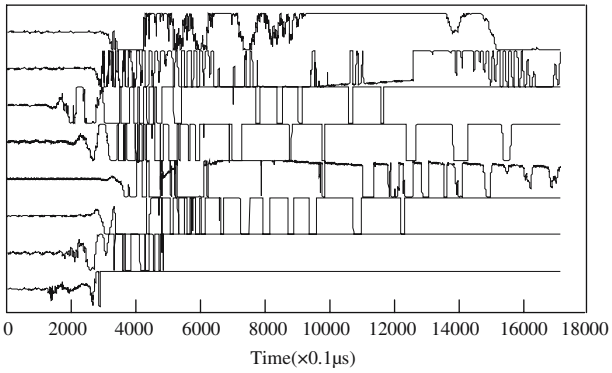


Figure 10
Records showing that low-frequency (wave) events were much nearer the extra large fracture.

accumulated in the relevant region; an earthquake ought to occur once the condition of instability failure has been satisfied. The duration in which the energy stays at a sufficiently high level is the seismic active period; in contrast, the duration in which the energy stays at an level insufficiently high is the seismic quiet period.

Fracture mechanics takes the fracture toughness K_c as a parameter of materials; it represents a critical value the SIF should have when the crack begins to propagate unstably. Correspondingly, the crack growth resistance G_c has also been specified. It represents the energy needed for creating a unit area of new crack surface or a critical value the energy release rate should have when the crack begins to propagate unstably. For (regular) cracks of pure modes I, II and III, these two quantities have a constant relation with each other (YIN, 1985). The energy release rate is

$$G = \partial U / \partial A.$$

It represents the strain energy released by the unit area of new surface created by crack propagation.

$$G = G_{\mathrm{I}} + G_{\mathrm{II}} + G_{\mathrm{III}} = \frac{K_{\mathrm{I}}^2}{E'} + \frac{K_{\mathrm{II}}^2}{E'} + (1+\upsilon)\frac{K_{\mathrm{III}}^2}{E},$$

where

$$E' = E$$

for plane stress, and

$$E' = E/1 - v^2$$

for plane strain.

The criterion for crack propagation is that the crack propagates unstably when $K \geq K_c$ or $G \geq G_c$. For cracks of mixed mode, the theory of strain energy density factor or maximum energy release rate is usually used as the criterion for crack propagation. The strain energy density U and strain energy density factor S are, respectively

$$U = (1/2)\sigma\varepsilon = S/r$$

and

$$S = a_{11}K_I^2 + 2a_{12}K_IK_{II} + a_{33}K_{III}^2$$

The theory of strain energy density assumes that the crack propagates in the direction of the minimum S value, S_{min}; the crack begins to propagate when $S_{min} \geq S_c$. The theory of maximum strain energy release rate assumes that the crack propagates in the direction of the maximum strain energy release rate; the crack begins to propagate when $G \geq G_c$. In fact, the two criteria are consistent with each other.

When the strain energy or SIF in a region is high enough to satisfy the condition of unstable fracture, an earthquake occurs. The duration of this sustaining high value is the seismic active period. After earthquake, the strain energy becomes less and the SIF lowers. The time length needed for the SIF to increase again to a level that satisfies the condition of unstable fracture depends on a number of factors. One is the amplitude of reduction of the strain energy or SIF; another is the rate of stress increase; and the other is the changed value of the fracture toughness or crack growth resistance of materials. This period of time before the condition for unstable fracture has been reached is the seismic quiet period. Because the strain energy, stress rate and fracture toughness are not all the same, the seismically active and quiet periods are also different. Figure 3 shows that, when the strain energy has reached a certain value, the fracture active period comes and extra large fracture occurs. The figure also shows that, in the three periods, the strain energy density factors or critical values of energy release rate for crack propagation are not the same. This means that the material experiences change in property after the occurrence of fracture; here, this change is that the strain energy density factors or critical values of energy release rate for crack propagation have decreased. Similarly, Figure 9 also indicates that SIF is not a constant, so long as the SIF is not smaller than the fracture toughness, fractures on the sample always remain in the active period. In this way, the periodical activity of earthquakes has been explained in terms of fracture mechanics. However according to results of those experiments, even if the sample is homogeneous material, the earthquake recurrence period is not a constant, which is relative to the stress situation, property variation of material and so on.

More than two decades before, many experts began to study the explanation of earthquake activity in terms of fracture mechanics. However, the studies they made

were mostly from the theoretical angle. Certain good methods they proposed could hardly be used; for example, the method of predicting earthquakes based on the fracture energy characteristics was not feasible then. With the development of science and technology, especially that of GPS, GIS and RS, it has enabled the precise measurement of the crustal deformation and calculation of the strain energy. This has also reopened a field of study for the prediction of earthquake activity by use of the strain energy.

3.2. On the Trend of Post-seismic Development

Regarding the trend of post-seismic development, there are two problems. One is whether or not there there would be even larger earthquakes to occur? The other is how large the aftershocks could be? It can be seen from the test results that if a large event is followed by events that contain more low-frequency components of a strong event in the spectrum, there might be an even larger fracture to come; in contrast, if there are more high-frequency components of a strong event, no larger events would occur. Figure 6b shows the low-frequency (wave) events a slightly longer time before the extra large event, while Figure 10 shows the low-frequency (wave) events a very short time before the extra large event. is like Figure 6b, In other words, when the extra large event is followed by low-frequency events (as Figs. 6b and 6d), there would be events with a fairly strong event to come. Furthermore, when there are too many events of low intensity but less or very few of high intensity, there would probably be events of higher intensity (in a relative sense) to come. In the upper part of Figure 5, the magnitude of events after the largest event decayed more uniformly; in contrast, in the lower part, the magnitude of events after the extra large event decayed none-uniformly, with large events being very few and small ones more, hence there were high-magnitude events to reoccur. Figure 4 shows that after the first extra event the AE rate decreased, but with the stress increasing the rate increased again, when it reached the maximum value, the second extra fracture occurred. In addition to the above, judgment must also be made relative to the state of stress and strain. After a large event has occurred, there may be large or even larger events to come when the stress (or strain energy) does not have obvious reduction or it even increases again very soon; in contrast, when the stress (or strain energy) reduces obviously or decreases considerably, there unlikely would be large or even larger events to occur later, as Figure 3 showed, the shear stress and strain still increased after an extra event, another extra fracture would occur again. Of course, the above comparison is related to the properties of materials; for example, when $K \geq K_c$ or $G \geq G_c$, the crack propagates unstably (earthquake occurs); otherwise, the crack does not propagate unstably (no earthquake occurs). Thus in the study of earthquake series, it is important to study the stress situation and the fracture mechanism of event.

The problems of how to determine the crustal stress in the field and how to determine precisely the strain energy need to be studied thoroughly. Though there are still many difficulties, this has provided us with a definite and feasible direction towards which efforts should be made.

Acknowledgement

This project is supported by the NSFC (10232050), MOST(2004CB418406) and Key Projects of 10th Five-Year Plan of Yunnan Province, documented as yunnandiqu [2002] 540202.

References

Cai Jingguan, and Liu Zhengrong (1990), *Forthcoming strong earthquakes calculated upon recent seismic activities*, Earthquake Research in China *3*, 47–52 (in Chinese with English abstract).

Hu Yili, Xu Zhaoyong, Wang Bin, Zhao Jinming, and Mei Shirong, (1997), *The variation of strain field on free plane of the sample containing hard inclusion in process of fracture development*, Earthquake *2*, 162–168 (in Chinese with English abstract).

Mei Shirong, (1960), *The seismic activity of China*, Acta Geophysica Sinica, *1*, 1–19 (in Chinese with Russian abstract).

Mogi, K. (1962), *Study of the elastic shocks caused by the fracture of heterogeneous materials and its relation to earthquake phenomena*, Bull. Earthq. Res. Inst., Univ. Tokyo *40*, 125–173.

Mogi, K. (1963), *Some discussions on aftershocks, foreshocks and earthquakes swarm — The fracture of a semi-infinite body caused by inner stress origin and its relation to the earthquake phenomena* (3) Bull. Earthquake Res. Inst. Univ. Tokyo *41*, 615–658.

Mogi, K., *Fundamental studies on earthquake prediction*, *A Collection of Papers of International Symposium on Continental Seismicity and Earthquake Prediction* (*ISCSEP*) (ed.The Organizing Committee of ISCSEP), (Beijing: Seismological Press,1984), pp. 619–652.

Shi Zhenliang, Huan Wenlin, Lu Shoude, and Yan Jiaquan, *Characteristics of continental seismicity in Central and Eastern Asia (Abstract)*, *A Collection of Papers of International Symposium on Continental Seismicity and Earthquake Prediction* (*ISCSEP*) (ed. The Organizing Committee of ISCSEP) (Beijing: Seismological Press,1984), pp. 155.

Su Xianji, and Lei Zhihui, (1987), *A preliminary application of the experimental method of dynamic caustics in fracture mechanics*, Acta Mechanica Sinica *19*, 357–365 (in Chinese with English abstract).

Su Xianji, and Li Zheng, (1997), Dynamic Measurement Techniques in Solid Mechanics (Beijing: Higher Education Press, 1997) (in Chinese).

Xiong Bingheng, Wang Zhengrong, Lü Xiaoxu, Zhong Liyun, Zhang Yong'an *et al.* (2002), *A noval optical testing system created for experimental seismological research*, Chinese J. Lasers *29*, 376–380 (in Chinese with English abstract).

Xu Shaoxie, (1984), *A review of the seismicity patterns*. In *A Collection of Papers of International Symposium on Continental Seismicity and Earthquake Prediction* (*ISCSEP*) (ed. The Organizing Committee of ISCSEP) (Beijing, Seismological Press, 1984), pp. 11–42.

Xu, Z., Yang, R., Zhao, J. *et al.* (2002a), *Experimental study of the process zone, nucleation zone and plastic area of earthquakes by the shadow optical method of caustics*, Pure Appl. Geophys. *159*, 1951–1966.

Xu Zhaoyong, Mei Shirong, Yang Runhai *et al.* (2001a), *Study on short-impending strain precursors before rock sample failure*, J. Seismol. Res. *24*, 191–196 (in Chinese with English abstract).

Xu Zhaoyong, Shu Yanhua, and Bao Yifeng, (1989), *The rupture characteristics of rock samples with cut crack filled by epoxy resin under uniaxial compression*. In *A Collection of Research Papers of Experimental Geophysics*, (ed. Editorial Committee of Chinese Journal of Geophysics) (Beijing, Seismological Press,1989) pp. 69–274 (in Chinese with English abstract).

Xu Zhaoyong, Yang Runhai, Wang Bin *et al.* (1998) *Burst strain disturbance — A new test study on fracture precursor*, Acta Seismologica Sinica *11*, 737–744.

Xu Zhaoyong, Mei Shirong, Yang Runhai *et al.* (2001b), *Experiment study of the precursor in far and ultra-far field*, J. Seismol. Res. *24*, 332– 338 (in Chinese with English abstract).

Xu Zhaoyong, Yang Runhai, Zhao Jinming *et al.* (2002b), *Experimental study of multiple fractures on a Y-shape boundary of blocks*, Chinese J. Geophy. *45*(Sup.), 214–224 (in Chinese with English abstract).

Yin Xiangchu, *Solid Mechanics* (Beijing: Seismological Press, 1985) (in Chinese).

Zhang Guomin (1987), *Rhythmic characteristics of high seismic activity in China mainland*, Seismol. Geol. *9*, 27–37 (in Chinese with English abstract).

Zhang Guomin, Geng Luming, and Shi Yaolin, (1993), *A computer model for cyclic activities of strong earthquake in continental seismic zones*, Earthq. Res. in China *9*, 20–32 (in Chinese with English abstract).

