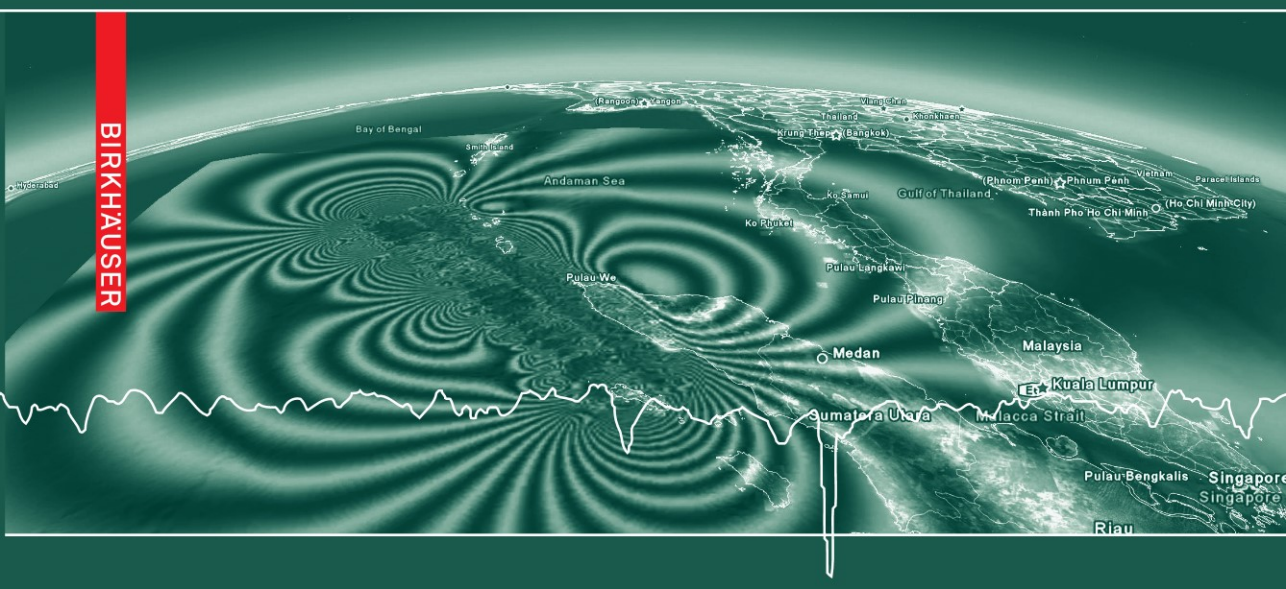
Earthquakes: Simulations, Sources and Tsunamis

Edited by
Kristy F. Tiampo
Dion K. Weatherley
Stuart A. Weinstein



pageoph topical volumes

BIRKHÄUSER

Earthquakes: Simulations, Sources and **Tsunamis**

Edited by
Kristy F. Tiampo
Dion K. Weatherley
Stuart A. Weinstein

Birkhäuser
Basel · Boston · Berlin

Reprint from Pure and Applied Geophysics
(PAGEOPH), Volume 165 (2008) No. 3 - 4

Editors:
Kristy F. Tiampo
Department of Earth Sciences
University of Western Ontario
N6A 5B7 London, Ontario
Canada
Email: ktiampo@uwo.ca

Dion K. Weatherley
Earth Systems Science
Computational Centre (ESSCC)
Department Earth Sciences
University of Queensland
4072 St. Lucia, Queensland
Australia
Email: dion@esscc.uq.edu.au

Stuart A. Weinstein
Pacific Tsunami Warning Center (PTWC)
91-270 Fort Weaver Road
Ewa Beach Hawaii 96706
USA
Email: stuart.weinstein@noaa.gov

Library of Congress Control Number: 2008924894

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

ISBN 978-3-7643-8756-3 Birkhäuser Verlag AG, Basel · Boston · Berlin

Basel · Boston · Berlin
P.O. Box 133, CH-4010 Basel, Switzerland
Part of Springer Science+Business Media
Printed on acid-free paper produced from chlorine-free pulp. TCF ∞
Cover photo: Xiaoru Yuan, Peking University, Beijing, P.R. China
Printed in Germany

ISBN 978-3-7643-8756-3
9 8 7 6 5 4 3 2 1

e-ISBN 978-3-7643-8757-0
www.birkhauser.ch

Contents

Pure appl. geophys. 165 (2008) 449–450
0033–4553/08/030449–2
DOI 10.1007/s00024-008-0332-x

Pure and Applied Geophysics

Earthquakes: Simulations, Sources and Tsunamis

K.F. Tiampo,[1] D.K. Weatherley,[2] and S.A. Weinstein[3]

1. Introduction

Large earthquakes pose an almost incalculable risk, causing great destruction and, often, loss of life. While large regions of the world are susceptible to this potentially devastating natural hazard, the Asia-Pacific region is particularly vulnerable. Poignant examples in the recent past include the M $\sim$ 9.2 Northern Sumatra earthquake and tsunami in December, 2004 which resulted in the deaths of nearly 300,000 people, the M$\sim$7.6 Taiwan earthquake of September, 1999, with a death toll exceeding 2000 and $20 billion in damage, and the M $\sim$ 7.2 Kobe, Japan earthquake of January, 1995, that killed 5500 people and induced losses of more than $150 billion.

In the 1990s, major advancements in seismic research greatly added to the understanding of earthquakes fault systems as complex dynamical systems. Large quantities of new and extensive remote sensing data sets provided information on the solid earth on scales previously inaccessible, and were integrated with a combination of innovative analysis techniques and advanced numerical and computational methods. A new paradigm in the advancement of earthquake science evolved through advances in high performance computing and numerical simulation methods, coupled with these new understandings of earthquake physics. The resulting advances and discoveries include the LSM (Lattice Solid particle simulation Model), the Australian Computational Earth Systems Simulator (ACcESS), Japan's Earth Simulator, GeoFEST, QuakeSim, MFEM (Multiscale Finite-Element Model), LURR (Load-Unloaded Response Ratio), PI (Pattern Informatics), friction law studies and their link to seismicity and stress, episodic tremor and slip, and studies on information transfer and the interaction between faults, among others. Many of these are included in or resulting from APEC Cooperation for Earthquake Simulation (ACES) related research and activities, and are presented in this volume.

ACES has offered a unique and innovated opportunity to scientists at all levels to gain new understandings of the earthquake nucleation process, precursory phenomena, and

[1] Department of Earth Sciences, University of Western Ontario, London ON, Canada.
E-mail: ktiampo@uwo.ca
[2] Earth Systems Science Computational Centre, ESSCC, The University of Queensland, St.Lucia., Brisbane Qld 4072, Australia.
[3] Pacific Tsunami Warning Center, 91-270 Ft. Weaver Rd, Ewa Beach, HI 96706, U.S.A.

space-time earthquake patterns, and has improved and enhanced earthquake hazard quantification and forecasting capabilities. The focus of the 5th ACES Workshop, April 4–6, 2006, Maui, Hawaii, was not only to continue the highly successful ACES collaboration, but to expand into studies of new satellite measurements, focusing on Interferometric Synthetic Aperture Radar (InSAR) data, and earthquake and tsunami modeling and forecasting, while still concentrating on advancing and enabling modeling and simulation techniques and computational methods, including web services and petascale computing.

This volume attempts to present the current state of the field, in a sampling of work from within this broad spectrum, by focusing not only on the modeling of earthquake and earthquake generated tsunamis, but also on practical comparisons of the resulting phenomenology. It also should be noted that work herein is indicative of both a broad spectrum of collaborative earthquake and tsunami research and the international nature of the contributions.

The volume begins with a variety of studies that focus on the modeling of tsunamis and earthquakes. These include both large-scale simulation and visualization programs, as well as detailed models of small-scale features. Particular attention is paid to the application of these techniques. Examples include the application of earthquake simulations for the support of tsunamis warning systems as well as the modeling and visualization of tsunamis themselves, the use of finite-element models to investigate mid-crustal faulting, deformation and gravity, and the genesis of phenomena such as slow earthquakes and evolving geometries. Several additional works discuss research into particular computational techniques, languages, and hardware that can be used to facilitate data analysis, visualization, and modeling of both the earthquake process itself and the associated phenomena, such as InSAR deformation fields. Studies of forecasting techniques such as the Stress Accumulation Method (SAM), the Pattern Informatics (PI) technique, and the Load-Unload Response Ratio (LURR) follow, including comparisons of their results with historic earthquake data. Finally, the volume concludes with theoretical analyses of statistical properties of seismicity by internationally recognized experts in the field.

The editors would like to personally thank all the reviewers whose energy and efforts were essential to ensuring the high standards of this volume—S. Abe, D. Arcas, A. Corral, J. Davidsen, M. Davies, D. DePaor, G. Fryer, E.L. Giest, S. Gross, T. Hayes, J. Holliday, L. Hood, A. Jimènez, L. Kellogg, B. Kennett, L. Kettle, S. Kohen-Kadosh, V. Kossobokov, P. Li, H. Muhlhaus, K. Olsen, P.L. Olson, J. Parker, R. Robinson, J.B. Rundle, S. Samsonov, V. Sardina, B. Shiro, S.N. Ward, C. Williams, J. Weertman, L. Wyborn, H. Xing and Y. Yang.

The editors hope that this issue will contribute to a better understanding of the link between critical point phenomena and precursory seismicity and, in particular, the premonitory changes resulting from stress transfer in this correlated system.

Pure appl. geophys. 165 (2008) 451–474
0033–4553/08/030451–24
DOI 10.1007/s00024-008-0316-x

Pure and Applied Geophysics

Finite Fault Modeling in a Tsunami Warning Center Context

STUART A. WEINSTEIN,[1] and PAUL R. LUNDGREN[2]

Abstract—The US NOAA/NWS tsunami warning centers have relied on earthquake location and depth, and scalar measures of earthquake size and slowness to assess the potential for the generation of a destructive tsunami by an earthquake. Recent earthquakes, such as Peru 2001, Sumatra 2004 and the Java 2006, manifest some of the difficulties the warning centers face as they try to cope with unusual earthquakes. We have undertaken a study of a simple teleseismic waveform inverse model and applied it to the earthquakes of June 23, 2001 in Peru and of July 17, 2006 in Java. Synthetic numerical experiments suggest that the most salient features of the rupture history of an earthquake can be recovered. Furthermore the calculations can be conducted quickly enough to be useful in a warning center context.

We have applied our technique to the Peru 2001 and recent Java 2006 earthquakes. Our overall results are consistent with those obtained from other studies. The results show why the Peru event initially looked slow to the US tsunami warning centers and that the Java event is a truly slow or tsunami earthquake. Clearly, the warning centers stand to benefit from an increased understanding of the earthquakes they monitor. Furthermore, better knowledge of the slip distribution along a fault will improve tsunami wave-height forecasts.

Key words: Finite fault modeling, earthquake rupture history, tsunami warning center.

1. Introduction

The Pacific Tsunami Warning Center (PTWC) and the West Coast, Alaska Tsunami Warning Centers are tasked with providing advanced warning of dangerous tsunamis to the US, Canada and many other Nations within our areas of responsibility (AOR). Collectively, this includes all US Mainland and Canadian coasts, US assets in the Caribbean and Pacific, as well as many international clients in the Indian, Pacific and Caribbean Ocean basins. PTWC informs its clients within minutes of a large earthquake occurring in these areas of the location, magnitude and assessment of the potential for tsunamigenesis. A principle difficulty the warning centers have had has been the inability to predict the size of tsunami waves when they arrive on distant shores. Due to the conservative posture of the warning system, this shortcoming has resulted in unnecessary evacuations. NOAA's PMEL (Pacific Marine Environmental Laboratory) has developed

[1] Richard H. Hagemeyers Pacific Tsunami Warning Center, Ewa Beach, HI, 96706.
E-mail: stuart.weinstein@noaa.gov
[2] Paul R. Lundgren, Jet Propulsion Laboratory, Pasadena, CA, 91109.

models and tools (SIFT) to give the US warning centers a nascent ability to make tsunami wave height forecasts.

The SIFT software breaks the major subduction zones of the Pacific into grids consisting of elements or fault blocks 100 km along strike and 50 km down dip. When a large earthquake occurs, the duty personnel input the hypocenter and magnitude of the earthquake into the SIFT software. Further, they also select the blocks excited by the earthquake. An elastic dislocation model is then used to compute the sea-floor displacement. A simulation database, amounting to a numerical Green's functions approach is then used to quickly generate the wave height forecast. The length of the rupture and amount of slip are determined by basic scaling relationships. Initially, the slip is used for all excited blocks. Clearly, this is not normally the case, and there are often complications with the rupture process that may render this source estimate procedure inadequate, resulting in a less than optimal forecast. For example "late" earthquakes like Peru 2001 or bilateral rupturing can result in large forecast errors in the near field, and significant errors in the far field. The slip distribution can be further refined through the inversion of monograms obtained by the DART network.

This study is a step towards the warning center's use of more realistic source parameters rather than prescriptions based on scaling laws that may be inadequate for complex events resulting in forecast errors. We use a variant of the method used by MENDOZA (1996) to perform waveform inversions in an effort to better characterize earthquake sources. We have performed waveform inversions using only P-wave broadband waveforms for the PERU (June, 2001) and the recent, JAVA (July, 2006) earthquakes. These inversions are performed using seismic data the warning centers are likely to receive quickly enough to be of use in a warning context.

Our results suggest that considerably more reliable source parameters than those determined using scaling relationships can be obtained with enough speed to be useful in a tsunami warning system context. These parameters include the distribution of slip, the source time function and length of rupture. The latter are important for use in diagnosing extreme events like Sumatra (December, 2004) and the so-called "slow" or "tsunami earthquakes" such as the Java (July, 2006) earthquake. Tsunami earthquakes result in tsunamis that are larger and more destructive than expected, based on the earthquake's moment (KANAMORI, 1972; KANAMORI and KIKUCHI, 1993; BILEK and LAY, 1999). These types of events present tough challenges to warning centers that must quickly recognize them, especially when they occur in the warning center's local area and where an urgent warning must be issued.

2. *Numerical Method*

The US warning centers receive approximately 150 broadband data streams from the GSN via the NEIC, ASL, IRIS and other vendors. Due to the current limitations of our communications, and a desire to maximize our coverage of the planet, we receive only

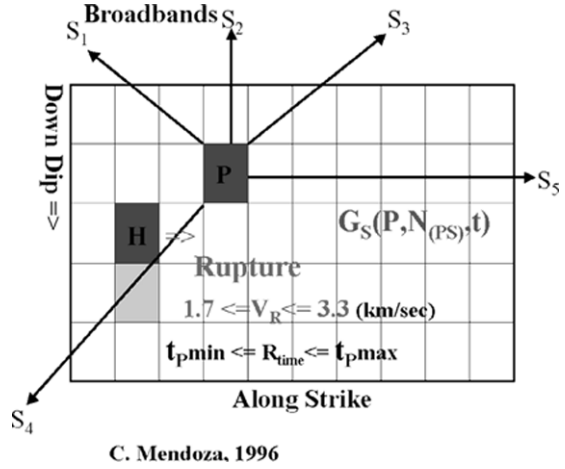


Figure 1
Diagram of finite fault model showing discetization of fault plane into teleseismic fault sources and the determination of rupture time-delay windows.

the vertical component. Any methods used by the warning centers must take into account the above limitations as well as provide results in a timely fashion.

We have explored the use of an inverse technique that is a variant of that used by MENDOZA (1996) to perform waveform inversions to better characterize the earthquake source. This method is one of a class of inversion methods (HARTZELL and HEATON, 1983, 1986; HARTZELL and MENDOZA, 1991; WALD and HEATON, 1994) in which the waveform inverse problem is solved by forming a matrix of Green's functions that is then inverted for the rupture history.

Our computational method draws its design from MENDOZA (1996), and uses P-wave data obtained from vertical component broadband waveforms. Using P waves allows the calculation to proceed sooner as opposed to using S or surface waves, for which we may need to wait 10–30 minutes or more to accumulate enough data to invert. In our method, the fault plane (Fig. 1) is broken up into a uniform grid of teleseismic point sources or "patches." Each patch has a number of time-delay windows over which faulting can occur. Given near source and receiver structure, a basic Green's function can then be determined for each patch-station couplet. The resulting mathematics reduces to a linear inverse problem to find the combination of scaling factors (M_o per patch that can be converted to slip) for each Green's function such that the synthetic waveforms resulting from the inversion best approximate the observed waveforms in a least-squares sense.

For each patch a time span over which the patch can potentially rupture is computed, based on the distance of the patch from the hypocenter, and on the minimum, maximum

rupture velocities assumed for the calculation. This time span is partitioned into 1 s intervals or time-delay windows. A time offset is built into the Green's functions, taking into account the delay resulting from rupture propagation and travel-time differences with respect to the hypocenter. The number of actual time-delay windows for each patch is additionally constrained by the requirement that for each patch-station couplet Green's function, at least 40 seconds are available for the computation. For example, if the total length of record is 170 s, and the total delay is $\geqq$ 170 s, then there is no contribution from the Green's function for this patch, time-delay widow, station triplet. No Green's function is used unless at least 40 s of the Green's function is available. The time span is divided into time-delay windows spaced one second apart. Therefore the elemental source time functions, which are triangular, can form longer duration and potentially more complex patch source time functions via superposition. The resulting system of equations is solved using the NNLS (non-Negative Least-Squares) algorithm of LAWSON and HANSON (1995). NNLS is a feasible set method that constrains the slips to zero or positive values.

The linear system is illustrated in Figure 2. Here G stands for Green's function and the first index is the time window number (N) (the number of which depends on patch-station couplet), the second is patch number (P), and the third is time (T). The subscript refers to the station for which the Green's function is computed (S). X is the weight (slip) vector where the first subscript is window number and the second is patch number. And finally W is the vector of waveforms, where the subscript is station and the index is time. The last equation constrains the moment and is optional. In an effort to keep computation time down, no smoothing is attempted.

We compute bodywave Green's functions for teleseismic point source waveforms (KIKUCHI and FUKAO, 1985) that include the contributions from the P, pP and sP "depth" phases as well as other reflected and transmitted near source and receiver phases. This

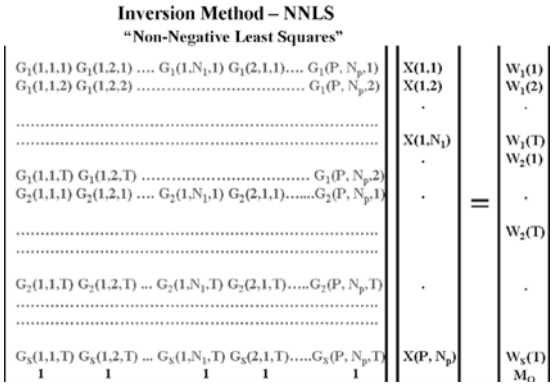


Figure 2
Description of the linear system solved in this study. For the Green's functions, the first index is the patch number, the second is the time window index, and the third is time.

Green's function requires several different parameters and other information as inputs. These are the near source and near receiver Earth structure, fault orientation, elemental source time function, geometric spreading and attenuation. The fault geometry can be determined by a quick CMT. The separation of the depth phases, and those due to near source and receiver structure, are functions of the take-off angle (distance) and azimuth of the receiving station. The relative amplitudes and polarities of the P and S components depend on the receiver location and the focal mechanism.

Performing the finite fault inversion problem requires specifying the number and dimensions of the fault patches, as well as the permissible range of rupture speeds. The inversions can be carried out with the moment constrained or as a free parameter, while the fault strike, dip and rake are fixed. We have a version of the algorithm in which rake is allowed to vary.

3. Synthetic Tests

In an effort to better understand the properties of our model and offer guidance in interpreting results obtained with actual earthquake data, we have performed a series of synthetic numerical experiments. To conduct these experiments, we have constructed a synthetic model with a fairly complicated rupture and a slip distribution similar to what has been obtained for the PERU (June, 2001) earthquake by previous studies. For these simulations, we have adopted the fault parameters given by Harvard's CMT solution for the Peru 2001 earthquake and the structure model used by GIOVANNI *et al*., (2002). The fault parameters are dip $= 18°$, rake $= 63°$, and strike $= 310°$. The hypocenter is placed at 23 km depth, and the rupture is assumed to propagate smoothly at 2.5 km/sec. An elemental source time function rise time of 3 seconds is used. Synthetic waveforms are computed, using the same Green's functions and parameters, at the same broadband sites used for the actual Peru earthquake data. The waveforms have a duration of 170 s with a sample interval of 0.2 s. The stations and their locations are shown in Figure 3a. We have conducted several sets of synthetic experiments. In the first set, the inversion is performed using synthetic elemental source time function rise times of 1, 3 and 6 seconds. In each case the inversion uses $M_o = 1.87 \times 10^{21}$ N-m (where M_o is a fixed parameter), and minimum, maximum rupture speeds of 1.75 km/s and 3.0 km/s, respectively. A value of 3.0×10^{10} Pa-s is assumed for the shear rigidity and is used to convert the moments obtained for each patch by the inversions into slip. The inversions use the same earth structure model as GIOVANNI and BECK *et al*. (2002) (see Table 1).

The synthetic slip distribution, rupture timing, and corresponding source time function are shown in Figures 4a and 5a. The synthetic slip distribution is comprised of two broad asperities; the left one with the greatest slip, and a small asperity with a modest slip equidistant between them. The hypocenter is in the center of the asperity on the right and the rupture propagates horizontally from right to left. In this study, the rupture timing

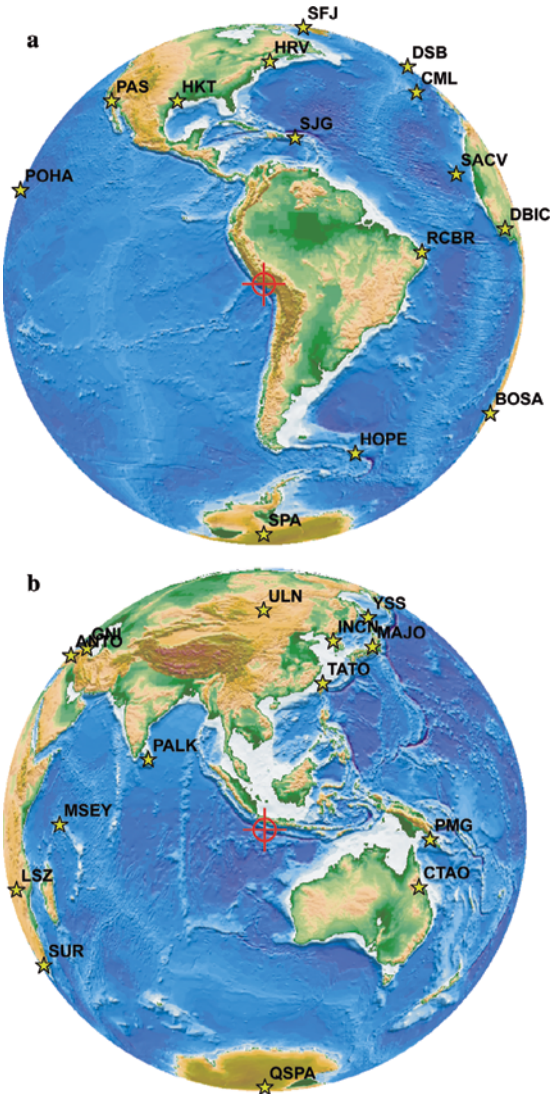


Figure 3
Station distribution used for the Peru 2001 (a) and Java (2006) (b) earthquakes.

for each patch is determined by computing the weighted average time based on the slip in each time window.

The results of the first group of synthetic experiments (Figs. 4–6) reveal that for the case in which an elemental source time function of 3 s is used in the inversion (1 s not

Table 1

Peru earthquake structure model

Vp km/s	Vs km/s	ρ kg/cm^3	Thickness (km)
Source			
1.5	0.0	1.0	4.4
6.5	3.7	2.8	30.0
8.1	4.64	3.2	
Receiver			
6.5	3.74	2.87	35.0
8.1	4.64	3.2	

shown, but are very similar to the 3 s waveforms), the waveform fits are precise, and the original slip distribution and event source time function are well recovered. The most notable difference is that the peak at 85 s is blunted. The rupture timing shows more differences, but the differences tend to concentrate in areas with little or no slip. The results are better for 3 s, as the slip distribution obtained with a 1 s elemental source time function shows slightly more defocusing for the large asperity on the left. When the elemental source time function of the inversion is increased to 6 s, the results shown a marked deterioration. This is not unexpected, as there is no way to combine triangular pulses with a rise time of 6 s to form a triangular pulse with a rise time of 3 s. Clearly, using an elemental source time function rise time that is larger than the actual one yields degraded results.

When the moment is a free parameter (Figs. 7–9), the resulting slip distributions and waveforms are similar to those obtained in the fixed $\mathbf{M_o}$ case. The moments resulting from the inverse for these three cases are 19.95, 20.32, and 30.53 $\times$ 10^{21} N-m. Given that the actual moment was 18.7 $\times$ 10^{21} N-m; these values represent deviations of 6.7%, 8.6% and 63.3%. Note the different slip scales on these figures. Despite having a slightly larger moment, the peak slips for each of the asperities are closer to the original fault model in the solution obtained with a 3 s rise time. All of the synthetic simulations suggest that using too small a rise time may result in defocusing and too large a rise time can cause errors in the moment and degrade the overall results of the inversion. These results instill confidence, that with the given distribution of stations, the most salient features of the slip distribution can be recovered.

3.1. Peru Earthquake of June 23, 2001

What was then the largest earthquake in over 25 years which occurred at 20:33 on June 23, 2001 in Peru. The PTWC watchstanders were paged nine minutes later at 20:42. Another nine minutes later PTWC issued a Tsunami Information Bulletin informing our clients that an earthquake occurred in Peru with a magnitude of 7.2. This magnitude estimate was arrived at using the Mwp magnitude method (Tsuboi *et al*., 1995). Another 20 minutes passed and at 21:12 PTWC, using the Mantle Magnitude method (Okal and

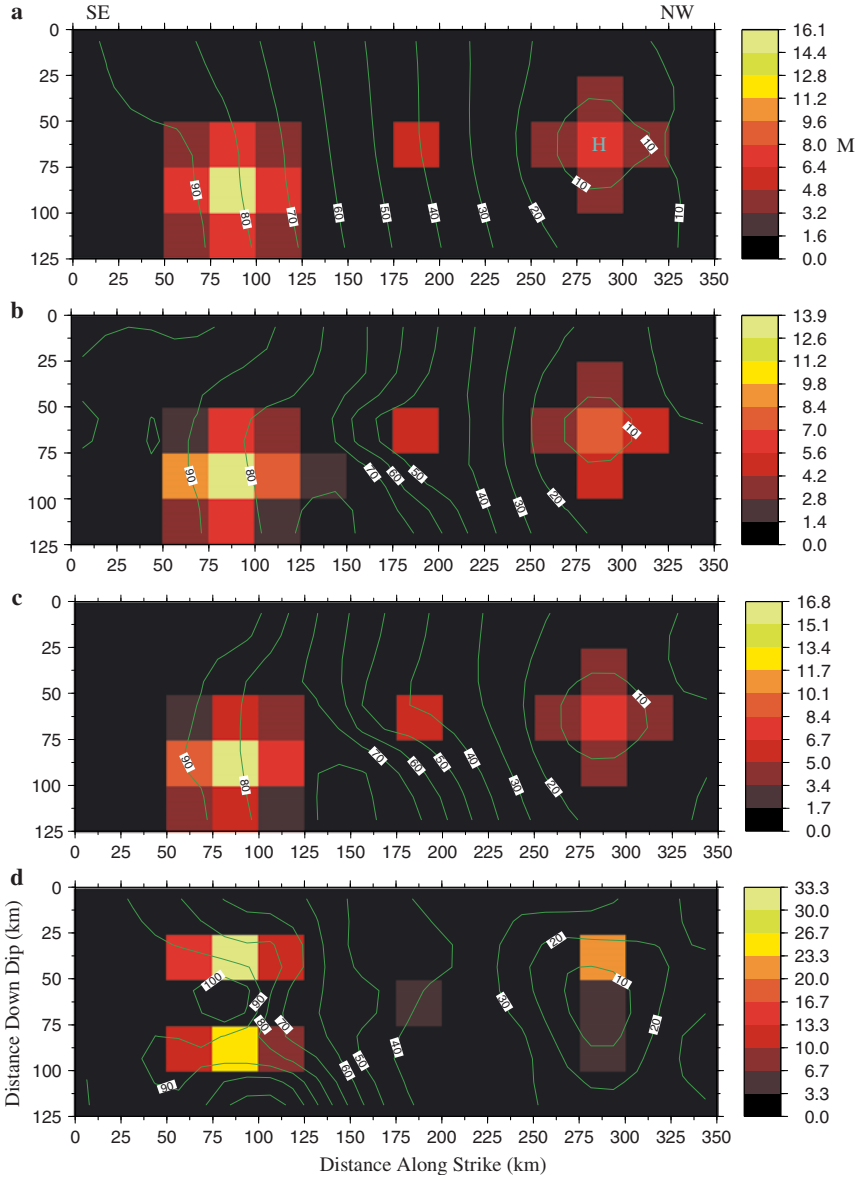


Figure 4

Results of synthetic experiments with moment fixed. These plots show the slip distributions and rupture timing for the original slip model a) original model, and those obtained by the inversions for b) 1 s rise time, c) 3 s rise time, and d) 6 s rise time. Contour interval is 10 s.

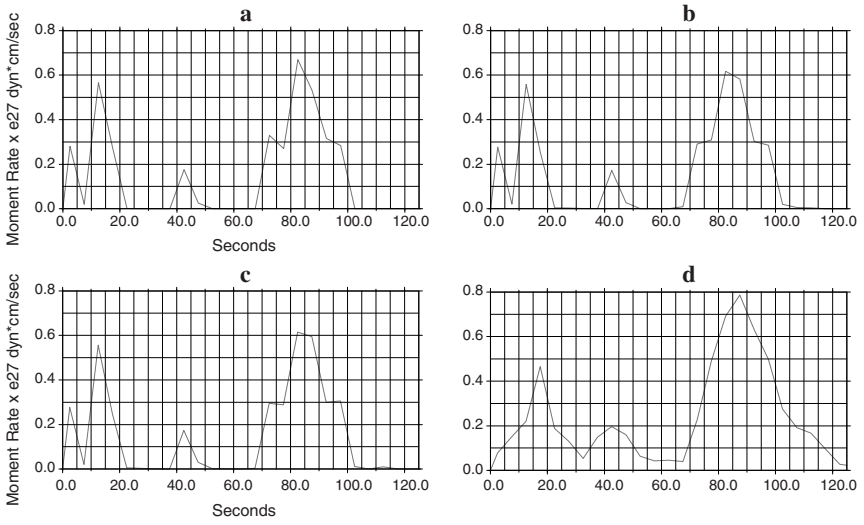


Figure 5
Results of synthetic experiments with moment fixed. These figures show the source time functions for the original slip model, and those obtained by the inversions for b) 1 s rise time, c) 3 s rise time, and d) 6 s rise time.

Talandier, 1988; Weinstein and Okal, 2005), revised the magnitude estimate sharply upwards to 8.1. Furthermore, the value of the energy discriminant Θ, devised by Okal and Newman (1998), was ~ -6.1, suggesting that not only was the earthquake large, but possibly a tsunami earthquake as well. On the basis of the revised magnitude, PTWC issued a Tsunami Watch/Warning Bulletin, warning of the possibility of destructive tsunami waves propagating away from the Peru region. The Watch/Warning was canceled approximately six hours later, when sea-level gauge readings suggested the tsunami waves moving away from Peru posed little threat to human life.

This event posed a couple of puzzles for the warning centers. First, why was the Mwp magnitude so deficient? Mwp can give accurate magnitudes for earthquakes as large as 8.5 (Nias Island earthquake, March 2005). We know now that the moment magnitude for the Peru earthquake was 8.4 (Harvard CMT). Second, despite the indications this was a large, tsunami earthquake, in the end, it did not appear so. Finite fault modeling reveals the reasons why. We note, before presenting our results, that several waveform modeling studies of the Peru earthquake have preceded this one (Giovanni *et al*., 2002; Kikuchi and Yamanaka, 2001; Bilek and Ruff, 2002).

We inverted the first 170 seconds of 14 vertical component broadband waveforms. The waveforms were bandpass-filtered between .01 and 1 Hz, and decimated to a sampling interval of 0.2 s. Displacement seismograms were then obtained by deconvolution. The fault plane was discretized into 70 patches, with dimensions of 25 km on a

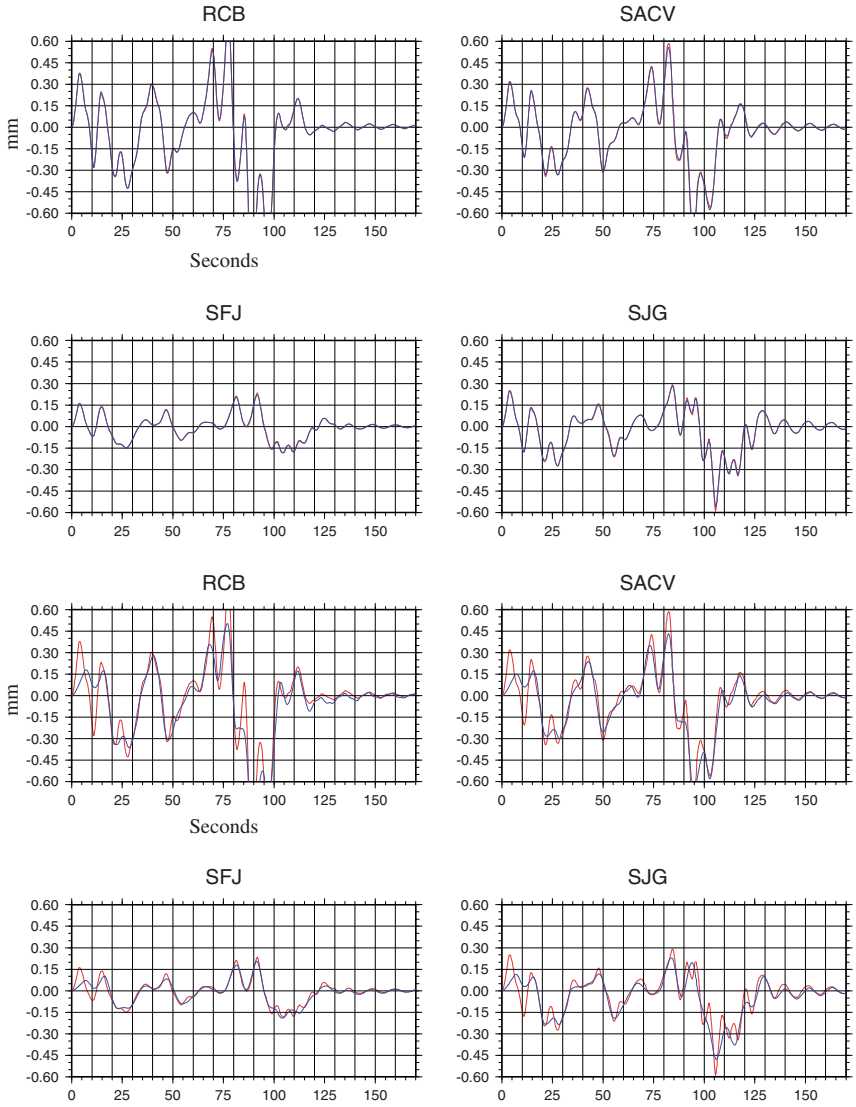


Figure 6

Select waveform comparisons for 3 s and 6 s rise times. The actual waveform and that obtained by the inversion are red and blue, respectively. For both 1 s and 3 s rise times, the actual and inversion waveforms are a close match. For a 6 s rise time, the differences between the actual and inversion waveform are significant. Displacements are in mm.

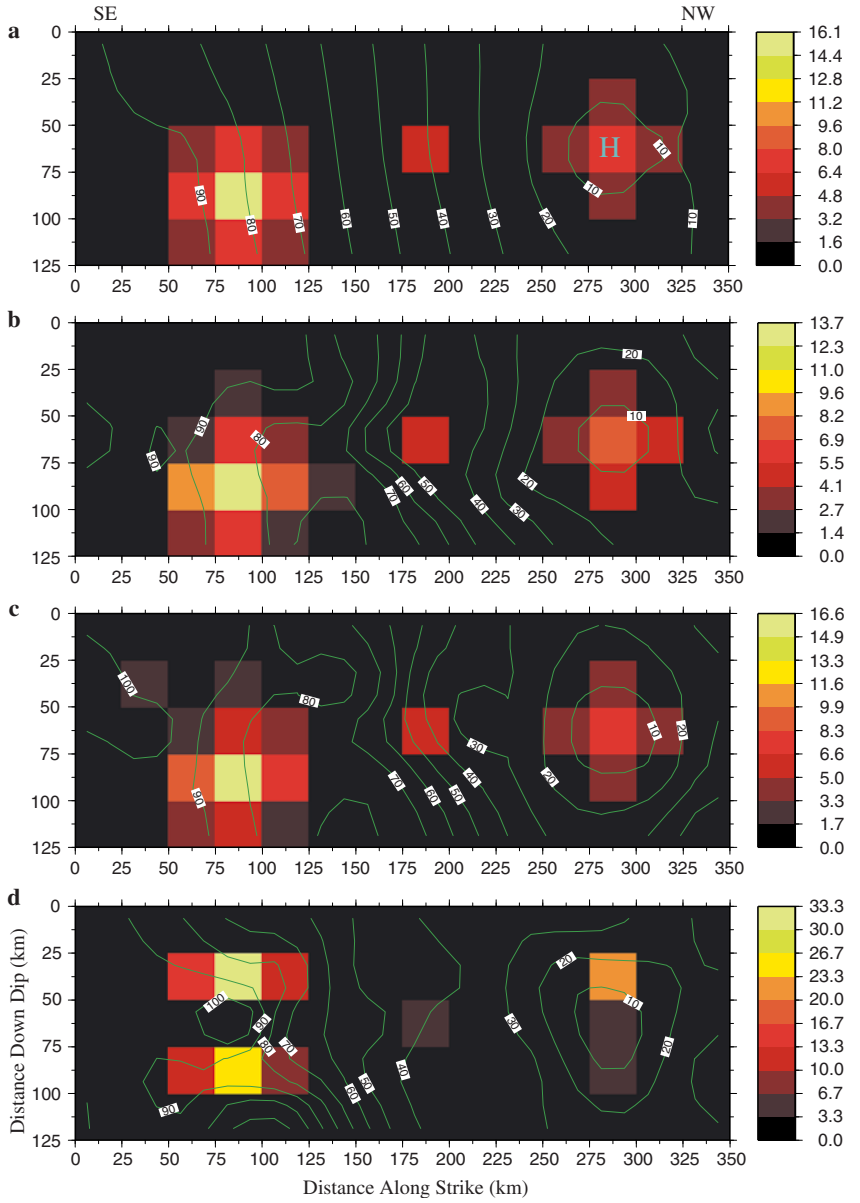


Figure 7
Results of the experiments with moment free. These plots show the slip distributions and rupture timing for the original slip model a) original model, and those obtained by the inversions for b) 1 s rise time, c) 3 s rise time, and d) 6 s rise time. Contour interval is 10 s.

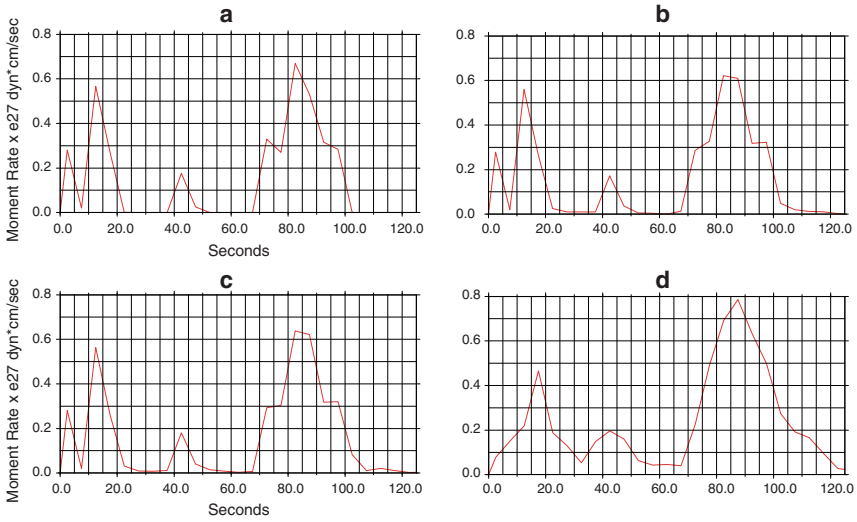


Figure 8
Results of synthetic experiments with moment fixed. These figures show the source time functions for the original slip model, and those obtained by the inversions for b) 1 s rise time, c) 3 s rise time, and d) 6 s rise time.

side organized in 14 columns and 5 rows. Hence the total dimensions of the model fault are 350 km along strike and 125 km down dip. The fault parameters (Harvard) are dip = 18°, rake = 63°, and strike = 289°. We use a location of 16.14°S, 73.3°W and a depth of 23 km for the hypocenter and the same earth structure as GIOVANNI and BECK (2002) (see Table 1.). We performed three inversions using elemental source time function rise times of 1, 3, and 6 seconds. Each inversion required 2317 Green's Functions resulting in a matrix with dimensions of 11900 × 2317.

The results (Figs. 10–12) indicate that the earthquake rupture propagates unilaterally from northwest to southeast and that most of the moment release occurs 70 s–75 s after the earthquake starts and well away (120 km – 160 km) from the hypocenter, as found in previous studies (GIOVANNI *et al.*, 2002; BILEK and RUFF; 2002; KIKUCHI and YAMANAKA, 2001). The moment release determined by the inversions ranges from 17.4 to 18.6 × 10^{28} ergs (M_w ∼ 8.1). These values are comparable to those obtained by GIOVANNI *et al.* (2002), and KIKUCHI and YAMANAKA (2001), and about 1/3 which was found by BILEK and RUFF (2002) (M_w ∼ 8.3). BILEK and RUFF (2002) obtained a better magnitude as a result of using surface waves for their inversions. The largest slip occurs in the broad asperity and is ∼10–15 m. Overall, the slip distributions show little variation until the elemental source time function rise time is 6 s. The event source time functions are similar, with an initial small event near 12–20 s and the maximum moment release occurring at ∼75 s. Table 2. shows CPU times, matrix dimensions and the number of Green's functions actually excited for each computation in this study.

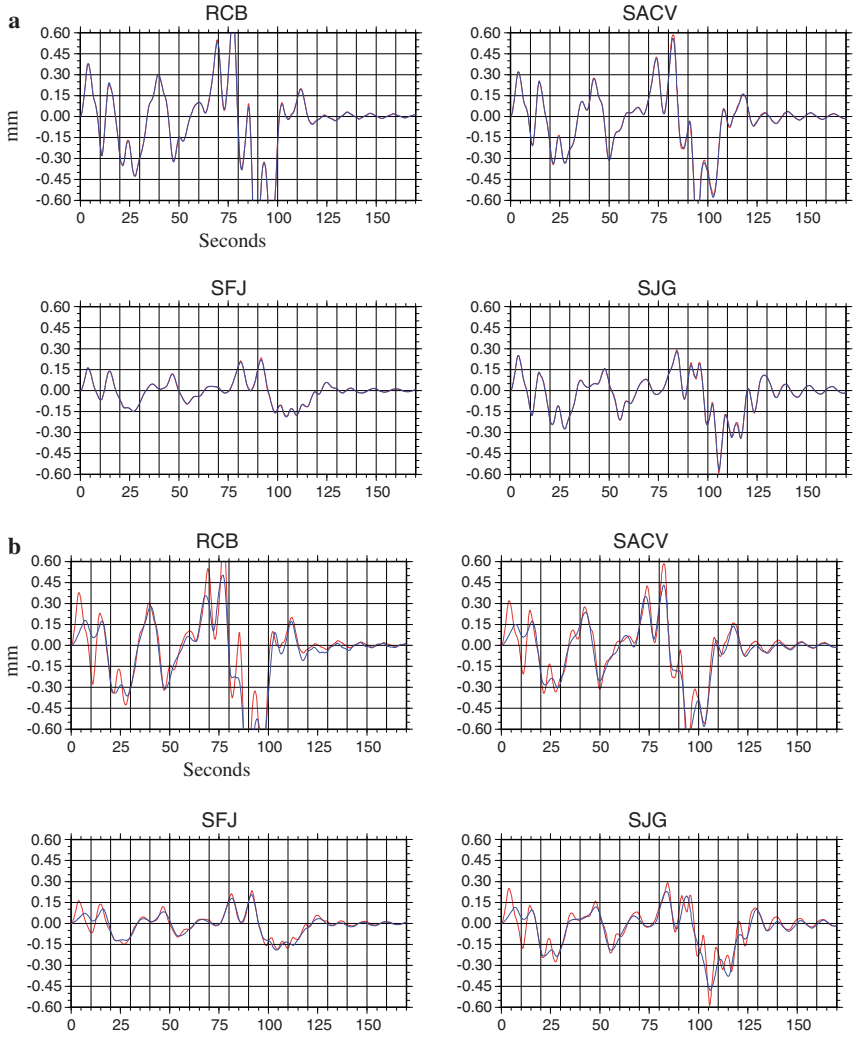


Figure 9
Select waveform comparisons for 3 s and 6 s rise times. The actual waveform and that obtained by the inversion are thin and thick lines, respectively. For both 1 s and 3 s rise times, the actual and inversion waveforms are a close match. For a 6 s rise time, the differences between the actual and inversion waveform are significant.

The waveform fits (shown for all stations in Fig. 12) seem to be as good as those obtained by Giovanni *et al*. (2002) and Kikuchi and Yamanaka (2001). We were unable to produce good fits to the HOPE waveform. The phases at 125 s for RCB, SACV,

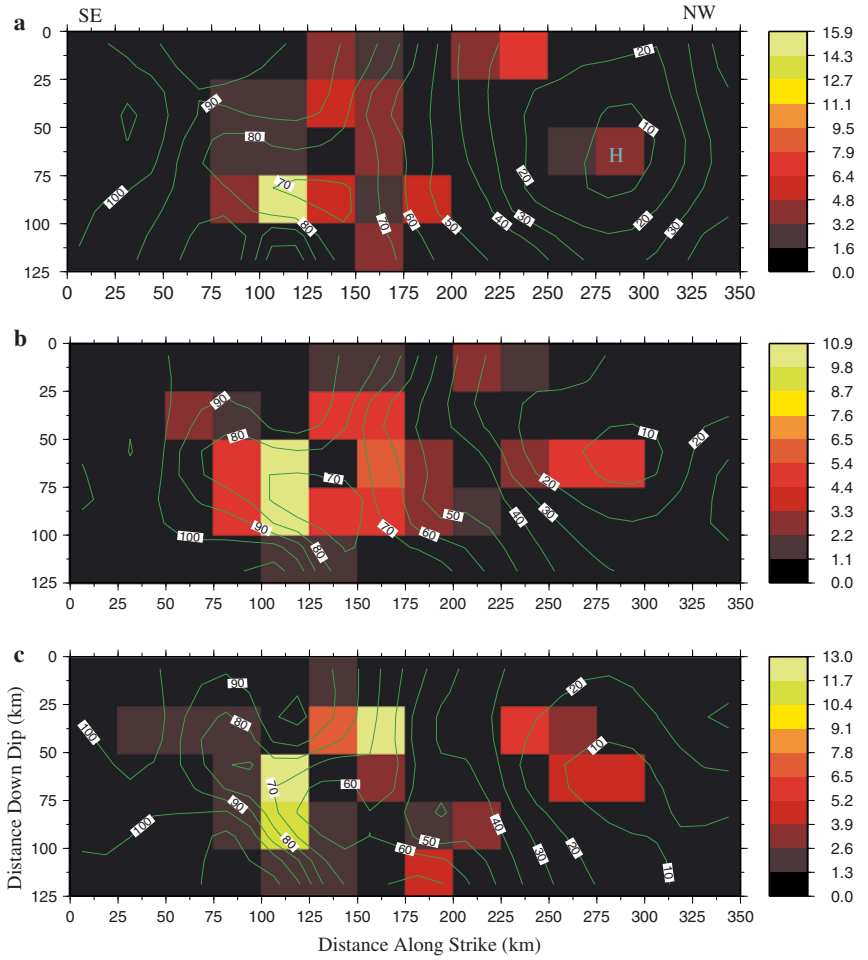


Figure 10
Slip distribution and rupture timing obtained for the Peru earthquake. The slip distributions and rupture timing obtained for rise times of 1 s, 3 s, and 6 s are shown in a, b, and c, respectively. Contour interval is 10 s.

HOPE, and at 150 s for SJG are not fit by any inversion. These phases do not appear to be related to pP, sP or even PP. We echo the suggestion of Giovanni *et al*. (2002) that these phases are perhaps the result of three-dimensional velocity structure in the vicinity of the Peru 2001 earthquake.

The finite fault modeling results reveal why Mwp was deficient and why Peru was not a slow earthquake. The complexity of this event caused Mwp to be deficient because in

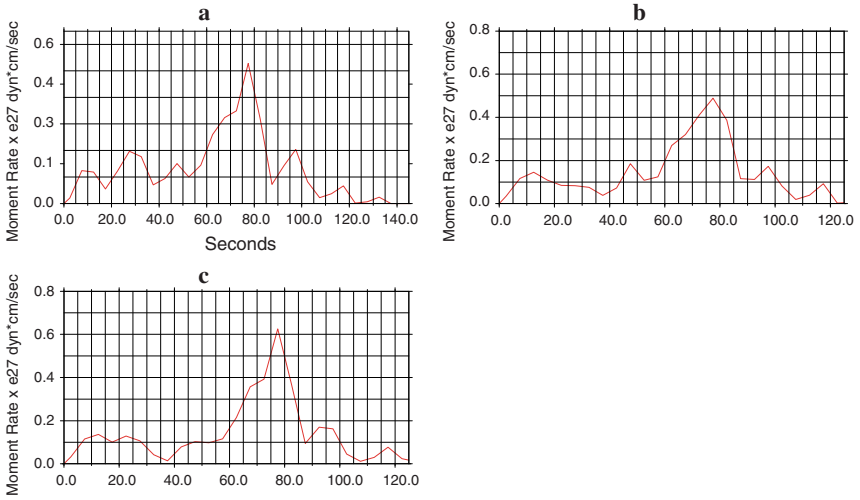


Figure 11
Source time functions obtained from Peru earthquake data for a) 1 s rise time, b) 3 s rise time, and c) 6 s rise time. All show that the bulk of the moment release occurred ~75 s after the earthquake started.

this method, only the first 100 seconds of P-wave coda are used. As a result, the Mwp method used P-wave coda that was primarily generated by the small precursor event that occurred between 12 and 25 seconds and not P waves generated by the rupturing of the large asperity that occurred ~60 seconds afterwards. Clearly, having these results during the event would have benefited the warning centers and given them a much better understanding of this earthquake.

3.2. Java Earthquake of July 17, 2006

On July 17, 2006 an earthquake occurred off the coast of Western Java generating a destructive tsunami that killed several hundred people. The initial estimate of the magnitude by PTWC was a 7.2, based on the Mwp method. PTWC issued a Local Tsunami Watch Bulletin for this earthquake to the Indian Ocean region 17 minutes after the earthquake. This bulletin warned of the possibility of destructive waves usually no more than 100 km from the epicenter. A second bulletin was issued 2.5 hours later reporting a small tsunami observed on a sea-level gauge in Benoa, Indonesia. Authorities in Indonesia also were alerted to the earthquake. Unfortunately no evacuation was ordered. Under ordinary circumstances, an earthquake this large near the coast would probably be felt over a wide area, yet many residents in closest proximity to the rupture did not report a strong shaking (WIDJO *et al*., 2006).

Using later arriving mantle waves, PTWC found $M_w \sim 7.4$ using the mantle magnitude method. Computations of Θ suggested a trend towards slowness, but that the

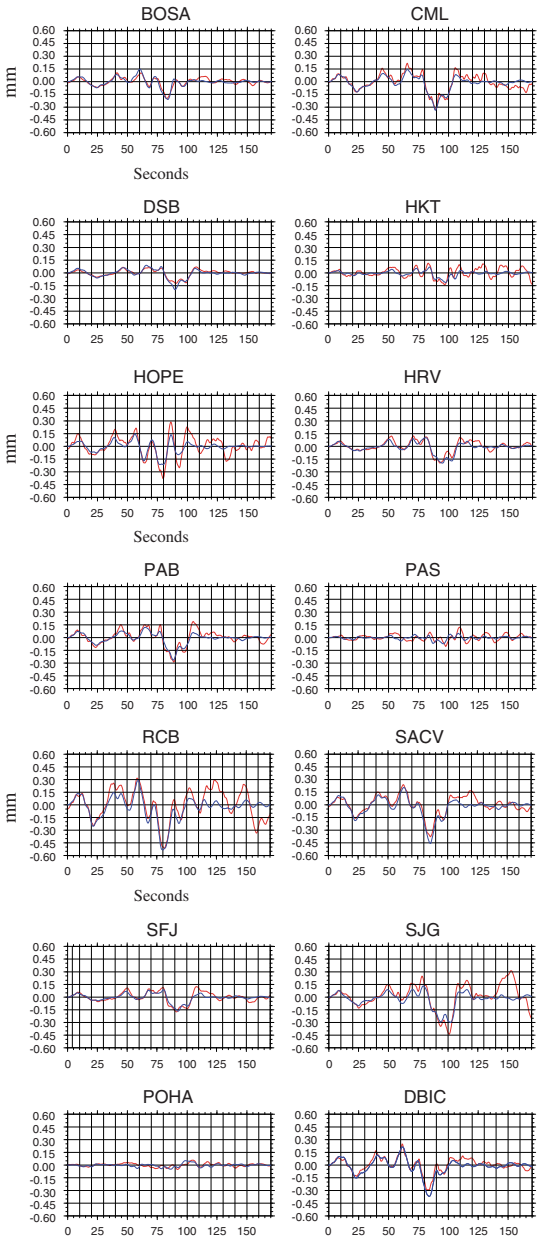


Figure 12
This figure shows the comparison between the actual waveforms (red) to those obtained by the inversion (blue) with a 3 s rise time.

Table 2

CPU times and number of excited Green's functions

Rise time	CPU time	Green's functions excited
1 s	290 s	375
3 s	175.0	161
6 s	102.0	69

energy radiated as body waves was not deficient enough to be considered a tsunami earthquake. It was later learned that M_w for this event was 7.7. Recalculating Θ with this value gave –6.1, a value indicative of tsunami earthquakes (OKAL and NEWMAN, 1998).

Using our technique, we have inverted 14 waveforms generated by the Java earthquake. The basic outline of the calculations is the same as Peru. The fault parameters (Harvard) are dip = 10°, rake = 95°, and strike = 289°. We use a location of 9.3 S, 107.3E and a depth of 15 km for the hypocenter. In this case we use 180 s of seismogram that was band pass-filtered between .01 Hz and 1 Hz, and decimated to 5 Hz. The rupture speed is constrained to be between 1.0 <= Vr <= 2.0 and the earth structure used is given in Table 3. We performed three inversions with elemental source time function rise times of 1 s, 3 s, and 6 s. We consider a larger fault model than the Peru model with 15 × 8 blocks with dimensions of 25 km on a side. The number of Green's functions required in each inversion is 5746, resulting in a matrix with dimensions of 12600 × 5746. Table 4. sets forth the CPU times, matrix dimensions and the number of Green's functions actually excited for each computation in the Java case.

The results, presented in (Figs. 13–15), indicate that the largest slip tends to occur in the patch containing the hypocenter. The results are broadly similar to those obtained by CHEN (2006). The earthquake rupture propagation is primarily unilateral to the East. However, our solutions show a larger bilateral component than is found by CHEN (2006). This is possibly a result of our only using P-wave teleseismic waveforms and our station distribution. At azimuths to the West and South, stations like MSEY are close to the CMT nodal plane for this earthquake. CHEN (2006) uses S waves as well (and more stations too) providing better constraints on the rupturing direction. As a result, in our calculations, the

Table 3

Java earthquake structure model

Vp km/s	Vs km/s	ρ kg/cm^3	Thickness (km)
Source			
1.5	0.0	1.0	4.0
6.5	3.7	2.8	23.0
8.1	4.64	3.2	
Receiver			
6.5	3.74	2.87	35.0
8.1	4.64	3.2	

Table 4

(Please provide caption)

Rise time	CPU time	Green's functions excited
1 s	938.0 s	638
3 s	374.0 s	261
6 s	228.0 s	116

timing constraints along the basically East-West striking fault may not be as strong as needed. Despite the limitations, the inversions yield useful results.

We used a value for the shear rigidity of 3.0×10^{10} Pa in computing the slips. However, given the slowness of the rupture, the appropriate value of the shear rigidity for the Java event may be considerably lower and therefore the actual slips along the Java fault may well have been larger than indicated in Figure 16a.

The source time function obtained by the inversion is what one would expect for a slow earthquake. The source time function for all three inversions shows a long, drawn-out, steady, rate of energy release. Even though our method is not likely to capture the total source process in this case, the inversions still yield an M_w of 7.6, very close to Harvard's CMT result with $M_w = 7.7$.

These earthquakes and our computations illustrate the potential of this technique for the tsunami warning centers. With results like those obtained above, we can clearly distinguish the difference between earthquakes that are simply "late" like Peru from those that are truly slow. Furthermore, we can use these more accurate descriptions of the slip distributions to drive tsunami propagation models.

4. Boot Strap Results

One problem with the results of waveform modeling in studies of this kind is a lack of understanding of the uncertainties inherent in the inversion results. We propose that one way of obtaining estimates of the uncertainties is through the boot-strap method (EFRON, 1984). The idea behind this method is to create numerous synthetic data sets from the real data, invert them, and use the statistics obtained by analyzing the resulting solutions as an estimater of the uncertainties. We have done this for the Peru case, since the calculations are somewhat quicker.

Given that we have 14 seismograms, we harness a random number generator to generate random numbers with integer values from 1 to 14. Each integer represents a seismogram from a particular station. 1000 sets of 14 random numbers are used to construct 1000 sets of 14 seismograms. In each set, seismograms from some stations may be included more than once, while others are absent. The calculations are conducted in the same manner as before, and with an elemental source time function of 3 s.The results of the 1000 inversions are summarized in Figure 16.

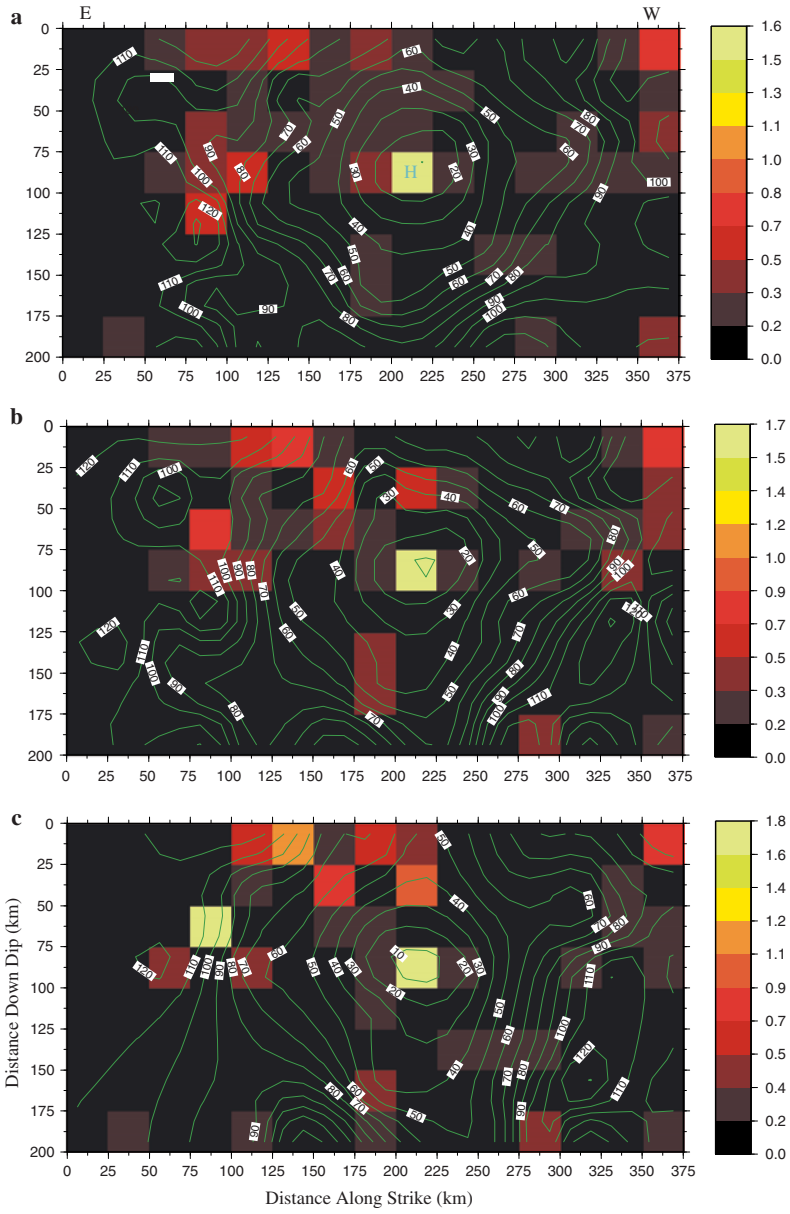


Figure 13
Slip distribution and rupture timing obtained for the Java earthquake. The slip distributions and rupture timing obtained for rise times of 1 s, 3 s, and 6 s are shown in a, b, and c, respectively. Contour interval is 10 s.

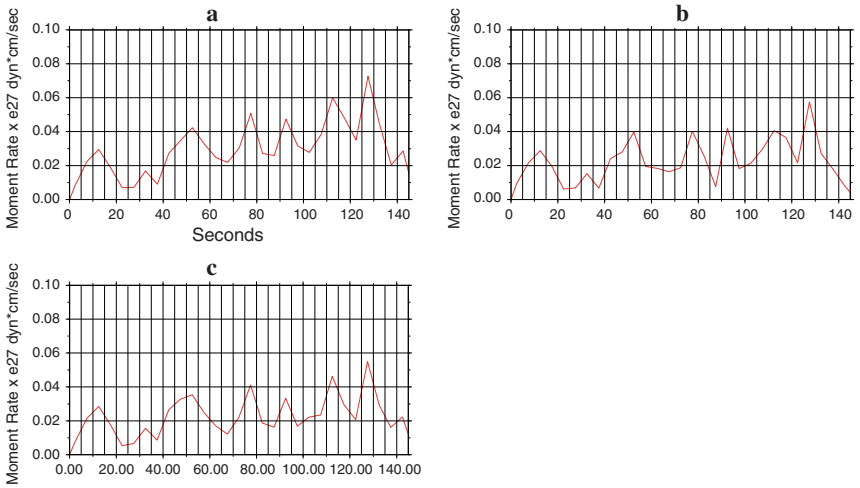


Figure 14
Source time functions obtained from Java earthquake data for a) 1 s rise time, b) 3 s rise time, and c) 6 s rise time. All plots show the slow, steady release of energy over 140 s.

The mean M_w for our distribution of results is $19.373 \pm 2.684 \times 10^{21}$ N-m. In comparison to the result obtained with just the single data set, the mean boot strapped finite fault model solution (Fig. 16a) is smoother although the outline of the slip distribution is basically the same. Note that a block in the left-most patch that has no slip in the single data-set solutions, has slip in the bootstrap solution. The rupture timing is also very similar. Figure 16b shows the corresponding 2σ errors in the slip and in the rupture timing. The estimated uncertainty in the slips is typically $\sim 35\%$, however, the uncertainty in timing is considerably better, on the order of 1%–10%. The uncertainty in timing generally increases with distance from the epicenter. This is perhaps not too surprising since other phases not accounted for by the Green's function and the time interval over which rupture is permitted increase with distance from the epicenter. The uncertainty in the recovered source time function is shown in Figure 16c. Here the uncertainty tends to increase with distance, for the same reason as with the rupture timing. However, the results of the bootstrap method suggest that the overall shape of the source time function is well constrained. In summary, the results of the bootstrap method suggest that the rupture timing is substantially better constrained than the spatial distribution of the moment release.

5. Conclusions

We have examined the behavior of a seismic waveform inverse method in an effort to evaluate the procedure's application to tsunami warning and the improvement of tsunami

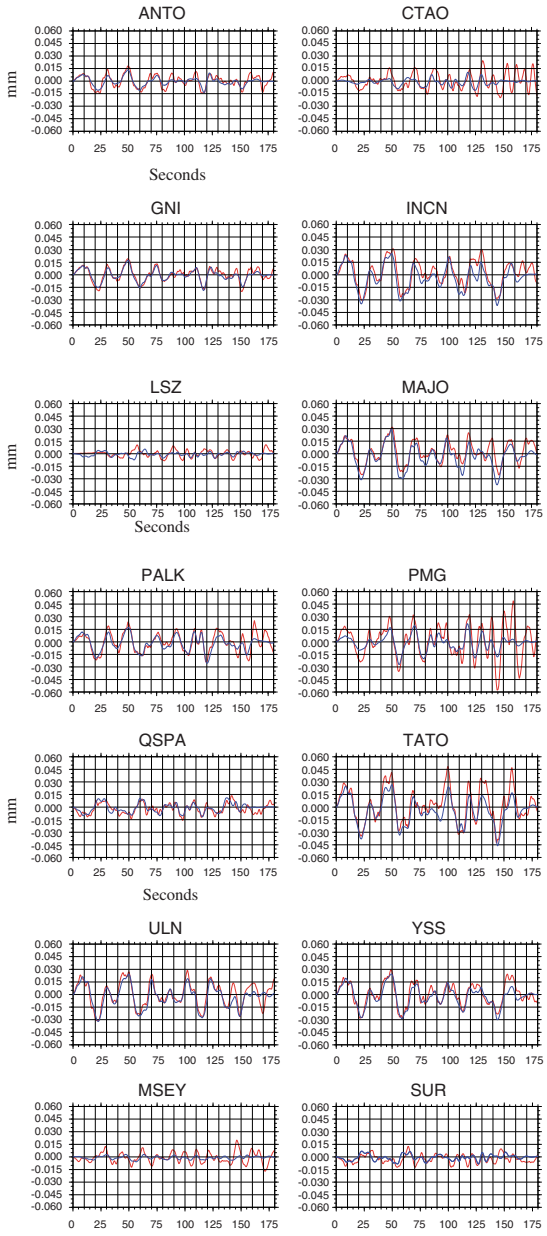


Figure 15
This figure shows the comparison between the actual waveforms (red) to those obtained by the inversion (blue) with a 3 s rise time.

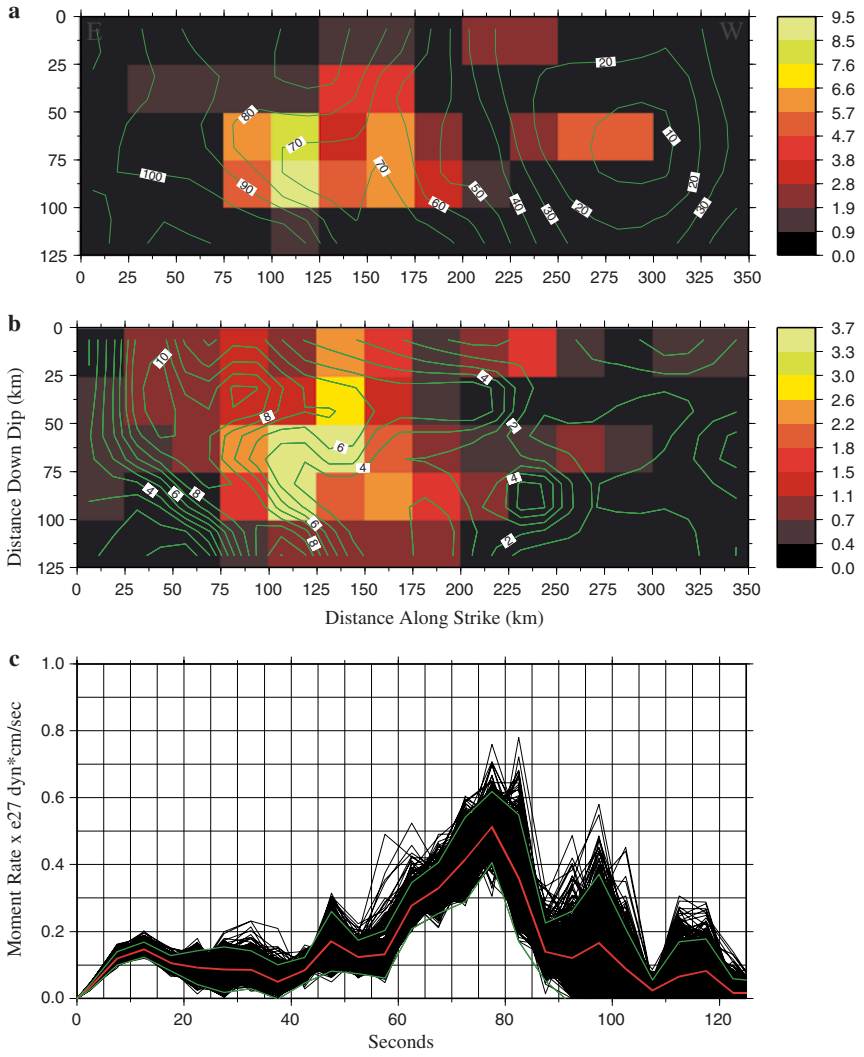


Figure 16
Results of the boot strap experiments for the Peru earthquake, a) average slip and rupture timing, b) slip and rupture timing 95% confidence intervals (contours interval is 1 s), c) each bootstrap result (black), mean (red), and 95% confidence limits (green) for the source time function.

wave height forecasts. Our method using just 14 broadbands yields a useful understanding of the rupture history of large earthquakes in a timely enough manner to be useful in a tsunami warning context

Our results for the Peru, June 23, 2001 and Java, July 17, 2006 earthquakes are broadly consistent with previous studies. We find that the Peru earthquake, despite early indications, was not a slow earthquake, but "late," with the main energy release occurring 75 seconds after the earthquake started. An analysis of the rupture history provided by this inversion very soon after the earthquake would have given tsunami warning center personnel a clearer understanding of the nature of this earthquake than they had on the basis of scalar measures of the earthquake source.

The results of the inversions for the Java event, as do previous studies, indicate this earthquake was a "slow" or a "tsunami earthquake." The rupture speed for this earthquake was only slightly greater than 1 km/s, and the moment determined by the inversion is 3.15×10^{21} N-m corresponding to $M_w = 7.6$. The source time function for this event shows a slow, but steady release of energy over at least 140 s. The results of our inversions reveal an interesting contrast between an earthquake (Peru) that initially looked "slow," and an earthquake that actually was "slow" (Java). An understanding of the rupture history will greatly add to a tsunami warning center watchstander's understanding of the properties of large earthquakes in near real-time, and improve tsunami warnings and tsunami wave height forecasts.

Finally, in an effort to better understand the properties of our inversions, we have used the bootstrap method to create 1000 synthetic waveform data sets. A statistical analysis of the results of the inversions of all of these data sets yields an understanding of what our numerical method is and is not resolving. We find that the rupture timing itself seems to be well constrained, however the moment release history is less so. This is primarily because the uncertainties in the slips can be as large or larger than 35%. Furthermore, the averaged results remove some of the disjointedness of the single data set inversions, and give what we feel is a more intuitive view of the earthquake.

Although it took $\sim$50 hours to complete 1000 inversions, we note that this was accomplished using only a single SUN U40 processor. PTWC will soon have at its disposal 40 processors from 10 individual DELL precision workstations connected together on a single switch. Hence, 1000 inversions could be accomplished in $\sim$1 hour, still more than quick enough to revise tsunami forecasts in the event of an ongoing tsunami warning.

Acknowledgements

The authors are grateful to NOAA/PTWC and JPL for the use of their facilities. The authors gratefully acknowledge support from NASA under the Pacific Disaster Center section of the Solid Earth and Natural Hazards program, grant SENH 99-0272-0109. SAW gratefully acknowledges the assistance of Melissa Giovanni with data and sac scripts.

References

Bilek, S.L. and Ruff, L.J. (2002), *Analysis of the 23 June 2001 $M_w = 8.4$ Peru underthrusting earthquake and its aftershocks*, Geophys. Res. Lett. *29*, 21–1, 21–4.

Bilek, S.L. and Lay, T. (1999), *Rigidity variations with depth along interplate megathrust faults in subduction zone*, Nature *400*, 443–446.

Centroid Moment Tensor (CMT) Catalog, www.seismology.harvard.edu/CMT/search.html.

Chen, J. (2006), *Preliminary Result of the 2006 July 17 Magnitude 7.7 – South of Java Indonesia Earthquake*, http://neic.usgs.gov/neis/eq_depot/2006/eq_060717_qgaf/neic_qgaf_ff.html.

Efron (1984).

Giovanni, M.K., Beck, S.L., and Wagner, L. (2002), *The June 23, 2001 Peru earthquake and the southern Peru subduction zone*, Geophys. Res. Lett. *29*, 14–1 – 14–4.

Hartzell, S. and Heaton, T. (1983), *Inversion of strong ground motion and teleseismic waveform data of the fault history of the 1979 Imperial valley, California, earthquake*, Bull. Seismol. Soc. Am. *73*, 1553–1583.

Hartzell, S. and Heaton, T. (1986), *Rupture history of the 1984 Morgan Hill, California, Earthquake from the inversion of strong motion records*, Bull. Seismol. Soc. Am. *76*, 649–674.

Hartzell, S. and Mendoza, C. (1991), *Application of an iterative least-squares waveform inversion of strong-motion and teleseismic records to the 1978 Tabas, Iran earthquake*, Bull. Seismol. Soc. Am. *81*, 305–331.

Kanamori, H. (1972), *The mechanism of tsunami earthquakes*, Phys. Earth Planet. Interiors *6*, 346–359.

Kanamori, H. and Kikuchi, M. (1993), *The 1992 Nicaragua earthquake: A slow tsunami earthquake associated with subducted sediment*, Nature, *361*, 714–715.

Kikuchi, M., and Fukao, Y. (1985), *Iterative deconvolution of complex body waves – The Tokachi-Oki earthquake of 1968*, Phys. Earth Planet. Int. *37*, 235–248.

Kikuchi, M. and Yamanaka, Y. (2001), *EIC Seismological Note Number 105*, www.eic.eri-u-tokyo.ac.jp/EIC/EIC_news/105E.html.

Lawson, C.L. and Hanson, R.J., *Solving Least-Squares Problems* (SIAM, Philadelphia 1995).

Mendoza, C. (1996), *Rapid derivation of rupture history for large earthquakes*, Seismol. Res. Lett. *67*, 19–26.

Okal, E.A. and Talandier, J. (1989), *M_m: A variable period mantle magnitude*, Pure Appl. Geophys. *134*, 355–384.

Okal, E.A. and Newman, A.V. (1998), *Teleseismic estimates of radiated energy: The E/M_O discriminant for tsunami earthquakes*, J. Geophys. Res. *103*, 26,885–26,898.

Tsuboi, S., Abe, K., Takano, K., and Yamanaka, Y. (1995), *Rapid Determination of Mw from Broadband P Waveforms*, Bull. Seismol. Soc. Am. *85*, 606–613.

Wald, D.J. and Heaton, T.H. (1994) Spatial and Temporal Distribution of Slip for the 1992 Landers, California, Earthquake, Bull. Seismol. Soc. Am. 84, 668–691.

Wald, D.J., Helmberger, D.V., and Hartzell, S.H. (1990), *Rupture process of the 1987 Superstition Hills earthquake from the inversion of strong-motion data*, Bull. Seismol. Soc. Am. *80*, 1,079–1,098.

Weinstein, S.A. and Okal, E.A. (2005) *The mantle magnitude M_m and the slowness parameter (Θ: Five years of real-time use in the context of tsunami warning*, Bull. Sesimol. Soc. Am. *95*, 779–799.

Widjo, K. *et al*. (2006), *Rapid Survey on Tsunami Java 17 July, 2006*, http://ioc3.unesco.org/itic/files.php?action=viewfile&fid=443&fcat_id=287.

(Received October 4, 2006, revised June 4, 2007, accepted June 8, 2007)

Pure appl. geophys. 165 (2008) 475–496
0033–4553/08/030475–22
DOI 10.1007/s00024-008-0324-x

Pure and Applied Geophysics

Modeling and Visualization of Tsunamis

HUAI ZHANG,[1] YAOLIN SHI,[1] DAVID A. YUEN,[1,2] ZHENZHEN YAN,[1] XIAORU YUAN,[3] and CHAOFAN ZHANG[1]

Abstract—Modeling tsunami wave propagation is a very challenging numerical task, because it involves many facets: Such as the formation of various types of waves and the impingement of these waves on the coast. We will discuss the different levels of approximations made in numerical modeling of 2-D and 3-D tsunami waves and their relative difficulties. In this paper new attempts are proposed to evaluate the hazards of tsunami's and visualization of large-scale numerical results generated from tsunami simulations. Specialized low-level computer language, based on a parallel computing environment, is also employed here for generating FORTRAN source code for finite elements. This code can then be run very efficiently in parallel on distributed computing systems. We will also discuss the need to study tsunami waves with modern software and visualization hardware.

Key words: Tsunami, wave propagation, parallel simulation environment, visualization.

1. Introduction

Following the great Sumatran earthquake on December 26, 2004, the Indian Ocean tsunami and the accompanying tsunami waves caused widespread damage and killed more than 225,000 people within a few hours and left millions of people homeless. This event has indeed awakened great scientific interest in tsunami wave propagation over undulated seafloor topography, and along irregular coastlines. Traditional analytical approximations are valid over long wavelengths in the far field. This can be used as a first measure for tsunami prediction and warning (http://tsunami.jrc.it/model/model.asp). But for near-field regions with complex geography and other complications, such as islands and harbors, "high resolution" numerical simulation must be employed to obtain accurate predictions in both space and time. Presently using 10 million to 100 million grid points becomes commonplace with improved dual-core laptops and also massively parallel computers with access to huge data and high-speed I/O support. Besides tsunamis,

[1] Laboratory of Computational Geodynamics, Graduate University of Chinese Academy of Sciences, Beijing 100049, China. E-mail: hzhang@gucas.ac.cn

[2] Department of Geology and Geophysics, University of Minnesota, Minneapolis, MN 55455, U.S.A.

[3] State Key Laboratory of Machine Perception, and Dept. of Machine Intelligence, School of EECS, Peking University, Beijing 100871, China.

turbulent river discharges from upstream events and tall waves driven by hurricanes or by huge tankers, will also cause severe damage to dams and the foundation of mountain slopes. This aspect is of societal relevance, especially the Three Gorges project in central China along the Yangtze River.

Although the frequency of earthquake-generated tsunamis around the globe is relatively low compared to many other natural hazards, such as earthquakes, volcanoes, and hurricanes, the terrible Sumatran tsunami event was still an unforgettable reminder that the damaging impacts of tsunamis may remain extremely high in human history. Especially, with the booming of the population along coastal regions in recent years around the world, this type of shocking disaster will pose even greater risk than ever before (TIBBETTS, 2002). Tsunami propagation is thus a problem with global dimensions, knowing no international boundaries across the sea. There is currently a great need for understanding better tsunami wave propagation, which calls for comparison of simulations with detailed data from observations. Fortunately, fast developments of geographical information systems (GIS), the global positioning system (GPS), remote sensing techniques, such as Interferometric Synthetic Aperture Radar (InSAR) (CHANG, *et al*., 2005; NAEIJE, *et al*., 2002) and other modern observational technologies, enable scientists in the research fields of tsunami sciences to obtain daily huge amounts of data. These data, together with numerical results, can help researchers to better understand this deadly natural disaster. The Sumatran tsunami was without doubt the best documented case in history (TITOV *et al*., 2005). From videos of the run-up processes to direct satellite observations of the waves propagating in the far field, research scientists now have an unprecedented opportunity to study these catastrophes (http://www.asiantsunamivideos.com/). One important task facing the earth science community is to develop reliable easy-to-use software tools for facile modeling and visualization of tsunamis.

The objective of tsunami modeling research is now focused on developing numerical models for more accelerated and more reliable forecasting of tsunamis propagating through vast oceans before they strike the coastlines (MEINIG *et al*., 2005; TITOV and Gonzalez 1997, TITOV *et al*., 1999; GONZALEZ *et al*., 1995; MOFJELD *et al*., 2000). Some models can easily be satisfied by two-dimensional shallow water equations, while other models use slightly modified Navier-Stokes equations, which can enable the researcher to proceed beyond just the first-order physical phenomena (GILL 1982; PEDLOSKY 1987). We simply arrange these models in Figure 1.

Due to the alteration of ray patterns over complex bathymetry, tsunamis can be significantly modified while they are propagating over transoceanic areas, leading to the alteration of wave fronts and wave groups, frequently dispersion effects, and changes in spatial distribution of wave energy. Under such circumstances, Boussinesq approximation and Boussinesq equations are well known for their descriptions of such phenomena (KENNEDY and KIRBY, 2003). Boussinesq equations are obtained from the Euler equations with rotation, which include the effects of weak dispersion and nonlinearity in a shallow water framework and allow accurate near-shore simulation of wave transformation processes. Up to date, the extended Boussinesq equation systems allow the models to be

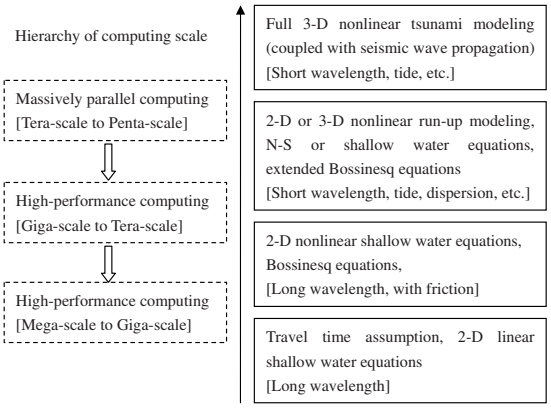


Figure 1
Hierarchy of tsunami models and computing scales.

applied in deeper water over relatively narrow and complex bathymetry so as to extend the range of applications, as well as increasing the accuracy of the linear dispersion characteristics of these models (WALKLEY and BERZINS, 2002). Parameters can be introduced to characterize horizontal wave packet length scale and aspect ratio which can describe dispersion effects. Those extended Boussinesq equations would be more appropriate to local wave evolution both near the tsunami source and in the final run-up stage. Those effects include representation of bottom motion, sea-bed friction and fully nonlinear treatment of surface conditions in order to represent large run-up amplitudes during inundation which can all be modeled and simulated by extended Boussinesq equation systems via retaining several aspects of parameterized formulations (KIRBY and DALRYMPLE, 1986; KENNEDY *et al*., 2000; CHEN *et al*., 2000).

There are several computational issues worthy of consideration. They deal with wave propagation over both short distances (near field) and long distances (far field). To date, most tsunami simulations have been carried out in two dimensions with the latitude and longitude being the independent variables. Three-dimensional simulations of tsunami waves including run-up, remain a grand-challenge problem because of the multiple-scale nature of the phenomenon (GICA and TENG, 2003; TITOV *et al*., 2005). Three-dimensional equations cannot be employed to solve real-time problems, due to the still inordinately long computational time.

Two-dimensional equations are more commonly used and can be used in places where warning must be issued in a timely fashion (TANG *et al*., 2006; GEIST *et al*., 2006; SMITH *et al*., 2005). Within the framework of two-dimensional tsunami equations, there are linear and nonlinear approximations with the linear shallow-water equations being the most popularly employed, since they are the simplest to implement and provide reliable answers regarding the travel time of tsunami waves in the far field. There are also

near-field and far-field regimes for the nonlinear regime. One must take the Coriolis force into account in the far field for wave propagation across the wide ocean (TITOV *et al.*, 2003). The advantage of the linear theory is that it allows one to explore the parameter space in earthquake faulting parameters and the spatial dependence of the impinging wave height along the coast on the earthquake faulting parameters.

A typical calculation for a 2000 $\times$ 2000 grid point configuration, using the shallow-water linear equations, takes around a few hours on a dual-core 2.3 GHz laptop. The elapsed time of the wave-propagation across a regional extent is about the same as the wall-clock time of the computer simulation.

In properly simulating a run-up process, which is part of the phenomenon that directly impacts society, one would need at least the two-dimensional nonlinear equations and better yet, 3-D nonlinear equations, which is one of the focuses in this paper (shown in Table 1). This is a challenging problem, as it involves very careful implementation of numerical schemes using the actual bottom topography. This procedure is also very expensive computationally and requires massively parallel computing with tens of processors to accomplish 3-D simulations on the order of a few days. We summarize the hierarchy of tsunami simulations in Figure 1, where we classify the ease of computation with the level of mathematical approximations of the tsunami equations of motion. They span from fully 3-D Navier-Stokes equations to the linear two-dimensional shallow-water

Table 1

Shows the hierarchy of tsunami numerical research in recent years

Model	Data Needed	Model Name and Reference
1-D equation, Travel time assumption	Topography, Earthquake Magnitude	JRC http://tsunami.jrc.it/
2-D equations	Topography	INGV
Shallow water theory, Finite difference	Fault Parameters, Earthquake Mechanism	http://www.ingv.it/
2-D Shallow water equations, linear and nonlinear wave propagation, Leap and frog finite-difference schemed	Topography, Fault Parameters, Earthquake Mechanism	TSUNAMI, http://www.tsunami.civil.tohoku.ac.jp/c-indexe.html
2-D Shallow equations, tsunami generation, propagation and inundation modeling; Extended Boussinesq equations	Topography, Fault Parameters, Earthquake Mechanism	MOST http://nctr.pmel.noaa.gov/model.html FUNWAVE, WAVESIM and etc.
2-D/3-D modeling, Finite differences, finite element, finite volume	Topography, Initial Wave Turbulence	Delft3D http://www.wldelft.nl/soft/d3d/intro/index.html
2-D/3-D modeling, Finite elements	Topography Fault Parameters, Earthquake Mechanism	Fastflo http://www.cmis.csiro.au/fastflo/
2-D/3-D Smoothed Particle Hydrodynamics modeling	Topography	SPH http://www.cmis.csiro.au/cfd/sph/index.html

equations, with the 2-D nonlinear equations in between. Table 1 shows some of tsunami numerical models developed in recent years. We also try to summarize certain efforts taken including those from one-dimensional empirical equations to full three-dimensional strongly coupled models. Although there are still other important models, and numerical methods are not included in this table, we emphasize here that using the full 3-D Navier-Stokes equations to simulate tsunami hazards needs to be emphasized in future research.

In section 2 we will lay out the shallow-water equations in both linear and nonlinear formats. Next we will discuss the construction of parallel numerical codes used for solving tsunami equations with new techniques from software engineering. In section 4 discuss the preparation of the topography data needed for the numerical simulation. In section 5 we discuss the numerical solution of the 3-D set of tsunami equations. In section 6 we show the results with an emphasis on current visualization techniques. Finally, in section 7 we give a summary and future perspectives.

2. Shallow-Water Equations

Physical modeling of tsunami wave propagation is a difficult and complex task. A full description and simulation require the use of proper numerical algorithms and corresponding reliable software run on parallel supercomputers (MAJDA 2003; ARBIC *et al*., 2004). This is far too time-consuming and not feasible for most real-time applications of tsunami warning, which needed to be precomputed, however. The simplified theory of tsunami waves that reasonably approximates the realistic behavior of ocean waves over vast open sea is the coupled partial differential equations known as the shallow-water equations (LAYTON and VAN DE PANNE, 2002; PELINOVSKY *et al*., 2001). Basically, this is nonlinear and satisfies not only the far-field but also the near-field tsunami propagation.

The shallow water equations are derived with the fundamental scaling parameter δ, which is relevant to the tsunami wavefield, i.e., water depth over wavelength,

$$\delta = \frac{D}{L} \ll 1. \tag{1}$$

here D is the vertical scale and L is the horizontal scale. With this condition, the 3-D equations can be reduced to 2-D and not pose a fundamental problem for application of the model. As shown by PEDLOSKY (1987), the major deficiency is the absence of density stratification present in the real ocean. Boussinesq approximation is also used where the disturbance of the dimensions is small compared with its mean value. The static fluid pressure assumes that gravity is balanced with the vertical pressure gradient,

$$0 \approx -\frac{1}{\rho}\frac{\partial p}{\partial z} - g \tag{2}$$

and the incompressible assumption,

$$\nabla \cdot \mathbf{v} = 0. \tag{3}$$

With these approximations, the motions of the ocean waves can be expressed in Cartesian coordinates as,

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + g\frac{\partial h}{\partial x} - 2\Omega v \sin\phi = 0, \tag{4}$$

$$\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + g\frac{\partial h}{\partial y} + 2\Omega u \sin\phi = 0, \tag{5}$$

$$\frac{\partial h}{\partial t} + \frac{\partial}{\partial x}[(h - h_B)u] + \frac{\partial}{\partial y}[(h - h_B)v] = 0. \tag{6}$$

Here, u and v are the horizontal components of water particle velocities $\mathbf{v}$ in the x and y direction, h in the continuity equation is the sum of water depth plus earthquake/landslide vertical displacement, h_B is described as water depth or the sea bottom topography. Ω is the angular velocity of Earth's rotation and ϕ is the latitude. g is the gravitational acceleration.

A more simplified linear theory can be expressed as:

$$\frac{\partial u}{\partial t} + g\frac{\partial h}{\partial x} = 0, \tag{7}$$

$$\frac{\partial v}{\partial t} + g\frac{\partial h}{\partial y} = 0, \tag{8}$$

$$\frac{\partial h}{\partial t} + \frac{\partial}{\partial x}[(h - h_B)u] + \frac{\partial}{\partial y}[(h - h_B)v] = 0. \tag{9}$$

For far-field tsunami wave propagation, linear theory is adequate, but for the near field and the run-up process, shallow water theory with the convection term is needed. The Coriolis force term can also be included to account for the spherical inertial effect.

The viscous stress term of the bottom friction is also included in the very popular TSUNAMI model (IMAMURA *et al.*, 2006). In this case, the equations can be expressed as

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y} + g\frac{\partial h}{\partial x} - 2\Omega v \sin\phi + \frac{1}{2g}\frac{f}{h}u\sqrt{(u^2 + v^2)} = 0, \tag{10}$$

$$\frac{\partial v}{\partial t} + u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y} + g\frac{\partial h}{\partial y} + 2\Omega u \sin\phi + \frac{1}{2g}\frac{f}{h}v\sqrt{(u^2 + v^2)} = 0, \tag{11}$$

$$\frac{\partial h}{\partial t} + \frac{\partial}{\partial x}[(h - h_B)u] + \frac{\partial}{\partial y}[(h - h_B)v] = 0, \tag{12}$$

where f is the friction coefficient, which can be spatially dependent. $H = h - h_{\mathrm{B}}$ is the thickness of the fluid layer.

In general, this type of shallow water equation can be solved with the finite-difference method using different schemes, such as upwind total variation diminishing (TVD)

scheme (YEE *et al*., 1983). Multigrid methods may also be utilized to obtain better performance in numerical computing (ADAMS, 2000; BREZINA *et al*., 2004).

Finite-volume methods are becoming increasingly more popular for strongly nonlinear hyperbolic cases, if the convection term dominates over the other terms, especially when the waves break upon arriving at the coast (WEI *et al*., 2006).

In this paper, we propose using a least-squares scheme in the finite-element method to take full advantage of unstructured meshes to portray fractal-like features, in order to represent coastal bathymetry more exactly. This will be discussed in the next section. More importantly, we introduce a novel way to generate FORTRAN source code for finite-element computing that can run on a distributed parallel system. We present this work based on a parallel computing environment which we have developed for many years.

3. Parallel Codes for Tsunami Wave Propagation Using Modern Software Engineering

In geosciences, the major aim is to obtain an accurate physical model to understand the physics correctly. Mathematics forms the basis of this link. For the governing partial differential equations, adding an extra term or changing an existing linear coefficient to include nonlinearity often means difficult and Laborious work for coding. This process is very tedious and is prone to errors.

There has been recent progress in software development, in which parallel finite-element (FEM) codes in FORTRAN language, suitable for massively parallel computing, can be readily generated by modern advances in software engineering. Using this type of approach, we have taken an initiative (ZHANG *et al*., 2005, 2007; SHI *et al*., 2006) in generating codes for a variety of geodynamical problems which include crustal deformation, mantle convection and now tsunami wave propagation.

In this section we demonstrate a modeling language-based parallel finite-element computing environment as the direct link between computational mathematics and geosciences. The FORTRAN source code generated from this system can be run on distributed parallel machines without any modifications. All the environment users need to input to this system are the expressions of PDEs and their corresponding algorithm expressions.

We can show this system and our method of coding as follows. First is the partial differential equations File (shallow.pde file). Figure 2 is an example specifically designed for the nonlinear tsunami equations.

The shallow.pde file is one of the input files in which we use the operator splitting method, in which the calculation process is divided into three steps. We use the Galerkin virtual displacement method, least-squares finite-element and Galerkin virtual displacement method to solve the elliptic terms, convection term and diffusion term, respectively. Although it seems to be complicated numerically, this procedure can handle the strong nonlinear terms associated with tsunami waves, especially for the run-up process. The

```
Shallow.pde

disp hu hv
coor x y
func fhu fhv
coef hun1 hvn1 hun hvn un1 vn1 un vn
hn1
mate rou 1.0
shap %1 %2
gaus %3
mass %1 1.0
vect hun hun hvn
vect x x y
vect fhun1 fhun1 fhvn1
vect un un vn
vect un1 un1 vn1
vect hu hu hv
vect fhu fhu fhv
```

```
func
$cv fhun1=+{un_j/x_j}*hun+{hun/x_j}*un_j
$cv fhun1=+fhun1*dt+hun1
$cv fhvn1=+{un_j/x_j}*hvn+{hvn/x_j}*un_j
$cv fhvn1=+fhvn1*dt+hvn1
fhu=+[hu]+[hu_j]*{un/x_j}*dt+[hu_j/x_j]*un*dt
      +[hu]*{un_j/x_j}*dt+[hu/x_j]*un_j*dt

fhv=+[hv]+[hu_j]*{vn/x_j}*dt+[hu_j/x_j]*vn*dt
      +[hv]*{un_j/x_j}*dt+[hv/x_j]*un_j*dt

stif
dist=+[fhu_i;fhu_i]

load=+[fhu_i]*fhun1_i

end
```

Figure 2

This is the English-like expressions of the convection terms of shallow water equations. We use vector expression and make the whole finite element weak form very briefly. Hu and hv are the variables (unknowns), hun, hvn, un1, vn1, un, vn, hn1 are all initial values of current time step unknowns. " Fun" section means that we are defining new functions. The "stiff" and "load" sections are the expressions for the stiffness matrix and right-hand side, respectively.

algorithm expression based on the computing environment can be written simply in the Generalized Coupling Nonlinear file (shallow.gcn), as shown in Figure 3.

When all files have been input to this computing environment, this system will automatically generate FORTRAN program segments from the information of partial differential equations and algorithm expressions, as shown in Figure 4.

These program segments will then be inserted into a common program stencil, to different locations respectively, as shown in Figure 5. In this manner the entire source code package is generated. This software package, which is based on a distributed parallel computing architecture machine and message passing interface (MPI) system (www.mpi-forum.org), together with parallel solvers for large-scale linear systems and

```
Shallow.gcn

defi
a shola &
b sholb
c sholc

startsin a
startsin b
startsin c
call trans
if exist stop del stop
```

a

```
:1
bft
solvsin a
copy unod unoda
if exist end del end
:2
solvsin b
if not exist end goto 2
solvsin c
call post
if not exist stop goto 1
```

b

Figure 3

(a) and (b) are the first and second parts of one GCN file. This file resembles a scripts file to communicate to the computing environment for generating various source codes, using different program stencils. Another scripts file will also be generated according to this input file, which can run all the programs generated after the compilation.

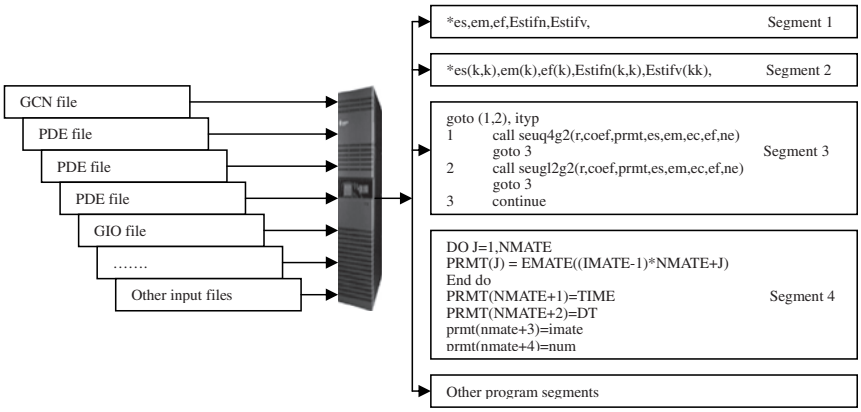


Figure 4
This schematic diagram illustrates how this computing environment generates all the source code according to the input files. The GIO file will describe the format of preprocessing files, such as the different element types, element factors, coordination system, initial values and constrained boundary information. The PRMT part of the generated program segment is derived basically from the GIO file.

automatic mesh and data partition system, can be compiled and run in parallel without any changes. Users can download the source code from the generation server via the client software interface. This is a typical prototype of the grid-computing environment. We will continue work on this area.

In this case, all the algorithmic expressions are already stored in the system library and can be used directly. A typical algorithm expression for elliptic type of partial differential equation is expressed as shown in Figure 5. More details of the modeling language and the computational environment can be seen in our recent paper (ZHANG *et al*., 2007).

4. Three-Dimensional Tsunami Modeling

Besides epidemic control and post-tsunami recovery, a timely and effective warning system is one of the most crucial elements to determine the threat to the coastal communities. This warning system can consist of gathering as much information as possible on the potential tsunamis, estimation of their frequency, detecting the dynamic process of fault rupturing and sea-floor deformations along the main thrusts of plate boundaries, tsunami formation, tsunami wave propagation and the coastal region inundated. Technical issues of tsunami modeling and forecasting, tsunami formation, tsunami wave propagation and run-up process are still the persistent research problems. Wave propagation over short distances (near field) and long distances (far field) are quite different because of Coriolis acceleration and the friction effects of sea-bed sediment

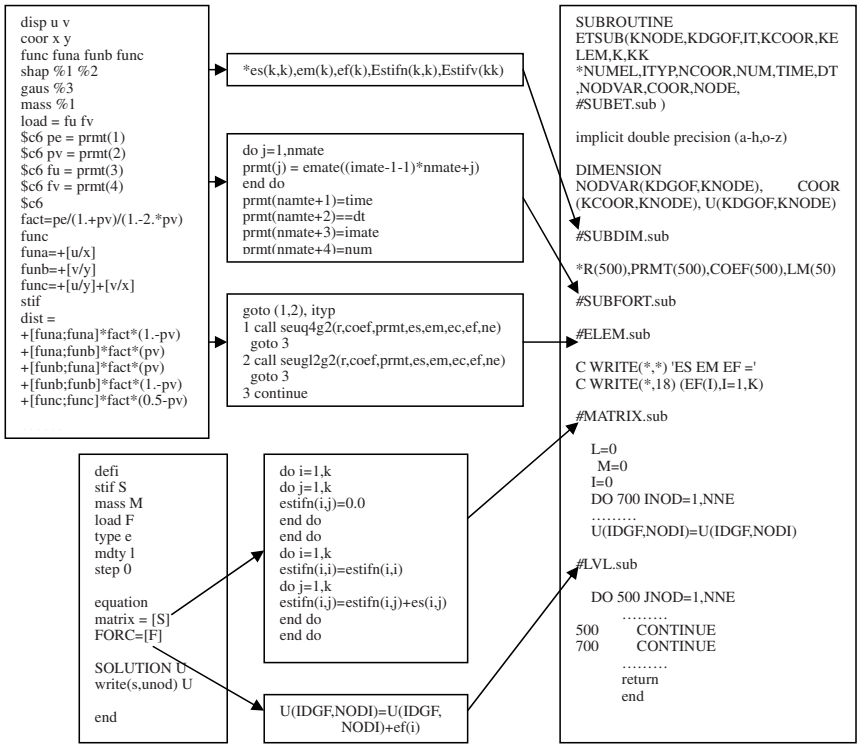


Figure 5

Generation of FORTRAN source codes which are solved by the FEM (finite-element) method. The left two columns show the input finite-element modeling language, the upper part is the expressions of partial differential equations, and the lower part shows the solving algorithmic expressions of the elliptic type PDEs. The modeling language-based computing environment will generate program segments (center column) according to these expressions, then all the program segments will be inserted into a program stencil for assembling as a FORTRAN-77 styled source code (the very right column).

layer. For the run-up processes, the shallow water equations are not feasible to model the highly nonlinear hyperbolic phenomena.

The construction of the initial water depth distribution is of vital importance for a full three-dimensional simulation of tsunami propagation. We have used the GTOP30 data for continental topography and SRTM30 data for bathymetry (sea-floor topography) to generate the finite-element mesh describing the sea-floor profile and landscape around them. This makes it feasible to obtain the water depth distribution within this area by means of generating water meshes over the bathymetry profile and local mesh refinement in the costal areas. We use unstructured mesh generation technology to produce the finite-element mesh grid describing the actual water body. Figure 6 shows this process of generating the highly irregular finite-element meshes.

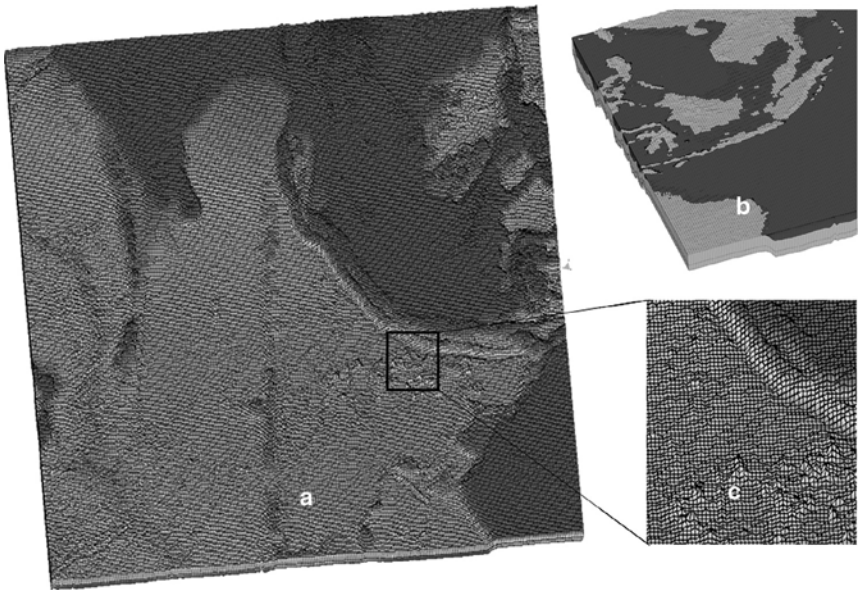


Figure 6
These three graphs show how we generate the finite-element mesh of sea-bed and sea-water layer distribution and make it consistent with the profile of sea-floor data. (a) shows the topography data from GTOP 30, (b) shows how to generate the finite-element mesh of the seawater over sea-floor bathymetry, (c) shows the zoomed-in section of one portion of the whole area. Because the upper layer of this water body is more important than the lower layer, we have used special mesh size distribution to refine the mesh on the top part of the sea-water layer.

The topography of the sea-bed can be regarded as a basin of seawater, thus we just delete the mesh elements of sea-bed and all water layer mesh elements whose thicknesses are less that 0.01 meter. These thin mesh elements will cause numerical singularities during the numerical simulation. Figure 7 shows the result of water body distribution and the finite-element mesh.

Sediment layer effect (or bottom friction effect) contributes greatly to tsunami wave propagation over long distance. The shallow water equations always have this mechanism as a friction term. In realistic three-dimensional model construction, we also need such data. In this case, we use the lowest layer of seawater as the mixture of sediment and seawater and assign different material factors, such as a higher rheological coefficient (MINOURA *et al*., 2005).

With the help of high performance computing infrastructure, it is also possible for us to take coupling process of sea-floor seismic wave propagation and tsunami wave propagation into consideration (CACCHIONE *et al*., 2002; GOWER, 2005; LUI *et al*., 2005). We show this kind of nonlinear coupling and the associated finite-element mesh generated as shown in Figure 8.

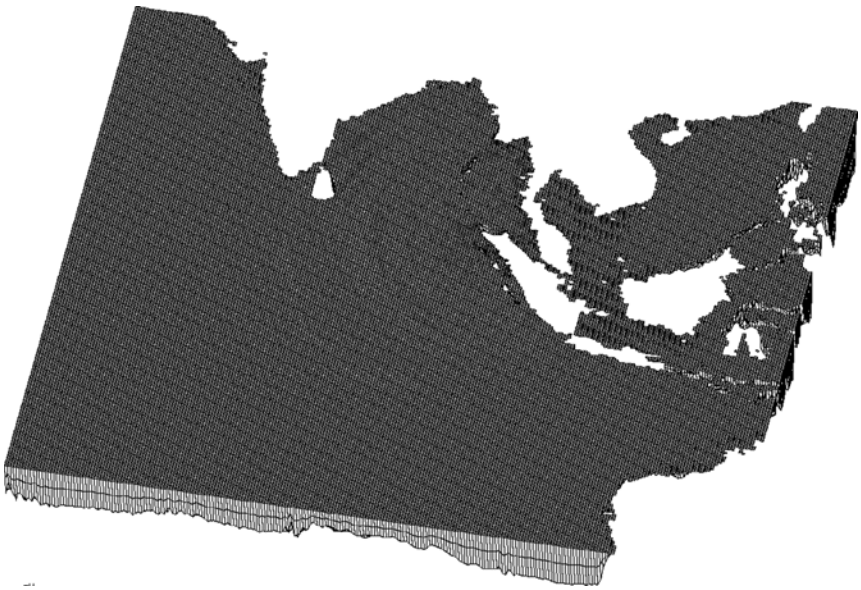


Figure 7
Water distribution and the finite-element mesh.

We developed both sequential and parallel versions of tsunami propagation. The sequential version of the 3-D finite-element model has more than 180,000 mesh nodes and the parallel 3-D version has more than 2 million finite-element nodes. We ran the parallel version on a 32-node PC cluster. The total run time was about 6 hours. In the next section we will show the simulation results from this type of modeling.

5. Numerical Solution of the Set of 3-D Tsunami Equations

Using automatic grid generation methods, we have devised a finite-element based code, for the three stages which culminates with the use of the augmented Lagrangian method (DANILOV *et al.*, 2004; FORTIN and GLOWINSKI, 1983; ARǴAEZ and TAPIA, 2002; BERTSEKAS, 1996) for the run-up process, as well as the Arbitrary Lagrange-Euler Configuration method (LONGATTE *et al.*, 2003) to address the free surface problem near shore. Our continuous efforts are focused on seeking novel algorithms and state-of-art techniques, in order to unravel the mysteries associated with tsunami wave propagation and wind-driven waves in 3-D. We have cast the Navier-Stokes equations within the framework of an incompressible model with an equation of state for the seawater. Our formulation allows the tracking and simulation of three principal stages, to the formation,

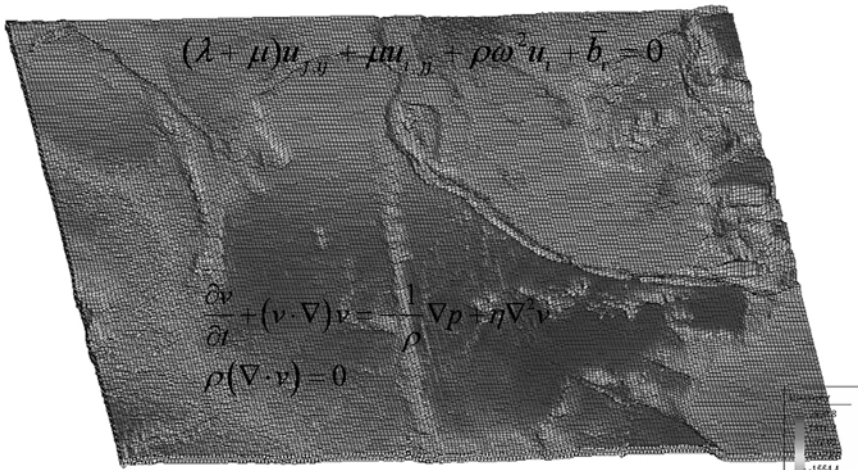


Figure 8
Finite-element mesh for coupled modeling of sea-floor seismic wave propagation and tsunami wave propagation, where and are the displacement and velocity vectors, respectively, and are the Lame elastic constants, is mass density, is circular frequency, and is the body force.

propagation and run-up stages of tsunami and waves coming ashore. These equations are written as the following:

$$\frac{\partial \mathrm{v}}{\partial t}+(\mathrm{v}\cdot\nabla)\mathrm{v}+2\Omega\times\mathrm{v}=-\frac{1}{\rho}\nabla p+f(h(x,y),\gamma(x,y))\mu\nabla^2\mathrm{v}+\kappa p+g, \qquad (13)$$

where $\boldsymbol{v}$, Ω , p and η are the velocity, rotation angular velocity, the dynamical pressure and rheology respectively, and g is the gravity acceleration.

$$f=f(h(x,y),\gamma(x,y)) \qquad (14)$$

is the factor of relationship between velocity and wavelength with the depth of water distribution, the seabed condition of sediment layer thickness which will absorb energy from tsunami waves while they are propagating. $h(x,y)$ is the parameterized coefficient of velocity and wavelength relationship. $\gamma(x,y)$ is referred to as the impact of sediment layer on the tsunami wave, in dissipating the wave energy.

The relationship of velocity and wavelength with depth of water distribution is shown as Figure 9.

We also need to explain kappa term; this term contains more numerical than physical meaning:

$$\kappa=k(\Delta t,\Delta h) \qquad (15)$$

is the augmented Lagrangian multiplier for the pressure term to make it more elliptical-like and stable numerically, when a first-order explicit method is deployed to update the time-dependent equations.

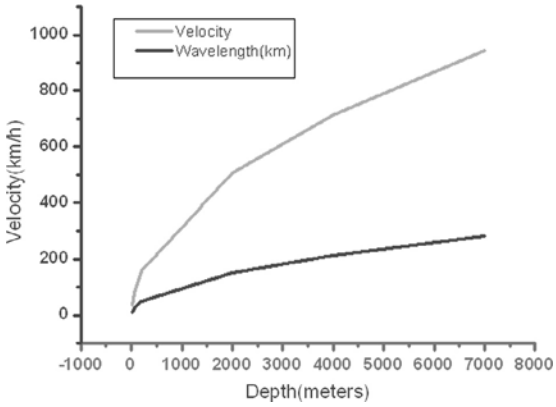


Figure 9
shows the relationship of velocity and wavelength with depth of water distribution. Basically, the velocity of idealized traveling waves on the ocean is wavelength-dependent and for shallow enough depths it also depends upon the depth of the water. This can be formulated as, here is tsunami wavelength, is the water depth (http://hyperphysics.phy-astr.gsu.edu/hbase/watwav.html#c3).

The operator-splitting algorithm (BRUSDAL *et al.*, 1998; MARINOVA *et al.*, 2003) is utilized to solve these Navier-Stokes (NS) equations. This numerical computing scheme uses different methods to solve different types of partial differential equations, respectively. All these equations are from the same NS equation set. We use this algorithm mainly because this kind of algorithm can allow the strong nonlinear processes of run-up processes while the tsunami reaches the seashore, which may lead to an unstable numerical solution of finite element and is very difficult to converge. In brief, we describe our algorithm as the following.

Firstly, we solve the diffusion equations, together with the incompressible condition equation,

$$\begin{cases} \dfrac{\partial \mathrm{v}}{\partial t} + 2\Omega \times \mathrm{v} = -\dfrac{1}{\rho}\nabla p + f(h(x,y),\gamma(x,y))\mu\nabla^2 \mathrm{v} + \kappa p + g \\ \nabla \cdot \mathrm{v} = 0 \end{cases} \tag{16}$$

the weak form of these equations is given in terms of inner-products and is written as

$$\begin{cases} \left(\dfrac{\partial \mathrm{v}}{\partial t}, \bar{\mathrm{v}}\right)_\Omega + (2\Omega \times \mathrm{v}, \bar{\mathrm{v}})_\Omega + \left(\dfrac{1}{\rho}\nabla p, \bar{\mathrm{v}}\right)_\Omega - f(h(x,y),\gamma(x,y))\mu(\nabla \mathrm{v}, \nabla \bar{\mathrm{v}})_\Omega = (g, \bar{\mathrm{v}})_\Omega \\ (\nabla \cdot \mathrm{v}, \bar{p})_\Omega - k(p, \bar{p})_\Omega = 0 \end{cases} \tag{17}$$

where

$$k = c \cdot \det . \tag{18}$$

Here c is an independent constant, det is the determinant of element Jacobian matrix.

Then we solve the convection-like equation in the mass-conservation balance,

$$\rho \frac{\partial \mathbf{v}}{\partial t} + \rho(\mathbf{v} \cdot \nabla)\mathbf{v} = 0, \tag{19}$$

using a first-order Euler backward difference scheme

$$\frac{\partial \mathbf{v}}{\partial t} = \frac{\mathbf{v}^{n+1} - \mathbf{v}^n}{dt} \tag{20}$$

and the Newton-Raphson method to linearize this nonlinear term,

$$(\mathbf{v} \cdot \nabla)\mathbf{v} = (\mathbf{v}^n \cdot \nabla)\mathbf{v}^n + (\mathbf{v}^n \cdot \nabla)(\mathbf{v} - \mathbf{v}^n) + ((\mathbf{v} - \mathbf{v}^n) \cdot \nabla)\mathbf{v}^n \tag{21}$$

from equation (22, 23 and 24) we get,

$$\rho \frac{\mathbf{v}^{n+1} - \mathbf{v}^n}{dt} + \rho(\mathbf{v}^n \cdot \nabla)\mathbf{v}^{n+1} + \rho(\mathbf{v}^{n+1} \cdot \nabla)\mathbf{v}^n = \rho(\mathbf{v}^n \cdot \nabla)\mathbf{v}^n \tag{22}$$

another form is

$$\rho \mathbf{v}^{n+1} + dt\rho(\mathbf{v}^n \cdot \nabla)\mathbf{v}^{n+1} + dt\rho(\mathbf{v}^{n+1} \cdot \nabla)\mathbf{v}^n = \rho \mathbf{v}^n + dt\rho(\mathbf{v}^n \cdot \nabla)\mathbf{v}^n. \tag{23}$$

We use the least-squares method (Bochev and Gunzburger, 1993), which is robust for solving hyperbolic equations, to solve this equation,

$$\begin{aligned}
&\rho(L(v_x), L(\bar{v}_x))_\Omega + \rho(L(v_y), L(\bar{v}_y))_\Omega + \rho(L(v_z), L(\bar{v}_z))_\Omega \\
&\quad = \left(\rho\left(v_x^n + dt\left(v_x^n \frac{\partial}{\partial x} v_x^n + v_y^n \frac{\partial}{\partial y} v_x^n + v_z^n \frac{\partial}{\partial z} v_x^n\right)\right), L(\bar{v}_x)\right)_\Omega \\
&\qquad + \left(\rho\left(v_y^n + dt\left(v_x^n \frac{\partial}{\partial x} v_y^n + v_y^n \frac{\partial}{\partial y} v_y^n + v_z^n \frac{\partial}{\partial z} v_y^n\right)\right), L(\bar{v}_y)\right)_\Omega \\
&\qquad + \left(\rho\left(v_z^n + dt\left(v_x^n \frac{\partial}{\partial x} v_z^n + v_y^n \frac{\partial}{\partial y} v_z^n + v_z^n \frac{\partial}{\partial z} v_z^n\right)\right), L(\bar{v}_z)\right)_\Omega
\end{aligned} \tag{24}$$

where L is given by,

$$\begin{aligned}
L(v_x) = {} & v_x^{n+1} + dt\left(v_x^n \frac{\partial}{\partial x} v_x^{n+1} + v_y^n \frac{\partial}{\partial y} v_x^{n+1} + v_z^n \frac{\partial}{\partial y} v_x^{n+1}\right) \\
& + dt\left(v_x^{n+1} \frac{\partial}{\partial x} v_x^n + v_y^{n+1} \frac{\partial}{\partial y} v_x^n + v_z^{n+1} \frac{\partial}{\partial y} v_x^n\right)
\end{aligned} \tag{25}$$

$$L(v_y) = v_y^{n+1} + dt\left(v_y^n \frac{\partial}{\partial x} v_y^{n+1} + v_x^n \frac{\partial}{\partial y} v_y^{n+1} + v_z^n \frac{\partial}{\partial y} v_y^{n+1}\right) + dt\left(v_y^{n+1} \frac{\partial}{\partial x} v_y^n + v_x^{n+1} \frac{\partial}{\partial y} v_y^n + v_z^{n+1} \frac{\partial}{\partial y} v_y^n\right) \quad (26)$$

$$L(v_z) = v_z^{n+1} + dt\left(v_z^n \frac{\partial}{\partial x} v_z^{n+1} + v_x^n \frac{\partial}{\partial y} v_z^{n+1} + v_y^n \frac{\partial}{\partial y} v_z^{n+1}\right) + dt\left(v_z^{n+1} \frac{\partial}{\partial x} v_z^n + v_x^{n+1} \frac{\partial}{\partial y} v_z^n + v_y^{n+1} \frac{\partial}{\partial y} v_z^n\right) \quad (27)$$

Here we use v_x,v_y,v_z as the three orthogonal components of **v** for a clear description, and $\bar{v}_x, \bar{v}_y, \bar{v}_z$ is the virtual displacement of v_x,v_y,v_z, respectively. Our developed parallel computing environment is used to generate all the Fortran source code.

Our formulation allows the tracking and simulation of three stages, principally the formation, propagation and the run-up stages of tsunami, culminating with the waves coming ashore. This formulation also allows for the wave surface to be self-consistently determined within a linearized framework and is computationally very fast. The sequential version of this code can run on a workstation with 4 Gbytes memory less than 2 minutes per time step for one million grid points. This code has also been parallelized with MPI-2 and has good scaling properties, nearly linear speedup, which has been tested on a 32-node PC cluster. We have employed the actual ocean sea-floor topographical data to construct oceanic volume and attempt to construct the coastline as realistically as possible, using 11 levels of structure meshes in the radial direction of the Earth. Our initial focus is on the East Coast of Asia. In order to understand the intricate dynamics of the wave interactions, we have implemented a visualization overlay based on the Amira package (http://www.amiravis.com/), a powerful 3-D volume rendering visualization tools for massive data post-processing. The ability to visualize large data sets remotely is also an important objective we are aiming for, as international collaboration is one of the top aims of this research. This part will be displayed in section 7.

6. Visualization

The dynamics of tsunami wave propagation are very rich and offer great opportunities for visual studies. Yet the visualization of tsunami wave propagation has not maintained its progress with the advances in visualization being made in mantle convection and earthquake dynamics. A review of the current status of visualization in the geosciences has been given in the COHEN Report (2005). We have employed an interactive system for advanced 3-D visualization and volume rendering, the package Amira (http://www.amiravis.com). This software takes advantage of modern graphics hardware and is available for all standard operating systems, ranging from Linux to MacOS and Windows. The extensive set of easy-to-use features includes data imaging on the Cartesian and

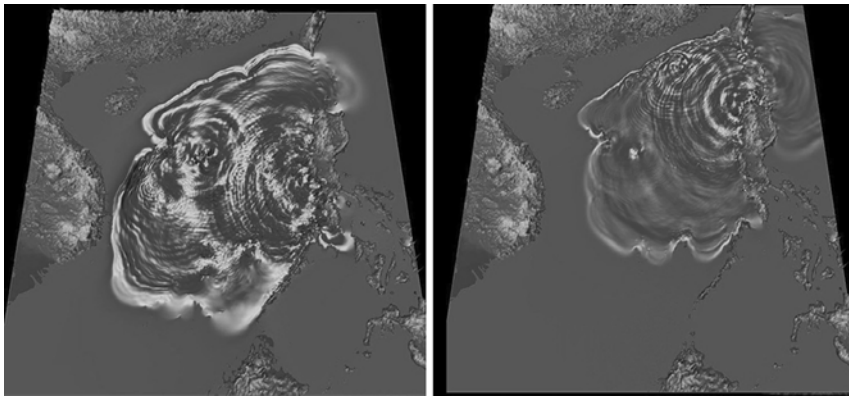


Figure 10
Visualization of tsunami wave propagation in the South China Sea at different time steps, which span less than one hour of real time.

finite-element grids, scalar and vector field visualization algorithms, computation of iso-surfaces, direct volume rendering, time-series manipulation and creating movies, support for the Tcl scripting language and remote data-processing.

Figure 10 shows rendered results computed by the Amira package. The propagation of the tsunami wave over time is clearly demonstrated through the visualization.

The dynamics of tsunami wave-propagation simulation result we are visualizing are always associated with specific geographical regions globally. In the visualization package such as Amira we mentioned above, it is possible to render neighboring terrains together with the simulation visualization. However, for such packages, the semantic geographical information, which is critical for hazard evaluation based on the tsunami wave propagation we simulated, is missing. In our research we start to integrate the visualization result of our simulation with the newly available software Google Earth [Google Earth] for contextual geological information for such visualizations (Fig. 11).

Google Earth is a server-client based virtual globe program (http://earth.google.com/). It maps the entire earth by pasting images obtained from satellite imagery, aerial photography and GIS over a 3-D globe. Many regions are available with at least 15 meters of resolution. Google Earth allows users to search for addresses, enter coordinates, or simply use the mouse to browse to a location. Google Earth also has digital terrain model data collected by NASA's Shuttle Radar Topography Mission. Users can directly view the geological features in three-dimensional perspective projection, instead of as 2-D maps. We use the image overlay function provided by the Google Earth software to integrate our visualization results into a virtual globe context. Image overlay function allows us to map rendered pictures of our visualization result on to the virtual globe with the geographical locations specified by the user. The transparency of the

Figure 11
Visualization of tsunami simulation results using Google Earth software with Power-Wall Display (http://www.lcse.umn.edu).

mapped pictures can be tuned to between 0 (totally transparent) and 1 (totally opaque). In Figure 11, one time frame of tsunami wave propagation simulation results is mapped to the Google Earth and projected to a multi-panel PowerWall display to allow careful, close inspections. Note in the figure, that the cities and other important geographical locations under the impact of the simulated tsunami simulation are clearly visible to the viewer. Such a tool could greatly, enhance the interpolation and presentation of our tsunami wave-propagation simulation results.

The seismic displacement itself is multi-scale in nature. Although the earthquake extends its deformation throughout the whole far field, only regions in the near field have large displacement.

The Google Earth enables the users to navigate the whole virtual globe with integrated visualizations at different level of details. As illustrated in Figure 12, the user can freely zoom to different levels of resolution to either obtain a global view or a close look. Note that in the most zoomed-out image, Google Earth provides the recorded earthquake information indicated by a red dot in the studied region. Such information is valuable for researchers seeking to understand the event. The visualization results we embed in Google Earth could also be saved in a data exchange format (KML) together with the geological context information. Such data could be directly opened by other parties who have the software.

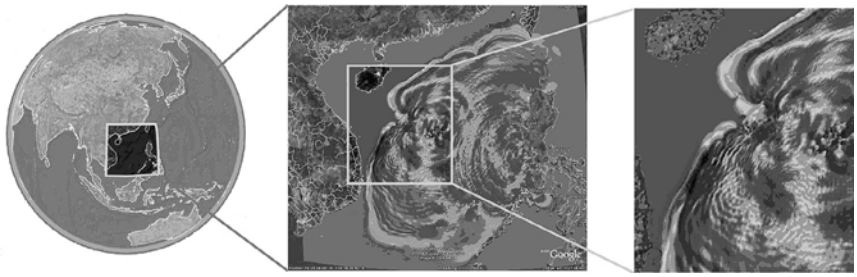


Figure 12
Visualization of tsunami simulation results in the South China Sea at multi-level of details.

7. Conclusions and Perspectives

We have laid out the hierarchy of the different levels of partial differential equations needed to solve the tsunami problem, ranging from the linear shallow-water equation to the fully nonlinear 3-D Navier-Stokes equations in which the role of sedimentary layer is introduced as a regularizing agent for stabilizing the numerical solution near the shore. We regard that the forecasting and tsunami-warning problem may be best attacked with the linear shallow-water equation, because of the enormous computational efforts needed for solving the nonlinear shallow water equations and the full 3-D equations. We cannot over stress the importance of using the physics of sedimentary processes in stabilizing the most vicious nonlinearities during the run-up stage in the 3-D problem.

We have described the visualization of both the seismic displacements and tsunami wave propagation using the Amira visualization package and our own developed method using the graphics processing unit (GPU), which offers a low-cost solution from which to solve a graphically intensive task such as the construction of InSAR images. We have also presented a technique for overlaying our calculations atop the map using Google Earth. This innovation will assist the reader to better understand the multi-scale physical phenomena associated with tsunami waves.

Acknowledgement

This work is supported by the National Science Foundation of China under Grant Numbers 40474038, 40374038 and by the National Basic Research Program of China under Grant Number 2004cb418406 and and U.S. National Science Foundation under the ITR and EAR programs. This work was conducted as part of the visualization working group at the laboratory of computational geodynamics supported by the Graduate University of Chinese Academy of Sciences. We thank Shi Chen and Shaolin Chen as members of the visualization working group who provided some of the figures. We thank Mark S. Wang for technical assistance.

References

Adams, M.F. (2000), *Algebraic multigrid methods for constrained linear systems with applications to contact problems in solid mechanics*, Numerical Linear Algebra with Applications *11*(2–3), 141–153.

Amira, http://www.amiravis.com.

Arbic, B.K., Garner, S.T., Hallberg, R.W., and Simmons, H.L. (2004), *The accuracy of surface elevations in forward global barotropic and baroclinic tide models*, Deep-Sea Res. *II* 51, 3069–3101.

Argáez, M., and Tapia, R.A. (2002), *On the global convergence of a modified augmented Lagrangian line search interior-point method for Nonlinear Programming*, J. Optimiz. Theory Applicat. *114*, 1–25.

ARSC, http://www.arsc.edu/challenges/2005/globaltsunami.html.

Asian Tsunami Videos, http://www.asiantsunamivideos.com/.

Barnes, W.L., Pagano, T.S., and Salomonson, V.V. (1998), *Prelaunch characteristics of the Moderate Resolution Imaging Spectroradiometer (MODIS) on EOS-AM1*, IEEE Trans. Geosci. Remote Sensing *36*, 1088–1100.

Bertsekas, D.P., *Constrained Optimization and Lagrange Multiplier Methods* (Athena Scientific, Belmont 1996).

Bochev, P. and Gunzburger, M. (1993), *A least-squares finite-element method for the Navier-Stokes equations*, Appl. Math. Lett. *6*, 27–30.

Brezina, M., Falgout, R., MacLachlan, S., Manteuffel, T., McCormick, S., Ruge, J. (2004), *Adaptive smoothed aggregation*, SIAM J. Sci. Comp. *25*, 1896–1920.

Brusdal, K., Dahle, H.K., Karlsen, K.H., and Mannseth, T. (1998), *A study of the modelling error in two operator splitting algorithms for porous media flow*, Comput. Geosci. *2*(1), 23–36.

Butler, D. (2006), *Virtual globes: The web-wide world*, Nature *439*, 776–778.

Cacchione, D.A., Pratson, L.F., and Ogston, A.S. (2002), *The shaping of continental slopes by internal tides*, Science *296*, 724–727.

Chang, C., and Gunzburger, M. (1990), *A subdomain Galerkin/least-squares method for first order elliptic systems in the plane*, SIAM J. Numer. Anal. *27*, 1197–1211.

Chang, H.C., Ge, L., and Rizos, C. (2005), *Asian Tsunami Damage Assessment with Radarsat-1 SAR Imagery*, http://www.gmat.unsw.edu.au/snap/publications/chang_etal2005e.pdf.

Chen, Q., Kirby, J. T., Dalrymple, R. A., Kennedy, A. B., and Chawla, A., 2000, *Boussinesq modeling of wave transformation, breaking and runup. II: Two horizontal dimensions*, J. Waterway, Port, Coastal and Ocean Engin. *126*, 48–56.

Cohen, R.E. (2005), *High-Performance Computing Requirements for the Computational Solid Earth Sciences*, http://www.geo-prose.com/computational_SES.html.

Danilov, S., Kivman, G., and Schroeter, J. (2004), *A finite-element ocean model: Principles and evaluation*, Ocean Modelling *6*, 125–150.

Fortin, M. and Glowinski, R., *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems* (Elsevier Science, New York, 1983).

Geist, E.L., Titov, V.V., and Synolakis, C.E. (2006), *Tsunami: Wave of change*, Scientific American, *294*(1), 56–63.

Gica, E. and Teng, M.H., *Numerical modeling of earthquake generated distant tsunamis in the Pacific Basin*, Proc 16 ASCE Engin. Mech. Conf., (University of Washington, Seattle 2003).

Gill, A.E., *Atmosphere-Ocean Dynamics* (Academic Press, New York 1982).

Gonzalez, F.I., Satake, K., Boss, E.F., and Mofjeld, H.O. (1995), *Edge wave and non-trapped modes of the 25 April 1992 Cape Mendocino tsunami*, Pure Appl. Geophys. *144*(3/4), 409–426.

Google Earth, http://earth.google.com/.

Gower, J. (2005), *Jason 1 detects the 26 December 2004 tsunami*, EOS Transact. Am. Geophy. Union *86*(4), 37–38.

Gower, J. (2005), *Jason 1 detects the 26 December 2004 tsunami*, EOS Transact. Am. Geophys. Union, *86*(4), 7–38.

Guo, J.Y., *Fundamental Geophysics* (Surveying and Mapping Press, China, 2001).

Hyperphysics, http://hyperphysics.phy-astr.gsu.edu/hbase/watwav.html#c3.

Imamura, F. *et al*. (2006), Tsunami Modeling Manual.

ISPRS, http://www.isprs.org/istanbul2004/.

JRC, http://tsunami.jrc.it/model/model.asp.

KEES, C.E. and MILLER, C.T. (2002), *Higher order time integration methods for two-phase flow*, Adv. Water Resources *25*, 159–177.

KENNEDY, A.B, CHEN, Q., KIRBY, J.T., and DALRYMPLE, R.A. (2000), *Boussinesq modeling of wave transformation, breaking and runup. I: One dimension*, J. Waterway, Port, Coastal and Ocean Engin. *126*, 39–47.

KENNEDY, A.B. and KIRBY, J.T. (2003), *An unsteady wave driver for narrow-banded waves: Modeling nearshore circulation driven by wave groups*, Coastal Engin. *48*(4), 257–275.

KIRBY, J. T. and DALRYMPLE, R. A. (1986), *Modelling waves in surfzones and around islands*, J. Waterway, Port, Coastal and Ocean Engin. *112*, 78–93.

LAYTON, A.T., and VAN DE PANNE, M. (2002), *A numerically efficient and stable algorithm for animating water waves*, Visual Comput. *18*, 41–53.

LONGATTE, E., BENDJEDDOU, Z., and SOULI, M. (2003), *Application of arbitrary Lagrange-Euler formulations to flow-induced vibration problems*, J. Pressure Vessel Technol. *125*(4), 411–417.

LUI, P. L.F., LYNETT, P., FERNANDO, H., JAFFE, B.E., HIGMAN, B., MORTON, R., GOFF, J., and SYNOLAKIS, C. (2005), *Observations by the international tsunami survey team in Sri Lanka*, Science, *308*, 1595.

MAJDA, A.J. (2003), *Introduction to PDEs and waves for the atmosphere and ocean*, Courant Lecture Notes 9, Am. Math. Soc.

MARINOVA, R.S., CHRISTOV, C.I., and MARINOV, T.T. (2003), *A fully coupled solver for incompressible Navier-Stokes equations using operator splitting*, Internat. J. Comp. Fluid Dyn. *17*(5), 371–385.

MEINIG, C., STALIN, S.E., NAKAMURA, A.I., GONZALEZ, F., and MILBURN, H.G. (2005), *Technology developments in real-time tsunami measuring, monitoring and forecasting in oceans*, MTS/IEEE, 19–23 September 2005, Washington, D.C.

MINOURA, K., IMAMURA, F., KURAN, U., NAKAMURA, T., PAPADOPOULOS, G.A., SUGAWARA, D., TAKAHASHI, T., YALCINER, A.C. (2005), *A tsunami generated by a possible submarine slide: Evidence for slope failure triggered by the North Anatolian fault movement*, Natural Hazards *36–3*(10), 297–306.

MOBAHERI, M.R. and MOUSAVI, H. (2004), *Remote sensing of suspended sediments in surface waters using MODIS images*, Proc. XXth ISPRS Congress, Geo-Imagery Bridging Continent, Istanbul, 12 –23.

MOFJELD, H.O., TITOV, V.V., GONZALEZ, F.I., and NEWMAN J.C. (2000), *Analytic theory of tsunami wave scattering in the open ocean with application to the North Pacific Ocean*, NOAA Tech. Memor. ERL PMEL-116, 38.

MPI, www.mpi-forum.org.

NAEIJE, M.E., DOORNBOS, L., *et al.* (2002), *Radar altimeter database system: Exploitation and Extension (RADSxx)*, Final Report, NUSP-2 Report 02–06.

NOURBAKHSH, I., SARGENT, R., WRIGHT, A., CRAMER, K., MCCLENDON, B., and JONES, M. (2006), *Mapping disaster zones*, Nature *439*, 787–788.

ORMAN, J.V., COCHRAN, J.R., WEISSEL, J.K., and JESTIN, F. (1995), *Distribution of shortening between the Indian and Australian plates in the central Indian Ocean*, Earth Planet. Sci. Lett. *133*, 35–46.

PEDLOSKY, J., *Geophysial Fluid Dynamics* (Springer-Verlag, New York 1987).

PELINOVSKY, E., TALIPOVA, T., KURKIN, A., and KHARIF, C. (2001), *Nonlinear Mechanism of Tsunami Wave Generation by Atmospheric Disturbances*, Natural Hazard and Earth Sci. *1*, 243–250.

SHI, Y.L., ZHANG, H., LIU, M., YUEN, D., and WU, Z.L. (2006), *Developing a viable computational environment for geodynamics*, WPGM, Beijing, (Abstract).

SMITH, W.H.F., SCHARROO, R., TITOV, V.V., ARCAS, D., and ARBIC, B.K. (2005), *Satellite altimeters measure tsunami*, Oceanography *18*(2), 10–12.

SWANSON, R.C. (1992), *On central-difference and upwind schemes*, J. Comp. Phys. *101*, 292–306.

TANG, L., CHAMBERLIN, C. *et al*. (2006), *Assessment of potential tsunami impact for Pearl Harbor, Hawaii*. NOAA Tech. Memo. OAR PMEL.

TIBBETTS, J. (2002), *Coastal Cities: Living on the Edge. In Environmental Health Perspectives*, 110(11), http://ehp.niehs.nih.gov/members/2002/110–11/focus.html.

TITOV, V.V. and GONZALEZ, F.I. (1997), *Implementation and testing of the Method of Splitting Tsunami (MOST) model*, NOAA Tech. Memo. ERL PMEL-112, 11.

TITOV, V.V., GONZALEZ, F.I., BERNARD, E.N., EBLE, M.C., MOFJELD, H.O., NEWMAN, J.C., and VENTURATO, A.J. (2005), *Real-time tsunami forecasting: Challenges and solutions*, Natural Hazards *35*(1), 41–58.

TITOV, V.V., GONZALEZ, F.I., MOFJELD, H.O., and VENTURATO, A.J. (2003), *NOAA time seattle tsunami mapping project: procedures, data sources, and products*, NOAA Tech. Memo. OAR PMEL-124.

TITOV, V.V., MOFJELD, H.O., GONZALEZ, F.I., and NEWMAN J.C. (1999), *Offshore forecasting of Alaska-Aleutian Subduction Zone tsunamis in Hawaii*, NOAA Tech. Memo. ERL PMEL-114, 22.

TITOV, V.V., RABINOVICH, A.B. *et al*. (2005), *The global reach of the 26 December 2004 Sumatra Tsunami*, Science, DOI: 10.1126/science, 1114576.

UNEP-WCMC IMAPS, http://nene.unep-wcmc.org/imaps/tsunami/viewer.htm.

VAN WACHEM, B.G.M., and SCHOUTEN, J.C. (2002), *Experimental validation of 3-D Lagrangian VOF model: Bubble shape and rise velocity*, AIChE J. *48*(12), 2744–2753.

WALKLEY, M. and BERZINS, M. (2002) *A finite-element method for the two-dimensional extended Boussinesq equations*, Internat. J. Num. Meth. in Fluids *39*(10), 865–885, 2002.

WEI, Y., MAO, X.Z., and CHEUNG, K.F. (2006), *Well-balanced finite-volume model for long-wave runup*, J. Waterway, Port, Coastal and Ocean Engin. *132*(2), 114–124.

YEE, H.C., WARMING, R.F., and HARTEN, A. (1983), *Implicit total variation diminishing (TVD) schemes for steady-state calculations*, Comp. Fluid Dyn. Conf. *6*, 110–127.

ZHANG, H., SHI, Y.L., LIU, M., WU, Z.L., and LI, Q.S. (2005), *A China-US collaborative effort to build a web-based grid computational environment for geodynamics*, Am. Geophys. Union, Fall (Abstract).

ZHANG HUAI, LIU MIAN, SHI YAOLIN, DAVID A. YUEN, YAN ZHENZHEN, and LIANG GUOPING, *Toward an automated parallel computing environment for geosciences*, Phy. Earth Planet. Inter., in press, accepted manuscript, available online 24 May 2007.

(Received October 31, 2006, revised August 1, 2007, Accepted August 6, 2007)
Published Online First: April 2, 2008

Pure appl. geophys. 165 (2008) 497–521
0033–4553/08/030497–25
DOI 10.1007/s00024-008-0325-9

Pure and Applied Geophysics

Geophysical Finite-Element Simulation Tool (GeoFEST): Algorithms and Validation for Quasistatic Regional Faulted Crust Problems

Jay Parker, Gregory Lyzenga, Charles Norton, Cinzia Zuffada, Margaret Glasscoe, John Lou, and Andrea Donnellan

Abstract—GeoFEST (Geophysical Finite Element Simulation Tool) is a two- and three-dimensional finite element software package for the modeling of solid stress and strain in geophysical and other continuum domain applications. It is one of the featured high-performance applications of the NASA QuakeSim project. The program is targeted to be compiled and run on UNIX systems, and is running on diverse systems including sequential and message-passing parallel systems. Solution to the elliptical partial differential equations is obtained by finite element basis sampling, resulting in a sparse linear system primarily solved by conjugate gradient iteration to a tolerance level; on sequential systems a Crout factorization for the direct inversion of the linear system is also supported. The physics models supported include isotropic linear elasticity and both Newtonian and power-law viscoelasticity, via implicit quasi-static time stepping. In addition to triangular, quadrilateral, tetrahedral and hexahedral continuum elements, GeoFEST supports split-node faulting, body forces, and surface tractions. This software and related mesh refinement strategies have been validated on a variety of test cases with rigorous comparison to analytical solutions. These include a box-shaped domain with imposed motion on one surface, a pair of strike slip faults in stepover arrangement, and two community-agreed benchmark cases: a strike slip fault in an enclosing box, and a quarter-domain circular fault problem. Scientific applications of the code include the modeling of static and transient co- and post-seismic earth deformation, Earth response to glacial, atmospheric and hydrological loading, and other scenarios involving the bulk deformation of geologic media.

1. Introduction

In order to simulate viscoelastic stress and flow in a realistic model of the Earth's crust and upper mantle, the modeling technique must be able to accommodate a complicated structure consisting of material layers and blocks, with imbedded faults which may cut at arbitrary angles. Stress and displacement features will vary most rapidly near the faults and particularly near fault-terminations. These features argue for fully three-dimensional finite-element modeling in the time domain. Two-dimensional modeling, semi-analytical techniques, finite-difference and semi-spectral methods either cannot model significant features or geometry of interest, or require gross oversampling in regions of little interest, leading to impossible computational requirements.

Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA 91109, U.S.A.
E-mail: Jay.W.Parker@jpl.nasa.gov

Finite-element modeling in three dimensions allows faithful modeling of complex faulting geometry, inhomogeneous materials, realistic viscous flow, and a wide variety of fault slip models and boundary conditions. While there are particular problems that are more efficiently expressed using alternative approaches, finite elements represent the most generally useful method for inhomogeneous elastostatic and viscoelastic problems. Because finite elements conform to (nearly) any surface geometry and support wide variations in mesh density, solutions may be made arbitrarily accurate with high computational efficiency. This flexibility comes with a price tag for the user or the tool-builder: That of generating and adapting the mesh of elements upon which the solution is computed. When such generation tools are primitive, researchers may spend substantially more time creating a mesh than solving their problem and interpreting the results. Therefore we describe automated tools for creating and adapting the mesh, including using an initial coarse mesh to generate a solution, whose computable error characteristics inform a further mesh generation cycle and produce an efficient and accurate solution.

GeoFEST uses stress-displacement finite elements to model stress and strain due to: 1) Elastostatic response to an earthquake event in the region of the slipping fault, 2) the time-dependent viscoelastic relaxation, and 3) the net effects from a series of earthquakes. The physical domain may be two- or fully three-dimensional and may contain heterogeneous rheology and an arbitrary network of faults. The software is intended to simulate viscoelastic stress and flow in a realistic model of the Earth's crust and upper mantle in a complex region such as the Los Angeles Basin, and has been used to explain features of the Ventura basin (HAGER *et al*., 1999) and the relaxation of the Great San Francisco Earthquake (GLASSCOE *et al*., 2006).

Many enhancements to GeoFEST have been recently added due to the NASA QuakeSim project (http://quakesim.jpl.nasa.gov), which has extended GeoFEST to the high-performance domain of message-passing parallel computer systems (PARKER *et al.,* 2003), added the functions of the PYRAMID parallel adaptive mesh refinement library (NORTON and CWIK, 2002; available at http://www.openchannelsoftware.org / projects/Pyramid), established availability of the source code with a no-fee license from http://www.openchannelsoftware.org /projects/GeoFEST, and placed it within a web-based problem solving environment (PIERCE *et al.,* 2003). The message-passing parallel code has been ported to many systems, including LINUX, HPUX, MacOS and SunOS.

Five annual workshops (under the name Workshop on Community Finite-Element Models for Fault Systems and Tectonic Studies, abbreviated CFEM Workshop) have been held for establishing computational benchmark cases for code validation and comparison, under the sponsorship of the Southern California Earthquake Center, the Computational Infrastructure for Geodynamics, and the National Aeronautics and Space Administration. Workshop material is currently hosted by http://www.geodynamics.org. These workshops have included model building and side-by-side comparison of modeling software, including definition of accuracy benchmark cases of community-wide interest.

Two of the validation cases presented here were formulated by that workshop. We do not here compare GeoFEST performance with other software, nonetheless that effort is in progress (for example, AAGAARD *et al*., 2006).

We describe the supported features of the GeoFEST code as of this writing in the next section. The precise description of the code's operation and mathematical specification follows in the Principles of Operation section. We give a detailed account of the relevant finite elements principles so that modelers from other fields (geodesy, tectonics) will better understand the abilities and limitations of GeoFEST, while finite-elements practitioners will see at once how the software may be classified among that diverse family of models. Recognizing that a software package must gain trust among potential users, we present the results of four comparisons between GeoFEST solutions and analytic solutions for equivalent problems.

2. *Features*

The primary quantity computed by GeoFEST is the displacement at each point in a domain. The stress tensor is also computed as a necessary byproduct. The computational domain represents a region of the Earth's crust and possibly underlying mantle. It is typically a square or rectangular domain in map view, with a flat upper free surface and constant depth, but the domain may deviate from this. The only requirement is that it be a bounded 3D domain with appropriate surface boundary conditions to render the problem well defined. These boundary conditions may be specified as surface tractions and/or displacements, which are usually specified on all surfaces and possibly on interior surfaces such as faults. Free surfaces have zero surface traction by definition. Faults are interior surfaces, and may have associated dislocation increments at set times, that is, a kinematic slip model. Dynamic or threshold-based slip is not currently implemented. The solid domain may contain layers or other distributions of material with associated rheological properties. Currently supported materials are isotropic elastic and viscoelastic (including linear and power-law Maxwell materials).

Elastostatic solutions are supported, such as computing the displacements and stresses immediately caused by a specified slip distribution on a fault or finding the interior displacement and stress distribution due to a surface traction or displacement. These solutions are not time-dependent.

Viscoelastic solutions are also supported, in which the material flows and relaxes in response to imposed stress, such as an earthquake event. One may compute the viscoelastic response to a single event, or to multiple events in a sequence. In the current version the sequence must be user-specified. Body forces are supported.

Boundary conditions and solutions apply to a finite-element discretized approximation to this domain. The domain is defined internally as a mesh of space filling tetrahedral or hexahedral elements, with three components of displacement at each mesh node constituting the solution. All numeric examples here utilize linear basis functions defined

on these elements. Stress is computed for each element, and is element-wise constant for the linear tetrahedral element type. Surface nodes carry special boundary conditions such as tractions or specified displacements. Nodes on faults are special split-nodes that define screw or tensile dislocation on the fault without perturbing the mesh geometry. Temporal evolution is by discrete time steps using an implicit solution technique, allowing large time steps without numerical instability.

The code may be used with all nodes, elements, faults, boundary conditions and time history control created or modified by word processor, constituting the only needed input file. For large meshes hand-construction becomes impractical, and we have sought to support tools for automated mesh generation. Initial attempts have focused on having the user specify fault rectangles, materials, and mesh density, and then semi-automated tools produce the ASCII input file.

Other supported features include:

- Specification of temporal epochs, each with differing steady boundary conditions
- Boundary velocity condition (steady change in the displacement, imposed)
- Controls for generating output on a subset of nodes/elements
- Control of implicit integration parameter
- Ability to shortcut temporal advance by fixing the sparse system for several time steps
- Control of check pointing, saving state and allowing restart
- An editable control file is checked at the completion of each iteration, allowing clean interruption by the user.

3. Principles of Operation

The domain of computation is a closed volume divided into simple volumes, typically tetrahedral in what follows, but generally other shapes such as hexahedra, wedges, or pyramids. The principal field of interest is the displacement, which is represented at nodes (such as the vertices of the tetrahedra) and assumed defined between the nodes by pre-arranged shape functions. Taking for example a linear tetrahedron, these are quite simple: For each node, define the function that is unity at that node and linearly declines to zero at the entire opposite face. We may also speak of global shape functions: For each element that shares a particular node, define the function that is unity at that node and declines linearly to zero at the opposite faces of each of the sharing tetrahedra. Given the values at the nodes, values in all space are defined by the shape functions as an interpolatory basis. Such interpolated displacements will be continuous across all element boundaries (tetrahedral facets).

Within the volume of an isolated finite element, appropriate derivatives of the element shape functions weight the spatial derivatives of the stress tensor (defined from the strain and generalized Hooke's law, relying on locally defined element rheology) to enforce local equilibrium. The shape functions constitute a finite degrees-of-freedom

approximation to the continuous system, and a volume integral enforces the equilibrium in a weighted average sense. Elements sharing a node contribute such weighted average terms to an equation for a single displacement. Away from boundary conditions and absent body forces, this set of displacement equation terms is set to zero. Body forces and surface tractions add forcing terms to the right-hand side.

The ensemble system of equations so defined is sparse, and if boundary conditions are sufficient the system is closed and solvable by standard sparse matrix techniques (typically by iterative techniques, although Crout factorization is supported on single-processor systems). Closed boundary conditions are usually easy to obtain; the user does have to ensure that enough of the boundary is constrained such that there are no unconstrained body translations or rotations permitted to the domain.

The solution is the elastostatic solution for the posed problem in terms of displacements at each node. Local linear combinations of these displacements with shape function derivatives yield the stress tensor in each element.

The elastostatic solution is required for any viscous relaxation computation. Once the static step is complete, the time evolution of quasi-static viscous relaxation may begin by computing the inelastic strain rate (termed "viscoplastic" in our derivation, in conformity with more general materials models; more general plastic behavior involving yield stress is not supported), which is directly determined by the stress and the viscosity parameters. Conceptually this rate adds a force term to the right-hand side of a sparse system similar to the elastostatic equilibrium. In practice, to obtain the advantages of implicit temporal development, terms are rearranged to modify the sparse equation coefficients as well. Each time step involves a solution to a sparse equation system, of similar cost to the elastostatic solution.

Faults are specified as either fault elements or as split nodes. With fault elements the user specifies fault properties such as failure criterion in an input record similar to that for a finite element. With split nodes the user specifies for each node on the fault its direction and amount of slip. The fault will slip at the initial time step, and also add the same increment of slip at each fault failure event, specified by the user. The split node formalism may be considered to represent the screw dislocation at a node as a separate entity from the displacement at that node, but is implemented here as an equivalent increment in the stress affecting the nodes immediately adjacent.

3.1. Viscoelastic Mechanics

We describe the quasi-static mathematical equations for viscoelastic materials, which is the assumed material type of the solid Earth being modeled. In the following, σ and ε denote second-order stress tensors for stress and strain fields, respectively, and u is the displacement field. The summation convention is used for repeated indices. A comma is used to denote a partial derivative with respect to a spatial dimension in a Cartesian coordinate system. For example, in R^3 the Cauchy stress tensor σ_{ij} has divergence

$$\sigma_{ij,j} \equiv \frac{\partial \sigma_{ij}}{\partial x_j} \equiv \sigma_{i1,1} + \sigma_{i2,2} + \sigma_{i3,3}. \tag{1}$$

When the two comma-separated indices are enclosed in parentheses, a symmetric combination is indicated, as in the total strain ε_{ij} defined as

$$\varepsilon_{ij} = u_{(i,j)} \equiv \frac{1}{2}\left(u_{i,j} + u_{j,i}\right), \tag{2}$$

We seek solutions to the quasistatic equilibrium equation:

$$\sigma_{ij,j} + f_i = 0, \tag{3}$$

where f_i is the given body force. The stress is subject to the constitutive equation

$$\sigma_{ij} = c_{ijkl}\varepsilon_{kl}^{elas}, \tag{4}$$

where c_{ijkl} is a tensor made up of material-specific constants, and ε^{elas} is the elastic strain. We assume viscoelastic relaxation may be modeled by a viscoplastic strain rate that is a known function of stress

$$\frac{\partial \varepsilon_{ij}^{vp}}{\partial t} = \beta_{ij}(\boldsymbol{\sigma}), \tag{5}$$

such as the function for Maxwell viscosity (expressed in (37–38)). Because this viscoplastic strain is defined as the inelastic part we assume total strain ε_{ij} from (2) is given by

$$\varepsilon_{ij} = \varepsilon_{ij}^{elas} + \varepsilon_{ij}^{vp} \tag{6}$$

(as in Owen and Hinton, 1980) and initially $\varepsilon^{vp} = 0$. Therefore, during relaxation time steps, (4) generalizes to

$$\frac{\partial \sigma_{ij}}{\partial t} = c_{ijkl}\left(\frac{\partial \varepsilon_{kl}}{\partial t} - \frac{\partial \varepsilon_{kl}^{vp}}{\partial t}\right). \tag{7}$$

In all cases the total strain ε_{ij} is fully specified by the displacement field according to (2).

The problem to be solved is formulated as an initial boundary value problem in a domain $\Omega \subset R^n$, where $n = 2$ or 3. This requires specified initial and boundary conditions. The spatial boundary may be partly type 1 ($\partial\Omega_1$, having displacements g_i) and partly type 2 ($\partial\Omega_2$, free surfaces with prescribed tractions h_i). Then the problem is to find a displacement field $u_i(\mathbf{x},t)$ and a stress tensor field $\sigma_{ij}(\mathbf{x},t)$ which satisfy equations (2–7) for all locations $\mathbf{x} \in \Omega$ and $t \in [0,T]$, $T > 0$, such that:

$$\begin{array}{rll} u_i(\mathbf{x},t) = u_{0i}(\mathbf{x}), & \mathbf{x} \in \Omega & \\ \sigma_{ij}(\mathbf{x},0) = \sigma_{0ij}(\mathbf{x}), & \mathbf{x} \in \Omega & \\ u_i(\mathbf{x},t) = g_i(\mathbf{x},t), & \mathbf{x} \in \Omega_1, & t \in [0,T] \\ \sigma_{ij}n_j = h_i(\mathbf{x},t), & \mathbf{x} \in \Omega_2, & t \in [0,T] \end{array}, \tag{8}$$

where u_{0i} and σ_{0ij} are the initial displacement and stress fields, respectively, n_j is an outward normal vector to $\partial\Omega_2$, and $g_i(\mathbf{x},t)$ and $h_i(\mathbf{x},t)$ are prescribed boundary displacement and tractions, respectively.

For isotropic material, the material constants in (4) can be expressed as:

$$c_{ijkl} = \mu(\mathbf{x})\left(\delta_{ik}\delta_{jl} + \delta_{il}\delta_{jk}\right) + \lambda(\mathbf{x})\delta_{ij}\delta_{kl}, \tag{9}$$

where λ and μ are known as *Lamé parameters*, which are related to Young's modulus, E and Poisson's ratio, v by

$$\lambda = \frac{vE}{(1+v)(1-2v)}, \quad \mu = \frac{E}{2(1+v)}. \tag{10}$$

3.2. Finite-Element Formulation

Equilibrium Formulation. Here we show the formulation for the initial elastic equilibrium problem, which represents an instantaneous condition. Many of the components will prove useful to the general relaxation problem, which has a time dimension.

In a finite-element approximate solution to (3), we seek an approximate displacement field $u_i(\mathbf{x},t) \in S$, where S is a finite-dimensional trial solution space with each u_i in S satisfying $u_i = g_i$ (the essential boundary condition) on $\partial\Omega_1$. We also define a finite-dimensional variation space V with each $w_i \in V$, satisfying $w_i = 0$ on $\partial\Omega_2$.

The following is the variational form of (3) subject to (8) (HUGHES, 2000). Find $u_i \in S$ such that for all $w_i \in V$

$$\int_{\Omega} w_{(i,j)}\sigma_{ij}\, d\Omega = \int_{\Omega} w_i f_i\, d\Omega + \sum_{i=1}^{n}\left(\int_{\partial\Omega_2} w_i h_i\, d\Omega\right), \tag{11}$$

where σ_{ij} is related to u_i through (2) and (4), and n is the spatial dimension. The field w_i is sometimes referred to as *virtual displacements* in solid mechanics. Under some mild smoothness assumptions on the involved variables, it can be shown that a solution to (9) is a solution to (3, 8) and *vice versa*.

To find a numerical solution to the finite-element problem (9), all the variables and the integral equation in (9) are discretized on a *finite-element mesh*. In the GeoFEST program implementation, the discrete displacement field u_i^h is defined at nodal points of the mesh, and stress field σ_{ij}^h and strain field ε_{ij}^h are defined at integration points inside each mesh element.

HUGHES (2000) indicates the combination $w_{(i,j)}\ \sigma_{ij}$ has important symmetries which lead to a convenient reduced expression which is part of our implementation, and allows the compact expressions in the viscoelastic algorithms below. From (4)

$$w_{(i,j)}\sigma_{ij} = w_{(i,j)}c_{ijkl}u_{(k,l)}. \tag{12}$$

The material properties c_{ijkl} are subject to major and minor symmetry conditions. This leads to a collapsed index notation, in which the elements of the stress and strain tensors may each be expressed as a 3×1 array in R^2 (or 6×1 in R^3), while c_{ijkl} may be replaced by a 3×3 array (or 6×6 for R^3). Use of capitalized indices indicates this collapsed index notation.

Briefly, define ε_I as entries of the collapsed strain array $\boldsymbol{\varepsilon}$ as

$$\varepsilon_J : \boldsymbol{\varepsilon}(\mathbf{u}) = \left\{ \begin{array}{c} u_{1,1} \\ u_{2,2} \\ u_{1,2} + u_{2,1} \end{array} \right\}, \quad J \in (1,2,3). \tag{13}$$

Note that the third component of ε_I is twice the value of ε_{12} from (2); this is a convenient definition as (in the combination (12)) it accounts for a redundancy in c_{ijkl}.

Similarly

$$\sigma_I : \boldsymbol{\sigma}(\mathbf{u}) = \left\{ \begin{array}{c} \sigma_{11} \\ \sigma_{22} \\ \sigma_{12} \end{array} \right\}, \quad I \in (1,2,3). \tag{14}$$

The symmetries of c_{ijkl} and the form of the strain tensor indicate many redundant and many zero terms in (12). These may be exploited by defining the D_{IJ} entries of $\mathbf{D}$ (in R^2) as

$$D_{IJ} : \mathbf{D} = \begin{bmatrix} c_{1111} & c_{1122} & c_{1112} \\ c_{1122} & c_{2222} & c_{2212} \\ c_{1112} & c_{2212} & c_{1212} \end{bmatrix} \tag{15}$$

and so

$$\sigma_I = D_{IJ}\varepsilon_J. \tag{16}$$

Then the left side of (11) may be expressed in the equivalent form

$$\int_\Omega w_{(i,j)}\sigma_{ij}d\Omega = \int_\Omega \boldsymbol{\varepsilon}^T(\mathbf{w})\mathbf{D}\boldsymbol{\varepsilon}(\mathbf{u})\, d\Omega \equiv a(\mathbf{w},\mathbf{u}) \tag{17}$$

where we define the vector space inner product a (,) here for later use. The isotropic case reduces to

$$\mathbf{D} = \begin{bmatrix} \lambda + 2\mu & \lambda & 0 \\ \lambda & \lambda + 2\mu & 0 \\ 0 & 0 & \mu \end{bmatrix}, \tag{18}$$

Similar expressions apply for ε_I, σ_J, and D_{IJ} in R^3.

Candidate solutions u_i^h consist of all linear combinations of locally-defined basis functions N_a(x) (that is, the scalar basis functions form an expansion for each spatial coordinate of $\mathbf{u}$, and may be linearly varying within elements as well as more general functions). N_a takes unit value at node a and vanishes on neighboring nodes of a; d_a is the displacement coefficient at node a, which indicates the value of one component of displacement, an unknown to be computed. To properly match the boundary conditions, we use the basis functions N_a to define the behavior adjacent to $\partial\Omega_1$ nodes in terms of the specified values g_a:

$$g_i = \sum_{a \in \eta_{eb}} N_a g_a \tag{19}$$

where η_{eb} is the set of nodes on $\partial\Omega_1$. The set of nodes on the mesh is A. Solve for displacement coefficients d_a, forming a solution away from the $\partial\Omega_1$ boundary as

$$v_i^h = \sum_{a \in (A - \eta_{eb})} N_a d_a, \tag{20}$$

creating the global solution u_i^h

$$u_i^h = v_i^h + g_i^h. \tag{21}$$

Also let

$$w_i^h = \sum_{a \in (A - \eta_{eb})} N_a c_a, \tag{22}$$

where c_a (entries of $\mathbf{c}$) are arbitrary constants.

Substituting the previous definitions into (11), it follows that

$$\mathbf{c}(\mathbf{Kd} - \mathbf{F}) = 0, \tag{23}$$

but since this holds for any arbitrary c_a then $\mathbf{d}$ (with entries d_a) must be the solution to

$$\mathbf{Kd} = \mathbf{F}, \tag{24}$$

where $\mathbf{K} = [k_{pq}] \in R^{m \times m}$ is called the stiffness matrix. $\mathbf{K}$ (arising from $a(\mathbf{w},\mathbf{u})$) is symmetric and positive definite, and the system dimension m (the range of indices p, q) is the sum over nodes, of the free components of $\mathbf{u}$ at each node. The index p indicates a coordinate component $w_{i(p)}$ at a node $a(p)$, while q indicates a coordinate component $v_{j(q)}$ at a node $b(q)$. An entry of matrix $\mathbf{K}$, k_{pq}, has the form

$$k_{pq} = a\left(N_a \mathbf{e_i}, N_b \mathbf{e_j}\right) = \mathbf{e_i^T} \int_{\Omega} \mathbf{B_a^T D B_b}\, d\Omega \mathbf{e_j}, \tag{25}$$

where $\mathbf{e_i}$ (or $\mathbf{e_j}$) is the coordinate unit vector associated with equation p (or q), and $\mathbf{B_a}$ is the strain-displacement mapping determined by substituting N_a $\mathbf{e_i}$ into (13) for $\mathbf{u}$. Therefore $\mathbf{B_a}$ in R^2 has dimensions 3×2, (that is, collapsed index and spatial dimension) and may be indexed with the mixed notation $B_{a:Ii}$:

$$B_{a:Ii} = \varepsilon_I(N_a \mathbf{e_i}). \tag{26}$$

Term $\mathbf{F}$ on the right-hand side of (24) is a vector in R^m made of quantities directly determined by the problem definition. It includes the contributions from the body force and boundary condition terms, and is easily derived from the right side of (11) substituting the basis function expansion (22). There is also a term involving g_i, that

arises from substituting (21) into the left side of (11), followed by moving terms in g_i to the right side.

Solving the linear system (24) yields the displacement field coefficients d_a (the values at the nodes), which defines the displacement u_i^h in all the domain via (19–21). Stress is evaluated subsequently, using (16).

Time-Stepping Formulation. In order to determine the displacements u_i and stresses σ_{ij} after the initial equilibrium solution, we repeatedly solve for changes in these quantities over discrete steps in time. Denote changes from one time step to the next with the prefix δ, and define $\delta u^{(n+1)}$, such that

$$u_i^{(n+1)} = u_i^{(n)} + \delta u_i^{(n+1)}, \tag{27}$$

where the superscript (n) indicates the value at discrete time t_n. For clarity we show the formulation in the limit of small steps (suitable for a first-order truncated Taylor expansion in time; this limitation is relaxed in the implicit scheme shown below).

From (3) linearity of (1) implies

$$\delta\sigma_{ij,j} + \delta f_i = 0, \tag{28}$$

while (7) results in

$$\delta\sigma_{ij} = c_{ijkl}(\delta\varepsilon_{kl} - \Delta t\beta_{kl}(\boldsymbol{\sigma})). \tag{29}$$

Therefore the variational form (compare (11)) is

$$\int_\Omega w_{(i,j)} c_{ijkl}\,(\delta\varepsilon_{kl} - \Delta t\beta_{kl}) d\Omega = \int_\Omega w_i \delta f_i d\Omega + \sum_{i=1}^{n} \left(\int_{\partial\Omega_2} w_i \delta h_i \, d\Omega \right). \tag{30}$$

The term in β_{kl} does not depend on δu, so we move it to the right side, arriving at

$$a(\mathbf{w}, \delta\mathbf{u}) = \delta f + F_2, \tag{31}$$

where δf is due to changes in boundary conditions and forces (the right side of (30) plus a term due to δg_i), and

$$F_2 = \int_\Omega w_{(i,j)} c_{ijkl} \Delta t \beta_{kl} \, d\Omega \tag{32}$$

is a force-like term that accounts for the inelastic relaxation.

Sampling via finite-element basis function as in (19–22) results in the explicit scheme similar to that in HUGHES and TAYLOR (1978, Table 1), but stability requires tediously small steps. Instead, we implement a variant of the implicit scheme of HUGHES and TAYLOR (1978, Table 2). In this case stability is ensured, while accuracy depends on limited step size (typically much larger than for the explicit solution) that depends on the problem features.

An Implicit Time-Stepping Scheme. GeoFEST adopted an implicit scheme because of its unconditional numerical stability with respect to time step sizes, but supports a user-selected implicit fraction α, which may be set in the range 0 (fully explicit) to 1 (fully implicit). For problems with fault slip followed by relaxation, we have generally used $\alpha = 1$ and have seen little need to use other values; HUGHES and TAYLOR (1978) show cases suggesting that other values of α may be useful for problems with more complicated viscoplastic models. They also include an inner iteration step which we do not currently implement, which may be useful for nonlinear problems. GeoFEST support for nonlinear problems uses a dynamic time step control, which we do not detail as it is not well tested.

The entire solution process begins with an initial solution of a pure elastic problem for which the viscoplastic strain rate is set to zero, as formulated in (11–26). The pure elastic solution provides an initial stress field, which is then relaxed over a time period in a viscoelastic solution for which an implicit stepping scheme is used. This algorithm is used by GeoFEST for each time step (HUGHES and TAYLOR, 1978, Table 2 simplified):

1. Form step stiffness matrix and right-hand side

$$\begin{aligned} k_{pq}^{(n+1)} &= \int_{\Omega} B_{Ii(p)} \left[(\mathbf{S} + \alpha \Delta t \boldsymbol{\beta}')^{-1} \right]_{IJ} B_{Jj(q)} \, d\Omega, \\ F_p^{(n+1)} &= \int_{\Omega} B_{Ii(p)} \left[(\mathbf{S} + \alpha \Delta t \boldsymbol{\beta}')^{-1} \right]_{IJ} \left(\Delta t \beta_J^{(n)} \right) d\Omega + \delta f_p^{(n+1)}, \end{aligned} \tag{33}$$

where $\mathbf{S} = \mathbf{D}^{-1}$, $0 < \alpha < 1$, $\delta f^{(n+1)}$ expresses changes in boundary conditions and forces over the time step. Note the $\Delta t \, \beta_J$ term in $F_p^{(n+1)}$ comes from (7), just as in (32).

2. Solve

$$\mathbf{K}^{(n+1)} \mathbf{d} = \mathbf{F}^{(n+1)} \tag{34}$$

which determines $\delta u^{(n+1)}$ from (19–21).

3. Construct stress increment:

$$\delta \sigma_I^{(n+1)} = \left[\left(\mathbf{S} + \alpha \Delta t \beta'^{(n)} \right)^{-1} \right]_{IJ} \left(B_{Jj} \delta u_j^{(n+1)} - \Delta t \beta_J^{(n)} \right). \tag{35}$$

4. Update displacement and stress fields:

$$\begin{aligned} u_i^{(n+1)} &= u_i^{(n)} + \delta u_i^{(n+1)}, \\ \sigma_{ij}^{(n+1)} &= \sigma_{ij}^{(n)} + \delta \sigma_{ij}^{(n+1)}. \end{aligned} \tag{36}$$

In (33, 35) $\boldsymbol{\beta}^{(n)}(\sigma)$ represents the Maxwell viscoplastic strain rate, and $\boldsymbol{\beta}'^{(n)}(\sigma)$ its Jacobian matrix. Dropping the superscript (n), in R^2 (the R^3 form is similar) they are:

$$\beta_I(\sigma) = \frac{\kappa^{\chi-1}}{4\eta}\begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 4 \end{bmatrix}\begin{bmatrix} \sigma_{xx} \\ \sigma_{yy} \\ \sigma_{xy} \end{bmatrix},$$
$$\beta'_{Ij}(\sigma) = \frac{\kappa^{\chi-1}}{4\eta}\begin{bmatrix} a & -a & c\sigma_{xy} \\ -a & a & -c\sigma_{xy} \\ c\sigma_{xy} & -c\sigma_{xy} & 4b \end{bmatrix}, \tag{37}$$

where

$$\kappa = \sqrt{\left(\frac{\sigma_{xx} - \sigma_{yy}}{2}\right)^2 + \sigma_{xy}^2}$$
$$a = 1 + (\chi - 1)\left(\frac{\sigma_{xx} - \sigma_{yy}}{2\kappa}\right)^2,$$
$$b = 1 + (\chi - 1)\left(\frac{\sigma_{xy}}{\kappa}\right)^2, \tag{38}$$
$$c = (\chi - 1)\left(\frac{\sigma_{xx} - \sigma_{yy}}{\kappa^2}\right),$$

where χ is the exponent for power-law viscosity (linear viscosity holds where $\chi = 1$). Note this has the effect of relying on the value of $\boldsymbol{\beta}$ and $\boldsymbol{\beta}'$ at the start of the time step; changes in σ_{ij} will have a magnified effect for $\chi > 1$. Therefore small time steps may be needed at critical times to obtain accurate solutions, to the degree that $\delta\boldsymbol{\sigma}$ and χ are large.

Solutions σ_{ij} to (36) do not exactly satisfy equilibrium (24), and as many time steps make their contributions this can give rise to substantial errors. We have found it inexpensive and useful to form a correction term

$$f_p^{(res)} = \int_\Omega B_{i(p)I}\sigma_I \, d\Omega, \tag{39}$$

creating a force-like vector that is subtracted from the succeeding time step (34), a kind of residual force correction (Owen and Hinton, 1980).

Fault Specification and Split Node Implementation. Fault conditions can be specified either as fault elements or as split nodes. With fault elements, one can specify fault properties such as failure criterion (not currently supported). With split nodes, one can represent the rate of displacement of a fault surface by assigning the direction and amount of slip for each node on the fault surface. Typically, a split node has different slip rates assigned to it on each side of the fault surface, which introduces a discontinuity in the displacement field to simulate real fault slip (Melosh and Raefsky, 1981). This idea can be illustrated by a simple one-dimensional example with two elements, as shown in Figure 1.

It is assumed that elements 1 and 2 (E1, E2) are located adjacent to the opposite sides of the fault surface represented by a dash line between the two elements, and u

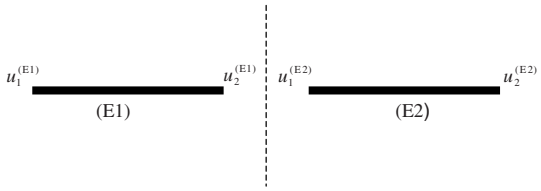


Figure 1
Two bar elements (one-dimensional) joined at a split node. Dashed line indicates split node, where single node in mesh has distinct displacement values $u_1^{(E2)}$ and $u_2^{(E1)}$. Subscript here denotes the node numbered local to the element.

is the displacement field (here a scalar field, but easily generalizing to vector field u_i). Away from the fault, u has a single value defined at each node of the 1-D finite element mesh, such as $u_1^{(E1)}$ on the left node of (E1) and $u_2^{(E2)}$ on the right node of (E2). The node between the two elements is considered a split node since it straddles the fault. The displacement field has different values at the split node, which are $u_2^{(E1)}$ on the side of element (E1) and $u_1^{(E2)}$ on the side of element (E2). Specifically we can write:

$$\begin{aligned} u_2^{(\mathrm{E1})} &= \overline{u_2^{(\mathrm{E1})}} + \Delta u_2^{(\mathrm{E1})}, \\ u_1^{(\mathrm{E2})} &= \overline{u_2^{(\mathrm{E2})}} + \Delta u_2^{(\mathrm{E1})}, \end{aligned} \tag{40}$$

where $\overline{u_2^{(\mathrm{E1})}} = \overline{u_1^{(\mathrm{E2})}}$ is the mean value of displacement at the split node, and $\Delta u_2^{(\mathrm{E1})} = -\Delta u_1^{(\mathrm{E2})}$ is the splitting part of displacement that has opposite signs on two sides of the fault. In a finite-element implementation, the contribution from the splitting displacements can be formulated as an additional forcing term. This fact can also be shown using the two-element example. The local stiffness matrix for element 1 can be written as:

$$\begin{bmatrix} k_{11}^{(\mathrm{E1})} & k_{12}^{(\mathrm{E1})} \\ k_{21}^{(\mathrm{E1})} & k_{22}^{(\mathrm{E1})} \end{bmatrix} \begin{bmatrix} u_1^{(\mathrm{E1})} \\ \overline{u_2^{(\mathrm{E1})}} + \Delta u_2^{(\mathrm{E1})} \end{bmatrix} = \begin{bmatrix} f_1^{(\mathrm{E1})} \\ f_2^{(\mathrm{E1})} \end{bmatrix} \tag{41}$$

which relates local displacements to local force terms. By moving the known quantities of the above equation to the right-hand side, we have:

$$\begin{bmatrix} k_{11}^{(\mathrm{E1})} & k_{12}^{(\mathrm{E1})} \\ k_{21}^{(\mathrm{E1})} & k_{22}^{(\mathrm{E1})} \end{bmatrix} \begin{bmatrix} u_1^{(\mathrm{E1})} \\ \overline{u_2^{(\mathrm{E1})}} \end{bmatrix} = \begin{bmatrix} f_1^{(\mathrm{E1})} - k_{12}^{(\mathrm{E1})} \Delta u_2^{(\mathrm{E1})} \\ f_2^{(\mathrm{E1})} - k_{22}^{(\mathrm{E1})} \Delta u_2^{(\mathrm{E1})} \end{bmatrix}. \tag{42}$$

Similarly for element 2, we have:

$$\begin{bmatrix} k_{11}^{(\mathrm{E2})} & k_{12}^{(\mathrm{E2})} \\ k_{21}^{(\mathrm{E2})} & k_{22}^{(\mathrm{E2})} \end{bmatrix} \begin{bmatrix} u_1^{(\mathrm{E2})} \\ \overline{u_2^{(\mathrm{E2})}} \end{bmatrix} = \begin{bmatrix} f_1^{(\mathrm{E2})} - k_{12}^{(\mathrm{E2})} \Delta u_1^{(\mathrm{E2})} \\ f_2^{(\mathrm{E2})} - k_{22}^{(\mathrm{E2})} \Delta u_1^{(\mathrm{E2})} \end{bmatrix}. \tag{43}$$

Assembling the local stiffness matrix equations into a global stiffness matrix equation, we obtain:

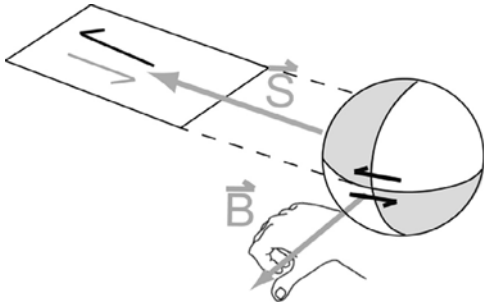


Figure 2
Right-hand rule convention used for defining the direction of the B vector for a given fault orientation and sense of slip.

$$\begin{bmatrix} k_{11}^{(E1)} & k_{12}^{(E1)} & 0 \\ k_{21}^{(E1)} & k_{22}^{(E1)} + k_{12}^{(E2)} & k_{12}^{(E2)} \\ 0 & k_{21}^{(E2)} & k_{22}^{(E2)} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} f_1 - k_{12}^{(E1)} \Delta u_2^1 \\ f_2 - k_{22}^{(E1)} \Delta u_2^{(E1)} - k_{11}^{(E2)} \Delta u_1^{(E2)} \\ f_3 - k_{21}^{(E2)} \Delta u_1^{(E2)} \end{bmatrix} \tag{44}$$

where u_i's (no superscript) are global displacements, which are related to the node local displacements by

$$u_1 = u_1^{(\mathbf{E}1)}, \quad u_2 = u_2^{(\mathbf{E}1)} = u_1^{(\mathbf{E}2)}, \quad u_3 = u_2^{(\mathbf{E}2)}. \tag{45}$$

The global force terms f_i are related to the local ones by

$$f_1 = f_1^{(\mathbf{E}1)}, \quad f_2 = f_2^{(\mathbf{E}1)} + f_1^{(\mathbf{E}2)}, \quad f_3 = f_2^{(\mathbf{E}2)}. \tag{46}$$

Equations (41–44) show that the effect of the slips on the split nodes is equivalent to adding the specified additional terms on the right-hand side of the finite-element matrix equations.

In three dimensions, the fault slip at each node is specified (e.g., in the input file) by two vectors, a slip direction (unit vector) **S**, a reference unit vector **B**, and a slip amplitude *s*. The vectors are related by the right-hand rule convention as shown in Figure 2.

As in the one-dimensional case above, an adjustment to the right-hand side is created with this information, with contributions coming from each element that shares the split node. Thus for element *e* the triple-product is formed of the vectors **S**, **B**, and **C**, where **C** is the vector from the split node to the center of *e*. The sign of this quantity determines the sign of the application of slip, which is applied to the finite-element system right-hand side. Each slip component at the split node contributes a set of terms (one for each adjoining finite element) to the right-hand side of every adjacent degree of freedom, as illustrated in (44).

Stress and displacement at each time are the accumulations of incremental stresses and displacements for past time steps. When a slip event occurs, the incremental displacements are found by applying the split nodes adjustments to the right-hand side of the stiffness equation. After the incremental displacement is obtained, the incremental stress is found by including the split node contribution to the stress for that time step. In this way the displacement and stress effects of a slip event are correctly carried forward into future time steps, without any need for additional storage for the slip history of the fault.

4. Validation

Domains for four validation studies are shown in Figure 3 (from above, or map view: All four problems have a substantial depth). Although GeoFEST computes with units–free values, for definiteness we will refer to spatial dimensions in km, and displacements (including fault slip) in m.

4.1. Simple Download Cases

Three simple validation examples are available for download with the GeoFEST code to offer an immediate test of a correct, fresh build from the downloaded source code. The first two ("valid1" and "valid2", which we shall not discuss in detail) verify correct response of a small cube model to a compressive displacement imposed on one face, and a second type of compression due to a constant body force.

The third download validation case ("valid3") displays the effects of viscoelastic relaxation. At $t = 0$, and following, an imposed displacement U_x is applied on the $z = 0$ surface. The elastic solution at $t = 0$ is computed, establishing a pattern of stress in the solid which subsequently relaxes according to the time-stepping algorithm (27–30). The hexahedral-element mesh fills a cube-shaped domain, has 2197 nodes, 1728 elements with material values $\mu = \lambda = 30$, $\eta = 31.69$ with power-law viscosity exponent $n = 1$. This material has Maxwell time τ = (approximately) 1, identical in all elements. The boundary conditions are simple shear (an imposed value of 1 for U_x on the $z = 0$ surface), followed by static relaxation. The simulation runs 100 time steps of 0.1 (approximately 10 τ). One may regard these quantities as unitless, hence the solution represents a family of scenarios. The set of values given above may represent a problem with $\mu = \lambda = 30$ GPa, $\eta =$ 1.00e18 GPa-s, $\tau \sim 1$ year, $U_x = 1$ m, duration 100 time steps of 0.1 year, with resulting stress reported in MPa. Figure 4 shows the relaxation of the σ_{xz} component of stress, following closely the expected exponential decay.

4.2. Reverse Fault

Another validation study is a reverse fault in a cube-shaped domain, and is specified as "Reverse Fault Benchmark, no gravity" (or Benchmark 5b) by the CFEM Workshop

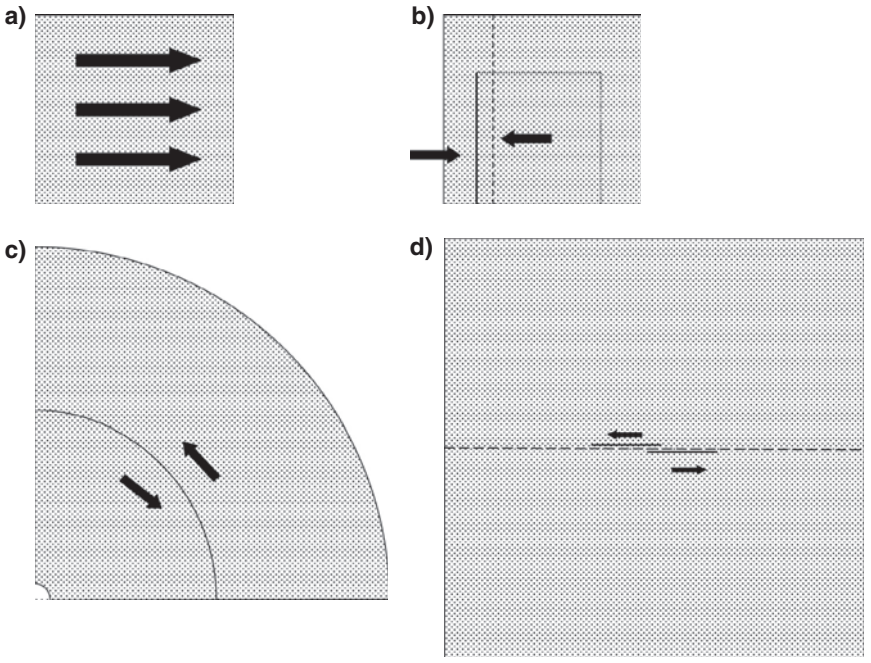


Figure 3
Map view of validation test cases detailed below. a) Viscoelastic 24 km cube with top face initially sheared. b) Reverse fault (Benchmark 5b), 45 degree dip fault in 24 km cube domain. Surface projection of sloping fault is indicated by shaded rectangle. Dashed line shows location of surface solution comparison in Figure 6. c) Circular fault, 100 km radius in 10 to 200 km annulus, 40 km domain depth. d) Stepover case, two 10 km parallel faults offset by 1 km with 1 km overlap, in 60 km square domain, 30 km domain depth. Dashed line shows location of surface solution comparison in Figure 10.

(Fig. 3b). By focusing on the initial elastic solution we are able to demonstrate fidelity with the analytic solution to the specified fault slip in an elastic half space. This test case is designed to test the accuracy of the solution near the fault tip, and is particularly challenging for using a small domain that has substantial solution strain at the truncation boundaries. The mesh is supplied courtesy of Carl Gable using the Los Alamos LaGrit software (http://lagrit.lanl.gov/), and has been used by several workshop participants; it uses 5772 nodes and 29412 elements, all of roughly the same size (typical edge length is about 0.75 km). The domain is a cube 24 km on a side, with uniform elastic material. The boundary conditions on four sides are samples of the elastic half-space analytic solution. The top boundary is a free surface, and one side (at $y = 0$) is a mirror plane (free in the tangential direction). The fault is a 45° dipping fault extending from the mirror plane at $y = 0$ to $y = 16$, with the top edge at the surface and $x = 4$ km. The bottom edge at $x = 20$ km. The slip is prescribed to be uniform (1 m thrust) for the top 12 km and

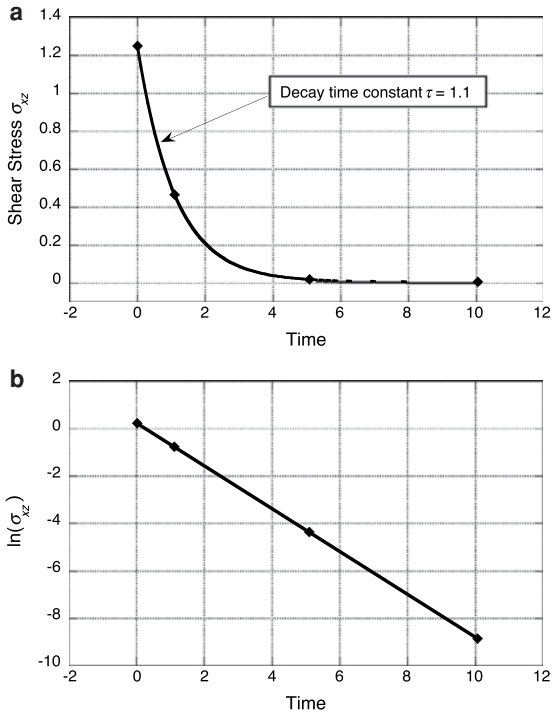


Figure 4
(a) Time evolution of shear stress in 'valid3' GeoFEST validation calculation. Best fitting exponential curve yields a decay time constant of 1.1, while the theoretical Maxwell relaxation time for the constitutive model is $\tau = 1.06$. Shear stress is unitless but may be taken as MPa (see text). (b) Logarithmic plot of the shear stress evolution with linear fit, showing close correspondence to expected exponential dependence.

$0 < y < 12$ km, and varies according to a linear taper in the depths from 12 to 16 km and also $12 < y < 16$.

Figure 5 shows the z-component of the solution displayed on the cube-shaped domain surface; the mirror-plane edge and top of the fault geometry are clearly seen. Some of the roughness of contours on the mirror plane is an artifact of displaying ambiguous split node values in an interpolated contour plot. Figure 6 shows comparisons of the three components of the solution and analytic half-space solution on the top surface along a line at $x = 6$. Analytic solutions here and in following validations are according to the QuakeSim application "Disloc," which implements equations from OKADA (1985). In light of the facts that the solution is weakly singular at the fault tip, the element size there is not adapted to the solution features, and that all error must be accommodated within the small domain, the degree of agreement is excellent.

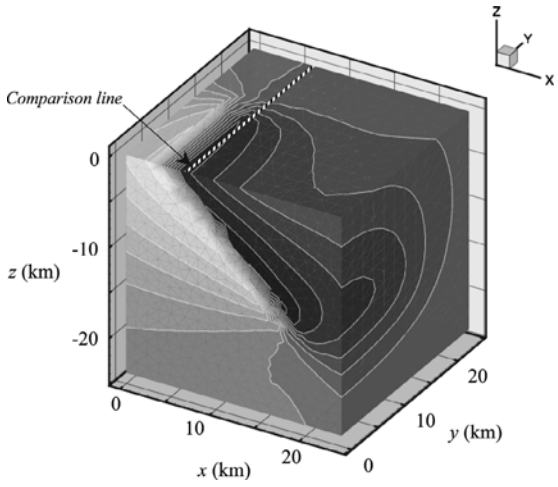


Figure 5
Reverse fault (Benchmark 5b elastic): z component displayed on surface of meshed domain. Dashed line indicates location of solution comparison in Figure 6.

4.3. Circular Fault

Another validation case from the CFEM Workshop solves for a slipping circular fault, represented on a quarter-disk shaped domain with boundaries appropriate for circular symmetry (Benchmark 7). The circular fault allows a test of solution quality for linear tetrahedral elements that cannot represent the geometry exactly. The solution also has the notable feature that the deformations inside and outside the circle are not of similar amplitude, due to the fault curvature and superposition of elemental solutions. We solve an elastic problem on the domain shown in Figure 3c, a quarter-disk with a circular fault of radius 100 km, in a radius 200 km domain. The fault is 16 km deep in a 40 km deep layer and is subjected to uniform 1 m strike-slip for the first 12 km of depth, tapering to zero in the remaining depth to 16 km. The boundary conditions for displacement on the bottom and circumference allow no motion; while the flat surfaces allow perpendicular displacement, hence use symmetry to match a full-circle disk solution. The elements are arranged mechanically by subdividing the quarter-disk into annuli and sectors, and finally dividing the small nearly-cube domains into 5 tetrahedra. In this way the mesh consists of 145866 nodes, 675000 elements.

The x component of the solution is displayed on the surface of the mesh in Figure 7. Note that it correctly tapers to zero at $y = 0$, and is larger in amplitude toward the inside of the fault.

In Figure 8 we compare the tangential component of the solution to an analytic half-space solution, although the finite-element domain is subject to fixed conditions at the

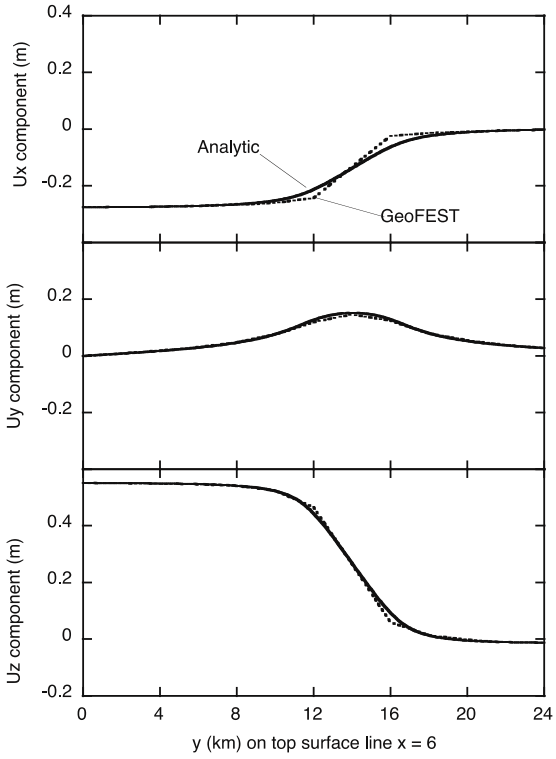


Figure 6
Reverse fault (Benchmark 5b elastic): Components of analytic and GeoFEST solutions along line $x = 6$ km, top surface.

outer and bottom boundaries, which is so close as to introduce notable error. The solutions agree in qualitative features, including the larger amplitude in the interior of the circular fault. This demonstrates that curved faults can be approximated by straight-edged elements and linear basis functions, to a reasonable degree.

4.4. Two-Fault Stepover

The final validation case uses a pair of uniform slip rectangular faults arranged in a stepover configuration (Fig. 3d). The faults run in the x direction, the first from -9 to 1 km (at y = 0.5 km), the second from -1 to 9 km (at y = -0.5 km); the elastic problem is solved with uniform slip of 1 m (right lateral) on both faults. The complexity of the solution in the stepover vicinity suggests this as a problem of interest for examining

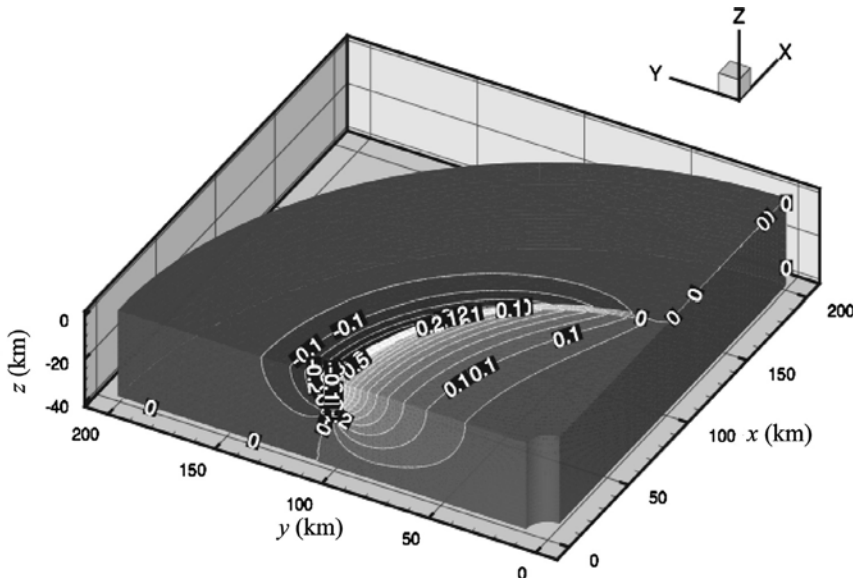


Figure 7
Circular fault (Benchmark 7): *x*-component of solution on surface of mesh.

the convergence rate for a series of locally refined meshes using an automated technique. The initial mesh represents the 1 km distance between the faults with a single finite-element: clearly a poor representation for obtaining any details there. Refinement was obtained using the adaptive *h*-refinement method of POLSTYANKO and LEE (2001) and uses an adaptation of their software. But in place of their (electromagnetics-based) error functional we apply a functional defined as the strain energy in each finite element from the elastic solution at each refinement level. That is, from the initial unreasonably coarse mesh we compute a poor-quality solution, and find the strain energy E_e within each finite element, defined as

$$E_e = a(\mathbf{u}, \mathbf{u}) = \int_{\Omega} \boldsymbol{\varepsilon}(\mathbf{u})^{\mathrm{T}} \mathbf{D} \boldsymbol{\varepsilon}(\mathbf{u})\, d\Omega, \tag{47}$$

where the integration is taken over the volume of element e.

This functional associated with each element is used to create an ordered list of priority for refinement for each element. Each refinement step adds about 35% more elements to the mesh in the priority region indicated by these elements, and the new mesh (Level 2) is used to compute a fresh solution. We repeat, using the new solution to determine E_e, and so obtain the Level 3 refined mesh, and so on for levels up to Level 6. The z-component of the surface solution for Level 6 is shown in Figure 9, with by-component comparisons to the analytic solution on the midline shown in Figure 10.

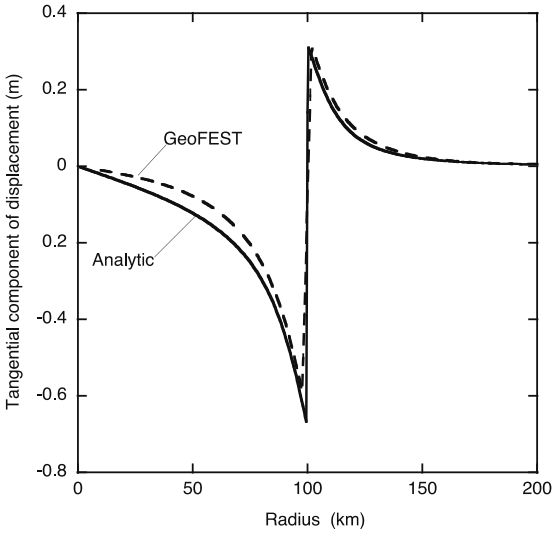


Figure 8
Circular fault (Benchmark 7): tangential component along y = 0 line, top surface, comparing GeoFEST and analytic solution.

The overall quality and convergence is shown in Figure 11, which shows the RMS difference (that is, the error) between the finite element solution and an analytic half-space solution for the same pair of uniform slip faults along the midline. This is computed for each refined mesh by summing the norm squared of the vector differences at each node location on the surface midline ($y = 0$), normalizing by the node count and taking the square root of the result. The convergence of the RMS error toward zero appears to be roughly a power-law with the number of elements, with a fit value of exponent of -1.8, but consistent with -2 for levels 1–4. It is likely that convergence stalls past level 6 due to the finite size of the domain and the differences in representation of the uniform slip due to the finite elements. Power-law exponents in this range indicate a convergence rate far faster (as a function of element count) than one would have by applying uniformly denser meshes at each stage (where error might be expected to asymptotically follow h^2 for characteristic element length h, and therefore element count to a power of $-2/3$).

5. Conclusion

The equations and methods for the GeoFEST application (available for source download at http://www.openchannelsoftware.org/projects/GeoFEST) include capabili-

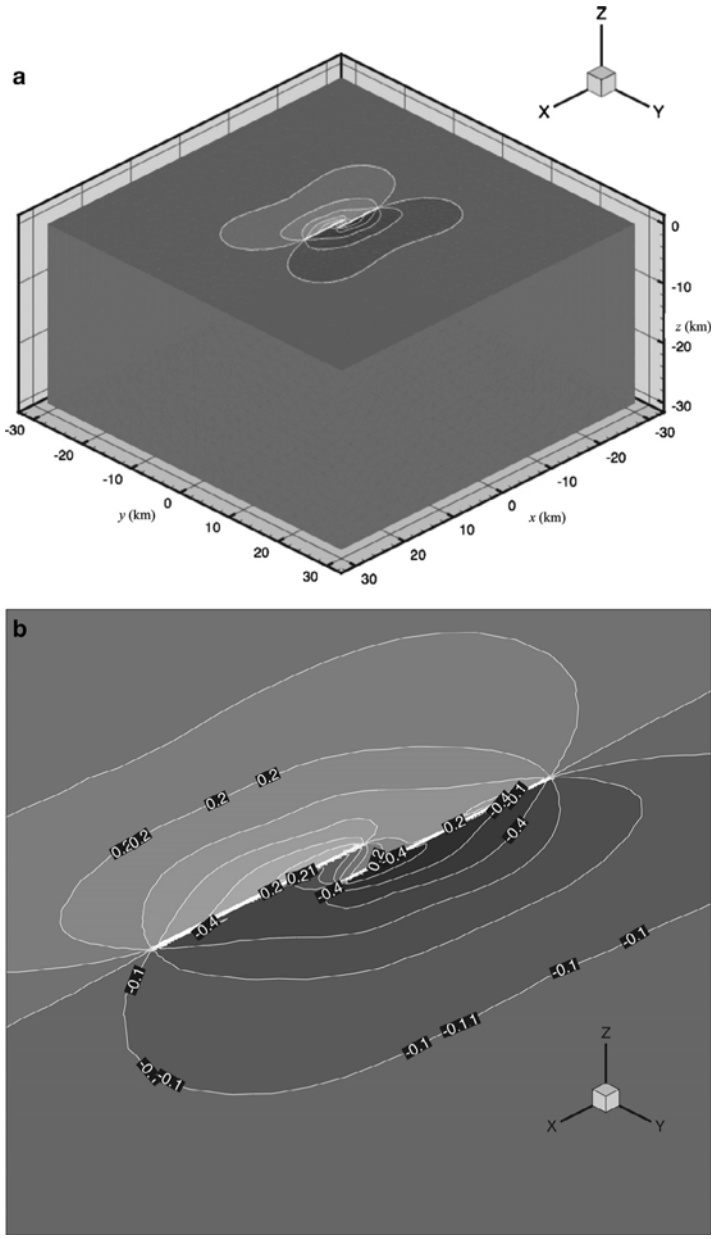


Figure 9
Two-fault stepover: Geometry and domain, with contours of U_z on the surface. a) Full domain, b) expanded view of fault region.

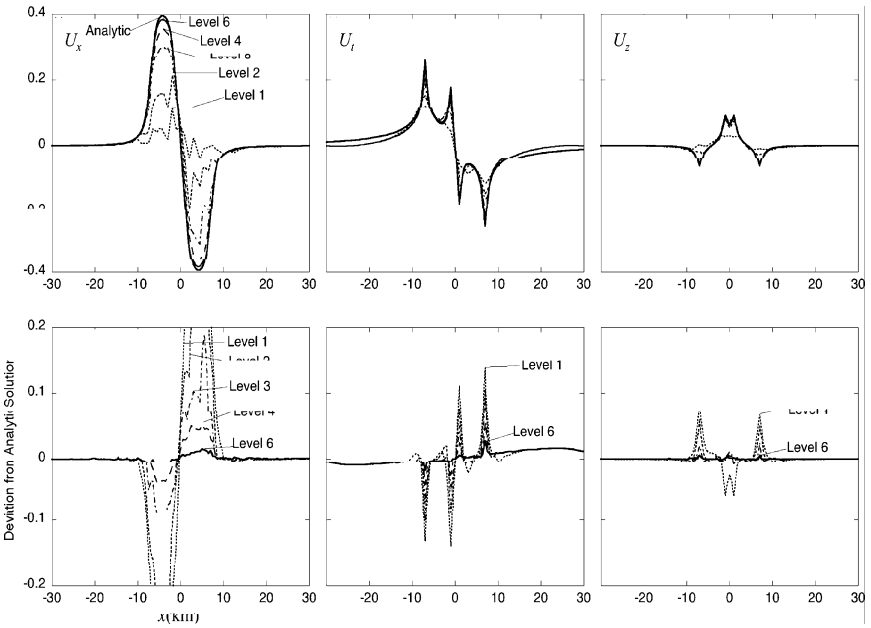


Figure 10
Two-fault stepover: Components compared to analytic solution. Top: Left, U_x (Analytic solution is dark line, refinement level solutions as labeled); Center, U_y; Right, U_z. Bottom: Corresponding differences from analytic solution for each component, refinement level.

ties for simulating elastic deformation and viscoelastic relaxation in domains including imposed-slip faults. Validation cases demonstrate essential correctness, including challenging phenomena of Maxwell relaxation, thrust-fault tip motion, curved fault slip, and systematic convergence in the vicinity of fault edges (in a stepover configuration). The energy-based functional used for refinement is not claimed to be optimal, nonetheless it proves very useful for its level of simplicity. It should be adequate for highly accurate solutions using computing common systems capable of solving millions of elements (currently a moderate cluster). In our experience the key advantage is delivery of good solutions with tolerable element count near fault edges where weakly singular stress tends to corrupt solutions relying on low-order basis functions.