(Received October 26, 2004, revised/accepted September 26, 2005)

Pure appl. geophys. 163 (2006) 2561–2574
0033–4553/06/122561–14
DOI 10.1007/s00024-006-0145-8

Pure and Applied Geophysics

State Vector: A New Approach to Prediction of the Failure of Brittle Heterogeneous Media and Large Earthquakes

HUAI-ZHONG YU,[1,2] XIANG-CHU YIN,[2] QING-YONG ZHU,[3] and YU-DING YAN[4]

Abstract—The concept of state vector stems from statistical physics, where it is usually used to describe activity patterns of a physical field in its manner of coarsegrain. In this paper, we propose an approach by which the state vector was applied to describe quantitatively the damage evolution of the brittle heterogeneous systems, and some interesting results are presented, i.e., prior to the macro-fracture of rock specimens and occurrence of a strong earthquake, evolutions of the four relevant scalars time series derived from the state vectors changed anomalously. As retrospective studies, some prominent large earthquakes occurred in the Chinese Mainland (e.g., the M 7.4 Haicheng earthquake on February 4, 1975, and the M 7.8 Tangshan earthquake on July 28, 1976, etc) were investigated. Results show considerable promise that the time-dependent state vectors could serve as a kind of precursor to predict earthquakes.

Key words: State vector, acoustic emission, precursor of macro-fracture, earthquake prediction.

1. Introduction

The catastrophe of brittle heterogeneous media involves many natural phenomena of which an earthquake may be the most representative example. The physical essence of an earthquake lies in the nonlinear far-from-equilibrium damage process along with the random distribution of initial defects within the source media, which causes the complexity in earthquake prediction. In addition, the information of seismic activities that can be derived from the crust is still relatively limited, which also greatly barricades the prediction of earthquakes. To date, though few earthquakes have ever been successfully predicted globally, considerably more cases

[1] State Key Laboratory of Earthquake Dynamics, Institute of Geology, China Earthquake Administration, Beijing 100029, China. E-mail: yuhz@lnm.imech.ac.cn
[2] State Key Laboratory of Nonlinear Mechanics, Institute of Mechanics, Chinese Academy of Sciences, Beijing 100080, China. E-mail: xcyin@public.bta.net.cn
[3] School of Mathematics and Computational Science, Zhongshan University, Guangzhou 510275, China. E-mail: mcszqy@mail.sysu.edu.cn
[4] Guangdong Seismological Bureau, China Earthquake Administration, Guangzhou 510070, China.

were not forecasted. Even though there were dense seismic monitoring networks and highly precise equipment in the regions, the robust precursors of earthquakes still failed to be detected.

Since earthquake prediction is so complicated and difficult, new physical clues or approaches, which reveal the formation and impendence of a large earthquake, should be introduced. Critical point hypothesis (SORNETTE and SORNETTE, 1990; SORNETTE and SAMMIS, 1995; BOWMAN *et al.*, 1998; SAMMIS and SMITH, 1999; RUNDLE *et al.*, 1999; JAUME and SYKES, 1999) may be an acceptable earthquake prediction model. According to the hypothesis, the earth's crust does not always stay in a critical state. Occurrence of a large earthquake in a region appears to remove the crust from a critical state. Thereafter as the tectonic stress accumulates, the crust will be driven back towards the critical state again. During the establishment of the criticality, seismic moment release accelerates in the region surrounding the epicenter of the ensuing event. The character may be identified by fitting cumulative energy release prior to a large earthquake to a power law time-to-failure relation. Such a fit provides an intermediate-term prediction for a large earthquake. Similar to CPH, the Load/Unload Response Ratio (LURR) method, which narrates the impendence of a large earthquake by anomalous change of LURR, was proposed by YIN *et al.* (1995; 2000; 2002), whereas, deeper research on the precursors needs to further analyze the damage mechanism of the heterogeneous media in the focal region.

A large earthquake may be considered as a local dynamic rupture process which appears to be a catastrophic transition. A series of recent studies suggest that the process might depend sensitively on the details of heterogeneous source media and stress transfer in the crust (STEIN, 1999). Consequently a possible strategy to identify clues for the prediction of an earthquake is to explore general features of the catastrophic rupture of brittle heterogeneous media. From the viewpoint of damage mechanics, the emergence, accumulation and clustering of micro-cracks will lead to the occurrence of the final failure. The nature of which may quantitatively be described by a special physical approach called state vector, which characterizes the spatial and temporal activity patterns of a physical field (RUNDLE *et al.*, 2002) in the manner of coarsegrain (REICHL, 1980). In recent years, we transplanted the idea of state vector from statistical physics into mechanics to investigate the damage evolution of the brittle heterogeneous systems (YIN *et al.*, 2004a, b).

In this paper, we use both acoustic emission (AE) experiments and seismic data (their catalog retrieved from Center for Analysis and Prediction, China Earthquake Administration) to examine the predictive power of the state vector approach. The experiments were conducted by means of international cooperation. The participants came from State Key Laboratory of Nonlinear Mechanics, Institute of Mechanics, Chinese Academy of Sciences; Center for Analysis and Prediction, Chinese Seismological Bureau; Ioffe Physical Technique Institute, Russian Academy of Sciences; Yunnan Province Seismological Bureau, etc.

2. Definition of State Vector

Figure 1 shows a continuum, which was divided into n sub-regions. The physical quantity $V_i(t_k)$ (e.g., the sum of Benioff strain) representing the damages in sub-region i within a time window T can be regarded as i-th component of an n-dimensional vector $\mathbf{V}_k$. This n-dimensional vector is called state vector, which slides with a time interval of Δt. Different state vectors at different times t_k ($k = 1, 2, \ldots$) can form a trajectory in the phase space. If a series of state vectors at different time steps has been acquired, the temporal and spatial evolution of the physical field may be obtained. To quantitatively measure the evolution of the physical field with state vector, four relevant scalars are defined as follows:

1) The modulus of state vector $\mathbf{V}_k$ (represented by M), meaning the entire level of the physical field at any time t_k,

$$\mathrm{M} = |\mathbf{V}_k|. \tag{1}$$

2) The angle between state vectors $\mathbf{V}_k$ and $\mathbf{V}_{k+1}$ (represented by φ), suggesting the change of the physical field between two consecutive times t_k and t_{k+1},

$$\varphi = \arccos\left(\frac{\mathbf{V}_{k+1} \cdot \mathbf{V}_k}{\mathrm{V}_{k+1}\mathrm{V}_k}\right). \tag{2}$$

3) The modulus of state vector increment $\mathbf{V}_{k+1} - \mathbf{V}_k$ (represented by ΔM), reflecting the change of the physical field between two consecutive times from another viewpoint,

$$\Delta\mathrm{M} = |\mathbf{V}_{k+1} - \mathbf{V}_k|. \tag{3}$$

1	2	3	⋯				⋯	n_1
		⋯	i-1	i	i+1	⋯		
					⋯	n-2	n-1	n

Figure 1
Division of a continuum into n sub-regions.

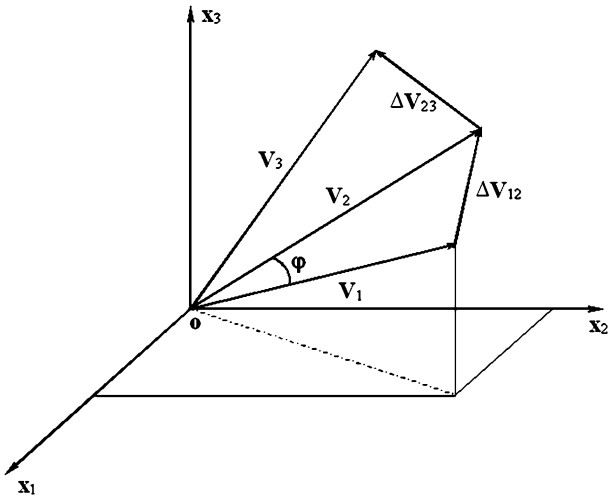


Figure 2
The spatial and temporal patterns of state vectors in three-dimensional phase-space. $\mathbf{V}_1$, $\mathbf{V}_2$, $\mathbf{V}_3$ denotes three state vectors at three consecutive times t_1, t_2, t_3. φ and $\Delta\mathbf{V}_{12}$ represent respectively the angle and the increment between two consecutive state vectors $\mathbf{V}_1$ and $\mathbf{V}_2$.

The angle between state vector $\mathbf{V}_k$ and equalized vector $\mathbf{V}_e$ (represented by φ_c), characterizing the deviation of the physical field from the average field at each time steps,

$$\varphi_c = \arccos\left(\frac{\mathbf{V}_e \cdot \mathbf{V}_k}{\mathrm{V}_e \mathrm{V}_k}\right), \tag{4}$$

where, the equalized vector V_e consists of equal components.

Figure 2 displays the patterns of state vectors simplified into 3-D phase-space (In practice, more sub-regions are usually provided). It is obvious that with increased damage within the media the value of M will increase. On the other hand, the value of φ, φ_c, and ΔM will abruptly increase when the damages within the media begin to cluster.

3. Experiments Research on State Vector

Four tri-axial compression acoustic emission (AE) experiments involving two kinds of rocks (Wuding gneiss and Wuding sandstone) were used to research the evolution of the state vectors. Wuding gneiss has a Young's modulus of E = 27 GPa, Poisson ratio $v = 0.26$, density $\rho = 2.6 \times 10^3$ kg/m^3 and compressional wave speed c = 3200 m/s. Wuding sandstone has a Young's modulus of E = 11 GPa, Poisson ratio $v = 0.26$, density $\rho = 2.2 \times 10^3$ kg/m^3 and compressional wave speed

c = 2150 m/s. The geometry of the specimens is $105 \times 40 \times 10$ cm^3 for gneiss and $105 \times 40 \times 15$ cm^3 for sandstone. In the experiments, the rock specimens of rectangular prisms were subjected to both lateral and axial loads simultaneously (Fig. 3). Lateral load was symmetrically applied to a specimen with steel plates on the two smaller sides and was retained at a constant pressure of 60 T. Axial load was applied to the specimen with the rams controlled by a hydraulic valve (the maximum of axial load is 500 T) until the final fracture took place.

AE signals were recorded continuously with an A-line 32D-AE system made by Ioffe. (Physical Technical Institute, Russian Academy of Sciences and Interunis Ltd.). Each channel consists of an AE sensor, a pre-amplifier and an AECB (Acoustic Emission Channel Board). Ten piezoelectric AE sensors were attached directly to the specimens (small circles represent the AE sensors in Fig. 3). They were used to monitor the sound signal generation from the specimens and convert them into electrical signals that were amplified by a preamplifier and converted into a digital data stream in an AECB (LOCKNER and BYERLEE, 1977). AE features such as arrival times, rise-times, duration, peak amplitude, energy and counts were extracted by a FPGA (Field Programmable Gate Array). The frequency of input filter ranged from 10 kHz to 500 kHz [−3 dB] and 24 dB/octave. The ADC (Analogue to Digital Converter) was operated at a 5 MHz sampling frequency and provided the 16-bit resolution. The threshold values ranged from 45 to 51 dB and the resolution was 1 μV. The working frequency of the sensor ranged from 100 to 500 kHz and its

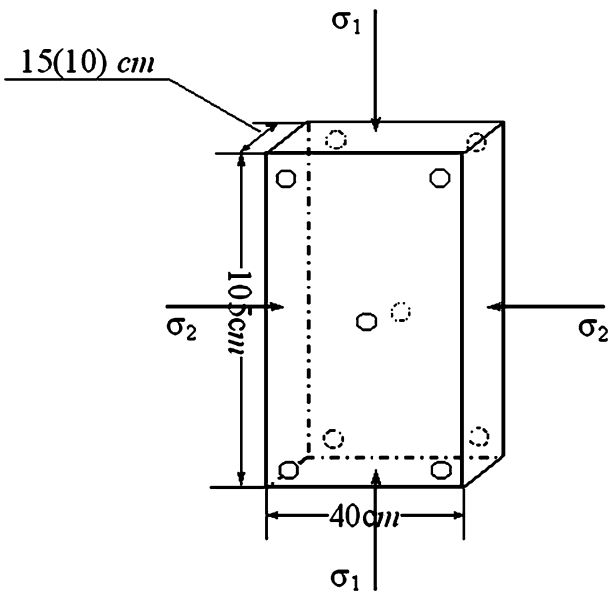


Figure 3
The geometry of rock specimens, the loading conditions, and the arrangement of AE sensors (circles).

sensitive factor on 175 kHz was 67 dB relative to 1 B/m/sec. AE events were 3-D-positioned (LOCKNER *et al.*, 1991) by arrival time of AE signals in real time and could be displayed on a screen and recorded for post-processing (YIN *et al.*, 2004a; YU *et al.*, 2005).