Additional geophysical and numerical features are in progress, including freely slipping and friction-controlled faults, and parallel adaptive mesh refinement suitable for both elastic and relaxation phases of simulations. GeoFEST is becoming increasingly useful for comparisons between simulated and observed displacements in complex faulted regions, observed by geodetic networks and interferometric radar.

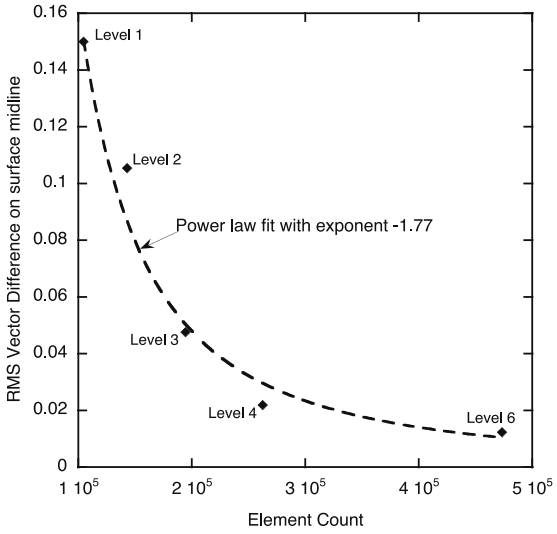


Figure 11
Stepover validation case: RMS convergence with refinement levels. Each point corresponds to a distinct refined mesh, and represents the RMS vector-norm difference between the finite-element solution on the midline, compared to an analytic solution.

Acknowledgements

Work done under the QuakeSim project was supported by the NASA Earth Science Technology Office through the Computational Technologies Program. As project manager, Michele Judd's attention to detail was essential to organizing meetings and the production of reports. We thank Carl Gable (Los Alamos National Laboratory) for providing a shared mesh representing the Benchmark 5b geometry, and Noah Fay (University of Oregon) for analytic near-field values used for boundary conditions in Benchmark 5b (elastic). We also express gratitude to the sponsors, leaders and participants in the five annual Workshop on Community Finite-Element Models for Fault Systems and Tectonic Studies put on by the Southern California Earthquake Center, the NSF Computational Infrastructure for Geodynamics, and NASA for establishing computational benchmark cases for code validation and comparison. This research was performed at the Jet Propulsion Laboratory, California Institute of Technology, under contract to the National Aeronautics and Space Administration.

References

Aagaard, B., Armendariz, L., Gable, C., Hager, B., Hetland, E., Knepley, M., Lu, J., Lyzenga, G., Parker, J., Simons, M., and Williams, C., *Benchmarking crustal deformation modeling software: Ensuring accurate and*

consistent results, Southern California Earthquake Center Annual Meeting, September 10–13, (Palm Springs, California, 2006).

GLASSCOE, M., DONNELLAN, A., GRANAT, R., LYZENGA, G., NORTON, C., and PARKER, J. (2006), *Simulation- and Statistics-Based Analysis of the 1906 Earthquake and Northern California Faults*, Seismol. Res. Lett. *77*, 247.

HAGER, B.H., LYZENGA, G.A., DONNELLAN, A., and DONG, D. (1999), *Reconciling rapid strain accumulation with deep seismogenic fault planes in the Ventura basin, California*, J. Geophys. Res. *104(B11)*, 25207–25220.

HUGHES, T.J.R., *The Finite Element Method: Linear Static and Dynamic Finite Element Analysis* (Dover, Publication, Inc., Mineola, New York, 2000).

HUGHES, T.J.R. and TAYLOR, R. (1978), *Unconditionally stable algorithms for quasi-static elasto-plastic finite element analysis*. Computers Struct. *8*, 169–173.

MELOSH, H.J. and RAEFSKY, A. (1981), *A simple and efficient method for introducing faults into finite-element computations*, Bull. Seismol. Soc. Am. *71*, 1391–1400.

NORTON, C. and CWIK, T., *Parallel Unstructured AMR and Gigabit Networking for Beowulf-Class Clusters*, *Parallel Processing and Applied Mathematics, 4th Internat. Conf., PPAM 2001 (LNCS 2328)*, (Springer-Verlag, Berlin, 2002) 552–563.

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

OWEN, R.J. and HINTON, E., *Finite Elements in Plasticity: Theory and Practice*. (Pineridge Press Ltd., Swansea, 1980).

PARKER, J.W., DONNELLAN, A., LYZENGA, G., RUNDLE, J.B., and TULLIS, T., *Performance modeling codes for the QuakeSim problem solving environment*, *Proc. Internat. Conf. on Comp. Sci. (Part III, LNCS 2659)* (Springer-Verlag, Berlin 2003), 845–862.

PIERCE, M., YOUN, C., and FOX, G., *Interacting data services for distributed earthquake modeling*, *Proc. Internat. Conf. Comp. Sci. (Part III, LNCS 2659)*, (Springer-Verlag, Berlin, 2003) 863–871.

POLSTYANKO, S. and LEE, J.F. (2001), Adaptive finite element electrostatic solver, IEEE Trans. Mag. *37*, 3120–3124.

(Received October 16, 2006, revised August 30, 2007, accepted October 21, 2007)

Pure appl. geophys. 165 (2008) 523–543
0033–4553/08/030523–21
DOI 10.1007/s00024-008-0326-8

Pure and Applied Geophysics

Multiscale Brittle-Ductile Coupling and Genesis of Slow Earthquakes

K. Regenauer-Lieb,[1] and D. A. Yuen[2]

Abstract—We present the first attempt to explain slow earthquakes as cascading thermal-mechanical instabilities. To attain this goal we investigate brittle-ductile coupled thermal-mechanical simulation on vastly different time scales. The largest scale model consists of a cross section of a randomly perturbed elasto-visco-plastic continental lithosphere on the order of 100 × 100 km scale with no other initial structures. The smallest scale model investigates a km-scale subsection of the large model and has a local resolution of 40 × 40 m. The model is subject to a constant extension velocity applied on either side. We assume a free top surface and with a zero tangential stress along the other boundaries. Extension is driven by velocity boundary conditions of 1 cm/a applied on either side of the model. This is the simplest boundary condition, and makes it an ideal starting point for understanding the behavior of a natural system with multiscale brittle-ductile coupling. Localization feedback is observed as faulting in the brittle upper crust and ductile shearing in an elasto-viscoplastic lower crust. In this process brittle faulting may rupture at seismogenic rates, e.g., at 10^2–10^3 ms^{-1}, whereas viscous shear zones propagate at much slower rates, up to 3×10^{-9} ms^{-1}. This sharp contrast in the strain rates leads to complex short-time-scale interactions at the brittle-ductile transition. We exploit the multiscale capabilities from our new simulations for understanding the underlying thermo-mechanics, spanning vastly different, time- and length-scales.

Key words: Fast instabilities, slow earthquakes, numerical modelling, rheology, brittle-ductile transition.

1. Introduction

The concept of slow earthquakes (Jordan, 1991; Pfluke, 1978; Sacks *et al*., 1978) has been around for nearly 30 years but there has been no serious work trying to explain the phenomenon from first principles in physics and mathematics. The recent Sumatran earthquake (Stein and Okal, 2005a, 2005b) followed by the 8.7 aftershock were slow earthquakes. It is the contention that many slow earthquakes may have been undetected (Ide *et al*., 2007). Slow earthquake instabilities, of the order of several days, have recently been reported to drop out of numerical simulations covering millions of years of thermal-mechanical deformation (Regenauer-Lieb and Yuen, 2006). The scaling and spectral behavior of slow earthquakes demonstrates that they can be thought of as different

[1] Department of Earth and Geographical Sciences, The University of Western Australia, and CSIRO, Exploration and Mining, 26, Dick Perry Ave, WA 6102 Perth, Australia. E-mail: Klaus.regenauer-lieb@csiro.au
[2] Department of Geology and Geophysics and Minnesota Supercomputer Institute, University of Minnesota, Minneapolis, MN 55455-0219, U.S.A.

manifestations of regular earthquakes (IDE *et al*., 2007). In this context our approach is corroborated by a 1-D model of shear heating feedback, which has indeed shown that the ductile mechanism alone is capable of generating regular intermediate-depth earthquakes in the mantle (KELEMEN and HIRTH, 2007).

We show here a first attempt from a solid-state point of view that can explain slow earthquakes as cascading thermal-mechanical instabilities through coupling shear heating and adiabatic heating to creep-acceleration which may explain the phenomenon. The approach is based on an extension of a classical geodynamic modelling approach, alleviating the basic shortcomings of classical, static lithosphere strength profiles, which are primarily overestimating the lithospheric strength and imply a very sharp brittle-ductile transition. The new method is based on the important positive feedbacks produced by the dissipation function, i.e., shear heating.

In order to illustrate where we might expect the maximum effect of dissipation we discuss the classical lithosphere strength profiles as proposed originally by GOETZE and EVANS (1979) and estimate where the maximum dissipation would be expected without considering feedback. The dissipation function is defined as the double dot product of the visco-plastic stress times the associated strain rate which constitutes the shear heating. Before going into a discussion of dissipation, we first present a qualitative description of the basic differences between steady-state and time-dependent results.

In Figure 1 we compare classical and new results shortly after application of the load showing the magnitude of the second invariant of the deviatoric stress tensor and its associated strain rate as an example for quartz/olivine rheology composite. The grey line is time-dependent and the black line is steady-state, based on the constitutive relationship under application of a constant extension background strain-rate of 2.5×10^{-15} s^{-1}. At this early stage of deformation the two strength profiles bear some similarity, particularly in the central part of the plate which does not yet differ much from the steady-state solution. The top and bottom of the plate, however, show significant departure from steady state. In the main body of the text we will focus on the later stages, where the central portion, i.e., the area around the brittle-ductile transition, also becomes weaker than the steady-state solution The differences at the early stage are caused by feedback processes which initially mainly affect the top and bottom portion of the plate. Weakening in the top 8 km is due to the pressure-sensitive yield which causes fast brittle failure. In the time-dependent solution pressure-sensitive yielding breaks the brittle crust sequentially from the top to the bottom within the first 1 k years of loading. This initial response unloads the top part followed by a drop in strain rates and stress (see Fig. 5). The process is rate controlled by an increase of the effective viscosity with depth over this pressure-dependent layer (see Fig. 6). The same unloading phenomenon applies to a depth greater than 50 km, at which the temperature-sensitive process has weakened the bottom part of the lithosphere. Note, that in the time-dependent solution weakening varies both through space and time. Instabilities within the top and bottom part are small and occur on a time scale shorter than 1 k year. A submodel (not shown) with 10 m local resolution suggests that brittle faulting mechanism is self-similar down to this scale. We

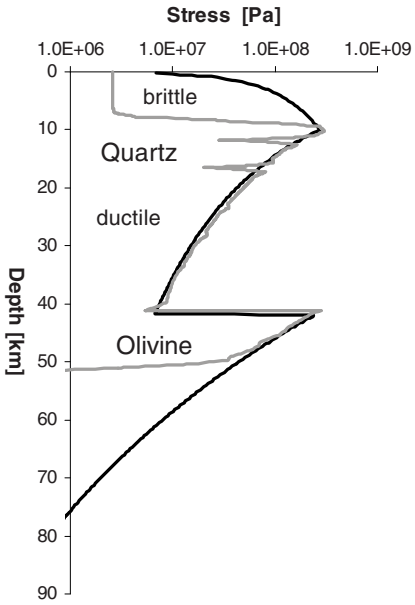


Figure 1

The classical Christmas tree (black) yield strength envelope versus a very early snapshot (130 k years after onset of extension) in the initial loading phase of a dynamic elasto-visco-plastic strength profile calculated in REGENAUER-LIEB (2006). The lithosphere is made up of a simple quartz < 42 km depths olivine composite > 42 km depth. At the early stage of deformation the main difference between the two approaches is an apparent low stress near the upper and lower frictionless interfaces. Upon initial loading the upper and lower interfaces are effectively shielded elastically through the incremental elastic stretching of the areas circumscribed by the stress peaks. In subsequent deformation these stress peaks are destroyed because they are also the sites of the maximum local dissipation (Fig. 2). While the Christmas tree is a good strength measure within the classical elasto-plastic framework, a dynamic elasto-visco-plastic Christmas tree structure also requires investigation of the effective viscosity, which gives a better measure of strength through time (see Fig. 5).

focus in the following on the large-scale models. This large-scale model resolves failure of the brittle-ductile transition, causing plate size events. These large-scale events are controlled by a critical temperature. In the following, the concept of the critical temperature for such plate scale instabilities is first discussed on the basis of the steady-state solution.

Figure 2 highlights the local critical temperature of maximum dissipation at around 900 Kelvin for olivine and 500 Kelvin for quartz, assuming a local strain-rate of 2.5×10^{-15} /s. At a higher strain-rate the critical temperature also increases. The diagrams shown in Figures 1 and 2 are the classical quasi-static interpretations of the strength of the lithosphere as initiated by GOETZE and EVANS (1979). In this work the concept of lithospheric strength was simplified in a static picture an elasto-plastic

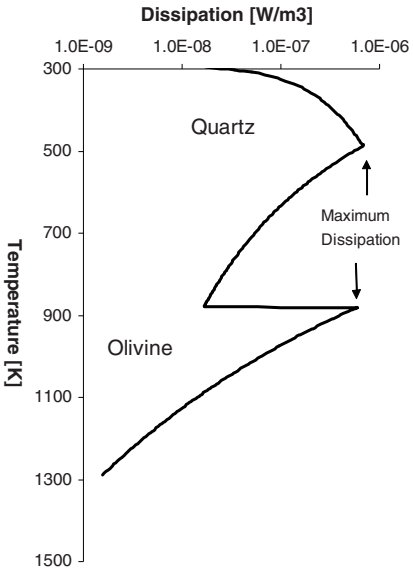


Figure 2

The associated dissipation of Figure 1 without feedback. The critical temperature is defined by maximum dissipation in the olivine/quartz -power-law creep regimes and is around 500 K for quartz rheology and 900 K for olivine. The classical view implies that thermal-mechanical instabilities, if they exist, are to be observed at these levels. We investigate more closely in the following the critical regime in fully coupled calculations and show that this simple concept of a critical temperature defines a gradual transition zone rather than a sharp boundary.

yield-strength envelope. The lithosphere was assumed to be elastic when loaded below the stress identified in the curve, and plastic when it reaches the curve. This obviously fixes the yield strength envelope to a particular strain rate most often 10^{-15} s^{-1}. Later additions turned the elasto-plastic approach into a visco-elasto-plastic approach from which the elastic, plastic and viscous rheological bodies are assumed to be used in series (Albert and Phillips, 2002; Chery *et al*., 1991; Gerbault *et al*., 2003), thus generalizing the problem for all strain rates.

This approach has, in theory, the potential of resolving a self-consistent brittle-ductile transition as a function of the imposed boundary conditions. It will for instance move downwards for a higher applied background strain rate. There is, however, one significant shortcoming. The approach predicts a very sharp transition from basically elasto-plastic behavior in the brittle crust to visco-elastic creep in the ductile regime, with no significant transition region in between. This shows the efficiency of the Arrhenius temperature dependence of the flow laws, which allows very rapid accommodation of permanent deformation by visco-elastic creep. In the extended visco-elasto-plastic transition the pressure-sensitive localization mechanism thus abruptly gives way to the thermally

accommodated creep. This is at odds with inferences from both laboratory (KOHLSTEDT *et al*., 1995) and the field experiments (MANCKTELOW, 2006). Transitional brittle-ductile behavior develops naturally, when one considers the positive feedback from the dissipation function.

The first fully coupled thermal-mechanical approach in geology was pioneered by CHERY *et al*. (1991), however, at that early stage the required spatial resolution was still sorely lacking and thermal expansion was not considered, hence the effects of feedback reported by this work were overlooked. The first model with potential relevance to slow earthquakes was published (REGENAUER-LIEB and YUEN, 2006). In that paper we analyzed a single quartz crust in extension and found self-consistent development of three distinctly different rheological layers: A fully brittle layer at the surface with fast time-scale instabilities, a bottom fully ductile layer with slow mylonitic shear zones and an intermediate semi-brittle/ductile layer of several km thickness, where processes communicate. For the case of quartz rheology this resulted in a single critical temperature with maximum dissipation governing the creep/brittle ductile layer instability system. Here, we extend the analysis to the lithosphere and perform a more detailed analysis of the brittle-ductile coupling.

2. *Modelling Framework*

We use a fully coupled thermo-mechanical framework for the generation of shear zones in geology and planetary sciences (REGENAUER-LIEB and YUEN, 2003). The key element of the theory is an incorporation of the energy equation as a master of the two driving forces for instability, which are the pressure pulses ΔP and temperature pulses ΔT acting through the filter of continuity and the quasi-static momentum equation as slaves. Both temperature and pressure pulses are positive feedback variables giving rise to the instabilities. The only negative feedback constraints preventing catastrophic failure are given through conduction, dissipating the thermal pulse, and imposed boundary conditions stabilizing the pressure through the quasi-static momentum and the continuity equations. Figure 3 shows the general framework of the brittle/ductile feedback loop where the equation of state is tied to the continuity and the rheology acts as the most important feedback filter. The temporal evolution of the process is thus controlled inherently by the thermodynamic evolution of the energy equation. A more detailed description of the feedback loops, the equations employed and a comparison to the classical engineering style approach, which only considers moment and continuity equation, can be found elsewhere (REGENAUER-LIEB and YUEN, 2004).

We want to point out at this stage that the classical theory neglects the energy evolution, which is at the heart of any nonlinear time scale, such as occurs in earthquakes. Attempts to incorporate time scale through empirical strain-, or strain-rate weakening laws or rate and state variable friction (DIETERICH, 1979a, 1979b) are not self-consistent within the framework of the energy equation. The empirical laws may be

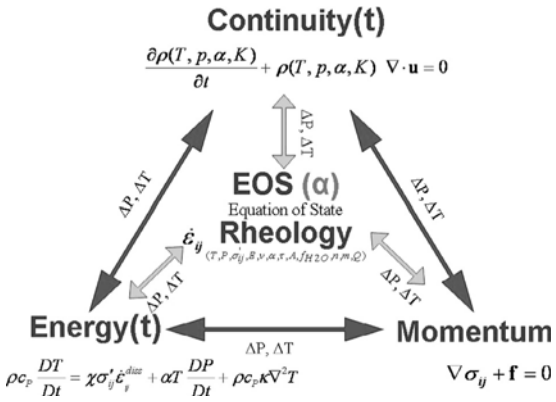


Figure 3

Small changes in pressure and temperature can have a large effect on the deformation of rocks. Non-linear feedback can cause pattern formation out of an initially unpatterned state. Here, two important mechanical feedback processes are described in the energy equation. Small perturbations in temperature and pressure are caused by shear heating (first term on the right side of the energy equation), thus changing the entropy through dissipation and thermal expansion (second term) being an isentropic process. Both dissipative and non-dissipative processes in conjunction with dissipation cause formation of shear zones. The time-dependence in both the energy and continuity equations is indicated by the symbol *t*.

extended to a continuum framework (ORD *et al*., 2004), however, careful examination of the energy equation needs to be performed, such as other sources of energy due to grain-boundary processes (BERCOVICI and RICARD, 2003). The main drawback is that an empirical plasticity framework does not prevent violation of the energy equation, which, incidentally may occur. The original CAM-Clay model (COLLINS and HILDER, 2002) is a typical example which has been shown to violate the second law of thermodynamics. Although the energy equation is not explicitly needed in such isothermal models, an amendment in a thermo-mechanical framework is mandatory. The incorporation of (isothermal) thermodynamics into mechanics pioneered in Zieglers book (ZIEGLER, 1983) has recently revolutionized plasticity theory and the modelling shallow deformation in soil and sands (COLLINS, 2005). This example shows the key role of the energy equation for geological applications where we are certainly not dealing with isothermal conditions.

3. Equations Solved

3.1. Continuity Equation

We express the continuity equation here in differential form where the scalar (dot) product of the velocity vector **u** with the nabla vectorial differential operator ∇, i.e., the

divergence of the velocity field is exactly equal to the negative rate of the fractional density change.

$$-\frac{1}{\rho}\frac{D\rho_{(T,P,\alpha,\ldots)}}{Dt} = \nabla \cdot \mathbf{u}. \tag{1}$$

Here *D/Dt* is the substantive Lagrangian (also called material, Lagrangian, convective) derivative. The density ρ of the solid Earth materials is itself a function of the thermodynamic state variables; the most important being temperature T, pressure P and strain caused by thermal expansion α. The substantive derivative is the rate of change of some material property (here the density) defined by:

$$\frac{D()}{Dt} = \frac{\partial()}{\partial t} + \mathbf{u} \cdot \nabla(),$$

where $\mathbf{u}$ is the local material velocity vector of the volume under consideration and the empty bracket is a placeholder for the particular material property. This can be rearranged considering the dependence of the density on the thermodynamic state variables as

$$\frac{\partial \rho_{(T,P,\alpha,\ldots)}}{\partial t} + \nabla \cdot \left(\rho_{(T,P,\alpha,\ldots)} \cdot \mathbf{u}\right) = 0.$$

Equation (1) incorporates time as a derivative, which is implicitly derived from the evolution of isentropic work to be discussed in conjunction with the energy equation. This subtlety of continuity equation enhances thermal-mechanical coupling, which is of course not present in the incompressible approaches, where time does not enter into the continuity equation. We emphasize that volumetric deformation cannot be neglected, especially in solid-state phase transitions, such as the basalt to eclogite transition for geodynamic conditions.

3.2. Momentum Equation

In its general form the material derivative also enters the momentum equation there describing the accelerating components experienced by the Lagrangian (i.e., convected) material volume. In geodynamic deformation it is a common practice to drop the inertial forces, which contain the time-dependence. Whether this approach still holds for slow earthquakes is questionable. However, we shall adopt here as a first approach the common practice of neglecting inertial terms. With this simplification the momentum equation therefore can be written independently of time as the equilibrium condition of the surface tractions (divergence of the Cauchy stress tensor σ_{ij}) applied over the surface of a volume element and the body forces $\mathbf{f}$ applied to the volume element.

$$\nabla \cdot \sigma_{ij}(T, P, \sigma_{ij}^{'}, E, \nu, \alpha, \tau, A, f_{H_2O}, n, m, Q) + \mathbf{f} = 0. \tag{2}$$

The Cauchy stress tensor is in itself a function of the rheology (eq. (3)), depending in our calculations on the variables temperature T, pressure P, deviatoric stress σ'_{ij}, Young's

modulus E, the Poisson's ratio v, the volumetric coefficient of thermal expansion α, the plastic yield stress τ, and the creep law parameters, such as the activation enthalpy and pre-exponential factor (eq. (3)). The most important quantities in terms of the creep law are the water fugacity f_{H_2O} and the activation enthalpy Q while A, m and n are the parameters for the phenomenological fit to the creep behavior.

3.3. Rheology

For ductile and brittle behavior, we use a composite rheology by assuming a modified von Mises plasticity, with a pressure-dependent yield stress (linear Drucker-Prager) combined with the elasto-visco-plastic coaxial flow rule. This gives the following

$$\dot{\varepsilon}_{ij} = \left(\frac{1+v}{E}\frac{D\tilde{\sigma}'_{ij}}{Dt} + \frac{v}{E}\frac{DP}{Dt} + \alpha\frac{DT}{Dt}\delta_{ij}\right)_{\text{elastic}} + \left(\dot{\varepsilon}^{pl}\frac{\sigma'_{ij}}{2\tau}\right)_{\text{plastic}} + \left(A f_{H_2O}^{m}\sigma'_{ij}J_2^{n-1}\exp\left(-\frac{Q}{RT}\right)\right)_{\text{creep}}. \tag{3}$$

Recall that E is Young's modulus and v is Poisson ratio. $\dot{\tilde{\sigma}}_{ij}$ is the objective corotational stress rate (see Appendix) and δ_{ij} is the Kronecker delta. J_2 is defined as the second invariant of the deviatoric stress tensor

$$J_2 \equiv \sqrt{\frac{3}{2}\sigma'_{ij}\sigma'_{ij}}.$$

The plastic yield potential Φ defines the elasto-plastic limit at low pressure. It is known as a linear Drucker-Prager Yield Criterion, whereby the coefficient for P is assumed to be one and the tensile strength is assumed zero.

$$\Phi = J_2 - P = 0.$$

Inside the potential function the material response is entirely elastic (plus some minor viscous flow) and when the stress reaches values defined by the yield, stress plastic strain rates are allowed.

3.4. Energy Equation

The substantive derivative makes a prominent reappearance in the energy equation (4) as both temperature T and pressure P convected derivatives. A Lagrangian mesh is therefore chosen for our numerical solution.

$$\rho c_P\frac{DT}{Dt} = \chi\sigma'_{ij}\dot{e}_{ij}^{diss} + \alpha T\frac{DP}{Dt} + \rho c_P\kappa\nabla^2 T. \tag{4}$$

The total change of heat balances in three important terms. The first two terms on the right-hand side of equation (4) describe the temperature change due to dissipative and

stored mechanical work, respectively. The first term on the right is also known as the shear-heating term where χ is the efficiency of converting mechanical work into heat at large deformation. χ is of the order of 0.9 for most materials (Chrysochoos and Dupre, 1992). Shear heating only describes the temperature changes owing to dissipation where $\dot{\varepsilon}_{ij}^{diss}$ is the plastic and viscous strain rates. The second (isentropic) term on the right describes the reversible temperature change due to elastic volumetric changes. It thus contains the effect of temperature changes due to elastic compressibility. The third term gives the familiar contribution of heat conduction as a function of diffusivity κ and ∇^2 is the Laplacian scalar differential operator. The specific heat of the Lagrangian convected volume element is c_p.

4. The Geodynamical Model

We investigate a very simple but basically fundamental model, in which a 100-km-long segment of continental lithosphere is subject to far-field extension of 0.5 cm/a applied on either side. The continental lithosphere consists of a 42-km-thick quartzitic crust underlain by peridotitic mantle material with a 38-km thickness. At the bottom and along the sides of the lithosphere free slip, i.e., zero tangential stress boundary conditions are assumed, while the displacements are constrained by zero vertical displacement on the bottom and on the side only periodic vertical displacements are allowed, i.e., the vertical displacements on the right are constrained to be exactly equal to the vertical displacement on the left-hand side of the model. The top boundary is an open boundary. For comparison see also Regenauer-Lieb *et al*. (2006) and Rosenbaum *et al*. (2005).

The thermal boundary conditions are also very basic. At the bottom of the lithosphere segment, a fixed base mantle heat flow of 30 mW/m^2 is assumed while within the crust there is a curvature following the steady state for continents. The initial thermal profile is allowed to be in the steady-state limit. Superposed on this classical pattern are small random thermal perturbations, allowing a white noise seed for nucleating the shear zones.

5. Results

5.1. Geological Time Scale ($O(10\ Ma)$)

Here we are describing only modelling results for the case of extension. The modelling results for compression underline the identification of (higher) critical temperature for transition from brittle to ductile processes, however they display different complex patterns not discussed here. Continental breakup has been followed by formation of an ocean (time scale, millions of years). For reasons of clarity we do not focus on geodynamic consequences of a switch from crustal quartz strength to a mantle olivine strength control observed for lower surface heat flow. In Figure 4 (a,b) we first

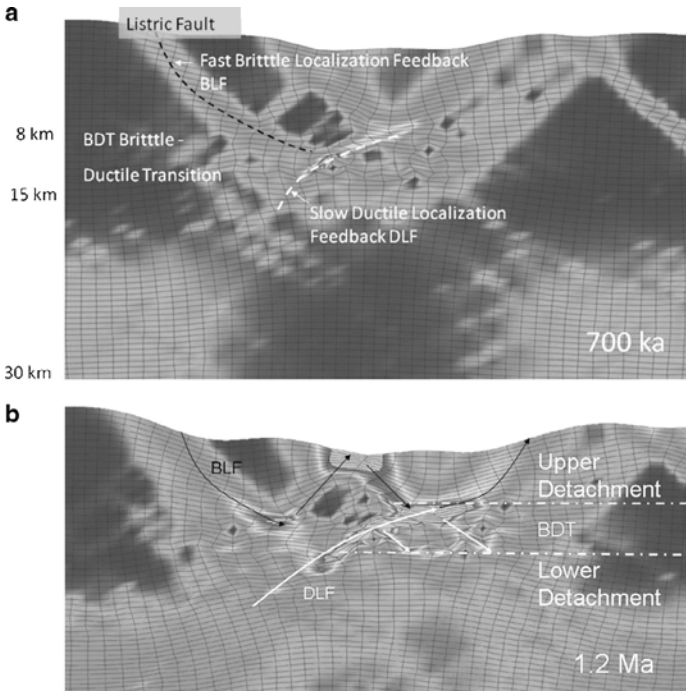


Figure 4

Brittle (P-dependent) and ductile (T-dependent) feedback for a small piece of continental crust in extension illustrating the brittle-ductile feedback. Two subsequent time slices are shown in a) and b) 700 k and 1.2 Ma after the elastic loading phase. The thermal conditions are 60 mW/m^2 surface heat flow and the total extension velocity for this particular model is 2 cm/yr. Only the crust is considered, for simplicity. The two detachments are delimiting an area to either side of the stress and maximum dissipation peaks in Figures 1 and 2, thus decoupling an "elastic core" (BDT = brittle-ductile transition) of the crust, i.e., an elastic stress-bearing portion of the crust. This maximum elastic energy and maximum dissipation area plays a significant role for slow earthquake instabilities.

discuss a simple calculation of quartz crust with 60 mW/m^2 surface heat flow and a frictionless base instead of a mantle layer (REGENAUER-LIEB and YUEN, 2006). We use this example to investigate in more details the possible thermal-mechanical mechanism underlying the generation of short-time scale instabilities.

For classical extension models that do not consider brittle-ductile energy feedbacks, steeply dipping normal faults are conventionally obtained (WIJNS *et al.*, 2005). Listric, normal faults, as observed in geology, are very difficult to achieve. Our fully coupled simulation overcomes this problem (REGENAUER-LIEB *et al.*, 2006). Localization feedbacks in and between the brittle and ductile layers can turn the direction of faults around. Listric faults are caused by efficient weakening on brittle-ductile detachment layers. These sub-horizontal layers develop through shear-heating forming brittle and

ductile flow localization. Two detachments are defined by a bottom mylonitic zone formed by ductile localization and a top detachment zone due to brittle localization. Upon inspecting the more detailed structure of the detachment layer, we encounter a surprising complexity all located within the brittle-ductile transition zone, which upon initial loading forms an elastic core. The core is ruptured from two sides (Fig. 4). From the top the propagation of a brittle fault is driven by the combination of loading and brittle localization feedback effects. Fault propagation is hampered, as it reaches the top of the elastic core, where brittle and ductile processes interact, because a viscous response blunts the brittle fault tip and eliminates the stress singularity. Further arrest of the brittle fault is driven by ductile thermal-mechanical feedback effects, which act to decrease the differential stress at the tip of the fault by lowering the effective viscosity. Simultaneously, ductile shear zones propagate upwards towards the elastic core. However, these shear zones die out because of increasing crustal strength and because they reach the area in which brittle localization feedback dominates. Their energy is translated into a local, positive amplitude disturbance of the elastic stress field. This extra stress, loaded to the core from below, eventually starts to interact with similar features loaded from above.

We now return to the discussion of the time evolution of the one-dimensional depth slice through the model presented in Figure 1. In a dynamical framework the strength envelope of Figure 1 involves not only a discussion of stress but also the strain rates. We propose that a suitable measure of strength is the effective viscosity. The stress/strain rate-depth curve for the dynamic curve in Figure 1 is shown in Figure 5 and the resultant viscosity profile is shown in Figure 6.

At the onset of deformation (e.g., 130 kyrs) the energy feedback affects mainly the maximum dissipation depth level shown in Figure 2 thus chiefly depends on depth/temperature. A 1-D depth profile might therefore give representative results for the entire model. The more evolved strength evolution curves depend not only on depth but also significantly on the horizontal dimension. For instance, near the surface in the vicinity of fully developed listric shear zones shown in Figure 4, the stress is significantly higher than in the center of the graben structure, away from this shear zone. For ease of comparison to Figures 5 and 6 we monitor the 1-D depth-profile in a vertical cross section through the stable middle of the graben structure, thus resolving mainly vertical changes in strength evolution for the evolved stage after 6.4 Ma of extension in Figures 7 and 8.

Figures 7 and 8 display a consistent evolution of features visible in Figures 5 and 6. The main feature is a vertical upwards movement of the detachment accompanied by a narrowing of the upper and lower detachments. The two shear zones sometimes coalesce through individual oblique shear zones, to the effect that the core is disrupted during these creep burst events. In this way the upper mid-crustal shear zones develops into two low viscosity zones, defined by two detachments separated by an elastic core. We argue that the rupture behavior of this elastic core holds the key to crustal earthquakes. In order to zoom into the dynamics associated with the critical behavior, we investigate a small

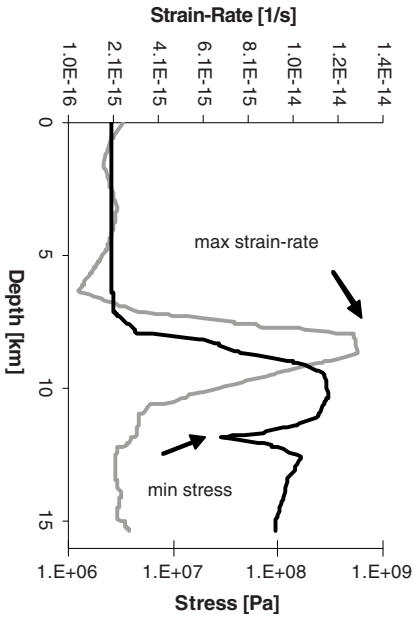


Figure 5
Log (stress) (black) and Log (Strain rate) (grey) versus depth curve showing data from the dynamic strength-evolution model shown in Figure 1, focussing on the brittle-ductile transition during the initial elastic loading phase after 130 kyrs extension with a background strain rate of 2.2 10^{-15} s^{-1}. A 2.5-km phase shift between the maximum strain rate and the maximum stress peak is visible. The upper detachment develops at the maximum in strain rate at around 8.5 km depth while the lower detachment develops at the minimum in the local stress at around 12 km depth.

sub-model representing a thin depth slice of the continental crust with a spatial resolution of 40 m.

5.2. Geodetic Time Scale (O(1 Ma)

Crustal deformation (time-scale years) is compared to high resolution geology. For this first analysis a depth slice is taken with isothermal boundary conditions using stress and displacement boundary conditions from the large-scale model. The thermal boundary conditions are different from the large-scale model. This model has random thermal perturbations of 70 K, corresponding to the approximate amplitude of a natural thermal gradient of a 3-km vertical slice of the crust. These perturbations are inserted as random white noise on 4% of the nodes with a maximum positive amplitude of 70 K. The wavelength of the thermal perturbations after deformation is not only controlled by conduction but also significantly affected by the thermal-mechanical feedback. We use a

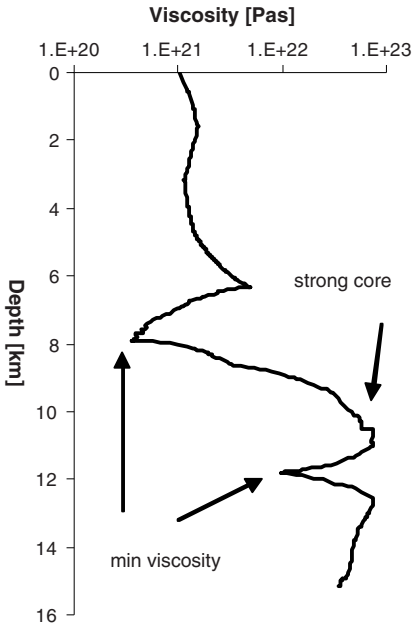


Figure 6
Log (viscosity) versus depth of the model shown in Figure 5. The effective (secant) viscosity is defined by dividing the 2nd invariant of the stress over its associated strain rate. The two detachments around the "elastic" stress-bearing core are clearly visible as local viscosity lows. At the early stage of deformation there is still significant strength in the ductile part of the crust.

homogeneous setup with a central layer of aplitic material in the middle of the crustal granitic slice. The background temperature is varied in individual model runs in 10 K increments from 430 K to 560 K. In this first test of the mechanical response, the submodel is not advected within a large-scale crustal model, i.e., not fully coupled to the large-scale simulation. It therefore lacks the dynamics of the brittle-ductile transition discussed above.

Here, only the material behavior around the critical temperature is investigated i.e., the material response that leads to the rupture of the elastic core. At this early stage we also wish to isolate the mesoscale material response from the large-scale hierarchical driver. An exactly analogous analysis is done under convergent boundary conditions to be reported elsewhere (REGENAUER-LIEB *et al.*, 2007). Future follow-up studies will also examine full coupling of geodetic and geodynamic time and length scales. Initially, we will search for exemplary regions of interest.

The mesoscale results clearly show that in addition to the above-mentioned P- and T-dependent brittle/ductile fault/shear zones above and below the brittle ductile

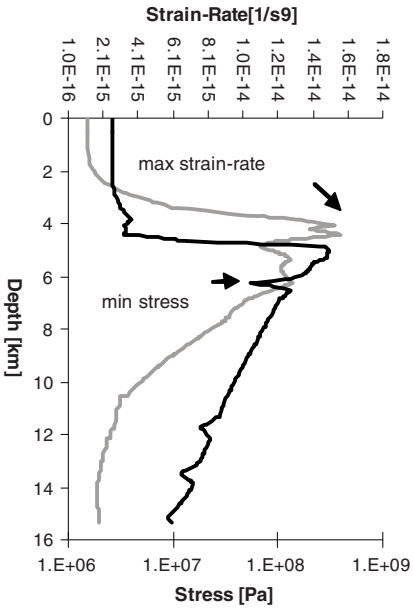


Figure 7
Log (stress) (black) and Log (Strain rate) (grey) versus depth curve showing data from the dynamic strength-evolution model shown in Figure 5 after 6.4 Myrs of extension. The phase shift from maximum in strain rate to maximum in stress still persists. However, the shift is reduced to 1 km. Consequently the two detachments near the maximum in strain rate and the other local maximum at the local minimum in stress are approaching one another.

transition, we record near a critical temperature an additional phenomenon that deviates significantly from the pure shear boundary conditions. Localized centres with high vorticity appear most prominently in the feldspar layer (Figs. 9 and 10). High elastic energy inside the vortices are the sources of short-time scale instabilities. Figure 9 illustrates that these centres are controlled by the thermal expansion feedback and relate to local temperature maxima. Figure 10 shows that the local vortices store a significant amount of elastic energy.

5.3. Towards Earthquake Timescales O (a)

We are now putting the above calculations into perspective for the type of time-scale relevant for earthquakes without presenting additional results. Obviously, a conclusive analysis can only be done if geodynamic and geodetic length and time scales are fully coupled in a multi-scale framework. This analysis would require robust automated adaptive meshing techniques, available for the entire critical shear zone, where the discretization needs to be smaller than the thermal diffusion length scale associated with

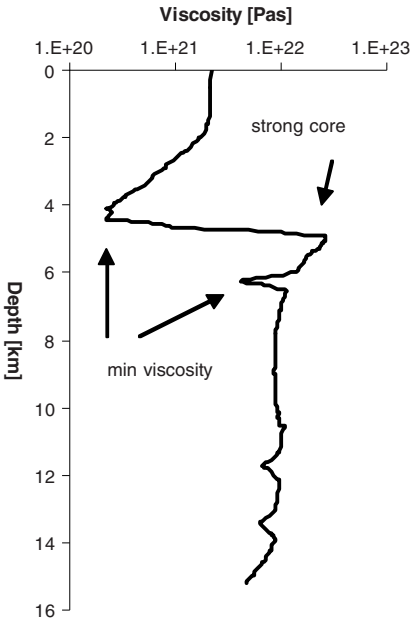


Figure 8
Log (viscosity) versus depth of the model shown in Figure 7. There is a pronounced viscosity minimum near the upper detachment. The lower crust contributes significantly to the strength of the crust, however, not as significant as the strong core, which is very small and located at shallow 5 km depth.

the thermal disturbances within the recurrence interval of major earthquakes, say 1000 years. This time interval equates to the length scale of around 100 m. However, we would additionally wish to consider inertial terms to explore the richer dynamics at smaller time and length scale. Our approach is clearly at a very early stage. We would argue that the results nonetheless point to a promising avenue of research and highlight possible application to the genesis of slow earthquakes, which we discuss below.

Extraction of the free-oscillation amplitude excited by the recent Sumatran earthquake by STEIN and OKAL (2005a) has shown a linearly growing trend in the semi-log plot between amplitude and frequency. This trend is reminiscent of the Rayleigh-Jeans portion of the Planck curve in radiation physics, which was posed as the so-called ultra-violet catastrophe. This signature shows that other physics must intervene to neutralize this singular tendency. In analogy to the Rayleigh-Jeans paradox we propose that a quantized energy state may offer a solution. This physics obviously controls the largest size of the slow slip event which we argue to be in turn controlled by the energy stored in the elastic core and the thermal-mechanical rupture behavior of the core. We postulate here that the centers of vorticity recorded in the mesoscale simulation could be a possible candidate for the analogy of a “Plancks Quant” for earthquakes. We call

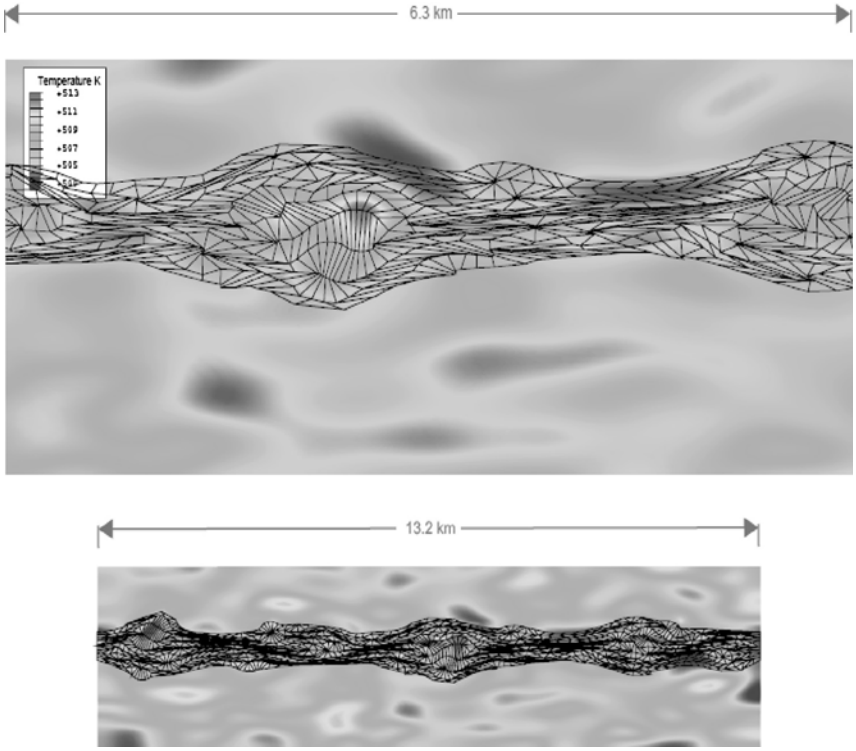


Figure 9
Temperature distribution after 2 Million years of extension. The original rectangular grid over the feldspar layer shows significant distortion and rotation around temperature maxima.

attention to the large-scale simulation shown in Figure 4 which also shows these vortices between the upper and lower detachment. It is hence a robust feature for a crust in extension for both geodynamical and geodetic time-scale simulations.

6. *Summary and Future Perspectives*

Our study presents thermal-mechanical shear interaction within the framework of a two-dimensional time-dependent model wherein a realistic visco-elastic-plastic rheology is employed, and the governing equations include the momentum equation without inertia, the rheological and energy equation. Earlier work on a one-dimensional model with the Peierls stress built into the rheology (KAMEYAMA *et al.*, 1999) has already shed light on these short time-scale instabilities. Other work pioneering the two-dimensional

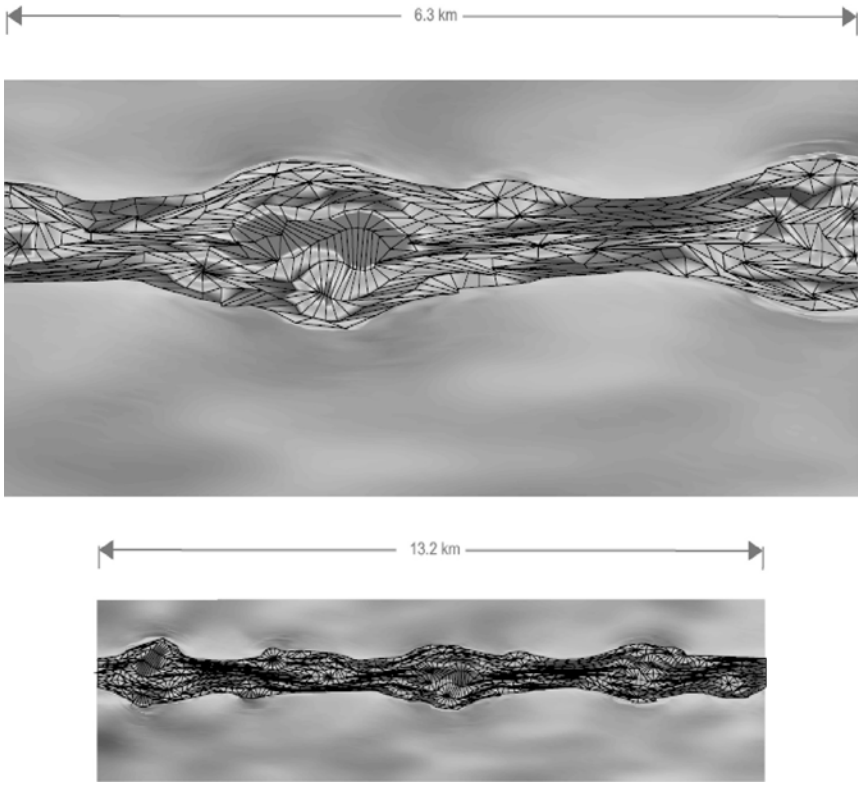


Figure 10
Elastic energy distribution (red = 1.0 10^6 J/m^3) corresponding to Figure 9. The central portion of the high temperature vortices stores a significant amount of elastic energy.

dynamics caused by rheological layering has also been suggested (LYAKHOVSKY *et al*., 2001). It is also our intention to explore the influence of temperature-dependent thermal conductivity on the critical temperature, since EMMERSON and MCKENZIE (2007) have recently found that the temperature associated with the concentration of seismicity in slabs is reduced by some 20% from temperature-dependent conductivity. The models presented here make the case more convincing. In the energy equation we have retained all mechanical heating terms and heating terms involving volumetric expansion. In our simulations we encounter two basically different bifurcation phenomena at the brittle-ductile transition-zone, which can be attributed to two different families of eigenmodes of the system. One in which the shear zone nucleates on thermal perturbations in the ductile field, and the second which is fully associated with elasto-plastic (brittle, pressure-dependent) displacements associated with thermal stresses. A quartz slab has all

two modes operating simultaneously in three different depth levels. The bottom of the crust is controlled by the elasto-visco-plastic mode while the top is controlled by the elasto-plastic mode. The exchange of the two modes appears to communicate on a subhorizontal layer within the brittle-ductile transition zone. We show that the isentropic work term, in the energy equation, encapsulating the effect of thermal-elastic expansion, is the key to fast time scales of deformation. Geodetic meso-scale simulations demonstrate that thermal expansion is particularly important in a critical temperature domain at the brittle-ductile transition where the two different eigenmodes interact. Near the critical temperature, thermal expansion forms local centers of vorticity with a length scale of the order of several hundreds of meters. In our calculations these centers are controlled by both heat conduction feeding back through shear heating and thermal expansion through individual creep burst events. We demonstrate here the underlying mechanism for fast time-scale instabilities to be the thermal expansion work from coupling to elasticity. Therefore, we have shown that with a properly formulated thermal-mechanical model, we can generate time scales now very close to those of earthquakes and of the same order as slow earthquakes. These results together with future data acquisition from GPS data would facilitate linkages to the two regimes of time scales and enable us to develop a uniformly valid rheological law for the lithosphere.

Acknowledgements

We are grateful for discussions with Bruce Hobbs and Charley Kameyama. This research has been supported by NSF's ITR and Math-Geo programs and the Western Australian Premiers Fellowship program, the University of Western Australia and the CSIRO Exploration and Mining Division.

APPENDIX: Further Details on Constitutive Equations

Elastic, Plastic and Creep Strain Rates

For calculating strain rates in equation (3) we use classical isotropic elasticity, considering in addition the thermal-elastic effect through thermal expansion. Parameters are given in Table A1. For the Drucker-Prager yield envelope we use the simple assumption that the yield stress is given by the ambient pressure. For calculating viscous strain rates we use power-law creep for a wet lithosphere in which the parameters are given in the Table A1.

$$\dot{\varepsilon}_{ij}^{cr} = A\sigma'_{ij}J_2^{n-1}\exp\left(-\frac{H^{\text{Power}}}{RT}\right). \tag{A1}$$

Table 1

Thermal, creep and elastic parameters used in the numerical model

Parameter	Name	Value	Units	Reference
χ	Shear heating efficiency	1	–	(Chrysochoos and Belmahjoub, 1992)
κ	Thermal diffusivity	Quartz = 0.7×10^{-6} Feldspar = 0.7×10^{-6} Olivine = 0.8×10^{-6}	m^2s^{-1}	
α_{th}	Thermal expansion	3×10^{-5}	K^{-1}	
c_P	Specific heat	Quartz = 1300 Feldspar = 1300 Olivine = 1300	Jkg^{-1} K^{-1}	
ρ	Density	Quartz = 2800 Feldspar = 2800 Olivine = 3300	kgm^{-3}	
v	Poisson ratio	0.25	–	
E	Youngs modulus	4.5×10^{10}	Pa	
A	Material constant pre-exponential parameter	Quartz = 1.3×10^{-34} Feldspar = 7.9×10^{-26} Olivine = 3.6×10^{-16}	$Pa^{-n}s^{-1}$	Quartz (Hirth *et al*., 2001) Feldspar (Shelton *et al*., 1981) Olivine (Hirth and Kohlstedt, 2004)
n	Power-law exponent	Quartz = 4 Feldspar = 3.1 Olivine = 3.5	–	
H	Activation enthalpy	Quartz = 135 Feldspar = 163 Olivine = 480	$kJmol^{-1}$	

We use a flow law for wet quartzite. Wet olivine rheology is used for the mantle. For feldspar hydrolytic weakening has not been reported. We use a classical flow-law by Shelton *et al*. (1981).

In our application of isotropic visco-plasticity the elastic spin, equivalent to vorticity in fluids, is considered through the use of the corotational rate of the Kirchhoff stress also known as the Jaumann stress-rate tensor (Regenauer-Lieb and Yuen, 1998).

References

Albert, R.A. and Phillips, R.J. (2002), *Time-dependent effects in elasto-visco-plastic models of loaded lithosphere*, Geophys. J. Internat. *151*(2), 612–621.

Bercovici, D. and Ricard, Y. (2003), *Energetics of a two-phase model of lithospheric damage, shear localization and plate-boundary formation*, Geophys. J. Internat. *152*, 581–596.

Chery, J., Vilotte, J.P., and Daigniers, M. (1991), *Thermomechanical evolution of a thinned continental lithosphere under compression: Implication for the Pyrenees*, J. Geophys. Res. *96*(B3), 4385–4412.

Chrysochoos, A. and Belmahjoub, F. (1992), *Thermographic analysis of thermomechanical couplings*, Archives Mechanics *44*(1), 55–68.

Chrysochoos, A. and Dupre, J.C. (1992), *An infrared set-up for continuum thermomechanics*, J. Societé Française des Thermiciens *27*, 129.

Collins, I.F. (2005), *Elastic/plastic models for soils and sands*, Internat. J. Mech. Sci. *47*(4–5 Spec. Issue.), 493–508.

COLLINS, I.F. and HILDER, T. (2002), *A theoretical framework for constructing elastic/plastic constitutive models of triaxial tests*, Internat. J. Num. Analyti. Methods in Geomech. *26*(13), 1313–1347.

DIETERICH, J.H. (1979a), *Modeling of rock friction. 1. Experimental results and constitutive equations*, J. Geophys. Res. *84* (NB5), 2161–2168.

DIETERICH, J.H. (1979b), *Modeling of rock friction . 2. Simulation of pre-seismic slip*, J. Geophys. Res. *84* (NB5), 2169–2175.

EMMERSON, B. and MCKENZIE M. (2007), *Thermal structure and seismicity of subducted lithosphere*, Physics of Earth and Plane. Inter. *163*(1–4), 191–208.

GERBAULT, M., HENRYS, S., and DAVEY, F. (2003), *Numerical models of lithospheric deformation forming the Southern Alps of New Zealand*, J. Geophys. Res. B: Solid Earth *108* (7).

GOETZE, C. and EVANS, B. (1979), *Stress and temperature in the bending lithosphere as constrained by experimental rock mechanics*, Geophys. J. Roy. Astron. Soc. *59*, 463–478.

HIRTH, G. and KOHLSTEDT, D. (2004), *Rheology of the upper mantle and the mantle wedge: A view from the experimentalists*. In (J. Eiler, ed.), *The Subduction Factory Geophysical Monograph*, Am. Geophys. Union, Washington, pp. 83–105.

HIRTH, G., TEYSSIER, C., and DUNLAP, W.J. (2001), *An evaluation of quartzite flow laws based on comparisons between experimentally and naturally deformed rocks*, Internat. J. Earth Sci. *90* (1), 77–87.

IDE, S., BEROZA, G.C., SHELLY, D.R., and UCHIDE, T. (2007), *A scaling law for slow earthquakes*, Nature *447* (7140), 76–79.

JORDAN, T. (1991), *Low-frequency characteristics of slow earthquakes and their implications for near-field monitoring of precursory phenomena*, Seismol. Res. Lett. *62* (1), 40.

KAMEYAMA, C., YUEN, D.A., and KARATO, S. (1999), *Thermal-mechanical effects of low temperature plasticity (the Peierls mechanism) on the deformation of a viscoelastic shear zone*, Earth Planet. Sci. Lett. *168*, 159–162.

KELEMEN, P. and HIRTH, G. (2007), *A periodic shear-heating mechanism for intermediate-depth earthquakes in the mantle*, Nature *446*, 787–790.

KOHLSTEDT, D.L., EVANS, B., and MACKWELL, S.J. (1995), *Strength of the lithosphere: Constraints imposed by laboratory measurements*, J. Geophy. Res. *100* (B9), 17587–17602.

LYAKHOVSKY, V., BEN-ZION, Y., and AGNON, A. (2001), *Earthquake cycle, fault zones, and seismicity patterns in a rheologically layered lithosphere*, J. Geophys. Res.-Sol. Earth *106* (B3), 4103–4120.

MANCKTELOW, N.S. (2006), *How ductile are ductile shear zones*? Geology *34* (5), 345–348.

ORD, A., HOBBS, B.E., and REGENAUER-LIEB, K. (2004), *A smeared seismicity constitutive model*, Earth, Planets and Space *56* (12), 1121–1133.

PFLUKE, J. (1978), *Slow earthquakes and very slow earthquakes*. In M.P.U.S. Geol. Surv., Calif., United States (USA) (ed.). Open-File Report — U. S. Geological Survey, pp. 447–468.

REGENAUER-LIEB, K., WEINBERG, R., and ROSENBAUM, G. (2006), *The effect of energy feedbacks on continental strength*, Nature *442*, 67–70.

REGENAUER-LIEB, K. and YUEN, D. (1998), *Rapid conversion of elastic energy into shear heating during incipient necking of the lithosphere*, Geophys. Res. Lett. *25* (14), 2737–2740.

REGENAUER-LIEB, K. and YUEN, D. (2006), *Quartz Rheology and short time-scale crustal instabilities*, Pure Appl. Geophys. *163* (9), 1915–1932.

REGENAUER-LIEB, K. and YUEN, D.A. (2003), *Modeling shear zones in geological and planetary sciences: solid- and fluid- thermal- mechanical approaches*, Earth Sci. Rev. *63*, 295–349.

REGENAUER-LIEB, K. and YUEN, D.A. (2004), *Positive feedback of interacting ductile faults from coupling of equation of state, rheology and thermal-mechanics*, Phys. Earth Plane. Inter. *142* (1–2), 113–135.

ROSENBAUM, G., REGENAUER-LIEB, K., and WEINBERG, R.F. (2005) *Continental extension: From core complexes to rigid block faulting*, Geology *33* (7), 609–612.

SACKS, I., SUYEHIRO, S., LINDE, A., and SNOKE, J. (1978), *Slow earthquakes and stress redistribution*, Nature *275*, 599–602.

SHELTON, G.L., TULLIS, J., and TULLIS, T. (1981), *Experimental high temperature and high pressure faults*, Geophys. Res. Lett. *8* (1), 55–58.

STEIN, S. and OKAL, E. (2005a), *Ultra-long period seismic moment of the great December 26, 2004 Sumatra earthquake and implications for the slip process*.

STEIN, S. and OKAL, E.A. (2005b), *Speed and size of the Sumatra earthquake*, Nature *434* (7033), 581.

WIJNS, C., WEINBERG, R., GESSNER, K., and MORESI, L. (2005), *Mode of crustal extension determined by rheological layering*, Earth Plane. Sci. Lett. *236* (1–2), 120.

ZIEGLER, H., *An Introduction to Thermomechanics* (North-Holland Publishing Company, Amsterdam, 1983.) 358 pp.

(Received November 2006, revised July 15, 2007, accepted October 4, 2007)
Published Online First: April 2, 2008

Pure appl. geophys. 165 (2008) 545–565
0033–4553/08/030545–21
DOI 10.1007/s00024-008-0323-y

Pure and Applied Geophysics

Parallel Fault Systems with Evolving Self-similar Sliding Zones

A. V. Dyskin

Abstract—Catastrophic fault sliding is preceded by the development of sliding zones which grow further driven by the excess of the shear stress over friction at the loci of initiation. This growth is strongly affected by the interaction between the sliding zones. We propose a model of development of such zones based on two major simplifications. Firstly, each sliding zone is modelled as a disc-like shear crack driven by a pair of concentrated forces representing the excess of the shear stress over friction at the loci of initiation. Secondly, the interaction between these cracks is modelled based on the assumption that the distribution of their sizes is self-similar and the self-similarity is maintained in the process of their growth. We show that for parallel cracks the latter is only possible if the sliding zones are localised in a narrow layer. In this case the exponent and the prefactor of the distribution function are uniquely determined. The addition of new sliding zones does not change the distribution but rather increases the upper cut-off. This happens either by instantaneous growth of each added sliding zone to the maximum size producing the strongest microseismic event or by initiating a cascade of intermediate growth producing a series of smaller events. We determine the energy distribution associated with the cascade and the probability of hazardous events. We show that measuring the statistical properties of seismic energy alone is not sufficient for determining the parameters of the model; monitoring of fault deformation is also needed.

Key words: Shear crack, crack interaction, self-similar distributions, effective characteristics, stress intensity factors.

1. Introduction

One of the main mechanisms of earthquakes is catastrophic sliding over a fault, which occurs rapidly and then propagates dynamically (e.g., Scholz, 1990). In principle, if the original stress state consisting of the gravitational and tectonic components is known together with the stress redistribution caused by the Earth's crust structure and properties, then complete determination of the magnitude of the resulting seismic event can be achieved provided that the strength parameters of the fault are known. The most difficult is the determination of the fault strength properties, since the parameters determined by retrospective analysis of the post failure situation (post-mortem examination) do not have to be relevant to a next sliding which will occur in conditions changed by the previous fault sliding. Also a possibility of some strength increase with time (healing, e.g., Scholz,

School of Civil and Resource Engineering, The University of Western Australia, 35 Stirling Hwy, Crawley, WA 6009, Australia.

1990; ABINANTE and KNOPOFF, 1995) should be taken into account. In this case the main tool for prediction of the sliding or, at least the risk assessment, is seismic monitoring. The major challenge in seismic monitoring is the development of interpretation methods which, in order to be realistic, must be based on understanding the mechanics of catastrophic fault sliding.

In many cases the mechanical behaviour of the fault is controlled by the rock mass structure encompassing a number of scales. Then the assumption of self-similarity in distributions of microstructural elements becomes a major simplifying factor in an otherwise usually intractable problem. There is strong evidence of self-similar properties of both rocks and the Earth's crust (e.g., SADOVSKIY, 1983; SCHOLZ and AVILES, 1986; SCHOLZ, 1990; REDNER, 1990; OLDING, 1992; BARTON and ZOBACK, 1992; TURCOTTE, 1993, 2004; GILLESPIE *et al*., 1993; YAMAMOTO *et al*., 1993; DUBOIS, 1998).

The appearance of self-similar structures is usually attributed to the critical state of the material (e.g., BAK and TANG, 1989; CHOPARD and DROZ, 1998), however, the particular mechanism of formation of self-similar distributions, particularly distributions of cracks and fractures, is poorly understood. The most popular approach is to consider the fractures as clusters of connected defects (e.g., SAHIMI and GODDARD, 1986; NISHIUMA *et al*., 1996; CHAKRABATI and BENGUIGUI, 1997; MISHNAEVSKY, 1998) which, near the critical state (i.e., at the percolation threshold), have self-similar distributions. It should however be noted that only in the 2D picture did these structures actually break the material. In the real 3D world the formation of such structures does not affect the connectedness of the body.

DYSKIN (2001, 2002) proposed a mechanism of developing isotropic self-similar distributions of disc-like cracks, based on crack interaction and leading to a self-similar distribution of crack sizes, with the distribution function proportional to the inverse fourth power of the crack radius. Essential in this model is the stable growth of the cracks, which is provided by a special type of loading, *viz.* by a couple of concentrated forces applied at the centre of every disc-like crack. DYSKIN (2005) proposed a 2D model of growth of sliding zones and demonstrated that the self-similar distributions of sizes of sliding zones can maintain themselves only in the case of isotropically oriented faults and in the case of sliding zones localised in a narrow zone. The 2D model may, however, have only restricted applications because the 3D effects can play a significant role in fracture propagation, especially in the presence of compression (e.g., DYSKIN *et al*., 2003).

In this paper we develop a 3D model of interaction of shear cracks in self-similar systems and the evolution of sliding zones. The paper is organised as follows. In *Section* 2 we introduce a model of evolution of a single sliding zone in a pre-existing fault. We assume that the sliding zone is developed due to a local deterioration of friction in the fault independently of the deformation rate or slip velocity. This locus is then considered as a region of excess of shear force and serves as a driving force to propagate the sliding zone further. We model such a sliding zone as a planar disc-like shear crack in homogeneous elastic medium, ignoring the real discontinuities as the purpose of this

study is to investigate a mechanism of emergence of self-similar distributions. We assume that the features that introduce lengths scales, such as fault curvature or blocks are very large as compared to the dimensions of the sliding zones. We use a conventional Linear Elastic Fracture Mechanics criterion of their propagation (see, for instance, SCHOLZ, 1990), the fracture energy being a macroscopic substitute for microscopic mechanisms of sliding over the fault. *Section* 3 treats the effect of interaction between these sliding zones. It is based on the assumption that because the sliding zones could not be born equal, the interaction only exaggerates the size inequality and produces a wide distribution of sizes (sliding zones of similar sizes are in low concentration and hence do not interact directly; the interaction is essential only between the zones of considerably different sizes). Under this assumption a model case of chaotically oriented disc-like shear cracks is considered and demonstrated that interaction leads the emergence of a self-similar crack distribution. To circumvent the technical difficulties pertinent to the realistic case of parallel cracks, a self-similar size distribution needs to be considered and checked whether it survives the process of evolution of the sliding zones. The necessary formalism is introduced in *Section* 4. Based on this, *Section* 5 demonstrates that in the case of distributed parallel sliding zones the self-similar distribution could not be formed. *Section* 6 considers another limiting case of coplanar sliding zones and demonstrates the stability of their self-similar size distribution. *Section* 7 infers the statistical parameters of seismic energy distribution associated with the evolution of self-similar distributions of sliding zones.

2. A Model of Sliding Zone

We consider a fault in a meta-stable state arising when sufficient time has elapsed after a previous sliding such that a cementation or another healing process has restored the cohesion. This condition will be expressed in terms of the large-scale (macroscopic with respect to the dimensions of the sliding zones) shear, τ and normal, σ_N, stress components, acting on the fault plane by assuming that it is the cohesion that makes the fault stable, while if the cohesion is removed, the fault is in moving equilibrium. This is expressed as follows (compressive stresses are assumed positive)

$$\tau < c + \sigma_N \tan\varphi, \quad \tau = \sigma_N \tan\varphi, \tag{1}$$

where φ is the friction angle and c is the current cohesion.

The stress variations with time and the time effects, like cyclic loading caused by tidal stress and neighbouring seismic events or delayed fracture producing friction deterioration can generate local sliding zones at weak places of the fault, which, as yet, do not affect the fault stability. The local sliding zones can also be formed if the fault is locally hit by a tensile wave from a neighbouring seismic event with a magnitude exceeding the normal compressive stress (this mechanism was considered by DYSKIN *et al*., 1998). We will call the sliding zones initiated by one of these mechanisms the

initial sliding zones. Such a zone is shown in Figure 1 where the fault is sketched with its middle section, AA'. The initial sliding zone (denoted by CC' in Fig. 1) is characterised by the sliding resistance, τ_I, which is considerably lower than the large-scale friction on the fault:

$$\tau_I = c_I + \sigma_N \tan \varphi_I < \sigma_N \tan \varphi. \tag{2}$$

The initial sliding zone will propagate to a sliding zone of a certain size (denoted by BB' in Fig. 1). We assume that the cementation in this zone is broken and the resistance to sliding is only provided by dilation on asperities in accordance to a conventional model (PATTON, 1966).

We firstly consider the case when the sliding zone is alone (the following sections will consider the interaction of self-similarly distributed sliding zones). Then its development can be modelled by a shear crack loaded at a central part (segment CC' in the middle section) by shear tractions $\tau - \tau_I$ (Fig. 2a). Since we are going to consider the potentially extensive development of the sliding zones to sizes much greater than the size of the initial zone we will, as a further simplification, model the sliding zone as a crack sheared by a pair of concentrated forces of average magnitude $F = S(\tau - \tau_I)$, where S is the area of the initial sliding zone (Fig. 2b).

The growth of such sliding zones is controlled by the fault microstructure, in particular the fault roughness and the strength of the gouge. Macroscopically, these

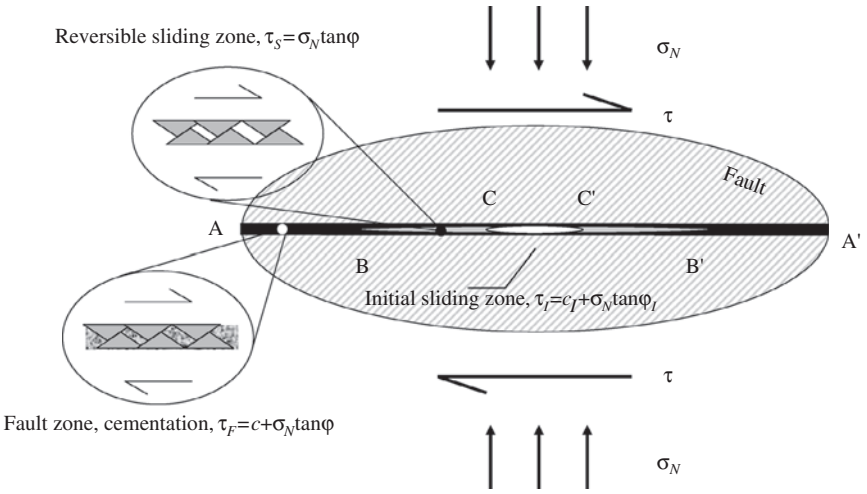


Figure 1

Schematic representation of a fault with a sliding zone. A cross section is shown. The sliding zone with cross-sectional dimensions BB' is developed from an initial sliding zone CC'. It is assumed that the large-scale friction in the sliding zone is provided by interacting asperities with negligible friction between the asperity surfaces, such that the sliding over this part of the fault could, in the first approximation, be assumed reversible.

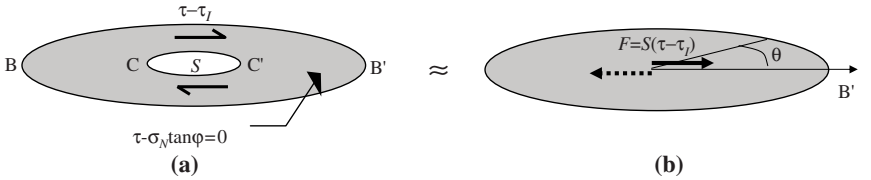


Figure 2
A model of the sliding zone: (a) A representation of the sliding zone as a shear crack loaded at the central part (marked in white) of area S which corresponds to the initial sliding zone; (b) modelling the sliding zone as a shear crack loaded by a pair of concentrated forces of magnitude $F = S(\tau - \tau_I)$, the angle θ is a polar angle.

factors will be represented by the macroscopic (with respect to the characteristic size of the fault asperities) fracture energy, Γ. Then, neglecting the contribution of the dilatant opening of the fault to the energy release rate, a crack propagation type criterion can be formulated for the growth of a sliding zone:

$$\frac{1-\nu^2}{E}K_{II}^2 + \frac{1+\nu}{E}K_{III}^2 = \Gamma, \tag{3}$$

where E, ν are Young's modulus and Poisson's ratio of the surrounding material of the Earth's crust which, in this simplified model, is assumed isotropic and homogeneous elastic.

For a disc-like crack shown in Figure 2b, the Mode II and III stress intensity factors K_{II} and K_{III} have at point B' the following form (e.g., TADA *et al.*, 1985)

$$K_{II} = \frac{F}{(\pi R)^{3/2}}\frac{2}{2-\nu}(1+\nu)\cos\theta, \quad K_{III} = -\frac{F}{(\pi R)^{3/2}}\frac{2}{2-\nu}(1-2\nu)\sin\theta, \tag{4}$$

where R is the crack radius, the polar angle θ is shown in Figure 2.

Substitution of Equations (4) into (3) shows that the propagation criterion for the sliding zone is not uniform over the crack contour, which means that even if the initial shape was circular, it will not be maintained in the process of the growth of the sliding zone. Nevertheless, since the present analysis is mainly concerned with the effects of interaction, as an initial approximation we consider a simplified model of the sliding zone. In this model the sliding zone will still be treated as a disc-like shear crack with the propagation criterion based on a certain average value of the stress intensity factors over the crack contour. Subsequently, since the expressions for both K_{II} and K_{III} have the same factor $F(\pi R)^{-3/2}$, the criterion of growth of the sliding zone can, in this approximation, be expressed as

$$\frac{F}{(\pi R)^{3/2}} = K_*, \tag{5}$$

where K_* is a force representation of the average fracture energy $\bar{\Gamma}$.

$$\frac{1-\nu^2}{E}K_*^2 = \bar{\Gamma}.$$

We assume that K_* is scale-independent (with respect to the scales exceeding the characteristic dimensions of the fault asperities).

As evident from (5), the sliding zone grows in a stable manner with the radius increasing with the force as $F^{2/3}$. This implies that the growth of sliding zones is controllable and hence excludes the dynamic phase. It means that the growth of separate sliding zones by themselves cannot cause catastrophic sliding. The mechanism of large seismic events should therefore be sought in the effect of interaction between the sliding zones. This mechanism is considered in the following section.

3. Interacting Shear Cracks. Emergence of Self-similarity

Interaction between sliding zones literally means that each zone propagates under the action of a superposition of the external stress and stress disturbances, generally non-uniform, generated by all other sliding zones. When the number of sliding zones is large, the modelling of propagation of interacting sliding zones becomes quite complicated.

A considerable simplification can be achieved by employing the notion of self-similarity in the distribution of the sliding zones, which is suggested by the Guttenberg-Richter law universally observed in the records of both natural and technogenic seismic events (e.g., SCHOLZ, 1990; GIBOWICZ and KIJKO, 1994). Self-similarity implies that the scaling is described by power law, which is the major simplifying factor in addressing the problem of interaction, as will be shown in the following section. More importantly, in some cases, including the case of sliding zones considered, the interaction itself can produce self-similar distributions as will be demonstrated below (the consideration below will make use of the method developed by DYSKIN, (2001, 2002) for Mode I cracks).

We consider a material with cracks growing in a stable manner, for instance driven by pairs of equivalent concentrated forces applied to the centres of the crack faces. Suppose that the cracks are distributed randomly. Then, even if all cracks were initially of the same size and were loaded by exactly the same forces, the interaction will make them grow differently such that a certain size distribution of cracks will emerge. One can then assume that the difference in the crack sizes will only increase with their growth. Thus the interaction of such cracks can be modelled in the asymptotics of large distribution of sizes (SALGANIK, 1973) assuming that: (i) cracks of close sizes do not interact directly and; (ii) the interacting cracks are very different in size. Then each crack can be considered in an equivalent medium with effective characteristics determined by all cracks of smaller sizes.

We start the consideration with a somewhat artificial case of isotropic distribution of sliding zones (shear cracks), meaning that the faults subject to sliding are randomly oriented. We will further assume that the asperity-resisted sliding can be reversible such

that small deviations of the shear stresses from the friction stress can cause both an increase and a reduction in the crack sliding. This assumption is essential for the proposed use of the effective characteristic theory since the basis of it—the consideration of a crack in an effective medium determined by smaller cracks—presumes that the effective medium is elastic at least incrementally at the stress near the equilibrium point, such that loading and unloading follow the same path. Modelling of this case will provide the concept which will then be extended to the case of parallel faults which is more realistic for describing fault sliding.

Direct modelling of the growth of large numbers of interacting cracks is quite complex, especially in 3D. A significant simplification can be achieved if the crack growth, more precisely its stable phase, is formulated in terms of the average stress intensity factors $\langle K_I \rangle$, $\langle K_{II} \rangle$, $\langle K_{III} \rangle$, where the averaging is presumed over all possible mutual locations and, possibly, orientations of the interacting cracks.[1]

We now consider each crack as being in an effective medium whose characteristics are determined by smaller cracks, the calculation of the average stress intensity factors involves two scales. The effective medium determines the overall displacement discontinuities over the crack surface, therefore the computations involving the effective medium correspond to the scale *macroscopic with respect to the small cracks*. The stress intensity factors on the other hand reflect the stress concentration in a vicinity of the crack contour at the distances closer than the dimensions of the small cracks. Hence the stress intensity factors represent the stress distributions at the scale *microscopic with respect to the small cracks*. Consequently, the average stress intensity factors should be determined from the distribution of the displacement discontinuity close to the crack contour. Since here the scale of consideration is microscopic with respect to the small cracks, the respective stress concentration corresponds to the part of the original material free from the small cracks. Hence, the elastic moduli relating to this part of displacement discontinuity and the stress intensity factors are at this stage the elastic moduli of the material. For the isotropic case the average stress concentrations can therefore be expressed in the following form (see DYSKIN, 2002 for details)

$$\frac{\langle K_I \rangle}{K_I^0}, \frac{\langle K_{II} \rangle}{K_{II}^0}, \frac{\langle K_{III} \rangle}{K_{III}^0} \sim \frac{E_0}{E}, \tag{6}$$

where E is the effective Young's modulus, E_0, is the Young's modulus of the material, K_I^0, K_{II}^0, K_{III}^0 are the stress intensity factors which would have characterised the crack had there

[1] The usual perception is that the fracture processes are described by extreme quantities, be it a maximum stress or minimum strength. This is of course relevant to the stage when a critical crack is formed and dynamically propagates producing the ultimate failure. However if the ultimate failure is preceded by a considerable stage of multiple crack accumulation/growth without dynamics, the use of average stress intensity factors would more adequate.

been no interaction with small cracks. Then the criterion of propagation of the sliding zone (5) assumes the form

$$K_* = \frac{F}{(\pi R)^{3/2}} \frac{E_0}{E}. \tag{7}$$

Presume that at a certain value of the load, F, the cracks have the distribution function $f(R)$, $R_0 \leqslant R \leqslant R_{\max}$. Here R_0 is the radius which the crack (sliding zone) would have without the interaction, $R_{\max}$ is the maximum radius attained under the given value F. It will also be assumed that this crack distribution can approximately be considered as the wide distribution of sizes.

Based on this assumption, the effective moduli for the material with cracks up to the given size can be calculated. Consider cracks with radii between R and $R + \mathrm{d}R$. Because the number of these cracks is low, they shall be considered as non-interacting and placed in an effective medium determined by cracks of radii smaller than R. Let the Young's modulus and Poisson's ratio of this medium be $E(R)$ and $\nu(R)$. Then, following SALGANIK's (1973) differential self-consistent method, the effective characteristics for the medium with the new cracks of radii between R and $R + \mathrm{d}R$ are obtained as the effective characteristics of a medium with moduli $E(R)$ and $\nu(R)$. The concentration of these new cracks is:

$$d\nu(R) = NR^3 f(R) dR, \tag{8}$$

where N is the number of cracks per unit volume.

The concentration of these new cracks being infinitesimal allows neglecting of the interaction between them. For the case of pure shear disc-like cracks (no opening or closure, i.e., normal displacement is continuous through the crack) the effective moduli in the approximation of low concentrations ($\nu \ll 1$) have the form (see Appendix):

$$E = E_0 \left[1 - \frac{64}{45} \frac{1 - \nu_0^2}{2 - \nu_0} \nu \right], \quad \nu = \nu_0 + \frac{32}{45} \frac{1 - \nu_0^2}{2 - \nu_0} (1 - 2\nu_0)\nu. \tag{9}$$

Equation (9) can be further simplified by taking into account that since $0 \leqslant \nu_0 \leqslant 0.5$,

$$0.5 \leqslant \frac{1 - \nu_0^2}{2 - \nu_0} \leqslant 0.536, \tag{10}$$

such that with the accuracy of 7.2% this fraction can be replaced with 1/2. Subsequently,

$$E \approx E_0 \left[1 - \frac{32}{45} \nu \right], \quad \nu \approx \nu_0 + \frac{16}{45} (1 - 2\nu_0)\nu. \tag{11}$$

According to the differential self-consistent method, at each step when we apply formula (9) or the approximation (11), the moduli obtained at the previous step play the role of E_0 and ν_0. Consequently,

$$E(\mathrm{v}+d\mathrm{v}) = E(\mathrm{v})\left[1-\frac{32}{45}d\mathrm{v}\right],$$
$$v(\mathrm{v}+d\mathrm{v}) = v(\mathrm{v})+\frac{16}{45}[1-2v(\mathrm{v})]d\mathrm{v}.$$

We now obtain from the first Equation of the above system and equation (8) that

$$\frac{dE}{dR} = -\frac{32}{45}ENR^3 f(R). \tag{12}$$

Expressing E from (7) differentiating it with respect to R and substituting into (12) one obtains

$$f(R) = \frac{135}{64N}\frac{1}{R^4}. \tag{13}$$

The distribution function should satisfy the usual normalisation condition

$$\int_{R_0}^{R_{\max}} f(R)\, dR = 1. \tag{14}$$

This gives

$$\frac{1}{R_0^3}-\frac{1}{R_{\max}^3} = \frac{64}{45}N. \tag{15}$$

Expressing R_0— the crack radius in the absence of interaction—from (5):

$$R_0 = \frac{1}{\pi}\left(\frac{K_*}{F}\right)^{2/3}, \tag{16}$$

and substituting into Equation (15) one has

$$\frac{R_{\max}}{R_0} = \left[1-\frac{64F^2}{45\pi^3 K_*^2}N\right]^{-1/3}.$$

As the expression in the square brackets tends to zero, the ratio $R_{\max}/R_0 \to \infty$ which can be interpreted as complete (catastrophic) sliding of the fault. The corresponding force magnitude is denoted as $F_{\max}$:

$$F_{\max} = \frac{3\pi K_*}{8}\sqrt{\frac{5\pi}{N}}. \tag{17}$$

From here one obtains

$$\frac{R_{\max}}{R_0} = \left[1-\frac{F^2}{F_{\max}^2}\right]^{-1/3}. \tag{18}$$

The power law (13) together with $R_{\max}/R_0 \to \infty$ can be interpreted as an emergence of a self-similar distribution in the crack arrangement considered. In more complex cases it is

technically more difficult to trace the emergence of self-similar distributions. What we propose instead is to determine whether the self-similar distributions are stable with respect to the described type of crack growth. This necessary condition of the self-similar distributions will be analysed in the following sections.

4. Mechanics of Materials with Self-similar Crack Distributions

4.1. General Considerations

Let the crack distribution be self-similar such that there is no characteristic size in the microstructure. According to DYSKIN (2004) a material with such crack distribution should be modelled simultaneously at all scales by a continuous set of continua (the H-continua) with the volume element sizes, H, assuming all values. In this case, all continuum quantities also should be functions of scale, H. Then all characteristics of the continua become the power functions of H. Furthermore, it was proven that all tensorial properties should scale isotropically, i.e., all tensorial components should scale with the same exponent. In particular, the tensors of elastic moduli, $\mathbf{C}$, and compliances, $\mathbf{A}$, in a Cartesian coordinate frame (x_1, x_2, x_3) must scale for any crack orientations and any material anisotropy as

$$C_{ijkl}(H) = c_{ijkl}H^{\alpha}, \quad A_{ijkl}(H) = a_{ijkl}H^{\beta}, \quad i,j,k,l = 1,2,3, \quad \alpha = -\beta \tag{19}$$

for all non-zero components of the prefactors. Here α and β are the common scaling exponents for the tensors of elastic moduli and compliances, respectively.

The prefactors and exponents can be determined from the following system of equations, if the expressions for the scaling for the contribution of cracks to the compliances or moduli, $\Delta A_{ijkl} = \Delta a_{ijkl}H^{\rho_A}$ or $\Delta C_{ijkl} = \Delta c_{ijkl}H^{\rho_C}$ are known for the case of non-interacting cracks. Then using the differential self-consistent method and taking into account that according to the dimension analysis the exponent $\rho_A = \beta - 1$ or ($\rho_C = \alpha - 1$) one obtains the following system of scaling equations (see details in DYSKIN, 2004)

$$\beta a_{ijkl} = \Delta a_{ijkl} \quad \text{or} \quad \alpha c_{ijkl} = \Delta c_{ijkl}. \tag{20}$$

Consider, for instance, the first system (the second system can be treated similarly). This is generally a system of 21 equations for 22 unknowns, a_{ijkl} and β. Since the prefactors for both compliances and the increments have the same units, one of the compliance prefactors can be chosen arbitrarily, while the other prefactors and the exponent can be found by solving this system.

For a special case of self similar crack distributions $f(R) = wR^{-4}$, to which distribution (13) belongs (with the concentration factor $w = 135/64N$), it is shown by DYSKIN (2002) that the assumptions of the wide distribution of sizes are satisfied and that the differential self-consistent method can be used to determine Δc_{ijkl}.

Furthermore, in line with (6), the average SIFs scale as

$$\langle K_I \rangle,\ \langle K_{II} \rangle,\ \langle K_{III} \rangle \sim H^{-\alpha}. \tag{21}$$

We consider now two important particular cases.

4.2. Isotropic Self-similar Distributions of Disc-like Shear Cracks

In the case of randomly oriented pure shear cracks the second system of scaling Equation (20) can be obtained from (11) by considering only the crack contribution to the moduli and then by formal replacement of v with w. We also take note that the Poisson's ratio being bounded has the scaling exponent equal to zero, i.e., $v = \text{const}$. This produces the following system of scaling equations:

$$\begin{cases} \alpha E = -E\dfrac{32}{45}w, \\ 0 = \dfrac{16}{45}(1-2v)w. \end{cases} \tag{22}$$

The solution of system (22) reads

$$E = eH^{\alpha}, \quad \alpha = -\frac{32}{45}w, \quad v = 0.5, \tag{23}$$

where e is a normalising prefactor. It is seen that both Young's modulus and Poisson's ratio scale according to power laws. The exponent for the Young's modulus is proportional to the concentration factor w. It is interesting that the value for the Poisson's ratio is found to be 0.5, suggesting that materials with self-similar distributions of isotropically oriented pure shear cracks are incompressible. This is because self-similar distributions are formally unbound. Hence, formally, for finite w, the total crack concentration is infinite such that the properties of materials with self-similar crack distributions are fully controlled by the crack behaviour. In this case the incompressibility is a result of the absence of any normal relative displacements of the faces of shear cracks.

4.3. Parallel Disc-like Shear Cracks

Consider now a case of parallel disc-like shear cracks which could model sliding zones over a distributed set of parallel faults. Assume the cracks are oriented perpendicular to the x_3 axis and distributed self-similarly with the distribution function $f(R) = wR^{-4}$. For this case the effective compliances can be found from the general solution by Vavakin and Salganik (1978) for a transverse-isotropic material with disc-like cracks parallel to the plane of isotropy. The Hooke's law for a transverse isotropic material is expressed in the coordinate set (x_1, x_2, x_3) through the compliances as follows

$$
\begin{aligned}
\varepsilon_{11} &= A_{11}\sigma_{11} + A_{12}\sigma_{22} + A_{13}\sigma_{33}\\
\varepsilon_{22} &= A_{12}\sigma_{11} + A_{11}\sigma_{22} + A_{13}\sigma_{33}\\
\varepsilon_{33} &= A_{13}\sigma_{11} + A_{13}\sigma_{22} + A_{33}\sigma_{33}\\
\varepsilon_{23} &= \frac{1}{2}A_{44}\sigma_{23}\\
\varepsilon_{13} &= \frac{1}{2}A_{44}\sigma_{13}\\
\varepsilon_{12} &= (A_{11} - A_{22})\sigma_{12}
\end{aligned}
\tag{24}
$$

VAVAKIN and SALGANIK's (1978) solution, after the contribution of the crack in the normal strain in the direction perpendicular to the cracks (the x_3 axis) is set to zero assumes the form:

$$
\begin{cases}
A_{11} = A^0_{11}, \quad A_{12} = A^0_{12}, \quad A_{13} = A^0_{13}\\
A_{33} = A^0_{33} + \dfrac{8v}{3A^0_{11}}\sqrt{B^0\left[A^0_{11}A^0_{33} - \left(A^0_{13}\right)^2\right]}\\
A_{44} = A^0_{44} + \dfrac{16v}{3}A^0_{44}\dfrac{\sqrt{B^0\left[\left(A^0_{11}\right)^2 - \left(A^0_{12}\right)^2\right]}}{A^0_{11}A^0_{44} + \sqrt{\frac{1}{2}B^0A^0_{44}\left[A^0_{11} + A^0_{12}\right]}}\\
B^0 = A^0_{11}A^0_{44} + 2A^0_{13}\left[A^0_{11} - A^0_{12}\right] + 2\sqrt{\left[A^0_{11}A^0_{33} - \left(A^0_{13}\right)^2\right]\left[\left(A^0_{11}\right)^2 - \left(A^0_{12}\right)^2\right]}.
\end{cases}
\tag{25}
$$

Here A^0_{11}, A^0_{12}, A^0_{13}, A^0_{33}, A^0_{44} are the compliances of the material.

The scaling Equations (the first system in (20)) can be obtained by replacing v with w bringing A^0_{ii} to the left-hand sides and then replacing $A_{ii} - A^0_{ii}$ with βa_{ii} and, finally, replacing A^0_{ii} with a_{ii} in the remaining parts. This yields the following scaling equations

$$
\begin{cases}
\beta a_{11} = 0, \quad \beta a_{12} = 0, \quad \beta a_{13} = 0\\
\beta a_{33} = \dfrac{8w}{3a_{11}}\sqrt{B\left[a_{11}a_{33} - (a_{13})^2\right]}\\
\beta = \dfrac{16w}{3}\dfrac{\sqrt{B\left[(a_{11})^2 - (a_{12})^2\right]}}{a_{11}a_{44} + \sqrt{\frac{1}{2}Ba_{44}[a_{11} + a_{12}]}}\\
B = a_{11}a_{44} + 2a_{13}[a_{11} - a_{12}] + 2\sqrt{\left[a_{11}a_{33} - (a_{13})^2\right]\left[(a_{11})^2 - (a_{12})^2\right]}
\end{cases}
\tag{26}
$$

Suppose $\beta \neq 0$. Then the first three equations of (26) yield $a_{11} = a_{12} = a_{13} = 0$. Then the third equation produces $\beta = 0$. This contradiction means that the scaling exponent vanishes. As a result we obtain the following scaling law:

$$
A_{ij} = a_{ij}H^{\beta}, \quad \beta = 0. \tag{27}
$$

5. Stability of Self-similarity in the Case of Distributed Sets of Growing Shear Cracks

We are now in a position to check whether the above self-similar distributions of shear cracks are stable with respect to the mechanism of crack growth described in Section 2. Assuming that each homogenisation scale $H \sim R$ (in the H-continuum only cracks of sizes $R > H$ can be seen) and substituting scaling (21) into the expression for the stress intensity factors (4), (6) and then the result into the criterion of shear crack growth (5) one obtains

$$K_* \sim R^{-\alpha-3/2}. \tag{28}$$

From here, since K_*, is assumed to be scale-independent, we obtain an equation for the scaling exponent $-\alpha - 3/2 = 0$. Therefore, if the cracks are to grow keeping the self-similarity, the exponent of the moduli scaling cannot be arbitrary, but should satisfy

$$\alpha = -3/2. \tag{29}$$

For isotropic distributions of disc-like shear cracks this, according to the scaling law (23) corresponds to

$$w = 135/64. \tag{30}$$

Thus the concentration factor of the crack distribution remains constant. This means that in order to maintain the concentration factor constant the crack growth must only affect the lower and upper cutoffs of the distribution (both $R_{\max}$ and R_0 become increased in accordance with the normalisation Equations (14)).

In the case of parallel shear cracks, because the scaling of non-vanishing components of compliances or moduli is characterised by the same exponent, the scaling of the average stress intensity factors still can be obtained from the Equation (6). Therefore the considerations of the previous section can be repeated and the condition of stability (29) can be obtained. However, for the case of a single set of parallel cracks this condition cannot be satisfied due to the trivial scaling ($\alpha = \beta = 0$). Therefore, for such crack distributions, the self-similarity cannot be maintained: It will be destroyed by crack growth. Consequently, one cannot expect that the crack growth will be accompanied by self-similar seismic emission which contradicts the Guttenberg-Richter law. This suggests that the case of distributed crack locations is not realistic and a condition of cracks localised in a narrow band should be considered. This will be accomplished in the following section.

6. Stability of Self-similar Distributions in the Case of Localised Sets of Parallel Sliding Zones

The instability of a self-similar distribution of parallel cracks with respect to the crack growth came from the fact that the scaling exponent vanishes. This, in its own turn, is a

consequence of the fact that these cracks do not contribute to some components of compliances that characterise the transverse-isotropic H-continua which model the material with a single set of distributed parallel cracks. Therefore, in order to find a stable arrangement of parallel cracks, one needs to find a circumstance in which the cracks influence all essential components of compliances. An obvious candidate for this is a localised distribution of parallel cracks, i.e., the distribution in which all cracks are concentrated within a thin layer (Fig. 3). (In order to maintain the self-similarity the layer should be infinitesimally thin; in reality its thickness should be much smaller than the lower cutoff of the crack radii, R_0.) We will model such a set of coplanar cracks as a Winkler layer with the shear stiffness k defined as $\tau = k\Delta u_s$, where Δu_s is the shear displacement discontinuity over the layer (difference in displacements at opposite boundaries of the layer) in response to shear load τ.

Under the assumption of self-similar crack distribution, the stiffnesses should scale with the same exponent (since they are components of a diagonal tensor relating the stress vector and displacement discontinuity vector):

$$k \sim H^{\alpha}. \tag{31}$$

Suppose that the cracks are distributed according to the following distribution function

$$f(R) = \frac{\lambda}{R^m}. \tag{32}$$

Consider an equivalent 2D H-continuum which models the Winkler layer. This H-Winkler layer comprises all cracks of the size up to $R \sim H$. Transition to the scale $H + dH$ leads to the addition of new cracks occupying a relative area of the fault $dw = \pi\lambda R^{2-m}\, dR$. These new cracks increase the average (effective) stress, τ_{eff}, by the factor of $(1 - dw)^{-1}$, because the presence of cracks simply reduces the intact area subjected to loading with a given total force. This results in the reduction of effective stiffness by the factor of $(1 - dw)$. Subsequently

$$\frac{dk}{k} = -\pi\lambda R^{2-m} dR \;\; dk/k = -\pi\lambda R^{2-m} dR. \tag{33}$$

Obviously, the power law is only possible if $m = 3$:

$$f(R) = \frac{\lambda}{R^3}. \tag{34}$$

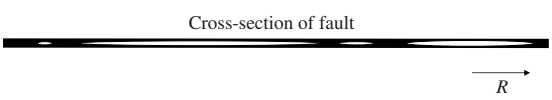


Figure 3
Coplanar sliding zones in a fault: A cross Sectional view.

Solution of the differential equation (33) gives

$$\alpha = -\pi\lambda, \quad k \sim R^{-\pi\lambda}, \quad \tau_{\text{eff}} \sim R^{\pi\lambda}. \tag{35}$$

The influence of the interaction on the average stress intensity factors scales inversely to the scaling of stiffness, therefore the criterion of crack growth reads

$$K_* \sim \frac{F}{(\pi R)^{3/2}} R^{\pi\lambda} \sim R^{\pi\lambda - 3/2}. \tag{36}$$

The self-similarity will be preserved if

$$\lambda = \frac{3}{2\pi}. \tag{37}$$

The total dimensionless concentration of cracks with sizes ranging from R_0 to $R_{\max}$ is

$$\Omega = \int_{R_0}^{R_{\max}} R^2 f(R)\, dR = \frac{3}{2\pi} \ln \frac{R_{\max}}{R_0}. \tag{38}$$

Thus, the self-similar distribution of parallel sliding zones in an infinitesimally thin layer is stable with respect to their growth.

This result can obviously be applied to a system of parallel faults, as long as the interaction between the sliding zones belonging to different faults can be neglected. This can for instance happen when sliding occurs in different parts of the faults such that the distance between the sliding zones belonging to different faults is greater than the upper cut-off of the size distribution of the sliding zones.

7. *Self-similar Cascades and the Resulting Distribution of Seismic Energy*

If we assume that the self-similarity of the size distribution of sliding zones is maintained, then the process of their evolution can be described as follows. The new sliding zones which are initially the zones of average radius R_0 induced in the fault will violate self-similarity. Therefore, the self-similarity will have to be restored by growing these new cracks (or some of them) to the highest sizes (radii) or by propagating them to some intermediate sizes followed by increasing the sizes of other cracks. This cascade of propagating sliding zones will result in a new distribution with the same minimal radius R_0 but with an increased maximal one, $R_{\max}$.

In order to model this cascade of propagating cracks consider a level R. Since the cracks grow such that the concentration factor, λ, remains constant (see (37)), the number of cracks grown to radius R should equal the number of cracks of radius R starting their growth in the process of rearrangement. Let $g(R)$ be a fraction of cracks grown to radius R. Due to self-similarity this distribution should be expressed by a power law with a certain exponent k:

$$g(R) = \frac{\gamma}{R^k}. \tag{39}$$

Subsequently the distribution of cracks grown to radius R is (see also summary in Table 1)

$$f(R)g(R) = \frac{\lambda\gamma}{R^{3+k}} \tag{40}$$

Consider now the energy distribution of seismic events emitted by the cascade. Assume the seismic event is associated with a growth of a crack (sliding zone) from a certain size R_1 to size R. We will assume that this growth happens dynamically such that the energy emitted by the crack $\mathrm{E}(R)$ is equal to the change in elastic energy associated with the displacement of the crack faces. Since the crack growth is caused by the crack interaction, represented in this model by some uniform effective stress τ_{eff}, the energy is $\mathrm{E}(R) \sim \tau_{\mathrm{eff}}\,(R^3 - R_1^3)$. Assuming that the majority of growing cracks considerably increase their sizes, $R \gg R_1$ and taking into account the last equation of (35) and equation (37) one obtains

$$\mathrm{E}(R) = \kappa R^{9/2}. \tag{41}$$

One can now express the energy distribution by the probability of events with energy greater that a certain level U:

$$P\{\mathrm{E} > U\} = P\left\{R > \left(\frac{U}{\kappa}\right)^{\frac{2}{9}}\right\} = \int\limits_{\left(\frac{U}{\kappa}\right)^{\frac{2}{9}}}^{\infty} \frac{\pi\lambda\gamma}{R^{3+k}}\,dR.$$

This results in

$$P\{\mathrm{E} > U\} = \frac{\pi\lambda\gamma}{k+2}\left(\frac{\kappa}{U}\right)^{\frac{2}{9}(k+2)}. \tag{42}$$

This is a Guttenberg-Richter law expressed in terms of the energy rather than magnitude. After relating the exponent $2(k+2)/9$ to the observed one it is possible to find k.

8. Determination of the Model Parameters. Risk Assessment

In order to determine the parameters of the model we consider statistical moments of the seismic energy:

Table 1

Summary of self-similar distributions introduced

Distribution	Exponent	Prefactor
$f(R)$ - distribution of sliding zones	-3	$\lambda = 3/(2\pi)$
$g(R)$ - fraction of sliding zones grown to radius$R$	$-k$	γ
$f(R)g(R)$ - distribution of sliding zones grown to radius R	$-(3+k)$	$\lambda\gamma$
$\mathrm{E}(R)$ - energy emitted by the sliding zone of radius R	9/2	κ

$$\langle \mathrm{E}^n \rangle = \int_{R_0}^{R_{\max}} \mathrm{E}(R)^n f(R) g(R)\, dR = \frac{\pi \lambda \gamma \kappa^n}{2 + k - 4.5n} \left[\frac{1}{R_0^{2+k-4.5n}} - \frac{1}{R_{\max}^{2+k-4.5n}} \right]. \tag{43}$$

Thus by measuring the consecutive energy moments $\langle \mathrm{E}^n \rangle$ one can determine remaining parameters, κ, R_0, $R_{\max}$ and the combination $\lambda\gamma$. One of the parameters in this combination, λ or γ, should be determined independently. One possibility would be to evaluate λ by monitoring the fault deformation, since this parameter together with R_0 and $R_{\max}$ determines the effective stiffness of the fault.