For the experiments, the state vector is computed using the sum of AE Benioff strain recorded within each of $n = 4 \times 6 = 24$ sub-regions divided from the larger side of the specimens. Figure 4a is a plot of the two kinds of axial loads (monotonous and cycling) and the variations of scalars M, φ with time for two gneiss specimens (G1 and G2), where the unit J meaning joule. The moving time window T and sliding

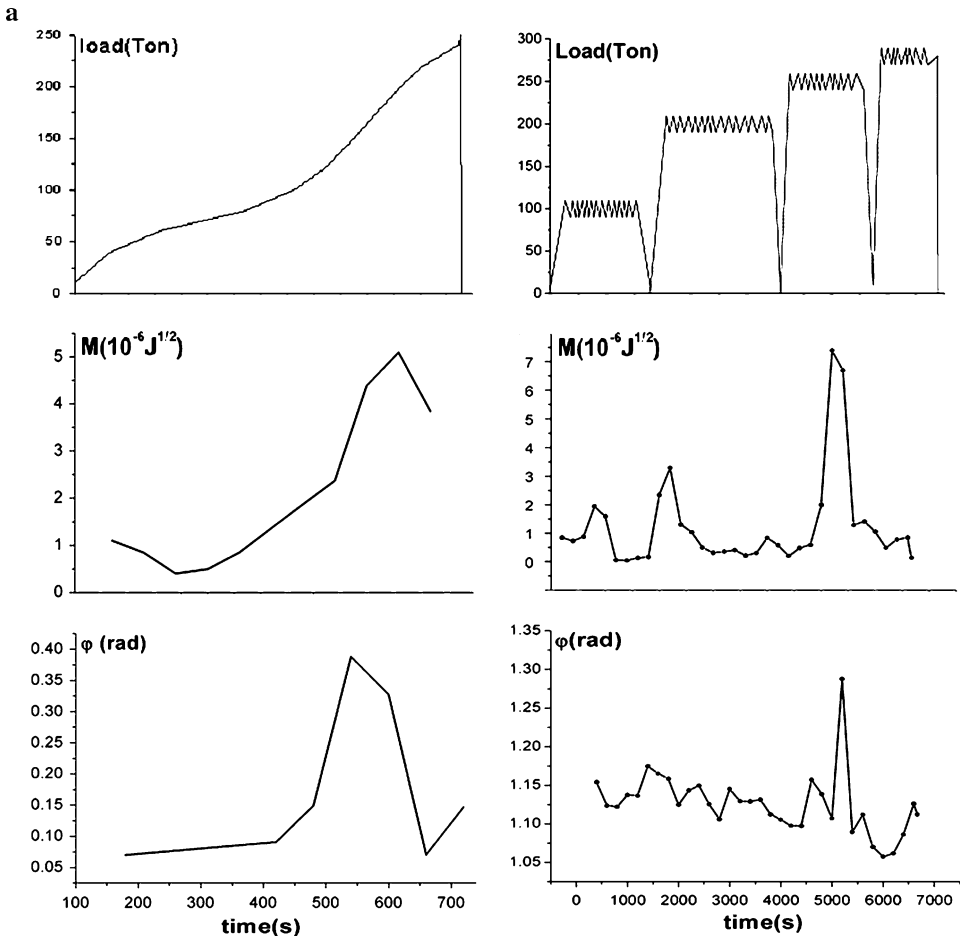


Figure 4
(a) Axial load histories and time series of scalars M, φ for two gneiss specimens. Left: G1 and right: G2. (b) Axial load histories and time series of scalars M, φ for two sandstone specimens. Left: S1 and right: S2.

step Δt for G1 was 120 s and 60 s, along with $T = 400$ s, $\Delta t = 200$ s for G2. Two sandstone specimens were subjected to the similar load process. Both the axial loads and the evolutions of scalars M, φ time series ($T = 150$ s, $\Delta t = 100$ s for S1 and $T = 400$ s, $\Delta t = 200$ s for S2) are shown in Figure 4b. During the first stage of the experiments, there was no coalescence of microcracks within the specimens so scalars φ and M simply fluctuate at a low level. When the microcracks began to coalesce into larger fractures, the damage state of the different sub-regions changed severely as evidenced by sharp peaks in both traces. Following the peak evolution, the macro-fracture took place.

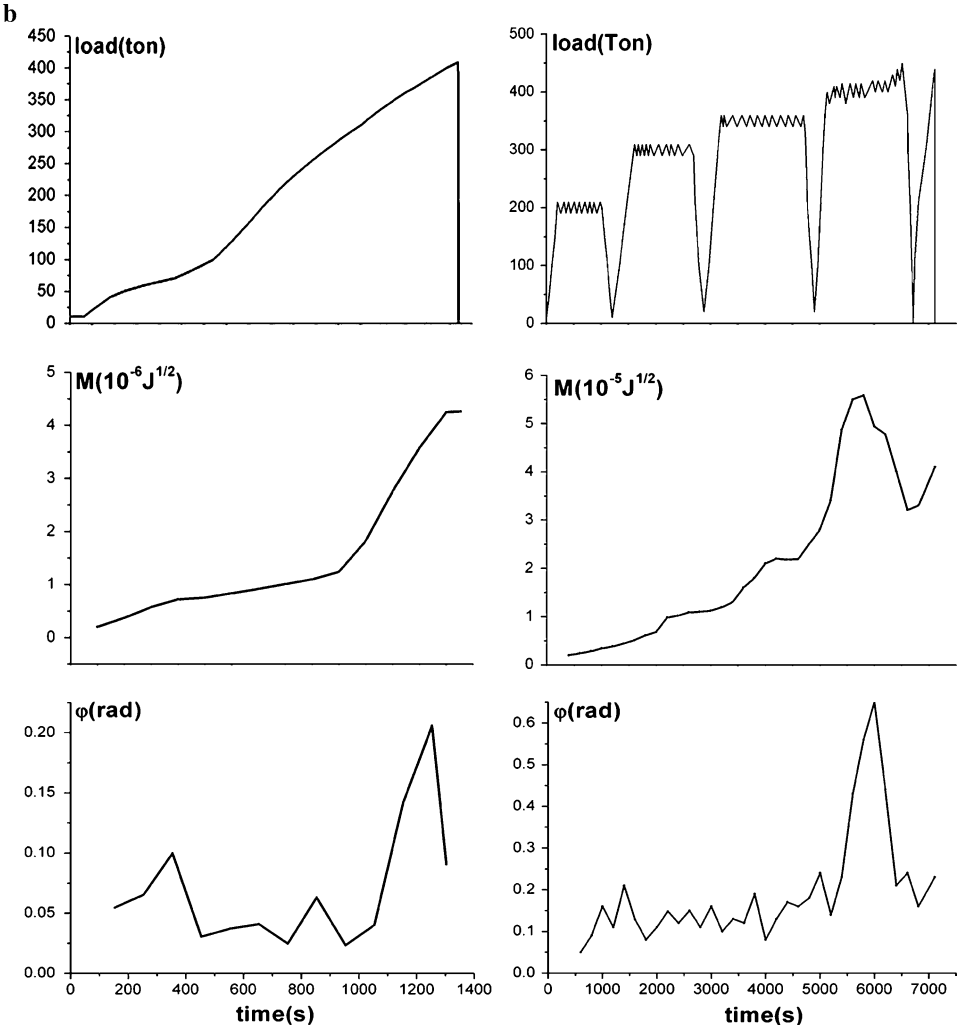


Figure 4
(contd.)

4. Application to Seismic Data

As a retrospective study, we apply the state vector to 'predict' some well known earthquakes which occurred in Chinese Mainland. Figs. 5 and 6 display the evolutions of the four relevant scalars time series before the 1975 Haicheng earthquake and the 1976 Tangshan earthquake, respectively. The spatial windows centered at the epicenter of these two earthquakes (40.7° N; 122.7° E and 39.63° N; 118.18° E) were taken to be $6° \times 6°$ and $8° \times 8°$, and were then divided by the optimal area of $1° \times 1°$ into 36 and 64 sub-regions. The sums of logarithm of earthquake energy of each earthquake in each sub-region within a period of one year were used to calculate traces of the four scalars by moving them with a time interval of 20 days. Results show that traces of M, ΔM, φ, and φ_c changed anomalously about one month prior to the Haicheng earthquake and three months prior to the Tangshan earthquake. From these two earthquake cases we suggest that the sharp changes of state vectors could be used as medium- or short-term precursors to predict large earthquakes.

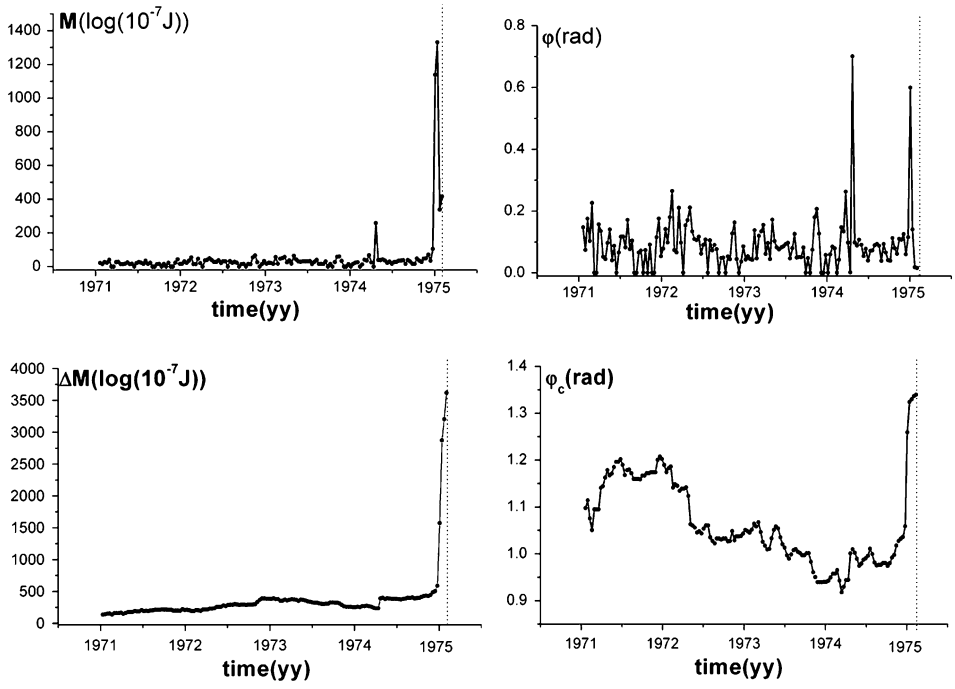


Figure 5
Time series of the four state vector scalars before the M 7.4 Haicheng earthquake. Dashline indicates occurrence of the earthquake.

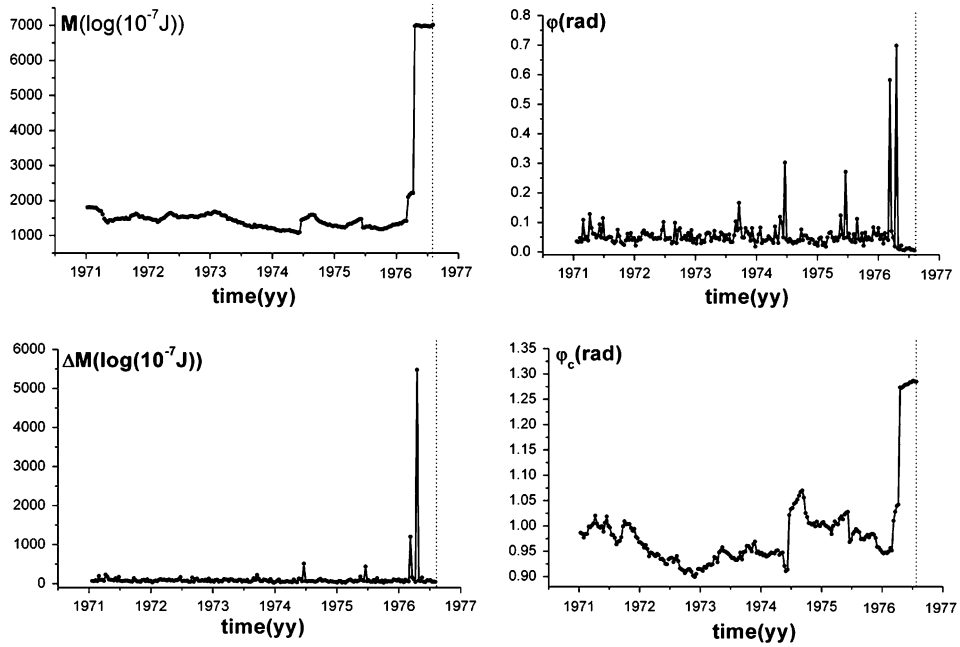


Figure 6
Time series of the four state vector scalars before the M 7.8 Tangshan earthquake. Dashline denotes time of the earthquake.

Figure 7 shows the four relevant scalars evolution time series before two M 6.1 Datong-Yanggao earthquakes. The spatial window centered at the epicenter of the event was $6° \times 6°$ with a sub-region area of $1° \times 1°$; and the moving time window was one year at a sliding step of 20 days. It is quite interesting that though the traces did change prior to the first (October 19, 1989) and the second (March 26, 1991) Datong-Yanggao earthquake, the changes were not as noticeable as that of the Haicheng and Tangshan earthquakes. Furthermore, the variation patterns of state vectors before many moderate ($M \sim 6$) earthquakes occurred in the Chinese Mainland, such as the 1976 Helingeer earthquake, and the 1981 Xingtai earthquake, etc. were also studied, but few of these earthquakes could be predicted by the state vector approach.

5. Predictive Applicability of State Vector

The fact that the change of state vectors correlates better with the occurrence of larger ($M \sim 7$) earthquakes may suggest the predictive power of the approach. To further determine the predictive applicability of the state vector method in earthquake prediction practices, the location precision of earthquakes selected for state vector evaluation should be taken into consideration. Figure 8 shows a selected

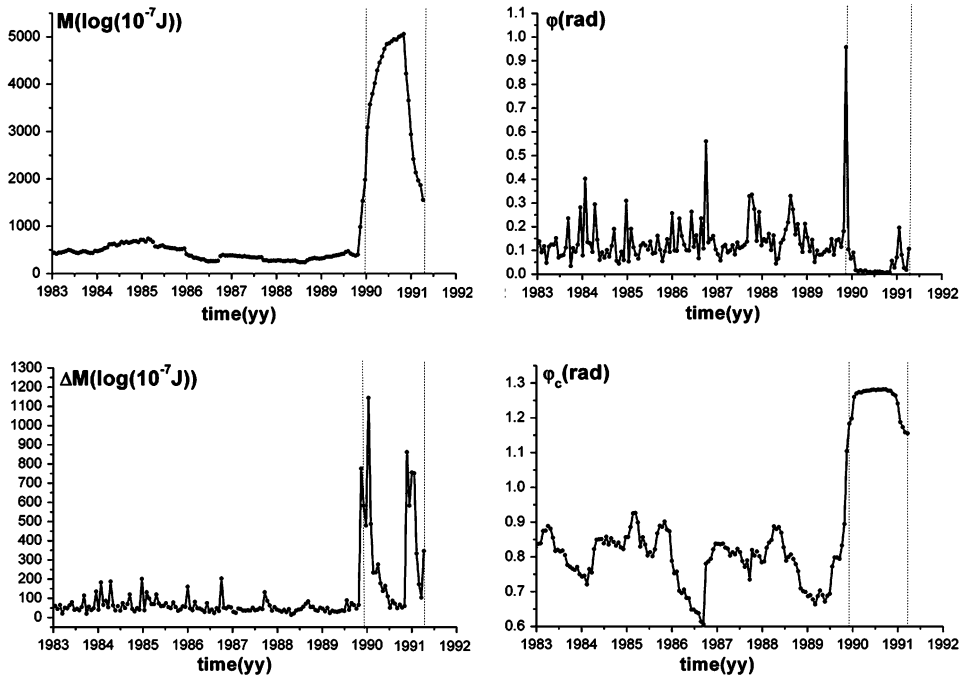


Figure 7
Time series of the four state vector scalars before two M 6.1 Datong-Yanggao earthquakes. Two dashlines represent occurrence of these two earthquakes.

sub-region whose area is determined by the size of length a, and the shadow area decided by size c is the precision-dependent area, where the location error impacts the results of the state vector; while in the blank area, the location error will have no effects on the results of the state vector. The higher the location precision of the events is, the smaller the shadow area is, and the less influence will affect the results of the state vector.

If we let

$$\frac{S_{\text{shadow}}}{S_{\text{total}}} < v, \tag{5}$$

where S_{shadow} and S_{total} represent respectively the area of the shadow region and the entire sub-region, and the ratio v is inverse-correlation with the reliability of the prediction, we can derive

$$a > \frac{2c(1+\sqrt{1-v})}{v}. \tag{6}$$

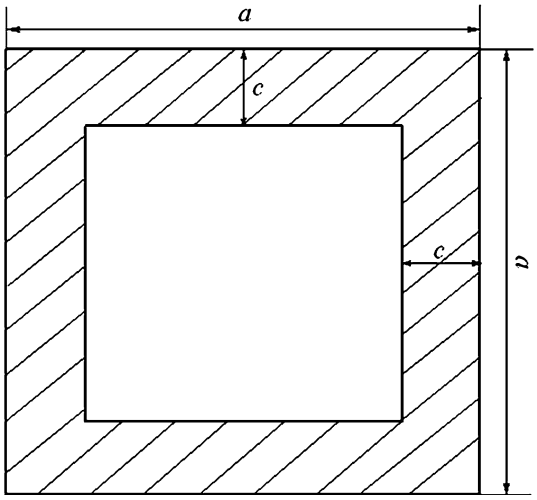


Figure 8
The extent of the location error influence on the state vector. In the shadow area the location error has impacts on the results of the state vector. In the blank area the location error has no effects on the results of the state vector.

On the other hand, the rectangular spatial window selected to compute state vectors is decided by the size of the critical seismogenic region of an earthquake. Thus, the number of sub-regions in a row (e.g., the n_1 indicated in Fig. 1) should meet

$$n_1 < \frac{2R}{a}, \tag{7}$$

where R is the radius of the critical seismogenic region, and can be defined as

$$\log(R) = 0.087 + 0.34\mathrm{M}, \tag{8}$$

where M denotes the magnitude of the earthquake (YIN *et al.*, 2002; 2004a). In earthquake prediction practice, the equation can also be adopted to evaluate the magnitude of an earthquake.

From equations (6) and (7),

$$R > \frac{n_1 c(1+\sqrt{1-v})}{v}. \tag{9}$$

Then combining equation (8) with equation (9), we have

$$\mathrm{M} > \frac{\log(\frac{n_1 c(1+\sqrt{1-v})}{v}) - 0.087}{0.34}. \tag{10}$$

Equation (10) dictates the relationship between the magnitude of the earthquake that can be predicted by state vector and three parameters of n_1, c and v. Suppose that the earthquake catalog has more precise locations of earthquakes (smaller c), the smaller earthquakes may be predicted (smaller M). However, for a definite precision of earthquake locations, only the larger earthquakes (greater M) can be predicted with higher predictive reliability (smaller v), and also the larger earthquakes (greater M) can be subdivided into more regions (greater n_1) to monitor the heterogeneous damage evolution of the source media.

For the Chinese earthquake catalog $c > 5$ km, if we let $n_1 \geq 3$, and $v < 0.1$ (the earthquakes could be well evaluated at this level), we can obtain

$$\mathrm{M} > 7.0. \tag{11}$$

The result indicates that with the available Chinese earthquake data, only those earthquakes with magnitude larger than 7.0 could be well predicted by the state vector approach.

6. Earthquake Prediction Practice

Figure 9 displays the evolutions of scalars M and ΔM time series in the Yunnan region, China (8° × 8° centered at 24° N; 102° E, optimal sub-region: 1.6° × 1.6°, moving time window: 1 year, and sliding step: 20 days). The anomalous change in evolutions M and ΔM occurred obviously during the period of 2001 to 2004. The results may suggest that a large earthquake with magnitude greater than 7.0 will possibly occur in this region in 1–2 years. According to equations (5) and (10), and

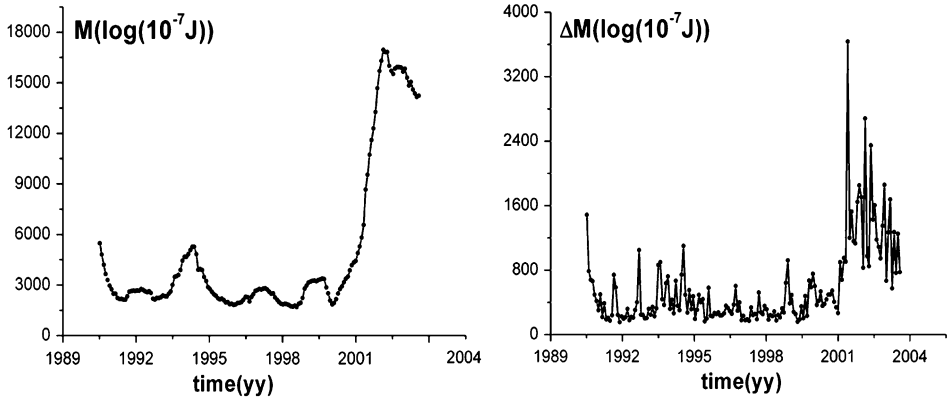


Figure 9
Time series of scalars M and ΔM in the Yunnan district, China.

the earthquake location precision in this region ($c \approx 10$ km), the reliability ($1-v$) of this prediction is slightly higher than 0.76, a level with about 50 % probability to properly predict an earthquake from our statistical evaluation (YIN *et al.*, 2004b).

7. Conclusion

A new approach to describe damage evolution of brittle heterogeneous media, designated as the state vector, was introduced. Using AE data recorded during rock experiments, it was demonstrated that the anomalous change of state vectors show sharp peaks during the period in which microcracks coalesce into larger fractures, prior to the final failure. Knowing this special character of the state vector, we may use that for testing various kinds of earthquakes. Presently, though the approach is still in its infancy, the attempts in this paper have opened new possibilities to systematically evaluate the hazard potential of future earthquakes. It is doubtless that further progress in the approach will also require more rigorous hypothesis testing and data analyzing (KAGAN and JACKSON, 2001).

Acknowledgments

The research was supported by the Joint Earthquake Science Foundation of China (Grant No. 606018), China Postdoctoral Science Foundation (Grant No. 2005037384), the National Natural Science Foundation of China (Grant Nos. 40334042, 10232050, 10572154, and 10572140), and special foundation of the Ministry of Science and Technology of China (Grant Nos. 2004CB418403, 2002CB412706 and 2004CB418406).

References

BOWMAN, D.D., OUILLON, G., SAMMIS, C.G., SORNETTE, A., and SORNETTE, D. (1998), *An observational test of critical earthquake concept*, J. Geophys. Res. *103*, 24359–24372.

JAUME, S.C. and SYKES, L.R. (1999), *Evolution toward a critical point: A review of accelerating Seismic moment/energy release prior to large great earthquakes*, Pure Appl. Geophys. *155*, 279–306.