If these parameters were determined, the earthquake hazard assessment could be performed circumstantial to when the mechanism of seismicity is dominated by the self-similar development of the distribution of sliding zones. We note that the maximum energy of a seismic event has the form:

$$\mathrm{E}_{\max} = \kappa R_{\max}^{9/2}. \tag{44}$$

If this energy exceeds the safe level, $\mathrm{E}_{\mathrm{safe}}$, then the seismic event should be considered as hazardous. The risk (probability) of a hazardous seismic event can be estimated as

$$P_{\mathrm{hazard}} = \int_{(\mathrm{E}_{\mathrm{safe}}/\kappa)^{2/9}}^{R_{\max}} g(R)\, dR = \begin{cases} \dfrac{\gamma}{k-1} \left[\left(\dfrac{\kappa}{\mathrm{E}_{\mathrm{safe}}} \right)^{\frac{2}{9}(k-1)} - \left(\dfrac{\kappa}{\mathrm{E}_{\max}} \right)^{\frac{2}{9}(k-1)} \right], & \text{if } k \neq 1 \\ \dfrac{2}{9} \ln \dfrac{\mathrm{E}_{\max}}{\mathrm{E}_{\mathrm{safe}}}, & \text{if } k = 1 \end{cases} \tag{45}$$

9. Conclusion

We have considered a specific mechanism of seismicity associated with the propagation of sliding zones within a fault in such a way that the distribution of sizes of the sliding zones maintains self-similarity. This is of course a very strong assumption. If the persistent observations of Guttenberg-Richter law constitute a sufficient reason to believe that the self-similarity is a prevailing characteristic of the fault sliding, then we have a method of hazard assessment of earthquakes. It is essential that not all parameters needed for the hazard assessment could be determined from the microseismic measurements; these must be complemented by independent observations of the process of the fault deformation.

Appendix

Effective Characteristics of an Isotropic Material with Randomly Oriented Disc-like Shear Cracks in Dilute Concentrations

Consider an isotropic material with chaotically distributed disc-like cracks and assume that the concentration of the cracks is low such that the interaction between the

cracks can be neglected. Introduce a global Cartesian coordinate set (x_1, x_2, x_3) and consider a volume element macroscopic with respect to the crack sizes. Suppose the volume element is loaded by uniform tractions which, in the absence of the cracks, would produce uniform stress field σ_{ij} within the volume element.

Let V_i be an ith component of displacement discontinuity through the crack integrated over the crack surface (this quantity has the units of volume). In the case of elastic material, vector V_i is linearly related to the applied stress which is considered to be uniform by virtue of the uniform tractions applied to the volume element and the presumed absence of the crack interaction

$$V_k = J_{klm}\sigma_{lm}, \tag{A1}$$

where J_{klm} is a tensor which depends on parameters of the crack and its orientation.

In accordance with the general theory (e.g., SALGANIK, 1973) the effective compliances of the cracked body have the form

$$S_{iklm} = S^0_{iklm} + \frac{1}{2}\int [n_i J_{klm} + n_k J_{ilm}]F(\mathbf{Y})\, d\mathbf{Y}. \tag{A2}$$

Here S^0_{iklm} is the tensor of compliances of the material, n_i is the unit vector normal to the crack surface, $\mathbf{Y}$ is the vector of parameters of the shape and orientation of the crack. Summation over repeated indexes is presumed. The distribution function $F(\mathbf{Y})$ satisfies the following normalisation condition:

$$\int F(\mathbf{Y})\, d\mathbf{Y} = N, \tag{A3}$$

where N is the number of cracks per unit volume.

Consider a crack and introduce a Cartesian coordinate set (x_1', x_2', x_3') with the x_3' axis normal to the crack surface. For a disc-like crack of radius R we have in this coordinate set (SALGANIK, 1973)

$$V_k' = \frac{8\pi}{3}\frac{1-\nu_0^2}{E_0}R^3\sigma_{k3}', \tag{A4}$$

where σ_{ij}' is the applied stress in this coordinate set, E_0 and ν_0 are the Young's modulus and Poisson's ratio of the material.

For shear cracks with $V_3' = 0$, tensor J_{klm} can be written in the coordinate set related to the crack as

$$J'_{klm} = \frac{4\pi}{3}\frac{1-\nu_0^2}{E_0}R^3 C_{(k)}\left[\delta_{(k)l}\delta_{m3} + \delta_{(k)m}\delta_{l3}\right], \quad C_1 = C_2 = \frac{4}{\pi}\frac{1}{2-\nu_0}, \quad C_3 = 0, \tag{A5}$$

where brackets around an index mean that no summation over this index is presumed. In the global coordinate set (x_1, x_2, x_3) this tensor can be written as follows

$$J_{klm} = J'_{pqr} g_{pk} g_{ql} g_{rm}, \tag{A6}$$

Here the directional cosines of the crack coordinate set with respect to the global one can, following SALGANIK (1973), be written as

$$[g_{ik}] = \begin{bmatrix} -\cos\psi\cos f\cos\theta - \sin\psi\sin f & -\sin\psi\cos f\cos\theta + \cos\psi\sin f & \cos f\sin\theta \\ -\cos\psi\sin f\cos\theta + \sin\psi\cos f & -\sin\psi\sin f\cos\theta + \cos\psi\cos f & \sin f\sin\theta \\ \cos\psi\sin\theta & \sin\psi\sin\theta & \cos\theta \end{bmatrix}, \tag{A7}$$

where $\theta \in [0\ \pi]$ is the angle between axes x_3' and x_3, $\Psi \in [0,2\pi]$ is the angle between planes $(x_3'\ x_1')$ and $(x_3'\ x_3)$, $f \in [0,2\pi]$ is the angle between planes $(x_3\ x_1)$ and $(x_3\ x_3')$.
For the isotropic original material the tensor of compliances can be expressed in the form

$$S^0_{iklm} = \frac{1}{E_0}\left\{-\nu_0\delta_{ik}\delta_{lm} + \frac{1}{2}(1+\nu_0)(\delta_{im}\delta_{lk} + \delta_{il}\delta_{km})\right\}. \tag{A8}$$

Substituting Equations (A5)–(A8) into equation (A2), taking into account that $g_{3m} = n_m$, that the columns of matrix (A7) are orthogonal and that for isotropic distribution of crack orientations

$$dY = \frac{1}{8\pi^2}\sin\theta d\theta df d\psi dR, \tag{A9}$$

one obtains

$$S_{iklm} = \frac{1}{E_0}\left\{-\left[\nu_0 + \frac{32}{45}\frac{1-\nu_0^2}{2-\nu_0}\nu\right]\delta_{ik}\delta_{lm} + \frac{1}{2}\left[1+\nu_0 + \frac{32}{15}\frac{1-\nu_0^2}{2-\nu_0}\nu\right](\delta_{im}\delta_{lk} + \delta_{il}\delta_{km})\right\} \tag{A10}$$

where

$$\nu = N\int R^3 F(R)\, dR = N\langle R^3\rangle \tag{A11}$$

is the dimensionless crack concentration; $\langle R^3\rangle$ is the cube of the crack radius averaged over all crack sizes and orientations.

Comparing (A10) and (A9) one finally obtains the expressions for the effective Young's modulus and Poisson's ratio:

$$\begin{aligned} E &= E_0\left[1 - \frac{64}{45}\frac{1-\nu_0^2}{2-\nu_0}\nu\right] \\ \nu &= \nu_0 + \frac{32}{45}\frac{1-\nu_0^2}{2-\nu_0}(1-2\nu_0)\nu \end{aligned}. \tag{A12}$$

Acknowledgment

The author acknowledges the financial support through the Australian Research Council Discovery Grant (DP0559737) and the financial support by the Australian Computational Earth Systems Simulator – Major National Research Facility (ACcESS MNRF). The author also wishes to express thanks to Elena Pasternak for valuable comments.

References

Abinante, M.S. and Knopoff, L. (1995), *Quasidynamic model for earthquake simulations*, Phys. Rev. E. *52*, 5, 5675–5678.

Bak, P., and Tang, C. (1989), *Earthquakes as a self-organised critical phenomenon*, J. Geophys. Res. *94B* (11), 15,635–15,637.

Barton, C.A. and Zoback, M.D. (1992), *Self-similar distribution and properties of macroscopic fractures at depth in crystalline rock in the Cajon Pass scientific drill hole*, J. Geophys. Res. *97B* (4), 5181–5200.

Chakrabati, B.K. and Benguigui, L.G., *Statistical Physics of Fracture and Breakdown in Disordered Systems* (Clarendon Press, Oxford 1997).

Chopard, B. and Droz, M., *Cellular Automata Modelling of Physical Systems* (Cambridge University Press 1998).

Dubois, J. *Non-Linear Dynamics in Geophysics* (John Wiley and Sons, Chichester, New York, Weinheim, Brisbane, Singapore, Toronto 1998).

Dyskin, A.V. (2001), *Self-similar crack distributions: Mechanics of formation and method of modelling*, Intern. J. Fracture *112*, L41–L46.

Dyskin, A.V. (2002), *Self-similar crack patterns induced by spatial stress fluctuations*, Fatigue and Fracture of Engin. Mater. and Struct. *25*, 187–200.

Dyskin, A.V. (2004), *Effective characteristics and stress concentrations in materials with self-similar microstructure*, Int. J. Solids Struct. *42* (2), 477–502.

Dyskin, A.V., *Mining-induced seismicity associated with self-similar propagation of sliding zones*, In *Controlling Seismic Risk. Proc. Sixth Internat Conf on Rockburst and Seismicity in Mines (RaSiM6)* (ed. Potvin, Y. and Hudyma, M.), (UWA 2005) pp. 319–325.

Dyskin, A.V., Galybin A.N., and Brady, B.H., *Catastrophic sliding over a fault caused by accumulation of dilation zones*. In *Mechanics of Jointed and Faulted Rock (MJFR-3)*. (ed. Rossmanith, H.P.), (Balkema 1998), pp. 69–74.

Dyskin, A.V., Jewell, R.J., Joer, H., Sahouryeh, E., and Ustinov, K.B. (2003), *Influence of shape and locations of initial 3-D cracks on their growth in uniaxial compression*, Eng. Fract. Mech. *70*, 15, 2115–2136.

Gibowicz, S.J. and Kijko A., *An Introduction to Mining Seismology* (Academic Press, San Diego, New York, Boston, London, Sydney, Tokyo, Toronto 1994).

Gillespie, P.A., Howard, C.B., Walsh, J.J., and Watterson, J. (1993), *Measurement and characterisation of spatial distribution of fractures*, Tectonophysics, *226*, 113–141.

Mishnaevsky, L., *Damage and Fracture in Heterogeneous Materials. Modelling and Application to the Improvement of Drilling Tool* (Balkema, Rotterdam, Brookfield 1998).

Nishiuma, S., Hasegawa, Y., and Miyazima, S. (1996), *Percolated cracks in a solid containing many initial defects*, Fractals *4* (3), 377–384.

Olding, N.E. (1992), *Network properties of a two-dimensional nature fracture pattern*, Pure App. Geophyse. *138*, 95–114.

Patton, F.D. *Multiple modes of shear failure in rock*. In *Proc. of the First Congress of ISRM* (Lisbone 1966), *1*, 509–513.

Redner, S. *Fragmentation*. In *Statistical Models for the Fracture of Disordered Media* (ed. Herrmann, H.J., Roux, S.) (North-Holland, Amsterdam, Oxford, New York, Tokyo 1990), 321–348.

Sadovskiy, M.A. (1983), *Distribution of preferential sizes in solids*, Transactions USSR Academy of Sciences. Earth Science Series *269* (1), 8–11.

SAHIMI, M. and GODDARD, J.D. (1986), *Elastic percolation models for cohesive mechanical failure in heterogeneous systems*, Phys. Rev. B *33* (11), 7848–7851.

SALGANIK, R.L. (1973), *Mechanics of bodies with many cracks*, Mech. Sol. *8*, 135–143.

SCHOLZ, C.H., *The Mechanics of Earthquakes and Faulting*, (Cambridge University Press, Cambridge, New York, Port Chester, Melbourne, Sydney 1990).

SCHOLZ, C.H. and AVILES, C., *The fractal geometry of faults and folding*. In *Earthquake Source Mechanics*, *Monograph Series* 37. (ed. Das, S., Boatwright, J., Scholz, C.H.) (Washington, D.C., American Geophysical Union 1986), 147–155.

TADA, H., PARIS, P.C. and IRWIN, G.R. *The stress analysis of cracks. Handbook*. Third edition. (New York: ASME Press 1985, vol. II.

TURCOTTE, D.L., *Fractals and Chaos in Geology and Geophysics* (Cambridge University Press 1993).

TURCOTTE, D.L. (2004), *The relationship of fractals in geophysics to "the new science"*, Chaos, Solitons and Fractals *19*, 255–258.

VAVAKIN, A.S. and SALGANIK, R.L. (1978), *Effective elastic characteristics of bodies with isolated cracks, cavities, and rigid nonhomogeneities*, Mech. Sol. *13* (2), 87–97.

YAMAMOTO, H., KOJIMA, K., and TOSAKA, H., *Fractal clustering of rock fractures and its modelling using cascade process*. In *Scale Effects in Rock Masses 93* (ed. Pinto da Cunha) (Balkema, Rotterdam 1993), 81–86.

(Received March 20 2007, accepted July 9, 2007)
Published Online First: April 2, 2008

Pure appl. geophys. 165 (2008) 567–583
0033–4553/08/030567–17
DOI 10.1007/s00024-008-0321-0

Pure and Applied Geophysics

3-D Simulation of Steady Plate Subduction with Tectonic Erosion: Current Crustal Uplift and Free-Air Gravity Anomaly in Northeast Japan

CHIHIRO HASHIMOTO,[1] TOSHINORI SATO,[2] and MITSUHIRO MATSU'URA[1]

Abstract—Free-air gravity anomaly in plate subduction zones, characterized by island-arc high, trench low and outer-rise gentle high, reflects the cumulative effects of long-term crustal uplift and subsidence. In northeast Japan the island-arc high of observed free-air gravity anomaly takes its maximum about the eastern coastline. On the other hand, the current vertical crustal motion estimated from geological and geomorphological observations shows a gentle uplift in the land area and steep subsidence in the sea area with the neutral point near the eastern coastline. Such a discrepancy in spatial patterns between the free-air gravity anomaly and current vertical crustal motion can be ascribed to a change in the mode of crustal uplift and subsidence associated with the initiation of tectonic erosion at the North American-Pacific plate interface. We developed a realistic 3-D simulation model of steady plate subduction with tectonic erosion in northeast Japan on the basis of elastic/viscoelastic dislocation theory. Through numerical simulations with this model we found that simple steady plate subduction brings about the crustal uplift characterized by island-arc high with its maximum about the eastern coastline, while steady plate subduction with tectonic erosion, which is represented by the landward retreat of the plate interface, brings about gentle uplift in the land area and steep subsidence in the sea area with the neutral point near the eastern coastline. Therefore, if we suppose that tectonic erosion started 3–4 million years ago after the long duration of simple steady plate subduction, we can consistently explain both patterns of free-air gravity anomaly and current crustal uplift in northeast Japan.

Key words: Plate subduction, tectonic erosion, crustal uplift, free-air gravity anomaly, plate interfaces, elastic dislocation theory.

1. Introduction

Free-air gravity anomaly in plate subduction zones is generally characterized by island-arc high, trench low, and outer-rise gentle high. In northeast Japan we can observe that the island-arc high of free-air gravity anomaly takes its maximum about the eastern coastline (TOMODA and FUJIMOTO, 1982; SANDWELL and SMITH, 1997). The free-air gravity anomaly reflects the cumulative effects of long-term crustal uplift and subsidence (surface mass excess and deficit) due to steady plate subduction (MATSU'URA and SATO, 1989; SATO and MATSU'URA, 1988, 1992, 1993). However, in northeastern Japan, we can observe the

[1] Department of Earth and Planetary Science, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. E-mail: hashi@eps.s.u-tokyo.ac.jp

[2] Department of Earth Sciences, Faculty of Science, Chiba University, 1-33 Yayoi-cho, Inage-ku, Chiba-shi, Chiba 263-8522, Japan.

significant discrepancy in spatial patterns between free-air gravity anomaly and current vertical crustal motion. Form fluvial-terrace heights, for example, TAJIKARA (2004) has estimated the average crustal uplift rates over northeast Japan arc to be 0.2–0.3 mm/yr for the last 0.12 million years (the late Quaternary period). From marine-terrace heights KOIKE and MACHIDA (2001) have reported the very slow crustal uplift along eastern coastlines in northeast Japan during the last 0.12 million years. From DSDP drilling and seismic reflection records von Huene and his coworkers (VON HUENE *et al*., 1982; VON HUENE and LALLEMAND, 1990; VON HUENE and SCHOLL, 1991) have revealed steep subsidence in a broad region along the landward slope of the Japan trench. On the basis of these geological and geomorphological data we may conclude that the current crustal motion in northeast Japan is characterized by gentle uplift in the land area and steep subsidence in the sea area, with the neutral point near the eastern coastline.

In order to explain the steep subsidence in a broad region along the landward slope of the Japan trench, VON HUENE and LALLEMAND (1990) proposed a model of tectonic erosion that consists of frontal and basal erosion of the overriding plate. The frontal erosion removes rock and sediment masses at the front of the landward trench slope. The basal erosion removes rock masses on the underside of the overriding plate. Both frontal and basal erosion thin the wedge portion of the overriding plate and cause the subsidence of the landward slope of the trench. From the tectonic point of view the basal erosion has more essential meaning than the frontal erosion, because it causes landward retreat of the plate interface. In northeast Japan the landward retreat rate of the plate interface has been estimated to be 2–3 km/Myr (VON HUENE and LALLEMAND, 1990). Recently, HEKI (2004) proposed a mechanical model of basal erosion to simulate crustal deformation in northeast Japan. In his model the process of removing a certain amount of mass on the underside of the overriding plate is represented by the negative crack opening (crack closing) at the plate interface. However, the negative crack opening cannot satisfy the mass conservation of the subduction zone system unless some mechanism of mass transfer to the depths is assumed.

In the present study we propose another mechanical model of basal erosion and carry out 3-D simulation of steady plate subduction with tectonic erosion. In Section 2 we show a basic idea for modelling basal erosion, and construct the mechanical model. In Section 3 we examine the effect of tectonic erosion on vertical crustal motion through numerical simulation. In Section 4 we compute vertical crustal motion in northeast Japan with a realistic 3-D model of plate interface geometry, and demonstrate that the pattern of current vertical crustal motion in northeast Japan can be rationally explained by the steady plate subduction with tectonic erosion.

2. 3-D Modelling of Steady Plate Subduction with Tectonic Erosion

2.1. Representation of Mechanical Interaction at Plate Interfaces

In subduction zones oceanic plates descend beneath overriding plates at constant rates on a long-term average. The occurrence of large interplate earthquakes can be regarded as

the perturbation of steady plate subduction. Thus, we may express crustal deformation in subduction zones as the superposition of the secular change due to steady slip over the whole plate interface and the cyclic change due to stick-slip motion in seismogenic regions. The steady slip motion along a plate interface brings about long-term vertical crustal deformation, characterized by island-arc uplift and trench subsidence (MATSU'URA and SATO, 1989; SATO and MATSU'URA, 1988, 1992, 1993). The stick-slip motion brings about cyclic crustal deformation around the seismogenic regions (HASHIMOTO and MATSU'URA, 2000, 2002).

Mechanical interaction at a plate interface is rationally represented by the increase of tangential displacement discontinuity across it (MATSU'URA and SATO, 1989). In general, the tangential displacement discontinuity (fault slip) is mathematically equivalent to the force system of a double-couple without moment (MARUYAMA, 1963; BURRIGE and KNOPOFF, 1964), which has no net force and no net torque. Such a property must be satisfied for any force system acting on plate interfaces, since it originates from dynamic processes within the Earth. Thus, given relative plate motion vectors, we can quantitatively describe the crustal deformation processes caused by plate subduction on the basis of elastic/viscoelastic dislocation theory.

2.2. Long-term Crustal Deformation due to Steady Plate Subduction

We model the lithosphere-asthenosphere system by a two-layered Maxwellian viscoelastic half-space. The constitutive equations of the surface layer and the underlying half-space are given by

$$\dot{\sigma}_{ij} + \frac{\mu^{(l)}}{\eta^{(l)}}\left(\sigma_{ij} - \frac{1}{3}\sigma_{kk}\delta_{ij}\right) = \lambda^{(l)}\dot{\varepsilon}_{kk}\delta_{ij} + 2\mu^{(l)}\dot{\varepsilon}_{ij}. \tag{1}$$

Here, σ_{ij}, ε_{ij}, and δ_{ij} are the stress tensor, the strain tensor, and the unit diagonal tensor, respectively, $\lambda^{(l)}$ and $\mu^{(l)}$ ($l = 1, 2$) are the Lamé elastic constants, and $\eta^{(l)}$ ($l = 1, 2$) denotes the viscosity of each medium. The dot indicates differentiation with respect to time. It should be noted that the viscosity of the lithosphere (the surface layer) is considerably larger than that of the asthenosphere (the underlying half-space). The standard values of structural parameters used for computation are listed in Table 1.

We introduce a 3-D curved interface Σ that divides the surface layer into two plates. In general, according to BACKUS and MULCAHY (1976), the surface moment tensor density corresponding to a tangential displacement discontinuity (fault slip) vector $\boldsymbol{u}^{(s)}$ at a point $\boldsymbol{\xi}$ on an interface Σ can be represented as

$$m_{pq}(\boldsymbol{\xi}) = \mu\, u^{(s)}(\boldsymbol{\xi})\, \left[\chi_p(\boldsymbol{\xi})n_q(\boldsymbol{\xi}) + \chi_q(\boldsymbol{\xi})n_p(\boldsymbol{\xi})\right], \tag{2}$$

where μ is the rigidity of the medium, $\boldsymbol{n}$ is the unit normal vector of Σ, $u^{(s)}$ is the magnitude of $\boldsymbol{u}^{(s)}$, and $\boldsymbol{\chi}$ is the unit slip-direction vector satisfying the orthogonality condition $\boldsymbol{n} \cdot \boldsymbol{\chi} = 0$. Then, the i-component of viscoelastic displacement (slip-response

Table 1

Structural parameters used in numerical simulations

	h [km]	ρ [kg/m^3]	λ [GPa]	μ [GPa]	η [Pa s]
Lithosphere	40	3000	40	40	5×10^{23}
Asthenosphere	∞	3400	90	60	5×10^{18}

h: thickness, ρ: density, λ and μ: Lamé elastic constants, η: viscosity.

function) $U_i (i = x, y, z)$ at a point $\boldsymbol{x}$ and time t due to a unit step slip at a point $\boldsymbol{\xi}$ and time τ is given by

$$U_i(\boldsymbol{x}, t; \boldsymbol{\xi}, \tau) = \mu F_{ipq}(\boldsymbol{x}, t; \boldsymbol{\xi}, \tau) [\chi_p(\boldsymbol{\xi}) n_q(\boldsymbol{\xi}) + \chi_q(\boldsymbol{\xi}) n_p(\boldsymbol{\xi})]. \tag{3}$$

Here, the function $F_{ipq}(\boldsymbol{x}, t; \boldsymbol{\xi}, \tau)$ denotes the i-component of viscoelastic displacement due to the pq-element of the moment tensor density with the magnitude $u^{(s)} = 1$. Then, we can compute the displacement field $u_i (i = x, y, z)$ due to arbitrary fault slip distribution $w(\boldsymbol{\xi}, \tau)$ on the plate interface Σ by using the technique of hereditary integral as

$$u_i(\boldsymbol{x}, t) = \int_{-\infty}^{t} \int_{\Sigma} U_i(\boldsymbol{x}, t - \tau; \boldsymbol{\xi}, 0) \frac{\partial w(\boldsymbol{\xi}, \tau)}{\partial \tau} d\boldsymbol{\xi} d\tau. \tag{4}$$

In the present problem we focus on vertical surface displacements u_z. For example, MATSU'URA and SATO (1989) and SATO and MATSU'URA (1993) give the explicit expressions of the viscoelastic slip-response function U_z for a viscoelastic multi-layered half-space under gravity.

Now we decompose the total fault slip motion w into the steady slip at a plate convergence rate v_{pl} and its perturbation Δw (forward slip due to interplate earthquakes and slip deficit due to interplate coupling) as

$$w(\boldsymbol{\xi}, \tau) = v_{pl}(\boldsymbol{\xi})\, \tau + \Delta w(\boldsymbol{\xi}, \tau). \tag{5}$$

Substituting this expression into Eq. (4), we obtain

$$u_z(\boldsymbol{x}, t) = \int_{t_0}^{t} \int_{\Sigma} U_z(\boldsymbol{x}, t - \tau; \boldsymbol{\xi}, 0) v_{pl}(\boldsymbol{\xi})\, d\boldsymbol{\xi} d\tau + \int_{t_0}^{t} \int_{\Sigma} U_z(\boldsymbol{x}, t - \tau; \boldsymbol{\xi}, 0) \frac{\partial \Delta w(\boldsymbol{\xi}, \tau)}{\partial \tau}\, d\boldsymbol{\xi} d\tau. \tag{6}$$

Here, we supposed that the fault slip motion (plate subduction) started at a time $t = t_0$. The first and the second terms on the right-hand side of Eq. (6) indicate the contributions from the steady slip motion and its perturbation, respectively. In the present problem where we are interested in long-term crustal deformation, we can neglect the second term, because it does not bring about secular changes. Thus the crustal uplift rate at a point $\boldsymbol{x}$ due to steady plate subduction after $t = t_0$ is given by

$$\dot{u}_z(\boldsymbol{x}, t) = V_\Sigma(\boldsymbol{x}, t - t_0) \equiv \int_\Sigma U_z(\boldsymbol{x}, t - t_0; \boldsymbol{\xi}, 0)\, v_{pl}(\boldsymbol{\xi})\, d\boldsymbol{\xi}, \tag{7}$$

where the dot indicates differentiation with respect to time t, and $V_\Sigma(\boldsymbol{x},t)$ is the abbreviated representation of the integral of the slip-response function $U_z(\boldsymbol{x}, t; \boldsymbol{\xi}, 0)$ over the plate interface Σ.

As shown in Table 1, the lithosphere-asthenosphere system consists of a high-viscosity surface-layer and a low-viscosity underlying half-space. Therefore, the viscoelastic vertical crustal deformation u_z has three different phases; instantaneous deformation of the elastic layered half-space, intermediate-term deformation of the elastic surface layer floating on the completely relaxed asthenosphere, and long-term deformation of the viscoelastic layered half-space after the completion of stress relaxation both in the asthenosphere and the lithosphere. Assuming the viscosity of the asthenosphere and the lithosphere to be $10^{18} - 10^{19}$ Pa s and $10^{23} - 10^{24}$ Pa s, respectively, the effective stress-relaxation times of the asthenosphere T_A and the lithosphere T_L become $10 - 10^2$ yr and $10^6 - 10^7$ yr, respectively (SATO and MATSU'URA, 1993). Then, we consider two representative cases; the case of young subduction zones, where the duration $t - t_0$ of steady plate subduction is much longer than T_A but much shorter than T_L, and the case of old subduction zones, in which the duration of steady plate subduction is much longer than T_L.

2.2.1. Young subduction zones. In the first case, $T_A << t - t_0 << T_L$, the asthenosphere has completely relaxed, although the lithosphere still behaves like an elastic plate. Then the slip response function U_z can be regarded as constant in time on this time scale. Denoting the slip response function on this time scale by $U_z^{(E)}(\boldsymbol{x}; \boldsymbol{\xi})$, we can rewrite Eq. (7) as

$$\dot{u}_z(\boldsymbol{x}, t) = V_\Sigma^{(E)}(\boldsymbol{x}) \equiv \int_\Sigma U_z^{(E)}(\boldsymbol{x};\boldsymbol{\xi}) v_{pl}(\boldsymbol{\xi}) d\boldsymbol{\xi} \quad \text{for} \quad T_A << t - t_0 << T_L. \tag{8}$$

The above equation means that the amount of crustal uplift increases with time at constant rates in young subduction zones.

2.2.2. Old subduction zones. In the second case, $t - t_0 >> T_L$, not only the asthenosphere but also the lithosphere has completely relaxed. In this case, since the slip response function U_z becomes zero, we can rewrite Eq. (7) as

$$\dot{u}_z(\boldsymbol{x}, t) = 0 \quad \text{for} \quad t - t_0 >> T_L. \tag{9}$$

The above equation means that the amount of crustal uplift no longer increases in old subduction zones.

2.3. *Effects of Tectonic Erosion*

Now, we consider the effect of tectonic erosion in plate subduction zones. The physicochemical process of tectonic erosion may be very complex. However, from a mechanical point of view, we can regard this process as a mass transfer process from the underside of the overriding plate to the depths. Such a mass transfer process inevitably accompanies a landward retreat of the interface between the descending and the overriding plates. The progress of tectonic erosion would also produce changes in plate interface geometry. In the present study, however, we simply assume that the plate interface migrates landward at a constant rate v_e in the direction of plate convergence $(\cos\theta,\ \sin\theta,\ 0)$ without change in its geometry. Then, assuming that tectonic erosion started at $t = t_i$ in a part Σ' of the plate interface Σ, we can incorporate the effect of tectonic erosion into the first term on the right-hand side of Eq. (6) (the second term has been neglected) by replacing the argument $\boldsymbol{\xi}$ with $\boldsymbol{\xi} + \boldsymbol{v_e}\,(t - t_i)$ for the migrating interface Σ' as

$$\begin{aligned} u_z(\boldsymbol{x}, t) = &\int_{t_0}^{t} \int_{\Sigma-\Sigma'} U_z(\boldsymbol{x}, t-\tau; \boldsymbol{\xi}, 0) v_{pl}(\boldsymbol{\xi})\, d\boldsymbol{\xi} d\tau \\ &+ \int_{t_0}^{t_i} \int_{\Sigma'} U_z(\boldsymbol{x}, t-\tau; \boldsymbol{\xi}, 0) v_{pl}(\boldsymbol{\xi})\, d\boldsymbol{\xi} d\tau \\ &+ \int_{t_i}^{t} \int_{\Sigma'} U_z(\boldsymbol{x}, t-\tau; \boldsymbol{\xi} + \boldsymbol{v_e}\,(t-t_i), 0) v_{pl}(\boldsymbol{\xi} + \boldsymbol{v_e}\,(t-t_i))\, d\boldsymbol{\xi} d\tau \end{aligned} \tag{10}$$

Here, $\boldsymbol{v_e} = (v_e \cos\theta,\ v_e \sin\theta,\ 0)^T$ is the landward-retreat rate vector of the plate interface Σ', and the superscript T denotes the transpose of the corresponding vector.

Using the approximation of $v_{pl}(\boldsymbol{\xi} + \boldsymbol{v_e}\,(t - t_i)) \cong v_{pl}(\boldsymbol{\xi})$ and a symmetric property of $U_z(\boldsymbol{x}, t-\tau; \boldsymbol{\xi} + \boldsymbol{v_e}\,(t-t_i), 0) = U_z(\boldsymbol{x} - \boldsymbol{v_e}\,(t-t_i), t-\tau; \boldsymbol{\xi}, 0)$, we can rewrite Eq. (10) as

$$u_z(\boldsymbol{x}, t) = \int_{t_0}^{t} V_{\Sigma-\Sigma'}(\boldsymbol{x}, t-\tau)\, d\tau + \int_{t_0}^{t_i} V_{\Sigma'}(\boldsymbol{x}, t-\tau) d\tau + \int_{t_i}^{t} V_{\Sigma'}(\boldsymbol{x} - \boldsymbol{v_e}(t-t_i), t-\tau)\, d\tau \tag{11}$$

with

$$V_{\Sigma}(\boldsymbol{x}, t-\tau) = \int_{\Sigma} U_z(\boldsymbol{x}, t-\tau; \boldsymbol{\xi}, 0) v_{pl}(\boldsymbol{\xi})\, d\boldsymbol{\xi}. \tag{12}$$

Differentiating both sides of Eq. (11) with respect to time t, we can obtain the expressions of crustal uplift rates due to steady plate subduction with tectonic erosion as follows. The differentiation of the first, second, and third terms on the right hand side of Eq. (11) with respect to time t yields, respectively,

$$\frac{d}{dt}\int_{t_0}^{t} V_{\Sigma-\Sigma'}(\boldsymbol{x}, t-\tau)d\tau = V_{\Sigma}(\boldsymbol{x}, t-t_0) - V_{\Sigma'}(\boldsymbol{x}, t-t_0), \tag{13}$$

$$\frac{d}{dt}\int_{t_0}^{t_i} V_{\Sigma'}(\boldsymbol{x}, t-\tau)d\tau = -V_{\Sigma'}(\boldsymbol{x}, t-t_i) + V_{\Sigma'}(\boldsymbol{x}, t-t_0), \tag{14}$$

and

$$\begin{aligned}\frac{d}{dt}\int_{t_i}^{t} V_{\Sigma'}(\boldsymbol{x}-\boldsymbol{v_e}\,(t-t_i), t-\tau) &= V_{\Sigma'}(\boldsymbol{x}, t-t_i)\\ &- v_e\cos\theta\int_{t_i}^{t} V_{\Sigma',x}(\boldsymbol{x}-\boldsymbol{v_e}\,(t-t_i), t-\tau)\,d\tau\\ &- v_e\sin\theta\int_{t_i}^{t} V_{\Sigma',y}(\boldsymbol{x}-\boldsymbol{v_e}\,(t-t_i), t-\tau)\,d\tau\end{aligned} \tag{15}$$

with

$$V_{,x(y)}(\boldsymbol{x}, t-\tau) = \frac{\partial}{\partial x(y)}V(\boldsymbol{x}, t-\tau). \tag{16}$$

Thus the crustal uplift rates due to steady plate subduction with tectonic erosion are given by

$$\dot{u}_z(\boldsymbol{x}, t) = V_{\Sigma}(\boldsymbol{x}, t-t_0) - \int_{t_i}^{t} \boldsymbol{v}_e\cdot\nabla V_{\Sigma'}(\boldsymbol{x}-\boldsymbol{v_e}(t-t_i), t-\tau)\,d\tau. \tag{17}$$

Here, the first term on the right-hand side of the above equation represents the effects of steady slip motion over the whole plate interface Σ, which is the same as that in Eq. (7), and the second term represents the effects of the landward retreat of the plate interface Σ' due to tectonic erosion. When the migration distance of the plate interface Σ' during $t - t_i$ is much smaller than the horizontal distance between the migrating plate interface and observation points, that is, $|\boldsymbol{v}_e\,(t-t_i)| << |\boldsymbol{x}-\boldsymbol{\xi}|$, we can obtain the following approximate expression of Eq. (17):

$$\dot{u}_z(\boldsymbol{x}, t) \cong V_{\Sigma}(\boldsymbol{x}, t-t_0) - \int_{t_i}^{t} \boldsymbol{v}_e\cdot\nabla V_{\Sigma'}(\boldsymbol{x}, t-\tau)\,d\tau \quad \text{for} \quad |\boldsymbol{v}_e\,(t-t_i)| << |\boldsymbol{x}-\boldsymbol{\xi}|. \tag{18}$$

Now we consider three representative cases: 1) Young subduction zones, in which the duration of both tectonic erosion $t - t_i$ and steady plate subduction $t - t_0$ are

much longer than the effective stress-relaxation time of the asthenosphere T_A but much shorter than the effective stress-relaxation time of the lithosphere T_L, 2) moderate-aged subduction zones, in which the duration of tectonic erosion is by much longer than T_A but much shorter than T_L, and that of steady plate subduction is much longer than T_L, and 3) old subduction zones, in which the duration of both tectonic erosion and steady plate subduction are much longer than T_L.

2.3.1. Young subduction zones. In the first case, $T_A << t - t_i << T_L$ and $T_A << t - t_0 << T_L$, the integral of the slip-response function over the plate interface $V_\Sigma(\boldsymbol{x},t - t_0)$ and its spatial derivative $\nabla V_{\Sigma'}(\boldsymbol{x}, t - \tau)$ in of Eq. (18) can be regarded as constant in time. Denoting the slip response function on this time scale by $U_z^{(E)}(\boldsymbol{x};\boldsymbol{\xi})$, and its surface integral over Σ by $V_\Sigma^{(E)}(\boldsymbol{x}) \equiv \int_\Sigma U_z^{(E)}(\boldsymbol{x};\boldsymbol{\xi}) v_{pl}(\boldsymbol{\xi}) d\boldsymbol{\xi}$, we can obtain the expression of crustal uplift rates in the case of young subduction zones ($T_A << t - t_i << T_L$ and $T_A << t - t_0 << T_L$) as

$$\dot{u}_z(\boldsymbol{x}, t) = V_\Sigma^{(E)}(\boldsymbol{x}) - \boldsymbol{v}_e \cdot \nabla V_{\Sigma'}^{(E)}(\boldsymbol{x})\ (t - t_i). \tag{19}$$

The above equation means that both effects of tectonic erosion on Σ' and steady plate subduction over Σ contribute to crustal uplift and subsidence in young subduction zones.

2.3.2. Moderate-aged subduction zones. In the second case, $T_A << t - t_i << T_L$ and $t - t_0 >> T_L$, the integral of the slip-response function over the plate interface $V_\Sigma(\boldsymbol{x},t - t_0)$ becomes zero, however its spatial derivative $\nabla V_{\Sigma'}(\boldsymbol{x}, t - \tau)$ takes the same constant value as in the first case. Then, we can obtain the expression of crustal uplift rates in the case of moderate-aged subduction zones ($T_A << t - t_i << T_L$ and $t - t_0 >> T_L$) as

$$\dot{u}_z(\boldsymbol{x}, t) = -\boldsymbol{v}_e \cdot \nabla V_{\Sigma'}^{(E)}(\boldsymbol{x})\ (t - t_i)\,. \tag{20}$$

The above equation means that only the effect of tectonic erosion contributes to crustal uplift and subsidence in moderate-aged subduction zones.

2.3.3. Old subduction zones. In the third case, $t - t_i >> T_L$ and $t - t_0 >> T_L$, both the integral of the slip-response function over the plate interface $V_\Sigma(\boldsymbol{x},t - t_0)$ and its spatial derivative $\nabla V_{\Sigma'}(\boldsymbol{x}, t - \tau)$ become zero. Then, in the case of old subduction zones ($t - t_i >> T_L$ and $t - t_0 >> T_L$), we obtain

$$\dot{u}_z(\boldsymbol{x}, t) = 0. \tag{21}$$

The above equation means that even tectonic erosion no longer brings about to crustal uplift and subsidence in old subduction zones.

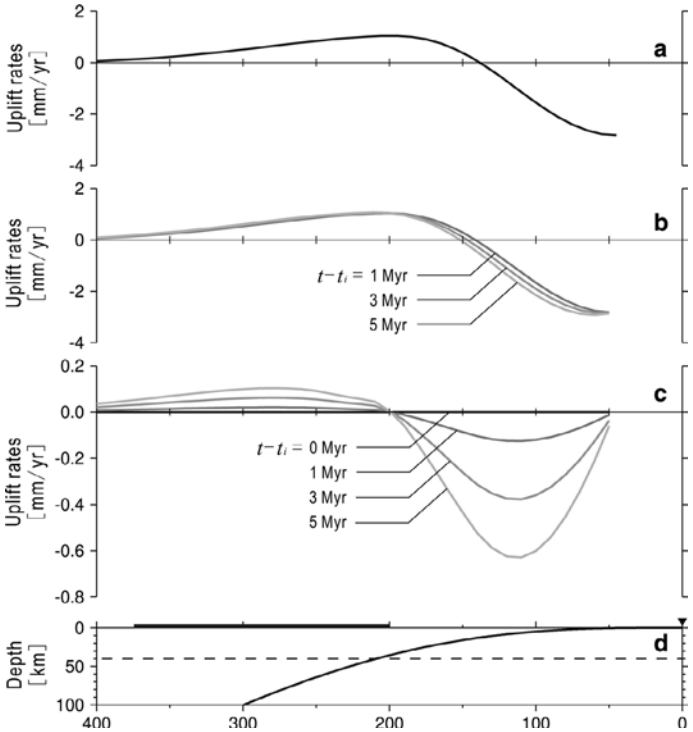


Figure 1

Computed vertical crustal motion due to steady plate subduction with tectonic erosion. (a) The surface uplift rates due to steady plate subduction. (b) The temporal change in surface uplift rates in young subduction zones with tectonic erosion. The uplift rates are given by the sum of the surface uplift rates due to steady plate subduction (a) and tectonic erosion (c). (c) The surface uplift rates due to tectonic erosion after the duration $t - t_i = 0$, 1, 3, and 5 Myr. The landward retreat rate of the plate interface is taken to be 3 mm/yr. (d) The trench-across vertical section of the plate interface. The thick horizontal line indicates the range of the land area in northeast Japan. The solid triangle indicates the position of the trench axis. The dotted line indicates the lithosphere-asthenosphere boundary.

3. Vertical Crustal Motion Caused by Tectonic Erosion

The vertical crustal motion due to steady plate subduction with tectonic erosion can be expressed in the form of Eq. (17). In this section, in order to quantitatively examine the effect of tectonic erosion, we carry out numerical simulation with simplified plate interface geometry.

We consider an infinitely long plate interface Σ along the trench axis. The trench-across vertical section of the plate interface is shown in Figure 1d. We take the rate of fault slip along the plate interface to be 8 cm/yr with its direction perpendicular to the trench axis. Furthermore, we suppose that the whole plate interface migrates landward at a constant rate

of 3 mm/yr. Consequently, we can treat the problem in a 2-D framework. Taking the y axis so as to coincide with the trench axis, the x axis on the surface to be perpendicular to the trench axis, and the z axis to be vertical upward, we can rewrite Eq. (17) as

$$\dot{u}_z(\boldsymbol{x}, t) = V_\Sigma(\boldsymbol{x}, t - t_0) - v_e \int_{t_i}^{t} \frac{\partial}{\partial x} V_\Sigma(\boldsymbol{x} - \boldsymbol{v_e}(t - t_i), t - \tau)\, d\tau. \tag{22}$$

Here, the first and second terms on the right-hand side of the above equation represent the effects of steady plate subduction and landward retreat of the plate interface due to tectonic erosion, respectively. On a time-scale where the lithosphere can be regarded as an elastic surface layer floating on the completely relaxed asthenosphere, Eq. (22) can be rewritten in the form of Eq. (19) as

$$\dot{u}_z(\boldsymbol{x}, t) = V_\Sigma^{(E)}(\boldsymbol{x}) - v_e(t - t_i) \frac{\partial}{\partial x} V_\Sigma^{(E)}(\boldsymbol{x}). \tag{23}$$

From the above equation we can see that the crustal uplift rate due to tectonic erosion (the second term) is given by the product of the migration distance $v_e\,(t - t_i)$ and the negative gradient of the crustal uplift rate due to steady plate subduction $\partial V_\Sigma^{(E)}(\boldsymbol{x})/\partial x$.

In Figure 1a we show the computed crustal uplift rates due to steady plate subduction, characterized by island-arc high and trench low. In Figure 1c we show the computed crustal uplift rates due to tectonic erosion ($v_e = 3$ mm/yr) for the duration $t - t_i = 0$, 1, 3, and 5 Myr. From this figure we can see that tectonic erosion brings about the steep subsidence of the wedge portion of the overriding plate and the gentle uplift of the landward portion. The zero crossing point of the uplift rate curves agrees with the point of the maximum crustal uplift rate due to steady plate subduction. Figure 1b shows the temporal change of crustal uplift rates in young subduction zones with tectonic erosion ($T_A << t - t_i << T_L$ and $T_A << t - t_0 << T_L$), which is given by the sum of the crustal uplift rates due to steady plate subduction (Fig. 1a) and tectonic erosion (Fig. 1c). Here, it should be noted that Eq. (23) is not applicable for the points near the migrating plate interface, because the approximate expression in Eq. (18) is not valid there. From Figure 1b we can see that the effect of tectonic erosion is less significant in young subduction zones. However, in old subduction zones ($T_A << t - t_i << T_L$ and $t - t_0 >> T_L$), the effect of tectonic erosion becomes dominant, because the crustal uplift rates due to steady plate subduction (the first term of Eq. (22)) tend to zero. In other words, the initiation of tectonic erosion in old subduction zones causes significant change in the mode of vertical crustal motion.

4. Vertical Crustal Motion in Northeast Japan

4.1. 3-D Plate Interface Geometry in and Around Japan

Recently, HASHIMOTO *et al.* (2004) constructed a realistic 3-D model of plate interface geometry in and around Japan on the basis of the topography of ocean floors

and the hypocenter distribution of earthquakes (Fig. 2). The modelled region extends from 125°E to 155°E in longitude and from 20°N to 50°N in latitude (about 3000 km $\times$ 3000 km). In this region the lithosphere is divided into four plates; the Pacific (PA), North American (NA), Philippine Sea (PH), and Eurasian (EU) plates. Since the interface between the North American plate and the Eurasian plate is indefinite, we discard it in modelling. The 3-D geometry of the upper boundaries of the Pacific and Philippine Sea plates is represented by the superposition of bi-cubic B-splines with 8 km equally spaced local supports. The expansion coefficients of plate interface geometry are determined from ISC (International Seismological Centre) hypocenter location data by using an inversion technique developed by YABUKI and MATSU'URA (1992).

In order to determine the direction of steady slip vectors $\boldsymbol{v}_{pl}$ on the four given plate interfaces, Σ_1 (PA-NA), Σ_2 (PA-PH), Σ_3 (PH-NA) and Σ_4 (PH-EU), first, we calculate the relative plate motion vector $\boldsymbol{v} = (v_x, v_y, 0)^T$ at each plate boundary from NUVEL-1A (DEMETS *et al*., 1994) as shown in Figure 3. Then, we project the calculated relative plate motion vector $\boldsymbol{v}$ onto the plate interface. The depth to the plate interface is defined by the

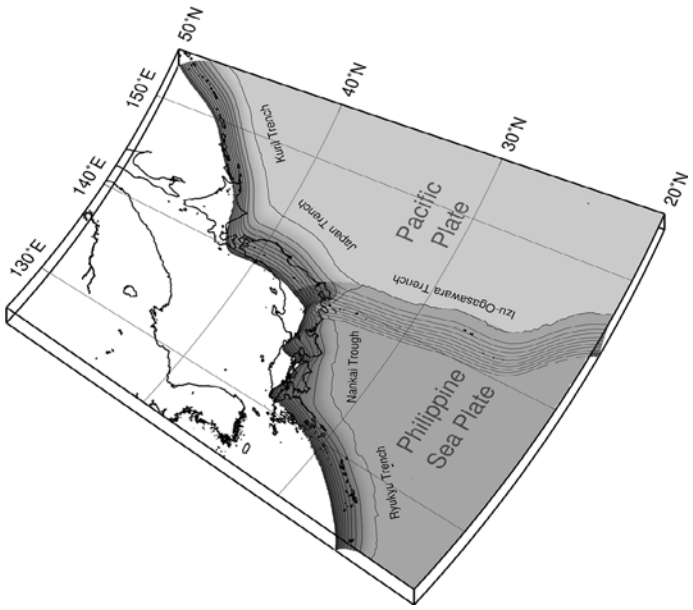


Figure 2
A 3-D model of plate interface geometry in and around Japan, used in numerical simulations. The upper boundaries of the descending Pacific and the Philippine Sea plates are shown. The intervals of the isodepth contours are taken to be 10 km. The model region extends from 125°E to 155°E in longitude and from 20°N to 50°N in latitude.

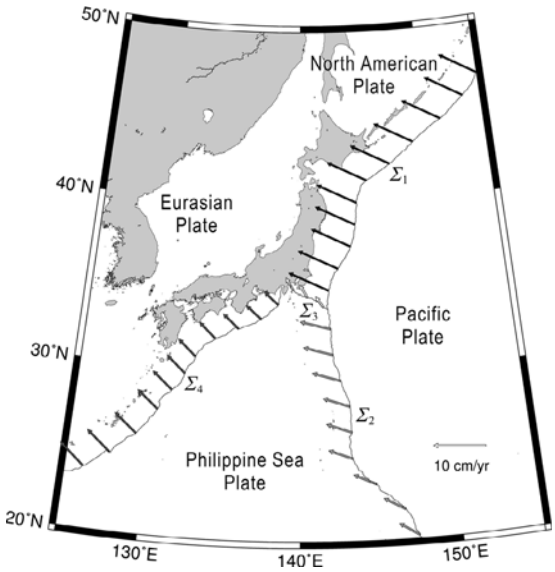


Figure 3
Plate boundaries and relative plate motion vectors in and around Japan. The elastic lithosphere is divided into four plates; the Pacific (PA), the North American (NA), the Philippine Sea (PH), and the Eurasian (EU) plates. These four plates are interacting with each other at four plate interfaces; Σ_1 (PA-NA), Σ_2 (PA-PH), Σ_3 (PH-NA) and Σ_4 (PH-EU). The relative plate motion vectors (thick arrows) are calculated from the global plate motion model NUVEL-1A (DeMets *et al.*, 1994).

continuous function $z = z(x, y)$, and therefore the unit normal vector $\boldsymbol{n}$ of the plate interface is given by

$$\boldsymbol{n} = \begin{pmatrix} n_x \\ n_y \\ n_z \end{pmatrix} = \frac{1}{\sqrt{(\partial z/\partial x)^2 + (\partial z/\partial y)^2 + 1}} \begin{pmatrix} -\partial z/\partial x \\ -\partial z/\partial y \\ 1 \end{pmatrix}. \tag{24}$$

The magnitude v_{pl} of the steady slip vector at the plate interface is taken to be the same as the magnitude of the relative plate motion vector $\boldsymbol{v}$ to satisfy the mass conservation of the descending slab. Then, the steady slip vector $\boldsymbol{v}_{pl}$ is given by

$$\boldsymbol{v}_{pl} = |\boldsymbol{v}|\,\boldsymbol{\chi} = \frac{\sqrt{v_x^2 + v_y^2}}{\sqrt{v_x^2 + v_y^2 + \left(v_x n_x + v_y n_y\right)^2 / n_z^2}} \begin{pmatrix} -v_x \\ -v_y \\ \left(v_x n_x + v_y n_y\right)/n_z \end{pmatrix}. \tag{25}$$

Here, $\boldsymbol{\chi}$ represents the unit slip-direction vector, which is parallel to the projection of the relative plate motion vector $\boldsymbol{v}$ onto the plate interface.

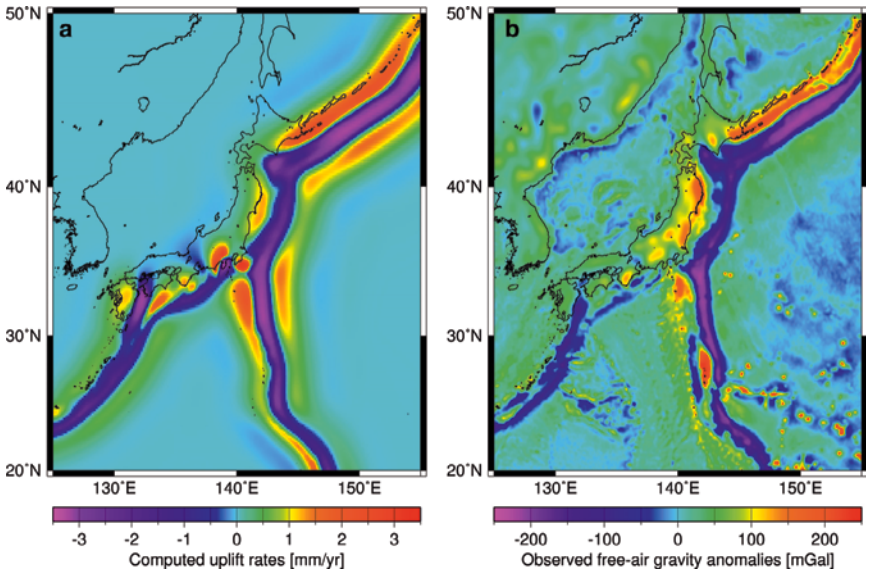


Figure 4

Computed surface uplift rates and observed free-air gravity anomalies in and around Japan. (a) The surface uplift rates due to simple steady subduction of the Pacific and the Philippine Sea plates. The maximum subsidence rates at the Japan trench and the Nankai through are about 2.5 mm/yr and 1.5 mm/yr, respectively. The maximum uplift rate is approximately 1.5 mm/yr, both in northeast Japan and southwest Japan. (b) Free-air gravity anomalies obtained by satellite altimetry (SANDWELL and SMITH, 1997). The color bar shows the amplitude of free-air gravity anomaly.

4.2. Vertical Crustal Motion in Northeast Japan

First, we compute long-term crustal motion in northeast Japan due to steady plate subduction without tectonic erosion. Carrying out the integration in Eq. (8) over all the plate interfaces $\Sigma = \Sigma_1 + \Sigma_2 + \Sigma_3 + \Sigma_4$, we obtain the surface uplift rates due to steady plate subduction. In Figure 4 we show the computed surface uplift rates together with the free-air gravity anomalies obtained by satellite altimetry (SANDWELL and SMITH, 1997). From comparison of the patterns of computed surface uplift rates (Fig. 4a) and observed free-air gravity anomalies (Fig. 4b), we can see that the 3-D model of simple steady plate subduction adequately reproduces not only the general pattern of free-air gravity anomaly (island-arc high, trench low and outer-rise gentle high) but also the regional patterns of negative anomaly at the Kuril-Japan trench junction in northeast Japan and the Nankai trough-Ryukyu trench junction in southwest Japan. This means that the essential cause of free-air gravity anomaly in and around Japan is in long-term crustal uplift and subsidence due to simple steady plate subduction. As mentioned in Subsection 2.2, Eq. (8) is an approximate expression in the case of young subduction zones; that is,

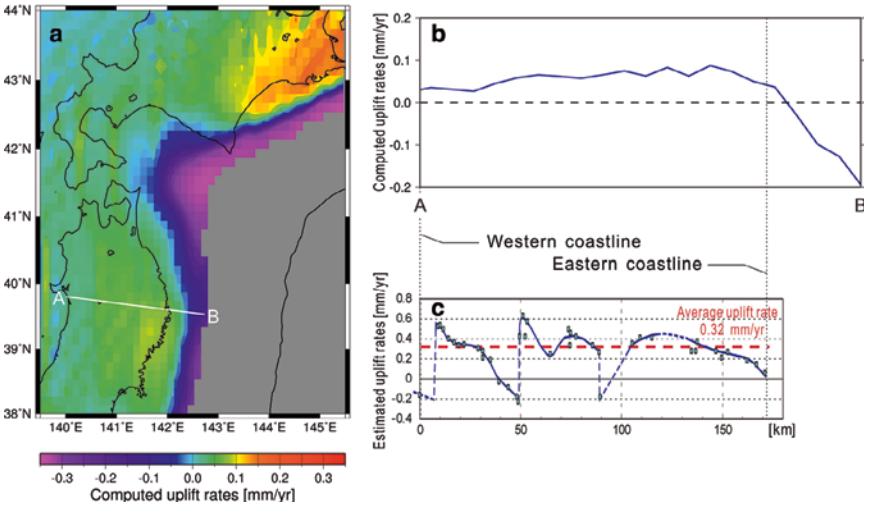


Figure 5

Computed and observed surface uplift rates in northeast Japan. (a) The pattern of computed surface uplift rates due to steady plate subduction with tectonic erosion. The landward-retreat rate and the duration of tectonic erosion are taken to be 3 mm/yr and 3.5 Myr, respectively. The color bar shows the amplitude of the computed surface uplift rates. (b) The profile of the computed surface uplift rates along the line A-B in Figure 5a. (c) The average crustal uplift rates during the last 0.12 million years, estimated from fluvial-terrace heights by TAJIKARA (2004). The average crustal uplift rate over the northeast Japan arc is about 0.3 mm/yr. The short-wavelength (10–50 km) deformation in the western part of the northeast Japan arc is ascribed to the thrust motion of active faults. The crustal uplift rate in the eastern part of the northeast Japan arc gradually decreases eastward to zero near the eastern coastline.

the duration of steady plate subduction is much shorter than the effective stress-relaxation time of the lithosphere. If the duration of steady plate subduction becomes longer, the surface uplift rates gradually decrease as a whole with time and finally tend to zero because of stress relaxation in the lithosphere (SATO and MATSU'URA, 1993). From the ages of islands and seamounts at the northwestern end of the Hawaiian Chain, the duration of the steady subduction of the Pacific plate beneath northeast Japan in the present direction is simply estimated about 40 Myr, which is much longer than the effective stress-relaxation time of the lithosphere. Thus, we can explain the stable pattern of free-air gravity anomaly by the simple steady plate subduction model, but not the pattern of current surface uplift rates observed in northeast Japan.

In northeast Japan, as shown in Figure 4b, the island-arc high of observed free-air gravity anomaly takes its maximum about the eastern coastline, while the current crustal motion estimated from geological and geomorphological observations shows gentle uplift in the land area and steep subsidence in the sea area with the neutral point near the eastern coastline. In order to explain this discrepancy, we compute long-term crustal uplift rates due to steady plate subduction with tectonic erosion. Now we consider the case of old

subduction zones ($T_A << t - t_i << T_L$ and $t - t_0 >> T_L$), and assume the landward migration of the Pacific-North American plate interface ($\Sigma' = \Sigma_1$) at a constant rate in the direction of relative plate motion (N63°W). From Eq. (20), carrying out the integration of the second term on the right-hand side in Eq. (18) over the migrating plate interface $\Sigma' = \Sigma_1$, we can obtain the surface uplift rates in northeast Japan due to steady plate subduction with tectonic erosion. The landward retreat rate v_e is taken to be 3 mm/yr on the basis of the geological estimate by VON HUENE and LALLEMAND (1990). According to SATO (1994), the mode of tectonic loading in northeast Japan has changed from extension to compression 3.5 million years ago. Then, we take the duration of tectonic erosion $t - t_i$ to be 3.5 Myr, which is considerably shorter than the effective stress-relaxation time of the lithosphere T_L. In Figure 5a we show the pattern of computed surface uplift rates in northeast Japan, characterized by gradual uplift over the northeast Japan arc and steep subsidence in the sea area with the neutral point near the eastern coastline. In Figures 5b and c we show the profile of the computed surface uplift rates along the line A-B (Fig. 5a) and that of the average crustal uplift rates during the last 0.12 million years estimated from fluvial-terrace heights by TAJIKARA (2004), respectively. From these figures we can see that the 3-D model of steady plate subduction with tectonic erosion adequately reproduces the pattern of current vertical crustal motion in northeast Japan.

The tectonic evolution model is based on the assumption that tectonic erosion at a constant rate has started 3–4 million years ago after the long duration of simple steady plate subduction. From Figures 5b and c we can find that the computed surface uplift rate is somewhat small in comparison with the observed crustal uplift rate. It should be noted that the observed crustal uplift rate includes not only the direct effect of tectonic erosion but also the indirect effect of crustal thickening. TAJIKARA (2004) suggested that his estimates of crustal uplift rates include the effect of crustal thickening of the northeast Japan arc by East-West compression. Actually, in northeast Japan, there exist many thrust faults and folds with their strikes parallel to North-South direction (e.g., RESEARCH GROUP FOR ACTIVE FAULTS OF JAPAN, 1991; SATO, 1994; TAJIKARA, 2004). The short-wavelength (10–50 km) deformation on the west side of the northeast Japan arc in Figure 5c can be ascribed to the thrust motion of such active faults (TAJIKARA, 2004).

5. Discussion and Conclusions

Free-air gravity anomaly in plate subduction zones reflects the cumulative effects of long-term crustal uplift and subsidence due to steady plate subduction. In northeast Japan, however, the island-arc high of free-air gravity anomaly takes its maximum about the eastern coastline, while the current vertical crustal motion shows gentle uplift in the land area and steep subsidence in the sea area with the neutral point near the eastern coastline. Such a discrepancy in spatial patterns between free-air gravity anomaly and current vertical crustal motion can be rationally explained by considering the effect of tectonic

erosion at the PA-NA plate interface, which started 3–4 Myr ago after the long duration of simple steady plate subduction.

We developed a mechanical model of steady plate subduction with tectonic erosion on the basis of elastic/viscoelastic dislocation theory. Through numerical simulations with this model we demonstrated that the pattern of free-air gravity anomaly in northeast Japan could be explained by simple steady slip at the PA-NA plate interface at the early stage of subduction. We also demonstrated that the pattern of current vertical crustal motion could be explained by steady slip with tectonic erosion at the late stage of subduction. From these simulation results we proposed a tectonic evolution model which can consistently explain both patterns of free-air gravity anomaly and current vertical crustal motion in northeast Japan.

The proposed tectonic evolution model is based on the assumption that tectonic erosion at a constant rate started 3–4 Myr ago after the long duration of simple steady plate subduction. At the late stage of subduction the surface uplift rate is proportional to the product of the tectonic erosion rate (the landward-retreat rate v_e of the plate interface) and the duration of tectonic erosion ($t - t_i$). In computation of the surface uplift rates in Figure 5b, which are significantly small in comparison with observed surface uplift rates, we took the landward-retreat rate and the duration of tectonic erosion to be 3 mm/yr and 3.5 Myr, respectively. As mentioned in subsection 2.4, the discrepancy between computed and observed uplift rates can be ascribed partly to the neglect of crustal-thickening effects and partly to the underestimate of the landward-retreat rate. If we ignore the effects of crustal thickening, we require the landward-retreat rate of about 12 mm/yr to account for the observed surface uplift rates. The surface uplift rates also increase with the duration of tectonic erosion, if it is much shorter than the effective stress-relaxation tine of the lithosphere T_L. In the present case, however, the duration of tectonic erosion (3.5 Myr) is of the same order as T_L (about 5 Myr), and thus we cannot expect the resolution of the discrepancy between computed and observed uplift rates by increasing the duration of tectonic erosion.

Acknowledgments

Computation of viscoelastic slip-response functions was performed on the Earth Simulator at the Earth Simulator Center, Japan Agency for Marine-Earth Science and Technology (JAMSTEC).

References

Backus, G. and Mulcahy, M. (1976), *Moment tensors and other phenomenological descriptions of seismic sources. II. Discontinuous displacements*, Geophys. J. R. astr. Soc. *46*, 301–329.

Burridge, R. and Knopoff, L. (1964), *Body force equivalents for seismic dislocations*, Bull. Seismol. Soc. Am. *54*, 1875–1888.

DeMets, C., Gordon, R. G., Argus, D. F., and Stein, S. (1994), *Effect of recent revisions to the geomagnetic reversal time scale on estimates of current plate motions*, Geophys. Res. Lett. *21*, 2191–2194.

Hashimoto, C. and Matsu'ura, M. (2000), *3-D physical modelling of stress accumulation and release processes at transcurrent plate boundaries*, Pure Appl. Geophys. *157*, 2125–2147.

Hashimoto, C. and Matsu'ura, M. (2002), *3-D simulation of earthquake generation cycles and evolution of fault constitutive properties*, Pure Appl. Geophys. *159*, 2175–2199.

Hashimoto, C., Fukui, K., and Matsu'ura, M. (2004), *3-D Modelling of plate interfaces and numerical simulation of long-term crustal deformation in and around Japan*, Pure Appl. Geophys. *161*, 2053–2067.

Heki, K. (2004), *Space geodetic observation of deep basal subduction erosion in northeastern Japan*. Earth Planet. Sci. Lett. *219*, 13–20.

Koike, K. and Machida, H., *Atlas of Quaternary marine terraces in the Japanese islands* (Univ. of Tokyo Press, Tokyo 2001).

Maruyama, T. (1963), *On the force equivalents of dynamical elastic dislocations with reference to the earthquake mechanism*, Bull. Earthq. Res. Inst., Tokyo Univ. *41*, 467–486.

Matsu'ura, M. and Sato, T. (1989), *A dislocation model for the earthquake cycle at convergent plate boundaries*, Geophys. J. Int. *96*, 23–32.

Research Group for Active Faults of Japan, *Active faults in and around Japan, sheet maps and inventories (revis. ed.)*, (Univ. of Tokyo Press, Tokyo 1991).

Sandwell, D. T. and Smith, W. H. F. (1997), *Marine gravity anomaly from Geosat and ERS 1 satellite altimetry*, J. Geophys. Res. *102*, 10039–10054.

Sato, H. (1994), *The relationship between late Cenozoic tectonic events and stress field and basin development in northeast Japan*, J. Geophys. Res. *99*, 22261–22274.

Sato, T. and Matsu'ura, M. (1988), *A kinematic model for deformation of the lithosphere at subduction zones*, J. Geophys. Res. *93*, 6410–6418.

Sato, T. and Matsu'ura, M. (1992), *Cyclic crustal movement, steady uplift of marine terraces, and evolution of the island arc-trench system in southwest Japan*, Geophys. J. Int. *111*, 617–629.

Sato, T. and Matsu'ura, M. (1993), *A kinematic model for evolution of island arc-trench systems*, Geophys. J. Int. *114*, 512–530.

Tajikara, M. (2004), *Vertical crustal movements of the northeast Japan arc in late Quaternary time*, Ph. D. Thesis, University of Tokyo.

Tomoda, Y. and Fujimoto, H. (1982), *Maps of gravity anomalies and bottom topography in Western Pacific*, Bull. Ocean Res. Int. Univ. Tokyo 14.

von Huene, R. and Lallemand, S. (1990), *Tectonic erosion along the Japan and Peru convergent margins*, Geol. Soc. Am. Bull. *102*, 704–720.

von Huene, R. and Scholl, D. W. (1991), *Observations at convergent margins concerning sediment subduction, subduction erosion, and the growth of continental crust*, Rev. Geophys. *29*, 279–316.

von Huene, R., Langseth, M., Nasu, N. and Okuda, H. (1982), *A summary of Cenozoic tectonic history along the IPOD Japan trench transect*, Geol. Soc. Am. Bull. *93*, 829–846.

Yabuki, T. and Matsu'ura, M. (1992), *Geodetic data inversion using a Bayesian information criterion for spatial distribution of fault slip*, Geophys. J. Int. *109*, 363–375.

(Received September 7, 2006, revised April 5, 2007, accepted May 30, 2007)
Published Online First: April 2, 2008

Pure appl. geophys. 165 (2008) 585–607
0033–4553/08/030585–23
DOI 10.1007/s00024-008-0318-8

Pure and Applied Geophysics

Development of Long-period Ground Motions from the Nankai Trough, Japan, Earthquakes: Observations and Computer Simulation of the 1944 Tonankai (Mw 8.1) and the 2004 SE Off-Kii Peninsula (Mw 7.4) Earthquakes

Takashi Furumura,[1] Toshihiko Hayakawa,[1] Misao Nakamura,[2] Kazuki Koketsu,[1] and Toshitaka Baba[3]

Abstract—Strong ground motions recorded in central Tokyo during the 1944 Tonankai Mw8.1 earthquake occurring in the Nankai Trough demonstrate significant developments of very large (>10 cm) and prolonged (>10 min) shaking of long-period (T > 10–12 s) ground motions in the basin of Tokyo located over 400 km from the epicenter. In order to understand the process by which such long-period ground motions developed in central Tokyo and to mitigate possible future disasters arising from large earthquakes in the Nankai Trough, we analyzed waveform data from a dense nation wide strong-motion network (K-NET and KiK-net) deployed across Japan for the recent SE Off-Kii Peninsula (Mw 7.4) earthquake of 5 September 2004 that occurred in the Nankai Trough. The observational data and a corresponding computer simulation for the earthquake clearly demonstrate that such long-period ground motion is primarily developed as the wave propagating along the Nankai Trough due to the amplification and directional guidance of long-period surface waves within a thick sedimentary layer overlaid upon the shallowly descending Philippine Sea Plate below the Japanese Island. Then the significant resonance of the seismic waves within the thick cover of sedimentary rocks of the Kanto Basin developed large and prolonged long-period motions in the center of Tokyo. The simulation results and observed seismograms are in good agreement in terms of the main features of the long-period ground motions. Accordingly, we consider that the simulation model is capable of predicting the long-period ground motions that are expected to occur during future Nankai Trough M 8 earthquakes.

Key words: Numerical simulation, finite-difference method, strong ground motion, long-period ground motion, Tonankai earthquake.

1. Introduction

Large M 8 earthquakes have occurred repeatedly in the Nankai Trough at a recurrence period of about 90–110 years. The 1944 Tonankai (Mw 8.1) earthquake

[1] Earthquake Research Institute, University of Tokyo, 1-1-1 Yayoi, Bunkyo-ku 113-0032, Japan. E-mail: furumura@eri.u-tokyo.ac.jp

[2] Information Service for Disaster Prevention, 230-7 Miroku-machi, Sakura, Chiba 285-003, Japan.

[3] IFREE, Japan Agency for Marine-Earth Science and Technology, 3172-25 Showa-machi, Kanazawa-ku, Yokohama, Kanagwa 236-0001, Japan.

resulted in more than 1250 deaths along the Pacific-facing southern coast of Japan. During this event, high seismic intensities (>5 on the 7-point scale of the Japanese Meteorological Agency (JMA)) were recorded along the southern coast of central Japan, more than 100 km from the source region of the earthquake (USAMI, 1996).

Tokyo, the largest population center in Japan, is located about 400 km from the hypocenter, and recorded a maximum intensity of 3–4 on the JMA scale. No significant damage was reported in Tokyo during the large but distant earthquake; however, ground motions recorded in Tokyo demonstrate the development of very large (>10 cm) and prolonged (>10 min) shaking, dominated by long-period (T0 = 10–13 s) ground-shakings generated by the earthquake (FURUMURA and NAKAMURA, 2006).

Such strong and prolonged shaking of long-period ground motions within sedimentary basins brings to mind the extensive damage and fires in oil-storage tanks during the 2003 Tokachi-oki (Mw 8.0) earthquake (e.g., HATAYAMA *et al.*, 2004; KOKETSU *et al.*, 2005). The 2004 Chuetsu Niigata earthquake (Mw 6.6) significantly effected on tall buildings of approximately 70 floors, with a resonance of long-period ground motions of approximately 7 s recorded in central Tokyo (FURUMURA and HAYAKAWA, 2007). These observations warn of the potential vulnerability of modern large-scale construction to large earthquakes occurring around Tokyo. Thus, it is urgent to investigate the processes by which long-period ground motions within the Kanto Basin originate and evolve. This will aid in the mitigation of future seismic-related disasters predicted to result from earthquakes in the Nankai Trough.

The Mw 7.4 SE Off-Kii Peninsula earthquake occurred in the Nankai Trough, southwest of Kii Peninsula, Japan, at 23:57 JST on 5 September, 2004. Although the maximum intensity recorded in Tokyo was less than 2 on the JMA scale, most of the Tokyo area was subjected to more than 5 minutes of intense (>5 cm) ground-shaking and a relatively long dominant period of about 7–12 s (e.g., HAYAKAWA *et al.*, 2005; MIYAKE and KOKETSU, 2005). These large and prolonged long-period ground motions led to significant resonance to large-scale oil-storage tanks in the area around Tokyo Bay and caused damage to the floating roofs of the tanks (HATAYAMA *et al.*, 2005), highlighting the risk to modern large-scale construction posed by large earthquakes.

Although the strong motions generated during the 1944 Tonankai earthquake were recorded on only a few distant stations such as installed in Tokyo (most of the near-field stations were clipped by the intense ground-shaking), ground motions associated with the 2004 SE Off-Kii Peninsula earthquake were clearly recorded by more than 500 stations of the high-density nation wide strong-motion network of K-NET and KiK-net across Japan. Thus, observations from the 2004 event provide us with a detailed understanding of the nature of long-period ground motions that affected central Tokyo during the earthquake in the Nankai Trough.

To complement the observational data and provide insight into the complexities of seismic behavior which result from interaction between complex 3D subsurface structure along the propagation path, we conducted large-scale computer simulations of seismic waves propagation during the 2004 SE Off-Kii Peninsula earthquake and the 1944 Tonankai earthquake. In conducting these simulations, we used the Earth Simulator

supercomputer with a high-resolution subsurface structure model of central Japan and an appropriate source-slip model for the earthquakes. Snapshots of seismic wave propagation during these events, as well as waveforms derived from the simulation and comparisons with observed seismograms provide for a detailed understanding of the nature of the seismic wavefield and the development of intense and prolonged long-period ground motions within central Tokyo.

2. *Long-period Ground Motions Associated with the 2004 SE Off-Kii Peninsula Earthquake*

Figure 1 shows snapshots of ground motions resulting from the SE Off-Kii Peninsula earthquake, as recorded by the K-NET and KiK-net strong-motion network using a

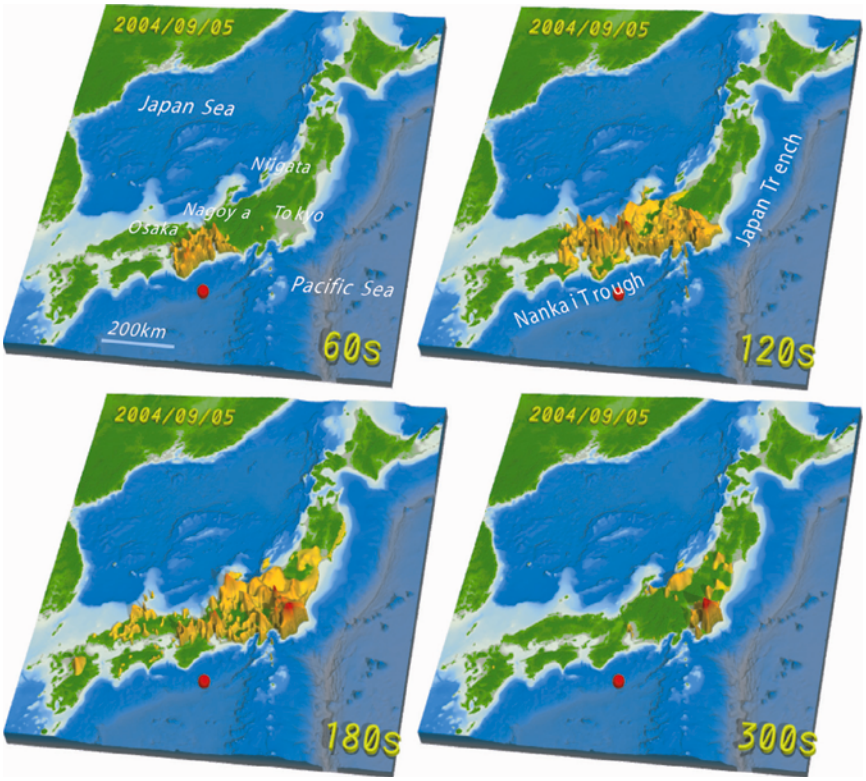


Figure 1
Snapshots of seismic wave propagation following the 2004 SE Off-Kii Peninsula earthquake, Japan. The hypocenter of the event is marked by a red circle. The distribution of horizontal velocity ground motions is shown for times of T = 60, 120, 180, and 300 s following the earthquake rupture.

visualization procedure for the seismic wavefield (FURUMURA *et al.*, 2003). Each snapshot shows the spatial and temporal distribution of ground motions at a given time from the earthquake rupture. The ground accelerations observed at 512 strong-motion stations are integrated into ground velocity motions in order to enhance the contribution of long-period ground motions with period of over 1 s. The amplitude of horizontal ground motion at each point is multiplied by the epicentral distance for roughly compensated geometrical attenuation and then logarithmically compressed in order to display a wide range of ground motions from source region to regional distances.

The first snapshot (T = 60 s) shows intense ground-shaking in the area closest to the source region. A significant amplification of ground motions in sedimentary basins resulted in very strong and prolonged ground motions, which is apparent in large basins such as in Osaka and Nagoya in the 120 s snapshot and in Toyama and Niigata in the 180 s snapshot.

The attenuation of short-period surface waves is evident as the seismic waves spread from the source region, However the amplitude of the ground motion is reinforced significantly when entering the large Kanto Basin (Tokyo region). Ground motion in the Kanto Basin is at least 10-times greater than that measured in surrounding areas. The 300 s snapshot shows that intense and prolonged ground motions occurred in the Kanto Basin for more than several minutes, clearly outlining the basin margins.

Figure 2 shows the distribution of ground motions associated with the earthquake in terms of seismic intensity, peak ground acceleration (PGA), and peak ground displacement (PGD). The attenuation properties of the short-period ground motions (with periods of less than about 1 s) are well illustrated by PGA (Fig. 2b) and seismic intensity (Fig. 2a). Both PGA and seismic intensity show subcircular distributions above the hypocentral area, with pockets of greater amplification in the Nagoya and Osaka Basins resulting from the amplification of short-period ground motions within shallow covering layers in the sedimentary basins.

The attenuation of short-period seismic waves in the Kanto area is significant, and the maximum intensity is less than 2; however, the PGD map shows surprisingly enhanced ground motions in the Kanto Basin, in excess of 5 cm (Fig. 2c). Large PGD is also recorded in other major sedimentary basins such as those in Osaka, Nagoya, Toyama, and Niigata, although the anomalously large PGD of more than 5 cm recorded for Tokyo is very surprising, given the distance from the hypocenter (>400 km) and the intermediate magnitude (Mw 7.4) of the event.

As is evident in the waveforms of ground displacement motion recorded at K-NET stations (OSK005, AIC003, TYM005, YMN005, and TKY025; see Fig. 2c) located upon major basins, the large PGD recorded in major population centers and in Tokyo is caused by large, long-period surface waves that developed within thick sedimentary basins (Fig. 2d).

Figure 3 shows the velocity response spectra of horizontal ground motion recorded in each station and assumed a damping coefficient of $h = 0.05$. As the K-NET and KiK-net recordings rely on an event-trigger system, weak initial P and S waves and later coda with relatively small acceleration levels were not recorded by the strong motion instruments.

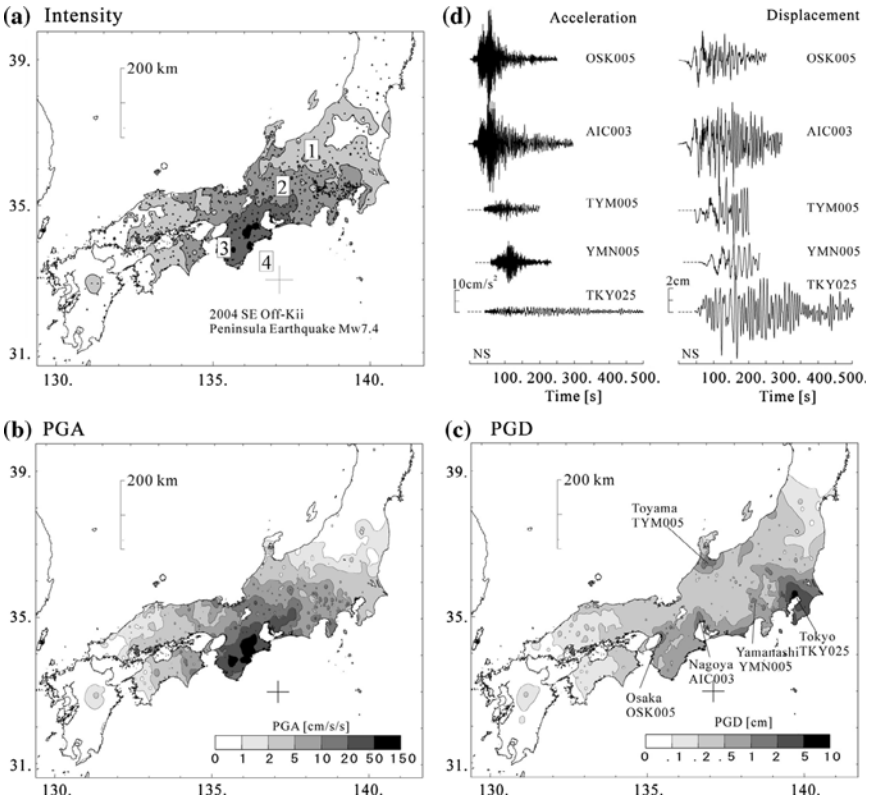


Figure 2

Distribution of (a) seismic intensity in terms of the JMA 7-point scale (small triangles denote K-NET and KiK-net stations that were triggered during the earthquake), (b) peak ground acceleration (cm/s/s), and (c) peak ground displacement (cm) during the 2004 SE Off-Kii Peninsula earthquake. (d) Wave form of NS-component ground acceleration and displacement recorded at five K-NET stations (OSK005, AIC003, TYM005, YMN005, and TKY025).

However, even in the short-duration recordings, significant resonance of the long-period ground motions in sedimentary basins is clearly displayed with a sharp peak in resonance periods of about 2 s for Nagoya (AIC003), about 3 s for Osaka (OSK005), and about 6 s for Toyama (TYM005). The very large response of over 10 cm/s in central Tokyo (TKY024) is almost as large as that observed near the source region.

MIYAKE and KOKETSU (2005) studied the spatial distribution of the dominant period of long-period ground motions generated during the 2004 SE Off-Kii Peninsula earthquake by employing high-density records of intensity meters located at e.g., the city government office and fire stations, as well as using K-NET and KiK-net data. The detailed map arising from their study demonstrates that long-period ground motions with a dominant

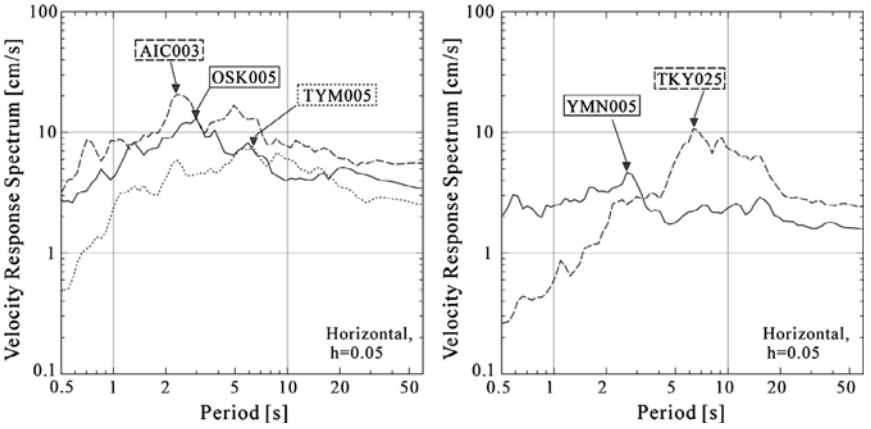


Figure 3
Velocity response spectra of observed ground motions recorded at stations AIC003, OSK005, TYM005, YMN005, and TKY025, assuming a damping coefficient of h = 0.05. The first peak of each resonance period is marked by a triangle.

period of around 7–12 s developed in the Tokyo Bay area during the earthquake. These results are consistent with reports that oil-storage tanks experienced a resonance period of about 6–12 s, resulting in sloshing of the oil and damage to the tanks during the earthquake (HATAYAMA *et al.*, 2004).

To study the progressive development of such long-period ground motions from the source area in the Nankai Trough toward the Kanto Basin more thoroughly, we compiled the observed waveforms of seven linearly aligned K-NET stations from the Kii Peninsula to central Tokyo (Fig. 4, line a–a').

Close to the source region, short-period (<1 s) S waves are the main contributor to horizontal ground motions, but attenuation of the short-period signals becomes severe with increasing distance from the hypocenter. In contrast, long-period ground motions with a dominant period of about 7–10 s that follows the S waves gradually acquire a greater amplitude with propagation along the coast, and these waves are the principle contributor to the observed large and prolonged ground motions, as depicted in the snapshots (Fig. 1) and PGD map (Fig. 2c). The group velocity of the, long-period, later signals is less than about 1 km/s, thus, we can recognize the long-period signals as surface waves which developed near the hypocentral area by the conversion of S waves radiating from the shallow (h = 8 km) earthquake source. As these later signals are clearly depicted in transverse and radial motions, we are able to recognize that the mixing of Rayleigh and Love waves is the major component of the surface wave.

A number of recent reflection and refraction experiments in the Nankai Trough (e.g., TAKAHASHI *et al.*, 2002; NAKANISHI *et al.*, 2002; KODAIRA *et al.*, 2005) have demonstrated a thick cover of oceanic sediments (>5–10 km) above the shallowly subducting Philippine Sea Plate, forming a low-wavespeed accretionary prism above the subduction zone. This

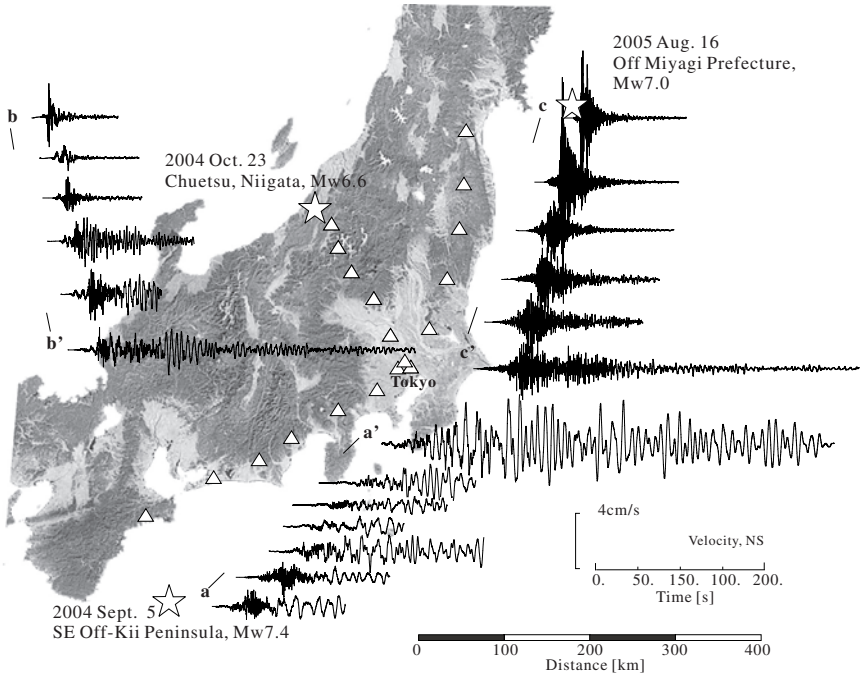


Figure 4

Recorded sections of NS-component ground velocity motions associated with the 2004 SE Off-Kii Peninsula earthquake (a–a'), the 2004 Chuetsu, Niigata earthquake (b–b'), and the 2005 Off Miyagi Prefecture earthquake (c–c'), illustrating the development of long-period surface waves during propagation of the waves from the source region to central Tokyo.

thick cover of sediments upon the subducting plate, and the large velocity contrast across the sediment–plate interface, are considered to contribute strongly to the efficient development of large surface waves and the guiding of seismic waves toward Tokyo. The importance of the low-wavespeed accretionary prism covering the subdicting plate on the developments of long-period surface waves has been noted by SHAPIRO *et al*. (2000) as an important mechanism of the development of prolonged ground motions observed in Mexico City during large subduction zone earthquakes.

As the seismic wave that propagates to the Kanto Basin along the Nankai Trough already has a large and very lengthy wavetrain, and as entering the thick cover of sediments below the Kanto Basin it changes to another type of surface wave, induced by conversion and a scattering of incidental waves at the edge of the basin. Consequently, the resulting surface waves within central Tokyo are at least several times larger and longer than the original motions outside the basin.

We compared the development of long-period surface waves in the Kanto Basin that originated from large recent earthquakes located near Tokyo: a shallow (h = 15 km)

inland Mw 6.6 earthquake at Chuetsu, Niigata on October 23, 2004 (Fig. 4, line b–b'), and a relatively deep (h = 42 km) Mw 7.0 subduction zone earthquake off Miyagi Prefecture on 16 August, 2006 (Fig. 4, line c–c').

A plot of the progression of seismic waves from the Chuetsu earthquake to central Tokyo clearly demonstrates the development of surface waves in the Kanto Basin induced by conversion from S waves at the northern edge of the Kanto Basin (Fig. 4, line b–b'). The amplitude and duration of the surface waves gradually increased as they approached central Tokyo, but the developing properties and the duration of the surface waves was very mild compared to those of the 2004 SE Off-Kii Peninsula earthquake. The Off Miyagi Prefecture earthquake, a relatively deep (h = 42 km) event that occurred upon the subducting Pacific Plate in the Japan Trench, produced a very weak surface wave following a large, short-period S wave with long coda. It is widely recognized that deep earthquakes upon the subducting Pacific Plate show anomalously large ground acceleration and intensity along the eastern seaboard of the Pacific Ocean due to the guiding of short-period (T < 0.5 Hz) signals along the subducting plate. This occurs because of the multiple forward scattering of short-period signals within small-scale heterogeneities inside the plate (e.g., FURUMURA and KENNETT, 2005), but long-period signals with longer wavelengths do not capture such small-scale heterogeneities.

Consequently, of the three earthquakes that have recently occurred around Tokyo, shallow earthquakes in the Nankai Trough are potentially the most disastrous in terms of generating strong, prolonged long-period surface waves in the direction of central Tokyo. Thus, we anticipate a significant impact in Tokyo associated with M8 earthquakes in the Nankai Trough.

3. Long-period Ground Motions Arising from the 1944 Tonankai Mw 8.1 Earthquake

Long-period ground motions that occurred in Tokyo during the Mw 8.1 Tonankai earthquake in 1944 were well documented by the Imamura-type strong-motion instruments (natural period of 3–5 s and magnification of 2) and the Central Meteorological Observatory (CMO)-type strong-motion instruments (natural period of 3–5 s and a magnification of 1) installed in the seismic observatory in central Tokyo (Otemachi), at Chiba (Togane), and at Yokohama (Fig. 5).

The JMA has compiled historical strong-motion seismograms for major damaging earthquakes since 1927, and the data are available on a microfilm database. The Earthquake Research Institute of the University of Tokyo also has an archive of historical recordings of strong-motion instruments since 1920. We obtained recordings of strong motions in Tokyo during the 1944 Tonankai earthquake (FURUMURA and NAKAMURA, 2006), and will make use of these data in studying the characteristics of long-period ground motions in Tokyo arising from large M 8 earthquakes in the Nankai Trough.

In contrast to the ease of analysis of modern, high-quality digital seismograms, the smoked-paper seismograms of early strong-motion instruments is rather perplexing;

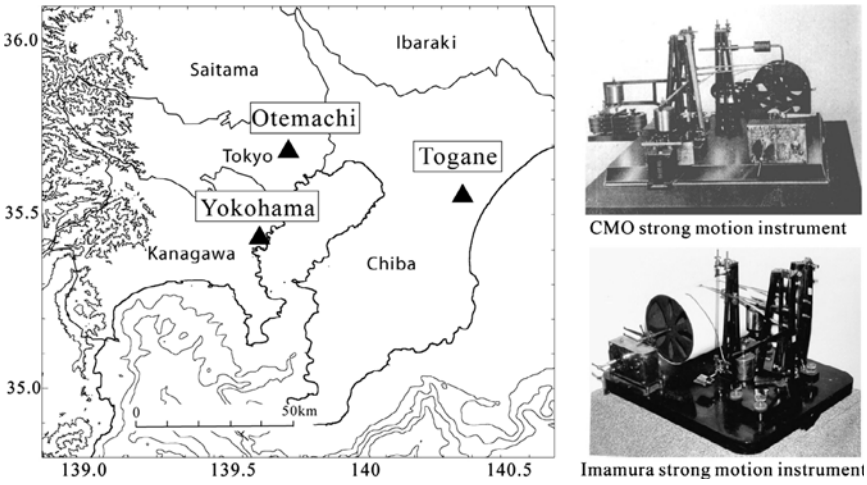


Figure 5
Map showing the locations of strong-motion stations that recorded strong ground motions during the 1944 Tonankai earthquake. The photographs show the CMO-type strong-motion instrument that was installed at Otemachi (upper) and the Imamura-type strong-motion instrument that was installed at Togane and Yokohama (lower).

however, recent developments of functional image-processing software made it possible to analyze these historical records (Fig. 6). The continuous recording method employed by these seismometers captured the entire record of the strong ground motions that occurred in Tokyo during the Tonankai earthquakes.

Images of the original smoke-paper seismographs were first captured using a scanner at high resolution (600 dpi) and are then digitized by hand tracing at small time intervals of less than 0.1 s. The digitized data set was corrected for pen-arc and base-line errors, and the digital waveform was then resampled with a uniform sampling interval of 10 Hz. Due to the effect of pen friction in the early seismometers, the effective frequency range of the seismogram is below about 1 Hz, so that we applied a bandpass filter to the waveform data with a bandpass frequency between 0.02 and 1 Hz.

Finally, the reproduced waveforms of the horizontal motions of the strong-motion instruments were translated to actual ground motions by removing the instrumental response from the recorded waveforms. Instrumental constants such as the natural period (T_0) and the damping constants (h) were obtained for each component of the seismometers by analyzing the time history of the impulse response curve recorded on the same seismograph paper.

We finally obtained the strong ground motions of horizontal ground displacement during the 1944 Tonankai earthquake at Togane, Otemachi, and Yokohama (Fig. 7). The reproduced ground motions resulting from the Tonankai earthquake, which were at least 2.5 times larger and several times longer than those recorded during the SE Off-Kii

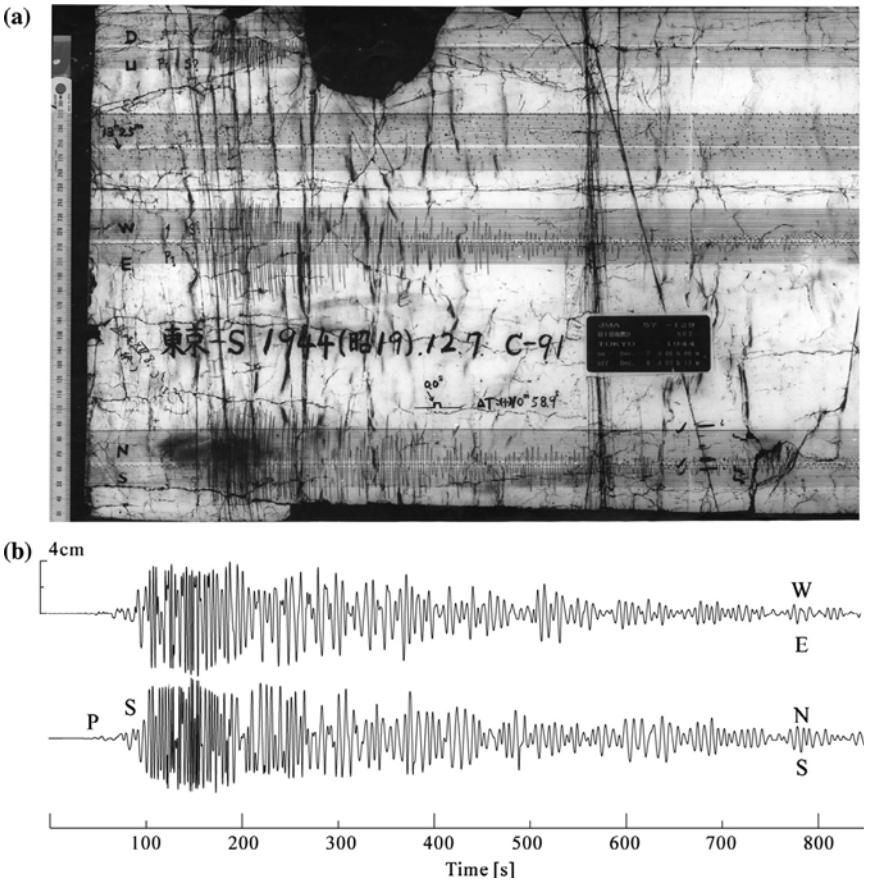


Figure 6
(a) Example of a smoked-paper seismogram from a CMO-type strong-motion instrument installed at Tokyo during the 1944 Tonankai Earthquake. (b) Digitized wave form data of the WE and NS components of displacement motions. Instrumental corrections were not applied to the seismograms.

Peninsula (Mw 7.4) earthquake in 2004, demonstrate a very large (>10 cm/s) and prolonged (>10 min) long-period ground-shaking at a dominant period of about 10–12 s.

4. Computer Simulation of Long-period Ground Motions Arising from Earthquakes in the Nankai Trough

To complement the observational data for the two earthquakes that occurred in the Nankai Trough, we conducted computer simulations of seismic wave propagation using a

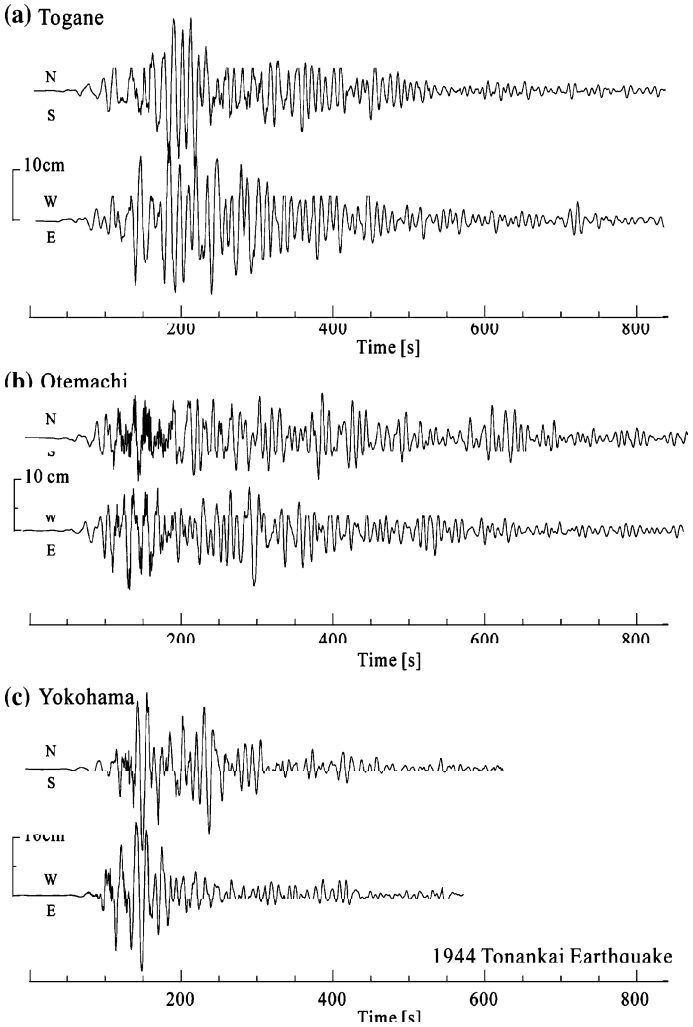


Figure 7
Wave forms of NS- and WE-component ground displacement motions during the 1944 Tonankai earthquake, as recorded at (a) Togan, (b) Otemachi, and (c) Yokohama. The instrumental responses of the strong-motion seismometers were corrected to reproduce the ground motions.

high-resolution structural model of central Japan and source-slip models for the earthquakes.

A structural model of the crust and upper-mantle beneath central Japan was recently developed from an analysis of reflection and refraction experimental data

(TANAKA *et al.*, 2006). A detailed structural model of the Nankai Trough subduction zone, with a shallowly dipping Philippine Sea Plate subducting beneath central Japan, has recently been investigated by compiling 29 profiles of reflection and refraction experiments (BABA *et al.*, 2006); this model was also included in the simulation model.

The simulation model covers an area of 496 km by 800 km, from the source region to Tokyo, and extends to a depth of 141 km. The model is discretized using a small mesh size of $0.4 \times 0.4 \times 0.2$ km for depths less than 10 km and a large grid size of $0.8 \times 0.8 \times 0.4$ km for areas from 10 to 141 km depth (Fig. 8).

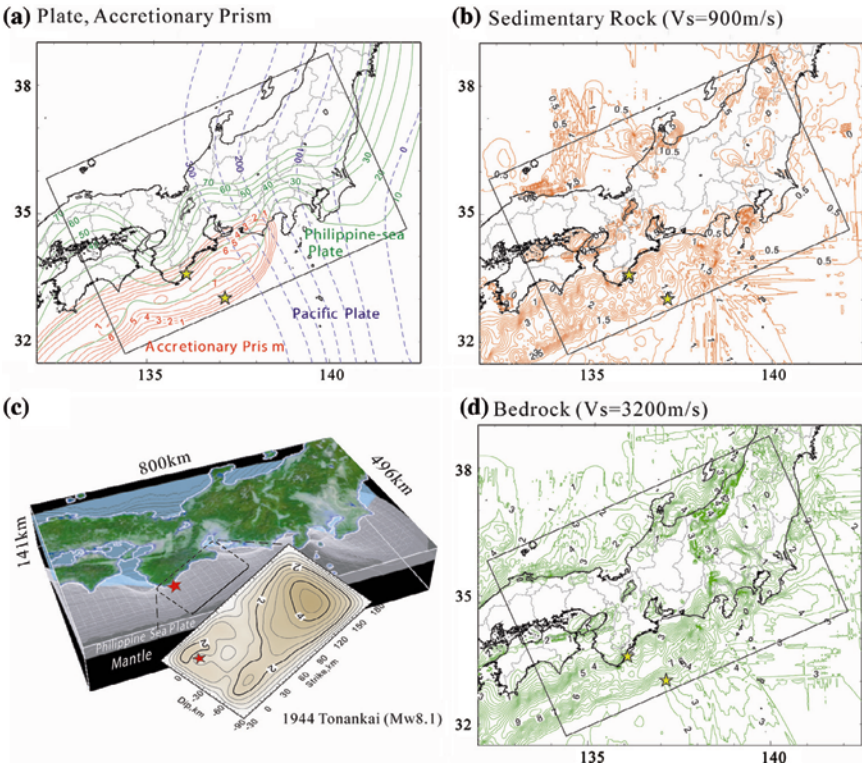


Figure 8

Structural model of central Japan used in the 3D wavefield calculations. (a) Outline of the study area, depth contours upon the subducting plates, and contours showing the thickness of the accretionary prism. (b) Depth distribution of the top of sedimentary layer with Vs = 900 m/s. (c) Depth distribution of the top of bedrock with Vs = 3200 m/s. (d) Details of the 3D model showing the configuration of the Philippine Sea Plate and the crust–mantle interface (Moho) and the source-slip model used for the 1944 Tonankai earthquake.

Table 1

Physical parameters of P- and S-wave velocity, rock density, and anelastic attenuation coefficients for each layer represented in the simulation

	Vp [km/s]	Vs [km/s]	ρ [t/m^3]	Qp	Qs
Sedimentary Layer					
Layer 1	1.8	0.5	2.0	100	50
Layer 2	2.3	0.9	2.1	200	100
Layer 3	3.0	1.5	2.3	300	150
Basement	5.5	3.2	2.6	600	300
Upper Crust	5.8	3.3	2.7	700	350
Lower Crust	6.5	3.7	2.7	800	400
Upper Mantle	8.0	4.5	2.8	1000	500

The sedimentary structure of the Kanto Basin is constructed using three layers (Vs = 0.5, 0.9, and 1.5 km/s) overlying a rigid bedrock of Vs = 3.2 km/s (TANAKA *et al*., 2006). The physical parameters of P and S wavespeed (Vp, Vs) and density (ρ) and attenuation coefficients (Qp, Qs) for the crust and upper-mantle structure and the subducting Philippine-Sea Plate are shown in Table 1. The Moho interface of the crust and upper-mantle boundary is based on the study of RYOKI (1999), and the depth of the mid-crustal (Conrad) interface is set to the middle of the crust.

The subducting Philippine Sea Plate consists of three layers of low-wavespeed oceanic layer 2 and layer 3 with a total thickness of 5 km, beneath which is a 25 km thickness of high-wavespeed oceanic mantle (BABA *et al*., 2006). The cover of low-wavespeed sediments (accretionary prism) along the trench is also accommodated within the simulation model (Fig. 8a). The wavespeed in each layer of the subducting plate is assumed to increase linearly with depth (Z) such that the physical parameters of the plate are assigned in the simulation model using the linear regression functions shown in Table 2.

The structural model employed in the simulation has been adjusted by matching predominant periods of observed ground motions at each station resulting from nearby earthquakes with theoretical estimates using physical parameters (Vp, Vs, ρ and thickness of each layer) of the structural model (TANAKA *et al*., 2006). The comparison between observations and theoretical estimates indicates that the model is capable of simulating the nature of a long-period wavefield of roughly over 1–2 s.

Table 2

Physical parameters of P- and S-wave velocity, rock density, and anelastic attenuation coefficients for each layer of the subducting plate and accretionary prism as a function of depth (Z) from sea level

	Vp [km/s]	Vs [km/s]	ρ ([t/m^3]	Qp	Qs
Accretionary Prism	3.5	1.5	1.8	90	50
Oceanic Layer 2	5.4 + 0.0055*Z	Vp/1.94	2.4	350	200
Oceanic Layer 3	6.5 + 0.0055*Z	Vp/1.87	2.5	500	300
Oceanic Mantle	8.1 + 0.0053*Z	Vp/1.76	2.8	850	500

The source-slip model for the 1944 Tonankai earthquake of YAMANAKA (2004a) has been estimated by an inversion using a near-field strong motion record and applying a low-pass filter with a cut-off frequency of 1 Hz. Thus the present simulation model of crust and upper-mantle structure in central Japan and source-slip model for the Tonankai earthquake is considered to be valid for simulating long-period ground motions of over about 1–2 s.

We did not include water column (Vp = 1.5 km/s, Vs = 0 km/s) in the present simulation model since such a low-wave speed layer below a free surface causes a serious instability problem during long-time FDM calculation due to a very large impedance contrast between water and low-wavespeed sedimentary layers at the coastal line area. Thus, our simulation underestimate the scattering and attenuation of P and S waves on the ocean side and it may underestimates the attenuation of P wave and elongation of P and S wave coda.

Seismic waves of relatively long-period ground motions over 1 s were calculated using a parallel, multi-grid Finite Difference Method (FDM; FURUMURA and CHEN, 2005), which uses a sixteenth -order staggered-grid FDM in horizontal (x, y) directions and a conventional fourth-order staggered-grid scheme in the vertical (z) direction. The parallel FDM simulation is based on the domain-partitioning procedure: The 3D model is partitioned horizontally into a number of subregions assigned to numerous processors, and the message-passing interface (MPI) is used for inter-processor communications.

5. Computer Simulation of the 2004 Off-Kii Peninsula Earthquake

The source slip model used for the 2004 Off-Kii Peninsula earthquake was derived from an inversion using far-field wave-form data (YAMANAKA, 2004a). The size of the fault is 60 × 24 km, with subfault segments of 4 × 4 km. To implement the source slip model in the fine simulation grid, the source slip distribution was resampled into 1 × 1 km subfaults using a linear interpolation function. A triangular source-slip function with a pulse width of 1 s was assigned to each subfault to radiate seismic waves with an assumed rupture propagation speed of Vr = 2.8 km/s.

The results of the computer simulation of the SE Off-Kii Peninsula earthquake are shown in Figure. 9 as snapshots of horizontal ground velocity motions. It is interesting to compare the simulation results (Fig. 9a) with the observed wavefield recorded from the dense K-NET and KiK-net seismic array (Fig. 9b). Whereas the seismic observations are largely restricted to land areas and dense seismic observations are unavailable in oceanic areas, the computer simulation provides details of the propagation of seismic waves from the offshore source region to inland areas. The modeling results are therefore very useful in understanding the complex seismic behavior that occurs during propagation through the 3D heterogeneous structure of the subduction zone. Note that the simulated wavefield is slightly dominated by the relatively short-period band because the snapshots are produced from wave forms recorded at densely and equally (in 1.6 km) spaced

Figure 9
Comparison of snapshots of ground velocity associated with the 2004 SE Off-Kii Peninsula earthquake at uniform time interval (T = 60, 120, 180, and 240 s) from earthquake rupture, as derived from (a) observations from the dense strong-motion network (K-NET and KiK-net), and (b) computer simulation.

observation points (stations), while the observations are restricted to a relatively low density (20–25 km) of recording stations.

In the first snapshots shown in Figure 9 (60 s), large ground motions radiate from the source and approach coastal areas around the hypocenter. Ground-shaking has just begun in coastal areas, but the computer simulation demonstrates a dramatic oscillation of ground motions above the source region associated with the amplification of ground motions within the thick cover of low-wavespeed sediments (accretionary prism) in the Nankai Trough. The results of the simulation indicate that land-based observations of the

Nankai Trough earthquake capture only a portion of the complicated wavefield generated by the earthquake.

The middle snapshots in Figure 9 show the striking development of surface waves as they propagate along the Nankai Trough toward Tokyo. The directional guidance of long-period surface waves during interaction with low-wavespeed sediments is clearly evident. As the surface waves enter the Kanto Basin, the ground motion is further amplified within the thick cover basin sediments, resulting in long and strong wavetrains of long-period ground motions in central Tokyo (T = 240 s).

This amplification and elongation of ground motions in the sedimentary basin is also clearly evident in other population centers, such as those in Osaka, Nagoya, and Toyama, in the T = 120 and 180 s frames, but the greatest ground motion is recorded for Tokyo. The strong motions in Tokyo are caused by the directivity effect of the rupturing fault from southwest to northeast and thus the directional guidance of long-period seismic waves toward Tokyo along the Nankai Trough. Ground motion in other populated area such as those in Toyama and Osaka area appears to be underestimated in the simulation snapshots, indicating that we need further improvements in the structure in those areas by close comparison between observed and simulated ground motions using this and other large events. The relatively smooth wavefield of computer simulation as compared with the snapshots of the observed wavefield also indicate that the present model is rather smooth to recover the scattering effect of the short-period wavefield. It should also be noted here that the amplitude of the P wave derived by the simulation is very weak as compared with observations such as those illustrated in 120 s frame in Tokyo. This is probably due to the source-slip model used in the simulation which is mainly constructed by using the S wave form and thus it was not so match constrained by the waveform of P wave.

In Figure 10, simulated wave forms of NS-component ground velocity motions are compared with observed K-NET and KiK-net strong-motion records for seven stations aligned from the source region to Tokyo. A low-pass filter with a cut-off frequency of 1 Hz was applied to both simulated and observed seismograms. The simulated and observed seismograms show good agreement in terms of major features such as P- and S-wave arrival times, peak ground velocity, dominant period of the surface waves, and the envelope shape of later S-wave coda. This demonstrates the reliability of the structure model and source model employed in the simulation. However, the simulation results underestimate high-frequency waves beyond about 1-0.5 Hz, especially for near source stations such as MIEH04, AIC017 and SZOH37, indicating the limitation of the present simulation model for applying for strong motion simulations of relatively higher-frequency wavefields of over 1 Hz.

6. Computer Simulation of the 1944 Tonankai Earthquake

We then evaluated long-period ground motions arising from the 1944 Tonankai earthquake using a source-slip model of YAMANAKA (2004b). The size of the inferred fault

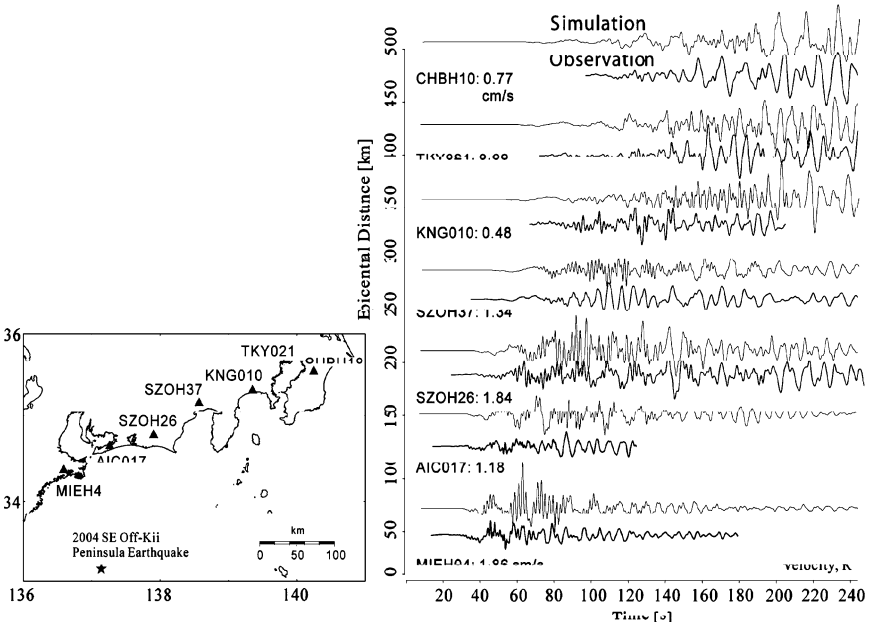


Figure 10
Comparison of simulated and observed wave forms of radial-component ground velocity motions at seven K-NET and KiK-net stations. Station locations are shown in the accompanying map.

is 180×90 km, which was resampled into 2×2 km subfaults using a linear interpolation function. A triangular slip-velocity function with a pulse width of 2 s was assigned to each subfault. As the inferred interpolated source-slip model might be producing a strong rupture-directivity effect along the direction of fault-rupture propagation, we introduced a random fluctuation in the fault-rupture speed at each point on the subfault. This random fluctuation is represented by a Gaussian function with a standard deviation of 5%, and the random fluctuation of rupture velocity is embedding over an average fault rupture speed of $Vr = 2.95$ km/s.

Snapshots of simulated horizontal ground velocity motions during the 1944 Tonankai earthquake are shown in Figure 11 at time intervals of 60 s. Each snapshot clearly demonstrates the spreading of S waves from the M 8 source and the development of long-period ground motions via interaction with sedimentary structures in the Nankai Trough. The effect of the low-wavespeed sedimentary wedge in terms of guiding long-period surface waves to the east is clearly evident in the middle frames ($T = 120$ and 180 s).

The snapshots capture the significant amplification of ground motions within large sedimentary basins such as those beneath Osaka, Nagoya, and Tokyo. This amplification was also detected in the observed wavefield during the 2004 SE Off-Kii Peninsula earthquake (Fig. 1, Fig. 8a).

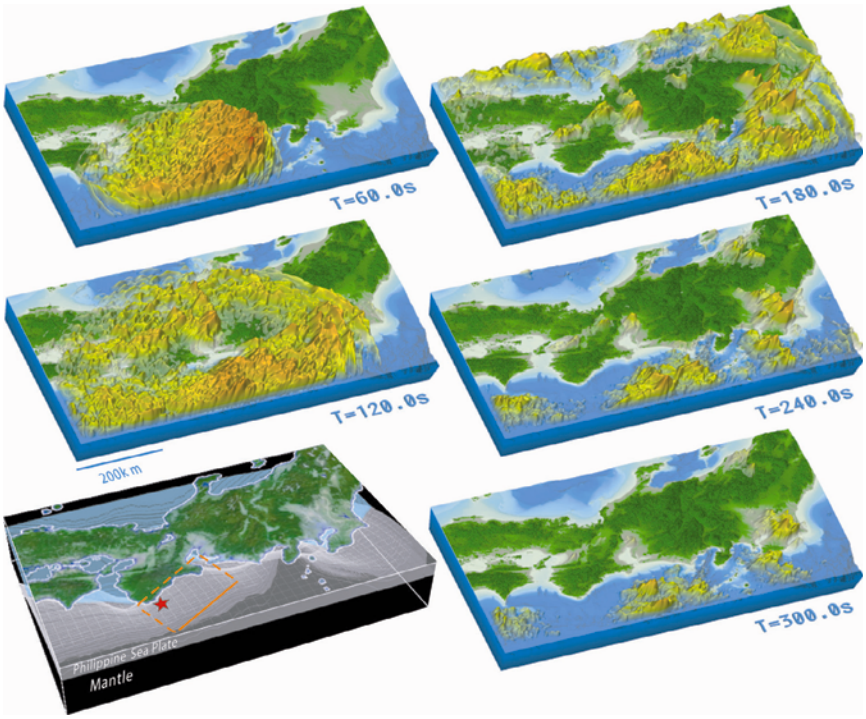


Figure 11
Snapshots of the propagation of simulated seismic waves arising from the 1944 Tonankai earthquake for T = 60, 120, 180, 240 and 300 s from source rupture.

Central Tokyo is affected by prolonged ground-shaking for several minutes leading up to the propagation of the S-wave front across the Kanto Basin to the north (T = 300 s). As the enhanced ground motions in the sedimentary basin are several times larger and longer than those observed during the 2004 SE Off-Kii Peninsula earthquake, so it might be a strong impact on modern large-scale constructions.

Figure 12 shows a comparison of the simulated waveforms of NS-component ground velocity motions and the velocity response spectra assuming $h = 0.05$ damping constant with observational data recorded at Togane, Otemachi, and Yokohama during the Tonankai earthquake. A low-pass filter with a band pass frequency of 0.02–0.5 Hz was applied to the simulated and observed seismograms. The simulation results demonstrate the major features of observed ground motions such as the dominant period of the surface waves and the shape of the S wave and its coda. The levels of the velocity response spectra of observed and simulated ground motions are in good agreement, but some differences in the dominant period of peak ground motion are recorded for Togane and

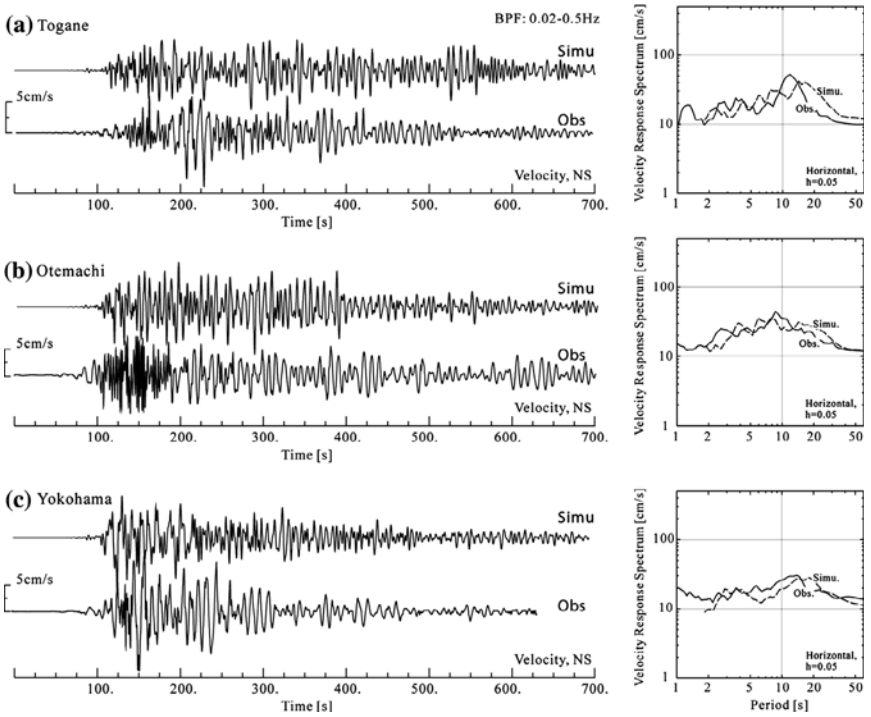


Figure 12
Comparison between the simulation (top) and observational data (bottom) in terms of the wave form of NS-component ground velocity motions and velocity response spectra recorded at (a) Togane, (b) Otemachi, and (c) Yokohama for the 1944 Tonankai earthquake.

Yokohama, probably due to limitations in the source model and uncertainties in the sedimentary structure model below these stations. However, the major features of the long-period ground motions developed within the Kanto Basin are well simulated by the present simulation, consequently we believe that the present model has largely attained the level of accuracy required to estimate the strength of long-period ground motions expected in central Tokyo during future earthquakes in the Nankai Trough.

The result of the computer simulation of the 1944 Tonankai earthquake also depicts strong ground motions within areas for which observational data were not available at the time due to a lack of seismic instruments or the fact that the seismometer was completely clipped due to intense ground-shaking (as occurred in near-fields such as in Osaka and Nagoya).

Figure 13 shows the simulated ground motions at stations located on major sedimentary basins, such as those at Osaka (OSK005), Nagoya (AIC003), Toyama

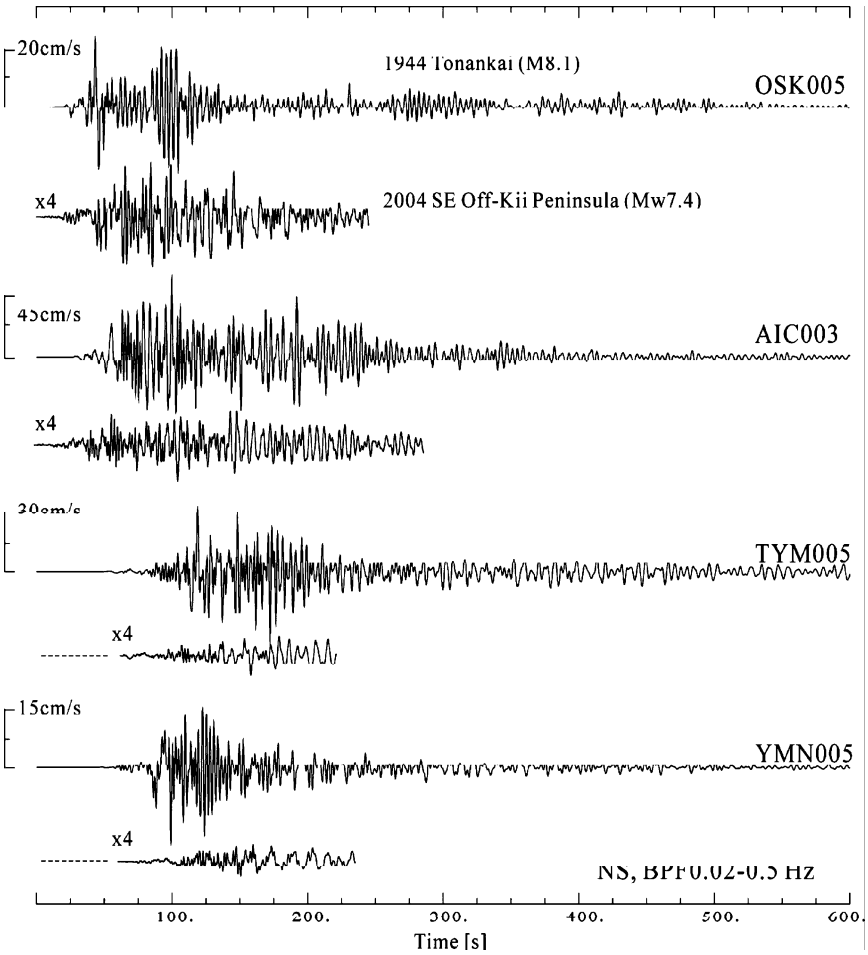


Figure 13
Simulated wave form of NS-component ground velocity motions for the 1944 Tonankai earthquake (top line) and comparison between observed wave forms from the 2004 SE Off-Kii Peninsula earthquake (bottom line, amplitude is multiplied by 4) at stations OSK005, AIC003, TYN005, and YMN005. A band pass filter of 0.02–0.5 Hz was applied to the wave forms.

(TYM005), and Kofu (YMN005). These results are compared with the observed ground motions for the 2004 SE Off-Kii Peninsula earthquake. It is apparent from the simulation that large and prolonged long-period ground motions occurred in major sedimentary basins following the occurrence of large earthquakes. The wave forms of the 1944 event are several times larger and longer than those of the 2004 earthquake. The dominant

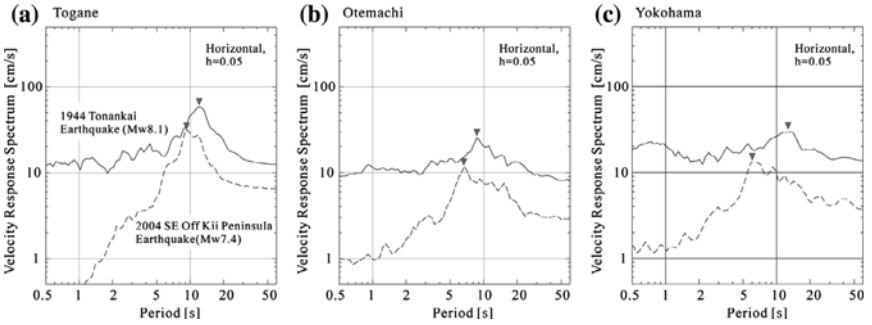


Figure 14
Comparison of the velocity response spectra for averaged horizontal motions from the 1944 Tonankai earthquake (solid lines) and the 2004 SE Off-Kii Peninsula earthquake, as recorded at stations OSK005, AIC003, TYM005, and YMN005.

period of long-period surface waves that developed in the basins is 2–3 s for Toyama and Kofu and 3–5 s for Osaka and Nagoya, approximately corresponding to the thicknesses of the basin sediments below the stations. The velocity response spectra for the 1944 Tonankai earthquake show a response that is approximately 10 times larger than that for the 2004 earthquake. This reflects the greater amplitude and longer duration of ground motions that arose from the Mw 8.1 earthquake.

7. Conclusions

The development of long-period ground motions with a predominant period of about 6–12 s by resonance in the thick cover of sedimentary layers beneath Tokyo is a common characteristic of all large events occurring around Tokyo, however the earthquakes in the Nankai Trough are expected to be the most disastrous in terms of producing extremely large and lengthy ground-shaking.

The strong motion record of the 1944 Tonankai Mw 8.1 earthquake demonstrates intense and prolonged shaking in central Tokyo associated with long-period ground motions with a dominant period of about 12–14 s. This finding highlights the potential vulnerability of modern large-scale construction to future earthquakes occurring in the Nankai Trough.

The present simulation model of crust and upper-mantle structure derived from the results of recent reflection and refraction experiments in central Japan and in the Kanto Basin and high-performance computing technology made possible the realistic simulation of long-period motions developing in the center of Tokyo during large earthquakes. However, the accuracy of the present model is rather limited in relatively longer-period bands over 1–2 s, and broadband simulation of strong ground motion including relatively

high-frequency signals over 1 Hz is still difficult due to uncertainties in the subsurface structure model and in the understanding of source-slip properties that affect the radiation and propagation of high-frequency seismic wavefields.

One of the important goals of strong-motion seismology is to predict the ground motions likely to occur in future earthquake scenarios and to assist in the development of appropriate building codes for different areas within sedimentary basins. Towards this goal, the improvements in the simulation model, especially the sedimentary structure beneath the Kanto Plain, are highly desired by maintaining a close link between computer simulations and observational studies that make use of the current high-density seismic network, archives of historical earthquakes.

Acknowledgements

This study was supported by the Multi-Scale and Multi-Physics Integrated Simulation project (CREST) of the Japan Science and Technology Agency, and the Special Project for Earthquake Disaster Mitigation in Urban Areas funded by the Ministry of Education, Culture, Sports, and Science, Japan. The computer simulation was conducted using the Earth Simulator. The author acknowledges the Earth Simulator Center for their support with computer resources. K-NET and KiK-net data were provided by the National Institute for Earth Science and Disaster Research, Japan. Photographs of Imamura-type and CMO-type strong motion instruments shown in Figure 5 were provided from the Hamamatsu Collection of the Japan Meteorological Agency. The authors would like to thank Kim Olsen and one anonymous reviewer for thoughtful review comments and suggestions for revising the manuscript.

References

Baba, T., Ito, A., Kaneda, Y., Hayakawa, T., and Furumura, T. (2006), *3D velocity structure model in the ocean around Japan inferred from controlled source seismic experiments*, Abst., Japan Geoscience Union Meeting. 2006.

Furumura, T. and Hayakawa, T. (2007), *Anomalous propagation of long-period ground motions observed in Tokyo during the 23 October 2004 Niigata-ken Chuetsu (Mw 6.6) earthquake*, Bull. Seismol. Soc. Am. 97, 863–880.

Furumura, T. and Nakamura, M. (2006), *Recovering of strong motion record of the 1944 Tonankai earthquake and long-period ground motion in Kanto region*, Geophys Exploration *59*, 337–351, in Japanese.

Furumura, T. and Kennett, B.L.N. (2005), *Subduction zone guided waves and the heterogeneity structure of the subducting plate: Intensity anomalies in northern Japan*, J. Geophys. Res., 110, B10302, dio:10.1029/2004JB003484.

Furumura, T., Kennett, B.L.N., and Koketsu, K. (2003), *Visualization of 3D wave propagation from the 2000 Tottori-ken Seibu, Japan, Earthquake: Observation and numerical simulation*, Bull. Seismol. Soc. Am. *93*, 870–881.

Furumura, T. and Chen, L. (2005), *Parallel simulation of strong ground motions during recent and histrical damaging earthquakes in Tokyo, Japan*, Parallel Computing, *31*, 149–165.

HATAYAMA, K., ZAMA, S., NISHI, H., YAMADA, M., and HIROKAWA, Y. (2004), *Long-period strong ground motion and damage to oil storage tanks dur to the 2003 Tokachi-oki earthquake*, Zisin *57*, 83–103.

HATAYAMA, K., and ZAMA, S. (2005), *Sloshing of liquid in oil storage tanks and long-period strong ground motions due to 2004 M 7-class earthquakes southeast off the Kii Peninsula*, Report of National Research Institute of Fire and Disaster *99*, 52–67.

HAYAKAWA, T., FURUMURA, F., and YAMANAKA, Y. (2005), *Simulation of strong ground motions caused by the 2004 off the Kii Peninsula earthquakes*, Earth Planets Space *57*, 191–196.

KODAIRA, S., IIDAKA, T., NAKANISHI, A., PARK, J.-O., IWASAKI, T., and KANEDA, Y. (2005), *Onshore-offshore seismic transect from the eastern Nankai Trough to central Japan crossing a zone of the Tokai slow slip event*, Earth Planets Space *57*, 943–959.

KOKETSU, K., HATAYAMA, K., FURUMURA, T., IKEGAMI, Y., and AKIYAMA, S. (2005), *Damaging long-period ground motions from the 2003 Mw 8.3 Tokachi-oki, Japan*, earthquake, Seismolol. Res. Lett. *76*, 1, 67–73.

MIYAKE, H. and KOKETSU, K. (2005), *Long-period ground motions from a large offshore earthquake: The case of the 2004 off the Kii peninsula earthquake, Japan*, Earth Planets Space *57*, 3, 203–207.

NAKANISHI, A., TAKAHASHI, N., PARK, J.-O. PARK, MIURA, S., KODAIRA, S., KANEDA, Y., HIRATA, N., IWASAKI, T., and NAKAMURA, M. (2002), *Crustal structure across the coseismic rupture zone of the 1944 Tonankai earthquake, the central Nankai Trough seismogenic zone*, J. Geophys. Res. *107*, doi:10.1029/2001JB000424.

RYOKI, K. (1999), *Three-dimensional depth structure of the crust and upper-most mantle beneath southwestern Japan and its regional gravity anomalies*, Zisin *52*, 51–63.

SHAPIRO, N.M., OLSEN, K.B., and SINGH, S.K. (2000), *Wave-guide effects in subduction zones: Evidence from three-dimensional modeling*, Geophys. Res. Lett. *27*, 433–436.

TAKAHASHI, N., KODAIRA, S., NAKANISHI, A., PARK, J.-O., MIURA, S., TSURU, T., KANEDA, Y., SUEHIRO, K., and KINOSHITA, H. (2002), *Seismic structure of western end of the Nankai Trough seismogenic zone*, J. Geophys. Res. *107*, doi:10.1029/200JB000121.

TANAKA, Y., MIYAKE, H., KOKETSU, K., FURUMURA, T., HAYAKAWA, T., BABA, T., SUZUKI, H., and MASUDA, T. (2006), *The DaiDaiToku integrated model of the velocity structure beneath the Tokyo metropolitan area (2)*, Abst., Japan Geoscience Union Meet. 2006, S116–P014.

USAMI, T., *Materials of Comprehensive List of Destructive Earthquakes in Japan* (University of Tokyo Press, 1996).

YAMANAKA, Y. (2004a), *Reanalysis of the 5 Sept. SE Off-Kii Peninsula earthquake (mainshock: Mj 7.4)*, EIC Seismological Note, Earthquake Research Institute, University of Tokyo, No.153.

YAMANAKA, Y. (2004b), *Source process of the 1944 Tonankai and the 1945 Mikawa earthquake*, Chikyu Monthly *305*, 739–745.

(Received September 6, 2006, revised January 1, 2007, accepted March 14, 2007)
Published Online First: April 2, 2008

Pure appl. geophys. 165 (2008) 609–620
0033–4553/08/030609–12
DOI 10.1007/s00024-008-0315-y

Pure and Applied Geophysics

Modeling Wing Crack Extension: Implications for the Ingredients of Discrete Element Model

YUCANG WANG,[1,2] and PETER MORA[2,1]

Abstract—In this study, we investigate what basic mechanisms a Discrete Element Model should have in order to reproduce the realistic wing crack extension, a widely observed phenomenon in uni-axial compression of brittle material with pre-existed crack. Using our Discrete Element Model—the Lattice Solid Model, we study how cracks propagate when different force-displacement laws are emplyed. Our results suggest that the basic features of crack propagation observed in laboratories cannot be reproduced under the following circumstances: 1) When only normal forces between two bonded particles exist and particle rotation is prohibited; 2) normal and shear stiffnesses are present and particle rotation is prohibited; 3) normal, shear stiffnesses and particle rotation are present and bending (rolling) stiffness is absent. Only when normal, shear and bending stiffness exist and particle rotation is permitted, is it possible to reproduce laboratory tests. We conclude that particle rotations and rolling resistance play a significant role and cannot be neglected while modeling such phenomenon. The effects of friction in the crack plane and confining pressure on extension of the cracks are also discussed.

Key words: Wing crack, discrete element method, the lattice solid model.

1. Introduction

The basic idea of Discrete Element Method (DEM, CUNDALL and STACK, 1979) is to treat the sample to be modeled as an assemblage of discrete particles interacting with one another. The typical calculations at each time step involve calculating forces by applying a force-displacement law at each contact and integrating equations of motion for each particle. Therefore, the force-displacement law between particles plays an important role in DEM simulations. Different contact laws have been used in various DEM models, either "real discrete" contact in which no tensile forces can be transmitted between particles, or "bonded" contact in which particles are bonded together so that tensile forces can be transmitted. The former model is used to model motions and behaviors of particulate materials (SHENG *et al.*, 2004; SITHARAM, 2000), and the latter is often used to model

[1] Earth Systems Science Computational Centre (ESSCC), The University of Queensland, St. Lucia, Brisbane, QLD 4072, Australia. E-mail: wangyc@esscc.uq.edu.au, or yucang_wang@hotmail.com

[2] Australian Computational Earth Systems Simulator (ACcESS), MNRF, The University of Queensland, St. Lucia, Brisbane, QLD 4072, Australia. E-mail: mora@esscc.uq.edu.au

fracture of intact material such as rocks (MORA and PLACE, 1993, 1994; MORA *et al.*, 2000; CHANG *et al.*, 2000; HAZZARD and YOUNG, 2000; POTYONDY and CUNDALL, 2004).

For the bonded models, there are different varieties according to interactions being transmitted between bonded particles: Only radial interaction is allowed (MORA and PLACE, 1993, 1994; MORA *et al.*, 2000; TOOMEY and BEAN, 2000), or only normal and shear forces are present (SCOTT, 1996; CHANG *et al.*, 2002; HAZZARD and YOUNG, 2000), or normal, shear forces and torques are transmitted (WANG *et al.*, 2000, 2006; IWASHITA and ODA, 1998, 2000; POTYONDY and CUNDALL, 2004).

Theoretically, to completely describe a 3-D rigid body, six degrees of freedom are required, three for the position and three for the orientation. Therefore six independent relative movements between two bodies can be permitted. The Lattice Solid Model (LSM), a bonded DEM, has been extended to include the complete set of interactions (WANG *et al.*, 2006). However, since extra complicity is introduced, the questions would arise: Is it necessary to include all these mechanisms? Can we get realistic results if we ignore some of mechanisms? The simple way to answer these questions is to investigate whether or not the typical laboratory experiments can be reproduced using DEMs when different mechanisms are included. Wing crack extension is selected as a benchmark test in this study.

Wing cracks are frequently observed in uni-axial compression of brittle materials with a pre-existed crack (BRACE and BOMBOLAKIS, 1963; ADAMS and SINES, 1978; INGRAFFEA and HEUZE, 1980; NEMAT-NASSER and HORII, 1982; HORII and NEMAT-NASSER, 1985, 1986; ASHBY and HALLAM, 1986; CANNON *et al*, 1990; GERMANOVICH *et al.*, 1994; SHEN, 1995; BOBET and EINSTEIN, 1998; ARCADY *et al.*, 1999, 2003; SCHULSON, 1999; RENSHAW and SCHULSON, 2001). It is found that ***tensile*** cracks nucleate at the tips of the flaw, grow in a ***stable*** manner with increasing compression, then tend to align with ***the direction of axial loads*** (BRACE and BOMBOLAKIS, 1963). Here stable means that with the increase of loads, cracks grow lowly, without causing abrupt overall fracture of the sample. The typical growth pattern of wing cracks is schematically shown in Figure 1. Wing crack extension is considered to be a possible mechanism for splitting along the loading axis of brittle solids (NEMAT-NASSER and HORII, 1982; CANNON *et al*., 1990). In fact, wing crack extensions are so widely observed that we think a reliable numerical model should be able to reproduce the basic features of such a phenomenon. In this study we will test the ability of our new model and discuss what ingredients a DEM should include in order to reproduce such observations. This paper is organized as follows: in Section 2 the Lattice Solid Model is briefly introduced, simulation results in different cases are presented in Section 3, and finally discussions and conclusions are given in Section 4.

2. A Brief Description of the Lattice Solid Model

The Lattice Solid Model (LSM) is a Discrete Element Model (DEM, Cundall and Stack, 1979) similar to the Molecular Dynamics (MD, ALLEN and TILDESLEY, 1987), but involves a different approach (PLACE and MORA, 1999). It was developed to simulate dynamics of

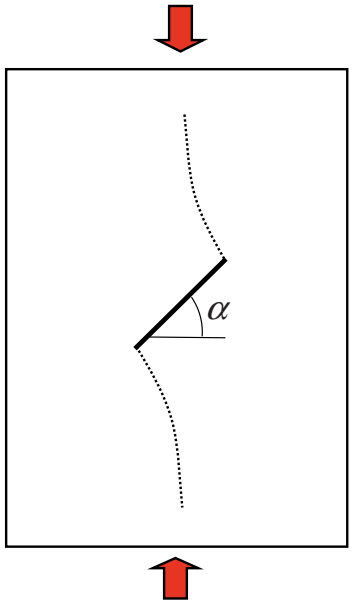


Figure 1
Schematic of typical wing crack propagation under compression.

earthquakes (MORA and PLACE, 1993, 1994; MORA *et al.*, 2000). The earlier version of the LSM only included radial forces between particles and single particle rotation is not allowed. Recently particle rotation and shear forces together with twisting and bending moments were incorporated (WANG *et al.*, 2006). Figure 2 shows six basic relative movements and interactions between two particles: Normal force, two shear forces, twisting torque and two bending torques. More details about representation of particle rotation, integration of rotational equations, decomposition of relative rotation and calibration of parameters can be found in WANG *et al.* (2006). In the 2-D case, six forces and torques in Figure 1 reduce to three: F_r, $F_s = F_{s1}$ and $M_b = M_{b2}$, and force-displacement law reduces to $F_r = K_r \Delta U_r$, $F_s = K_s \Delta U_s$ and $M_b = K_b \Delta \theta_b$, where K_r, K_s and K_b are normal, shear and rolling stiffness respectively, and ΔU_r, ΔU_s and $\Delta\theta_b$ are relative normal, shear and rotational displacement between two particles.

The criterion to judge whether or not a bond breaks is

$$\frac{F_r}{F_{r0}} + \frac{|F_s|}{F_{s0}} + \frac{|M_b|}{M_{b0}} \geq 1, \tag{1}$$

where F_{r0}, F_{s0} and M_{b0} are the thresholds for a bond to break under pure extension, pure shearing and pure bending, respectively. We set F_r positive under extension and negative

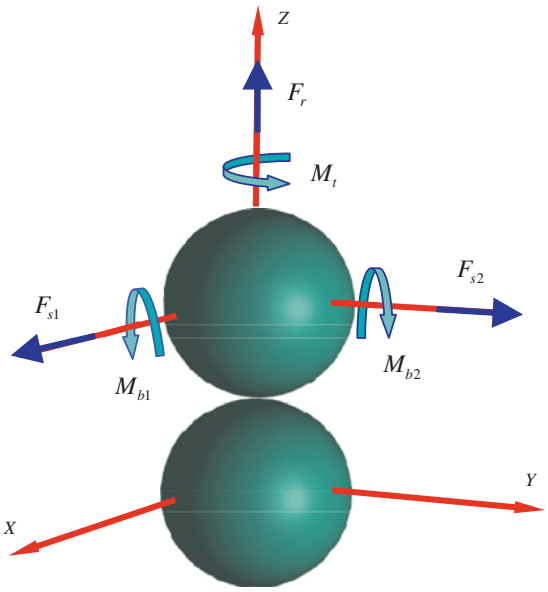


Figure 2
Six kinds of interactions between bonded particles. F_r is force in radial direction, F_{s1} and F_{s2} are shear forces, M_t is twisting torque, and M_{b1}, M_{b2} are bending torques.

under compression, therefore the effect of the normal force on breakage of a bond has been taken into account.

In this paper the LSM is used to model 2-D wing crack extension under quasi-static compression. We mainly focus on the exact pattern and stability of crack extensions. We are also interested in the effects of friction on the crack plane and lateral confining pressure on the crack extensions.

The simulation procedure is as follows: First we generate the rectangular sample of 20 cm $\times$ 40 cm using 6879 random-sized particles with radii r ranging from 0.5 to 5 mm (PLACE and MORA, 2001). Initially contacted particles are bonded by elastic-brittle bonds which can transfer all three interactions mentioned above, representing intact materials. A bond breaks if Eq. (1) is satisfied. Then we create a pre-existed crack of length $=$ 5 cm in the direction determined by angle α (Fig. 1). This is done by removing the bonds in the crack area such that only frictional and repulsive forces exist between particles when they contact. Finally we load the model at constant velocity to study how cracks extend in four different cases: Only normal stiffness exists without particle rotation; or both normal and shear stiffnesses exist without particle rotation; or normal, shear stiffnesses and particle rotation are present and rolling resistance is absent; or normal, shear, rolling stiffnesses and particle rotation are all permitted. We also examine the effects of crack orientation, frictional forces on crack plane and confining pressure on extensions of the cracks.

Except for the certain mechanisms that are missing, the model parameters are the same for all cases: density $\rho = 2000$ kg/m^3, mass $M = \rho\pi R^2$ (R is particle radius), stiffness parameter $K_r = 4.0 \times 10^7$ N/m, $K_s = 3.2 \times 10^7$ N/m, $K_b = 100$ Nm/rad, fracture parameter $F_{r0} = 6000$ N, $F_{s0} = 24000$ N, $M_{b0} = 10$ Nm, time step $dt = 1.0 \times 10^{-6}$ s, compression velocity at two ends $V_{comp} = 0.015$ m/s.

3. Simulation Results

A group of typical simulation results in four cases mentioned above are shown in Figures 3–7. Table 1 lists some measured macroscopic parameters including macroscopic strength, time of crack starting to extend t_1, time of overall fracture t_2, difference $\Delta t = t_2 - t_1$, or extension time period of cracks, which roughly represent the stability of cracks extension.

3.1. Only Normal Stiffness Exists without Rotation: $K_r \neq 0$, $K_s = 0$, $K_b = 0$

When only normal stiffness between two bonded particles exists without shear stiffness and particle rotation, it is found that in all three cases with different orientations ($\alpha = 30°$, 45° and 60°), cracks always propagate along the direction close to the original direction of the pre-existed cracks, as is shown in Figure 3. The durations of crack

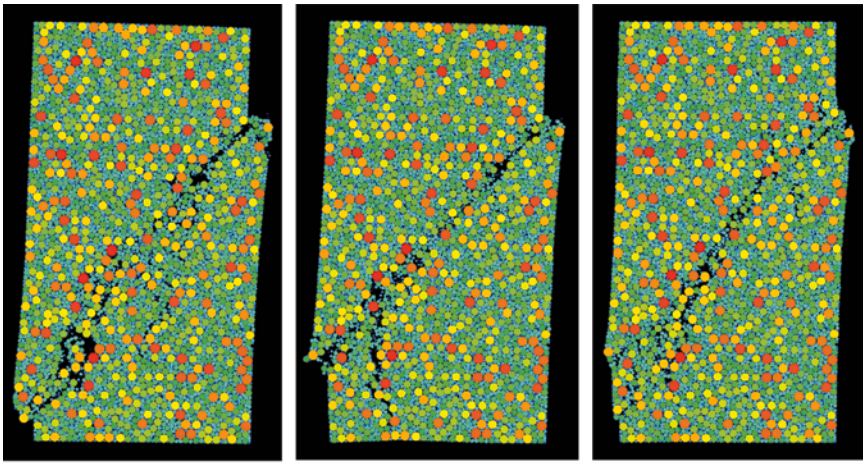


Figure 3
Crack extension in case of only normal stiffness is present without shear stiffness and single particle rotation.

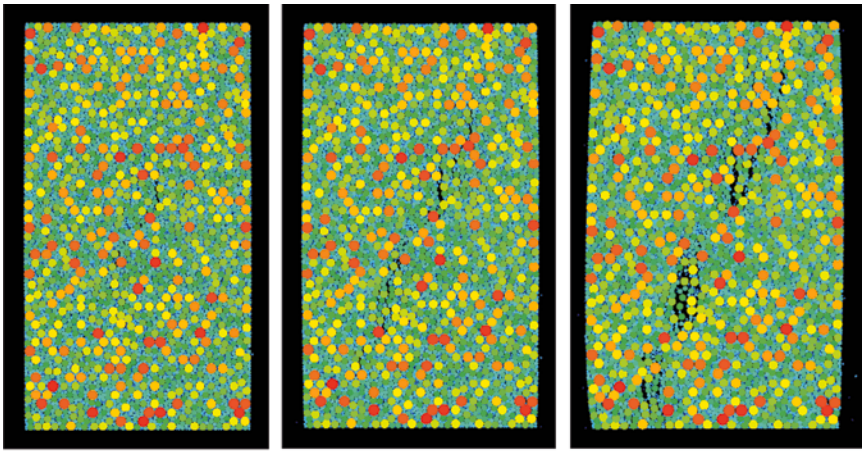


Figure 4
Crack propagation when normal, shear stiffness exist without single particle rotation.

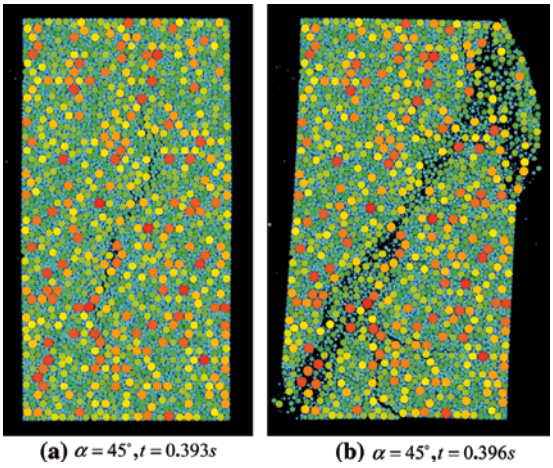


Figure 5
Crack propagation when normal, shear stiffness and single particle rotation exist but without bending stiffness.

extension are very short (ranging from 0.009 s to 0.078 s, see Table 1), indicating unstable propagation. These suggest that the model is oversimplified, and realistic crack extension paths cannot be obtained.

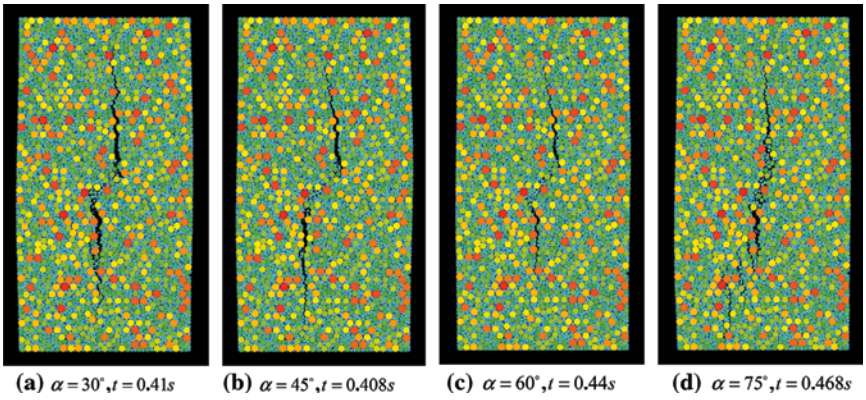


Figure 6
Crack propagation when normal, shear stiffness, bending stiffness and single particle rotation are all present.

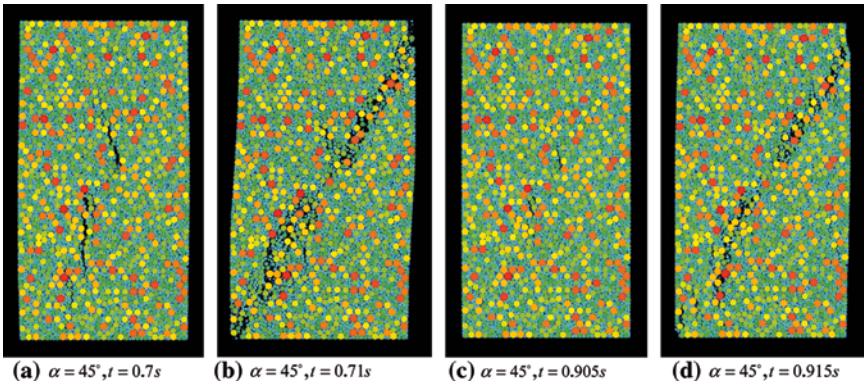


Figure 7
Effects of confinement on crack extensions. Confining pressure $\sigma_2 = 6.25 \times 10^4$ Pa (*a* and *b*) and $\sigma_2 = 1.875 \times 10^5$ Pa (*c* and *d*).

3.2. Normal, Shear Stiffness Exists, with Single Particle Rotation Prohibited, $K_r \neq 0$, $K_s \neq 0$

Figure 4 shows crack patterns at time $t = 1.33$ s, 1.8 s and 2.27 s in case of $\alpha = 45°$. In this case, tensile cracks start at the tips of the flaw at $t = 0.7466$ s, but they grow slowly and discontinuously, and several *en echelon* cracks appear at $t = 1.59$ s. Finally macroscopic shear band is formed along the direction between the original direction of flaw and the direction of axial compression, followed by macroscopic failure at

Table 1

Macroscopic measurement of cracks starting time, overall failure time, and crack extension time and macroscopic strength in different cases.

		Crack starting time t_1 (s)	Overall failure time t_2 (s)	Duration $\Delta t = t_2 - t_1$ (s)	Macroscopic strength (N)
Fig. 3	$\alpha = 30°$	0.366	0.375	0.009	71250
	$\alpha = 45°$	0.321	0.348	0.027	66625
	$\alpha = 60°$	0.297	0.375	0.078	71750
Fig. 4	$\alpha = 45°$	0.7466	2.282	1.535	1480812
Fig. 5	$\alpha = 45°$	0.393	0.396	0.003	148500
Fig. 6	$\alpha = 30°$	0.402	0.714	0.312	401688
	$\alpha = 45°$	0.208	0.592	0.384	323750
	$\alpha = 60°$	0.268	0.616	0.348	336875
	$\alpha = 75°$	0.464	0.672	0.208	367500
Fig. 7 $\alpha = 45°$	Smaller confinement	0.44	0.71	0.27	451500
	Larger confinement	0.56	0.915	0.355	579375
Friction $\alpha = 45°$	$\mu = 0.0$	0.208	0.592	0.384	323750
	$\mu = 0.1$	0.28	0.596	0.296	353062
	$\mu = 0.2$	0.30	0.584	0.284	326000
	$\mu = 0.4$	0.314	0.58	0.266	335375

$t = 2.282$ s. The crack propagation is stable, but the correct path cannot be reproduced. We also noticed that the macroscopic strength of the sample is greatly increased (Table 1). Similar results are observed when $\alpha = 30°$ and $60°$.

3.3. Normal, Shear Stiffness and Single Particle Rotation Exist but without Rolling Resistance $K_r \neq 0$, $K_s \neq 0$, $K_b = 0$

When rolling resistance is absent, cracks grow first along the direction between their original direction and the direction of the axial compression at $t = 0.393$ s, quickly tend to the direction of axial compression (Fig. 5a). The extension is quite unstable, characterized by very short extension time ($\Delta t = 0.003$ s) before an abrupt overall failure of the sample at $t = 0.396$ s (Fig. 5b).

3.4. Normal, Shear Rigidity, Single Particle Rotation and Rolling Resistance all Exist, $K_r \neq 0$, $K_s \neq 0$, $K_b \neq 0$

If all of the four mechanisms are present, we found that for the different orientations ($\alpha = 30°$, $45°$, $60°$ and $75°$), tensile cracks are observed to appear, propagate in a stable manner and then bend towards the direction of axial compression (Fig. 6). It is always easy to reproduce the laboratory test within a relatively wider range of parameters when normal, shear stiffness, particle rotation and rolling resistance are all present.

3.5. *Effects of Crack Orientation, Friction and Confining Pressure*

Generally we found that realistic wing crack extension modes are reproduced for α from 30 to 75 degrees (Figs. 6 a–d). From Table 1 it is interesting to note that wing cracks appear earlier in the case of $\alpha = 45°$ (at $t = 0.208$ s) than $\alpha = 75°$ ($t = 0.464$ s). This phenomenon was also reported in laboratory tests (ADAMS and SINES, 1978). The reason is that surfaces of the flaw slide due to the resolved shear stress caused by axial compression. This sliding generates tensile stress at the tips of the flaw. $\alpha = 45°$and $\alpha = 75°$ correspond to the maximum and minimum resolved shear stress in this case, therefore cracks appear earlier in the case of $\alpha = 45°$. It is also noted that crack propagation is more stable in the case of $\alpha = 45°$ than in the case of $\alpha = 75°$ (Table 1).

In Figure 6 there are no frictional forces between particles in the crack area. Similar simulations have been done with different frictional coefficients. It is found that when the frictional coefficients increase from 0 to 0.4, the tensile cracks tend to appear later (from $t = 0.208$ s to $t = 0.314$ s) and the propagation of cracks is more unstable. The crack propagation patterns similar to those of Figure 6 are observed. It seems that the frictional coefficient has only one minor effect on the macroscopic strength (Table 1).

Figure 7 shows the effects of confining pressure in the case of $\alpha = 45°$. When the smaller confining pressure ($\sigma_2 = 6.25 \times 10^4$ Pa) is applied in the horizontal direction (Figs. 7a, b), tensile cracks start at flaw tips at $t = 0.44$ s, although not as developed as that without confinement (see Fig. 6b). The cracks grow slowly to the size of the original crack and stop until unstable macroscopic shear band appears, followed by an abrupt failure at $t = 0.71$ s. With larger confining pressure applied (Figs. 7c, d, $\sigma_2 = 1.875 \times 10^5$ Pa), the tensile cracks appear later at $t = 0.56$ s, are even smaller and less developed (just half the size of the original flaw) before the overall collapse at $t = 0.915$ s. It is obvious that the confining pressure hampers wing crack extension, the larger the confining pressure is, the later cracks appear and less developed the cracks are before overall failure and the larger the macroscopic strength is (Table 1).

4. *Discussions and Conclusions*

We modeled the widely observed wing crack phenomenon using our Discrete Element Model and studied the effects of the force-displacement law on the capability of the model. The results in this study show that: 1) When only a normal interaction exists, theory predicts that cracks always propagate along the original crack direction, contrary to experiments; 2) when both normal and shear stiffnesses are present and single particle rotation is prohibited, tensile cracks start at the tips of the flaw, but grow in a jump-like manner, forming several en echelon cracks. Finally an abrupt shear band appeared propagating between its original direction and the direction of axial compression; 3) without rolling resistance, cracks propagate between their original direction and the

direction of axial compression, then tend to the direction of axial compression. The resulting extension is quite unstable, followed by abrupt failure of the model; 4) only when normal, shear and rolling stiffnesses are present and particle rotation is permitted, is it possible to reproduce the laboratory wing crack tests (including the tensile crack at the tips of the flaw, the correct crack extension paths, stable mode) within a wider range of parameters.

We also found that higher frictional coefficient between particles delays the nucleation time of tensile cracks and causes unstable propagation of cracks. A possible reason is that higher friction could increase the initial stress required to slide the crack plane, therefore accumulating more energy before the overall failure occurs. Orientation of the initial flaw also effects the nucleation time of the tensile cracks. Increasing the confining pressure hampers wing crack extension. Some simulations with different arrangements of random particles were also made, and all results were similar to those presented here, suggesting that our simulations are statistically robust.

This study suggests that shear stiffness cannot be neglected in particle-based models of fracture. Single particle rotation and rolling resistance must also be included due to their important roles in reproducing phenomena observed in laboratory tests. Neglecting these important mechanisms results in inaccurate transmission of forces and torques between particles, especially when stress concentration appears due to local fracture. Consequently if these mechanisms are neglected, the exact crack paths cannot be reproduced accurately.

Acknowledgments

Funding support is gratefully acknowledged by the Australian Computational Earth Systems Simulator Major National Research Facility, The University of Queensland and SGI. The ACcESS MNRF is funded by the Australian Commonwealth Government and participating institutions (Univ. of Queensland, Monash U, Melbourne U., VPAC, RMIT) and the Victorian State Government. 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. OpenDX is used to visualize the results in this paper. The first author would like to thank Dr. Fernando Alonso-Marroquin and Dr. Louise Kettle for their valuable suggestions to the manuscript.

References

Adams, M. and Sines, G. (1978), *Crack extension from flows in a brittle material subjected to compression*, Tectonophysics *49*, 97–118.

Allen, M.P. and Tildesley, D.J., *Computer Simulation of Liquids* (Oxford Science Press, Oxford, 1987).

Arcady, V.D., Germanovich, L.N. and Ustinov, K.B. (1999), *A 3-D model of wing crack growth and interaction*, Engin. Frac. Mech. *63*, 81–110.

ARCADY, V.D., SAHOURYEH, E., JEWELL, R.J., JOER, H., and USTINOV, K.B. (2003), *Influence of shape and locations of 3-D cracks on their growth in uniaxial compression*, Engin. Frac. Mech. *70*, 2115–2136.

ASHBY, M.F. and HALLAM, S.D. (1986), *The failure of brittle solids containing small cracks under compressive stress states*, Acta Metallurgica 34, 497–510.

BOBET, A. and EINSTEIN, H.H. (1998), *Fracture coalescence in rock-type materials under uniaxial and biaxial compression*, Int.J. Rock Mech. Sci. *35*(7), 863–888.

BRACE, W.F. and BOMBOLAKIS, E.G. (1963), *A note on brittle crack growth in compression*, J. Geophys. Res. *68*, 3709–3713.

CANNON, N.P., SCHULSON, E.M., SMITH, T.R., and FROST, H.J. (1990), *Wing cracks and brittle compressive fracture*, Acta Metall. Mater. *38*(10), 1955–1962.

CHANG, S.H., YUN, K.J., and LEE, C.I. (2002), *Modeling of fracture and damage in rock by the bonded-particle model*, Geosystem Eng. *5*(4), 113–120.

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

GERMANOVICH, L.N., SALGANIK, R.L., DYSKIN, A.V., and LEE, K.K. (1994), *Mechanisms of brittle fracture of rock with pre-existing cracks in compression*, Pure Appl. Geophys. *143*, 117–149.

HAZZARD, J.F. and YOUNG, R.P. (2000), *Simulation acoustic emissions in bonded-particle models of rock*, Inter.J. Rock Mech. Min. Sci. *37*, 867–872.

HORII, H. and NEMAT-NASSER, S. (1985), *Compression-induced microcrack growth in brittle solids: Axial splitting and shear failure*, J. Geophys. Res. *90*(B4), 3105–3125.

HORII, H. and NEMAT-NASSER, S. (1986), *Brittle failure in compression: splitting, faulting, and brittle-ductile transition*, Phil. Trans. R. Soc. Lond. A *319*, 337–374.

INGRAFFEA, A.R. and HEUZE, F.E. (1980), *Finite-Element Models for rock fracture mechanics*, Int. J. Numer. Anal. Meth. Geomech. *4*, 25–43.

IWASHITA, K. and ODA, M. (1998), *Rolling resistance at contacts in simulation of shear-band development by DEM*, J. Eng. Mech. *124*, 285–292.

IWASHITA, K. and ODA, M. (2000), *Micro-deformation mechanism of shear bending process based on modified distinct element method*, Power Tech. *109*, 192–205.

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. and PLACE, D. (1994), *Simulation of the frictional stick-slip instability*, Pure Appl. Geophys. *143*, 61–87.

MORA, P., PLACE, D., ABE, S., and JAUMÉ, S. (2000), *Lattice solid simulation of the physics of earthquakes: The model, results and directions*. In *GeoComplexity and the Physics of Earthquakes* (Geophysical Monograph Series; no. 120) (eds. Rundle, J.B., Turcotte, D.L. and Klein, W., pp. 105–125 (Am. Geophys. Union, Washington, DC).

NEMAT-NASSER, S. and HORII, H. (1982), *Compression-induced nonplannar crack extension with application to splitting, exfoliation, and rockburst*, J. Geophys. Res. *87*(B8), 6805–6821.

PLACE, D. and MORA, P. (1999), *A lattice solid model to simulate the physics of rocks and earthquakes: Incorporation of friction*, J. Comp. Phys. *150*, 332–372.

PLACE, D. and MORA, P. *A random lattice solid model for simulation of fault zone dynamics and fracture process*. In *Bifurcation and Localization Theory for Soil and Rock'99*, (eds. Muhlhaus, H.B., Dyskin, A.V., and Pasternak, E.) (AA Balkema, Rotterdam/ Brookfield (2001)).

POTYONDY, D. and CUNDALL, P. (2004), *A bonded-particle model for rock*, Int. J. Rock Mech. Min. Sci. *41***,** 1329–1364.

RENSHAW, C.E. and SCHULSON, E.M. (2001), Universal behavior in compressive failure of brittle materials, Nature *412*, 897–900.

SCHULSON, E.M. (1999), *The structure and mechanical behavior of ice*, J. Min. Met. Mater. Soc. *51*, 21–27.

SCOTT, D.R. (1996), *Seismicity and stress rotation in a granular model of the brittle crust*, Nature 381, 592–595.

SHEN, B. (1995), *The mechanism of fracture coalescence in compression – experimental study and numerical simulation*, Engin. Frac. Mech. 51, 73–85.

SHENG, Y., LAWRENCE, C.J., BRISCOE, B.J., and THORNTON, C. (2004), *Numerical studies of uniaxial powder compaction process by 3D DEM*, Eng. Comp. *21*, 304–317.

SITHARAM, T.G. (2000), *Numerical simulation of particulate materials using discrete element modeling*, Current Science *78*(7), 876–886.

TOOMEY, A. and BEAN, C.J. (2000), *Numerical simulation of seismic waves using a discrete particle scheme*, Geophys. J. Inter. *141*, 595–604.

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

WANG, Y.C., ABE, S., LATHAM, S., and MORA, P. (2006), *Implementation of particle-scale rotation in the 3-D Lattice Solid model*, Pure Appl. Geophys. *163*, 1769–1875.

(Received October 31, 2006, revised September 20, 2007, accepted September 20, 2007)

Pure appl. geophys. 165 (2008) 621–633
0033–4553/08/030621–13
DOI 10.1007/s00024-008-0317-9

Pure and Applied Geophysics

Interactive Visualization to Advance Earthquake Simulation

Louise H. Kellogg,[1] Gerald W. Bawden,[2] Tony Bernardin,[3] Magali Billen,[1] Eric Cowgill,[1] Bernd Hamann,[3] Margarete Jadamec,[1] Oliver Kreylos,[3] Oliver Staadt,[3] and Dawn Sumner[1]

Abstract—The geological sciences are challenged to manage and interpret increasing volumes of data as observations and simulations increase in size and complexity. For example, simulations of earthquake-related processes typically generate complex, time-varying data sets in two or more dimensions. To facilitate interpretation and analysis of these data sets, evaluate the underlying models, and to drive future calculations, we have developed methods of interactive visualization with a special focus on using immersive virtual reality (VR) environments to interact with models of Earth's surface and interior. Virtual mapping tools allow virtual "field studies" in inaccessible regions. Interactive tools allow us to manipulate shapes in order to construct models of geological features for geodynamic models, while feature extraction tools support quantitative measurement of structures that emerge from numerical simulation or field observations, thereby enabling us to improve our interpretation of the dynamical processes that drive earthquakes. VR has traditionally been used primarily as a presentation tool, albeit with active navigation through data. Reaping the full intellectual benefits of immersive VR as a tool for scientific analysis requires building on the method's strengths, that is, using both 3D perception and interaction with observed or simulated data. This approach also takes advantage of the specialized skills of geological scientists who are trained to interpret, the often limited, geological and geophysical data available from field observations.

Key words: Interactive visualization, virtual reality, earthquake simulation, active tectonics, virtual mapping.

1. Visualization of Geoscience Data

The human brain excels at visually identifying patterns, and as a result the best interpretations arise when scientists can fully visualize their data. As the expert on informational graphics Edward Tufte wrote two decades ago: "At their best, graphics are instruments for reasoning about quantitative information. Often the most effective way to describe, explore, and summarize a set of numbers—even a very large set—is to look at

[1] Department of Geology and W.M. Keck Center for Active Visualization in the Earth Sciences, University of California, Davis, CA 95616, USA . E-mail: kellogg@geology.ucdavis.edu

[2] USGS Western Remote Sensing and Visualization Center, U.S. Geological Survey, Sacramento, CA 95819, USA.

[3] Institute for Data Analysis and Visualization (IDAV), Department of Computer Science, and W.M. Keck Center for Active Visualization in the Earth Sciences, University of California, Davis, CA 95616, USA.

pictures of those numbers" (TUFTE, 1983). Earth science data sets have now increased in size and complexity to the extent that they can no longer be represented adequately in numerical form (ERLEBACHER *et al*., 2001; COHEN, 2005). Although statistical distributions and correlations can yield insight, by definition such an approach reduces the amount of information conveyed. As it becomes increasingly easy both to model interacting, nonlinear processes and measure natural systems, developing new ways of understanding and interpreting these expanding data sets is critical to making significant scientific advances (FOSTER, 2006; BUTLER, 2006] and responding to natural disasters (NOURBAKHSH, 2006). Using our innate abilities to interpret vast amounts of very complex visual information and focus our attention on the most salient features is the best technique for gaining the scientific insights that produce breakthroughs in difficult problems. Advanced visualization technology allows scientists to use their full visual capacity, helping them to identify previously unrecognized processes and interactions in complex systems (see e.g., CARLSON, 2006 for a discussion of recent advances in imaging geological materials).

2. *Immersive Visual Data Analysis*

However, it is not the case that visualization ends with a picture; on the contrary, visual data analysis just begins at this point. A picture should be the starting point for exploration, and visual exploration software should make it easy and fast to generate a picture that shows a feature of interest, and then provide the analysis tools to classify, measure, and understand the feature. This combined approach often leads to a deeper understanding of the scientific problem at hand.

As an example, consider a researcher who suspects an anomaly in one of her FEM (finite-element method) simulations. Using an interactive visual data analysis program, she explores her data by standard visualization techniques, such as creating and manipulating slices or contour surfaces, until she finds, say, a surface exhibiting an unexpected outgrowth. She then uses basic navigation to zoom in on the feature and look at it from different angles, until she obtains an understanding of its shape and relation to the surrounding data. Finally, she measures the location of the feature, e.g., the index of the FEM grid cell containing it, and the data values surrounding it. By repeating this process she gathers a list of "problem zones" that can lead her to ways of adjusting her model parameters, or that she can even use to debug the simulation code itself. As a result, she can fix or improve the accuracy of her models.

Importantly, pictures such as those shown throughout this paper are just transient products in this analysis process, and the novelty and benefit of interactive visualization does not lie in the pictures themselves, but in how easily they are generated and how they are used to gain understanding.

Making visual data analysis effective imposes constraints on both the underlying display hardware and software. The process relies on nonspecialist users being able to explore complex data and quickly find those visualization elements, e.g., slices or contour

surfaces that show the features they are looking for. Furthermore, once features are found, users must be able to examine them easily, and measure them accurately. User interfaces should be transparent, i.e., users should be able to completely focus on their data, and not have to worry about how to tell the software to do the things they want done. Immersive visualization, or virtual reality (VR), is a combination of hardware and software that is very appropriate for such tasks (KREYLOS, 2006; KREYLOS *et al*., 2006). In this context, "immersive" means that users can see, and interact with, virtual 3-D objects as if they were real. A virtual globe looks like a real globe, including the user's ability to walk around it and get closer looks, and can be picked up and moved around using the user's hands (via hand-held devices or "data gloves"). The crucial point is that users can interact with virtual objects using real-world actions, which enables user interfaces for visualization that are considerably more efficient and intuitive than those typically used in nonimmersive environments such as desktop computers. Measuring, say, the position of a point in (virtual) space, or the distance between two such points, involves merely touching those points in real space. We have found that the natural interfaces offered by VR allow geoscientists to fully utilize their training in the interpretation of 3-D Earth structure and 4-D (space plus time) reconstructions of geological processes to interpret computational data sets.

In other words, immersive visualization does not mean that users are surrounded by imagery on all sides, it means that users perceive virtual space as real. Technically, immersion requires stereoscopic display, head tracking, and 3-D input devices (KREYLOS, 2006). If any of those are missing, the illusion of real space breaks down, leading to a degradation of ease of use. For example, IMAX 3-D theaters offer only stereoscopy, and most 3-D projection systems such as Geowalls offer only stereoscopy and 2-D input devices, e.g., mice. The benefit of CAVEs (multi-screen walk-in environments) over single-screen immersive environments is that users are additionally surrounded by imagery, such that their peripheral vision improves the understanding of complex multiscale data, and such that measured data like surface topography can be displayed up to 1:1 scale. On the other end of the spectrum, nonimmersive systems can still be used for interactive visual data analysis, but at reduced efficiency depending on their features. For example, our software works on Geowalls and even "vanilla" desktop computers, but uses more involved user interfaces to interact with and evaluate 3-D data using only mouse and keyboard. Still, a recent informal user study shows that the software is useful for its purpose even in this most limited environment (BILLEN *et al*., 2006).

An immersive visualization system is ideal for Earth scientists: Earth processes are intrinsically complex; nonlinear systems and critical phenomena associated with earthquake simulation alone typically span more than six orders of magnitude in spatial scales with abrupt variations in behavior both in space and through time. For example, the deformation during an earthquake takes place on a distinctly human scale of time and space: ruptures of a fault can take seconds to minutes and cause shaking over a few to many kilometers. Models such as TeraShake require large-scale computing resources to simulate the shaking (OLSEN *et al*., 2006). In contrast, the interseismic deformation,

measured by geodetic methods, occurs at substantially lower rates. Crustal deformation at the intermediate to long-time scale can be modeled using numerical simulation: For example, interaction of large-scale fault systems (RUNDLE *et al.*, 2006) generates sequences of slip events over time, while simulations of damage in the crust (e.g., MANAKER *et al.*, 2006) generate a full stress and strain-rate field for the modeled system. The entire earthquake process is driven by plate motion, which takes place over millions of years and thousands of kilometers. Simulations hold some hope of providing insight into how these processes are linked.

Current observations and models push the limits of available interpretive methods. Yet new, larger observational data sets are rapidly becoming available, providing the opportunity to significantly advance our understanding of how the Earth works. In such a data-rich environment, rapid advances in knowledge are commonly limited by ideas rather than information. New modeling techniques are poised to provide the means to interpret these data, when coupled to increases in computational power and efficiency that have been gleaned using advanced IT methods. Visualization is already used to convey knowledge obtained from data and models from the scientific and engineering community to the general public. Use of fully immersive 3-D visualization is beginning to substantially change our perspective of these datasets and models in much the same way that going from presenting data as still images to movies fundamentally changed our scientific focus from the characterization of static distributions of parameters to understanding the dynamics of how distributions change.

In this paper, we describe an interdisciplinary approach to exploring earthquake-related geoscience data using an immersive, 3-D visualization and data manipulation environment. The work is motivated by the need to understand specific scientific problems that span many orders of magnitude in space and time, from millimeters to thousands of kilometers, and from seconds to billions of years. The three investigations described here include using VR as a tool for mapping geologic structures in remote, inaccessible locations, using immersive visualization to construct models of a subducting slab from earthquake locations in preparation for a full dynamical simulation, and using Tripod-based Light Detection and Ranging (T-LiDAR) to model structures and extract time-sequences of deformation. These applications are linked by the need to understand the multidimensional evolution of complicated boundaries interacting with larger and smaller systems.

3. VR in Neotectonic Mapping and Interpretation of Earth's Structure

Neotectonic geologists use field-based observations, digital elevation data and multi-spectral satellite or photographic imagery to record, measure, and reconstruct geologic structures such as faults and folds from deformed geomorphic features such as stream channels, abandoned shorelines, fluvial and marine terraces, or abandoned alluvial fan surfaces. Such reconstructions of geologic features are used to interpret present and past

deformation of the Earth's crust as it responds to the forces of plate tectonics and is modified by processes of erosion and deposition. The recent availability of almost global coverage of intermediate (10–90 m) and high (1–10 m: U.S. Geological Survey EROS data center: http://seamless.usgs.gov/) resolution digital elevation and imagery data has created new opportunities to study regions of the world inaccessible to the neotectonic geologist due to the scale of the structures of interest (e.g., thousands of kilometers long), or due to the remoteness of the locality (e.g., the Tibetan Plateau, the ocean floor, or another planet). At the same time this wave of new data poses a formidable visualization challenge.

The goal for the neotectonic geologists working with digital terrain data sets is to use remote-sensing data to observe and measure the detailed features (10–100 m long) that attest to active deformation of the Earth's surface over areas spanning an entire continental collision zone or active plate margin ($\sim 1 \times 10^5$ to 1×10^6 km^2). We therefore require a highly efficient, yet sensitive system to enhance analysis and interpretation of data collected through field mapping — an experience that normally includes viewing the region of interest from many perspectives and at different scales, detailed study and analysis of focus regions, and direct measurement of the location and orientations of often complex planar and undulatory 3-D structures, defined solely by their intersection with the 3-D surface topography.

A group of us (Bernardin *et al*., 2006) developed the *Real-time, Interactive Mapping System* (RIMS) to allow geologists to visualize and map structures in an intuitive and natural 3-D space (Fig. 1). RIMS provides interactive, textured height field-rendering capability for very large terrain data sets (tens of gigabytes) with full roaming and viewpoint manipulation and mapping of attributed points, polylines and polygons directly onto the terrain model. RIMS renders terrain data employing view-dependent, level-of-detail (LOD) and out-of-core data management techniques, using preprocessed quadtrees of the elevation and texture data. Google Earth is a similar tool that uses variable resolution and has been used, for example, by USGS scientists to provide a "virtual tour" of the 1906 San Francisco Earthquake (USGS, 2006). In contrast to Google Earth and other such software, RIMS is unique in its ability to provide users with a tool for efficiently mapping and directly measuring structure on the 3-D terrain models. In particular, users can attribute and then edit geo-referenced mapping elements using points and polylines, measure the orientation of surfaces such as bedding or folded alluvial fan surfaces using a virtual compass (an adjustable plane that tracks its own orientation with respect to geographic north and dip angle), and generate interpolated surfaces to facilitate geometric reconstructions using deformable surfaces. While these tools were developed with the terrestrial Earth scientist in mind, RIMS can also be used to explore high-resolution seafloor-bathymetry data and has potential applications for planetary geology data.

To interpret the 3-D geometry of a surface (for example, a folded sedimentary layer) based on its intersection with the surface of the Earth, geologists traditionally construct many 2-D cross sections along vertical slices through their map data, projecting the data

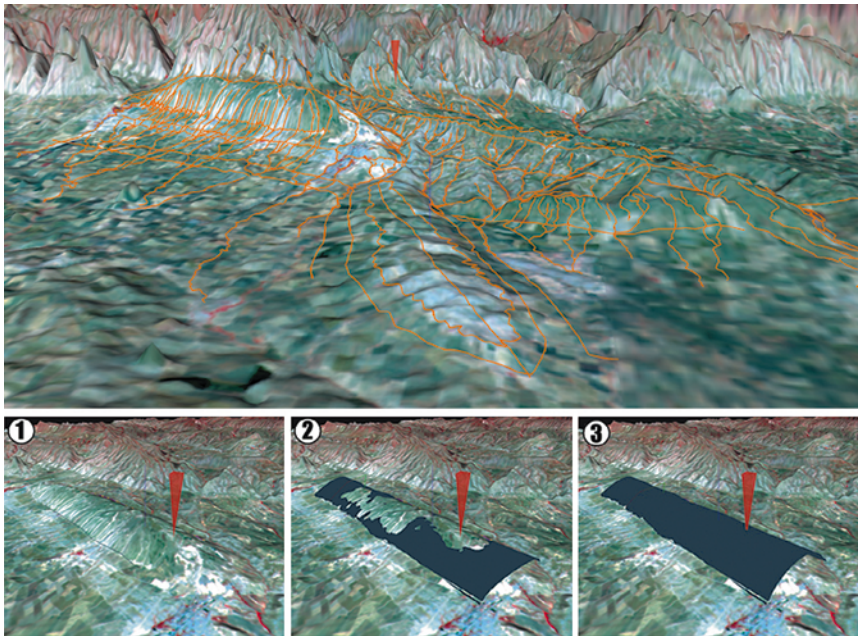


Figure 1
Top: Using the 3-D view of RIMS with a twofold vertically exaggerated textured DEM, structures were more easily identified and could be directly mapped. Bottom: Interpreting the 3-D geometry of a surface (a fold). Modified after Bernardin *et al*. (2006).

into the plane of the cross section. Using RIMS, the geologists can generate and manipulate curved or planar surfaces through their mapping of the structure at the surface, and thus display surfaces that would otherwise only be mentally imaged. Curved surfaces are typically used to fit to structures such as folds, while planes can be used to fit a fault by matching the plane to an exposed scarp and rotating it into the correct strike and dip. The resulting strike and dip can then be recorded. A measuring tool enables quantitative measurement of, for example, the height of a fault scarp or other features. Such structure-matching tools provide quantitative constraints on the minimum amplitude of the fold, which in turn can be interpreted physically as the minimum amount of deformation (shortening) in the region of the fold. Figure 1 illustrates this process in action. The bottom row of images shows the steps involved in measuring the shape of a fold. Image (1) shows the intersection of a distinct layer of rock mapped along opposite sides of a ridge. In image (2), an automatically generated reconstructed surface intersects topography where the layer is no longer present because it has been removed by erosion of the crest of the fold, indicating that this initial model is in error and that the

fold amplitude must be larger than the reconstructed surface. To correct this, the surface is adjusted manually (3) to appropriately represent the amplitude of the fold.

4. Volume Visualization in Geodynamics Models

Geodynamical modeling (TACKLEY, 2000; BILLEN *et. al*., 2003; MCNAMARA and ZHONG, 2005), seismological models (e.g., ROMANOWICZ 1991) and geodetic observations (e.g., RUNDLE *et al*., 2002) all generate large, multidimensional data sets that require analysis and interpretation. Carrying out high-resolution numerical models of geodynamics in 3-D presents a number of technical and scientific challenges that require numerical methods capable of handling extreme changes in rheology, vector and tensor data such as velocities and stress, and development and evolution of complex, heterogeneous structures. Moreover, we do not typically know the initial or boundary conditions (stress state, temperature) within the Earth, yet computer simulations are typically solving initial and boundary value problems that require specifying both *a priori*. A common approach is to carry out repeated simulations using different starting conditions to determine how sensitive results are to variations in the initial and boundary conditions. It would therefore increase the efficiency of the modeling process if we could use interactive tools to rapidly generate and evaluate starting models from which to run each geodynamical calculation. Furthermore, during a calculation or in post-processing analysis, we typically need to track a deforming interface to understand the progress of a calculation. Interactive feature extraction tools allow the measurements of specific features, such as the morphology of deformed surfaces, to facilitate comparison with seismic models and other geophysical data sets.

Many subduction zone plate boundaries (the location of many great earthquakes) are characterized by geometry of the subducted plate that varies both along strike and with depth (TASSARA *et al*., 2006; MILLER and KENNETT, 2006; MILLER *et al*., 2006). Processes such as surface deformation in the overriding plate and flow patterns in the underlying mantle may be sensitive to the 3-D geometry of the slab, particularly where the slab dip is shallow in the first process, or where the slab contains a cusp or edge in the latter process (FISCHER *et al*., 2000; BILLEN *et. al*., 2003]. Thus, adequately representing the 3-D shape of the plates at plate boundaries is likely important to understanding the processes of the govern deformation in these regions of the Earth. However, rendering smooth input for 3-D finite-element models while maintaining the complexity of the geological system under study poses a challenge. VR greatly improves the efficiency with which input for geodynamic models based on complex shapes can be generated and refined as well as the efficiency with which model output can be visualized and understood.

As an example, here we show how the 3-D visualization software is used to visualize input for a 3-D FEM model of a subduction zone, where the geometry of the subducted plate changes along the length of the subduction zone. In the model, the initial shape of the subducted plate is based on seismic observations (PAGE *et al*., 1989; BROCHER *et al*.,

1994; GUDMUNDSSON and SAMBRIDGE, 1998; DOSER *et al.*, 1999; RATCHKOVSKI and HANSEN, 2002). Superposition of the seismic data and the smoothed slab surface allows evaluating the fit between the data and the idealized, yet complex, initial slab shape (Figs. 2a and 2b). We use "Visualizer," an interactive volume visualization software for immersive and nonimmersive environments developed at KeckCAVES (BILLEN *et al.*, 2006), to analyze the initial temperature and viscosity fields associated with the FEM model by extracting isosurfaces and 2-D slices of these fields (Figs. 2c and 2d). Visualizer's interactivity enables rapid assessment of the character and quality of the input data; these illustrations apply equally to the assessment and visualization of the model output as well.

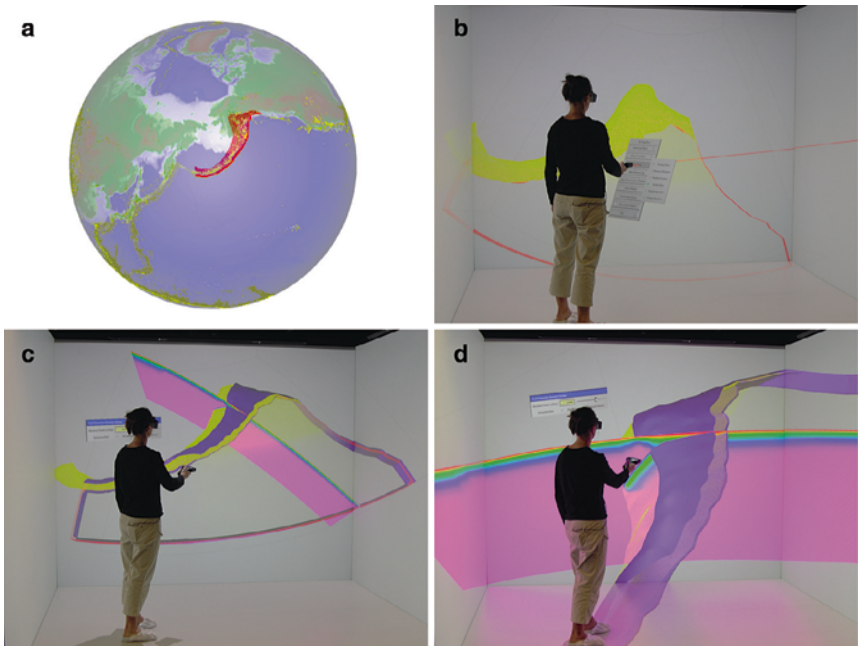


Figure 2

Constructing, viewing, and refining a model of a slab (adapted from the work of JADAMEC and BILLEN , 2006). (a) Global earthquake distribution in a transparent globe. (b) Surface through the Benioff zone (yellow) constructed from seismic data marks the top of the subducting plate (see text for data references). 3-D menus and dialogs are used to assign interactive tools. (c) Superposition of multiple data fields from the FEM model of the subduction zone in the CAVE. An arbitrary slice through the initial thermal field (rainbow color) is generated using the slicer tool. The isosurface tool generates an isosurface of constant viscosity (purple) along the plate interface and model edges. (d) Data are easily manipulated in the CAVE as shown here in the rotated and zoomed in view from (c). The slice through the initial thermal field (rainbow colors) illustrates the Pacific plate descending beneath the North American plate. The isosurface of constant viscosity (purple) delineates the plate interface. This isosurface enables evaluation of the cusp in the slab (upper right of the figure) where the dipping subduction zone joins at a near right angle with the vertical transform plate boundary.

In a recent study designed to assess the effectiveness of visualization tools in geodynamics applications (BILLEN *et al*., 2006) we asked a group of researchers and students to search for specific features in a model of a subducting slab using two different visualization methods on two different platforms. We compared Visualizer to the commercial TecPlot visualization package [http://www.tecplot.com]. TecPlot was used on a desktop system, while Visualizer was used both on a desktop and in a CAVE. Using Visualizer on both platforms allowed us to assess the experience of immersive visualization independent of the software. The users evaluated a model prepared as initial conditions for the finite-element program CitcomT (ZHONG *et al*., 1998; BILLEN *et al*., 2003), a widely used code designed for mantle convection models (e.g., MORESI and PARSON, 1995).

The user study showed that Visualizer, used in the CAVE, made data exploration (navigating, identifying and locating features) the easiest and was the easiest to learn and use overall. Visualizer used on the desktop also made data exploration easier than TecPlot, although users found Visualizer more difficult to learn initially. A key feature of Visualizer is the ability of users to create seeded slices and surfaces located and oriented arbitrarily using a handheld input device, in contrast to most commonly available software packages, in which users must typically decide *a priori* where to place a slice or an isosurface, often by entering numerical values. The interactivity provided by Visualizer is a more effective means of exploring data when the location and characteristics of features of interest are yet to be discovered. Movies showing a user creating slices and isosurfaces with Visualizer using geophysical data sets are available on the KeckCAVES project web site http://www.keckcaves.org/movies.

5. Tripod-LiDAR

Light Detection and Ranging measurements taken from the ground using a tripod (T-LiDAR) provide the capability to rapidly acquire ultra high-resolution surface (sub-millimeter) data on the outcrop scale (1 m^2 to 10 km^2), complementing space-based and airborne geodetic measurements including data acquisition underway and being planning by EarthScope [http://www.earthscope.org] and GeoEarthScope. T-LiDAR has impressive potential as a tool for neotectonics, quantitative geomorphology, and geological hazards assessment by allowing rapid, high-resolution measurements of a changing landscape (BAWDEN *et al*. 2004). A T-LiDAR instrument bounces light off surfaces, recording a "data cloud" of points that together make a 3-D image of the landscape. A complete view is created by collecting data from multiple directions. This new technology rapidly generates millions of scattered points from which features must be extracted and interpreted, but the development of analytical software to manipulate and analyze the data lags behind the hardware technology. We have developed a VR LiDAR viewer, which rapidly displays and allows navigation through large T-LiDAR data sets. Thus for example, the user can view the landscape from a perspective that would not be obtainable

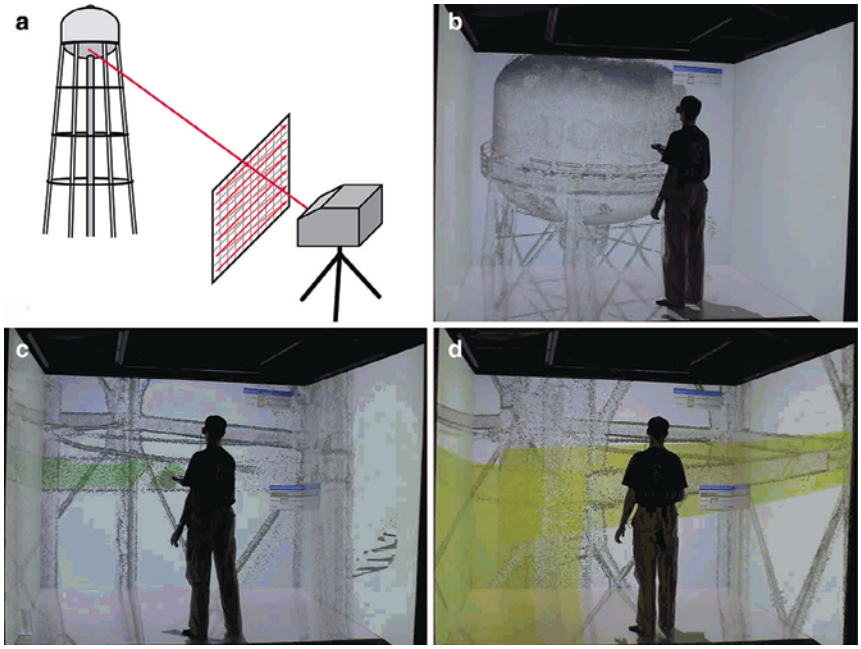


Figure 3
Working with tripod LiDAR (T-LiDAR) data in an immersive interactive visualization environment. (a) Schematic of tripod LiDAR data acquisition on, in this example, an engineered structure. (b) T-LiDAR scan of a watertower on the UC Davis campus: The user can take an otherwise unobtainable perspective next to the structure. (c) A tool allows the user to select parts of the structure (here, a beam). The selection tool is represented by the sphere at the end of the hand-held wand (in the user's right hand). The selected points are green. (d) The final step in creating a model involves fitting a geometric function (a plane) to the selected points. Here two planes (yellow) have been fitted to points selected on two beams.

in the field. By adapting the feature extraction tools described above in the context of the RIMS software package to LiDAR data, we enable the user to rapidly identify features, select sets of points that are of interest, such as the landscape being mapped (Fig. 3) while leaving behind irrelevant points (such as vegetation covering the landscape).

Feature extraction allows the user to fit geometric objects to a subset of the data and to make quantitative measurements. For example, fitting a plane to a wall, or fitting cylinders to fenceposts, within a sequence of LiDAR images taken over a period of time allows the user to determine offset and strain of a structure that crosses a fault plane. Thus, repeated T-LiDAR scans following an earthquake can be used to understand the 4-D deformation field. Fitting best-fit surfaces to select features in the T-LiDAR imagery (planes to building walls, vectors to fence post, cylinders to posts and bridge supports, etc.) improves the position accuracy of the target feature and provides a unique method for tracking how features move in 3-D space over time (Fig. 3). A movie showing a user

exploring and analyzing a sample data set, a high-resolution tripod LiDAR laser scan of part of the UC Davis campus, using the multiresolution point set visualization program, is available on http://www.keckcaves.org/movies. The visualized point set contains about 4.7 million 3-D points with intensity values that randomly sample all surfaces in the scanned area. The program uses a 3-D paintbrush interface, a sphere attached to the user's hand, to allow a user to select subsets of points and determine low-degree analytically defined approximations such as planes, cylinders, spheres, etc.

6. Conclusions

We have discussed how VR technology can benefit diverse geological applications for characterizing faults and simulating earthquakes by providing intuitive interactive visualization and analysis tools that focus on exploration and human interaction with data to make scientific progress on all aspects of earthquake simulation, from identifying seismogenic faults and their properties to constructing fully dynamical models. By moving beyond static pictures, and instead giving users the ability to rapidly create 3-D pictures showing features of interest, and then to evaluate and measure those features, we take advantage of the skills of geoscientists, and also more fully exploit the capabilities of VR environments. Previous approaches have mostly tried adapting available desktop-based visualization software to VR, but, due to those realms' very different constraints, have mostly not led to effective analysis tools. Conversely, we found out that software developed from the ground up to focus on the strengths of VR, i.e., 3-D perception and intuitive interaction, can also work well in a wide range of nonimmersive environments, including desktops, and even compete with native desktop software. The software's portability is enabled by an underlying VR "operating system" (KREYLOS, 2006) that, unlike previous approaches, hides vastly different capabilities in display hardware and, more importantly, input device hardware. As a side effect, this portability creates a way for researchers already owning a lower-end visualization system to scale up in a cost-effective and incremental manner: The software users already knowing becomes more effective as new components, such as tracking systems, are added to the system. The value of immersive, interactive data exploration is growing importantly with the explosion of large data sets created by imaging, large observational efforts, and high-resolution computer simulations. The ability to rapidly create complex objects for use in models allows the use of more realistic boundary conditions and objects in Earth science modeling. One of the most difficult aspects of developing forward models and simulations of Earth science processes is identifying the spatial distribution of critical behaviors and the temporal framework of changes. Proper resolution is critical to modeling realistic behaviors in fault zones. An ability to interactively adjust critical parameters in 3-D models substantially increases the appropriateness of boundary conditions during model development, promoting rapid advances in model sophistication, accuracy, and relevance to earthquake processes.

Acknowledgements

This work was supported by the W. M. Keck Foundation and the University of California, Davis. We thank the members of the Visualization and Computer Graphics Research Group of the UC Davis Institute for Data Analysis and Visualization.

References

Bawden, G. W., Kayen, R. E., Silver, M. H., Brandt, J. T., and Collins, B. (2004), *Evaluating Tripod Lidar as an earthquake response tool*, EOS, Trans. AGU, Fall Meeting 2004, 2004AGUFM.S51C0170B.

Bernardin, T., Cowgill, E., Gold, R. D., Hamann, B., Kreylos, O., and Schmitt, A. (2006), *Interactive mapping on 3D terrain models*, Geochem., Geophy., Geosystems *7*, Q10013, DOI 10.1029/2006GC001335.

Billen, M. I., Gurnis, M., and Simons, M. (2003), *Multiscale dynamic models of the Tonga- Kermadec subduction zone*., Geophys. J. Int. *153*, 359–388.

Billen, M. I., Kreylos, O., Kellogg, L. H., Hamann, B., Staadt, O., Sumner, D. Y., and Jadamec, M. (2006), *Study of 3D Visualization Software for Geo-Science Applications*, KeckCAVES Technical Report TR06–01.

Brocher, T. M., Fuis, G. S., Fisher, M. A., Plafker, G., and Moses, M. J. (1994), *Mapping the megathrust beneath the northern Gulf of Alaska using wide-angle seismic data*, J. Geophy. Res. *99*, 11,663–11,685.

Butler, D. (2006), *Virtual globes: The web-wide world*, Nature *439*, 776–778, doi:10.1038/439776a.

Cabral, B., Cam, N., and Foran, J., *Accelerated volume rendering and tomographic reconstruction using texture mapping hardware*. In: *Proc. 1994 Symp. Volume Visualization* (ACM Press, New York, New York, 1994), pp. 91–98.

Carlson, W. D. (2006), *Three-dimensional imaging of earth and planetary materials*, Earth Planet. Sci. Lett. *249*, 133–147.

Cohen, R. E., ed., *High-Performance Computing Requirements for the Computational Solid Earth Sciences*, 96 pp, (http://www.geo-prose.com/computational_SES.html, 2005).

Doser, D. I., Veilleux, A. M., and Velasquez, M. (1999), *Seismicity of the Prince William Sound region for over thirty years following the 1964 Great Alaskan Earthquake*, Pure Appl. Geophys. *154*, 593–632.

Erlebacher, G., Yuen, D. A., and Dubuffet, F. (2001), *Current trends and demands in visualization in the geosciences*, Electronic Geosciences, ISSN 1436–2511, DOI 10.1007/s10069–001-1019-y.

Fischer, K. M., Parmentier, E. M., and Stine, A. R. (2000), *Modeling anisotropy and plate-driven flow in the Tonga subduction zone back arc*, J. Geophys. Res. *105*(B7), 16181–16191.

Foster, I. (2006), *2020 Computing: A two-way street to science's future*, Nature *440*, 419, doi:10.1038/440419a.

Gudmundsson, O. and Sambridge, M. (1998) *A regionalized upper mantle (RUM) seismic model*, J. Geophys. Res. *103*, 7121–7136.

Jadamec, M. and Billen, M. I. (2006), *Influence of slab geometry on diffuse plate boundary deformation: 3D numerical models of the plate boundary corner in southern Alaska*, Eos Trans. AGU *87*(52), Fall Meet. Suppl., Abstract T23B–0491.

Kadlec, B. J., Yuen, D. A., and Erlebacher, G. (2006), *Visualization and analysis of multi-terabyte geophysical data sets in an interactive setting with remote webcam capabilities*, Pure Appl. Geophys. (still in press).

Kreylos, O. (2006), *Environment-Independent VR Development*, KeckCAVES Technical Report TR06–03 pp.

Kreylos, O., Bawden, G. W., Bernardin, T., Billen, M. I., Cowgill, E. S., Gold, R. D., Hamann, B., Jadamec, M., Kellogg, L. H., Staadt, O. G., and Sumner, D. Y. (2006), *Enabling scientific workflows in virtual reality*. In: (Hong Wong, K., Baciu, G. and Bao, H., eds.), Proc. ACM SIGGRAPH Interna. Conf. Virtual Reality Continuum and Its Applications, 2006 (VRCIA 2006) (ACM Press, New York, New York, 2006) pp. 155–162.

LaMar, E. C., Hamann, B., and Joy, K. I. (1999), *Multiresolution techniques for interactive texture-based volume visualization*. In: (Ebert, D.S., Gross, M. and Hamann, B., eds.), IEEE Visualization '99, IEEE Comp. Soc. Press, Los Alamitos, California, pp. 355–361.

Mankar et al. (2006).

MAX, N. (1995), *Optical models for direct volume rendering*. In: IEEE Transactions on Visualization and Computer Graphics *1*, 99–108.

MILLER, M. S. and KENNETT, B. L. N. (2006), *Evolution of mantle structure beneath the northwest Pacific: Evidence from seismic tomography and paleogeographic reconstructions*, Tectonics *25*, TC4002.

MILLER, M. S., GORBATOV, A. and KENNETT, B. L. N. (2006), *Three-dimensional visualization of a near-vertical slab tear beneath the southern Mariana arc*, Geochem. Geophys. Geosystems *7*, Q06012.

MCNAMARA, A. and S. ZHONG (2005), *Thermochemical structures beneath Africa and the Pacific*, Nature *437*, 1136–1139.