KAGAN, Y.Y. and JACKSON, D.D. (2001), *Probabilistic forecasting of earthquakes*, Geophys. J. Int. *143*, 438–453.

LOCKNER, D. and BYERLEE, J.D. (1977), *Hydrofracture in weber sandstone at high confining pressure and differential stress,* J. Geophys. Res. *82*, 2018–2026.

LOCKNER, D.A., BYERLEE, J.D., and KUKSENKO, V. (1991), *Quasi-static fault growth and shear fracture energy in granite*, Nature *350*, 39–42.

REICHL, L.E., *A Modern Course in Statistical Physics* (University of Texas 1980).

RUNDLE, J.B., KLEIN, W., and GROSS, S. (1999), *A physical basis for statical patterns in complex earthquake populations: models, predictions and tests*, Pure Appl. Geophys. *155*, 575–607.

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. Natl. Acad. Sci. U. S. A. *99*, 2514–2521.

SAMMIS, C.G. and SMITH, S.W. (1999), *Seismic cycles and the evolution of stress correlation in cellular automation models of finite fault network*s, Pure Appl. Geophys. *155*, 307–334.

SORNETTE, A. and SORNETTE, D. (1990), *Earthquake rupture as a critical point: consequences for telluric precursors*, Tectonophysics *179*, 327–334.

SORNETTE, D. and SAMMIS, C.G. (1995), *Complex critical exponents from renormalization group theory of earthquake prediction*, J. Phys. I. France *5*, 607–619.

STEIN, R.S. (1999), *The role of stress transfer in earthquake occurrence*, Nature *402*, 605–609.

YIN, X.C., CHEN, X.Z., SONG, Z.P., and YIN, C. (1995), *A new approach to earthquake prediction–The Load/Unload Response Ratio (LURR) Theory*, Pure Appl. Geophys. *145*, 701–715.

YIN, X.C., WANG, Y.C., PENG, K.Y., BAI, Y.L., WANG, H., and YIN, X.F. (2000), *Development of a new approach to earthquake prediction: Load/Unload Response Ratio (LURR) Theory*, Pure Appl. Geophys. *157*, 2365–2383.

YIN X.C., MORA, P., PENG, K.Y., WANG, Y.C., and WEATHERLY, D. (2002), *Load-unload response ratio and accelerating moment/energy release, critical region scaling and earthquake prediction*, Pure Appl. Geophys. *159*, 2511–2524.

YIN X.C., YU, H.Z., KUKSHENKO, V., XU, Z.Y., WU, Z.S., LI, M., PENG, K.Y., ELIZAROV, S., and LI, Q. (2004a), *Load-Unload Response Ratio (LURR), Accelerating Moment/Energy Release (AM/ER) and state vector evolution as precursors to failure of rock specimens*, Pure Appl. Geophys. *161*, 2405–2416.

YIN, X.C., YU, H.Z., and ZHANG, Y.X. (2004b), *State Vector—a New method to analysis seismicity*, Earthquake Research in China *18*, 263–270.

YU, H.Z., ZHU, Q.Y., YIN, X.C., and WANG, Y.C. (2005), *Moment Tensor Analysis of the acoustic emission source in Rock damage process*, Progress in Natural Science *15*, 609–613.

(Received October 13, 2004, revised September 13, 2005, accepted September 15, 2005)
Published Online First: December 20, 2006

Pure appl. geophys. 163 (2006) 2575–2583
0033–4553/06/122575–9
DOI 10.1007/s00024-006-0144-9

Pure and Applied Geophysics

Crustal Movement Observed by GPS and Earthquake Activity in the Chinese Mainland and its Neighborhood

GUOHUA GU

Abstract—The general characteristics of the recent horizontal crustal movement in the Chinese Mainland and the horizontal crustal movements before the Kunlun Mountain earthquake of M8.1 on Nov. 14, 2001 and two earthquakes of M 8.0 and M 7.9 occurred in Sept. 2003 around the Chinese Mainland are analyzed with GPS data obtained in the Crustal Movement Observation Network of China (CMONOC). Relative vertical displacements observed in the last five years at fiducial stations show that there was a significant correlation between the length of day and the vertical displacements at most stations in western China.

Key words: Crustal movement, earthquake, GPS, displacement.

1. Introduction

In the Crustal Movement Observation Network of China (CMONOC) there are twenty-five fiducial stations (Fig. 1) of continuous GPS observation, 56 basic stations for repeated observations at regular intervals of 1 to 2 years and 1000 regional stations for repeated observations according to earthquake activity. The major aim of CMONOC is to serve earthquake prediction.

In this study the results of horizontal displacement velocities for the regional and basic stations and weekly coordinates of fiducial stations in ITRF 2000 obtained at the Data Center of CMONOC are further processed and analyzed. The data of fiducial stations were reprocessed in early 2003 due to problems in data collection and previous processing from the raw data.

2. Horizontal Crustal Movements and Deformations of Blocks in the Chinese Mainland

The first GPS observation campaign of 25 fiducial stations, 56 basic stations, 1000 regional stations and a small number of other stations was carried out from March 10 to October 9, 1999 and the second campaign was carried out from March 8

Institute of Earthquake Science, China Earthquake Administration, Beijing, 100036, China.

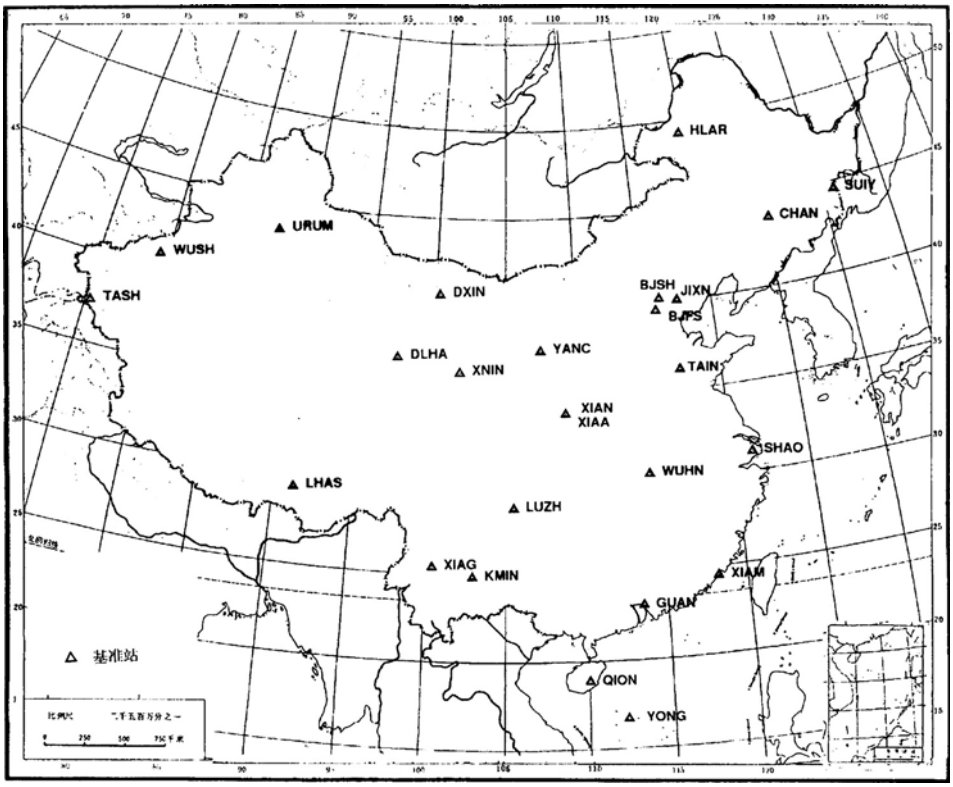


Figure 1
Map of fiducial stations in CMONOC.

to September 5, 2001. Shortly after the completion of the campaign in 2001, the great Kunlun Mountain earthquake of M 8.1 on November 14, 2001, occurred which was the largest earthquake on the Chinese mainland in the last fifty years. 928 stations are used in this study (GU *et al*., 2004). On the block of the Tancheng-Lujiang fault zone with 70 stations, the RMS in displacement velocities is only about 1.8 mm/a. It shows the precision of the displacement result.

Reference frame for the displacement solution formed by a group of stations is referred to as datum definition in this paper. For the study of regional crustal movement a regional datum definition is preferred to show clearly and directly the internal relative movement in the region. The change of the configuration of stations selected in datum definition should be as small as possible and the stations scatter in an area as large as possible to control possible rotations in the horizontal displacement solutions. A datum definition of a group of stable stations with small mutual displacements is used in this paper to gain a more favorable solution of displacement velocity than other solutions (WANG *et al*., 2003) a in the Chinese

Mainland (GU *et al.*, 2000, 2001). 163 stations are selected in datum definition for the displacement velocity solution, most of them are in the area east to E 100 and 57 stations of 163 stations are in northern China where there are 161 stations. Figure 2 is the map of horizontal displacement velocity vectors in Chinese Mainland. In Figure 2 the displacement velocities on blocks in the eastern part of China were quite small.

Though block motions in the Chinese Mainland were studied in other studies based on different displacement solutions and models, a different result can be obtained based on the displacement result in this paper. At the level of the precision of present GPS observations, it seems to be proper to identify the blocks in north, south and northeast China as parts of a larger block, the north and east China block. However there was a significant difference between the displacement velocities in the northern part of the Xinjiang block and those in the southern part of the Xinjiang block, so the block is divided into the northern Xinjiang block and the southern Xinjiang block. Equivalently the Sichuan-Yunnan block is divided into

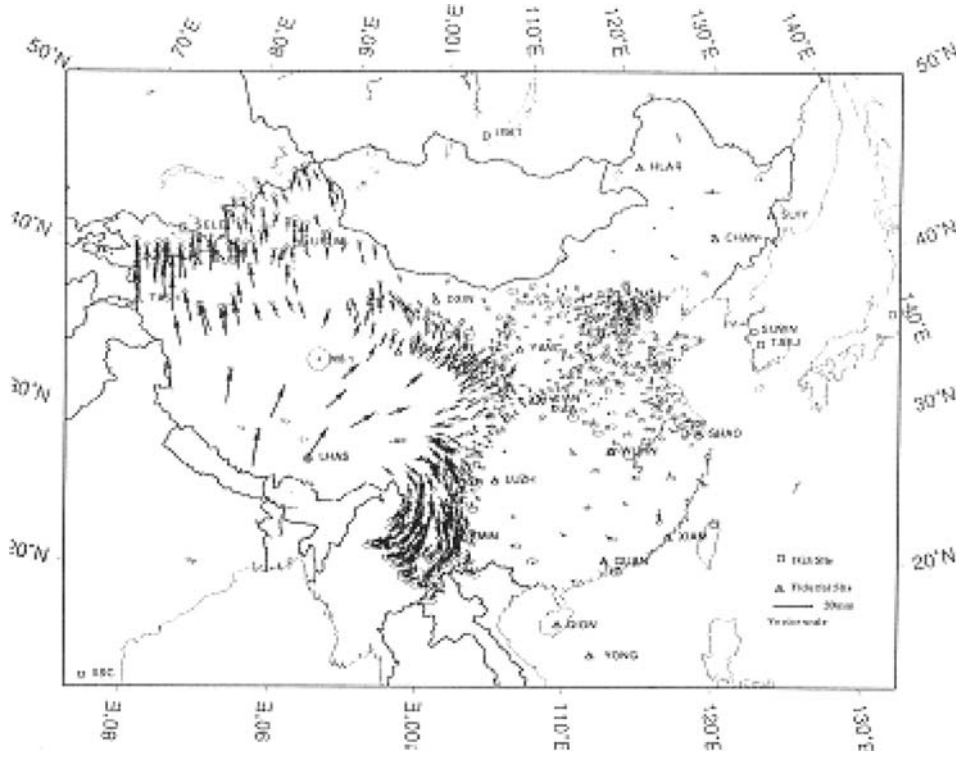


Figure 2
Map of horizontal displacement velocity vectors in the regional networks with datum definition of a group of stable stations with small mutual displacements in eastern China (from 1999 to 2001).

the northern Sichuan-Yunnan block and the southern Sichuan-Yunnan block, and the southwestern part of the Altyn-Qilian Mountains-Alxa block is considered as an individual block. Some stations in the eastern part of Sichuan and Yunnan are assigned to the north and east China block.

Analysis on the displacement velocities of the blocks indicates that the relative left lateral displacement velocity of 10.3 mm/a in the east component between the southern Xinjiang block and the Qinghai-Tibet block was significantly larger than the tensile displacement velocity of 3 mm/a in the north component (GU *et al.*, 2004). Slips of major or great earthquakes in this area were mainly left lateral strike slips in the eastern direction. The largest relative displacement velocity in the north component between two neighboring blocks is that between the Sichuan-Yunnan block and the Qinghai-Tibet block with a right lateral displacement velocity of 20.3 mm/a. Actually, the southern part of the Xinjiang block and the northern part of the Qinghai-Tibet block have formed a shear deformation zone extending from east to west, while the eastern part of the Qinghai-Tibet block and the Sichuan-Yunnan block have formed a huge area of rotational tectonic movement. On the Sichuan-Yunnan block and the eastern part of the Qinghai-Tibet block the rotational tectonic movement has been huge with a clockwise Euler rotation velocity of $0°.49 \pm 0°.0432$ /(m.a) and the rotation pole at $N27°.35 \pm 0°.33$,$E92°50 \pm 1°.02$.

The dispersion of displacements at all stations on a block could be used as a numerical measure of the inhomogeneous deformation. In this paper RMS (rooted-mean-squares) of displacements is adopted as such a measure (GU *et al.*, 2004). The deformations of blocks (except the northern Xinjiang block) in western part of China were large. The deformation of the Qinghai-Tibet block was the largest with a displacement velocity dispersion of 8.7 mm/a. The great Kunlun Mountain earthquake of M 8.1 on November 14, 2001 occurred within this block with deformation in the east component larger than that in the north component.

3. Horizontal Displacement Time Series and Horizontal Crustal Movement before and after some Great Earthquakes

A group of stable stations with small mutual horizontal displacements are also used in datum definition for the solution of time series of horizontal displacements at fiducial stations (GU *et al.*, 2002). The weekly combined solutions for fiducial stations from early September, 1998 to mid-May, 2004 (GPS week 1270) are used. Time series of horizontal displacements at fiducial stations selected in datum definition show quite small changes during this time period. Figure 3 shows time series of horizontal displacement components at fiducial stations. Anomalous horizontal crustal movements were detected before several great earthquakes of M 8.0 occurred in the Chinese Mainland or its neighborhood at some stations in China.

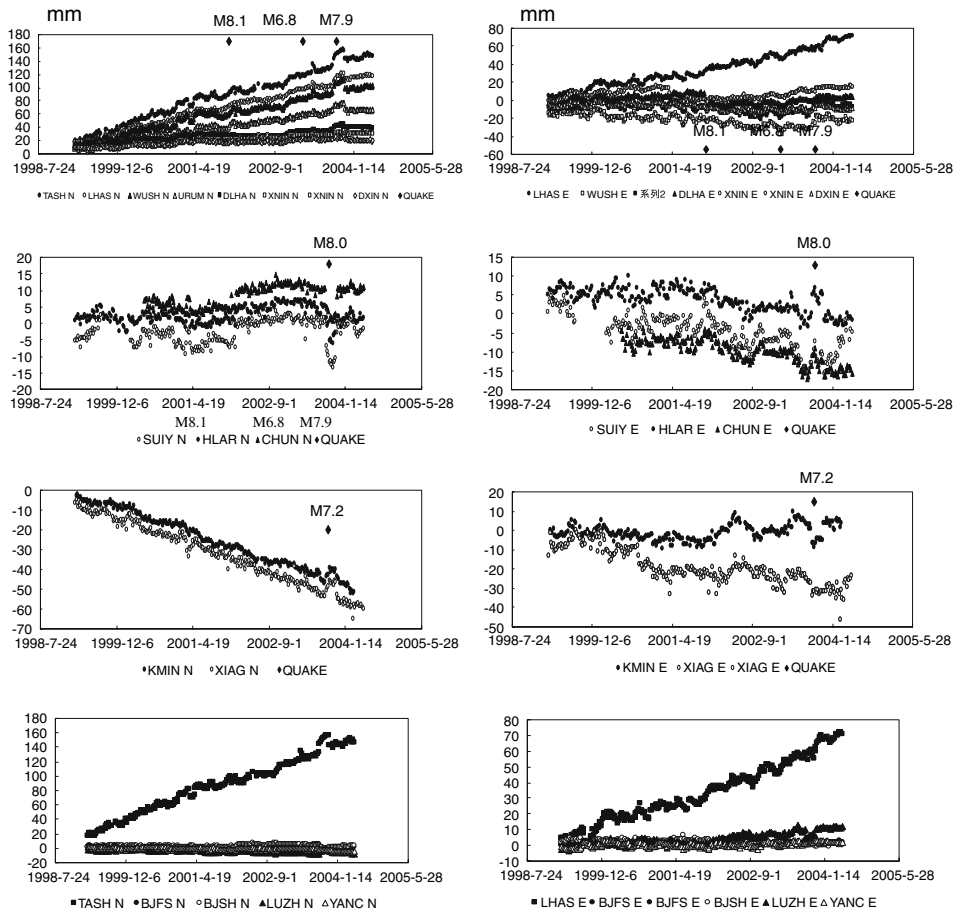


Figure 3
Time series of horizontal displacement components at fiducial stations. (from early September, 1998 to early May, 2004).

The great Kunlun Mountain earthquake of M 8.1 occurred in GPS week 1140 on Nov. 14, 2001 in western China. Analysis on the time series of displacements at the fiducial stations in western China shows that the north displacement components at most fiducial stations in west China had exhibited changes quite different from the previous period simultaneously since mid-April, 2001. These changes were detected in September and October, 2001. The time series of displacements, particularly the north components of displacements, at the fiducial stations in the western part of the Chinese Mainland demonstrated clearly significant movements in an opposite sense or decreased velocities. Among these stations, the changes at DLHA (590 km from the epicenter) and TASH (1400 km from the epicenter) were typical. The coseismic

displacement about 10 mm was detected only for the east component at DLHA, but no coseismic displacements were detected at other fiducial stations.

The anomalous horizontal displacements of N component at TASH, WUSH, DLHA and other stations before the Bachu earthquake of M 6.8 on Feb. 24, 2003, were simple in pattern with the maximum change of 20 mm. The earthquake was about 150 km away from TASH or WUSH. The anomalies lasted for about four months. An earthquake of M 6.6 occurred on March 17, 2003, 52 km from DLHA. However it is hard to distinguish the anomalies before the two events in Figure 4.

An earthquake of M 7.2 occurred on Sept. 22, 2003 in Myanmar, 826 km from XIAG. There were anomalous displacements in N component at least at XIAG and KMIN with a maximum change of about 5 mm and with a duration of only about two weeks.

The Hokkaido earthquake of M 8.0 occurred on Sept. 26, 2003 (GPS week 1237) in Japan, 1096 km from SUIY. There were rather clear displacement anomalies in E component at SUIY, CHUN and even at HLAR before the event. They all manifested westward displacements for about five months with a maximum change of 10 mm at SUIY. The earthquake occurred when they displayed eastward displacements. It should be noted that only the relative displacements in time at CHUN are meaningful. They do not show actual displacements in the datum definition because the coordinates of this station used in the reference epoch are arbitrary.

An earthquake of M 7.9 occurred on Sept. 27, 2003 (GPS week 1237) in Russia, near the border with Mongolia and China, 677 km from URUM. Anomalous displacements in N and E components appeared simultaneously at TASH, URUM,

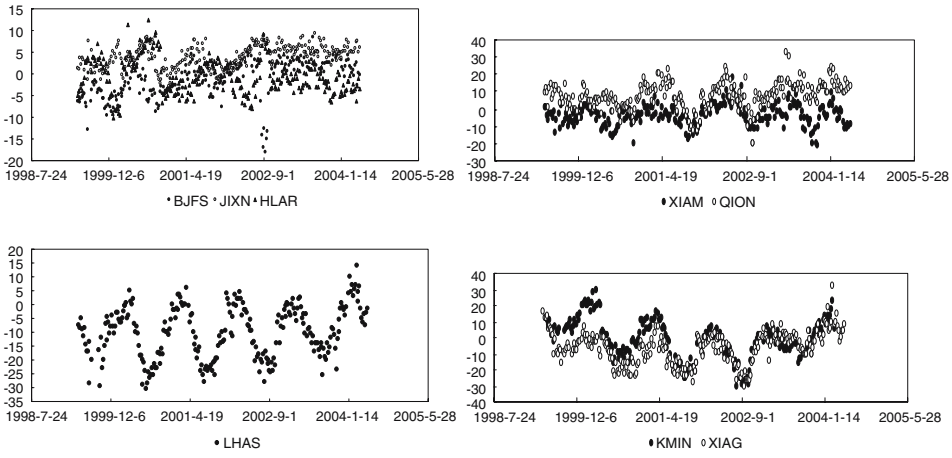


Figure 4
Time series of vertical displacements at some fiducial stations.

WUSH and LHAS, two weeks before the event. There were anomalies in N components with a maximum change of 10–15 mm. The displacements in E component at TASH, URUM and WUSH showed a long-term increase of westward displacements for a year or even longer, and rapid eastward displacements shortly before the event with a maximum change of 10 mm. Analysis on the crustal movement before and after the event suggests that the long-term increasingly westward displacements were closely related with this earthquake.

The crustal movements detected at the fiducial stations showed crustal movements of rotation of Chinese Mainland in a rather large area before the three earthquakes which occurred in September 2001. It has been a rare phenomenon since the continuous observations of fiducial stations in China.

4. Vertical Displacement Time Series

Regional datum definition for vertical displacements is used in this paper for the same reason as for the horizontal displacements. A group of six stable stations with small annual vertical displacement velocities and similar variations are selected for datum definition for vertical displacement time series in this study (Gu *et al*., 2002). Similarity transformation in one dimension is performed to obtain the solutions from the solutions of the global plate motion model or other models. These stations are BJSH, BJFS, JIXN, SHAO, HLAR and SUIY. Figure 4 shows the time series of vertical displacements at some stations. In general there are three types of stations based on differences of variations in the time series of vertical displacements: Stations with annual periodic variations, stations with nonperiodic variations and stations with annual periodic variations in some years. The annual periodic variations have been studied by many authors in ITRF reference frame (Dong *et al*., 2002).

Stations of annual periodic variations are LHAS, KMIN, XIAG, LUZH, XIAA, TASH, WUSH, XNIN and DLHA with nearly the same phase but different amplitudes of 20–50 mm and with a minimum in July to September and a maximum in January to March. Among them the variation at LHAS is quite typical with an amplitude of about 50 mm. All these stations are in western China. These annual variations cannot be explained by annual changes in temperature or rainfall because rain is quite rare in this area. They showed annual uplift and subsidence in a large area. The long-term subsidence at XIAA is due to the subsidence of the monument built in the soil. Small annual periodic variations in horizontal displacement only appear at KMIN with an amplitude about 5 mm.

Stations of nonperiodic variations are BJSH, BJFS, JIXN, TAIN, SHAO, YANC, WUHN, HLAR, SUIY DXIN and URUM. The precision of the weekly solutions of vertical displacement at a station can be estimated from RMS (rooted-mean-squares) of the linear fit of time series of vertical displacements at JIXN and BJSH. It is about 3 mm, better than previous estimates.

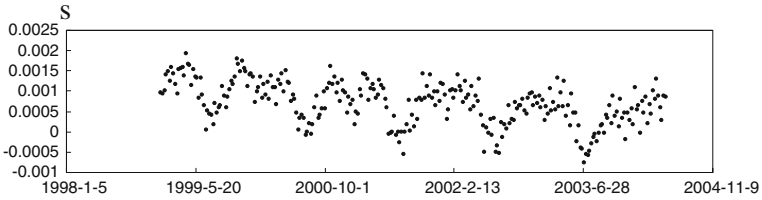


Figure 5
The length of day (LOD) from Jan. 1999 to early June. 2004.