MORESI, L. and PARSONS, B. (1995), *Interpreting gravity, geoid, and topography for convection with temperature dependent viscosity: Application to surface features on Venus*, J. Geophys. Res. *100*, 21155–21171.

NOURBAKHSH, I., SARGENT, R., WRIGHT, A., CRAMER, K., MCCLENDON, B. and JONES, M. (2006), *Mapping disaster zones*, Nature *439*, 787–788, doi:10.1038/439787a.

OLSEN, K. B., DAY, S. M., MINSTER, J. B., CUI, Y., CHOURASIA, A., FAERMAN, M., MOORE, R., MAECHLING, P., and JORDAN T. (2006), S*trong shaking in Los Angeles expected from southern San Andreas earthquake*, *Geophys. Res. Lett. 33*, L07305, doi:10.1029/2005GL025472.

PAGE, R. A., STEPHENS, C. D., and LAHR, J. C. (1989), *Seismicity of the Wrangell and Aleutian Wadati-Benioff Zones and the North American Plate along the trans-Alaska crustal transect, Chugach Mountains and CopperRiver Basin, Southern Alaska*, J. Geophys. Res. *94*, 16,059–16,082.

RATCHKOVSKI, N. A. and HANSEN, R. A. (2002), *New evidence for segmentation of the Alaska Subduction zone*, Bull. Seismol. Soc. Am. *92*, 1754–1765.

ROMANOWICZ, B. (1991), *Seismic tomography of the Earth's mantle*, Ann. Rev. Earth and Planet. Sci. *19*, 77–99.

RUNDLE, J. B., RUNDLE, P. B., DONNELLAN, A., Li, P., KLEIN, W., MOREIN, G., TURCOTTE D. L., and GRANT. L. (2006), *Stress transfer in earthquakes, hazard estimation and ensemble forecasting: Inferences from numerical simulations*, Tectonophysics *413*, 109–125.

RUNDLE , J. B., RUNDLE, P. B., KLEIN, W., DE SA MARTINS, J., TIAMPO, K. F., DONNELLAN A., and KELLOGG L. H. (2002), *GEM plate boundary simulations for the Plate Boundary Observatory: A program for understanding the physics of earthquakes on complex fault networks via observations, theory and numerical simulation*, Pure Appl. Geophy. *159*, 2357–2381.

SABELLA, P. (1988), *A rendering algorithm for visualizing 3D scalar fields*. In: (Dill, J., ed.) *Computer Graphics* (Proce. ACM SIGGRAPH 88) *22*(4), pp. 51–58.

STEREO ANALYST for ARCGIS. http://gis.leica-geosystems.com/Products/StereoAnalyst/.

TACKLEY, P. (2000), *Mantle convection and plate tectonics: Toward an integrated physical and chemical theory*, Science *288*, 2002–2007.

TASSARA, A., GOTZE, H. J., SCHMIDT S,, and HACKNEY R (2006) Three-dimensional density model of the Nazca plate and the Andean continental margin, J. Geophys. Res. *111*, B09404.

TUFTE, E., The Visual Display of Quantitative Information (Graphics Press, 1983).

USGS (2006), http://earthquake.usgs.gov/regional/nca/virtualtour/.

ZHONG, S., GURNIS, M. and MORESI. L. (1998), *Role of faults, nonlinear rheology, and viscosity structure in generating plates from instantaneous mantle flow models*, J. of Geophy. Res. *103* (B7), 15255–15268.

(Received November 1, 2006, revised September 9, 2007, accepted September 25, 2007)

Pure appl. geophys. 165 (2008) 635–651
0033–4553/08/030635–17
DOI 10.1007/s00024-008-0319-7

Pure and Applied Geophysics

The QuakeSim Project: Web Services for Managing Geophysical Data and Applications

Marlon E. Pierce,[1] Geoffrey C. Fox,[1,2] Mehmet S. Aktas,[1,2] Galip Aydin,[1,2] Harshawardhan Gadgil,[1,2] Zhigang Qi,[1,2] and Ahmet Sayar[1,2]

Abstract—We describe our distributed systems research efforts to build the "cyberinfrastructure" components that constitute a geophysical Grid, or more accurately, a Grid of Grids. Service-oriented computing principles are used to build a distributed infrastructure of Web accessible components for accessing data and scientific applications. Our data services fall into two major categories: Archival, database-backed services based around Geographical Information System (GIS) standards from the Open Geospatial Consortium, and streaming services that can be used to filter and route real-time data sources such as Global Positioning System data streams. Execution support services include application execution management services and services for transferring remote files. These data and execution service families are bound together through metadata information and workflow services for service orchestration. Users may access the system through the QuakeSim scientific Web portal, which is built using a portlet component approach.

Key words: Cyberinfrastructure, Web Services, Grid computing, geographical information systems, science portals, Workflow.

1. Introduction: Cyberinfrastructure for the Geosciences

The danger to human life and property posed by earthquakes has been demonstrated many times in the last several years. Earthquakes in Bam, Iran in December 2003, Sumatra in December 2004, and Kashmir in October 2005 all resulted in tens of thousands of deaths. The scale of these tragedies amply demonstrates the importance of understanding the immediate and long-term causes and behaviors of interacting earthquake fault systems. Death tolls for recent large earthquakes in developed countries such as the United States and Japan have not been as catastrophic, but the economic cost of events such as the 1994 Northridge, California earthquake reinforces the need for increased understanding.

[1] Community Grids Laboratory, Indiana University, 501 North Morton Stret, Bloomington, IN 47404, U.S.A. E-mail: (marpierc, gcf, maktas, gaydin, hgaolgi, zqi, asayar)@indiana.edu

[2] Department of Computer Science, Indiana University, Lindley Hall, Room 215, 150 S. Woodlawn Ave., Bloomington, IN 47405-7104, U.S.A.

Geophysicists have developed a range of tools for investigating earthquake fault systems. Techniques include: a) Data mining techniques for analyzing time series data produced by sensors such as Global Positioning Systems; b) modeling codes that theoretically simulate the effects of the stresses associated with faults; c) data assimilation codes that attempt to find best-fit models to historical data records, with the best-fitting models then used to forecast areas of increased hazard; and d) large-scale probabilistic simulations that can model the behavior of extensive interacting fault systems such as the San Andreas fault system and its neighbors.

Such tools, which are typically driven by observational data inputs, are obviously excellent candidates for incorporating into a Grid (FOSTER and KESSELMAN, 2004) for geophysicists. A Grid, in brief summary, is a collection of Web Services that provide capabilities such as securely accessing supercomputers (and their scheduling systems); accessing, moving, and replicating data resources; accessing information services that describe available resources; accessing databases; services for accessing specific scientific applications; and so on. For our purposes here, we need only specify that a Grid is a collection of Web Services (ATKINSON *et al.*, 2005) that have precise, XML-defined remote programming interfaces (typically WSDL, CHRISTENSEN *et al.*, 2001) describing how to interact with the service, and that they communicate over network connections with XML messages (SOAP, GUDGIN *et al.*, 2003). Reliability of messages, security, service discovery, service information models, management of distributed session state, message communication patterns, and other distributed computing topics are important, although for our purposes not fundamental. Building-persistent Grids with managed resources (supercomputers and high-speed networks) with a consistent administration is a large task that has been undertaken by several countries. In the United States, TeraGrid and the Open Science Grid are two prominent examples. Other large Grid efforts are underway in Australia, China, the United Kingdom, the European Union, and other countries. For a survey of these international efforts, see the Open Grid Forum web site (www.ogf.org).

Our project, QuakeSim, is developing the applications, the distributed computing infrastructure, the data management tools, and the Web portal user interfaces, for building this cyberinfrastructure (ATKINS *et al.*, 2003). QuakeSim is led by the NASA Jet Propulsion Laboratory and includes collaborators at the University of California-Davis, the University of Southern California, the University of California-Irvine, and Indiana University.

As with many other Grid projects, we began QuakeSim with a focus on services to support computational applications, but as the project has evolved, we have realized that providing programming interfaces to remote data sources (both real time and archival) through Web Services is at least as important as managing the execution of remote applications. We thus consider QuakeSim to be an example of a "Grid of Grids" (FOX *et al.*, 2005a) in which we must integrate "Data Grid" and "Execution Grid" services (which are themselves independently useful) into a larger, cooperating system. In a Grid of Grids, services drawn from Data Grid and Execution Grid families must be integrated

into composite applications. The development of tools for managing these distributed services is another important area of distributed systems research. An important subset of Grid management systems, known as Grid Workflow (GANNON, 2005), shepherds the execution of a composite application across several different, distributed services. Information management of Grids also falls with in the service management category.

This paper summarizes the distributed computing systems research efforts of the QuakeSim project. In Section 2 we provide a summary of requirements and describe a general Web Service architecture that can meet these requirements. We next summarize the Execution Grid and Geographical Information System Data Grid components that we have developed in Sections 3 and 4. User interface development using Web portals is an important component of our work and is summarized in Section 5. Service information and execution management research efforts are next discussed in Section 6. Approaches for monitoring deployed science portals and their supporting services are described in Section 7. Finally, open service architectures for science Grids imply broader community development models that have become pervasive in the current Internet but have not yet been fully exploited in scientific Grids. We conclude with a discussion of these issues in Section 8.

2. *Applications, Data Requirements, and Architecture*

To understand the QuakeSim architecture, we first consider some of its geophysical application components and their data requirements. These codes are typically initially developed on desktops and workstations and then may in some cases be parallelized to run on clusters and supercomputers to solve problems of increasing size and complexity.

- **Disloc** models multiple dipping dislocations (faults) in an elastic half-space (OKADA, 1985). It uses fault models (such as may be obtained from the QuakeTables database, described below) as input.
- **Simplex** is an inversion code based on Disloc. Using observational surface stress values and an initial geometric model for one or more causative faults, Simplex can determine optimum fault model parameters.
- **GeoFEST** (PARKER *et al*., 2005) is a three-dimensional viscoelastic finite-element model for calculating nodal displacements and tractions. It allows for realistic fault geometry and characteristics, material properties, and body forces. GeoFEST relies upon fault models with geometric and material properties as input. These geometric models are converted into finite-element meshes through supporting finite-element mesh refining programs.
- **Virtual California** (RUNDLE *et al*., 2002) is a program to simulate interactions between vertical strike-slip faults using an elastic layer over a viscoelastic half-space. Virtual California relies upon fault and fault friction models as input.

- **Pattern Informatics** (TIAMPO *et al.*, 2002) calculates regions of enhanced probability for future seismic activity based on the seismic record of the region. Pattern Informatics uses seismic data archives as input.
- **Regularized Deterministic Annealing Hidden Markov Model (RDAHMM)** (GRANAT, 2004) is a time-series analysis program that produces feature vectors and probabilities for transitioning from one (abstract) class to another. RDAHMM is typically used to analyze GPS and seismic catalog archives, however it can be applied to any time series data.

As can be seen from the above summary, QuakeSim applications span a large range of computational geophysical research and rely heavily on observational input data.

2.1. Architecture

QuakeSim is implemented as a collection of Web Services for accessing data sources, execution codes, visualization and mapping applications, and other tools. User interfaces to these services are implemented as Java Server Pages wrapped as portlets (PIERCE *et al.*, 2002a; ABDELNUR *et al.*, 2003), which are in turn aggregated into a central portal. Figure 1 depicts the basic system architecture.

As we describe in more detail in subsequent sections, we have built a specific set of Web Services that can be adapted to manage the various applications and their data sources. Figure 1 illustrates several sample interactions: The QuakeSim portal mediates the user's interactions with remote servers, and remote services must also communicate with each other. Figure 1 also indicates a simple Grid workflow, which may be managed by the portal or through a separate workflow engine. As we describe below, we use Apache Ant-based Web Services for managing various local (to a specific host) applications. We manage distributed system state, such as is required to step through several tasks on different hosts, through the Context Manager Service, described below. This allows complicated tasks to run independently of the user's portal session: Users can receive updates of their project's progress when they log back into the portal later.

As is more thoroughly discussed in Section 6, we chose the portlet approach for building our Web portal. This had the primary advantage of allowing us to plug other portlets into our system. Thus it is not difficult to create a user interface that combines portlets to our Web Service Grid components as well as portlets to various Globus Toolkit services, collaboration tools, news and information services, and so forth.

3. Execution Grid Services

QuakeSim is based on the so-called "WS-I+" approach to building Web Service-based Grid systems (ATKINSON *et al.*, 2005). We have implemented all services using standard Web Service specifications for WSDL and SOAP and have introduced

no nonstandard extensions to these specifications. Our initial focus in QuakeSim was to support and develop application management tools for the codes listed in the previous section. All services are implemented in Java and use the Apache Axis Web Service engine, although these implementation details are not needed to invoke the services.

Application Descriptor Web Services. Describing applications and hosts with a data model (or, more ambitiously, an ontology) has been part of the design of many projects. Our approach to this problem is to use linked, independent XML schemas for separately describing the hosts and applications (PIERCE *et al*., 2002b; YOUN *et al*., 2003). The XML schemas are designed primarily to simplify the management of user interfaces to QuakeSim applications. We thus focused on modeling the codes' execution details, such as number and type of input and output parameters. These descriptions of metadata can then be used to build portlet interfaces for managing codes that are available to portal users.

Our primary usage of the XML data models is in constructing job descriptions for remotely executing the applications using the complementary execution management services described below. We have not attempted to develop a taxonomy or ontological description of the codes or to describe the class of problems a particular code may be used to solve. To address this latter problem, we have explored extensions of our application and host modeling using Semantic Web and Case-Based Reasoning techniques, as described in AKTAS *et al*., (2004).

Execution Management Services. In our design of execution services, we realized that the Apache Ant project offered many powerful capabilities that could be adapted to the problem of running executables. It already implements numerous tasks useful for interacting with the file system, including tasks for executing programs and for creating, renaming, and deleting files and directories. These tasks can be arranged to express dependencies and simple control logic, making it an ideal tool for handling the simple pre- and post-processing jobs as well as running the executable. In QuakeSim, we use this service to wrap the applications described in Section 2. We have also found it useful to use this service to wrap such tools as IDL, Matlab, and GMT to provide simple visualization. Ant is written in Java and is open source, so it is a simple matter to build an Apache Axis-based Web Service with it.

Ant also offers a simple extension mechanism for adding custom tasks. We use this capability to define tasks that are wrappers to Web Service clients for file transfer and context management services. Such tasks can be executed as part of the Ant dependency chain to pull in files from remote hosts and to call back state changes to the portal in long-running applications. This allows us to implement the basic backend-to-backend communications shown in Figure 1. Ant tasks also allow us to define simple workflows and job sequences, and (unlike basic Globus GRAM services) give us some control of the

user's actions on the server side; i.e., the user only gets access to specific tasks and cannot run arbitrary commands on the server. This is enforced on the Web Service side as well as the user interface side.

To wrap a specific application, we develop an Ant *build.xml* file template. This XML file specifies all the tasks that are to be executed, their execution order, and their interdependencies. This template may be used for managing pre- and post-processing steps as well as for running the specific application. For example, we use this approach to couple GeoFEST with finite-element mesh generating applications. We may invoke the service by passing string arrays of parameter values to the template. More typically, we extend this general service to build WSDL interfaces that are specialized to specific services.

The execution service is not associated with user authentication. Users are authenticated through the portal, and the portal is responsible for keeping track of individual users's jobs and data. The codes themselves run on a generic account on the selected server.

One of the core difficulties with this approach to application management services is that the services depend directly on the robustness of the application. Unless the service is properly constrained, it is very easy for naïve users to set up potentially troublesome inputs that cannot be handled by the application's logic. A common example is input that results in the code entering an infinite loop in which error messages are written repeatedly to a file until the host computer's file system is filled. These issues must be addressed by the service implementation, which must validate input or be able to recognize and respond to runaway jobs. We have not found a general strategy to handle these problems.

File Transfer and Management Services. We have implemented file transfer Web Services for moving data between various backend machines (see Fig. 1). These services can also be used to upload and download data between the user's desktop and the backend host machines, using the portal server as an intermediary. This service may also work behind the scenes, called by the Ant execution service described above to move an output file to a GMT plotting service, for example. By coupling the file transfer service with the context management service, we have created virtual file spaces for the user: The context manager keeps track of the file location for a specific user project. These metadata can then be used to construct invocations of the file transfer service, which is responsible for actually transferring the file. Typically this is used to allow the user to directly download data or to transfer data from the execution host to a visualization service. This service has the advantage of being able to serve files that are not directly exposed through a public URL on a Web Server. However, we have more recently adopted this latter, URL-based approach as being simpler, easier to integrate with portals and workflow engines, and independent of specific Axis implementation libraries.

Context Management Services. The QuakeSim portal and its codes allow users to execute geophysical applications on a Grid, however we must go beyond this to provide a

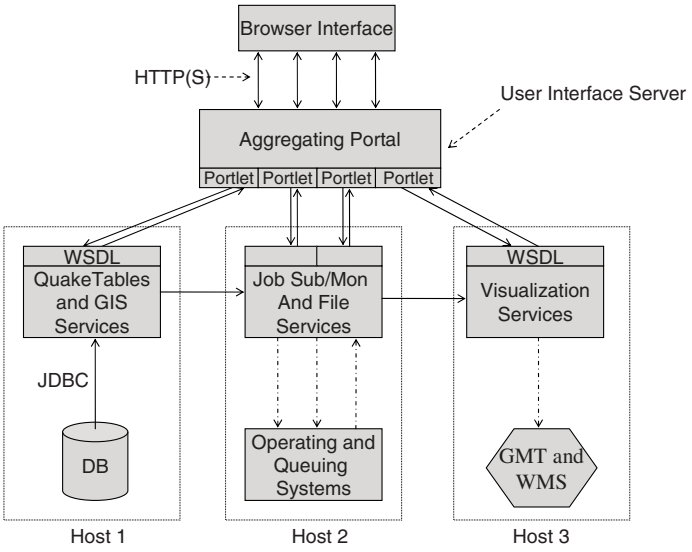


Figure 1
The QuakeSim architecture consists of distributed Web Services accessed through the QuakeSim user portal. Solid arrows represent network connections (SOAP over HTTP unless otherwise indicated). Broken arrows represent interactions with the local operating system. As shown in the figure, the portal may serve (in effect) as the state and workflow management system for the services. We also provide these as separate services through the HPSearch and WS-Context development work.

useful user experience. One way in which we address this is to provide an "electronic notebook" style user interface that lets users set up their problems in the form of projects. This project interface provides a way of maintaining user information (the codes selected, the input files used, the output generated, and so forth). The user interfaces are backed by a persistent metadata service that we call the context manager. We store user information using structured name-value pairs, which can be constructed as arbitrarily deep trees. We refer to these groups of metadata as context (PIERCE, 2002b). In our terminology, a context is an instance of a simple XML schema, and these XML instances are linked through parent-child relationships. A context has a simple, URI-like name that can be used to access its data. This service is part of the QuakeSim portal deployment, but it has been superseded in later projects by our implementation of the WS-Context specification (AKTAS, 2006a).

4. Data and Information Grid Services

4.1. Quake Tables Fault Database

The QuakeTables Web Service and Web accessible database (CHEN *et al*., 2003; GRANT *et al*., 2005) act as a data repository for earthquake fault data in the QuakeSim

project. It includes location, geometric and material characteristics, and provenance information such as the source (author, journal information, etc.) for a particular fault entry. QuakeTables provide both a human usable Web interface (wrapped as a portlet) and a WSDL-based Web Service programming interface. The former interface is useful for delivering data to human users while the latter interface is useful for delivering data to applications. Using the latter interface, we have integrated QuakeTables with execution management services for GeoFEST, Disloc, and Simplex through the QuakeSim portal.

4.2. Geographical Information System Services for Archival Data

Geographical Information System (GIS) standards may be adapted to meet many of the data and metadata requirements surveyed in Section 2. The Open Geospatial Consortium [OGC] defines an extensible (and extensive) XML data model, the Geographic Markup Language (GML) (Cox, 2003), for describing geospatial and geotemporal features. This common data model is then integrated with a number of services. The OGC defines standard service definitions and interoperability guidelines.

We have implemented the following "Information and Data Grid" Web Services:

- **Data Services:** We implemented the Web Feature Service (Vretanos, 2002) to store and retrieve seismic data archives, GPS data archives, and faults (in a GIS variant of the QuakeTables service). An OGC feature is a GML description of a map object. The description may include both information on how to render the feature as well as other useful metadata, such as the name of the feature and other observation data associated with it. More information regarding our work here is available from Aydin *et al*. (2005).
- **Map Generation Services:** We implemented the OGC's Web Map Service specification (Beaujardierre, 2004) as a Web Service. The Web Map Service is used to generate vector and raster maps in various formats (SVG, JPEG, GIF, etc.). These maps typically include realizations of abstract feature data obtained from Web Feature Services. Our Web Map Service can also integrate maps from other Web Map Servers as an overlay. This is described in more detail in Sayar *et al*. (2006).
- **Information Services:** One useful feature of the OGC service specifications is that they include a standard XML metadata description ("capabilities") and query method that can be used to describe and obtain information about what the service does. The OGC defines information services (catalogs) for aggregating capability information. We decided, however, that these specifications were too GIS specific and could be substituted with more general, UDDI-based systems. We developed an extension of UDDI along these lines to support general metadata extensions and XPath queries, with specific realizations for the OGC capabilities file. See Aktas *et al*. (2006a, b).

4.3. Real-Time Data Grids

In addition to request-response style access of archived and curated data, geophysical applications also need to support the real-time and near real-time analysis of data streams. The specific problem within the QuakeSim project is to provide the infrastructure for coupling real-time GPS data to time series analysis codes. The Scripps Orbital and Permanent Array Center maintains the California Real Time Network (CRTN) (see http://sopac.ucsd.edu/projects/realtime/), which we may directly access to obtain station position measurements at a rate of 1 Hz.

As described in Section 1, RDAHMM can be used to detect underlying mode changes in archived GPS signals. These modes, which can be determined independently of any *a priori* modeling, may be associated with underlying physical processes such as earthquakes and more subtle aseismic events. Typically, RDAHMM has been applied to much longer time scales (daily time series values over several months or years), thus the use of RDAHMM in the analysis of real-time data is an area of research. From the distributed systems research point of view, however, we must investigate the computing infrastructure that will enable real-time RDAHMM.

We have developed a real-time stream filtering infrastructure by using topic-based publish/subscribe software, NaradaBrokering (Pallickara and Fox, 2003). Our system currently supports 70 individual GPS stations in CRTN. The system works as a sequence of filters that transforms binary GPS data streams and is depicted in Figure 2. A detailed description is available from Fox *et al*. (2005b) and is summarized here. Each network source stream (which typically corresponds to 5–10 individual GPS stations) is initially published in binary RYO format. Each network's binary message is then received by an ASCII decoder, which republishes on a new subtopic. We may add further filters, such as a de-multiplexing filter that can separate the 5–10 interleaved GPS station position values in each RYO message and republish them on 5–10 different topics. This allows other filters and applications to register for individual station feeds. We have also implemented signal archive filters that can store and replay network time series on demand. Application filters for stream analysis can be built similarly to these simpler filters. We have prototyped the integration of an RDAHMM filter with the real-time data streams, as shown in Figure 3. This is a proof of concept: The system deployment is stable and persistently available, however the RDAHMM results (the modes identified in the signal) should not currently be interpreted as meaningful. We are in the process of evaluating the proper usage of RDAHMM on these time series.

Our current system is deployed and stable for the 70 stations of the Southern California network. We are in the process of a systematic series of throughput and scaling tests of our software. Our goal in throughput testing is to test the performance of the system to determine the overhead associated with simple filters under various publish and subscribe loads. Our scaling test is to determine the overall load of signals that can be supported by a single message broker and to investigate scaling strategies that use multiple brokers that behave as a single logical message broker.

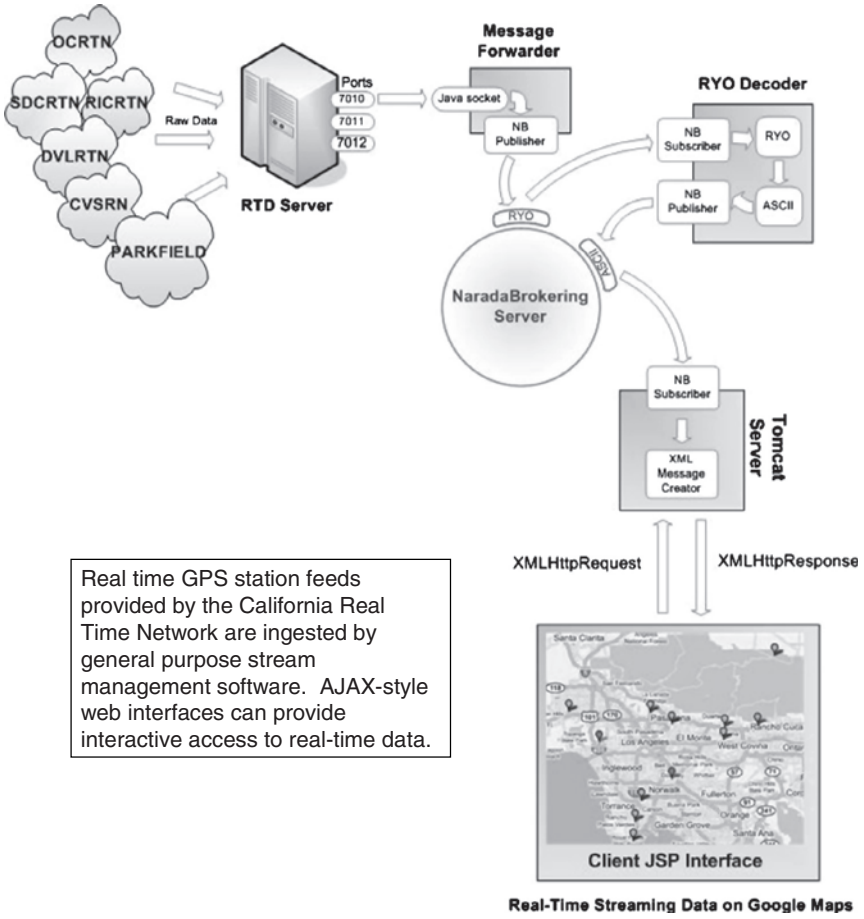


Figure 2

Live GPS data streams are processed using message filters that publish and subscribe to messaging topics. Open arrows represent TCP/IP connections. GPS network data (typically from 5–10 individual GPS stations per network) are streamed from sockets from the RTD server. Message forwarding publishers connect to these sockets and publish to topics on a publish/subscribe broker (NaradaBrokering in the figure). Message filters such as the RYO Decoder in the figure are subscribers to topics and can be used to transform or analyze messages. The results of these messages are published on new topics. For example, the individual stations' position data can be extracted from the composite signal, and each individual station's real time positions can be made available on a new topic.

5. Science Web Portals and AJAX Techniques

The QuakeSim portal has been developed around the portlet component model, which allows a Web portal to be decoupled into an aggregating container and a set of

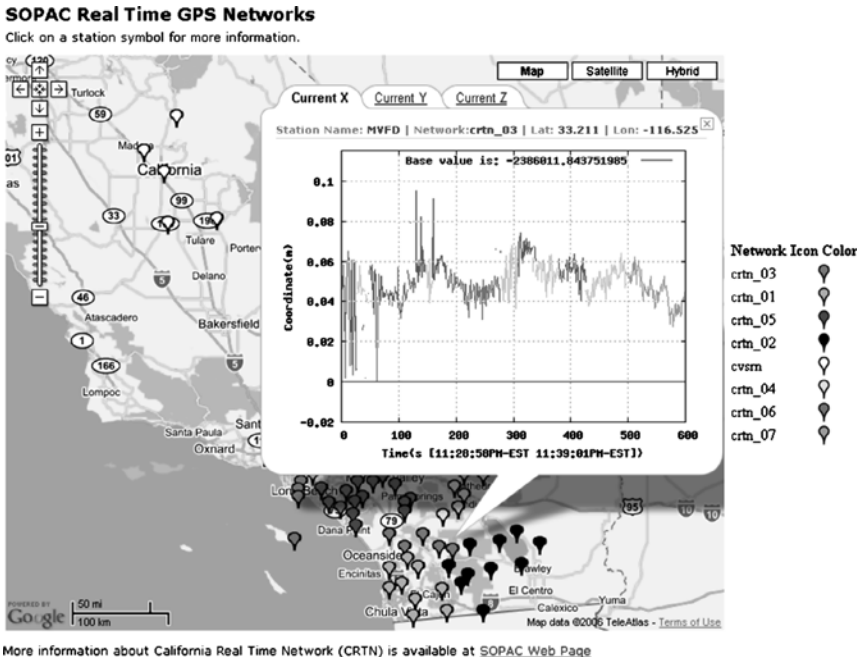


Figure 3
Real-time GPS position data for an individual GPS station are analyzed on-the-fly with RDAHMM and displayed with Google maps. The colors in the graph indicate underlying modes in the GPS signal determined by RDAHMM's analysis.

semi-independent portlet plugins. Each portlet is a self-contained Web application that (directly or indirectly) implements the portlet programming interface. This allows portlet components to be easily exchanged between developers and deployed in multiple Web portals. The original QuakeSim has been implemented using the Jetspeed 1 portlet container.

Our initial development concerns when working with Jetspeed were to support development environments that were not entangled with the container; that is, portlets should be relatively free of calls to the parent container services. We still advocate this design principal. We also wanted to ensure that portlets could be developed as (more or less) standard Java Server Pages, rather than relying on Jetspeed specific templates. To meet these requirements, we developed a general purpose *WebFormPortlet* as an extension of the Jetspeed 1 portlet base class. WebFormPortlet features include

- Support for portlets running independently (on other web servers) of the portlet container.
- Support for shared session state (through Tomcat's JSESSIONID cookie) between the container and remote portlets.

- Support for HTML <form> processing: POST parameters can be passed from the aggregating portlet container to the remote portlet.
- Portlet navigation: URL links and form actions are rewritten to allow navigated pages to appear within the portlet container rather.
- Support for both HTTP and HTTPS connections.

Our Jetspeed-based approach and WebFormPortlet have subsequently been superseded by the Java Specification Request (JSR) 168 (ABDELNUR *et al.*, 2003) and the Apache Portlet Bridges projects, respectively. We are currently updating the QuakeSim portal to comply with these standards.

Portlets are primarily a server-side Java technology and have been widely adopted in a number of science gateway projects. In parallel development, improved support in major browsers for JavaScript standards has enabled the prominent re-emergence of client-side browser applications, with a number of examples such as Google Maps. These rich browser client interfaces are developed using techniques that are collectively dubbed Asynchronous JavaScript and XML (AJAX). AJAX itself is a methodology, not a specific piece of software. AJAX is not purely a browser-side technology but instead makes use of the now widely supported *XMLHttpRequest* object in JavaScript to decouple user interactions in the browser from the browser requests to remote services. This greatly increases the interactive capabilities of Web browser user interfaces while maintaining server-side data. We have very actively pursued integration of our GIS Web Services (particularly our Web Feature Services and their GPS, seismic, and fault data) with Google Maps. We find this particularly useful for simulating "server push" in our real-time GPS data work (described in Section 4), as shown in Figure 3.

6. Service Coordination and Workflow

General purpose Web services for data access and application management need to be combined into composite applications capable of performing specific scientific tasks. Such coordination in Grid computing is typically called scientific workflow and is an active research area for numerous groups. Example scientific workflow research projects include Kepler (BOWERS *et al.*, 2005), Taverna (OINN *et al.*, 2004), and SciFlo (WILSON *et al.*, 2005). See also GANNON and FOX (2006) and LUDÄSCHER and GOBLE (2005) for overviews.

Workflow may be generally understood as the execution of a flowchart or (more simply) a directed acyclic graph of operations. Workflow systems at their most basic are ways for coupling (either graphically or through scripting) reusable components into composite applications. In Grid system, the components of the graph are proxies to remote services. Workflow systems for Grids thus must address the challenges of distributed execution environments: They must maintain a directory of available services,

they must support nonlocal data, and they must be able to handle distributed state; that is, they must be able to monitor the stage of an execution workflow.

Typically, the above Grid workflow requirements are addressed within a single workflow engine. In our approach, we have researched the decoupling of execution and state management, and we have addressed the data transportation problems. Our workflow engine research project is called HPSearch and is described in greater detail in GADGIL *et al*. (2005). HPSearch uses JavaScript as a workflow-scripting engine and can bind Web service proxies into flows. One of HPSearch's distinguishing features is its ability to handle and manage workflow components that communicate with data streams, or data sent continuously over network connections. This is particularly important for scientific workflows, which must process and transfer data between distributed components. Typically, Grid workflow engines do not take this directly into account, therefore data is often inefficiently passed directly back to the workflow engine, where it is resent to the next component. We have used the Pattern Informatics application (see Section 2) in several tests cases for our workflow system. This integration and our evaluation are described in more detail in AYDIN (2005).

Distributed workflows are composed of Web Service components collected for a specific task. This is in fact the strength of the Web Service approach: Services can be collected to solve a wide variety of problems. Sophisticated, specialized composite applications are composed of general purpose service components. To realize this in a distributed system, we need dynamically updated information systems that describe services so that they can be assembled into workflows, and we need information systems that can manage the internal state of these composite applications. The latter case is best understood as a communication of the stage in the execution of the workflow. Different components in the system will need to be updated as each service completes or fails in its execution, and we need a place to put various pieces of supporting metadata, such as the URLs for generating output files. Our research in this area is based on the Web Service specifications UDDI and WS-Context. More detailed information on architecture, implementation, and testing are available from ATKAS (2006a, b).

7. Unit and System Testing of Grid Deployments

One of the difficulties in managing Grids and Grid portal deployments is monitoring the system for failures. Similarly, science portal systems do not easily fit within a traditional "white-box" style of unit testing, which assumes knowledge of the source code of all parts of the system, since they depend both on remote clients (users' browsers) and remote Web services that in turn wrap binary executables. The Inca project (SMALLEN *et al*., 2004) is an example effort in Grid testing and monitoring. User interface-driven testing, such as is needed by science portals, is an open area of systems research, but can be very useful as a way of encapsulating composite tests that simulate real usage scenarios. We describe here some of our pragmatic solutions to these problems.

In QuakeSim, we have developed a "black box" system testing environment for the portal. This functionally tests the running portal instance to simulate user interactions with the portal. Unit tests verify that the HTML returned by the browser contains expected responses and does not contain error messages. Since the responses are generated through interactions with remote services, we can indirectly test these services to make sure they respond correctly to known inputs. We have found the Apache Ant, Apache JMeter, and HttpUnit projects to be especially useful for this task. JMeter and HttpUnit provide "black-box" testing that simulates a user's interactions with a portal. We have developed an HttpUnit-based test suite that enables us to simulate portal usage: The tests can be used to log into the portal, navigate user pages, and launch applications. These tests are run daily to monitor the system.

8. Conclusions and Future Work

This article has reviewed the development work of the QuakeSim project. We have adopted a Web Service-based architecture for encapsulating geophysical applications and data sources as simple Web services with well-defined programming interfaces. This allows diverse clients to interact with our services and provides an open architecture for other application developers to reuse our services and our portlet components.

Web services and portlet-based portals have been adopted by a number of projects, with GEON (YOUN *et al.*, 2005) and LEAD (PLALE *et al.*, 2005) serving as two relevant examples in geosciences and atmospheric sciences, respectively. In our opinion, one of the shortcomings of all of these projects is not in the basic architecture but in the lack of reuse of the projects' running service instances in clients built by third party groups. In the general Web development community, so-called "mash-up" web applications are widely used. These are novel, easily-developed user interfaces built by using Web client programming interfaces (typically written in JavaScript) to connect to remote services from general service providers. An interesting list of publicly available service interfaces is available from http://www.programmableweb.com/apis. Enabling the easier development of science mashups is a potentially important future development for broadening the impact of cyberinfrastructure developments. The QuakeSim project has followed an open, service-oriented architecture compatible with the "programmable web" paradigm, but we need to investigate lighter-weight bridges to our services.

The other major challenge for QuakeSim and for science portals and gateways in general is the technical challenges of integrating AJAX-style Web development with portlet development. AJAX, when done well, provides a substantial improvement in the user interface capabilities over purely server-side technologies. AJAX relies heavily on JavaScript in the client browser, while JSR 168 portlets are a server-side technology. It is an important activity for Web portals in the geosciences to develop JavaScript libraries for easily integrating services into such technologies as Google Maps. The current JSR 168 standard has some important limitations in dealing with JavaScript library imports,

but this should be addressed in the next version of the standard (JSR 286) that is currently being developed. More generally, integrating the two represents a challenge in state management: The information in the user's browser must be kept synchronized with the information on the server side, since the server side objects act as proxies to multiple remote Grid resources.

The technical issues we have raised in this section are solvable problems and are currently under investigation by several groups. More importantly the implications of open architectures adopted by our project and others require fuller examination. We hope that the next generation of science cyberinfrastructure will be driven by a considerably more community-driven development environment. We believe the geosciences will be an important example in this new development model.

Acknowledgements

This work was supported by the National Aeronautics and Space Administration's Computational Technologies (CT) and Advanced Information System Technology (AIST) programs. We thank Dr. Yehuda Bock and his group at the Scripps Orbital and Permanent Array Center for their assistance with the California Real Time Network.

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. E., Fox, G. C., and Leake, D. B. (2004), *A Web based Conversational Case-Based Recommender System for Ontology aided Metadata Discovery*, GRID 2004, 69–75.

Aktas M. S., Fox, G. C., and Pierce, M. E. (2005), *Information services for dynamically assembled semantic grids*, Proc. 1st Internat. Conf. on SKG2005, *Semantics, Knowledge and Grid Beijing China*, November 27–29 2005. In IEEE Proc. 1st Internat. Conf. on SKG2005 Semantics, Knowledge and Grid Beijing China, November 27–29 2005, to appear.

Aktas, M. S., Fox, G. C., and Pierce, M. E.: *Fault Tolerant High Performance Information Services for Dynamic Collections of Grid and Web Services*. In *Semantic Grid and Knowledge Grid: Systems and Applications*, Special Issue of Future Generation Computer Systems, The Internat. J. Grid Comp.: Theory, Models and Applications.

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

Atkinson, M. P., De Roure, D., Dunlop, A. N., Fox, G. C., Henderson, P., Hey, A. J. G., Paton, N. W., Newhouse. S., Parastatidis, S., Trefethen, A. E., Watson, P., and Webber, J. (2005), *Web Service Grids: an evolutionary approach. Concurrency*, Practice and Exper. *17*(2–4), 377–389.

Aydin, G., Aktas, M. S., Fox, G. C., Gadgil, H., Pierce, M. E., and Sayar, A. (2005), *QuakeSim Complexity Computational Environments (CCE) Integrated Performance Analysis*, Proc. Grid Computing Conf, 2005. The 6th IEEE/ACM International Workshop 13–14 Nov., 2005. 256–261 DOI.

Beaujardiere, de La, J. (2004), *Web Map Service*, OGC Project Document Reference Number OGC 04–024. .

Bowers, S. and Ludäscher, B. (2005), *Actor-Oriented Design of Scientific Workflows*, ER 2005: 369–384 (Efrat Jaeger, Ilkay Altintas, Jianting Zhang, Bertram Ludäscher, Deana Pennington, William Michener) *A Scientific Workflow Approach to Distributed Geospatial Data Processing Using Web Services*, SSDBM 2005: 87–90.

Chen, A., Donnellan, A., McLeod, D., Fox, G., Parker, J., Rundle, J., Grant, L., Pierce, M., Gould, M., Chung, S., and Gao, S., *Interoperability and Semantics for Heterogeneous Earthquake Science Data*, International Workshop on Semantic Web Technologies for Searching and Retrieving Scientific Data, Sanibel Island, FL (October 2003).

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.

Foster, I. and Kesselman, C. (eds.), *The Grid 2: Blueprint for a new Computing Infrastructure* (Morgan Kaufmann, 2004).

Fox, G. C., Lim, S., Pallickara, S., and Pierce, M. E. (2005a), *Message-based cellular peer-to-peer grids: Foundations for secure federation and autonomic services*, Future Generation Comp. Syst. *21*(3), 401–415.

Fox, G. C., Aydin, G., Gadgil, H., Pallickara, S., Pierce, M. E., and Wu, W. (2005b), *Management of Real-Time Streaming Data Grid Services*, GCC 2005, 3–12.

Gadgil, H., Fox. G. C., Pallickara, S., Pierce, M. E., and Granat, R. (2005), *A scripting based architecture for management of streams and services in real-time grid applications*., CCGRID 2005: 710–717.

Gannon, D. and Fox, G. C. (2006), Special Issue: *Workflow in Grid Systems*, Concurrency and Computation: Practice and Experience 18(10), 1009–1019.

Granat, R. A., *Regularized Deterministic Annealing EM for Hidden Markov Models*, Ph.D. Thesis, University of California, Los Angeles (2004).

Grant, L.B., Donnellan, A., McLeod, D., Pierce, M. E., Fox, G. C., Chen, A.Y., Gould, M. M., Sung, S. S., and Rundle, P. B. A, *Web-Service Based Universal Approach to Heterogeneous Fault Databases*, Comp. Sci. Engg. Special Issue on Multi-Physics Modeling, in press.

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/.

Ludäscher, B. and Goble, C. A. (2005), *Guest editors' introduction to the special section on scientific Workflows*, SIGMOD Record *34*(3), 3–4.

Oinn, T. M., Addis, M., Ferris, J., Marvin, D., Senger, M., Greenwood, M., Carver, T., Glover, K., Pocock, M. R., Wipat, A., and Li, P. (2004), *Taverna: A tool for the composition and enactment of bioinformatics workflows*, Bioinformatics *20*(17), 3045–3054.

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

Pallickara, S. and Fox, G. C. (2003), *NaradaBrokering: A Distributed Middleware Framework and Architecture for Enabling Durable Peer-to-Peer Grids*, Middleware 2003, 41–61.

Parker, J., Lyzenga, G., Donnellan, A., Judd, M., Baker T., Norton, C., Tisdale, E., and Li, P., *Using the GeoFEST faulted region simulation system*, Proc. 4th ACES Workshop, Beijing China, in Press.

Pierce, M. E., Fox, G. C., Youn, C.-H., Mock, S., Mueller, K., and Balsoy, O. (2002), *Interoperable Web services for computational portals*, SC 2002, 1–12.

Pierce, M. E., Youn, C.-H., and Fox, G. C. (2002), *The Gateway computational Web portal*, Concurrency and Computation: Practice and Experience *14*(13–15), 1411–1426.

Plale, B., Gannon, D., Reed, D. A., Graves, S. J., Droegemeier, K., Wilhelmson, R., and Ramamurthy, M. (2005), *Towards Dynamically Adaptive Weather Analysis and Forecasting in LEAD*, Internat. Conf. Comput. Sci. 2, 624–631.

Rundle, J. B., Rundle, P. B., Klein, W., Martins, J., Tiampo, K. F., Donnellan, A., and Kellogg, L. H. (2002), *GEM plate boundary simulations for the plate boundary observatory: Understanding the physics of earthquakes on complex fault systems*, *Pure Appl. Geophys. 159*, 2357–2381.

Sayar, A., Pierce, M. E., and Fox, G. C. (2006), *Integrating AJAX Approach into GIS Visualization Web Services*, AICT/ICIW 2006: 169.

Sensor Model Language (SensorML) Project, Web Site: http://vast.nsstc.uah.edu/SensorML/.

SMALLEN, S., OLSCHANOWSKY, C., ERICSON, K., BECKMAN, P., and SCHOPF, J. M. (2004), *The Inca Test Harness and Reporting Framework*, SC 2004, 55.

TIAMPO, K.F., RUNDLE, J.B., MCGINNIS, S., and KLEIN, W. (2002), *Pattern dynamics and forecast methods in seismically active regions*, Pure Appl. Geophys. *159*.

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

WILSON, B., TANG, B., MANIPON, G., MAZZONI, D., FETZER, E., ELDERING, A., BRAVERMAN, A., DOBINSON, E. R., and YUNCK, T. (2005), *GENESIS SciFlo: Scientific Knowledge Creation on the Grid using a Semantically-Enabled Dataflow Execution Environment*, SSDBM, 83–86.

YOUN, C.-H., PIERCE, M. E., and FOX, G. C. (2003), *Building Problem Solving Environments with Application Web Service Toolkits*, Internat. Conf. Comp. Sci. 403–412.

YOUN, C.-H., KAISER, T., SANTINI, C., and SEBER, D. (2005), *Design and Implementation of Services for a Synthetic Seismogram Calculation Tool on the Grid*, Internat. Conf. Comput. Sci. (1) 469–476.

(Received October 6, 2006, Revised April 4, 2007, accepted April 22, 2007)

Pure appl. geophys. 165 (2008) 653–670
0033–4553/08/030653–18
DOI 10.1007/s00024-008-0327-7

Pure and Applied Geophysics

A New Design of Scientific Software Using Python and XML

Lutz Gross,[1] Hans Mühlhaus,[1] Elspeth Thorne,[1] and Ken Steube[1]

Abstract—In this paper we advance the development of our *python*-based package for the solution of partial differential equations using spatial discretization techniques such as the finite element method (FEM) via two approaches. First we define a Model class object which makes it easy to break down a complex simulation into simpler sub-models, which then can be linked together into a highly efficient whole. Second, we implement an XML schema in which we can save an entire simulation. This allows implemention of check-pointing in addition to graphical user interfaces which enables non-programmers to use models developed for their research. These features are built upon our *escript* module, a software package designed to develop numerical models in a very abstract way while still allowing the use of computational components implemented in C and C++ to achieve extreme high-performance for time-intensive calculations.

Key words: Partial differential equations, mathematical modelling, XML schema, Drucker–Prager flow.

1. Introduction

Presently numerical simulations require a team effort unifying a variety of skills. In a very simple approach we can identify four groups of people involved: Researchers using numerical simulation techniques to improve the understanding and prediction of phenomena in science and engineering, modelers developing and validating mathematical models, computational scientists implementing the underlying numerical methods, software engineers implementing and optimizing algorithms for a particular architecture. Each of these skill levels employ their individual terminology: A researcher uses terms such as stress, temperature and constitutive laws, while modelers express their models through functions and partial differential equations. The computational scientists work with grids and matrices. Software engineers work with arrays and data structures. Finally, an object such as stress used by a researcher is represented as a suitable, maybe platform-dependent, data structure after the modeler has interpreted it as a function of spatial coordinates and the computational scientists as values at the center of elements in a finite-element mesh. When moving from the software engineer's view towards the view of the researcher these data structures undergo an abstraction process until finally only seeing

[1] Earth Systems Science Computational Center, The University of Queensland, St. Lucia, QLD 4072, Australia. E-mail: gross@uq.edu.au; h.muhlhaus@uq.edu.au; e.thorne@uq.edu.au; k.steube@uq.edu.au

the concept of stress and ignoring the fact that it is a function in the L^2 Sobolev space and represented using a finite-element mesh.

It is also important to point out that each of these layers has an appropriate user environment. For the researcher, this is a set of input files describing the problem to be solved, typically in XML (http://www.w3.org/XML). Modelers, mostly not trained as software engineers, prefer to work in script–based environments, such as *python* (LUTZ, 2001) or MATLAB while computational scientists and software engineers work with programming languages such as C, C++ or FORTRAN in order to achieve the best possible computational efficiency.

Various efforts have been made to provide tools for computational scientists to develop numerical algorithms, for instance PETSc (PACHECO, 1997) which is widely used. These tools provide linear algebra concepts such as vectors and matrices while hiding data structures from the user. The need of researchers for a user-friendly environment has been addressed in various projects with the development of problem solving environments (PSEs) (HOUSTIS *et al.*, 2000) some of which take the form of a simple graphical user interface. In contrast there were relatively few activities engaged in the development of environments for modelers. Two examples for partial differential equation (PDE) based modeling are ELLPACK (RICE and BOISVERT, 1985) and FASTFLO (LUO *et al.*, 1996). Both products use their own programming language and as such are not powerful enough to deal with complex and coupled problems in an easy and efficient way.

The *escript* module (GROSS *et al.*, 2005; DAVIES *et al.*, 2004) is an environment in which modelers can develop PDE based models using *python*. It is designed to solve general, coupled, time-dependent, non-linear systems of PDEs. It is a fundamental design feature of *escript* that is not tied to a particular spatial discretization technique or PDE solver library. It is seamlessly linked with other tools such as linear algebra tools (GREENFIELD *et al.*, 2001) and visualization tools (KITWARE, Inc., 2003). We refer to (https://shake200.esscc.uq.edu.au) to access the *escript* users guide, example scripts and the source code.

In the first part of the paper we will give an overview of the basic concepts of *escript* from a modelers point of view. We will illustrate its usage in the implementation of the Drucker–Prager flow model. In the second part we will present the *modelframe* module within *escript*. It provides a framework to implement mathematical models as *python* objects which then can be plugged together to build simulations. We illustrate this approach for the Drucker–Prager flow and demonstrate how this model can be linked with a temperature advection–diffusion model without modifying any of the cods for the models. We will then discuss how XML can be used to set not only simulation parameters but also to define an entire simulation from existing models. The XML files provide an ideal method to build simulations out of PSEs or from web services. The presented implementation of the Drucker–Prager flow has been validated on some test problems but a detailed discussion of these results is beyond the scope of this paper.

2. Drucker–Prager Flow Model

2.1. The General Framework

We will illustrate the above-mentioned concepts for the deformation of a three-dimensional body loaded by a time-dependent volume force, f_i, and by a surface load, g_i. If u_i and σ_{ij} define the displacement and stress state at a given time t and dt gives a time increment, then the general framework for implementing a model in the updated Lagrangian framework is given below:

0. start time integration, set $t = 0$
1. start iteration at time t, set $k = 0$
   1.0. calculate displacement increment du_i from σ_{ij}^- and current force f_i.
   1.1. update geometry
   1.2. calculate stretching D_{ij} and spin W_{ij} from v_i.
   1.3. calculate new stress σ_{ij} from σ_{ij}^- using D_{ij} and W_{ij}.
   1.4. update material properties from σ_{ij}
   1.5. if not converged, $k \leftarrow k + 1$ and goto 1.0.
2. set $t \leftarrow t + dt$ and goto 1.

The superscript $'-'$ refers to values at the previous iteration or time step. To terminate the iteration process at a time step one can use the relative changes in the velocity $v = \dfrac{u - u^-}{dt}$:

$$\|du\|_\infty \leq \zeta_{iter} \|u\|_\infty, \tag{2.1}$$

where $\|.\|_\infty$ denotes the maximum norm and ζ_{iter} is a specified, positive relative tolerance. Alternatively, one can check the stress change.

To ensure that the relative time integration error is forced below a specified tolerance ζ_{time} the time step dt for the next time step is bounded by dt_{max} given below:

$$dt_{max} = \frac{dt\ \|u\|_\infty}{\|v - v^-\|_\infty} \zeta_{time}, \tag{2.2}$$

which controls an estimate of the local time discretization error for the total displacement u.

The stretching D_{ij} and spin W_{ij} are defined as the symmetric and non-symmetric part of the gradient of du_i:

$$D_{ij} = \frac{1}{2}(du_{i,j} + du_{j,i}), \tag{2.3}$$

$$W_{ij} = \frac{1}{2}(du_{i,j} - du_{j,i}), \tag{2.4}$$

where for any function Z, $Z_{,i}$ denotes the derivative of Z with respect to x_i.

To calculate the displacement increment du_i one has to solve a partial differential equation on the deformed domain Ω^- which in tensor notation takes the form:

$$-(S_{ijkl}\, du_{k,l})_{,j} = (\sigma^-_{ij} - K\, \epsilon^{th}\delta_{ij})_{,j} + f_i \tag{2.5}$$

where K is the bulk modulus and S_{ijkl} is the tangential tensor which depends on the rheology that is used. The argument of the divergence expression on the right hand side is the Cauchy stress at time $t - dt$ with the mechanical stress σ^-_{ij} at time $t - dt$. The term $K\, \varepsilon^{th}\, \delta_{ij}$ is the thermal stress where the thermal volume strain ε^{th} is given as

$$\epsilon^{th} = \theta(T - T_{ref}) \tag{2.6}$$

with thermal expansion coefficient θ, current temperature T and the reference temperature T_{ref}. In geoscience applications, f_i typically takes the form of the gravitational force

$$f_i = -\rho\, g\delta_{id} \tag{2.7}$$

which acts oppositely to the positive d direction. The constants g and ρ are the gravity and density, respectively. The density is often given in the form

$$\rho = \rho_0(1 - \theta(T - T_0)), \tag{2.8}$$

where ρ_0 is the density at the reference temperature T_0.

The displacement increment has to fulfill the conditions

$$n_j(S_{ijkl}\, du_{k,l} + \sigma^-_{ij} - K\, \epsilon^{th}\delta_{ij}) = g_i \tag{2.9}$$

on the boundary of the domain where g_i is a time-dependent surface load. Moreover, the displacement increment has to meet a constraint of the form

$$du = dt \cdot V_i \text{ on } \Gamma_i, \tag{2.10}$$

where Γ_i is a subset of the boundary of the domain. At the first iteration step V_i gives the i-th velocity component acting on the subset Γ_i of the deformed domain and is set to zero for the following steps.

2.2. The Drucker–Prager Model

For the Drucker–Prager model the stress is updated in the following procedure: First the stress state at time $t + dt$ due to elastic deformation $\sigma^e_{i,j}$ is calculated:

$$\sigma^e_{ij} = \sigma^-_{ij} + K\, \epsilon^{th}\delta_{ij} + dt\left(2G\, D_{ij} + \left(K - \frac{2}{3}G\right)D_{kk}\delta_{ij} + W_{ik}\sigma^-_{kj} - \sigma^-_{ik}W_{kj}\right), \tag{2.11}$$

where G is the shear modulus. The yield function F is evaluated as

$$F = \tau^e - \alpha\, p^e - \tau_Y, \tag{2.12}$$

where

$$p^e = -\frac{1}{3}\sigma^e_{kk}, \tag{2.13}$$

$$\tau^e = \sqrt{\frac{1}{2}(\sigma^e_{ij})'(\sigma^e_{ij})'} \tag{2.14}$$

with the deviatoric stress

$$(\sigma^e_{ij})' = \sigma^e_{ij} + p^e\delta_{ij}. \tag{2.15}$$

The value τ_Y is the current shear length and α is the friction parameter, both of which are given as a function of the plastic shear stress γ^p.

We require a non–negative yield function. The factor χ marks when the yield condition is violated:

$$\chi = \begin{cases} 0 & \text{for } F < 0 \\ 1 & \text{else} \end{cases}. \tag{2.16}$$

With current hardening modulus $h = \dfrac{d\tau_Y}{d\gamma^p}$ and dilatancy parameter β, which is again a given function of the plastic shear stress γ^p, we can define the plastic shear stress increment to

$$\lambda = \chi\frac{F}{h + G + \beta K}. \tag{2.17}$$

We then can calculate a new stress as

$$\tau = \tau^e - \lambda G, \tag{2.18}$$

$$\sigma_{ij} = \frac{\tau}{\tau^e}(\sigma^e_{ij})' - (p^e + \lambda\beta K)\delta_{ij}. \tag{2.19}$$

Finally we can update the plastic shear stress

$$\gamma^p = \gamma^{p-} + \lambda \tag{2.20}$$

and the hardening parameter

$$h = \frac{\tau_Y - \tau^-_Y}{\gamma^p - \gamma^{p-}}. \tag{2.21}$$

For the Drucker–Prager model the tangential tensor is given as

$$\begin{aligned} S_{ijkl} = {} & G(\delta_{ik}\delta_{jl} + \delta_{jk}\delta_{il}) + \left(K - \frac{2}{3}G\right)\delta_{ij}\delta_{kl} \\ & + (\sigma^-_{ij}\delta_{kl} - \sigma^-_{il}\delta_{jk}) + \frac{1}{2}(\delta_{ik}\sigma^-_{lj} - \delta_{jl}\sigma^-_{ik} + \delta_{jk}\sigma^-_{il} - \delta_{il}\sigma^-_{kj}) \\ & - \frac{\chi}{h + G + \alpha\beta K}\left(\frac{G(\sigma_{ij})'}{\tau} + \beta K\delta_{ij}\right)\left(\frac{G(\sigma_{kl})'}{\tau} + \alpha K\delta_{kl}\right). \end{aligned} \tag{2.22}$$

In the implementation of the model we have to monitor the total stress σ, plastic stress γ^p, the shear length τ_Y as well as displacement u and velocity v.

It is also worth mentioning that the model is temperature-dependent through the thermal strain, the temperature-dependent density, and the standard advection-diffusion equation, the so called heat equation (WILLIAMSON, 2001).

3. Implementation

The key problem in the Drucker–Prager model is soving a linear, steady partial differential equation. The coefficients of the PDE are given by arithmetic expressions of the PDE solutions. In this section we outline the concept of the *python* module *escript* (GROSS *et al*., 2005; DAVIES *et al*., 2004) and demonstrate its usage in the Drucker–Prager model.

3.1. Spatial Functions

The solution of a PDE as well as the PDE coefficients are functions of the spatial coordinates varying within a domain. In *escript* a domain is represented in a `Domain` class object. The `Domain` class object not only holds information about the geometry of the domain, but also about the PDE solver library used to solve PDEs on this `Domain`. In particular, the `Domain` class object defines the discretization technique, such as the finite-difference method (FDM) or finite-element method (FEM) (ZIENKIEWICZ and TAYLOR, 2000), which is used to represent spatial functions on the domain. In the current implementation *escript* is linked to FEM solver library *finley* (DAVIES *et al*., 2004). However this design feature is open and other libraries could also be implemented.

A spatial function is represented through its values on sample points and is represented in a Data class object (see Fig. 1). For instance, in the context of the FEM the solution of a PDE is represented by its values on the nodes while PDE coefficients are typically represented by their values at the element centres. The type of representation is indicated by the `FunctionSpace` object attached to a `Data` class object (see Fig. 1). The appropriate representation of a function is chosen by *escript* or, if required, can be controlled by the user.

Spatial functions can be manipulated by applying unary operations, for instance cos, sin, log) and be combined through binary operations, for instance $+$, $-$,* , /. When combining spatial functions with different representations *escript* uses interpolation to make the representations of the arguments compatible for combination. For example if in the FEM context a function represented by its values on the element centers and a function represented by its values on the nodes are added together, interpolation from nodes to element centers will be performed before the addition takes place.

In the current version of *escript*, three internal storage schemes for spatial function representation are used. In the most general case individual values are stored for each sample point. Typically this expanded storage scheme is used for the solution of a PDE. In many cases PDE coefficients are constant. In this case, it would be a waste of memory

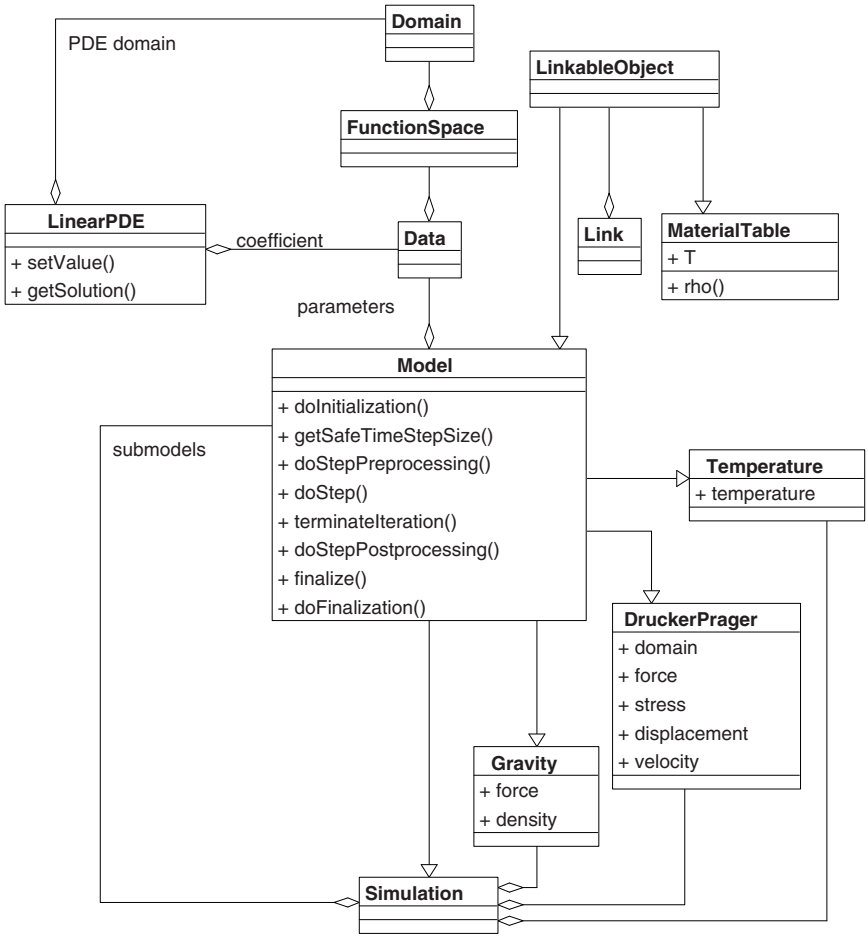


Figure 1
The *escript* Class Diagram. Boxes represent classes where the upper part defines the name, the middle part the attributes and the lower part the methods. An arrow indicates that the target is a subclass of the source. A line ending with a diamond indicates that the target class refers to the source

and computation time to use the expanded storage scheme. For this case *escript* stores a single value only and refers to this value whenever the value for a sample point is requested. The third form represents piecewise constant functions that are stored as a dictionary. The keys of the dictionary refer to tags assigned to the sample points at the time of mesh generation. The value at a sample point is accessed via the tag assigned to the sample point and the associated value of the tag in the dictionary. Typically, this

technique is used when a spatial function is defined according to its location in certain subregions of the domain, for instance in the region of a certain rock type. When data objects using different storage schemes are combined, *escript* chooses an appropriate storage scheme in which to store the resulting combined data.

The following *python* function implements the stress update for the Drucker–Prager model presented in section 2.2. It takes the last stress σ_{ij}^{-} (=s_old), the last plastic stress γ^{p-} (=gamma_p_old), displacement increment du_i (=du), the thermal volume strain increment ϵ^{th} (=deps_th) and the relevant material parameters to return the new stress and plastic stress:

```
from escript import *
def getStress(du,s,gamma_p_old,deps_th,tau_Y,G,K,alpha,beta,h):
  g=grad(du)
  D=symmetric(g)
  W=nonsymmetric(g)
  s_e=s+K*deps_th*kronecker(3)+\
    2*G*D+(K-2./3*G)*trace(D)*kronecker(3)\
    +2*symmetric(matrix_mult(W,s)
  p_e=-1./3*trace(s_e)
  s_e_dev=s_e+p_e*kronecker(3)
  tau_e=sqrt(1./2*inner(s_e_dev,s_e_dev))
  F=tau_e-alpha*p_e-tau_Y
  chi=whereNonNegative(F)
  l=chi*F/(h+G+alpha*beta*K)
  tau=tau_e-l*G
  s=tau/tau_e*s_e_dev+(p_e+l*beta*K)*kronecker(3)
  gamma_p=gamma_p+l
  return s, gamma_p
```

Each of the arguments, with the exception of du, can be handed over as a *python* floating point number or an *escript* spatial function. The argument du must be an *escript* spatial function. The Domain attached to it is the Domain of the returned values. Their representations are determined by the grad call, which returns the gradient of its argument. In the FEM context, the argument of grad must be represented by its values on the nodes while the result of grad is represented by its values at the element centers. This representation is carried through to the return values. Note that this function does not depend on whether values are stored on the nodes or at the centers of the elements, and so it may be re-used without change in any other model, even a finite-difference model.

3.2. Linear PDEs

The LinearPDE class object provides an interface to define a general, second-order, linear PDE. The Domain class object attached to a LinearPDE class object defines the domain of the PDE as well as the PDE solver library to be used to solve the PDE. Data class objects define the values of the coefficients of PDE (see Fig. 1).

The general form of a PDE for an unknown vector-valued function u_i represented by the LinearPDE class is

$$-(A_{ijkl}u_{k,l} + B_{ijk}u_k)_{,j} + C_{ikl}u_{k,l} + D_{ik}u_k = -X_{ij,j} + Y_i. \tag{3.1}$$

The coefficients A, B, C, D, X and Y are functions of their location in the domain. Moreover, natural boundary conditions of the form

$$n_j(A_{ijkl}u_{k,l} + B_{ijk}u_k - X_{ij,j}) = y_i \tag{3.2}$$

can be defined. In this condition, (n_j) defines the outer normal field of boundary of the domain and y is a given function. To restrict values of u_i to a specified value, r_i, on certain locations in the domains one can define constraints of the form

$$u_i = r_i \text{ where } q_i > 0, \tag{3.3}$$

and q_i is a characteristic function of the locations where the constraint is applied.

From inspecting equation (2.5) and equation (2.9) we can easily identify the values for the PDE coefficients needed to calculate the displacement increment du_i in the Drucker–Prager model:

$$A_{ijkl} = S_{ijkl}, Y_i = f_i, X = \sigma_{ij}^{-} - K\epsilon^{th}\delta_{ij} \text{ and } y_i = g_i. \tag{3.4}$$

The constraints for du_i (see equation (2.10)) is defined by

$$r_i = dt \cdot V_i \text{ and } q_i(x) = \begin{cases} 1 & \text{for } x \in \Gamma_i \\ 0 & \text{otherwise} \end{cases}. \tag{3.5}$$

If we define function `getTangentialTensor` which returns the tangential operator, the following function returns a new displacement increment `du` by solving a linear `PDE` for the `Domain` class object `dom`:

```
from escript import LinearPDE
def getDu(dom,dt,s,dt,deps_th,tau_Y,G,K,alpha,beta,h,f,g,Gamma,V):
 pde=LinearPDE(dom)
 S=getTangentialTensor(s,tau_Y,G,K,alpha,beta,h)
 pde.setValue(A=S, X=s-K*deps_th*kronecker(3),\
   Y=f, y=g)
 pde.setValue(q=Gamma, r=dt*V)
 du=pde.getSolution()
 return du
```

It would be more efficient to create an instance of the `LinearPDE` class object and reuse it for each new displacement increment with updated coefficients. In this way information that is expensive to obtain, for instance the pattern of the stiffness matrix, can be preserved. The following function returns the tangential tensor `S`:

```
from escript import *
def getTangentialTensor(s,tau_Y,G,K,alpha,beta,h):
  k3=kronecker(3)
  p=-1./3*trace(s)
  s_dev=s+p*k3
  tau=sqrt(1./2*inner(s_dev,s_dev))
  chi=whereNonNegative(tau-alpha*p-tau_Y)
  tmp=G*s_dev/tau
  sXk3=outer(s,k3)
  k3Xk3=outer(k3,k3)
  S= G*(swap_axes(k3Xk3,0,3)+swap_axes(k3Xk3,1,3))\
     + (K-2./3*G)*k3Xk3\
     + (sXk3-swap_axes(swap_axes(sXk3,1,2),2,3))\
     + 1./2*(swap_axes(swap_axes(sXk3,0,2),2,3) \
    -swap_axes(swap_axes(sXk3,0,3),2,3)         \
    -swap_axes(sXk3,1,2)                        \
    +swap_axes(sXk3,1,3)           )            \
   - outer(chi/(h+G+alpha*beta*K)*(tmp+beta*K*k3),tmp+alpha*K*k3)
  return S
```

In this implementation the `chi`, `s_dev` and `tau` are recalculated. A slight modification would allow these previously in the `getStress` calculated values to be reused.

4. Modeling Framework

From the functions presented in the previous section 3 one can easily build a time integration scheme to implement the Drucker–Prager flow model through two nested loops. The outer loop progresses in time, while the inner loop iterates for each time step until the stopping criterion (2.1) is fulfilled. The time step size is controlled by the criterion defined by Equation (2.2). This straight-forward approach has the disadvantage that the code needs to be modified when the temperature dependence becomes time-dependent. In this case, an advection-diffusion model for the temperature has to be incorporated into the code which may use a different time step size. In addition, the thermal strain ϵ^{th} changes over time, and material parameters, for instance density (see Equation (2.8)) may become temperature-dependent. The coding effort to implement these components is not large but has the big drawback that each submodel, such as the Drucker–Prager model, the temperature advection-diffusion model and the material property tables, loses its independence. However, from a software development point of view it is highly desirable to keep the models independent in order to test and maintain them separately, and reuse models in another context.

4.1. Linkable Objects

To resolve this problem, a model, such as Drucker–Prager flow model, is implemented in a *python* class. Model parameters, such as external force F_i or state variables, such as stress σ_{ij}, are defined as object attributes. For instance if the class `DruckerPrager` implements the Drucker–Prager flow, and `Gravity` the gravity force (2.7) one uses

```
g=Gravity()
g.density=1000
flow=DruckerPrager()
flow.force=g.force
```

to assign a value to the density of the gravity force and then defines the gravity force as the external force of the Drucker–Prager flow model. In the case of convection model, the density is instead being calculated from a temperature field updated through the temperature model over time. However, the `Gravity` object `g` has to use the most recent value of the density at any time of the simulation. To accomplish this the *escript* module provides two classes `LinkableObjects` and `Link` (see Fig. 1). A `Link` class object is a callable object that points to an attribute of the target object. If a `Link` class object is called, the current value of the attribute of the target object is returned. If the value is callable, the value is called and the result is returned. So, similar to a pointer in C, a `Link` class object defines another name for a target value, and is updated along with the target object.

A `LinkableObjects` class object is a standard *python* object with a modified `__getattr__` method. When accessing its attributes it behaves like a usual *python* object except when the value is a `Link` class object. In this case, the link is followed, that means the value is called and the result returned. The mechanism allows an attribute of `LinkableObjects` class object to be linked with the attribute of another target object such that at any time the object uses the value of the target object.

The following script shows how a gravity force in a flow model can be defined where the density of the gravity model is temperature-dependent:

```
mat=MaterialProperties()
mat.T_0=20
temp=Temperature()
g=Gravity()
flow=DruckerPrager()
mat.T=Link(temp,''temperature'')
g.density=Link(mat,''rho'')
flow.force=Link(g,''force'')
```

The four classes `MaterialProperties Temperature, Gravity` and `DruckerPrager` implement a material property table, a temperature advection-diffusion model, a gravity force model and a Drucker–Prager flow model, respectively. The temperature parameter `T` of the material property table is provided by the temperature

model attribute `temperature`. The density `rho` provided by the material property table is fed into the gravity model which then provides the external force for the flow model.

The model defined by equation (2.8) of a temperature-dependent density can be implemented in the following way:

```
class MaterialProperties(LinkableObject):
   def rho(self):
   return self.rho_0*(1.-self.theta*(self.T-self.T_0))
```

If the temperature `T` is linked to a time-dependent temperature model via a `Link` object the method `rho` will always use the most recent value of the temperature even if it is updated by the target temperature model. On the other hand, if an attribute of another model, for instance the attribute `density` of a `Gravity` class object, is linked against the method `rho` of an instance of the `MaterialProperties` class, accessing the attribute will call the method `rho` and use its return value calculated for the current temperature. Through the chain of links the Drucker–Prager flow model is coupled with the temperature modules. We call attention to the fact that if the temperature model considers advection, the velocity updated by the Drucker–Prager flow model is referenced by the temperature model which produces a loop of references requiring an iteration over a time step.

4.2. Models

Implementing models as `LinkableObject` class objects solves the problem of defining the data flow of coupled models. However, we also need an execution scheme for individual models in order to be able to execute a set of coupled models without changing any of the models involved. An appropriate execution scheme is defined through the `Model` class of the *modelframe* module (see Fig. 1). The `Model` class which is a subclass of the `LinkableObject` class provides a template for a set of methods that have to be implemented for a particular model. The `Model` class methods are as follows:

- `doInitialization`: This method is used to initialize the time integration process. For the flow model this method initializes displacements and stress at time $t = 0$ and creates an instance of the `LinearPDE` object used to update the displacement through equation (2.5).
- `getSafeTimeStepSize`: This method returns the maximum time step size that can be used to safely execute the next time step of the model. For the flow model the time step size defined by condition (2.2) is returned.
- `doStepPreprocessing`: This method prepares the iterative process to be executed on a time step. Typically, this function will set an initial guess for the iterative process.
- `doStep`: This method performs a single iteration step to update state variables controlled by the model due to changes in input parameters controlled by other models.

For the Drucker–Prager flow model the `doStep` method performs a single iteration step by solving the PDE (2.5) and updating the stress via (2.19).

- `terminateIteration`: The method returns `True` if the iterative process on a time step can be terminated. Typically, a stopping criterion of the form (2.1) is checked.
- `doStepPostprocessing`: This method finalizes a time step, typically by shifting the current values of state variables into the buffer of the previous time step.
- `finalize`: The method returns `True` if the time integration is finalized. Typically this method checks the current time against a given end time.
- `doFinalization`: This method finalizes the whole modeling run, typically by closing files and cleaning up.

Any Model class object can be run in the following way:

```
m=Model()
m.doInitialization()
while not finalize():
  dt=m.getSafeTimeStepSize(dt)
  m.doStepPreprocessing(dt)
  while not m.terminateIteration(): m.doStep(dt)
  m.doStepPostprocessing(dt)
m.doFinalization()
```

This scheme is executed by the `Simulation` class, which will be discussed in the next section 4.3.

The following sub-class of the Model class implements the Drucker–Prager flow model:

```
class DruckerPrager(Model):
  def doInitialization(self):
    self.pde=LinearPDE(self.domain)
    self.stress=Tensor(0.,Function(self.doamin))
    self.displacement=Vector(0.,Solution(self.doamin))
    self.velocity=Vector(0.,Solution(self.doamin))
    self.t=0
  def getSafeTimeStepSize(self):
    return dt*Lsup(u)/Lsup(self.velocity-self.velocity_old)*self.zeta
  def doStep(self,dt):
    S=self.getTangentialTensor()
    pde.setValue(A=S, X=self.stress,Y=self.force)
    self.du=self.pde.getSolution()
    self.displacement=self.displacement+self.du
    self.velocity=(self.displacement-self.displacement_old)/dt
    self.stress=self.getStress()
  def terminateIteration(self,dt):
    return Lsup(self.du)<=self.zeta*Lsup(self.displacement)
  def doStepPostprocessing(self,dt):
    self.velocity_old=self.velocity
    self.t+=dt
  def finalize(self):
    return self.t<self.t_end
```

Omitted `Model` class methods are empty. The functions `getStress` and `getTangentialTensor` introduced in section 3.2 are transformed into class methods. Function parameters are now accessed as class attributes.

4.3. Simulations

The `Simulation` class controls the execution of a list of submodels (see Fig. 1). It is an implementation of the `Model` class, where each `Model` method executes the corresponding method of all the submodels. The `getSafeTimeStepSize` returns the minimum of all step sizes required by the submodels. If all submodels terminate their iterative process `terminateIteration` returns `True`. The simulation is terminated by the `finalize` method if any of the submodels are to be `finalized`. The `doStep` method iterates over all submodels until all submodels indicate convergence.

The following script outlines the usage of the `Simulation` class for constructing and running a coupled temperature diffusion and Drucker–Prager flow model:

```
mat=MaterialProperties()
temp=Temperature()
g=Gravity()
flow=DruckerPrager()
s=Simulation([temp,g,flow])
s.run()
```

The couplings between the models as shown in section 4.1 have been omitted. As the object `mat` is not an instance of the `Model` it is not required to be included in the `Simulation`. The `run` method of the `Simulation` class executes its `Model` methods as shown in the previous section 4.2. At a time step, the simulation iterates over the submodels until all submodels have converged. The execution order is defined by the order of the models in the `Simulation` class argument. Alternatively, one can iterate over the flow model only and use the temperature from the previous time step. This can be achieved through

```
Simulation([temp,g,Simulation([flow])).run()
```

where the convergence tolerance in the temperature model has to be chosen appropriately. Note that visualization functionality, for example creating a movie of the velocity field, can be implemented as a `Model` class. This can be inserted into the `doStepPostprocessing` method, where the velocity field is rendered after each time step to create a movie frame, while the `doFinalization` method put the frames together to creates a movie.

We stress that the coupling of the models does not require any knowledge of the algorithms being used to implement the model. Furthermore, the codes implementing a model are not altered to perform the coupling. The coupling is based purely on the

submodel parameters. This high–level approach is possible because of the abstraction level that is provided by the *escript* environment.

5. Simulations and XML

Concurrent to the modeling framework, an XML dialect called EsysXML has been developed with the objective of storing simulations in a transportable form, defining the coupling of models, and allowing the creation of dynamic templates of model interfaces.

5.1. XML Model Interface

The parameters used to link models are called external parameters, and are stored in a `Model` class object. These parameters may be set by the programmer or by another model. The `Model` also stores internal parameters used in the various methods implementing the model. For instance for the `Model` class `DruckerPrager` introduced in section 4.2, the parameters `domain, force, stress , displacement` and `velocity` are external parameters while `pde` and `t` are internal parameters that are unavailable to other models. External parameters are exposed via the `declarePa-rameter` method called at `Model` class object initialization. For example, the `Model` class `DruckerPrager` is written in the following way:

```
class DruckerPrager(Model):
 def __init__(self):
   Model.__init__(self)
      self.declareParameters(domain=None, force=0.,\
      stress=0., displacement=0.,\
      velocity=0.)
```

This explicit declaration restricts the necessary input and output parameters that are exposed on a user–interface level. The XML representation of a `Model` class object includes all parameters and their current values, which could be `Link` class objects. The *python* implementation of a model itself is not stored in XML, only the data and the links between models.

A `Simulation` class object can be written into an XML file using the EsysXML schema. This includes all submodels, their external parameter and values, and possibly links between the submodels. This file can be edited using any text editor to change value of external parameters, introduce new models and set or remove couplings between models. A `Simulation` class object can be built from the XML file. The parser uses the standard *python* XML tools, packaged in PyXML (http://pyxml.sourceforge.net). Primarily, `minidom` is used to parse and generate XML content. It is a lightweight, simple implementation of the standard Document Object Model (DOM) interface, and is fast. An important application of writing `Simulation` class objects to XML is automated check–pointing.

5.2. Technical specifications of EsysXML

A diagram of the XML schema (SPERBERG-MCQUEEN and THOMPSON, 2006) used by EsysXML is included in Figure 2.

The XML schema begins with an `Esys` tag, which includes a single `Simulation`. This `Simulation` contains one or more components, each of which may be either a `Model` or another `Simulation`. The ranks of the `Component` objects define the order in which they are to be executed. Each `Model` includes one or more `Parameter` objects defined by a type (`float`, `string` or `Link`) and the name and value of the parameter. Whether or not a parameter is internal or external (available to other models) is determined by the use of `declareParameters` in the python script that implements the model.

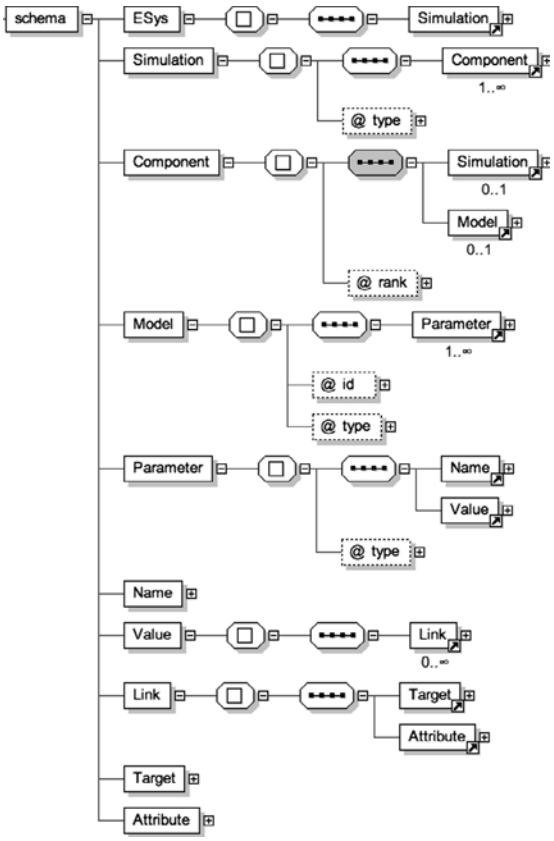


Figure 2
EsysXML Schema Tree.

The `Link` items require a persistent, deterministic way of preserving associations between linked `Model` class objects. This is achieved by generating a unique reference identification number for each `Model` class object which defines the id attribute. For a `Link` item the `Target` XML attribute specifies the reference identification number of the target `Model` item while the XML attribute `Attribute` specifies the name of the `Parameter` item of the target `Model`.

6. Conclusion

The design we have chosen for *escript* simplifies the development of complex yet efficient models by breaking them down into a series of simpler steps. In the example provided, we developed a Drucker–Prager flow model and a gravity model, and then linked them together with a temperature advection-diffusion model to create a more complex simulation.

By modifying the standard *python* mechanism for accessing object attributes, we allow sharing values between models without creating any burden on the modeler. The sharing of parameters between models is very efficient due to our use of links.

By adding the ability to store an entire simulation in XML format we have created a new level of control over a model. Using this technique we will be able to check-point simulations, share simulations with collaborators as well as providing a Graphical User Interface, such as in a web page, to define the parameters of a simulation and adjust models in a simulation. Once a modeler has developed a simulation in *python* it may then be used and adjusted by a researcher who is not encumbered with further programming.

Acknowledgments

Project work is supported by the Australian Commonwealth Government through the Australian Computational Earth Systems Simulator Major National Research Facility, by the Australian Partnership for Advanced Computing, by the Queensland State Government through the Smart State Research Facility Fund, by the Queensland Cyberinfrastructure Foundation and by SGI Ltd.

References

Davies, M., Gross, L. and Mühlhaus, H.-B. (2004), *Scripting high-performance Earth systems simulations on the SGI Altix 3700*. Proc 7th Internat. Conf. on High-Performance Computing and Grid in the Asia Pacific Region.

Greenfield, P., Miller, J.T., Hsu, J. and White, R.L., (2001) *An Array Module for Python*. In *Astronomical Data Analysis Software and Systems XI*.

Gross, L., Cochrane, P., Davies, M., Mühlhaus, H. and Smillie J. (2005), *Escript: Numerical modelling in python*. Proc. Third APAC Conf. *Advanced Computing, Grid Applications and e-Research* (APAC05).

HIGHAM, D.J. and HIGHAM, N.J. (2000) *MATLAB Guide*., SIAM.
HOUSTIS, E.N., GALLOPOULOS, E., RICE, J.R. and BRAMLEY, R. (2000), *Enabling Technologies for Computational Science* (Kluwer Academic Publishers, 2000).
http://pyxml.sourceforge.net (2006), [on-line].
https://shake200.esscc.uq.edu.au/twiki/bin/view/ESSCC/EsysUser, (2006), [on-line].
http://www.w3.org/XML/ (2006), [on-line].
LUO, X.-L., STOKES, A.N. and BARTON, N.G. (1996), *Turbulent flow around a car body - report of Fastflo solutions* Proc. WUA-CFD Conf., Freiburg.
LUTZ, M., *Programming Python*, 2nd Edition (O'Reilly, 2001).
PACHECO, P., *Parallel Programming with MPI* (Morgan-Kaufmann, 1997).
RICE, J.R. and BOISVERT, R.F., *Solving Elliptic Problems Using ELLPACK* (Springer Series in Computational Software **2**, 1985).
SPERBERG-MCQUEEN, C.M. and THOMPSON, H., *XML Schema*, http://www.w3.org/XML/Schema (2006), [on-line].
THE KITWARE, INC., *Visualization Toolkit User's Guide* (Kitware, Inc Publishers., 2003).
WILLIAMSON, R.E., *Differential Equations and Dynamical Systems* (McGrawHall, 2001).
ZIENKIEWICZ, O.C. and TAYLOR, R.L., *The Finite Element Method*, 5th Edition, (Butterworth Heinemann, 2000).

(Received October 31, 2006, Revised July 6, 2007, accepted July 9, 2007)

Pure appl. geophys. 165 (2008) 671–691
0033–4553/08/030671–21
DOI 10.1007/s00024-008-0328-6

Pure and Applied Geophysics

An Efficient System for Creating Synthetic InSAR Images from Simulations

Xiaoru Yuan,[1] Yingchun Liu,[2] David A. Yuen,[3] Baoquan Chen,[6] Tomas Pergler,[4] and Yaolin Shi[5]

Abstract—In this work we visualize tsunami and earthquake simulation results with graphics hardware acceleration. The rapid improvement in the computational power of graphics hardware and its programmability has made general computation on Graphics Processing Units (GPUs) very compelling. We generate Synthetic InSAR images using GPUs. Interference phenomena have formed the underlying theory for Interferometric Synthetic Aperture Radar (InSAR) in unveiling dynamical Earth movements. In our approach light path differences are defined by the surface values to be visualized. These path differences then modulate the lighting intensity to generate the interference patterns. We can interactively visualize surface deformation patterns by leveraging the computational power of GPUs. Our visualization method is applied to simulations of rupture fault displacements during the tsunamogenic earthquake events, which are vital to understanding the subsequent wave propagation. We also integrate the visualization results into Google Earth virtual globe to provide the geological context of the visualized regions.

Key words: Earthquake, GPU, InSAR, interference, tsunami, visualization.

1. Introduction

Underwater seismic events can produce tsunamis. Tsunamis can originate thousands of miles from land (traveling for hours before striking the coastline), or near shore (taking only minutes before impact). A massive undersea earthquake in the Indian Ocean measuring approximately moment magnitude (M_w) 9.3, known by the scientific community as the Sumatra-Andaman earthquake, hit the waters northwest of Sumatra on December 26th, 2004 and left more than 300,000 people dead or missing (USAID, 2006). It occurred along 1300 km of the oceanic subduction zone located 100 km west of

[1] Key Laboratory of Machine Perception, Minister of Education, and Department of Machine Intelligence, School of EECS, Peking University, Beijing 100871, China. Correspondence E-mail: xiaoru.yuan@gmail.com
[2] South China Sea Institute of Oceanology, Guangzhou 510301, China.
[3] Department of Geology and Geophysics and Minnesota Supercomputing Institute, University of Minnesota at Twin Cities, Minneapolis 55455, U.S.A.
[4] Department of Geophysics, Faculty of Mathematics and Physics, Charles University, Prague 180 00, Czech Republic.
[5] Department of Earth Sciences, Graduate University of Chinese Academy of Sciences, Beijing, China.
[6] University of Minnesota of Twin Cities, Minneapolis 55455, U.S.A.

Sumatra and the Nicobar and Andaman Islands in the eastern Indian Ocean (STEIN and OKAL, 2005). This undersea earthquake is one of the largest earthquakes ever recorded on a seismograph. It was reported to be the longest duration of faulting ever observed, lasting between 500 and 600 s. It was large enough that it caused the free oscillation of the entire planet (PARK *et al*., 2005). It also triggered earthquakes in other locations, even as far away as Alaska (WEST *et al*., 2005). Simulations of fault displacements during a tsunamogenic earthquake event are vital to understanding the subsequent wave propagation (BILHAM, 2005; SUBARYA *et al*., 2006). For example, numerical model simulations, combined with tidal-gauge and satellite altimetry data, reveal that wave amplitudes, directionality, and global propagation patterns of the Sumatra tsunami were primarily determined by the orientation and intensity of the offshore seismic line source and subsequently by the trapping effect of mid-ocean ridge topographic waveguides (TITOV *et al*., 2005). Visualization has been very useful in interpreting such simulation results, comparing them with data obtained from other measures (VAN PUYMBROECK *et al*., 2000), and helping scientists gain insights into such events (KRIKKE, 2005). In this paper we show the visualization of numerically generated post-seismic deformation by means of the computational power of burgeoning power due to graphics processing hardware.

The remainder of the paper is organized as follows. We briefly discuss graphics hardware in section 2. Details of our proposed interference inspired visualization method are given in section 3. Visualization applications of simulation data and discussions are described in section 4, followed by implementation details and performance measures in section 5. Finally we present our conclusions in section 6.

2. *Programmable Graphics Processing Units (GPUs)*

We exploit the recent advances in graphics hardware to accelerate our visualization applications. In this section, we briefly introduce the background of the current state of graphics hardware.

A Graphics Processing Unit (GPU) is a single chip processor with integrated rendering engines which is capable of processing a vast amount of geometry primitives. Although originally designed as a dedicated graphics rendering device for real-time rasterization of geometric primitives, modern GPUs are very efficient at solving computational intensive problems. The emergence of programmable GPUs makes general purpose computation (GPGPU) (GPGPU, 2006) feasible for a large range of complex scientific computational problems.

GPUs have been the most complex chips manufactured. One Nvidia GeForce 7800 GTX GPU chip has as many as 302 million transistors on board. The programmable pixel shaders of current GPUs can achieve a much higher theoretic peak performance than that of a currently available CPU. The ATI Radeon X1900 XTX has a claimed performance of 360 GFLOPS (Giga Floating Point Operations Per Second) of processing power (ATI, 2006), while a 3.7 GHz Intel Pentium4 SSE CPU can only achieve a theoretical peak of

14.8 GFLOPS (BUCK, 2004). A GeForce 7900 GTX chip can achieve a sustained 51.2 GB per second of memory bandwidth with a fill rate of 15.6 billion pixels per second and a vertex processing rate of 1.4 billion per second. As shown in Figure 2, it has 24 pixel shader pipelines and 8 vertex shader pipelines (NVIDIA, 2006). In general, current GPUs can offer more than an order of magnitude higher memory bandwidth and achieve up to an order of magnitude more computation speed up than current CPUs. Furthermore, GPU performance has improved faster than Moore's Law over the last decade. The graphics hardware performance is roughly doubling every six months. It takes the great advantages of utilizing GPUs for making scientific computation even more attractive.

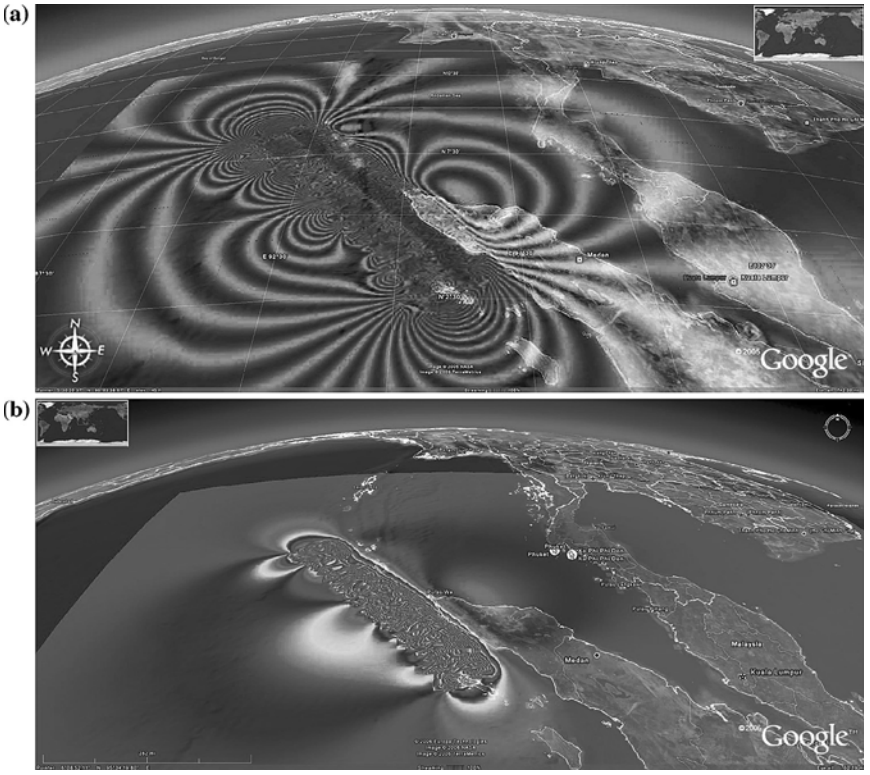


Figure 1
Our visualization of the vertical deformation of a fault rupture simulation of the Dec. 26th, 2004 Sumatra earthquake using GPUs with two different styles. (a) Red and blue areas are the rupture front and downdip slip regions with the most vertical displacement distances, respectively. Other regions have relatively smaller displacements. The interference pattern indicates the rate of change. Each interference period corresponds to 2.5 cm displacement in the vertical direction. (b) Using interference pattern. Each color cycle corresponds to 15.7 cm displacement in the vertical direction.

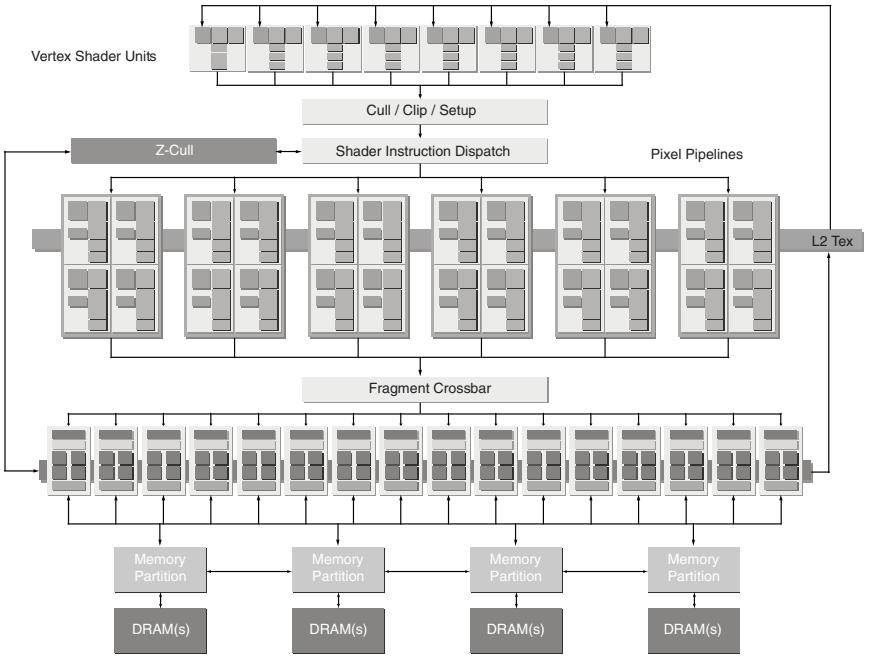


Figure 2
The chip layout of the Nvidia GeForce 7800 graphics processor. It has 24 pixel shader pipelines and 8 vertex shader pipelines.

GPUs are specialized computational devices. Computation in a GPU is done within the "Graphics Hardware Pipeline". The pipeline contains several steps in which the 3-D application sends a sequence of graphics primitives to the GPU to be processed by the programmable vertex processor and the programmable fragment processor subsequently. A representation of such a pipeline is shown in Figure 3.

To utilize GPUs to solve scientific computational problems, users need to transform the original computation setup to fit the graphics hardware pipeline, especially in the fragment processing stage. High level languages, (e.g., Cg, MARK *et al*., 2002; GLSL, OpenGL, 2006), have been developed to facilitate implementing sophisticated computation on hardware. Cg is a high-level 3-D graphics programming language developed by Nvidia and Microsoft. Microsoft's implementation of Cg is High-Level Shading Language (HLSL), and both versions are identical. Cg is based on C++ and uses C++ syntax but is specialized for GPUs. OpenGL Shading Language (GLSL), also known as GLSlang, is another high level shading language based on the C programming language and independent of graphics hardware. In our implementation of this work, we have employed the Cg language.

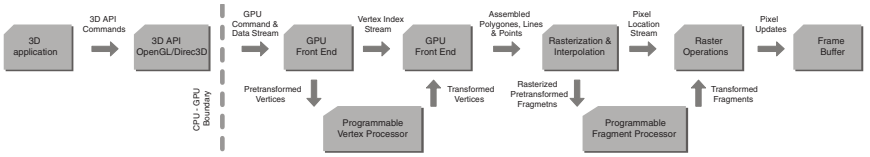


Figure 3
Graphics hardware pipeline.

Facilitated by the above high-level languages, programmable graphics processing units (GPUs) have been successfully applied to many domains beyond graphics, especially in scientific computation. We will demonstrate our applications of using GPUs to compute Synthetic InSAR images based on light interference for tsunami and earthquake simulation results.

3. Interference and INSAR

Interference phenomena are pervasive in nature. The physics of interference has been studied for more than 300 years. The first observation of interference was made by Robert Boyle in 1663 (GUENTHER, 1990), who observed what are now called Newton's Rings in which waves reflecting from two surfaces can interfere constructively and destructively. That interference pattern is associated with Newton because he performed a number of experiments on the effect. Young first generalized the principle of interference (YOUNG, 1802) and later developed his celebrated two-slit experiment (YOUNG, 1807). There are also rendering developments focused on simulating interference effects (DIAS, 1991; GONDEK, 1994; SMITS and MEYER, 1994; SUN, 1999). Our work is not intended to simulate the interference phenomena, but to use these patterns to visualize values and obtain results similar to InSAR images.

Interference is the interaction of two or more waves passing the same point. Constructive interference (Fig. 4(a)) occurs when the waves add in phase, producing a larger peak than either wave alone, whereas destructive interference (Fig. 4(b)) occurs when waves add out of phase, producing smaller peaks than one of the waves alone.

Consider the interference of two coherent monochromatic waves with wave functions ψ_1 and ψ_2,

$$\psi_n = \sqrt{I_n} e^{i(\mathbf{k}\mathbf{r}_n - \omega t + \varepsilon_n)}, \tag{1}$$

where I is irradiance, $\mathbf{k}$ is the propagation vector, $\mathbf{r}$ is the position vector, ω is the angular temporal frequency and ε is the initial phase. $n = 1, 2$. If they interfere coherently, then

$$\psi = \psi_1 + \psi_2 = \sqrt{I_1} e^{i(\mathbf{k}\mathbf{r}_1 - \omega t + \varepsilon_1)} + \sqrt{I_2} e^{i(\mathbf{k}\mathbf{r}_2 - \omega t + \varepsilon_2)}. \tag{2}$$

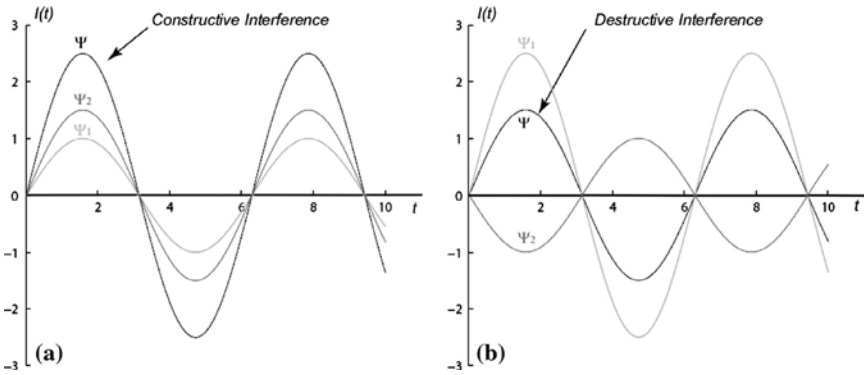


Figure 4
Light Interference. Red and blue curves are two coherent waves and the black curve is the interference results. (a) Constructive interference, (b) destructive interference.

The intensity is given by

$$I = |\psi|^2 = \psi\bar{\psi} = I_1 + I_2 + \sqrt{I_1 I_2}\cos\delta, \tag{3}$$

where

$$\delta = \mathbf{k}(\mathbf{r}_1 - \mathbf{r}_2) + (\varepsilon_1 - \varepsilon_2) = \mathbf{k}\Delta\mathbf{r} + \Delta\varepsilon. \tag{4}$$

δ is the phase difference arising from path length difference if the initial phases are the same.
If $I_1 = I_2 = I_0$ then

$$I = 2I_0(1 + \cos\delta) = 2I_0\cos^2\delta/2. \tag{5}$$

Constructive interference occurs when the phase difference between the two waves is an even integer multiple of π. Destructive interference occurs when the phase difference between the two waves is an odd integer multiple of π.

We apply the above interference patterns to visualizations of deformation on surfaces. We develop our visualization method by imitating the way interference is used to measure short distance variations by Newton's Rings. In our visualizations, light path differences are defined by the surface values to be visualized. These path differences then modulate the lighting intensity to generate the interference patterns. Assume for one point, the movement or deformation is d. We consider two coherent light waves traveling at a distance proportional to **d**. We define $\delta = 2\pi v\mathbf{d}$, where v is the **frequency** parameter used to control the frequency of the interference pattern. Then the interference intensity I will vary according to $1 + \cos(2\pi v\mathbf{d})$. Since the modulation term is a function with a period of 2π, each pair of peaks in the interference pattern indicates a difference of $1/v$ in **d**. We will refer to the value $1/v$ as the **resolution** of the subsequent interference rendering. This property is very useful for visualization purposes.

There are a few variations we could apply to use such interference patterns for visualization. In Figure 5, we demonstrate such visualizations on an artificially created data set. The data set is a plane with two peaks generated with the peak function in Matlab. As shown in Figure 5(a), the surface patch is a square with a size 255×255. The vertical deformation of two peaks varies from -0.0655 to 0.0811. The vertical elevation in the figure is exaggerated by 1000 times to make it visible.

As illustrated in Figure 5(b), we modulate the deformation value to an interference pattern and display the modulated values as intensity values in gray levels. With a v value of 100, (**resolution 1/v = 0.01**) each neighboring band indicates a deformation difference of 0.01. To display the modulated values, we scale them to be in the range of [0,1] by using the formula of $((1 + \cos(2\pi v\mathbf{d})) / 2)$. The same scaling is always applied through out this paper.

In Figure 5(c), we apply different frequencies for each color channel to visualize the deformation. Three frequencies of (100, 110 and 120) are applied. In this way, the result closely resembles the interference patterns generated with white light. The color order of transition regions indicates the decrease or increase of the value to be visualized. We also apply a color map (Fig. 5(g)) on top of the interference pattern as shown in Figure 5(d) with a formula of $colormap(normalize(\mathbf{d}))\cdot((1 + \cos(2\pi v\mathbf{d})) / 2)$. The function $colormap(\,)$ maps

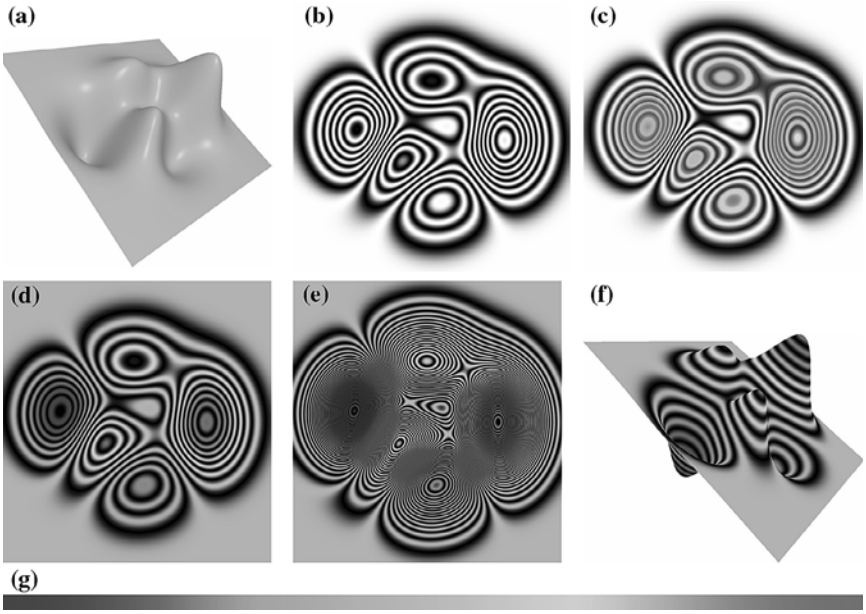


Figure 5
Interference patterns generated with variations. (a)–(d) all represent a surface with two peaks rendered in (e). In (f) the interference patterns are mapped to the 3-D surface.

values between 0 and 1 to a specified color map. The function *normalize*(**d**) linearly map **d** to range of [0,1]. Such an operation is also applied for Figure 1(b). Figure 5(e) has the same configuration as Figure 5(d) except that a higher frequency v value of 500 is applied. More details in low deformation regions can be revealed with higher frequency values. Such a visualization pattern is applied in Figure 1(a). In Figure 5(f), the rendered interference pattern is mapped on the exaggerated geometry model. We can also apply a color map to each cycle of interference pattern with formula of $colormap((1 + \cos(2\pi v\mathbf{d})) / 2)$. Since the function $(1 + \cos(2\pi v\mathbf{d})) / 2)$ is not a monotonic function, a color map will be mapped twice within one single interference cycle in two opposite directions. We can convert it into monotonic function using a formula of $colormap(((1 + \cos(\pi v\mathbf{d}))/2) \cdot sign((1 + \sin(\pi v\mathbf{d})) / 2))$. Function $sign(x)$ returns the value 1 for positive variables while it returns –1 for negative variables. With the factor $sign((1 + \sin(\pi v\mathbf{d})) / 2)$, the whole function has a period of π. In this case, each color period still indicates a difference of $1/v$ in **d**. This formula is used to visualize an earthquake simulation in Figure 1(b). In all the above renderings, the computation of interference modulation is performed in GPUs. We will discuss the details of the computation at the end of this paper after we present all visualization results.

The above interference-based method generates similar images to Interferometric Synthetic Aperture Radar (InSAR) (Fig. 6). Indeed, interference phenomena have formed the underlying theory for InSAR in unveiling earth movements. Radar satellites constantly shoot beams of radar waves towards the Earth and record their return after they have bounced back off the Earth's surface (MASSONNET, 1993). The phase shift of the bounced wave could be used to indicate the deformation that has occurred. Such techniques are utilized to provide direct and precise measurement of the vertical changes in ground level. The strength of InSAR lies in its ability to provide observations of ground displacements with a precision as high as a few millimeters with 20 meters of

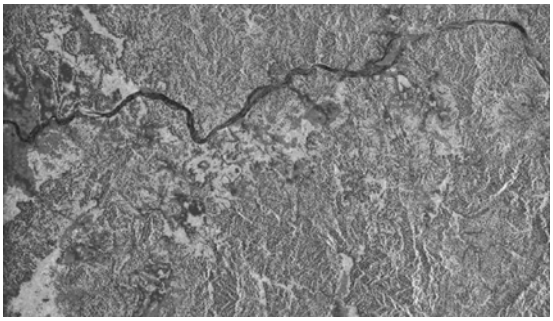


Figure 6
A sample interferometric image obtained with the Synthetic Aperture Radar (SAR) instrument on-board the ESA ERS satellites. Each cycle of the colors represents a change in the ground height which is dependent on the satellite geometry. The image is from the European Space Agency. (http://earth.esa.int/images/INSI/).

spatial resolution and cover hundreds of kilometers in spatial extent. InSAR has revolutionized the seismic and volcanic study.

Figure 6 is a sample interferometric image obtained with the Synthetic Aperture Radar (SAR) instrument on-board the ESA ERS satellites. The data are collected when the ERS-1 and ERS-2 satellites were flying in "tandem-mode" observing the same part of the ground with a 1-day interval. Each cycle of the colors (for example, going from yellow to purple to turquoise and back to yellow again) represents a change in the ground height, which is dependent on the satellite geometry. The introduction of InSAR makes the production of displacement maps of an entire region with sub-centimeter accuracy possible. The InSAR process involves the acquisition of two or more single images from an airplane or satellite. The radar sensors carried aboard remote-operates in the microwave range of the electromagnetic spectrum. To create an interferogram, two images must be acquired at different points in time.

After the powerful 2004 Sumatra earthquake, significant efforts have been devoted to measuring, simulating and visualizing the scenario and understanding this largest seismic event since the introduction of digital seismology (Ammon *et al*., 2005; Krikke, 2005; Lay *et al*., 2005). Steven Ward (2006), a geophysicist produced an animation of the Indian Ocean tsunami hours after the event. Ammon *et al*. (2005) studied the rupture process of the Sumatra-Andaman Earthquake. The research results showed that the earthquake was initiated slowly, with small slip and a slow rupture speed. Then the rupture expanded quickly toward the northwest, extending 1200 to 1300 kilometers along the Andaman trough. Lay *et al*. (2005) and Bilham (2005) also indicated a slow slip. The InSAR imaging is also perfect for measuring the Earth's deformation during a tsunamogenic earthquake event. There is a need to carry out a rapid comparison between modeling and observations from InSAR images. Responding to such demands, our research focuses on visualizing the simulated rupture process of the earthquake. A major motivation for developing this interference inspired visualization is its similarity to InSAR, which is a widely applied remote sensing technique for measuring deformation of the Earth's surface. The interference patterns generated in our approach are easier to understand for geophysicists who have previous experience with InSAR images. We generate such Synthetic InSAR images from the simulation results to allow researchers a quick assessment of the results. It is desirable for researchers to have quick access to InSAR images for comparison purposes. Since conversion of an InSAR image to an elevation map is still computationally expensive, our method provides a faster way for direct comparison.

4. Visualization for Tsunami Event Simulations

In this section we illustrate how our interference inspired method could be applied to various geophysical simulation data sets to better understand tsunami events. For the sake of completeness, we also briefly discuss the models used in these simulations.

4.1. Case 1: Visualization of a Hybrid Simulation Model

In this first numerical simulation, we use a hybrid method which is basically set up from finite-element methods in a 2-D area with a Fourier series decomposition as the third dimension. We deal with the elastic problem for simplicity.

$$\nabla \cdot \tau + \rho_0[(\nabla \cdot \boldsymbol{u})\boldsymbol{g}_0\boldsymbol{e}_z - \nabla(\boldsymbol{g}_0\boldsymbol{e}_z \cdot \boldsymbol{u})] = 0, \tag{6}$$

$$\tau - \lambda(\nabla \cdot \boldsymbol{u})\boldsymbol{I} - 2\mu\varepsilon(\boldsymbol{u}) = 0. \tag{7}$$

We can decompose the 3-D problem this way if the equation coefficients are independent in one horizontal direction and also if it is possible to rotate the domain according to the rupture. In this case the spectral decomposition is used as the y axis and the rupture function is placed to the computation domain along this axis. The rupture function has an important role in our solution technique, thus we explain this part in greater detail.

This function describes displacements on the 2-D fault itself. It has to be defined to be continuous on some part of the 3-D domain with non-zero measurement because of our mathematically weak problem formulation. It means this function is dependent on all three coordinates $f(x, y, z)$ and when we transform it to the spectral domain we have function $f(x, k_y, z)$ where k_y is now only a parameter. After solving simultaneous equations (6) to transform to the spectral domain we solve just 2-D problems with multiple parameters k_y. We just use backward transformation of 2-D FEM solutions to transfer the solution back to the 3-D domain. The numerical simulations were performed by, 2-D finite-element code together with the Fourier decomposition method to create adapted mesh. The GMRES method with ILU preconditioning is used as the linear solver. The code was written in C-programming language. To increase code effectiveness, OpenMP parallelization was used for the backward transformation and for the creation of surface mesh.

The domain of the subduction model is sized 1000 km $\times$ 1000 km $\times$ 600 km. The fault strike is 0°, the dip is 60° and the rake is $-90°$. The fault is 50 km deep and amplitude is 10 meters, its length in y coordinate is 100 km and its second dimension is 50 km. For the Fourier decomposition 600 k_y parameters were used. The computation mesh has 563 $\times$ 564 $\times$ 100 grid points. With respect to the concentration of points close to the rupture, it was worthwhile to increase the number of points to 1001×1001 during recalculation of equidistant rotated surface mesh. The computation was run at the Minnesota Supercomputing Institute on Altix workstations, where parallelism was utilized for the most time-consuming parts of the computations.

In Figure 7 we visualize the computational results using interference patterns. A color map (Fig. 7(d)) is mapped to each period of the interference pattern. Figures 7(a), (b) and (c) show the x, y and z components of the fault deformation simulations, respectively. In each group, the top row shows the interference-based visualization with v value of 2, 4, 8, 16 and 32, respectively. The bottom row images are zoom-ins of the center part of their corresponding top row images. Note that the boundary between the color blue and red is due to the two ends of the texture for which we used different colors. If a cyclic texture is

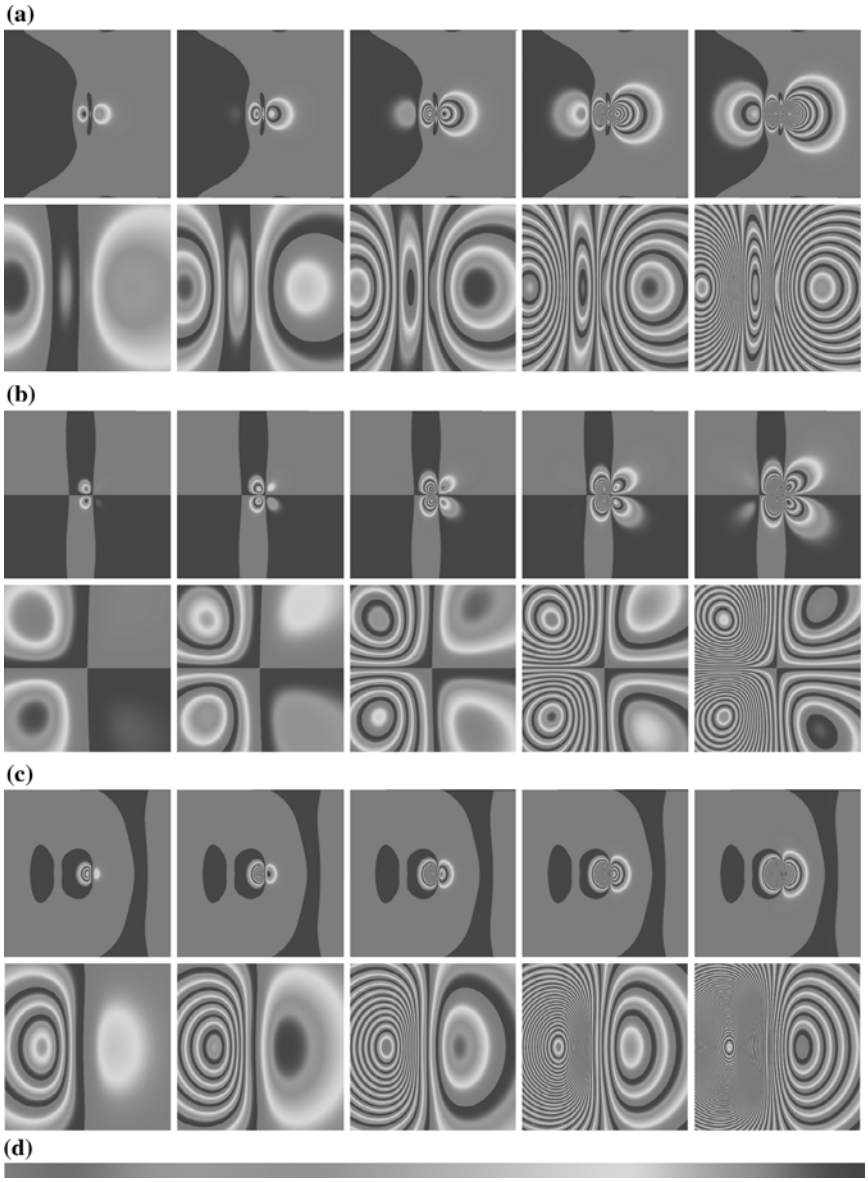


Figure 7
(a) x component, (b) y component, (c) z component of the fault deformation simulated by a Hybrid Simulation Model. In each group, the top row shows the interference-based visualization with v value of 2,4,8,16 and 32, respectively. The bottom row images are zoom-ins of the center part of their corresponding top row images.

used (in Fig. 9), there will be no such sharp color boundaries. Note in all examples shown in this paper, the displacement calculations are carried out in a Cartesian coordinate system. When the curvature of the Earth is considered, it is also possible to do the interference pattern computation in a spherical coordinate system, in which case, latitude and longitude are used to substitute vertical and horizontal values in the Cartesian geometry.

4.2. Case 2: Earth Rupture Simulations of Sumatra Earthquake

In the second case, we use the PSGRN/PSCMP code (OKADA, 1992; WANG *et al.*, 2006) to calculate coseismic and postseismic deformation for layered rheological media. In recent years continuous GPS data sets have been frequently used. With this data researchers can study the coseismic and postseismic deformation after the earthquake. In fact, the Earth can be described using two different material models in which the upper layer is elastic and the deeper is viscoelastic. The postseismic deformation is a very complicated process, especially as some factors are neglected in relaxations with long time duration.

The PSGRN/PSCMP code is based on the extension of the elastical dislocation theory with the gravity effect. The theory is applied to the governing equations for the infinitesimal static deformation in a self-gravitating, hydrostatically prestressed Earth. This approach extended the earlier elastic dislocation theory with the gravity effect reflected in the following equations:

$$\nabla \cdot \Gamma + \rho \nabla(\mathbf{u} \cdot \mathbf{g}) - \mathbf{g} \nabla \cdot (\rho \mathbf{u}) = 0, \tag{8}$$

$$\nabla^2 \psi - 4\pi G \nabla \cdot (\rho \mathbf{u}) = 0, \tag{9}$$

where Γ is the Lagrangian incremental stress tensor, $\mathbf{u}$ is the displacement vector, ψ is the Eulerian incremental potential, g is acceleration due to gravity, G is the gravitational constant, and ρ is the density. For an isotropic and elastic medium, Hook's linear constitutive relation between the stress and strain is valid,

$$\Gamma = (\lambda \nabla \mathbf{u})\mathbf{I} + \mu[\nabla \mathbf{u} + (\nabla \mathbf{u})^T], \tag{10}$$

where λ and μ are the two Lame constants, $\mathbf{I}$ is the unit tensor, and $(\nabla \mathbf{u})^T$ denotes the tensor transpose of $\nabla \mathbf{u}$. With consideration of various practical mathematical and geophysical issues, the governing equation can be reduced to:

$$\nabla \cdot \Gamma + \rho g \nabla u_z - (\rho \nabla \cdot \mathbf{u} + \frac{\rho^2 g}{\kappa} u_z) g e_z = 0, \tag{11}$$

where κ is constant and $\mathbf{g} = g e_z$ has been used. This approach considers the coupling between the deformation and the Earth's gravity field and produces the postseismic deformation data. The output includes the complete deformation field consisting of 3 displacement components, 6 stress (strain) components and 2 tilt components, and the

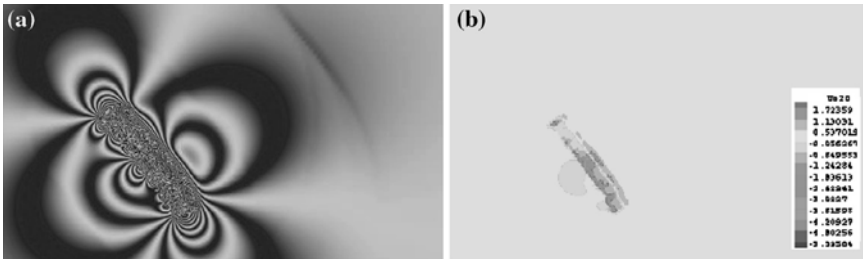


Figure 8
Comparing Interference Visualization with linear color mapping. (a): Interference patterns modulate the color map. (b): Color mapping using Tecplot. The field is a simulation of the displacements of the fault of the Sumatra earthquake on Dec. 24, 2004.

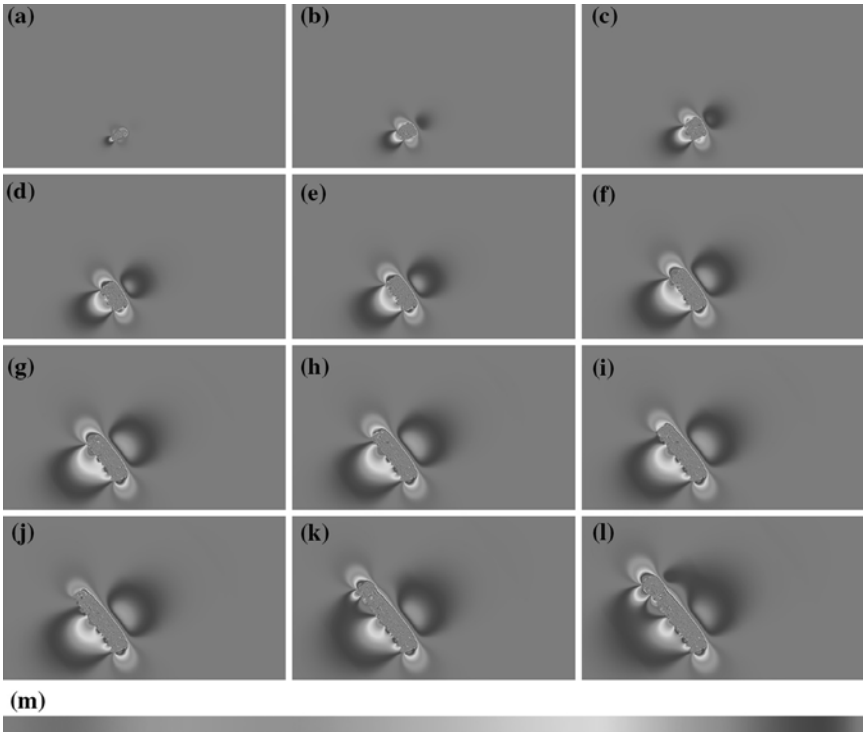


Figure 9
Interference Pattern Visualization of z displacement component of a simulation of a coseismic fault of a region of 0–14°N, 86–110°E. (a)–(l) 12 time steps are visualized. (m) The color map used in this figure.

geoid and gravity changes. For details of this computational model and implementation, please refer to Wang's recent work (WANG *et al.*, 2006). In this visualization work, we are mainly interested in visualizing the displacement components generated by such simulations.

The study field is a much larger region simulation for the Sumatra earthquake of Dec. 26, 2004. The fault length of the earthquake was 1300 km. The magnitude was M_w 9.3. The epicenter coordinates were 3.1°N;95.7°E. The fault strike was 310° and the dip was 10°. The fault width was 100 km. The displacements of the fault were computed by coseismic dip. The program code PRCMP3.2 and PSGRN3.2 was written by RONG-JIANG WANG (2006). We set the material of the fault to be elastic. The size of the field is 0–14°N, 86° –110°E. The rupture of the fault is 23 time steps along the strike and 7 layers in the vertical direction. The grid is 121×121 evenly spaced points. The maximal displacement is 13 m.

We visualize the vertical displacement component of such time-varying deformation surfaces with the parameter of $v = 50$, equivalent to 0.02 m in resolution. In Figure 8 we compare the visualization result using traditional color mapping and the interference pattern. Our method can clearly illustrate the details of deformation computed. In Figure 8(b), a color map is used to map the z displacement of the fault in the Sumatra earthquake simulations using TecPlot. In Figure 8(a), we modulate the visualization color mapped image with our interference pattern. The rings indicate the rate of the displacement change. The red color regions have positive vertical displacements, and the green color regions have negative vertical displacements. It is obvious that traditional color mapping methods problematically deal with such data sets due to a lack of differentiable color levels. The interference pattern clearly visualizes the deformation propagation in the surrounding areas.

In Figures 9(a)–(l), we illustrate 12 different time steps of the fault simulation visualization. The vertical displacements are visualized. The color mapping used in this visualization is shown in Figure 9(m). The same visualization is also mapped to Google Earth to show the geological context that is shown in Figure 1(a). From the visualization, it is clear that the rupture developed from the epicenter and expanded quickly toward the northwest which is the *a priori* constraint we set on the modeling (AMMON, 2005; BILHAM, 2005; LAY *et al.*, 2005). The fault can be categorized as a reverse fault. Data collected at 60 Global Positioning System (GPS) sites in Southeast Asia show the crust deformation caused by the 26 December, 2004 Sumatra-Andaman earthquake at an unprecedented large scale (VIGNY *et al.*, 2005). Small but significant coseismic jumps are clearly detected more than 3,000 km from the earthquake epicenter. Such small deformations are very obvious in our visualization. The nearest sites, still more than 400 km away, show displacements of 10 cm or more. Here we show how the rupture plane for this earthquake must have been atleast 1,000 km long and that nonhomogeneous slip is required to fit the large displacement gradients revealed by the GPS measurements. Our kinematic analysis of the GPS recordings indicates that the centroid of the released deformation is located at least 200 km north of the seismological epicenter. It also provides evidence that the rupture propagated northward with sufficient speed such that stations in northern

Thailand reached their final positions less than 10 min after the earthquake, ruling out the hypothesis of a silent slow seismic rupture.

4.3. Case 3: Tsunami Wave Simulation of South China Sea

We have also applied our visualization method to the computation of wave height generated by the tsunamogenic earthquake in the South China Sea (LIU *et al*., 2007). The trenches of the South China Sea stretch in a north to south direction, from south of the Taiwan Island, passing west of Luzon and extending to the west of Mindanao. The sea-bottom topography of the Manila trench is very complicated. The large variation of water depth and sea-bottom topography indicates the possibility of an earthquake-tsunami in that region. The Manila trench is the fault-graben trough dominated by normal fault, on the boundary of the Epicontinental graben system of the South China Sea and the Philippine Island-arc fault-fold system. The western part of the Manila trench fault zone is formed by the steepened fault and depression of the Central oceanic basin of the South China Sea. In terms of the fault movement, the Manila trench fault zone is the region most likely to have experience a tsunami in the South China Sea.

Based on such reasoning, a simulation of the sea-water wave height in that region has been conducted. As shown in Figure 10(a), we apply interference based visualization to the wave-height data computed. The wave height is stored in its absolute value. A color texture is mapped to each period of the interference pattern. The regions with denser interference fringes are associated with higher wave height and in larger danger of a tsunami. Figure 10(b) is a color texture mapped to the whole data range and then modulated by the interference pattern. Figure 10(c) uses a direct color mapping without interference pattern enhancement.

4.4. Integrating Interference Visualization Results with Google Earth

The tsunami and earthquake data sets visually at our disposal are associated with specific regions of the globe. It is therefore very important that we have the contextual geological information necessary for detailed visualizations. Therefore, we integrate our visualization results into GOOGLE EARTH (2007) using the image overlay function provided by its software. Google Earth is a server-client based virtual globe program (http://earth.google.com/). It maps the entire earth by pasting images obtained from satellite imagery, aerial photography and GIS over a 3-D globe. Google Earth allows users to search for addresses, enter coordinates, or simply use the mouse to browse to a location. Google Earth also has digital terrain model data collected by NASA's Shuttle Radar Topography Mission. Users can directly view the geological features in three-dimensional perspective projections, instead of as 2-D maps.

Many scientists have found Google Earth to be a powerful tool for scientific visualization (BUTLER, 2006; NOURBAKHSH, 2006). In our experiment, we use the image overlay function of the Google Earth software to integrate our visualization results into a

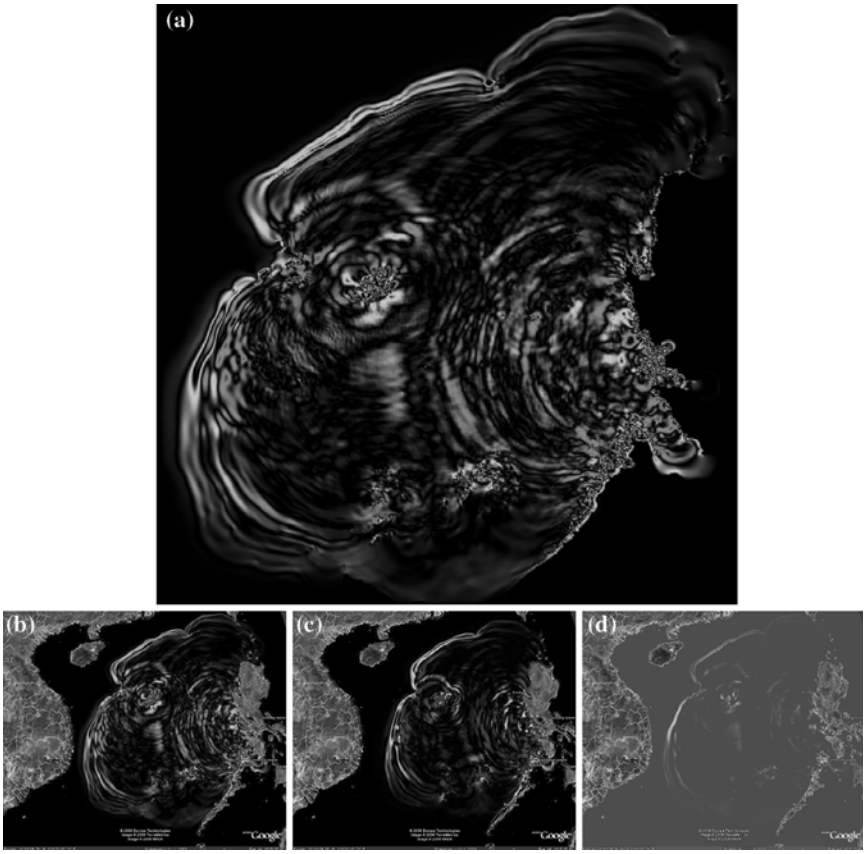


Figure 10
Visualization of wave height generated by a tsunamogenic earthquake in the South China Sea. (a) and (b) color mapped to each period of the interference pattern, (c) color mapped to the whole data range and then modulated by interference pattern, (d) direct color map.

virtual globe context. Figure 1 illustrates the integration of our visualization of the fault simulation of the Sumatra earthquake with Google Earth. The interference pattern visualization images are overlayed in Google Earth. The geological context of the terrain is rendered in 3-D and local cities are annotated. Such visualizations help users to associate the scientific data with a geological context which is familiar to ordinary users. In Figure 10, 3 other examples are also shown.

Figure 11 shows how users can navigate the virtual globe with integrated visualizations. Note that in the most zoomed-out image, Google Earth provides the recorded earth quake information indicated by a red dot in the studied region. Such

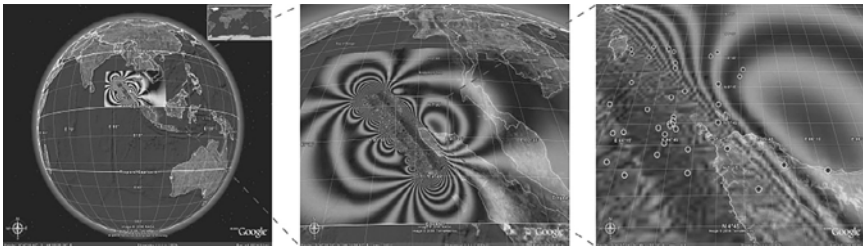


Figure 11
Navigation of Google Earth with integrated visualization results.

information is valuable for researchers seeking to understand the event. The software provides a convenient way for researchers to share their results. The visualization results we embed in Google Earth could be in a data exchange format (KML) together with the geological context information. Such meta data could be directly opened by other parties who also have the software. The Google Community service could make these visualization results available to the public, allowing the sharing of research results between research groups.

5. Implementation Details and Performance

All visualization computations including rendering have been performed on a Dell Precision 530 workstation with single Intel Xeon 2.20 G Hz CPU, 1 GB RAM, 4 × AGP motherboard and a 256 MB GeForce 6800 Ultra graphics card. Cg is used for GPU programming. We implement our rendering pipeline with deferred shading to improve interactivity for large data set visualization. The concept of deferred shading was introduced by DEERING *et al*. (1988) and was later used by PixelFlow (MOLNAR *et al*., 1992.) With advances in commodity graphics hardware, deferred shading is becoming practical for normal rendering (NVIDIA, 2004; SHISHKOVTSOV, 2005).

For the geophysical simulation data we are dealing with, there are multiple propensities associated with each grid point. Before the deferred shading, we render all values into multiple textures illustrated as the first rendering step in Figure 12. Current Multiple Render Targets (MRT) features enable us to simultaneously render up to four textures in a single pass. With one color channel in a texture reserved for the valid rendered area, there are 15 channels available for information storage. We use floating point textures to store the rendered results. After the first rendering pass, we create intermediate textures with color channels storing deformation values. Users can freely select the surface properties they wish to visualize in the second deferred shading pass illustrated as the second rendering step in Figure 12. The program computes the

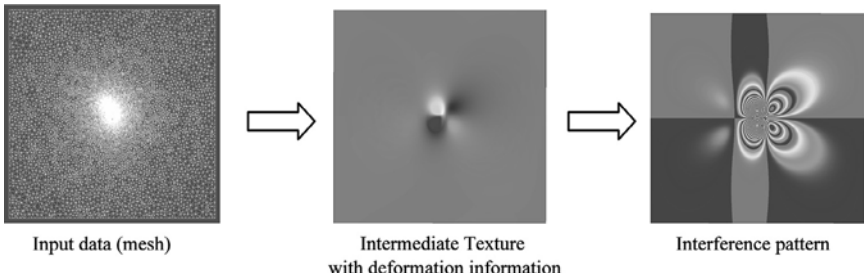


Figure 12
Rendering pipeline for interference based-visualization.

Table 1

Performance of our rendering method for various models. Two different screen sizes are used to test the performance. T_{geo} is the run time before the deferred shading. T_{def} is the time for the deferred shading.

Model	Grid Size	Screen Size	T_{geo}(ms)	T_{def} (ms)
Peaks (Fig. 5)	256 × 256	512 × 512	22	5.5
	256 × 256	1200 × 900	23	8
Sumatra earthquake (Fig. 11)	121 × 121	512 × 512	11	5.6
	121 × 121	1200 × 900	12	7

interference patterns using a fragment shader together with an optional color map lookup. We leverage the current commodity graphics hardware to perform the computation and rendering. Our method is very efficient. Table 1 shows the timing for the rendering of different data sets. Two different screen sizes are used to test the performance. T_{geo} (in milliseconds) is the run time before the deferred shading. T_{def} (in milliseconds) is the time for deferred shading. Note that because the deferred shading is executed on the final frame image, it is independent of model resolution and dependent only on the screen resolution as shown in the table.

6. Conclusions and Future Endeavors

We present a novel visualization method for visualizing dynamically changing surfaces, which is one type of highly dynamic range data (YUAN *et al.*, 2005, 2006, 2007). Our method is superior to traditional color mapping and can illustrate subtle but important surface value variations. We can interactively visualize surface deformation patterns by leveraging commodity graphics hardware to achieve interactivity. Our method is simple, easy to implement and very effective at illustrating subtle but critical surface changes to surface features in deformation processes. As compared with the

visualization using existing software tools, we gain advantage in the quality and speed with graphics hardware acceleration. The integration of our seismic visualization with Google Earth provides an easy way to include global geological context. Users can easily exchange research results using this tool.

It is possible to combine interference patterns with other existing visualization methods. In the future we will test our method on more complex models which incorporate bends and other geometric complexity. We also plan to extend our work to other simulation data sets. For one immediate step, we plan to use the same setting of the InSAR images and generate interference patterned images for direct comparison with InSAR satellite images. We would also like to compare our calculated displacement values with the observational values.

Acknowledgement

The authors would sincerely thank Google Inc. for providing the software Google Earth Pro. to Peking University through the Google Earth Education Initiative. We wish to thank the anonymous reviewers for their valuable suggestions. David Yuen would like to acknowledge support from Math-Geo and ITR grants from the National Science Foundation. Other funding includes NSF CAREER ACI-0238486, DMS-0528492 and the Army High Performance Computing Research Center under the auspices of the Department of the Army, Army Research Laboratory cooperative agreement number DAAD19-01-20014. We especially thank Rongjiang Wang for his assistance on simulation data generation. We also appreciate our discussions with Professor Huai Zhang. In the images of Figures 1, 10 and 11, the copyright of the geological contexts of our visualization results in the above images belongs to the Google Inc. and other geological information providers.

References

Ammon, C.J., Ji, C., Thio, H-K., Robinson, D., Ni, S., Hjorleifsdottir, V., Kanamori, H., Lay, H., Das, S., Helmberger, D., Ichinose, G., Polet, J., and Wald, D. (2005), *Rupture Process of the 2004 Sumatra-Andaman Earthquake*, Science *308*, 1133–1139.

ATI Technologies Inc. (2006), http://www.ati.com.

Buck, I. (2004), *GPGPU: General-Purpose Computation on Graphics Hardware–GPU Computation Strategies and Tricks*, ACM SIGGRAPH Course Notes (Aug. 2004).

Bilham, R. (2005), *A flying start, then a slow slip*, Science *308*, 1126–1127.

Butler, D. (2006), *Virtual globes: The web-wide world*, Nature *439*, 776–778.

Deering, M., Winner, S., Schediwy, B., Duffy, C., and Hunt, N., *The triangle processor and normal vector shader: A VLSI system for high performance graphics*. In *SIGGRAPH'88* (New York, NY, USA, ACM Press 1988) pp. 21–30.

Dias, M.L. (1991), *Ray tracing interference color*, IEEE Computer Graphics and Applications *11* (2), 54–60, March/April.

Google Earth (2007), http://earth.google.com/.

GONDEK, J.S., MEYER, G.W., and NEWMAN, J.G., *Wavelength dependent reflectance functions*. In *SIGGRAPH'94* (New York, NY, USA, ACM Press 1994) pp. 213–220.

GPGPU (2006), *General-purpose computation on GPUs*, http://www.gpgpu.org.

GUENTHER, R.D., *Modern Optics* (Wiley 1990).

KRIKKE, J. (2005), *Near Real-time tsunami computer simulations within reach*, IEEE Computer Graphics and Applications *25* (5), 16–21, Sept/Oct.

LAY, T., KANAMORI, H., AMMON, C.J., NETTLES, M., WARD, S.N., ASTER, R.C., BECK, S.L., BILEK, S.L., BRUDZINSKI, M.R., BUTLER, R., DESHON, H.R., EKSTROM, G., SATAKE, K., and SIPKIN, S. (2005), *The Great Sumatra-Andaman Earthquake of 26 December 2004*, Science *308*, 1127–1133.

LIU, Y., SANTOS, A., WANG, S.M., SHI, Y., LIU, H., and YUEN, D.A. (2007), *Tsunami hazards along Chinese coast from potential earthquakes in South China Sea*, Phys. Earth Planet. Inter. (PEPI) *163*, 233–244.

MARK, W.R., GLANVILLE, R.S., AKELEY, K., and KILGARD, M.J., *Cg: A System for Programming Graphics Hardware in a C-like Language*. In *SIGGRAPH'02* (New York, NY, USA, ACM press 2002) pp. 896—907.

MASSONNET, D., ROSSI, M., CARMONA, C., ADRAGNA, F., PELTZER, G., FEIGL, K., and RABAUTE, T. (1993), *The displacement field of the Landers earthquake mapped by radar interferometry*, Nature *364*, 138–142.

MOLNAR, S., EYLES, J., and POULTON, J. *Pixelflow: High-Speed Rendering Using Image Composition*. In *SIGGRAPH'92* (New York, NY, USA, ACM Press 1992) pp. 231–240.

NVIDIA CORP. (2004), *Deferred Shading*, http://developer.nvidia.com/.

NVIDIA CORP. (2006), http://www.nvidia.com.

NOURBAKHSH, I., SARGENT, R., WRIGHT, A., CRAMER, K., MCCLENDON, B., and JONES, M. (2006), *Mapping disaster zones*, Nature *439*, 787–788.

OKADA, Y. (1992), Internal Deformation Due to Shear and Tensile Faults in a Half-space. Bulletin of the Seismological Society of America, *82* (2), 1018–1040.

OPENGL SHADING LANGUAGE (2006), http://www.opengl.org/documentation/glsl/.

PARK, J., SONG, T.A., TROMP, J., OKAL, E., STEIN, S., ROULT, G., CLEVEDE, E., LASKE, G., KANAMORI, H., DAVIS, P., BERGER, J., BRAITENBERG, C., VAN CAMP, M., LEI, X., SUN, H., XU, H., and ROSAT, S. (2005), *Earth's free oscillations excited by the 26 December 2004 Sumatra-Andaman Earthquake*, Science *308*, 1139–1144.

SHISHKOVTSOV, O., Chapter 9. *Deferred shading in STALKER*. In *GPU Gems II: Programming Techniques for High-Performance Graphics and General-Purpose Computation* (Addison Wesley, 2nd edition 2005) pp. 143–545.

SMITS, B.E. and MEYER, G. (1994), *Newton's Colors: Simulating Interference Phenomena in Realistic Image Synthesis*. In Proc. *Eurographics Workshop on Photosimulation, Realism and Physics in Computer Graphics*, pp. 185–194.

STEIN, S., and OKAL, E. (2005), *Seismology: Speed and size of the Sumatra earthquake*, Nature *434*, 581–582.

SUBARYA, C., CHLIEH, M., PRAWIRODIRDJO, L., AVOUAC, J-P., BOCK, Y., SIEH, K., MELTZNER, A.J., NATAWIDJAJA, D.H., and MCCAFFREY, R. (2006), Plate-boundary deformation associated with the Great Sumatra-Andaman earthquake, Nature *440*, 46–51.

SUN, Y., FRACCHIA, F.D., CALVERT, T.W., and DREW, M.S. (1999), *Deriving spectra from colors and rendering light interference*, IEEE Computer Graphics and Applications *19* (4), 61–67.

TITOV, V., RABINOVICH, A.B., MOFJELD, H.O., THOMSON, R.E., and GONZALEZ, F.I. (2005), *The global reach of the 26 December 2004 Sumatra tsunami*, Science *309*, 2045–2048.

USAID (2006), http://www.usaid.gov/locations/asia near east/tsunami/trm.html.

VAN PUYMBROECK, N., MICHEL, R., BINET, R., AVOUAC, J.P., and TABOURY, J. (2000), *Measuring earthquakes from optical satellite images*, Appl. Optics Inform. Processing *39* (23), 3486–3494.

VIGNY, C., SIMONS, W.J., ABU, S., BAMPHENYU, R., SATIRAPOD, C., CHOOSAKUL, N., SUBARYA, C., SOCQUET, A., OMAR, K., ABIDIN, H.Z., and AMBROSIUS, B.A.C. (2005), *Insight into the 2004 Sumatra-Andaman earthquake from GPS measurements in Southeast Asia*, Nature *436*, 201–206.

WANG, R., LORENZO-MARTIN, R., and ROTH, F. (2006), *PSGRN/PSCMP - A new code for calculating co- and post-seismic deformation, Geoid and gravity changes based on the viscoelastic-gravitational dislocation theory*, Comp. Geosci., *32* (4), 527–541.

WARD, S.N. (2006), http://es.ucsc.edu/ ward/.

WEST, M., Sachez, J.J., and McNutt, S.R. (2005), *Periodically triggered seismicity at Mount Wrangell, Alaska, after the Sumatra earthquake*, Science *308*, 1144–1146.

YOUNG, T. (1802), *A Syllabus of A Course of Lectures on Natural and Experimental Philosophy*, London: Royal Institution.

YOUNG, T. (1807), *A Course of Lectures on Natural Philosophy and the Mechanical Arts*, London: J. Johnson.

YUAN, X., LIU, Y., CHEN, B., YUEN, D.A., and PERGLER, T. (2007), *Visualization of high dynamic range data in geosciences*, Phys. Earth Planet. Inter. (PEPI) *163*, 312–320.

YUAN, X., NGUYEN, M.X., CHEN, B., and PORTER, D.H. (2005), *High dynamic range volume visualization*, Proc. IEEE *Visualization 2005*, pp. 327–334. .

YUAN, X., NGUYEN, M.X., CHEN, B., and PORTER, D.H. (2006), *HDR VolVis: High dynamic range volume visualization*, IEEE Transact. Visualization and Computer Graphics *12* (4), 433–445.

(Received November 1, 2006, revised April 10, 2007, accepted May 30, 2007)

Pure appl. geophys. 165 (2008) 693–709
0033–4553/08/030693–17
DOI 10.1007/s00024-008-0329-5

Pure and Applied Geophysics

The Stress Accumulation Method and the Pattern Informatics Index: Complementary Approaches to Earthquake Forecasting

KRISTY F. TIAMPO,[1] DAVID D. BOWMAN,[2] HARMONY COLELLA,[3] and JOHN B. RUNDLE[4]

Abstract—In this work, we provide a joint study of the stress accumulation method (SAM) (KING and BOWMAN, 2003) and the Pattern Informatics (PI) index (TIAMPO *et al*., 2002b). We examine the theoretical underpinnings for the similarities between the two techniques, as well as the differences in their application. The SAM technique is employed to determine likely mechanisms for smaller areas of increased probability identified by the PI index, while a modified version of the PI index can be used to locate regions where the smaller magnitude associated with the anomaly is below the resolution of the SAM. Finally, we present three case studies from different regions of the San Andreas fault system to illustrate both their complementary nature, as well as the advantages to combining them in one synthesized analysis.

Key words: Seismicity, earthquake forecasting, seismic hazard, stress change.

1. Introduction

In recent years, two new techniques for the study of seismicity and its temporal evolution have been developed and applied to the forecasting of large events. The pattern informatics (PI) method is based upon recent observational evidence that earthquake faults are characterized by strongly correlated space-time dynamics, resulting in seismicity pattern structures that are non-local, a consequence of strong correlations in the underlying dynamics. The second, the stress accumulation method (SAM), models the evolution of the stress field relative to the failure stress. Both methods are based on the fundamental assumption that changes in the a-value, the y-intercept of the Gutenberg-Richter relation, are a measure of the underlying change in stress, and can be used as a sensor for current and future seismic moment release (TURCOTTE, 1997; RICHTER, 1958).

While a long-sought goal of earthquake research has been the reliable forecasting of the great events, until recently very little progress has been made in developing a successful, consistent methodology (RICHTER, 1958; SCHOLZ, 1990; GELLER *et al*., 1997).

[1] Department of Earth Sciences, University of Western Ontario, London, Ontario, Canada.
E-mail: ktiampo@uwo.ca.
[2] Department of Geological Sciences, California State University, Fullerton, CA, U.S.A.
[3] Department of Earth Sciences, University of California, Riverside, CA, U.S.A.
[4] Center for Computational Science and Engineering, University of California, Davis, CA, U.S.A.

Earthquake faults occur in topologically complex, multi-scale networks that are driven to failure by external forces arising from plate tectonic motions (BAKUN and LINDH, 1985; SCHOLZ, 1990; JONES, 1996). The primary problem that arises with this class of systems is that the true stress-strain dynamics are unobservable. Conversely, the space time patterns associated with the time, location, and magnitude of the earthquakes are easily observable. Recent retrospective studies provide evidence that there may be correlations associated with the future source region. Nonetheless it is difficult to apply these techniques if one does not know where the future source will be located (MOGI, 1969,1979; EVISON, 1977; HOUSE *et al*., 1981; KANAMORI, 1981; SCHWARTZ, 1984; WYSS and HABERMAN, 1988; SHAW *et al*., 1992; BUFE and VARNES, 1993; PRESS and ALLEN, 1995; DODGE *et al*., 1996; SALEUR *et al*., 1996; KATO *et al*., 1997; POLLITZ and SACKS, 1997; BOWMAN *et al*., 1998; BREHM and BRAILE, 1999; KAGAN and JACKSON, 2000).

The two methods presented in this paper are based on the identification of the characteristic patterns associated with the shifting of small earthquakes from one location to another prior to large earthquakes, as a result of changes in the underlying stress field (BOWMAN and KING, 2001; KING and BOWMAN, 2003). Theoretical analysis suggests the evolution of these seismicity structures is the surface expression of changes in underlying stress level (DIETERICH, 1994). Recent recognition 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; STEIN, 1999; DIETERICH *et al*., 2002; TODA *et al*., 2002, TIAMPO *et al*., 2006a) has been borne out by analysis of the data (RUNDLE *et al*., 2000; 2002a,b; 2004, 2006b). At the same time, there has been considerable success in linking the Coulomb stress changes and the evolution of the stress field to the temporal evolution of seismicity around the various fault structures (e.g., STEIN, 1999; BOWMAN and KING, 2001; SAMMIS *et al*., 2004). In this work we discuss results from a comparison of these two techniques for earthquake forecasting that rely on this relationship between the observable space-time patterns of earthquakes and the unobservable, underlying stress-strain dynamics.

The analysis of three separate regions identified with the PI index, using SAM, suggests that the two methods, while based on the same basic premise, can provide complementary information. The PI method can identify areas at risk for events of $M \geq 5$, but must rely on interpretation of the tectonic setting to estimate the potential hazard. While the SAM technique cannot forecast events of $5 \leq M \leq 6.5$, it can confirm that there is a substantial risk of a future large event, and it also can provide a mechanism that corresponds to the rupture dimension forecast by the modified PI index. The specific process involves additional analysis using SAM for areas of interest detected using the PI technique. A PI analysis identifies areas of systematic space-time variations in seismicity, linked to an increased probability of an event at a given location, during a specific time interval (TIAMPO *et al*., 2002a,b). The additional analysis uses the method of BOWMAN and KING (2001) to search for accelerating moment release for specific rupture scenarios associated with the PI anomalies. Finally, areas that the stress cycle evolution model

identifies as having increased seismic hazard are examined in more detail using the PI method, in order to provide insight into the nature of the anomaly.

2. Stress Accumulation Model

Recently two areas of study have been gaining increasing attention from the seismological community: Coulomb stress interactions between earthquakes, and accelerating seismic moment release before large earthquakes. In the first, it is observed that over periods of minutes to decades, future events generally occur in regions where the Coulomb stress has been enhanced by previous events (SMALLEY and TURCOTTE, 1985; BAKUN *et al.*, 1986; BAK *et al.*, 1987; RUNDLE, 1989; STEIN *et al.*, 1997; STEIN, 1999). The second area of study concerns observations of increased seismic activity in a wide region around an upcoming event epicenter. These distances typically extend to as much as five times the length of faulting associated with the future earthquake (BUFE and VARNES, 1993; DODGE *et al.*, 1996; BOWMAN *et al.*, 1998; BREHM and BRAILE, 1999). The dimensions of such a region can be directly related to the extent of increased Coulomb stress prior to the event, and observations of accelerating moment release in California become much clearer when the critical region is defined using Coulomb stress (BOWMAN and KING, 2001; MIGNAN *et al.*, 2006). This method produces a simple physical model that explains the observations in the context of the evolution of regional seismicity associated with large earthquakes (KING and BOWMAN, 2003).

Unlike conventional Coulomb stress techniques, which calculate stress changes, this stress cycle evolution method, or stress accumulation method (SAM), models the evolution of the stress field relative to the failure stress. Background seismicity is attributed to inhomogeneities in the stress field which are caused by a random field of fault patches, resulting in local regions above the failure stress. The inhomogeneous field is chosen such that when these patches fail, the resulting earthquake size distribution follows a Gutenberg-Richter law. Following a large event, the model produces both regions of increased seismicity (aftershocks), where the overall stress field has been elevated, and regions of reduced seismicity, where the stress field has been reduced (stress shadows). The high stress levels in the aftershock regions decrease following the main event. Combined with the stress shadow from the main event, the result is a broad seismically quiet region of lowered stress around the epicenter. Pre-event seismicity appears as the original stress shadows finally fill in over time as a result of tectonic loading. This increase in seismicity initially begins several fault lengths away from the main fault and moves inward and the seismic moment release in the region around the future epicenter increases as the event approaches. Synthetic catalogs generated by this model are virtually indistinguishable from real earthquake sequences in California and Washington (BOWMAN and KING, 2001; KING and BOWMAN, 2003).

While the stress cycle field associated with major earthquakes is simplified in this model, it includes the main elements to be expected for postseismic, coseismic, and preseismic stress fields over an apparent earthquake cycle. A pictorial representation of

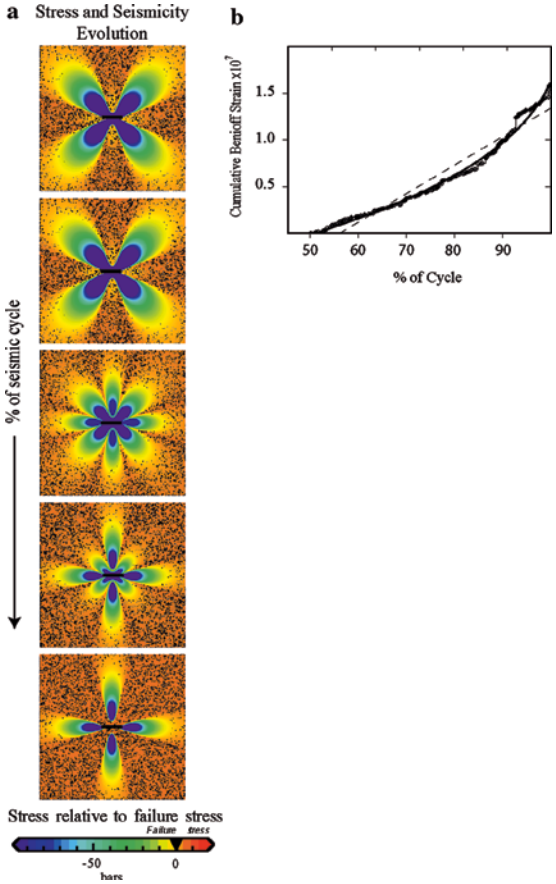


Figure 1

Stress and seismicity through the earthquake cycle for the stress accumulation model. a) Stress changes for a finite length fault 80 km long and 15 km wide. The maximum stress level is the failure stress (orange), all other stress levels are lower (the color bar indicates values). The stress field is generated by adding the stress change due to loading of the main fault to a random inhomogeneous regional stress field. Black areas indicate where earthquakes have occurred due to the stress rising above a failure threshold. The removal of the stress shadow from the previous event as a result of loading on the main fault leads to accelerating seismicity in the diagonal lobes (see KING and BOWMAN 2003) for a detailed description). b) Accelerating seismicity before a M = 6.5 earthquake from the model of KING and BOWMAN (2003). The result is typical of the synthetic seismicity catalogs generated by the model. The solid line indicates the fit of the synthetic data to a power-law time-to-failure equation, while the dashed line is a linear fit to the data. The time scale is given as a percentage of the loading cycle.

this model is shown in Figure 1. This model shows that seismic activity is greatest during the aftershock sequence and before the next event, and minimum in the intervening period. The variation in activity is predominantly due to changes in the y-intercept, or

a-value, of the Gutenberg-Richter relation with associated fluctuations in the maximum magnitude of events. While changes in the slope of the Gutenberg-Richter relation, commonly called the b-value, can occur as well, they are a result of spatiotemporal changes in the a-value relative to a constant number of large events, at least over the observation time (TURCOTTE, 1997; RICHTER, 1958).

BOWMAN *et al*. (1998) showed that a simple search algorithm can be used to define circular regions of accelerating moment release before large earthquakes. They noted that if the region being investigated is too large, then the accelerating moment release is masked by unassociated random background seismicity, while the selection of a region that is too small excludes events that are important in defining the acceleration. This region scales with the seismic moment of the upcoming large event (BOWMAN *et al*., 1998; JAUME and SYKES, 1999). The physical basis for these observations can be understood in terms of the model presented above.

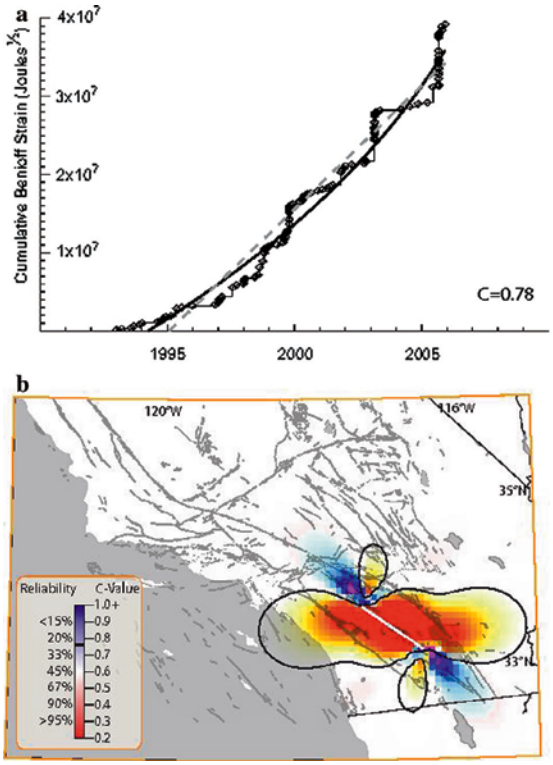


Figure 2
AMR calculated for a large, magnitude 7.5 event on the southern San Jacinto fault, as shown in b). a) shows the best fit to the cumulative Benioff strain release for this scenario.

The cumulative Benioff strain within a series of typical rupture mechanisms such as that shown in Figure 2b for the southern San Jacinto fault, centered on the epicenter of the scenario earthquake, is fit to both a power-law time-to-failure equation (Bufe and Varnes, 1993; Bowman *et al*., 1998) and a straight line. The ratio of the residuals to these fits (C = power-law residuals/linear residuals) is calculated for each radius, and is called the C value (Figure 2a). The greater the curvature of the accelerating seismic moment release, the smaller the C value and the greater the likelihood of an event. A series of these rupture scenarios of varying size is plotted for different time periods, and the minimum of the resulting plot is the region with the greatest acceleration.

In Figure 2 we have applied the region optimization algorithm of Bowman *et al*. (1998) to the historic catalogs in southern California. The data used are from the ANSS catalog, for events greater than or equal to magnitude 4.5, from 1932 to 2004. The catalog is not declustered, because it is the spatial and temporal clustering that identifies the region of interest. The scenario fault in Figure 2b has source parameters equivalent to a magnitude 7.5 earthquake along the southern San Andreas fault. The best C value for this event is 0.78. Recent work by Mignan *et al*. (2006) suggests that only C values less than 0.6 are considered reliable. This is shown in the scale on Figure 2b, where a C value of 0.78 is denoted by a reliability value of approximately 25%. This suggests that the event is unlikely to occur in the near future although, given that the physics of the SAM methodology does not lend itself to the forecasting of events of magnitude less than M = 6.5 (King and Bowman, 2003).

3. The Pattern Informatics Index

The PI index is a methodology for quantifying the spatiotemporal seismicity rate changes in historic seismicity (Tiampo *et al*., 2002b; Rundle *et al*., 2002) The quantitative basis for describing earthquake patterns here lies in phase dynamics analysis, where the temporal evolution of a dynamical quantity is described by changes in the amplitude and phase angle of a state vector, or phase function (Holmes *et al*., 1996; Mori and Kuramoto, 1998; Rundle *et al*., 2000b). Simulations strongly suggest that earthquake seismicity can be described by phase dynamics, in which the important changes in seismic activity are associated only with changes in the phase angle of this vector function (Rundle *et al*., 1999). By integrating these rotations over time, and differencing them at different observation times, space-time patterns of anomalous activity or quiescence emerge that seem to precede large events. The spatial locations of these coherent seismicity structures can be related to the locations of large earthquakes that occur a few years after formation of the structures (Tiampo *et al*., 2002b, c). Recently, this analysis has been extended to identify the rupture dimension of the upcoming events directly from the observed seismicity data prior to that event (Tiampo *et al*., 2006b). Here the rupture dimension is defined as the area of the main shock and subsequent aftershocks on the contiguous fault plane.

To calculate the PI index, seismicity data were taken from existing observations in southern California between 1932 and the present. This data is a combination of the SCEC catalog for historic seismicity in southern California and the NCSN catalog for seismicity in northern California. In order to ensure that all the data are consistent with the completeness of the catalog, only events of magnitude larger than $M \geq 3$ were included in this analysis. The choice in catalogs is related to the fact that the PI analysis uses numbers of events, and therefore relies on the relatively large numbers of smaller earthquakes as a proxy for stress, while the SAM analysis uses cumulative Benioff strain, and the smaller events contribute relatively little to this total. The original PI calculation used the two catalogs for all of California because they were more complete for smaller events in the early years of this research, but they are virtually identical for earthquakes of $M \geq 4.5$.

The data are decimated by location into boxes of 0.1 degrees, latitude and longitude, approximately 11 km to a side. Time series are created for each of these locations containing the total number of events at each location that occurred in each year. Each location is denoted x_i, where i ranges from 1 to N total locations. Note that, as in the case of the SAM analysis, the catalog was not declustered, because the space-time clustering carries the information about the important space-time correlations. Also, while this method deviates from that of the SAM analysis described above, in that it uses smaller events, it also attempts to objectively quantify changes in the a-value of the Gutenberg-Richter relation over time.

Using this technique, we construct the phase function $\mathbf{S}'(\mathbf{x}_i,t_0,t)$. Physically, the important changes in seismicity will be given by $\Delta\mathbf{S}'(\mathbf{x}_i,t_1,t_2) = \mathbf{S}'(\mathbf{x}_i,t_0,t_2) - \mathbf{S}'(\mathbf{x}_i,t_0,t_1)$, which is a rotation of an N-dimensional unit vector $\mathbf{S}'(\mathbf{x}_i,t_0,t)$ in time. Dividing by the constant variance normalizes the seismicity to its background, revealing only the small, local changes in seismic activity. The physical picture is that $\Delta\mathbf{S}'(\mathbf{x}_i,t_1,t_2)$ is proportional to a mean drift angle (or vector difference) over the time period t_1 to t_2, where changes in spatial correlations measured as temporal changes in seismicity rate influence the direction and amount of that change. The direction in which $\Delta\mathbf{S}'(\mathbf{x}_i,t_1,t_2)$ points has physical meaning because the changes in spatial and temporal activity are related to the underlying stress correlations in the system. Finally, we average $\Delta\mathbf{S}'(\mathbf{x}_i,t_1,t_2)$ over all possible base years, t_0, in order to reduce the effects of noise in the seismic catalog.

In phase dynamical systems, probabilities are related to the square of the associated vector phase function (Holmes *et al*., 1996; Mori and Kuramoto, 1998; Rundle *et al*., 2000b). Thus the PI index, $\Delta\mathbf{P}$, is $\Delta\mathbf{P}(\mathbf{x}_i,t_1,t_2) = \{\Delta\mathbf{S}'(\mathbf{x}_i,t_1,t_2)\}^2 - \mu_p$, where μ_P is the spatial mean of $\{\Delta\mathbf{S}'(\mathbf{x}_i,t_1,t_2)\}^2$, and represents physically the time-dependent background. For any given catalog up to the present, it is possible to compute the relative change in probability of an event over any given time period.

The PI technique is a means of processing seismicity data to reveal the underlying space-time structure, much as one uses a Fourier transform to analyze time series. In recent years it has been applied to California seismicity for the purpose of identifying systematic space-time variations that are an effective measure of current and future seismic moment

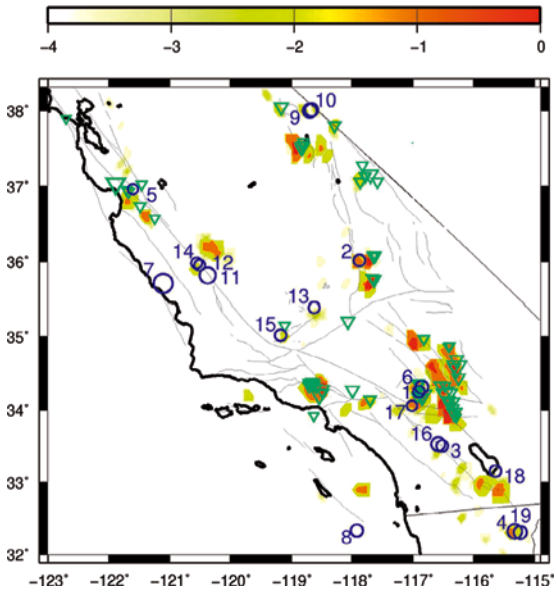


Figure 3
Color contour plot of PI index, $\Delta P(xi, t1, t2)$, for locations at which $\Delta P(xi, t1, t2) > 0$, 1989–1999. Color scale is logarithmic, where the number represents the exponent. Green triangles represent those events that occurred during the period's prior 10 years, 1989–1999. Circles show those events, $M \geq 5$, that occurred between 2000 and 2007.

release (TIAMPO *et al.*, 2002b,c). These variations, or anomalies, identify the areas of increased probability, $\Delta\mathbf{P}$ $(\mathbf{x}_i,t_1,t_2)$, of an event at a location x, during a time interval, t_2-t_1. The forecast skill of the technique is supported by thousands of statistical likelihood ratio tests (TIAMPO *et al.*, 2002b; HOLLIDAY *et al.*, 2006).

Figure 3 shows the PI index, $\Delta\mathbf{P} = \Delta\mathbf{P}(\mathbf{x_i},1989,1999)$, for detecting anomalous spatial correlations at sites in southern California *prior* to January 1, 2000. It is a plot of all $\Delta\mathbf{P} > 0$, scaled to the largest value of $\Delta\mathbf{P}$ in the region. This shows the increase in $\Delta\mathbf{P}$ above the background level and indicates spatially coherent regions of seismic activity, which in turn are indicators of an increased chance of a major earthquake happening near those locations. Inverted green triangles represent events of magnitude $M \geq 5.0$ that occurred during the time period covered by the plot, indicating locations associated with large events that are present in the data used to construct $\Delta\mathbf{P}$. Blue circles represent those earthquakes that have occurred since 2000, inclusive. Of interest are regions of increased $\Delta\mathbf{P}$ with no triangles near them, since these locations represent a region of higher probability of an event. Seventeen out of the nineteen earthquakes that have occurred between the origin date of this forecast (January, 2000) and September, 2006 were coincident with the forecast anomalies within the $\pm$ 11 km margin of error that is the

coarse graining box size (RUNDLE *et al*., 2002; TIAMPO *et al*., 2002b). However, it should be noted that no theoretical or practical link has been made between the intensity of these anomalies and the magnitude or timing of the upcoming event.

In recent work, HOLLIDAY *et al*. (2005) demonstrated that the results of this forecast through the first five years of testing are significant. This particular verification methodology is based upon the hypothesis that the PI forecasts are binary forecasts, i.e. there will be an earthquake for any PI value above some given threshold. These recent developments suggest that the PI index represents a threshold probability, in which a value may be chosen above which there is 100% expectation of a large event, providing the basis for modifications to the PI calculation that identifies those smaller regions where the PI value reaches extremely high local values, and thus can be used to delineate the fault rupture zone.

We begin by taking small subsets of the California fault zone, each centered on a particular, known epicenter. These subregions begin at a small but reasonable size, on the order of 0.1 to 0.4 degrees, and they increase systematically, again in steps of 0.05 to 0.2. The PI calculation is then performed for each of these subregions. Again, we do not take any seismicity that occurs after the actual event. The box size for the gridding of the seismicity and the calculation of the PI index remains separate, as detailed above, and can be different from the subregion size or its increasing step size. In this particular variation of the PI index, denoted the roughness index, we do not normalize the PI values to the maximum. Based on the discussion above, that the PI index is a binary forecast, quantifying, based on a threshold level, whether an event will or will not occur in the future, we wish to evaluate these smaller subregions based upon whether they exceed a particular threshold. That threshold was set at a value of 1.0, i.e. rupture is a certainty if the value in the box is above 1.0 , and all values of the PI index in that region that are greater than 1.0 in the original calculation then are set equal to 1.0. This particular value was chosen based upon those early studies, when we examined the statistical distribution of the PI values over the entire region, and determined that low-level fluctuations existed below a PI value of 1.0. Subsequently, the mean and variance of the PI values in each gridded box location is calculated for these subregions, over all potential subregions, and the region for which those values are a constant of 1.0 and 0.0, respectively, outline the upcoming fault rupture dimension. Retrospective testing of this method is outlined in TIAMPO *et al*., (2006b).

In the following section we examine in conjunction with SAM, as described above, a number of the smaller regions in Figure 3 that have yet to experience large events, in an attempt to learn more about the underlying stress levels and associated activity.

4. Case Studies

While the PI index can identify locations of events greater the $M \sim 5$, the SAM technique can be employed to quantify the mechanism and potential magnitude. Here we investigate that possibility using specific anomalies identified by the PI index.

In Figure 4 are shown three smaller regions extracted from Figure 3 above, for the northern, central, and southern San Andreas region. These three regions are not only at significant distances from each other, but are examples of the widely varying dynamics that exist in the San Andreas fault network — the northern San Andreas with its lower historical background seismicity rate, the Big Bend region of the San Andreas, and the very southern tip of the California fault system.

Note that in Figure 4a there are two anomalies located on the Calaveras fault east and north of San Francisco. One of these is located at approximately 38.75° latitude, −122.75° longitude, while the second occurs at the juncture of the Calaveras, Hayward, and northern San Andreas faults. Also, in Figure 4b, there is another anomaly in the Carrizo plain, near the tip of the White Wolf fault, while in Figure 4c there are a number of anomalies located at the southern end of the San Andreas, San Jacinto and Elsinore faults. Each of these locations is examined using SAM.

A stress accumulation scenario was devised for each of the above locations. In the first scenario, along the Calaveras fault (Fig. 4a), both anomalies were investigated. The northern anomaly is located on the Rodgers Creek Fault, near the Geyser's geothermal area, where substantially persistent seismicity occurs that could be affecting the results from both methods. Because the original formulation for the PI index did not allow for discrimination between the different anomalies based upon intensity, a modification of the PI index that allows for the identification of a rupture zone was applied to both anomalies. This method investigates smaller subregions of the fault system, and is detailed in Tiampo *et al*. (2006b). This analysis identifies an area centered directly on the Geysers, and led us to discount the likelihood of a large event centered on the Rodger's Creek fault. As a result, a SAM analysis was performed along the Calaveras, and a right-lateral strike slip event of approximately 60 km in length was modeled as shown in Figure 5b. The resulting best-fit residual, C value equal 0.54, corresponds to the cumulative Benioff strain shown in Figure 5a.

The rupture dimension calculation, again using a modified version of the PI index, is shown in Figure 6. This particular method identifies the nucleation zone of the upcoming event using changes in the rate of the smaller seismicity. Note the good correspondence with the rupture scenario from the SAM analysis shown in Figure 5. The rupture dimension is slightly shorter in the PI analysis, but both delineate a strike-slip event corresponding to an event of M $\sim$ 6–6.5 or greater (Wells and Coppersmith, 1994).

In the second scenario, a right-lateral strike-slip earthquake was simulated along the San Andreas fault through the Carrizo plain, corresponding to part of the segment that ruptured during the 1857 Fort Tejon earthquake. The stress accumulation results are shown in Figure 7. Here the optimal residual value is approximately 0.27, for a reliability of more than 95%. The associated AMR curve is shown in Figure 7a. Figure 7b shows the optimal rupture scenario, an event of M $\geq$ 7.2.

Figure 4c details a number of anomalies in the southern region of the San Andreas fault network, in addition to that already shown in Figure 2, for the Elsinore and San Jacinto faults. However, the SAM analyses for these regions do not produce a rupture

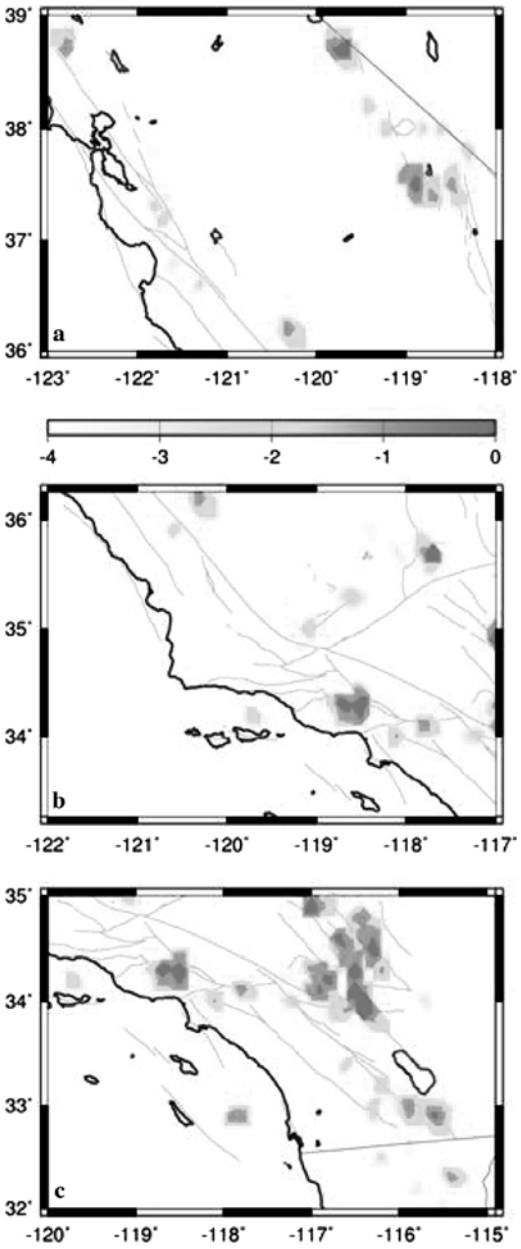


Figure 4
PI index for smaller regions from Figure 3, for the time period 1989–1999.

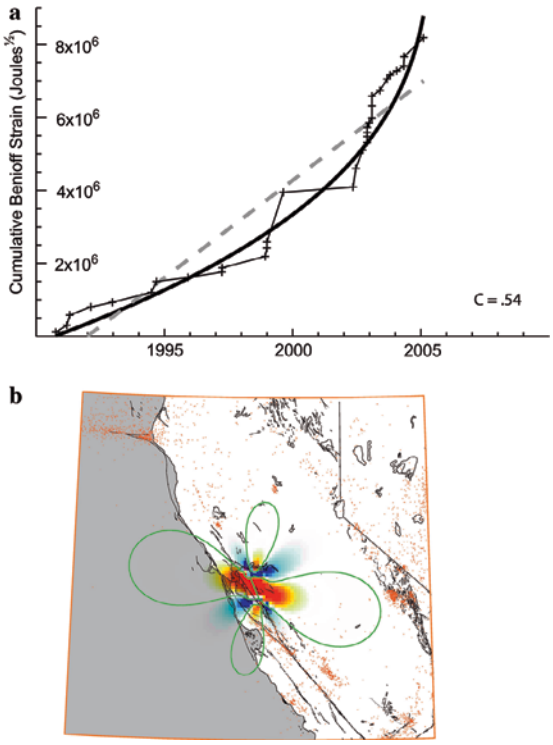


Figure 5
AMR calculation for the Calaveras in northern California, as taken from Figure 4a. (a) shows the cumulative Benioff strain calculation, while (b) shows the associated fault mechanism.

scenario with a good level of reliability. The best is that shown in Figure 2, for an event of M ~ 7.5 on the San Jacinto fault, with a C value of 0.78 and a corresponding reliability of less than 25%. However, the physics of the SAM methodology does not lend itself to the forecasting of events of magnitude less than M = 6.5, thus it was hypothesized that the PI anomalies delineated areas of increased probability for smaller magnitude earthquakes (KING and BOWMAN, 2003).

To test this hypothesis, the modified PI index for identifying rupture dimension was performed on the anomalies shown in Figure 4c (TIAMPO *et al.*, 2006b). The stable solution for an event located near the Anza region is shown in Figure 8. The magnitude of an event corresponding to a rupture dimension of approximately 20 km, as calculated, is only 6.25 (WELLS and COPPERSMITH, 1994), suggesting that the lack of reliability resulting from the SAM analysis is due to the low magnitude of the anticipated events.

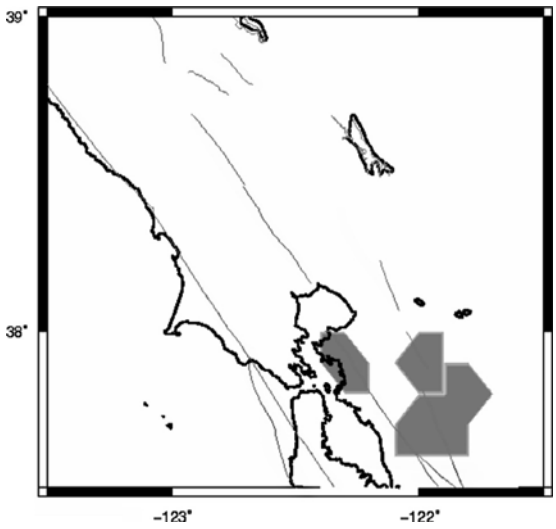


Figure 6
Rupture dimension for the Calaveras fault (TIAMPO *et al*., 2006b), corresponding to an event of M ~ 6.

5. Discussion and Conclusions

The analysis of three separate regions identified with the PI index, using SAM, suggests that the two methods, while based on the same basic premise, can provide complementary information. For example, in the first case shown in Figure 4a, the PI method can identify areas for study not previously examined, while SAM can not only confirm that there is a substantial risk of a future event, but also can provide a mechanism that corresponds to the rupture dimension forecast by the modified PI index.

The biggest difference between the analysis techniques is in the size of the events used to perform the analysis – SAM relies on events of magnitude greater than 4.5, while the PI index uses all events greater than magnitude 3.0. Figures 2, 4c, and 8 illustrate that, as a result, SAM does not forecast with any reliability events along the southern San Jacinto fault, due to the minimum limit of M ~ 6.5 for a SAM forecast, while the PI index can identify an event of M ~ 6.25.

The successful integration of these two analyses, the PI index and SAM, provides support for the hypothesis that change in the relative number of seismic events, or the Gutenberg-Richter a-value, is an important measure of change is the underlying stress field. This can result in an apparent change in the b-value, depending on the temporal and spatial region of study. Each method, however, identifies a region with a different spatial scale, resulting in differing information about the magnitude of the upcoming events. Finally, we have demonstrated that a comparison of their results in the same regions can provide important complementary information for the forecasting of future large earthquakes.

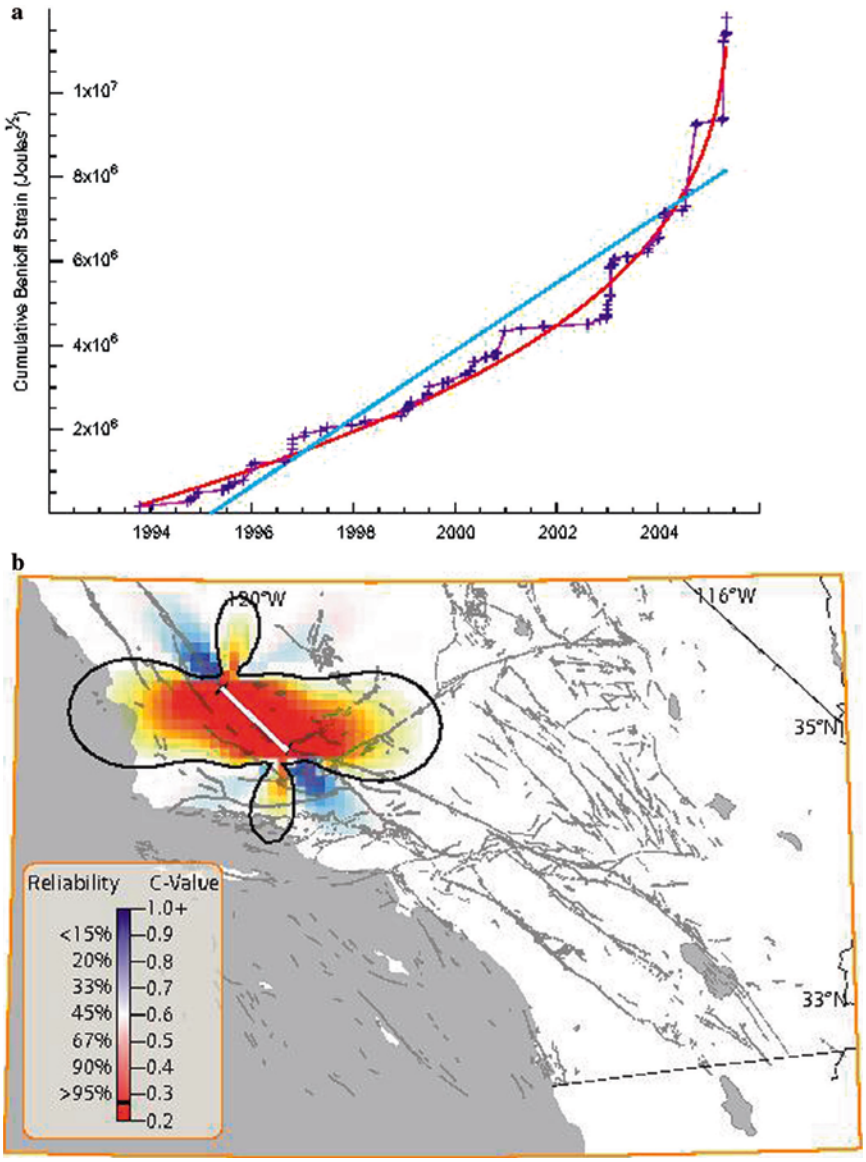


Figure 7
AMR calculation for the Carrizo plain, as taken from Figure 4b. The fault mechanism is shown in (b), while the cumulative Benioff strain calculation is shown in (a).

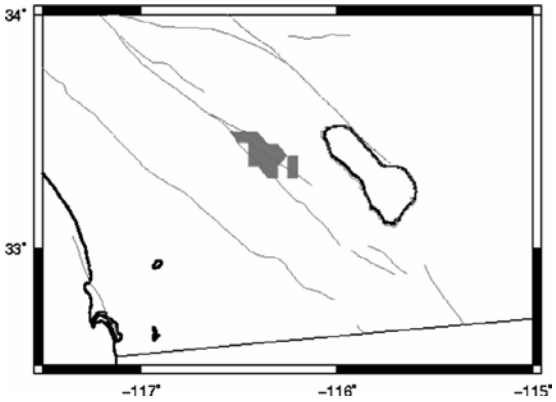


Figure 8
The rupture dimension calculated for an event along the southern San Jacinto fault using the modified PI index (TIAMPO *et al*., 2006b).

Acknowledgements

This research was supported by the Southern California Earthquake Center. SCEC is funded by NSF Cooperative Agreement EAR-0106924 and USGS Cooperative Agreement 02HQAG0008. The SCEC contribution number for this paper is 814. Research by KFT is also supported by an NSERC Discovery grant and the NSERC and Benfield/ICLR Industrial Research Chair in Earthquake Hazard Assessment. 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.

REFERENCES

BAK, P., TANG, C., and WEISENFIELD, K. (1987), *Self-organized criticality: An explanation of the 1/f noise*, Phys. Rev. Lett. *59*, 381–384.

BAKUN, W.H., KING, G.C.P., and COCKERHAM, R.S. (1986), *Seismic slip, aseismic slip, and the mechanics of repeating earthquakes on the Calaveras fault, California*, Earthquake Source Mechanics, AGU Monograph, 195–208.

BAKUN, W. and LINDH, A. (1985), *The Parkfield, California earthquake prediction experiment*, Science *89*, 3051–3058.

BOWMAN, D. D. and KING, G.C.P. (2001), *Accelerating seismicity and stress accumulation before large earthquakes*, Geophys. Res. Lett. *28*, 4039–4042.

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.

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.

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.

DODGE, D.A., BEROZA, G.C., and ELLSWORTH, W.L. (1996), *Detailed observations of California foreshock sequences: implications for the earthquake initiation process*, J. Geophys. Res. 101, 22371–22392.

EVISON, F.F. (1977), *The precursory earthquake swarm*, Phys. Earth Planet. Int., *15*, 19–23.

GELLER, R.J., JACKSON, D.D., KAGAN, Y.Y., and MULARGIA, F. (1997), *Enhanced: earthquakes cannot be predicted*, Science *275*, 1616–1620.

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.

HOLLIDAY, J.R., NANJO, K.Z., TIAMPO, K.F., RUNDLE, J.B., and TURCOTTE, D.L. (2005), *Earthquake forecasting and its verification*, Nonlinear Processes in Geophysics *12*, 1607-7946/npg/2005-12-965.

HOLLIDAY, J.R., RUNDLE, J.B., TIAMPO, K.F., KLEIN, W., and DONNELLAN, A. (2006), *Systematic procedural and sensitivity analysis of the Pattern Informatics method for forecasting large (M>5) earthquake events in southern California*, Pure App. Geophys. doi:10.1007/s00024-006-0131-1.

HOLMES, P., LUMLEY, J.L., and BERKOOZ, G., *Turbulence, Coherent Structures, Dynamical Systems and Symmetry* (Cambridge, University Press, U.K. 1996).

HOUSE, L.S., SYKES, L.R., DAVIES, J.N., and JACOB, K.H., *Identification of a possible seismic gap near Unalaska Island, eastern Aleutians.* In Earthquake Prediction: an International Review (eds D. W. Simpson, II, and P. G. Richards), 81–92 (AGU, Washington, D.C. 1981).

HUBERT-FERRARI, A., BARKA, A., JACQUES, E., NALBANT, S.S. MEYER, B., ARMIJO, R., TAPPONNIER, P., and KING, G.C.P. (2000), *Seismic hazard in the Marmara Sea following the 17 August 1999 Izmit earthquake*, Nature *404*, 269– 273.

JAUME, 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. Geophs. *155*, 279–306.

JONES, L.M. (1996), *Earthquake prediction: The interaction of public policy and science*, Proc. Nat. Acad. USA *93*, 3721–3725.

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 (eds D.W. Simpson, II, and P. G. Richards), 1–19 (AGU, Washington, D.C. 1981).

KATO, N., OHTAKE, M., and HIRASAWA, T. (1997), *Possible mechanism of precursory seismic quiescence: Regional stress relaxation due to preseismic sliding*, Pure Appl. Geophys. *150*, 249–267.

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.

KING, G.C.P., STEIN, R.S., and LIN, J. (1994), *Static stress changes and the triggering of earthquakes*, BSSA *84*, 935–953 (1994).

MIGNAN, A., BOWMAN, D.D., and KING, G.C.P. (2006*), An observational test of the origin of accelerating moment release before large earthquakes***,** J. Geophys. Res. *111*, B11304, doi:10.1029/2006JB004374.

MOGI, K. (1969), *Some features of recent seismic activity in and near Japan, 2. Activity before and after great earthquakes*, Bull. Earth. Res. Inst. *47*, 395–417.

MOGI, K. (1979), *Two kind of seismic gaps*, Pure Appl.Geophys. *117*, 1172–1186.

MORI H. and KURAMOTO, Y., *Dissipative Structures and Chaos* (Springer-Verlag, Berlin 1998).

PACHECO, J.F., SCHOLZ, C.H., and SYKES, L.R. (1992), *Changes in frequency-size relationship from small to large earthquakes*, Nature *355*, 71–73.

POLLITZ, F.F., and SACKS, I.S. (1997), *The 1995 Kobe, Japan, earthquake: A long-delayed aftershock of the offshore 1944 Tonankai and 1946 Nankaido earthquakes*, BSSA *87*, 1–10.

PRESS F. and ALLEN, C.R. (1995), *Patterns of seismic release in the southern California region*, J. Geophys. Res. *100*, 6421–6430.

RICHTER, C.F., *Elementary Seismology* (Freeman, San Francisco 1958).

RUNDLE, J.B. (1989), *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., KLEIN, W., and GROSS, S. (1999), *Physical basis for statistical patterns in complex earthquake populations: Models, predictions and tests*, Pure. App. Geophys. *155*, 575–607.

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

SALEUR, H., SAMMIS, C.G. and SORNETTE, D. (1996) *Discrete scale invariance, complex fractal dimensions, and Log-periodic fluctuations in seismicity*, J. Gephys. Res. *101*, 17661–17677.

SAMMIS, C. G., BOWMAN, D. D., and KING, G. (2004), *Anomalous seismicity and accelerating moment release preceding the 2001 and 2002 earthquakes in northern Baja California, Mexico*, Pure. Appl. Geophys. *161*, 2369–2378.

SCHOLZ, C.H., *The Mechanics of Earthquakes and Faulting* (Cambridge University Press, Cambridge, UK 1990)

SCHWARTZ, D.P. (1984), *Fault behavior and characteristic earthquakes – Examples from the Wasatch and San Andreas fault zones*, J. Geophys. Res. *89*, 5681–5698.

SHAW, B. E., CARLSON, J.M., and LANGER, J.S. (1992), *Patterns of seismic activity preceding large earthquakes*, J. Geophys. Res. *97*, 479–488.

SMALLEY, R.F., and TURCOTTE, D.L. (1985), *A renormalization group approach to the stick-slip behavior of faults*, J. Geophys. Res. *90*, 1894–1900.

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.

SYKES, L.R., SHAW, B.E., and SCHOLZ, C.H. (1999), *Rethinking earthquake prediction*, Pure. Appl. Geophys. *155*, 207–232.

TIAMPO, K.F., RUNDLE, J.B., GROSS, S.J., MCGINNIS, S. and KLEIN, W. (2002a), *Eigenpatterns in southern California seismicity*, J. Geophys. Res. *107*, 2354, doi:10.1029/2001JB000562.

TIAMPO, K.F., RUNDLE, J.B., MCGINNIS, S., GROSS, S., and KLEIN, W. (2002b), *Mean-field threshold systems and phase dynamics: An application to earthquake fault systems*, Eur. Phys. 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.

TIAMPO, K.F., RUNDLE, J.B., SÁ MARTINS, J., KLEIN, W., and MCGINNIS, S. (2004), *Methods for evaluation of geodetic data and seismicity developed with numerical simulations: Review and applications*, Pure. Appl. Geophys. *161*, 1489–1507.

TIAMPO, K.F., RUNDLE, J.B., and KLEIN, W. (2006a), *Premonitory seismicity changes prior to the Parkfield and Coalinga earthquakes in southern California*, Tectonophysics *413*, 77–86.

TIAMPO, K.F., RUNDLE, J.B., KLEIN, W., and HOLLIDAY, J. (2006b), *Forecasting rupture dimension using the Pattern Informatics technique*, Tectonophysics *424*, 367–376.

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.

TURCOTTE, D.L., *Fractals and Chaos in Geology and Geophysics*, 2nd ed. (Cambridge University Press, 1997).

VERE-JONES, D, ROBINSON, R., and YANG, W. (2001), *Remarks on the accelerated moment release model; Problems of model formulation, simulation, and estimation*, Geophys. J. Int. *144*, 517–531.

WELLS, D.L. and COPPERSMITH, K.J. (1994), *Empirical relationships among the magnitude, rupture length, rupture width, rupture area, and surface displacement*, BSSA *84*, 974–1002.

WYSS, M. and HABERMAN, R.E. (1988), *Precursory quiescence before the August 1982 Stone Canyon, San Andreas fault, earthquakes*, Pure. Appl.Geophys. *126*, 333–356.

(Received October 1, 2006, revised May 5, 2007, accepted June 7, 2007)

Pure appl. geophys. 165 (2008) 711–722
0033–4553/08/030711–12
DOI 10.1007/s00024-008-0314-z

Pure and Applied Geophysics

The Newest Developments of Load-Unload Response Ratio (LURR)

XIANG-CHU YIN,[1,2] LANG-PING ZHANG,[2] YONGXIAN ZHANG,[2,3] KEYIN PENG,[2,3] HAITAO WANG,[2,5] ZHIPING SONG,[2,6] HUAIZHONG YU,[2] HUI-HUI ZHANG,[4] CAN YIN,[7] and YUCANG WANG[2,7]

Abstract—The Load Unload Response Ratio (LURR) method is an intermediate-term earthquake prediction approach that has shown considerable promise. It is inspiring that its predictions using LURR have been improving. Since 2004 we have made a major breakthrough in intermediate-term earthquake forecasting of the strong earthquakes on the Chinese mainland using LURR and successfully predicted the Pakistan earthquake with magnitude M 7.6 on October 8, 2005. The causes for improving the prediction in terms of LURR have been discussed in the present paper.

Key words: LURR (Load Unload Response Ratio), earthquake prediction, Chinese mainland, M 7.6 Pakistan earthquake.

1. Introduction

An earthquake is a very complicated phenomenon. 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 of energy. Therefore the preparation process of an earthquake is exactly the damage process of the focal media.

From the microscopic viewpoint, the damage process of geo-material has incredible richness in complexity. In any rock block there must be a large number of disordered defects (cracks, fissures, joints, faults and caves) with different sizes, shapes and orientations. The damage process of the focal media includes initiation, growth,

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

[2] LNM (State Key Laboratory of Nonlinear Mechanics), Institute of Mechanics, Chinese Academy of Sciences (CAS), Beijing 100190, China.

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

[4] School of Civil Engineering and Mechanics, Yanshan University, Qinhuangdao 066004, China.

[5] Xinjiang Earthquake Bureau, CEA, Urumchi 830011, China. Emails: whtxj@xj.cninfo.net

[6] Shanghai Earthquake Bureau, CEA, Shanghai 200062, China. Emails: zpsong@sohu.com

[7] ESSCC, University of Queensland, Australia, the University of Queensland, St. Lucia, Brisbane QLD 4072, Australia. Email: wangyc@esscc.earth.uq.edu.au

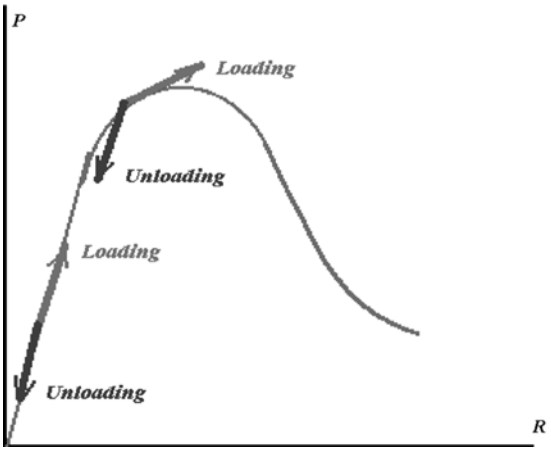


Figure 1
The constitutive curve of a focal zone.

interaction, coalition and cascade of cracks. It is an irreversible, far-from-equilibrium and nonlinear process which has been intensively studied for nearly half century; however numerous fundamental questions still remain unsolved. The inherent difficulty of earthquake prediction stems mainly from this complexity.

From the macroscopic viewpoint the constitutive relation is a comprehensive description of the mechanical behaviors of the materials. A typical constitutive relation for focal media (rock) is shown in Figure 1. We define a more general coordinate system where the general load P is used instead of stress σ and the general response R to load is used instead of strain ε in Figure 1. If the load acting on a specimen increases monotonously, the specimen 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 loading modulus and the unloading one are equal to each other. Contrary to the elastic regime, the damage one is irreversible and the unloading response is different from the loading one, or the loading modulus is different from the unloading one. This difference indicates the deterioration of material due to damage. In terms of seismology, 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. It involves calculating the ratio of a specified energy release measure during loading and unloading where loading and unloading periods are determined from the earth tide induced perturbations in the CFS (Coulomb Failure Stress) on optimally oriented faults. In the lead-up to large earthquakes, high LURR values are frequently observed a few months or years prior to a strong earthquake. To date the

LURR method is mainly applied for intermediate-term earthquake prediction. We have predicted some strong events successfully using the LURR approach, with few false predictions, especially in the early stage.

2. The Newest Developments of LURR

It is inspiring that the prediction results using LURR have been continuously improving. Since 2004, we have made a major breakthrough in intermediate-term earthquake forecasting in terms of LURR. Figure 2 shows the anomalous LURR regions in the Chinese mainland, calculated at the end of 2003, namely the regions where the LURR value is equal to or larger than 1 and marked with color.

In fact, the real area of the *seismogenic region* should be larger than the anomalous LURR regions (color regions). If P is a point at the boundary of the anomalous LURR regions. This means that the LURR value Y in the special window (a circle region centered at P with radius R) is equal to or larger than 1 and is assigned the Y value at point P (Fig. 4). In other words, the whole circle region centered at P with radius R is in a seismogenic state. At every point of the boundary of LURR anomalous regions we can draw a circle with radius R, and then an envelope of all the circles can be drawn. The areas in the envelope should be the seismogenic regions. We predicted that there would be a high probability of occurrence for $M_s \geq 5$ earthquakes in the next year (2004) in

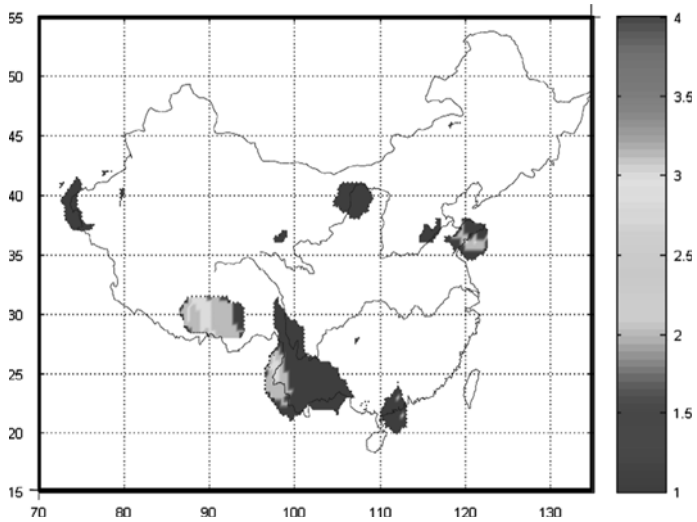


Figure 2
The anomalous LURR regions in the Chinese mainland calculated at the end of 2003 will have a high probability of occurrence for earthquakes with $M \geq 5$ in the next year (2004).

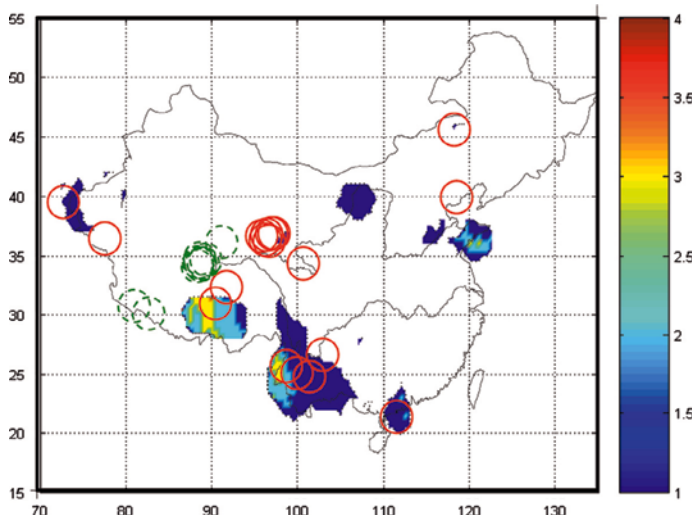


Figure 3
The map of epicenters of earthquakes with magnitude $M_s \geq 5$ in 2004 in the Chinese mainland. The red solid circles denote the events which occurred in regions in which the observed data are good enough to calculate LURR, and the green dashed circles denote the events which occurred in regions where the scarcity of observed data make the LURR calculation impossible.

these seismogenic 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 officially published in the end of 2003 by

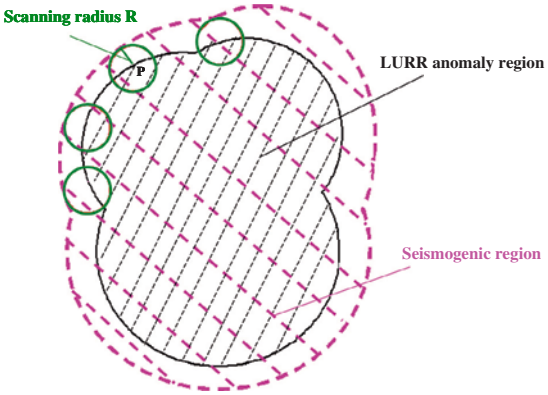


Figure 4
The sketch of an LURR anomaly region and a seismogenic region.