Stations of annual periodic variations in some years are YONG, GUAN and XIAM. There are phase lags of about 2–3 months at these stations in reference to the stations in western China.

The largest change in vertical displacement occurred at KMIN with a total amount of 60 mm. The time series of vertical displacements at KMIN and XIAG are quite similar. They changed from subsidence to uplift in 2003.

The most interesting result of the time series of vertical displacements at fiducial stations in western China except URUM and DXIN is their significant correlation with the length of day (LOD) shown in Figure. 5. LOD increased and or decreased while the highest plateau in the world in western China subsided or uplifted. It is a fact that can be expected from the conservation of the momentum of the earth's rotation.

Acknowledgements

Institutions from the China Seismological Bureau, the Bureau of Surveying and Mapping under the General Staff and the National Bureau of Surveying and Mapping carried out the GPS field observations at the regional and basic stations and most of the fiducial stations. Technicians from the Chinese Academy of Sciences carried out observations at some fiducial stations. The results of data processing at the Data Center of CMONOC have been further processed and analyzed in this study.

References

Dong, D., Fang, P., and Bock, Y. *et al.* (2002), *Anatomy of apparent seasonal variations from GPS derived site position time series,* JGR V107, B4, ETG9-1, ETG9-16.

Gu, Gou-Hua, Shen, Xu-Hui., and Wang, Min. *et al.* (2001), *General Characteristcs of the Recent Horizontal Crustal Movement in Chinese Mainland Acta Seismologica Sinica, 23*(4) 384–393.

Gu, Guohua. and Zhang Jing. (2002), *Time series of displacements from GPS observation at fiducial stations in the Crustal Movement Observation Network of China*, J. Geodesy Geodyn. *22*(2) 61–67.

GU, GUOHUA., FU YANG., and WANG WUXING. (2004), *Analysis on the intensity of horizontal crustal movement in China obtained from GPS observation,* Earthquake, *24*(2) 1–6 (in Chinese).

WANG, MIN., SHEN, ZHENG-KANG., and NIU, ZHI-JUN. *et al*. (2003), *Contemporary crustal deformation of the Chinese continent and tectonic block model [J]*. Science in China (Series D) *46* (Suppl): 25–40.

(Received September 22, 2004, revised December 26, 2005, accepted December, 28, 2005)

Pure appl. geophys. 163 (2006) 2585–2593
0033–4553/06/122585–9
DOI 10.1007/s00024-006-0141-z

Pure and Applied Geophysics

Thermo-hydro-mechanical Modeling of CO_2 Sequestration System Around Fault Environment

Qi Li,[1,2] Zhishen Wu,[1] Yilong Bai,[3] Xiangchu Yin,[3,4] and Xiaochun Li[5,6]

Abstract—Geological sequestration of CO_2 (carbon dioxide) shows great potential to reduce Greenhouse gas emissions. However, CO_2 injection into geological formations may give rise to a variety of coupled chemical and physical processes. The thermo-hydro-mechanical (THM) impact of CO_2 injection can induce fault instability, even possibly lead to seismic activities in and around the disposal reservoir. A sequential coupling approach under some assumptions was proposed in the numerical study to investigate the THM behavior of the CO_2 sequestration system concerning the temperature, initial geological stress, injection pressure and CO_2 buoyancy. The fault was treated as a flexible contact model. The effects of CO_2 injection on the mechanical behavior of the faults were investigated. The Drucker-Prager model and the cap model were used to model the constitutive relationship of formations. The numerical results show that injection pressure sensitively affects the relative slip change of the fault. At the initial stage of the sequestration process, the injection pressure plays a key role in affecting the pore pressure of the formations. However, as time continues, the influence of CO_2-induced buoyancy becomes obvious on the pore pressure of the formations. In general, The THM effects of CO_2 geosequestration do not affect the mechanical stability of formations and faults.

Key words: Sequestration, Greenhouse gas, fault, thermo-hydro-mechanical modeling, Drucker-Prager model, cap model.

1. Introduction

Geological sequestration (geosequestration) of captured CO_2 from large-scale emission sources, such as power stations and cement plants, is becoming one of the effective options to mitigate progressively the global Greenhouse effect. A conceptual

[1] Department of Urban and Civil Engineering, Ibaraki University, 4-12-1 Nakanarusawa, Hitachi, Ibaraki 316-8511, Japan.
[2] Now with the Institute of Geology and Geoinformation, National Institute of Advanced Industrial Science and Technology (AIST), AIST Central 7, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8567, Japan. E-mail: qi.li@aist.go.jp
[3] Institute of Mechanics, Chinese Academy of Sciences, Beijing 100080, China.
[4] Center for Analysis and Prediction, China Earthquake Administration, Beijing 100036, China.
[5] Research Institute of Innovative Technology for the Earth (RITE), Kizugawadai 9-2, Kyoto 619-0292, Japan.
[6] Now with the Institute of Rock and Soil Mechanics, Chinese Academy of Sciences, Wuhan 430071, China.

illustration of CO_2 geosequestration is depicted in Figure 1. The captured CO_2 is injected into the geological formations, such as unminable coal beds, depleted oil or gas reservoirs, and deep saline aquifers. In particular, the saline aquifers in sedimentary basins have a great storage capacity and most extensive distribution in Japan (TANAKA *et al.*, 1995). However, Japan Island Arc is located in a tectonically active region, with many major and minor faults intersecting the area. Consequently, the effects of CO_2 injection on faults must be evaluated with respect to the possibility of induced seismicity and leakage (KAYA *et al.*, 2001; LI *et al.*, 2002; STREIT and HILLIS, 2004).

Although considerable research on the CO_2 geosequestration has been done worldwide, the thermo-hydro-mechanical (THM) behavior of the sequestration system around the fault environment during the CO_2 injection has not yet been thoroughly studied (RUTQVIST *et al.*, 2002). Significant in restricting the research is the lack of detailed understanding of the THM modeling and CO_2 phase change. In this paper, an easy-to-accomplish THM numerical approach was proposed to model the sequestration system around the fault environment. The THM behavior of the CO_2 geosequestration system was investigated concerning the temperature, initial geological stress, injection pressure and CO_2 buoyancy. The effects of CO_2 injection on the mechanical behavior of the fault were investigated by a sequential coupling scheme. The simulation results show that the injection pressure seriously affects the relative slip change of the fault. At the initial stage of the injection, the pore pressure of the sequestration formations is obviously affected by the injection pressure. However, as time passes, the CO_2 plume induced buoyancy plays a key role in the influence of the pore pressure of the geosequestration system.

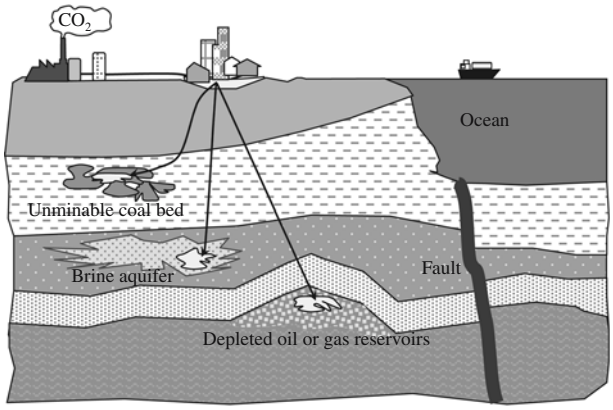


Figure 1
Schematic illustration of CO_2 geosequestration.

2. Structural Geometry and Computing Procedure

A two-dimensional plane strain supposition is applied to model the geosequestration site with a vertical shallow fault and four different formation layers, as depicted in Figure 2. The size of the structural model is 5000 m laterally and 2000 m vertically. The storage formation is at the depth of 1100 m and the injected CO_2 is kept in supercritical status. The CO_2 plume is assumed to extend to 1000 m wide along the bottom of the sealed cap formation around the injection well and 100 m thick. The depth of the injection part of the well extends from 1200 m to 1400 m.

In the finite-element analysis, modeling of the status development of the fault can be treated as a contact problem. In the present research, a classical spring model is adopted to consider the mechanical changes of the fault. Two flexible joint springs are used to devote to the simulation of the normal and tangential mechanical behavior of the fault. The normal stiffness 2.0×10^7 N/m and shear stiffness 1.0×10^7 N/m of the fault surfaces are used in the analysis (Li *et al.*, 2002).

The constitutive relationship of the overlying formation and the host formation are modeled by using the Drucker-Prager plasticity model, which assumes the nonassociated flow. The linear form of the Drucker-Prager model with no intermediate principal stress effect is used. A low permeability is assumed for the overlying formation, while a high permeability is assigned to the host formation. The cap formation and storage formation are modeled by using the cap (modified Drucker-Prager) plasticity model. The material properties used in the analysis are listed in Table 1. Because of a lack of field and experimental data, some of the formation properties were estimated from the literature data (National Astronomical Observatory, 2004).

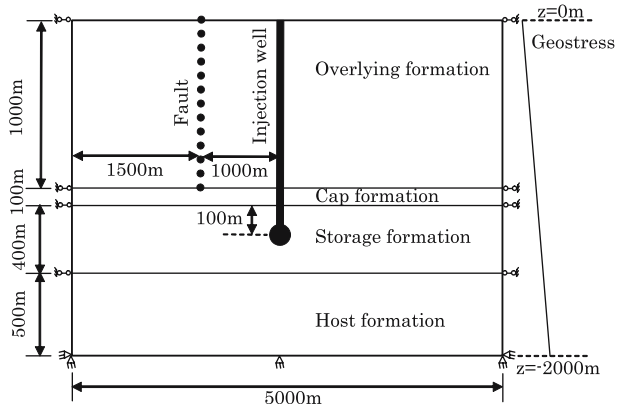


Figure 2
Computational domain of geosequestration modeling (disproportional scale).

Table 1

Material properties

Formation	Young's modulus [10^6 Pa]	Poisson's ratio [–]	Density [kg/m^3]	Permeability [10^{-5} m/s]
Overlying layer	434	0.25	1900	1.1760
Cap layer	546	0.25	1900	0.4939
Storage layer	494	0.25	1900	0.4939
Host layer	2482	0.25	2600	4.4685

In order to investigate the THM behavior of the geosequestration system around the fault environment, a simple sequential coupling process is taken into the simulation studies. The procedure of numerical simulation is run in five major steps. The first step is a self-respect analysis to equilibrate geostatic gravity loading of the computational domain. This step also contributes to establish the initial distribution of the porosity and the temperature field (10 degrees Celsius). The effective stress principle is adopted in the computation, which can be accomplished by defining the pore fluid pressure as the pore pressure in excess of the hydrostatic pressure required to support the weight of pore fluid above the elevation of the material mass point. The second step is a one-month fast consolidation process to further equilibrate any inelastic effects induced from the initial gravity loading of step one. The third step is to simulate the CO_2 injection by prescribing an excess pore pressure (1×10^6 Pa) at the injection part of the well. The fourth step simulates the CO_2 plume induced buoyancy as distributed forces (0.4×10^6 Pa) around the well and along the bottom of the cap formation (Li *et al.*, 2002). The final step consists of a consolidation analysis performed over half a year period to investigate the pore pressure dispersion in the sequestration system around the fault environment.

3. Constitutive Model of Geological Materials

The Drucker-Prager model and the cap model are classical options to simulate the constitutive behavior of geological materials in the numerical analysis. As we known, the Drucker-Prager model provides for a possibly noncircular yield surface in the deviatoric plane (π plane) to match different yield values in triaxial tension and compression, associated inelastic flow in the deviatoric plane, separate dilation and friction angles. Input data parameters define the shape of the yield and flow surfaces in the meridional and deviatoric planes as well as other characteristics of inelastic behavior. As shown in Figure 3, the Drucker-Prager model is intended primarily for applications in which the stresses are for the most part compressive. The yield criterion is written as

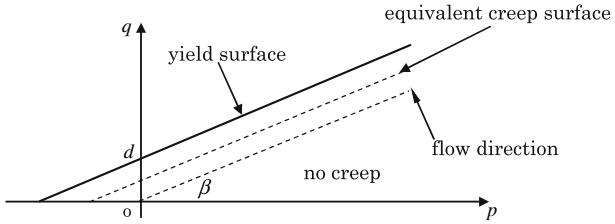


Figure 3
Linear Drucker-Prager model: Yield surface, flow direction and equivalent creep surface.

$$F_s = q - p\tan\beta - d = 0, \tag{1}$$

where p is the equivalent pressure stress, q is the Mises equivalent stress, d is the cohesion if hardening is defined by the uniaxial compression yield stress, β is the slope of the linear yield surface in the $p-q$ stress plane and is commonly referred to as the friction angle of the material (CHEN and HAN, 1988).

On the other hand, the cap model is actually a modification product of the Drucker-Prager model. The addition of the cap yield surface to the Drucker-Prager model serves two main purposes: it bounds the yield surface in hydrostatic compression, thus providing an inelastic hardening mechanism to represent plastic compaction; and it helps to control volume dilatancy when the material yields in shear by providing softening as a function of the inelastic volume increase created as the material yields on the Drucker-Prager shear failure surface. The yield surface has two principal segments: A pressure-dependent Drucker-Prager shear failure segment and a compression cap segment, as shown in Figure 4. The Drucker-Prager failure segment is a perfectly plastic yield surface. Plastic flow on this segment produces inelastic volume increase (dilation) that causes the cap to soften. On the cap surface plastic flow causes the material to compact. The shear failure surface is defined as

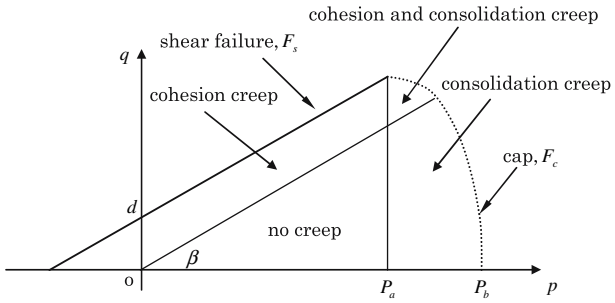


Figure 4
Cap model: Yield surfaces and regions of activity of creep mechanisms.

Table 2

Inelastic parameters of constitutive models

Formation	d [10^6 Pa]	R [–]	β [Degree]	K [–]	φ [Degree]	α [–]
Overlying layer	–	–	36.0	1.0	0.0	–
Cap layer	1.38	0.34	42.0	1.0	–	0.0
Storage layer	1.38	0.30	40.4	1.0	–	0.0
Host layer	–	–	38.0	1.0	0.0	–

similar to the Drucker-Prager model. But the cap yield surface has an elliptical shape with constant eccentricity in the meridional $p-q$ plane and also includes dependence on the third stress invariant in the deviatoric plane. The cap surface begins hardening or softening as a function of the volumetric inelastic strain: Volumetric plastic and creep compaction (when yielding on the cap and creeping according to the consolidation mechanism) causes hardening, while volumetric plastic and creep dilation (when yielding on the shear failure surface and creeping according to the cohesion mechanism) causes softening. The cap yield surface is

$$F_c = \sqrt{(p-p_a)^2 + (Rq)^2} - R(d + p_a \tan\beta) = 0, \tag{2}$$

where R is a material parameter that controls the shape of the cap, p_a is an evolution parameter that represents the volumetric inelastic strain driven hardening or softening. The hardening or softening law is a piecewise linear function relating the hydrostatic compression yield stress p_b and volumetric inelastic strain. The evolution parameter p_a is given as

$$p_a = \frac{p_b - Rd}{(1 + R\tan\beta)}. \tag{3}$$

This model has two possible creep mechanisms that are active in different loading regions: One is a cohesion mechanism, which follows the type of plasticity active in the shear-failure plasticity region, and the other is a consolidation mechanism, which follows the type of plasticity active in the cap plasticity region (ZIENKIEWICZ and TAYLOR, 2000).

In the computation, the values of parameters used, which are discussed in this section, are listed in Table 2.

4. Results and Discussion

A sequential approach was presented to investigate the THM coupling process of the geosequestration system around a vertical fault environment, and this numerical

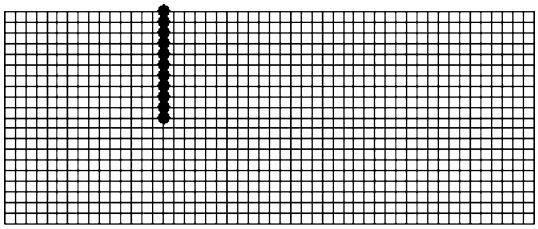


Figure 5
Finite-element mesh and node position of fault.

scheme was accomplished by the finite-element method. The Drucker-Prager model and the cap model were applied to simulate the constitutive behavior of formations. The flexible contact model was used to model the fault in the simulation. Figure 5 plots the finite-element mesh and node position of the fault.

The effects of CO_2 injection on the relative change magnitude of the fault slip are clearly presented in Figure 6. The injection pressure has a larger influence on the relative slip change of the fault than the buoyancy induced by the CO_2 plume. It can be further understood that the deep end of the fault shows broad change variation because of its proximity to the disposal zone of CO_2. After half a year consolidation of the formations, the total change of relative slip along the fault becomes small and weak. It should be noted that the injection pressure should be well controlled in practice to avoid causing the sharp fault slip during the sequestration process. However, after the injection is finished and as time continues, the influence of CO_2 buoyancy becomes obvious on the sequestration system. Figure 7 shows the pore pressure distribution of the sequestration system at three typical stages. The distinct change of the pore pressure of the formations caused by the injection process, i.e., the injection pressure and the CO_2 buoyancy, sharply dissipates with the consolidation process over six months.

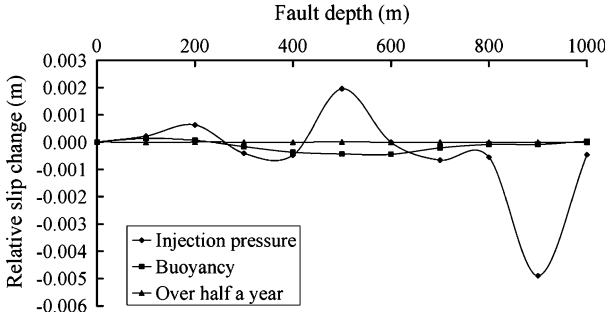


Figure 6
Relative slip change of fault.

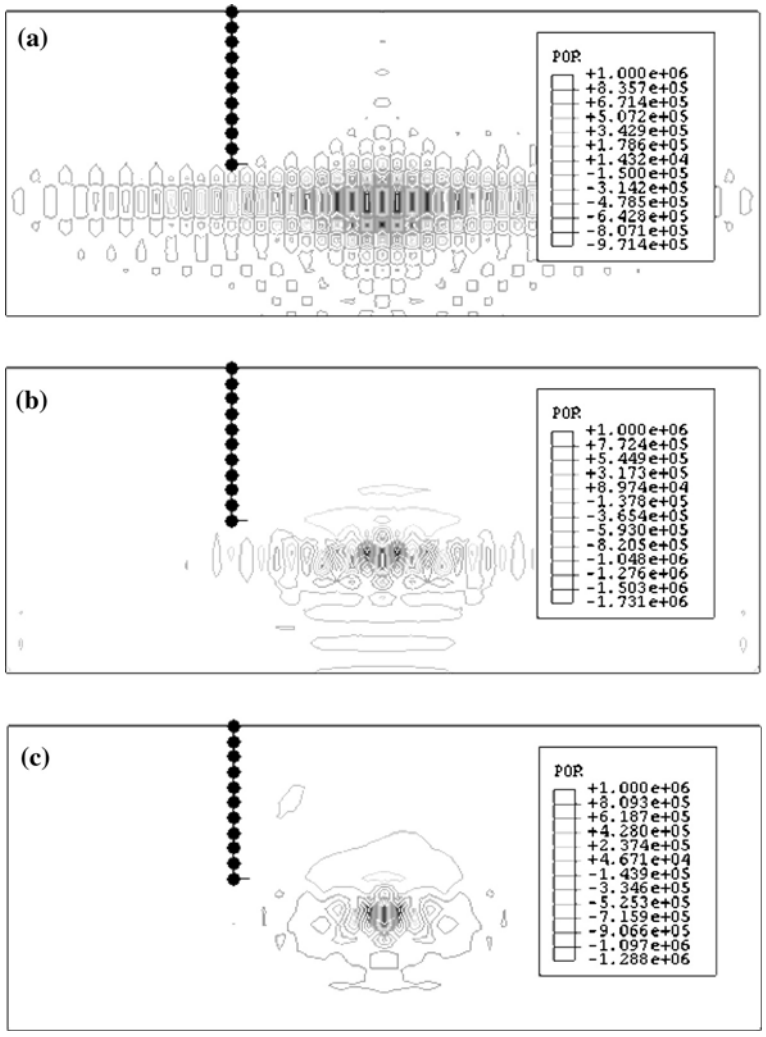


Figure 7
Contour plot of pore pressure distribution, (a) under injection pressure, (b) under buoyancy, (c) over half a year.

5. *Concluding Remarks*

Numerical studies indicate that it is important to control the injection pressure during the sequestration so as to avoid causing the sharp slip of the fault and the serious change of the pore pressure in formations. The results also provide a preliminary outlook on the THM analysis of the CO_2 geosequestration around the fault environment. Despite somewhat simplified suppositions of the analysis, and a lack of field and experimental data, this work is important not only for studies

concerning artificial reservoir problems such as CO_2 geosequestration or underground oil storage, but also for researches of reservoir-induced earthquake processes.

Acknowledgements

This work was jointly supported by the Chinese Academy of Sciences (CAS), China and the Research Institute of Innovative Technology for the Earth (RITE), Japan.

References

Chen, W.F., and Han, D.J., *Plasticity for Structural Engineers* (Springer-Verlag, New York 1988).

Kaya, Y., Koide, H., Nakaya, S., Ohsumi, T., Takita, K., and Yoshimura S., *Underground sequestration of carbon dioxide in tectonically active area*. In *Greenhouse Gas Control Technologies* (eds. Williams D., Durie B., McMullan P., Paulson C., and Smith A.) (CSIRO Publishing, 2001) pp. 296–298.

Li, Q., Wu, Z.S., Li, X.C., Ohsumi, T., and Koide, H. (2002), *Numerical simulation on crust deformation due to CO_2 sequestration in deep aquifers*, J. Appl. Mech., JSCE *5*, 591–600.

National Astronomical Observatory, *Rika nenpyo (Chronological Scientific Tables 2005)*, (Maruzen Co. Ltd., Tokyo 2004) (in Japanese).

Rutqvist, J., Wu, Y.S., Tsang, C.F., and Bodvarsson, G. (2002), *A modeling approach for analysis of coupled multiphase fluid flow, heat transfer, and deformation in fractured porous rock*, Int. J. Rock Mech. Mining Sci. *39*, 429–442.

Streit, J.E. and Hillis, R.R., *Building geomechanical models for the safe underground storage of carbon dioxide in porous rock*. In *Greenhouse Gas Control Technologies* (eds. Gale, J. and Kaya, Y.) (Pergamon Publishing, 2004) pp. 495–500.

Tanaka, S., Koide, H., and Sasagawa, A. (1995), *Possibility of underground CO_2 sequestration in Japan*, Energy Conversion and Management *36*, 527–530.

Zienkiewicz, O.C. and Taylor, R.L., *Finite-Element Method, vol. 1, 3* (Butterworth Heinemann, London 2000).

(Received December 30, 2004, revised December 25, 2005, accepted December 30, 2005)
Published Online First: December 20, 2006