Seismological Press, Beijing. Figure 3 shows epicenters of earthquakes with magnitude $M_s \geq 5$ which occurred in the Chinese mainland in 2004. The red solid circles denote the events of $M_s \geq 5$ (total number 17) which occurred in the regions in which the observed data are available to calculate LURR, and the green dashed circles denote the events of $M_s \geq 5$ (total number 16) which occurred in regions such as Tibet, where the observed data are not available to calculate LURR. According to the standard of the Chinese seismological circle, the earthquakes which occurred in the data scarcity regions should be ignored statistically. If an earthquake occurred in the seismogenic regions, namely if an earthquake occurred in the anomalous LURR regions or its epicenter distance from the anomalous LURR regions is less than the radius of the spatial window, this event is considered a successfully predicted one. Figure 3 shows us that there are 15 events among 17 (*a percentage of 88%*) which occurred in the predicted LURR anomalous regions.

In 2005 the predicted result in terms of LURR was even better than 2004 (YIN *et al*., 2004). There were twelve strong earthquakes with $M_s \geq 5$ among total 13 earthquakes (*a percentage of 92%*) which occurred in the seismogenic regions. The only missing forecasted one-Jiujiang earthquake (Jiangxi Province in eastern China) with magnitude M 5.7 (November 26, 2005; 29.7°N, 115.7°E) is not in the LURR anomalous regions, but is located in the data scarcity regions (Fig. 5). Noting that the minimum value of the color bar (color key) in Figure 5 is 0 instead of 1.0 in Figures 2 and 3. Therefore all the blank areas (uncolored) in Figure 5 are the areas in which the observed data are not available to calculate LURR.

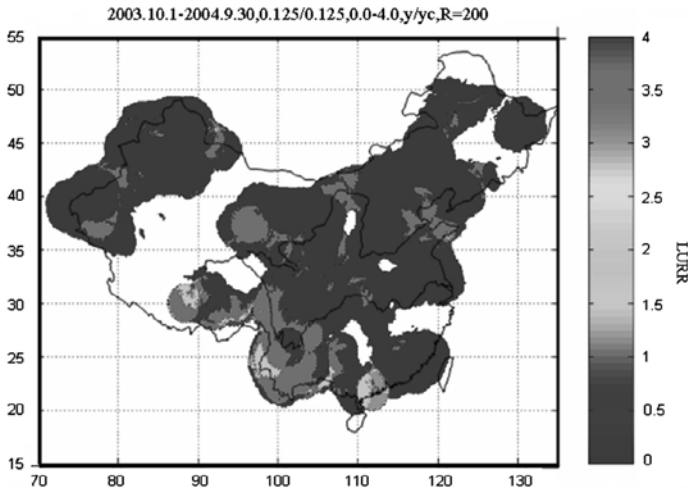


Figure 5
The definition of a data scarcity region. All the blank areas are data scarcity regions.

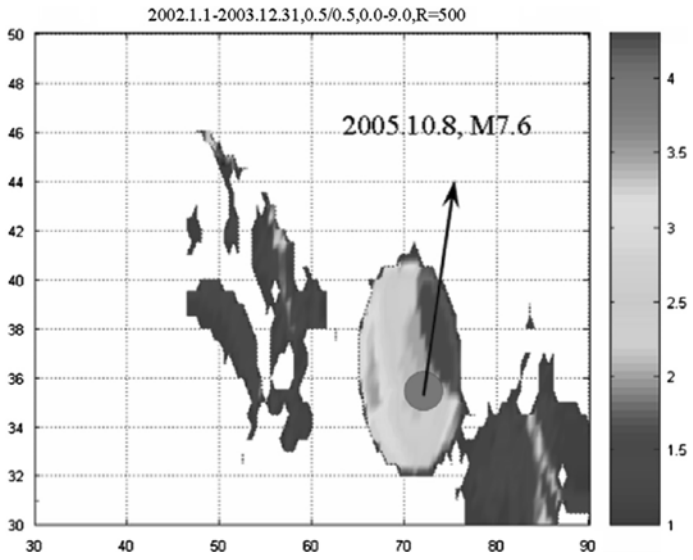


Figure 6
A spatial scanning of LURR in the Europe-Asia seismic belt during the period from 2002.1.1 to 2003.12.31.

The prediction of the Pakistan earthquake with a magnitude of M 7.6 on October 8, 2005 deserves particular mention here. The LURR special scanning of the Euro-Asian seismic belt (30°–50°N, 30°–90°E) was conducted in 2003. The outcomes (Fig. 6) were issued in the Journal *Mechanics in Engineering* (YIN, 2004) and *Supercomputing Newsletter*, sponsored by the Computer Network Information Center, Chinese Academy of Sciences. The Pakistan earthquake which occurred on October 8, 2005 was exactly sited fallen in the high LURR region.

It is also worth mentioning the prediction of the Yangjiang (Guangdong Province in southern China) earthquake with a magnitude of M_s 5 on October 17, 2004. We conducted a LURR spatial scanning of the Chinese mainland on the super-computer DeepComp 6800 in the Supercomputing Center of Computer Network information Center, Chinese Academy of Science on September 1, 2004 (Fig. 7). It is obvious that of all the abnormal areas the Yangjiang region was the most remarkable. Based on this result, we predicted that an earthquake of approximately M 5 would occur very soon in this region. We issued our prediction at the *DeepComp 6800 User Workshop* on 11 September, 2004. Just six days following the Workshop, the Yangjiang region was hit by an M 5 earthquake. This case suggested that LURR probably has the potential ability for short-term earthquake prediction.

3. Discussion

As mentioned above the prediction results of LURR have been continuously improving in recent years. According to our understanding, the reasons for the improvement comprise the following aspects:

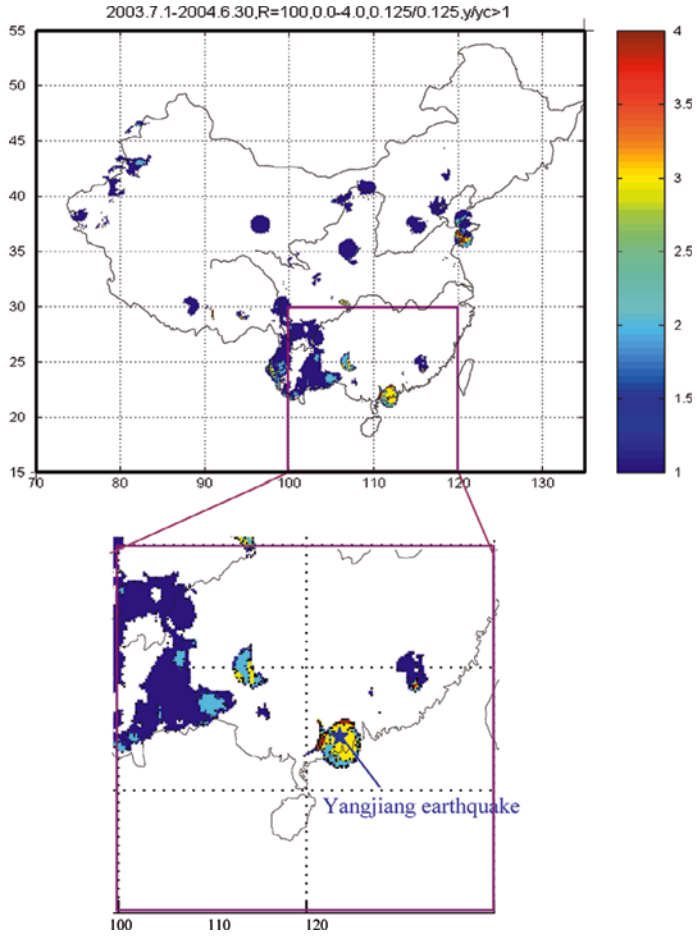


Figure 7
Spatial scan of mainland China on September 1, 2004, we predicted a magnitude 5 earthquake would soon hit the Yangjiang region of the Guangdong province in southern China. Such an event occurred within a month of our prediction.

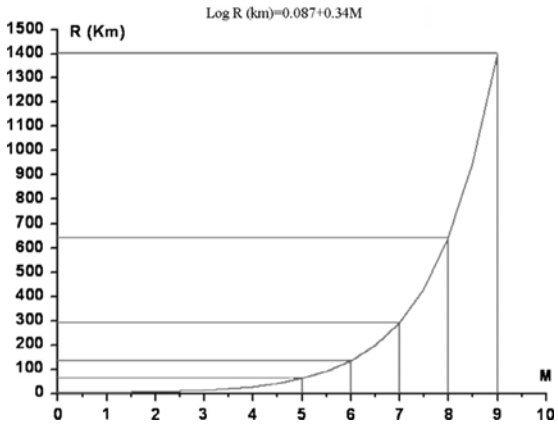


Figure 8
The scaling relation between R (the radius of a circular spatial window) and M (the magnitude of a predicted forthcoming earthquake).

3.1. Proper Scaling

Any physical problem has its scaling relation. It is very important to select proper temporal and spatial windows in researching any physical problem. We determined the scales of temporal and spatial windows according to our previous experience and now they are determined by the scaling laws shown in Figure 8 (YIN *et al*., 2002; PENG *et al*., 2006) for the spatial window and Figure 9 (ZHANG *et al*., 2005; ZHANG, 2006) for the temporal window.

Fourteen earthquakes that occurred in China have been studied with magnitudes ranging from 5.1 to 8.1 (PENG *et al*., 2006). It is found that the critical region size scales with the magnitude of the earthquake. Fitting the data, the function between critical region size and the magnitude is

$$\log R(\mathrm{km}) = 0.087 + 0.34M. \tag{1}$$

Figure 8 shows the function above in a more direct way.

Figure 9 shows the definitions of seismogenic duration. T is the total seismogenic duration which denotes the period from the beginning of the LURR anomaly (LURR $\geq$ 1) to the occurrence of a strong earthquake, T_2 is the period from the peak point of LURR to the occurrence time and T_1 is from the beginning of LURR to the peak point. Obviously

$$T = T_1 + T_2. \tag{2}$$

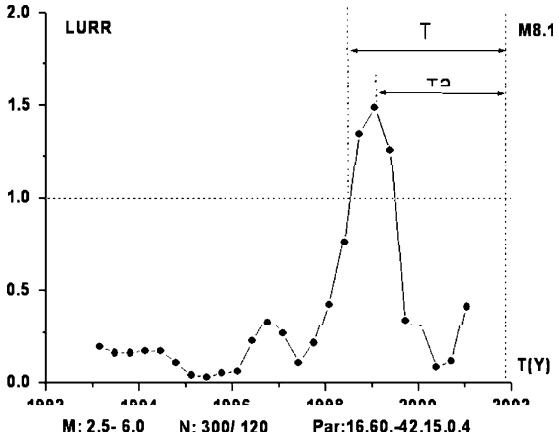


Figure 9
The definitions of seismogenic duration (measured in the month) using a large earthquake with M = 8.1 (Nov. 14, 2001) as an example. T denotes the total seismogenic duration, T_2 denotes the duration from the peak point of LURR to the occurrence time of the earthquake and $T = T_1 + T_2$.

According to our recent study, the peak point of LURR implies the nucleation of the earthquake. Therefore T_2 is the nucleation period which has such manner of function as below (YIN and ZHENG, 1983; ZHENG and YIN, 1984; OHNAKA, 2004),

$$\Delta T = A(1 - B \times 10^{-CM}), \tag{3}$$

where ΔT denotes the nucleation period, *M is the* magnitude of earthquake and *A, B, C* are material constants. According to the data of 30 earthquakes with magnitudes ranging from 5.0 to 8.1 which occurred in the Chinese mainland, we obtained

$$T = 80(1 - 2.5 \times 10^{-0.09M}) \tag{4}$$

and

$$T_2 = 60(1 - 2.3 \times 10^{-0.08M}) \tag{5}$$

where the time is measured in month. Figure 10 shows the function of T and the broken lines defines the error bar.

3.2. Better Parameters Selection

In order to calculate LURR, many parameters are required, such as fault mechanism, depth etc. In order to felicitously select these parameters we have been searching for references from considerable literatures, accumulating experience and conducting the scanning in parameter space according to MFO proposed by Can YIN (YIN, 2005).

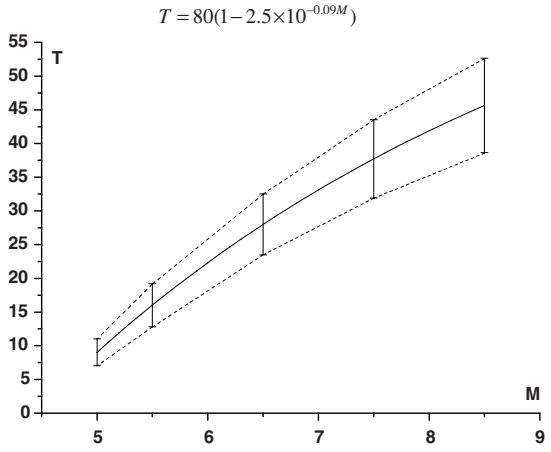


Figure 10
The scaling relation between T (the temporal window measured in the month) and M (the magnitude of a predicted forthcoming earthquake). The broken lines define the error bar.

3.3. Supercomputing

All the works above, especially MFO, need supercomputing. We conducted most of the computing in the Supercomputing Center of Computer Network Information Center, Chinese Academy of Sciences (CAS).

3.4. Earthquake Physics

The last, nonetheless the most important issue, is the deep understanding of the earthquake physics through basic research (RUNDLE *et al*., 2000; MORA *et al*., 2002; WANG *et al*., 2002; SAMMIS and SORNETTE, 2002; XIA *et al.*, 2002; WANG *et al.*, 2004; YIN, 2005; YIN *et al.*, 2004; YU, 2004; ZHANG *et al*., 2005; RONG *et al*., 2006; ZHANG, 2005; YIN *et al*., 2006; YIN, MORA *et al*., 2006). Any deeper advance into earthquake physics within, enhances LURR.

Acknowledgement

The authors gratefully acknowledge the support from MOST under Grant No.2004 CB418406, NSFC under Grant No.10572140, the Informalization Construction of Knowledge Innovation Projects of the Chinese Academy of Sciences “Supercomputing Environment Construction and Application” (INF105-SCE-2-02).

REFERENCES

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.

OHNAKA, M. (2004), *A Constitutive scaling law for shear rupture that is inherently scale-dependent, and physical scaling of nucleation time to critical point*, Pure Appl. Geophys. *161*, 1915–1929.

PENG, P., YIN, X.C., and ZHANG, L.P. (2006), *A statistical investigation of the earthquake predictions using LURR*, Pure Appl. Geophys. *163*, 2353–2362.

RONG, F. (2006), *Multiscale Simulation of Damage Evolution in Heterogeneous Brittle Media*, Ph.D. Thesis, the Graduate School, Chinese Academy of Sciences.

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

SAMMIS, C.G. and SORNETTE, D. (2002), *Positive feedback, memory, and the predictability of earthquakes*, PNAS *99*, 2501–2508.

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. *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*, 2491–2509.

XIANG-CHU YIN and ZHENG TIAN-YU (1983), *A rheological model for the process of preparation of an earthquake*, Scientia Sinica *26*, 285–296.

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

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 The Research on Seismic Tendency of China in 2004*, (ed. Center for Analysis and Prediction, CSB), pp 282–285, Seismological Press, Beijing, 2003 (in Chinese).

YIN, X.C., ZHANG, H.H., YU, H.Z., ZHANG, Y.X., and PENG, K.Y. (2004), *Prediction of seismic tendency of Chinese mainland in 2004 in terms of LURR. In The Research on Seismic Tendency of China in 2005*, pp. 282–285 (ed. Center for Analysis and Prediction, CSB), Seismological Press, Beijing, 2004 (in Chinese).

YIN, X.C., YU, H.Z., KUKSHENKO, V., XU, Z.Y., WU, Z.S., LI, M., PENG, K.Y., ELIZAROV, S., and LI, Q. (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*, 2405–2416.

YIN, X.C. (2004), *A new approach to earthquake prediction-Load/Unload Response Ratio (LURR) theory*, Mechanics in Engin. *26*, 1–7 (in Chinese).

YIN, X.C., ZHANG, L.P., ZHANG, H.H., YIN, C., WANG, Y.C., ZHANG, Y.X., PENG, K.Y., WANG, H.T., SONG, Z.P., YU, H.Z., and ZHUANG, J.C. (2006), *LURR's twenty years and its perspective*, Pure Appl. Geophys. *163*, 2317–2341.

YIN, X.C., MORA, P., DONNELLAN, A., and MATSUURA, M. (2006), *Introduction to computational earthquake physics*, Part I, Pure Appl. Geophys. *163*, 1737–1741 and *Part II*, Pure Appl. Geophys. *163*, 2359–2261.

YU, H.Z. (2004), *Experimental Research on Precursors of Brittle Heterogeneous Media and Earthquake Prediction Methods*, Ph.D. Thesis, Institute of Mechanics, Chinese Academy of Sciences (in Chinese).

ZHANG, H.H., YIN, X.C., and LIANG, N.G. (2005), *Statistic study of LURR anomaly temporal scale before moderately strong earthquakes in Chinese mainland*, Earthquake Research in China. *21*, 486–495.

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, Chinese Academy of Sciences (in Chinese with English abstract which can be found in the National Library of China, Beijing).

ZHANG, H.H., YIN, X.C., LIANG, N.G., YU, H.Z., LI, S.Y., WANG, Y.C., YIN, C., KUKSHENKO, V., TOMILINE, N., and ELIZAROV, S. (2006), *Acoustic emission experiments of rock failure under load simulating the hypocenter condition*, Pure Appl. Geophys. *163*, 2389–2406.

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.

ZHENG TIAN-YU and XIANG-CHU YIN (1984), *The subcritical extension of faulting and process of preparation of earthquake*, Bull. Sci. China *29*, 8, 1081–1086.

(Received September 6, 2006, revised March 30, 2007, accepted April 27, 2007)

Pure appl. geophys. 165 (2008) 723–736
0033–4553/08/030723–14
DOI 10.1007/s00024-008-0322-z

Pure and Applied Geophysics

Earthquake Trend Around Sumatra Indicated by a New Implementation of LURR Method

CAN YIN,[1,2] HUILIN XING,[1,2] PETER MORA,[1,2] and HEHUA XU[1]

Abstract—The current implementation of LURR (Load/Unload Response Ratio) theory has shown promise in intermediate to short-term earthquake forecasting through past practice, but has also met difficulties at the same time. In this paper a new implementation of LURR method is presented to address one of the problems. The major change in the new implementation is that LURR values will result from a calculated maximum faulting orientation (MFO) instead of any focal mechanism selected by users. After comparison with the world stress map, the calculated MFO has been found to be in good agreement with the observation from the regional tectonic stress regime. The MFO pattern in the Indonesia region has a special feature which may be related to the unique subduction complexity. The MFO pattern in the Sumatra region in general is different from that in the Java region after the 2004 M 9.0 Sumatra Andaman Islands earthquake. This phenomenon may be supported by the evidence of the recent observation that a section in the southern part of the Sumatran arc is locked. Furthermore, the MFO pattern before the 2004 main shock is different from that after the event. Retrospective examination of the Indonesia region by means of this implementation can produce significant LURR anomaly not only prior to the 2004 main shock but also before the 2006 M 7.7 South Java earthquake. Therefore future great earthquakes might favorably be forecasted if the LURR anomaly detected by MFO method could be considered a precursor.

Key words: LURR, earthquake prediction, Sumatra, MFO, critical instability.

1. Background

The 2004 Boxing Day Earthquake/Tsunami in Sumatra was one of the most destructive disasters in modern history. Comparatively the magnitude 7.7 earthquake which occurred in South Java on 2006.07.17 and killed hundreds of people seems trivial. In the aftermath, a fundamental question is inevitably raised: Could the earthquake have been forecasted? This question is so difficult to answer that there may be no confirmative answer with the current level of science and technology. Nonetheless, earth scientists must investigate this issue, otherwise the huge loss of life will become meaningless.

[1] Earth Systems Science Computational Centre (ESSCC), The University of Queensland, St. Lucia, QLD 4072, Australia. E-mail: canyon@esscc.uq.edu.au; h.xing@uq.edu.au

[2] Australian Computational Earth Systems Simulator (ACcESS), Major National Research Facility, The University of Queensland, St. Lucia, QLD 4072, Australia

Probing such a question involves many disciplines of sciences, including geology, geodesy, geophysics and of course some prediction methods.

The region under investigation covers a $20° \times 20°$ square area delimited by latitude 10°S to 10°N and longitude 90°E to 110°E. Because this area is within the subduction zone formed by the Indo-Australian plate and the Burma microplate, the tectonic stress field is very complicated (Fig. 1). The Burma plate moves southwesterly while the Indo-Australian plate moves towards the north. Such oblique collision generates not only

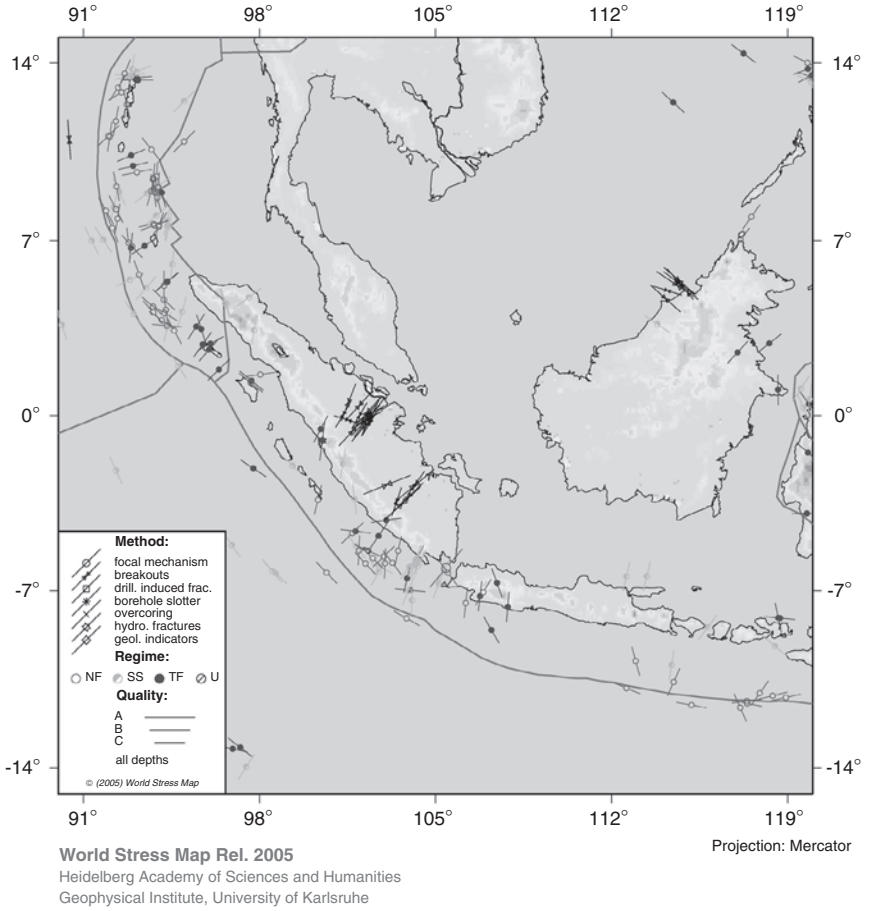


Figure 1
Stress map of the Sumatra region (generated through the World Stress Map, website: http://www.world-stress-map.org).

strike-slip events but also thrust earthquakes, some of them being megathrust earthquakes like the 2004 Sumatra-Andaman Earthquake (Fig. 2). The focal mechanism of the 2004 main shock displays a thrust faulting with the compression axis in NE-SW direction, indicating a southwesterly abrupt movement of the hang wall that caused the huge tsunami. The relationship between tsunami and earthquake is also a complicated topic as manifested by the insignificant tsunami generated by the M 8.6 earthquake which occurred close by 3 months later. However, that topic is beyond this paper.

Geodetic observations of the crustal deformation in the Sumatra region indicates that a section south of the epicenter (from 0.5°S to 6°S) has been fully locked for the past 50 years (SUBARYA *et al*., 2006), implying that stress concentration may have been

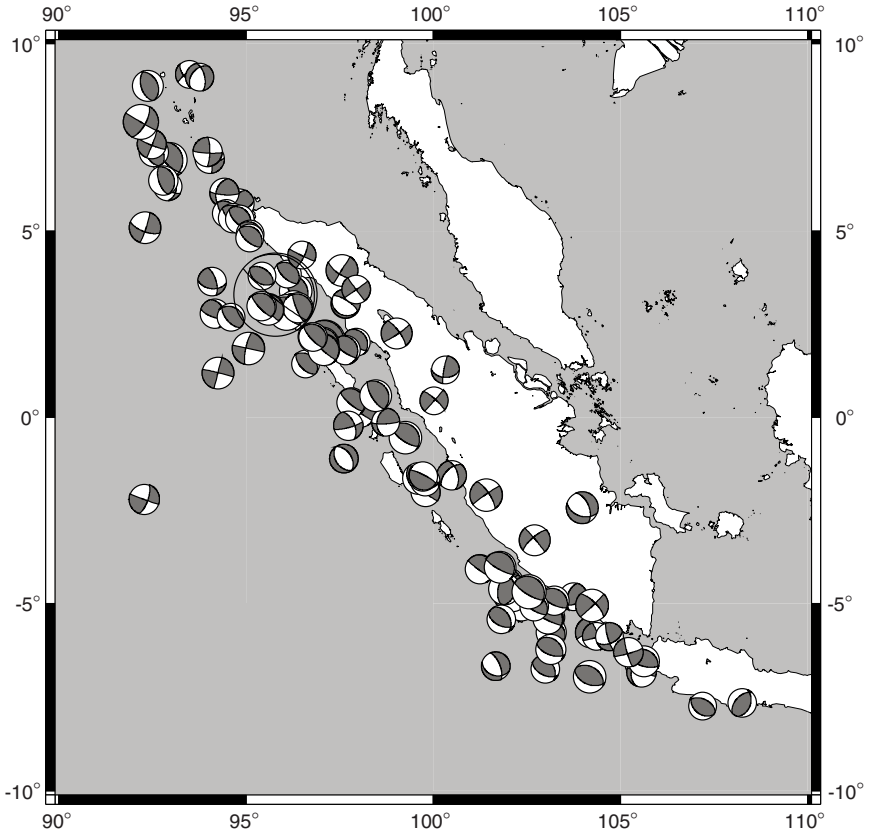


Figure 2
Focal mechanisms of $M \geq 6$ earthquakes between 1990 and 2005. (Data from Harvard CMT project). The main shock of the 2004 M9.0 Sumatra Andaman Islands earthquake is plotted without fill (transparent).

built-up in this section. Is this the place where the next great earthquake will occur? On the other hand, because of the nonlinear nature of tectonic movement, a locked-up block of crust may either break up very soon or remain locked for a very long time. It is difficult to determine when a future earthquake will occur.

To assess the earthquake trend in the near future a certain prediction method is required to identify unambiguous precursors shortly before the main events. The LURR (Load/Unload Response Ratio) theory is exactly an earthquake prediction method for this purpose. Even though the predictability of earthquakes is still a controversial topic and a perfect prediction method is yet to emerge, the practice of retrospective and prospective forecasting using LURR method over the past 20 years has shown very good promise. It aims at intermediate-term to short-term forecasting and is well positioned to assess the earthquake trend in the next couple of years.

2. Introduction of the LURR Method

The Load/Unload Response Ratio theory has been proposed for nearly two decades (YIN and YIN, 1991; YIN *et al*., 1995 2000). The basic idea was developed from critical sensitivity study of nonlinear systems which states that: When a mechanical system, such as a block of crust, is stable, its response to loading is similar to its response to unloading; but when the system is close to instability, its response to loading is much greater than its response to unloading. By comparing such difference at different stages, the stability status of the system will be indicated. The parameter that quantifies such a difference is termed Load/Unload Response Ratio. If LURR value reaches its peak, a major earthquake will probably occur within a short time, usually several months to two years according to past experience (YIN and YIN, 1991; Yin *et al*., 1995, 2000), though a clearly quantified timing relationship is yet to be determined due to the problematic knowledge regarding the damage process.

In the current implementation, where the exponential seismic energy release, E, is used as response and the duration of loading, T, as general loading, LURR is defined as follow:

$$Y \equiv \frac{\sum_{t\in T_+} E^m(t)/T_+}{\sum_{t\in T_-} E^m(t)/T_-} = \frac{(\sum_{i=1}^{N_+} E_i^m)_+}{(\sum_{i=1}^{N_-} E_i^m)_-} \Big/ \frac{T_+}{T_-}, \tag{1}$$

where the sign means loading and unloading. The sum of energy E denotes total energy release that can be calculated according to Gutenberg-Richter formula $\log E = a + bM$ (b = 1.5 in general) while N_+ and N_- denote the number of earthquakes which occurred during the loading and unloading period respectively. The total number of earthquakes must be sufficient to obtain a statistically stable result. Besides, the reliability of the results requires that the earthquake catalogue should be as complete as possible so that all response is guaranteed inclusion. The exponential factor m = 0, 1/3, 1/2, 2/3 or 1 and

usually takes 1/2 to denote the Benioff strain. The benefit of using Benioff strain is that the effect of deviation from G-R relationship at either end of the magnitudes will be neutralized. Using $m = 1/2$ is also a compromise because $m = 0$ (simple count) means smaller events have the same weight as larger ones, thus being unfair to larger ones, while $m = 1$ (energy) puts too much weight on larger events, thus resulting unfair to smaller ones.

Using Benioff strain still favors larger events and is one of the reasons why the LURR method is still controversial. SMITH and SAMMIS (2004) argue that the LURR peak prior to the main shock is simply caused by some larger foreshock(s). Although their implementation of LURR was quite different from the authors', which results in very different conclusions in many cases, their observation may lead to an interesting deduction: If foreshocks would always occur during tidal loading instead of unloading, the predictability of LURR method would have been proven. If a large earthquake is found to occur during tidal loading, it is very likely a larger one will follow shortly.

Moreover, the assessment of earthquake trend using LURR is not simply looking at the LURR vs. time curve only, it also involves the study of spatial variation pattern of LURR (ZHANG *et al*., 2004). The combined study of the spatial and temporal pattern of LURR contributed to successful predictions (ZHANG *et al*., 1999).

The near periodical deformation of the Earth induced by gravitational forces of celestial bodies (Sun, Moon, etc.) serves as loading and unloading after special processing. The stress due to both solid earth tides and ocean loading effect can be calculated with relatively high precision. The Coulomb failure hypothesis is adopted to judge loading/unloading according to the sign of the increment of tidal Coulomb Failure Stress (CFS), which is defined as

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

where f, τ_n and σ_n refer to intrinsic frictional coefficient, shear stress and normal stress (positive in tension) respectively, $\mathbf{n}$ the normal of a specific fault plane on which the tidal loading and unloading is defined.

The only uncertainty in (2) is the choice of $\mathbf{n}$, which is critical to the LURR method. Ideally, each earthquake should be assigned its own fault plane. However, it is impossible to obtain fault planes for most small earthquakes used in the LURR method because of the limitation of the current technique. Harvard CMT can provide focal mechanisms of earthquakes with a magnitude as small as 4.0, which nonetheless is not small enough to allow the sufficient number of events required for LURR calculation. Considering this difficulty, a uniform fault plane is used for ALL earthquakes instead. It is surmised that the majority of earthquakes occur on this plane or closely orientated planes. Subsequently the problem becomes how to find this uniform plane.

The would-be focal mechanism should facilitate earthquake occurrence most. If LURR theory is correct, this plane should make LURR value the maximum among all possible focal mechanisms. This particular plane is termed as MFO (maximum faulting

orientation or most facilitating orientation). Because MFO is calculated using past earthquakes, it does not require any prerequisite assumption on fault plane and thus makes prospective prediction more practical (YIN and MORA, 2006).

3. Retrospective Examination by LURR Method

Retrospective examination of the 2004 main shock was conducted first. The procedure of LURR calculation is as follow:

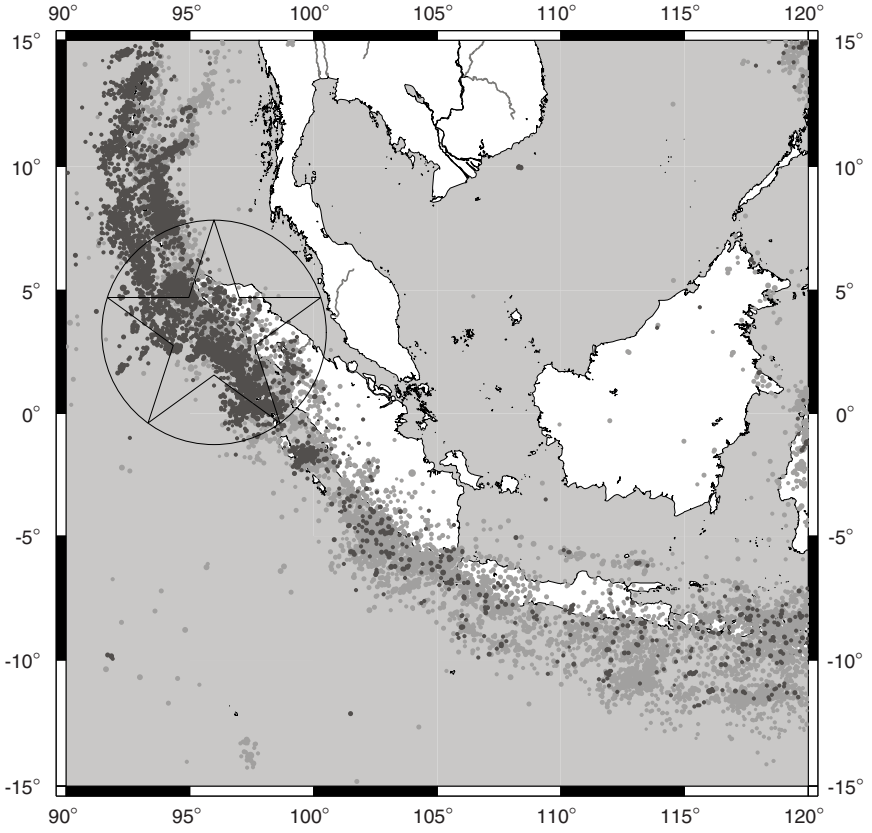


Figure 3
Earthquake distribution in the studied area. Darker spots represent those which occurred on and after 26/12/2004 while the lighter ones before 26/12/2004. The large starred circle indicates the region within 500 km radius of the 2004 main shock epicenter.

- Firstly, earthquakes which occurred within the concerned region are collected as completely as possible. The data used in this paper came from the ANSS (Advanced National Seismic System) composite catalog (http://www.ncedc.org/anss/cnss-catalog.html). The distribution of events is illustrated in Figure 3. Because of insufficient data for this region, a large window with a radius of 500 km around the epicenter of the main shock (E95.85°, N3.32°) is used. The time window is 10 years with a step of two months. No limit on magnitude.
- Next the tidal stress at the epicenter of each event at the time of its occurrence is calculated.
- Then the LURR values for all possible focal mechanisms are calculated and compared in order to identify the focal mechanism which produces the maximum, i.e., MFO. The MFO obtained over the past 24 years (from 1980 to 2004) is (130°,60°,110°), where the three angles refer to strike, dip and rake, respectively as in convention (Top of Fig. 4). The result is obtained using 10° increments of strike, dip and rake in the search process. Smaller increments may produce more accurate but likely unnecessary MFO at the cost of much longer computational time.
- Then the tidal Coulomb Failure Stress (CFS) on MFO is calculated for every event. Based on whether tidal CFS is positive or negative, the events that occurred in the loading period are separated from those in the unloading period.
- Finally the ratio of the total seismic energy released in the loading period to that in the unloading period is calculated. After being modified by the ratio of loading time duration to unloading time duration, the result is taken as our LURR value.

4. Results and Discussion

The calculated LURR vs. time curve is shown in Figure 5. An outstanding feature of the plot is that the LURR value is rarely lower than 10. This pattern may be caused by some systematic bias and should be a localized phenomenon.

Nonetheless, the more important pattern is the shape of the LURR curve. The result shows that since 2000, the LURR value kept increasing until reaching its peak by the end of 2002 and then started decreasing prior to the main shock. Such a sequence (increase - peak - decrease - main shock) is a typical LURR pattern prior to most other major earthquakes while no such sequence exists in stable regions (YIN and YIN, 1991; YIN *et al*., 1995, 2000), suggesting LURR can serve as a precursor. The obvious LURR anomaly demonstrates that the 2004 Sumatra earthquake might have been forecasted. On the other hand, because prolonged time (10 years length with 2 months step) and space (500 km radius) windows were used due to the difficulty of data acquisition, our result could not distinguish any precursor for the March 2005 M 8.6 earthquake which was just three months later and 100 km away from the 2004 M 9.0 earthquake.

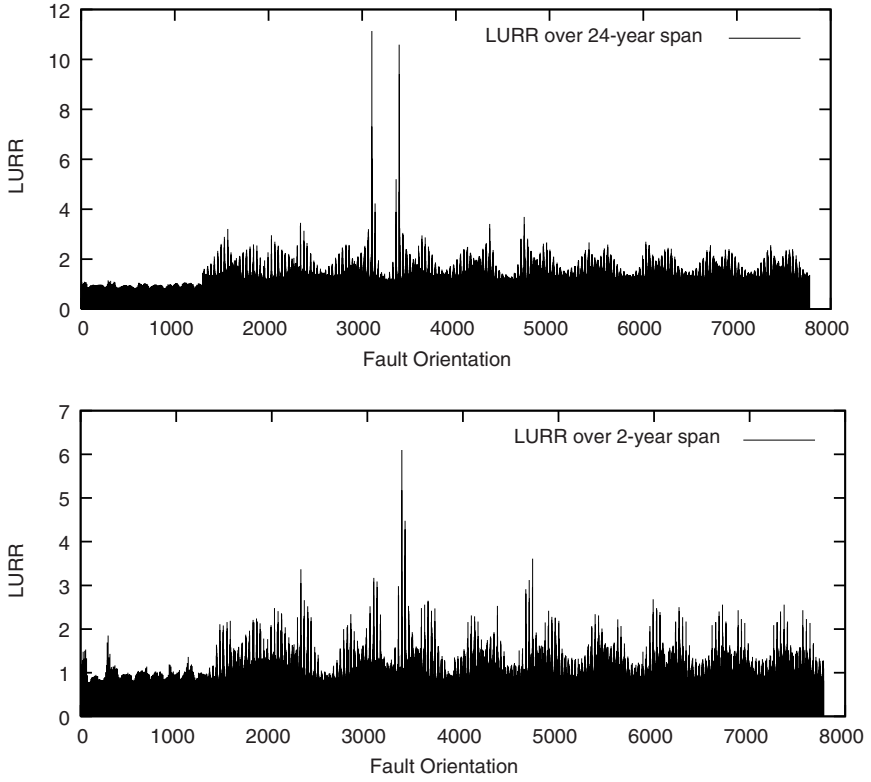


Figure 4
LURR values calculated for all possible focal mechanisms. Radius of window is 500 km, no limitation on magnitude. (Top) The focal mechanism corresponding to the maximum LURR is (60°, −40°, −110°) for a time span of 24 years (26/12/1980–25/12/2004); (bottom) The focal mechanism corresponding to the maximum LURR is (60°, 60°, 60°) for a time span of 2 years (26/12/2002–25/12/2004).

The observed focal mechanism of this earthquake is (329°, 8°, 110°) while the auxiliary nodal plane is (129°, 83°, 87°) (Harvard CMT solution). The calculated 24 year MFO is similar to the auxiliary plane. Interestingly, MFO over 1 or 2 years prior to the main shock is (60°, 60°, 60°) (bottom of Fig. 4), very different from the 24-year MFO.

Compared with the stress map in this region (Fig. 1), it is clear that although the short-term MFO is very different from the long-term one (and the actual focal mechanism of the main shock), it does not contradict the stress regime of this region. It agrees better with the stress regime of the west of the epicenter than with that of the east part. The main shock focal mechanism agrees with the eastern stress regime much

better. The mechanism of (60°, 60°, 60°) means a nearly north-south orientated compression direction. If LURR theory is justifiable, the above scenario may be explained as this: The long-term NE-SW orientated compression brings the two plates into a state of lockup. To break up the lock, a lateral push might be the most effective way utilizing common sense. Probably there was a significant change in the tectonic stress field where the north-south orientated component came into play rather than NE-SW direction which always dominated the stress field. It would be desirable if the stress field around the epicenter over the past decade could be continually monitored. The observed stress evolution pattern would probably reveal some clues leading to this great event. Before this idea could be realized, part of the stress evolution may be revealed by studying MFO pattern only.

Because MFO reflects the orientation in which the upcoming major event is mostly facilitated, investigation of the MFO distribution before and after the main shock may reveal some pattern of tectonic stress evolution. In Figure 6 the MFO calculated over 5000 days before 2004.12.25 for every position within the concerned area where MFO is calculable is plotted. The radius of window is 200 km. Because LURR method is only dependent on earthquake catalogue, the location where LURR method is applicable is also dependent on the availability of earthquake catalogue there. A cutoff is set on the amount of events: Only locations with more than 50 earthquakes within a 200 km radius are examined. According to Figure 6, the 2004 main shock occurred within a small area where the MFO is quite uniform but differs from surrounding ones. After the main shock the MFO became more uniform in a much larger area around the epicenter as shown in Figure 7, which is calculated using events between 01/01/2005 and 01/07/2006.

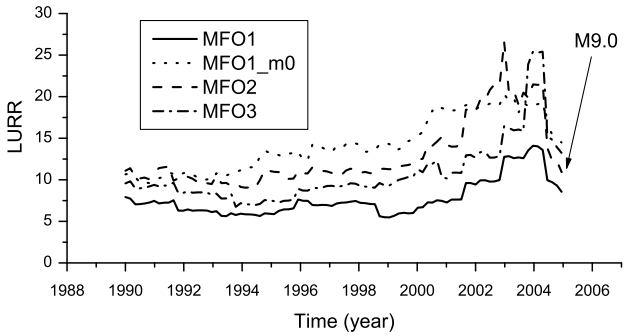


Figure 5
LURR vs. time curves before the 2004 M 9.0 Sumatra-Andaman Islands earthquake. Time window is 10 yrs with step of 2 mths, radius 500 km. Since the end of 2000, LURR continually rises for over a decade and falls just before the main shock. MFO1 refers to (130°, 60°, 110°), MFO2 (50°, 60°, 70°) and MFO3 (50°, 60°, 60°). MFO1-m0 is the result of simple count (m = 0).

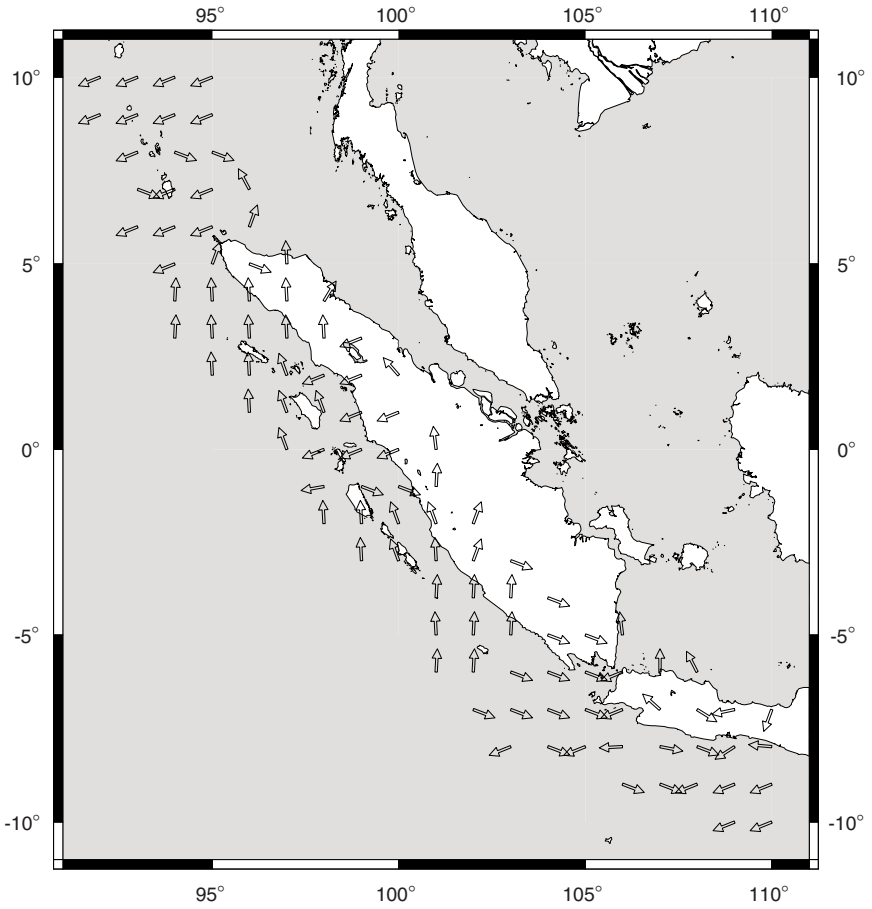


Figure 6
Horizontal projections of MFO for a 5000-day period which ended on 25/12/2004.

While the orientation of MFO can indicate the tectonic stress change, the place where LURR reaches its greatest value will define the position of upcoming major events. Figure 8 shows the values of LURR that have been calculated in Figure 6, while Figure 9 depicts the values of LURR that are calculated using events between 01/01/2005 and 01/07/2006. In both cases, the epicenters of the main shocks fall within the area delineated by the highest LURR.

To date most discussions regarding the targeted main shock are only qualitative. We are fairly confident in that the highest LURR will correspond to the location of the upcoming main shock. However, the magnitude and time of the target earthquake are yet

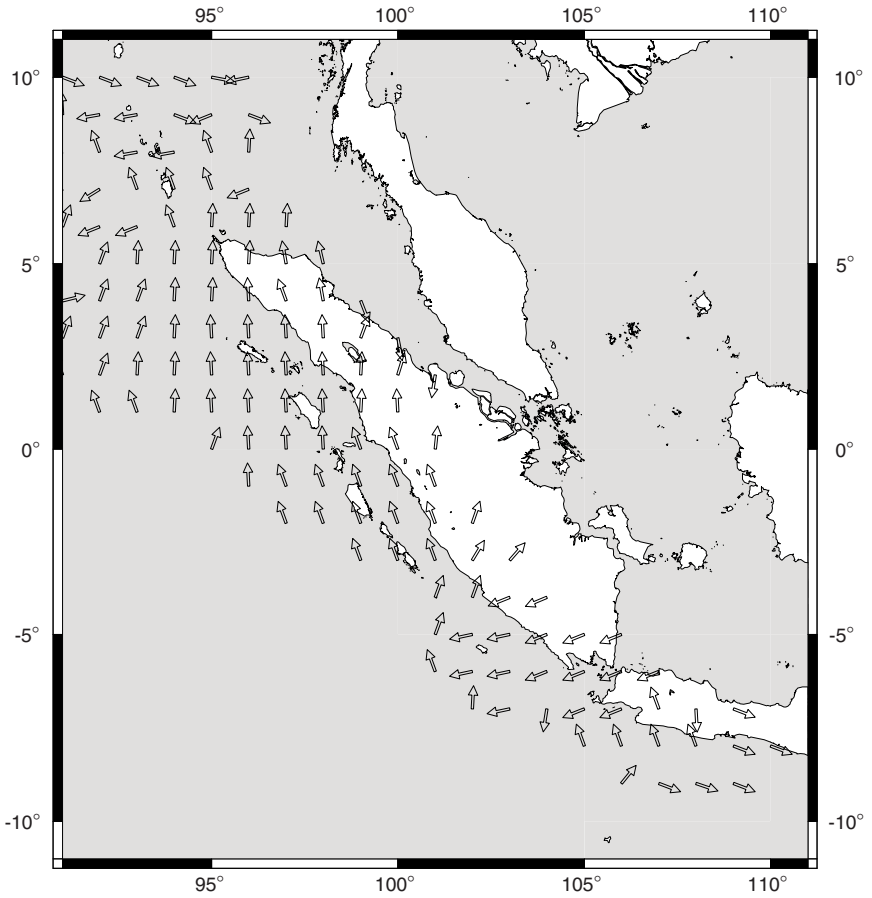


Figure 7
Horizontal projections of MFO for the time span from 01/01/2005 to 01/07/2006.

to be quantified. A good quantitative relationship between LURR and the targeted major earthquake is still elusive. Possibly the combination of the reorientation pattern and size of MFO will provide a new path towards the solution of this issue.

5. Conclusion

The introduced MFO brings new meanings to the LURR method. Taking both the temporal and spatial patterns of MFO variation into consideration, we conclude that the MFO contains richer information about the tectonic plate movement and the formation of

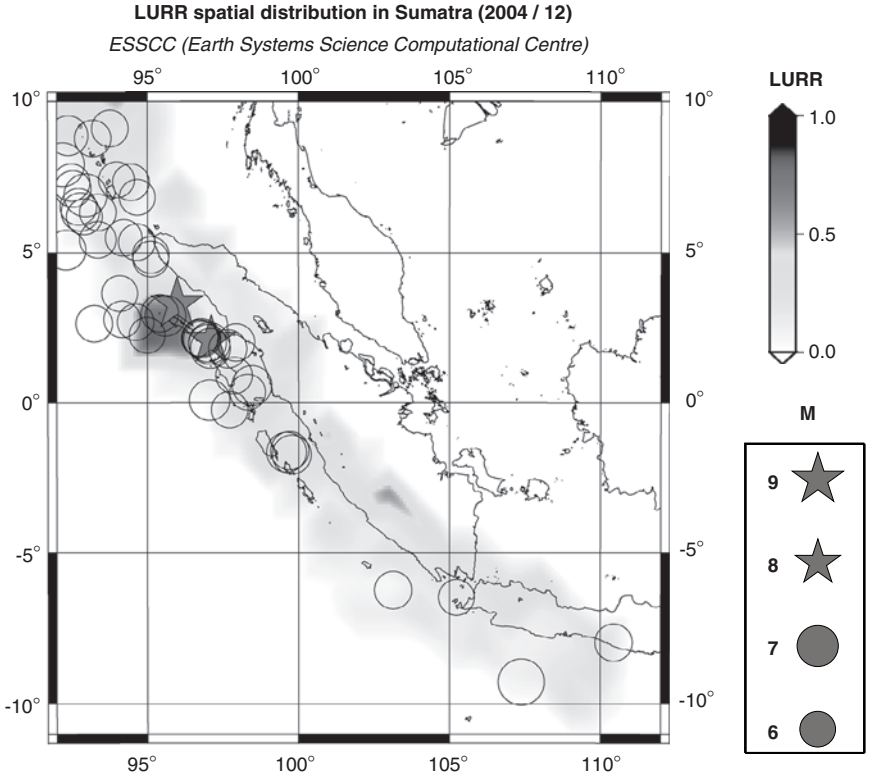


Figure 8
Spatial scanning of LURR for the Sumatra region before the main shock (up to 25/12/2004). The value of normalized LURR (Y value divided by the maximum LURR value in this region) is indicated by its color with the darkest for the highest. Notice that the upcoming main shock would fall into the darkest zone.

earthquakes than the original implementation of LURR theory. It is more robust and easier to apply, although it is still in the experimental stage because more examples are needed to substantiate its credentiality. The authors have to indicate that due to the very limited source in the Sumatra region, a more complete data set may reveal different scenarios. The conclusion presented here is preliminary and subject to such change.

Acknowledgements

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

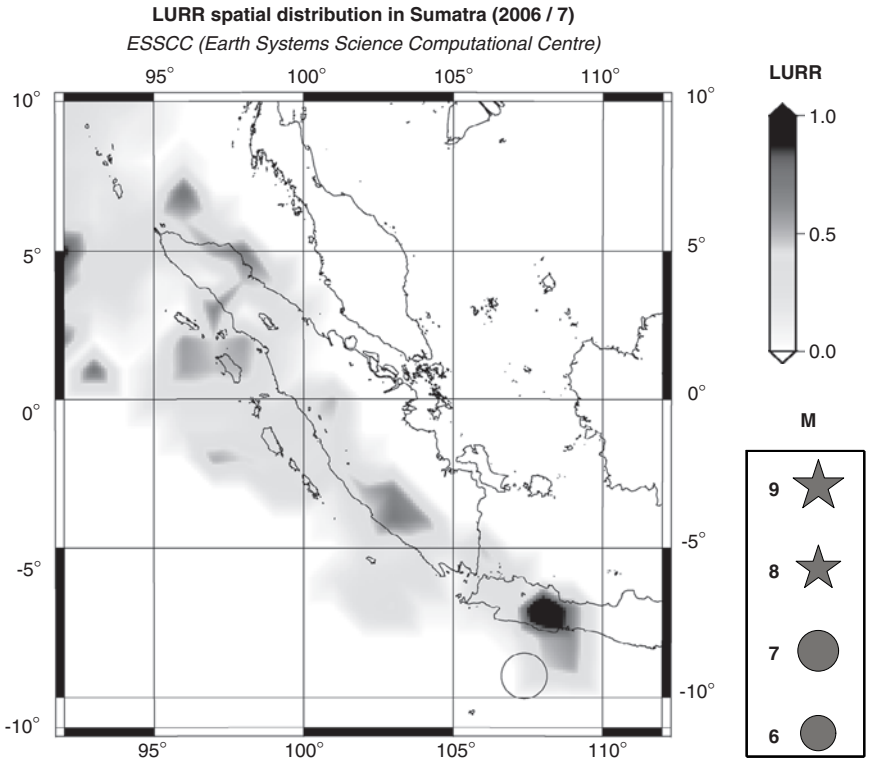


Figure 9
Spatial scanning of LURR for the Sumatra region before the 2006 South Java earthquake (earthquakes include from 2005.01.01 up to 01/07/2006). LURR values are normalized as in Fig. 8.

the Queensland State Government Smart State Research Facility Fund and SGI. All plots are generated by GMT (Generic Mapping Tool) version 3.4.

References

Reinecker, J., Heidbach, O., Tingay, M., Sperner, B., and Muller, B. (2005), *The release 2005 of the World Stress Map* (available online at http://www.world-stress-map.org).

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

Subarya, C., Chlieh, M., Prawirodirdjo, L., Avouac, J., Bock, Y., Sieh, K., Meltzner, A., Natawidjaja, D., and McCaffrey, R. (2006), *Plate-boundary deformation associated with the great Sumatra-Andaman earthquake*, Nature *440*, 46.

Yin, C., and Mora, P. (2006), *Stress reorientation and LURR: Implication for earthquake prediction using LURR*. Pure Appl. Geophys. *163*, 2363–2373.

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

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

ZHANG, G., ZHU, L., SONG, X., LI, Z., YANG, M., SU, N., and CHEN, X. (1999) *Predictions of the 1997 strong earthquakes in Jiashi, Xinjiang, China.* Bull. Seismol. Soc. Am. *89*, 1171–1183.

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

(Received September 30, 2006, revised August 14, 2007, accepted August 27, 2007)

Pure appl. geophys. 165 (2008) 737–748
0033–4553/08/030737–12
DOI 10.1007/s00024-008-0320-1

Pure and Applied Geophysics

Comparison Between LURR and State Vector Analysis Before Strong Earthquakes in Southern California Since 1980

YONGXIAN ZHANG,[1,2] YONGJIA WU,[1] XIANGCHU YIN,[2,3] KEYIN PENG,[1,2] LANGPING ZHANG,[2] AIQIN YU[4], and XIAOTAO ZHANG[3]

Abstract—There are seven strong earthquakes with M $\geq$ 6.5 that occurred in southern California during the period from 1980 to 2005. In this paper, these earthquakes were studied by the LURR (Load/Unload Response Ratio) method and the State Vector method to detect if there are anomalies before them. The results show that LURR anomalies appeared before 6 earthquakes out of 7 and State Vector anomalies appeared before all 7 earthquakes. For the LURR method, the interval between maximum LURR value and the forthcoming earthquake is 1 to 19 months, and the dominant mean interval is about 10.7 months. For the State Vector method, the interval between the maximum modulus of increment State Vector and the forthcoming earthquake is from 3 to 27 months, but the dominant mean interval between the occurrence time of the maximum State Vector anomaly and the forthcoming earthquake is about 4.7 months. The results also show that the minimum valid space window scale for the LURR and the State Vector is a circle with a radius of 100 km and a square of 3°×3°, respectively. These results imply that the State Vector method is more effective for short-term earthquake prediction than the LURR method, however the LURR method is more effective for location prediction than the State Vector method.

Key words: LURR, state vector, characteristics of anomaly, earthquake prediction, Southern California.

1. Introduction

LURR (Load/Unload Response Ratio) is an earthquake prediction method put forward by Yin (1987). LURR (Load/Unload Response Ratio) is defined as (YIN, 1987)

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

[1] China Earthquake Networks Center, Beijing 100045, China. E-mail: yxzhseis@sina.com; wuyongj@mails.gucas.ac.cn; pengkeyin@seis.ac.cn

[2] LNM, Institute of Mechanics, Chinese Academy of Sciences, Beijing 100080, China. E-mail: yin@lnm.imech.ac.cn; zhanglp@lnm.imech.ac.cn

[3] Institute of Earthquake Science, China Earthquake Administration, Beijing 100036, China. E-mail: yinxc@btmail.net.cn; qdzxt_116@163.com

[4] China Earthquake Disaster Prevention Center, Beijing 100029, China. E-mail: yuaiqin2006@126.com

where X^+ and X^- are the response rates during loading and unloading measured by some method. The intention of this concept is that 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 stronger than that to unloading.

Based on the theory of LURR and its recent development (YIN, 1987; YIN and YIN, 1991; YIN *et al*., 1995, 2000; WANG *et al*., 1999; MORA *et al*., 2000a,b, 2002), spatial and temporal variation of Y/Y_c 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). Here Y_c is the critical value of LURR that depends on the number of earthquakes under different specified confidence levels (ZHUANG and YIN, 1999). 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). Here we chose the confidence level as 99%. The scanning results show that obvious Y/Y_c anomalies occurred before 5 of the total 6 earthquakes with $M \geq 6.5$, the area with Y/Y_c anomalies is near the epicenters of the strong earthquakes, and the Y/Y_c anomalies occur months to years prior to the earthquakes. After March, 2002, another earthquake with $M \geq 6.5$ occurred in Southern California; that is the San Simeon M 6.5 (35.7°N, 121.1°W) earthquake on Dec. 22, 2003. In order to obtain the evolutionary process of LURR before this quake, ZHANG *et al*. (2004, 2006) calculated Y/Y_c in Southern California with earthquake catalogue of ANSS from April 2002 to June 2004 under the same calculation parameters. The results show that the characteristic of LURR before the San Simeon earthquake is similar to the former ones.

In recent years, a new earthquake prediction method named State Vector was put forward by YIN *et al*. (2004a,b). Their results show that significant anomalies occurred several months prior to the Tangshan M 7.8 and Haicheng M 7.3 earthquakes. YU *et al*. (2004) confirmed the phenomena of State Vector anomaly before rock failure by a rock mechanical test. WU *et al*. (2006) applied this method to study 25 strong earthquakes with $M \geq 6.8$ in the China mainland and obtained the results that there are obvious State Vector anomalies three years prior to 19 out of 25 strong earthquakes and among these 19 strong earthquakes there are obvious state vector anomalies 60 days before 10 of them. These results show considerable promise for short-term earthquake prediction.

Is the State Vector effective in other tectonic regions? What is the difference between the characteristics of the LURR and the State Vector? In order to obtain the answers, we apply the State Vector method to the Southern California region to see if there are obvious anomalies before strong earthquakes in this region, and compare the characteristics of State Vector anomalies to those of LURR.

The earthquake catalogue we used in this paper is downloaded from the web site http://www.ncedc.org/anss/catalog-search.html, which is the *ANSS* (*A*dvance *N*ational *S*eismic *S*ystem) composite earthquake catalogue.

2. Method to Calculate LURR and State Vector

2.1. LURR

LURR (Load/Unload Response Ratio) is an approach to testing for crustal criticality put forward by YIN (1987). Its hypothesis is that an earthquake is the failure or instability of the focal media or seismogenic system. When the 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). YIN (1987) defined LURR in formula (1).

In the LURR theory, Y is defined directly by means of seismic energy as follows:

$$Y_m = \frac{[\sum_{i=1}^{N^+} E_i^m]_+}{[\sum_{i=1}^{N^-} E_i^m]_-}, \tag{2}$$

where E denotes seismic energy which can be calculated according to the Gutenberg-Richter formula (KANAMORI and ANDERSON, 1975; BULLEN and BOLT, 1985), the "+" sign means loading and "−" means unloading, $m = 0$ or 1/3 or 1/2 or 2/3 or 1. When $m = 1$, E^m is exactly the energy itself. For $m = 1/2$, E^m denotes the Benioff strain. For $m = 1/3$ and 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^-, and N^+ and N^- denote the number of earthquakes which occur during the loading and unloading periods. In this paper, m is chosen as 1/2, which means that Y is determined by the ratio of Benioff strain during the loading period over the unloading period, and Y_c is chosen as Y value under the confidence level of 99%.

Earthquake energy E_i in formula (2) is related to magnitude by the following formula (GUTENBERG and RICHTER, 1956)

$$\log_{10} E_i = 11.8 + 1.5 M_i, \tag{3}$$

where M_i is the magnitude of i-th earthquake, and the unit of energy E_i is erg ($\times 10^{-5}$ J).

The periods of loading and unloading are determined by calculating perturbations in the Coulomb Failure Stress ΔCFS induced by earth tides (YIN *et al*., 1995; ZHANG *et al*., 2006). We divided the Southern California region into 11 units (ZHANG *et al*., 2004, 2006), and in each of them the stress distribution (ZOBACK, 1992) and fault parameters SCECDC (Southern California Data Center) are generally uniform.

In order to speed up the calculations and avoid the perturbation caused by outstanding earthquakes, we chose magnitude thresholds according to the linear part of the Gutenberg–Richter relation in each unit area.

According to the relationships between the spatial window, time window and the magnitude obtained in the earthquake case studies conducted by YIN *et al*. (2002a,b) and ZHANG *et al*. (2005), when the spatial window is chosen as a circular region with the radius of about 100 km, and the time window is one year, the magnitude of the forthcoming earthquake could be predicted as larger than M 6. In this region a value of

Y/Y_c (LURR/critical LURR) was calculated for a specific time window (1 year), and the time step is 1 month. The circle center was moved step by step in both latitude and longitude by increments of 0.25 degrees so the contour of Y/Y_c in each month could be obtained.

2.2. State Vector

The State Vector is an idea from statistical physics (REICHL, 1980). YIN *et al*. (2004b) extended this idea to describe the evolution of the damage of rock specimens. The whole specimen is divided into n regions. The AE (Acoustic Emission) energy or AE event rate at time t_k in region i denotes the i-th component e_i, and the whole n components construct a state vector $\vec{V}_t$ of n dimensions (YIN *et al*., 2004b).

$$\vec{V}_t = (e_1(t), e_2(t), \ldots, e_n(t)). \tag{4}$$

In earthquake prediction practice, e_i in formula (4) is defined as the sum of the logarithms of each earthquake energy E_j in the i-th subsquare during the period from $t-T$ to t, as the following

$$e_i(t) = \sum_{j=1}^{k} \lg E_j \Big|_{t-T}^{t}. \tag{5}$$

Here T refers to the time window for calculation, and the earthquake energy E_j is calculated by formula (3).

Figure 1 illustrates the schematic diagram of time window T and time step Δt in the calculation.

In our study, the time widow is chosen as $T = 1$ year and the time step is $\Delta t = 30$ days.

The following parameters associated with State Vector could be obtained:

a. Modulus of the State Vector at time t

$$M_t = \left|\vec{V}_t\right|. \tag{6}$$

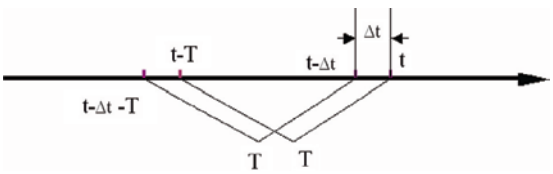


Figure 1
Schematic diagram of time window T and time step Δt.

b. Modulus of the increment of the State Vector during time t-Δt to t

$$\Delta M_{t-\Delta t,t} = \left|\Delta \vec{V}_{t-\Delta t,t}\right|. \tag{7}$$

c. Angle between State Vectors from time t-Δt to t

$$\varphi_{t-\Delta t,t} = \arccos\left(\frac{\vec{V}_{t-\Delta t} \cdot \vec{V}_t}{V_{t-\Delta t} V_t}\right). \tag{8}$$

When $n = 3$, the State Vector could be illustrated by 3D phase shown in Figure 2. $\vec{V}_1$ stands for the State Vector at time t_1, $\vec{V}_2$ is the State Vector at time t_2, and $\vec{V}_3$ is the State Vector at time t_3. $\Delta\vec{V}_{12}$ is the increment of the State Vector from t_1 to t_2, and $\Delta\vec{V}_{23}$ is the increment of the State Vector from t_2 to t_3. φ_{12} is the angle between $\vec{V}_1$ and $\vec{V}_2$.

In order to study the evolution of the State Vector before strong earthquakes in Southern California, we chose a square region of $m^\circ \times m^\circ$ subsquares with the epicenter of the target strong earthquake at the center, where m is varied from 3 to 11 to obtain the result with the best goodness of fit. To determine the best goodness of fit, we calculate the parameters of State Vector in different space scales. If obvious variation occurs before the quake and there are no other major changes in other periods (say, the change of the parameter is larger than two times the standard deviation), we take the space scale as the critical scale for the seismogenic region (Wu *et al*., 2006).

Considering the completeness of the earthquake catalogue and avoiding the turbulence of the result affected by major earthquakes, we chose the cutoff of the magnitude as $2.0 \leq m_j \leq m_0 - 0.5$, where m_0 is the magnitude of the target earthquake.

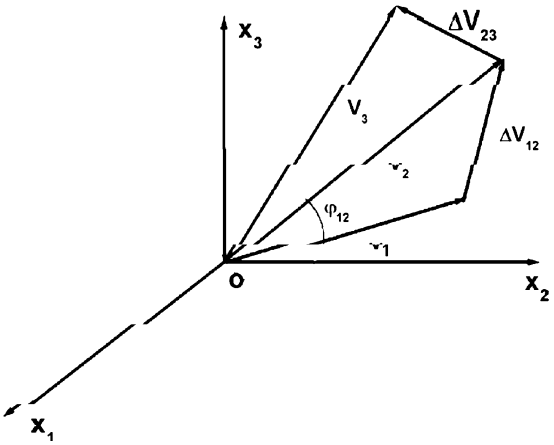


Figure 2
State Vector in 3D phase.

3. Characteristics of the LURR Anomaly before Strong Earthquakes in Southern California

According to ZHANG *et al*. (2004, 2006), among the seven strong earthquakes with $M \geq 6.5$ in Southern California since 1980, obvious LURR anomalies occurred prior to six of them. The characteristics of the LURR anomalies are listed in Table 1.

We can see from Table 1 that except for the Coalinga earthquake in 1983, Y/Y_c anomalies occurred before the other six earthquakes in Southern California. The forthcoming strong earthquake occurred 1 to 19 months after the time when Y/Y_c reached its peak value. According to the common conception of long-term (several years to ten years), medium-term (one year to several years), short-term (several months) and imminent (several days to ten days) earthquake prediction in China (MEI *et al*., 1993), Table 1 implies that LURR is valid for medium-term to short-term prediction. The dominant Δt is from 8 to 13 months, and the mean value is about 10.7 months. This might be applied to earthquake forecasting in this region.

Table 1 also shows that the distance between the epicenter of the forthcoming earthquake and the maximum Y/Y_c point is from 0 to 200 km, and the dominant Δ (km) is about 100 km. This implies that if the location of the maximum Y/Y_c point is detected, we might forecast that there probably will be a strong earthquake with $M \geq 6.5$ occurring within 200 km from the maximum Y/Y_c point.

4. Characteristics of the State Vector Anomaly before Strong Earthquakes in Southern California

For the seven strong earthquakes in Southern California in Table 1, we calculated the time series of the modulus of the State Vector, the modulus of the increment of

Table 1

Characteristics of Y/Y_c before strong earthquakes in Southern California since 1980.

Date	Epicenter and location	Magnitude /Δ (km)	Max Y/Y_c	Lasting time of anomalous Y/Y_c (month)	$\Delta\tau$ (month)
1983.5.2	(36.23°N, 120.32°W) Coalinga	6.7/None	None	None	None
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

Notes: Δ (km) is the distance between the forthcoming earthquake and the maximum Y/Y_c point.
$\Delta\tau$ is the interval between the occurrence time of the maximum Y/Y_c anomaly and the occurrence time of the forthcoming earthquake.
* The distance between the maximum point of ΔM and the forthcoming 1994 earthquake is 200 km, not 80 km as mentioned in the early paper (ZHANG *et al*., 2004). We corrected it in the later paper (ZHANG *et al*., 2006), but did not give the annotation.

the State Vector, and the angle between the State Vectors according to formula (4) to (6).

Since the parameter of the angle between State Vectors does not show a regular feature before these earthquakes, we only discuss the characteristics of the modulus of the State Vector and the modulus of the increment of the State Vector in this paper.

Here we give examples of the evolution of the modulus of the State Vector and the modulus of the increment of the State Vector before the Coalinga earthquake.

Figure 3 shows the curve of ΔM with time before the Coalinga M 6.7 earthquake. For this earthquake we chose square regions with scales ranging from $3^\circ \times 3^\circ$ to $11^\circ \times 11^\circ$ with the epicenter of the Coalinga earthquake at the center. The curve with the scale of $3^\circ \times 3^\circ$ shows the obvious anomaly of the State Vector before the quake, so we take the region of $3^\circ \times 3^\circ$ ($m = 3$) as the critical scale for the seismogenic region.

From this figure we can see that the maximum ΔM value occurred on Jan. 1, 1983, and the Coalinga earthquake occurred 4 months later on May 2, 1983. Two smaller peaks of ΔM appeared on May 1, 1980 and May 1, 1981 without earthquakes larger than M 6.5 occurring after them. However, what interests us is that, during the period from 1970 to May 2, 1983, two earthquakes of M 6.2 (37.50°N, 118.81°W) and M 5.9 (37.50°N, 118.81°W) occurred on May 27, 1980 and Sept. 30, 1981, respectively. In fact, there were only these three earthquakes with $M \geq 5.9$ in this selected region since 1970. What is the relationship between the two peak values and the M 6.2 and M 5.9 earthquakes? Are there any ΔM anomalies before smaller earthquakes like the two earthquakes of M 6.2 and M 5.9? We will study this phenomenon in the future.

For the modulus of the State Vector M, the result also shows positive prospect. Figure 4 shows the curve of M with time before the Coalinga earthquake. From this figure, we can see that the maximum M value occurred on April 1, 1983, and the Coalinga

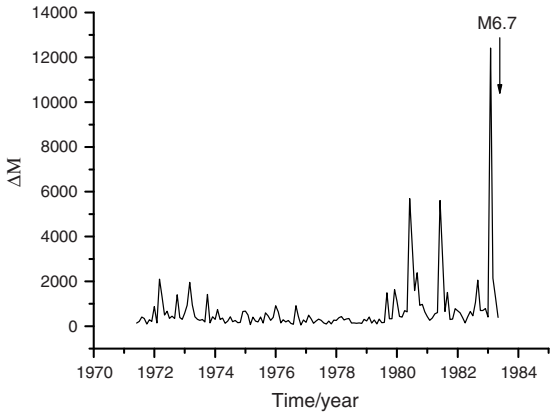


Figure 3
Evolution of the modulus of state vector increment before the Coalinga earthquake.

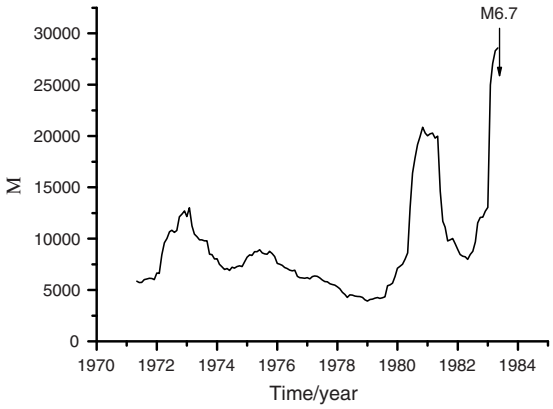


Figure 4
Evolution of the modulus of State Vector before the Coalinga Earthquake.

earthquake occurred 1 month later on May 2, 1983. For the second peak of M, which appeared on Oct. 1, 1980, there is a M 6.2 earthquake (37.50°N, 118.81°W) that occurred on May 27, 1980, that did not occur after the peak. However, M rises from the base value of about 13052 (on Jan.1, 1983) to 28591 (on April 1, 1983) before the M 6.2 earthquake. In other words, M rises rapidly before strong earthquakes, however some earthquakes might occur after M reaches the peak value, and others might occur before M reaches the peak value.

The critical scale of $3° \times 3°$ shows the obvious $\Delta M_{\max}$ and $M_{\max}$ anomalies preceding the Coalinga earthquake. The $\Delta M_{\max}$ anomaly occurred on Jan. 1, 1983, about 4 months before the earthquake, and $M_{\max}$ occurred on April 1, 1983, about 1 month before the earthquake.

In the same way, we obtained the characteristics of $\Delta M_{\max}$ and $M_{\max}$ before the seven strong earthquakes in Southern California since 1980, as shown in Table 2.

From Table 2, we can summarize that for strong earthquakes with $M \geq 6.5$ in Southern California, the interval between $\Delta M_{\max}$ and the forthcoming strong earthquake is from 3 to 27 months, and the average of $\Delta\tau_1$ is about 9 months. The interval between $M_{\max}$ and the forthcoming strong earthquakes is from 1 to 71 months, and the average of $\Delta\tau_2$ is 22.4 months.

Table 2 also shows that the critical scale for different strong earthquakes is different. The largest scale is $8° \times 8°$ for the Loma Prieta M 7.0 earthquake, while for the Hector Mine M 7.1 earthquake, the critical scale is only $3° \times 3°$. Why the critical scales for different earthquakes with similar magnitude differ so much is reserved as a question.

Table 2

Characteristics of ΔM_{max} and M_{max} before strong earthquakes in Southern California since 1980.

Date	Epicenter and location	Magnitude	Date 'of ΔM_{max} / $\Delta\tau_1$ (month)	Date of M_{max} / $\Delta\tau_2$ (month)	m
1983.5.2	(36.23°N, 120.32°W) Coalinga	6.7	1983.1.1/4	1983.4.1/1	3
1987.11.24	(33.01°N, 115.85°W) Superstition Hills	6.6	1987.6.24/5	1987.5.24/6	4
1989.10.18	(37.04°N, 121.88°W) Loma Prieta	7.0	1987.7.18/27	1987.05.18/29	8
1992.6.28	(34.20°N, 116.44°W) Landers	7.3	1992.03.28/3	1992.05.28/1	4
1994.1.17	(34.21°N, 118.54°W) Northridge	6.6	1993.06.17/7	1993.03.17/10	5
1999.10.16	(34.59°N, 116.27°W) Hector Mine	7.1	1999.02.6/8	1996.07.16/39	3
2003.12.22	(35.7°N, 121.1°W) San Simeon	6.5	2003.01.22/11	1998.01.22/71	5

Note: ΔM_{max} is the maximum value of the modulus of increment of state vector.
M_{max} is the maximum value of the modulus of the State Vector.
$\Delta\tau_1$ is the interval between time of the modulus of the maximum increment State Vector and the occurrence time of the forthcoming earthquake.
$\Delta\tau_2$ is the interval between time of the modulus of the maximum State Vector and the occurrence time of the forthcoming earthquake.
m is the number of grids along latitude or longitude direction of the square region.

5. Comparison of the Characteristics of the LURR and the State Vector

Table 1 and Table 2 list the characteristics of LURR and State Vector anomalies before 7 strong earthquakes with M $\geq$ 6.5 in Southern California since 1980. Here we compare three kinds of intervals, $\Delta\tau$, $\Delta\tau_1$ and $\Delta\tau_2$, before each strong earthquake. $\Delta\tau$ denotes the interval between the time of the peak LURR and the occurrence time of the forthcoming earthquake, $\Delta\tau_1$ is the interval between the time of the modulus of the maximum increment State Vector and the occurrence time of the forthcoming earthquake, and $\Delta\tau_2$ is the interval between the time of the modulus of the maximum State Vector and the occurrence time of the forthcoming earthquake.

Figure 5 shows that the maximum LURR values appear within 12 months before the Superstition Hills, Loma Prieta, Landers and Hector Mine earthquakes. The maximum *M* values appear within 12 months before the Coalinga, Superstition Hills, Landers and Northridge earthquakes, while the maximum ΔM values appear within 12 months before Coalinga, Superstition Hills, Landers, Northridge, Hector Mine and San Simeon earthquakes. The above results suggest that the ΔM parameter is more valid for earthquake prediction within 12 months, than the *M* parameter, and for almost half of the earthquakes (e.g., Superstition Hills, Landers, Northridge), the LURR and State Vector methods both work well.

In order to provide a simple and easy understanding of the effects of LURR and State Vector in time forecasting, we draw the spectrum of $\Delta\tau$, $\Delta\tau_1$ and $\Delta\tau_2$ in Figure 6. We can see from this figure that 4/7 of $\Delta\tau$ distribute from 9 to 13 months, and the mean value of the dominant $\Delta\tau$ is 10.7 months; 4/7 of $\Delta\tau_1$ distribute mainly from 3 to 11 months, and the mean value of $\Delta\tau_1$ is 4.7 months; 4/7 of $\Delta\tau_2$ distribute from 1 to 10 months, and the mean value of $\Delta\tau_2$ is 4.5 months. The results imply that the State Vector is a valid

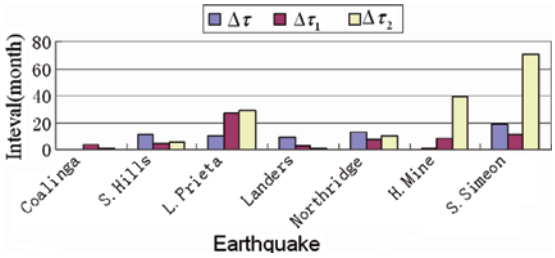


Figure 5
$\Delta\tau$, $\Delta\tau_1$ and $\Delta\tau_2$ for the strong earthquakes in Southern California since 1980.

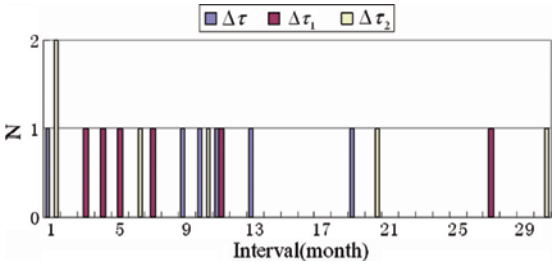


Figure 6
Spectrum of $\Delta\tau$, $\Delta\tau_1$ and $\Delta\tau_2$.

method for short-term earthquake forecasting, and the LURR is a valid method for medium-term earthquake forecast in Southern California since 1980. The ΔM parameter seems to be more stable than the M parameter.

6. Conclusions and Discussion

There are only seven strong earthquakes with M $\geq$ 6.5 in Southern California since 1980. From the above results, we can draw the following conclusions:

(1) LURR anomalies occurred before 6 earthquakes out of 7 and State Vector anomalies occurred before all of these 7 earthquakes.
(2) For LURR, the interval between the maximum Y/Y_c value and the forthcoming earthquake ranges from 1 to 19 months, and the dominant mean interval is about 10.7 months. For the State Vector method, the interval between the maximum modulus of the increment State Vector and the forthcoming earthquake ranges from 3 to 27 months, but the dominant mean interval between the occurrence time of the maximum State Vector anomaly and the forthcoming earthquake is about 4.7 months.

(3) The minimum valid space window scale for the LURR and State Vector is a circle with a radius of 100 km and a square of $3° \times 3°$, respectively.

From the above results, the State Vector method seems to be more effective for short-term earthquake prediction than the LURR method, however the LURR method is more effective for location prediction than the State Vector method.

The above conclusions are based on limited earthquake cases in Southern California. Are they tenable for other strong earthquakes in other regions? YIN *et al*. (2004a,b) and WU *et al*. (2006) studied the earthquake cases in China, and their results show that several months before many strong earthquakes, significant State Vector anomalies did occur. YU *et al*. (2004) also confirmed the phenomena of the State Vector anomaly before rock failure by rock mechanical testing, which proved the solid mechanical base for the State Vector method. However, more earthquake cases should be studied before drawing the confirmed conclusions.

Some questions are left behind in this study, such as why the critical scales for different earthquakes with similar magnitude differ so much. Is the square region the best to calculate parameters related to the State Vector? How should the threshold for the State Vector calculation be chosen? Is the subdivision of Southern California into 11 areas reasonable? There are many details to be concerned with and statistical checks to be done before the LURR and the State Vector could be used as earthquake predictors.

Acknowledgements

The authors gratefully acknowledge the support from Special Funds for Major State Basic Research Project under Grant No.2004 CB418406, 004DKA50740, 2002CB412706, 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) and Chinese Joint Seismological Foundation (305016). We also express our thanks to ANSS for earthquake data.

We especially thank the revisers who raised questions and offered good suggestions to improve our paper in grammar, organization, logic, etc.

REFERENCES

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

GUTENBERG, B. and RICHTER, C.F. (1956), *Magnitude and energy of earthquakes*, Annali di Geofisica *9*, 1–15.

KANAMORI, H. and ANDERSON, D.L. (1975), *Theoretical basis of some empirical relations in seismology*, Bull. Seismol. Soc. Am. *65*(5), 1073–1095.

MEI, S.R., FENG, D.Y., ZHANG, G.M., ZHU, Y.Q., GAO, X. and ZHANG, Z.C., *Introduction to Earthquake Prediction in China* (Beijing, Seismological Press, 1993) 35–36 (in Chinese with English Abstract).

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

MORA, P., WANG, Y.C., YIN, C., PLACE, D., and YIN, X.C. (2002), *Simulation of the Load-Unload Response ratio and critical sensitivity in the lattice solid model*, Pure Appl. Geophys. *159*(10), 2525–2536.

REICHL, L.E. (1980), *A Modern Course in Statistical Physics* (University of Texas 1980).

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.

Wu, Y.J., Zhang, Y.X., and Yin, X.C. (2006), *Characteristics of State Vector anomaly before 25 earthquakes ($M \geq 6.8$) in China mainland since 1970*, Earthquake (in Chinese with English abstract) *26* (2), 61–67.

YIN, X.C. (1987), *A new approach to earthquake prediction*, Earthq. Res. in 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., 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., 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*, 2511–2514.

YIN, X.C., ZHANG, H.H., PENG, K.Y., 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. and ZHANG, Y.X. (2004a), *State Vectors—A new method to analysis seismicity*, Earthq. Res. in China *18*(3), 263–270.

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*, 2405–2416.

YU, H.Z., YIN, X.C., ZHANG, Y.X. *et al*. (2004), *Experimental study of State Vector*, Earthquake *24*(3), 1–7 (In Chinese with English Abstract).

ZHANG, H.H., YIN, X.C., and LIANG, N.G. (2005), *Statistic study of LURR anomaly temporal scale before moderately strong earthquakes in Chinese mainland*, Earthq. Res. in China *21*, 486–495.

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*, 11–12, 2359–2367.

ZHANG, Y.X, YIN, X.C. *et al*. (2006), *LURR and the San Simeon M 6.5 earthquake in 2003 and the seismic tendency in CA*, Pure Appl. Geophys. *163*, 2343–2351.

ZHUANG, J.C. and YIN, X.C. (1999), *Random distribution of the Load/Unload Response Ration (LURR) under assumptions of Poisson model*, Earthq. Res. in 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 September 6, 2006, revised March 31, 2007, accepted August 30, 2007)

Pure appl. geophys. 165 (2008) 749–759
0033–4553/08/030749–11
DOI 10.1007/s00024-008-0313-0

Pure and Applied Geophysics

Fast Computing for LURR of Earthquake Prediction

YANGDE FENG,[1] XUEBIN CHI,[1] WU WANG,[1] JIANG CHEN,[2] and XIANGCHU YIN[3]

Abstract—The LURR theory is a new approach for earthquake prediction, which achieves good results in earthquake prediction within the China mainland and regions in America, Japan and Australia. However, the expansion of the prediction region leads to the refinement of its longitude and latitude, and the increase of the time period. This requires increasingly more computations, and the volume of data reaches the order of GB, which will be very difficult for a single CPU. In this paper, a new method was introduced to solve this problem. Adopting the technology of domain decomposition and parallelizing using MPI, we developed a new parallel tempo-spatial scanning program.

Key words: Load-unload response ratio (LURR); temporal-spatial scanning, optimization schemes.

1. Introduction

The Load-Unload Response Ratio (LURR) Method, which was invented by Prof. Yin (YIN, 1987), has achieved successful results in earthquake prediction within the China mainland as well as other regions in America, Japan and Australia (ZHANG *et al*., 2004; YIN *et al*., 1904; MORA *et al*.).

The main idea of the LURR earthquake prediction approach is that when a system is in a stable state, its response to a small loading is nearly the same as that to unloading, but when the system is near failure or in an unstable state, the response to loading and unloading becomes quite different (YIN *et al*., 1995, 2000). LURR is defined according to this difference. P and R are the load and response of a system respectively, if P has a small change ΔP resulting in a small change to R of ΔR, then we can define X as

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

[1] Supercomputing Center Computer Network Information Center Chinese Academy of Sciences, P.O. Box 349, Beijing 100080, China. E-mail: ydfeng@sccas.cn
[2] Software Engineer, Intel Asia-Pacific Research and Development Ltd., Shanghai, China.
[3] LNM, Institute of Mechanics, Chinese Academy of Sciences, Beijing 100080, China.

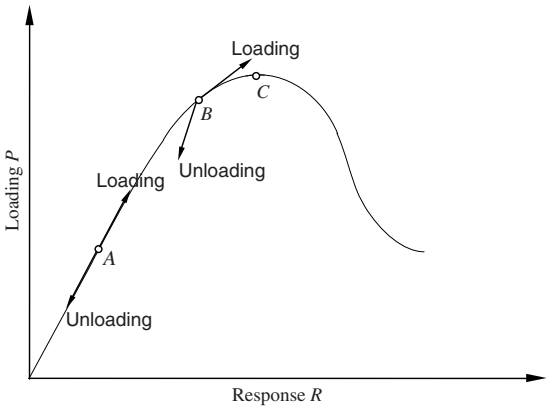


Figure 1
Constitutive relation of systems.

This is thus called the response ratio, and LURR is defined as

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

where X^+ and X^- are response ratio during loading and unloading. As shown in Figure 1, when a system is in a stable or linear state, $X^+ \approx X^-$ then $Y \approx 1$. When a system lies beyond the linear state, $X^+ > X^-$ and $Y > 1$. Hence, Y can be used as a criterion to judge the degree of stability of a system.

In LURR theory, Y is defined directly by the seismic energy as follows:

$$Y = \lim_{\Delta p \to 0} \frac{E^+}{E^-},$$

where E^+ and E^- are seismic energy during loading and unloading (ZHANG *et al*, 2000). 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 judge whether a high Y value can be considered an earthquake precursor at a specified confidence level. They gave the critical value of LURR Y_c that depends on the number of earthquakes at different specified confidence levels. For instance, at the confidence level of 90%, Y_c equals 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 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 LURR).

In this paper we give the critical region of LURR by Y/Y_c instead of Y at a confidence level of 99%.

With the expansion of the prediction region, the refinement of its longitude and latitude, and the increase of the time period, the computation overburden will be very high, and the volume of data reaches the order of GB, which will be very difficult for a single CPU to deal with. In this paper, a new method was introduced to solve this problem. Adopting the technology of domain decomposition and parallelizing using MPI, we developed a new parallel tempo-spatial scanning program based on Yin's previous work.

2. Optimization and Testing of LURR

We analyzed the algorithm as well as the source code of the tempo-spatial scanning program of LURR and optimized it. It was developed in four aspects: Eliminating redundant computation, eliminating unnecessary computation, improving algorithm formula and parallelizing. The code sample data set provided by Prof. Yin is used to measure the optimization of each aspect.

In this sample, the spatial region is from longitude 70.00°E to 135.00°E, and latitude 20.00°N to 55.00°N, earthquake data are from April 1, 2004 to March 31, 2005, and the earthquake spatial scale is R = 200 km, while the spatial precision for scanning is 0.1^{o}.

2.1. Eliminating Redundant Computation

The reason for eliminating redundant computation is that some values used in calculations are fixed during a calculation. This recomputation wastes massive time. Therefore, we redesigned the program by letting these values be computed in advance and be stored in an array. When needed, they will be recalled from the array directly instead of being recomputed.

According to the source code analysis and the program performance profiling, it is known that calculating the distance between a location and epicenter takes the longest time in the whole procedure. This distance is two points' distance of great circles on the spherical surface. The following formula can be used in the procedure to calculate the distance:

$$d = R_e \times \sqrt{\cos^2(lat_z) + \cos^2(lat_c) - 2\cos(lat_z)\cos(lat_c)\cos(lon_z - lon_c) + (\sin(lat_z) - \sin(lat_c))^2}, \quad (3)$$

where R_e is the Earth's radius, lon_c and lat_c are longitude and latitude of the epicenter, and lon_z and lat_z are longitude and latitude of the assigned location, respectively. The latitude and longitude are transformed from the angle value to the radian value. We discover that sine and cosine of latitude or longitude are independent between the

assigned location and the epicenter mutually. Thus these trigonometric functions can be calculated in advance.

In the example, the program will compute 4.25 billion times of d value in total, and each computation will call trigonometrical function 5 times, so that there will be 21.25 billion times of trigonometrical functions calls. However, the assigned location only has 651 longitudes and 351 latitudes; and there are 18,604 historical epicenters, thus 18,604 latitude values and 18,604 longitude values in total. Therefore, only 76,420 times of trigonometrical function calls are needed for all 38,210 of the sine and the cosine values of latitude and longitude values, which is just 1/2,780,000 of original calls.

In the procedure, another long-time costing operation is the calculation of the assigned location's loading or the unloading response in the earthquake regions. The following equation is used,

$$E = \sqrt{10^{(11.8+1.5M)}} = 10^{(11.8+1.5M)/2} = 10^{5.9+0.75M}, \tag{4}$$

where M is Richter magnitude scale. E must be computed once for each location in the earthquake spatial regions. In the present procedure, the LURR is only related to the Richter magnitude scale, namely, LURR of an earthquake is a constant and needs to be calculated once only. In the example, LURRs need to be computed 22,826,006 times, but after optimization only 18,604 times are needed.

2.2. Eliminating Unnecessary Computation

In each point, when the LURR of tempo-spatial scanning is computed, the program will judge this point whether in the spatial scope of the earthquake first, and then call formula (3) to calculate the distance between this point and the epicenter. But an earthquake spatial scale is very small compared to the entire scanning region. Therefore, a place will be judged whether it is in the spatial scope of the earthquake by comparing its latitude and longitude to the epicenter's. If it is not in the scope, its distance will not be computed. Obviously, this will save computing time.

In Figure 2, the pink region is the spatial scope of the earthquake. The latitude and longitude of earthquake spatial scope which is the green region that is obtained by the earthquake spatial scale R and the latitude and longitude of the epicenter; it is circumscribed quadrilateral of the earthquake spatial scope. Because it is on the spherical surface, the pink region is not circularity but the upper part is wider than the lower part, and the green region is not a square. In order to guarantee that the green region can contain the entire pink region, a toleration is given while computing. For a point, the program will test the latitude and longitude to judge this point as to whether it is in the green region first. If not, d will not be computed, as 'Spot 2' shown in Figure 2. If in the green region, the distance d between the point and epicenter can be calculated. Then the program will judge the point as to whether it is in the pink region based on d or R. If it is in, the LURR is calculated, as 'Spot 1' shown in Figure 2. This can save massive time to compute the spherical surface distance. In the example, there are more than 4.25 trillion times of

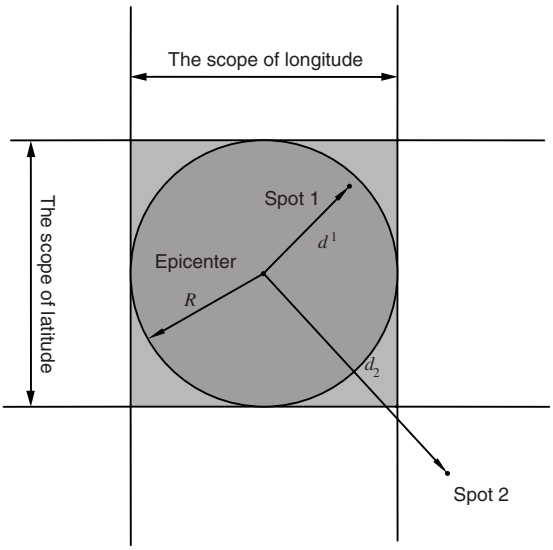


Figure 2
Estimating some point in the earthquake region. The quadrilateral is the spatial scope of being scanned; the pin region is the spatial scope of earthquake.

judging the point as to whether it is in the earthquake spatial scope, after optimization, the times of spherical surfaces distance computation is 32,092,365, which is just 1/132 of the original times.

2.3. Improving Algorithm Formula

In the procedure, to calculate the distance between two points on the spherical surface will require a lot of time. Thus, it is possible to make some improvement to obtain the same or the approximate result by using less operation. The spherical surface is approximated to the plane in the original formula to compute the distance d. The times of calling trigonometrical functions are 5, the times of calling the extraction of square root are one, and the times of floating point multiplications are 7, the times of floating point plus-minus operations are 5. The exact formula of the distance d of two points on the spherical surface is given as the following:

$$d = R_e \times \arccos\left(\sin(lat_z)\sin(lat_c) + \cos(lat_z)\cos(lat_c)\cos(lon_z - lon_c)\right). \quad (5)$$

In this formula, the times of calling trigonometrical functions are 5, the times of call anti-trigonometric functions are one, and the times of floating point multiplications are 4, the times of floating point plus-minus operations are 2.

After the approximate formula is replaced by equation (5), the times of floating point multiplications decrease 3 and the times of floating point plus-minus operations decrease 3, as well as the extraction of the square root which is replaced by inverse cosine function; the operating times are reduced. The LURR effects a slight change, whereas the, results of earthquake prediction cannot be affected.

2.4. Parallel Computing

Based on MPI (Message Passing Interface) library, we parallelized computing of Temporal and Spatial Scanning of LURR. The actual spatial region is divided into small spatial regions, each of which is computed by one processor. There are many different methods by which to partition the actual spatial region, such as, block distribution, cyclic distribution, block-cyclic distribution, and continue.

The computation volume of each point in the spatial region is different; in particular, after the program is optimized by several steps above, the computation mainly concentrates in the earthquake spatial scope. According to the block distribution, there are different volumes of computation in each small region which can cause the load imbalance between processors, and decrease the parallel performance of the procedure. If the cyclic distribution is adopted, two neighboring longitudes and latitudes can be computed by the different processor, and it makes the load balanced and boosts the procedure performance. The nonoverlapped domain decomposition method is used; this is because the partitioned small domains do not tie up each other. The communication amount is decreased on the boundary.

Because the data of the each step of the longitude and the latitude scanning to be computed are recorded, considerable data need high frequency storage which will require large amounts of communication time dealing with the data reduction. In order to solve this problem, many temporary files are opened to record data by each processor separately. Once all processes are finished, the output document will be written by one main processor.

3. Performance Testing

We have tested the original program and the optimized procedure in two kinds of different platforms, respectively. One platform is that CPU is Pentium 4 E 2.8 GHz, with level 1 cache of 32 KB. Level 2 cache of 1 MB, and memory of 384 MB. The operating system is Windows XP, and the compiler is Intel Visual FORTRAN Compiler 9.0 for Windows. We used optimization compiling option "/Ox".

Another platform is DeepComp 6800 Cluster which has 265 nodes, and each of them is four Itanium II 1.3 GHz CPU, with level 1 cache of 32 KB, level 2 cache of 256 KB, level 3 cache of 3 MB, and memory of 8 GB or 16 GB, Network is QSnet. The bandwidth of point-to-point communication is bigger than 300 MB/s, the latency is less

Table 1

Running times of eliminating before and after redundant computation on different platforms

Platform	Before	After
Pentium 4 E 2.8 GHz	1133.766 s	78.625 s
Itanium II 1.3 GHz	1328.989 s	146.037 s

than 7 microseconds. The operating system is Red hat Linux Advanced the Server 2.1 for IA-64, and the compiler is Intel FORTRAN Compiler 7.1 for IA-64. The optimization compiling option "-O3" is used.

3.1. Eliminating Redundant Computation

Table 1 shows running time before and after eliminating redundant computation on different platforms.

After the first step of optimization, on the Pentium 4 platform, the procedure running speed is 14.4 times faster than the procedure without optimization. On the Itanium II platform, the procedure running speed boosting is slightly lower, which is 9.1 times faster than without optimization. This indicates that the floating point performance of the Itanium II processor is extremely formidable. Therefore the speed boosting by the eliminating redundant floating point computing is small.

3.2. Eliminating Unnecessary Computation

Table 2 shows running time before and after eliminating unnecessary computation on different platforms.

After the second step of optimization, the procedure running speed is further enhanced, on the Pentium 4 platform; of the procedure running speed which is 2.64 times faster than the procedure of first step of optimization, is 38 times faster than the procedure without optimization, On the Itanium II platform, the procedure running speed which is 20.2 times faster than the procedure without optimization is 2.22 times faster than the procedure of the first step of optimization.

Table 2

Running times of before and after eliminating unnecessary computation on different platforms

Platform	Before	After
Pentium 4 E 2.8 GHz	78.625 s	29.828 s
Itanium II 1.3 GHz	146.037 s	65.908 s

Table 3

Running times of before and after improving the algorithm formula on different platforms

Platform	Before	After
Pentium 4 E 2.8 GHz	29.828 s	27.172 s
Itanium II 1.3 GHz	65.908 s	65.535 s

3.3. Improving Algorithm Formula

Table 3 shows running time before and after improvement of the algorithm formula on different platforms.

After the third step of the optimization, running speed of the procedure is enhanced slightly. On the Pentium 4 platforms, the running speed of the procedure is 41.7 times faster than the procedure without optimization. It is 9.77% faster than the previous step of using the approximate method procedure. On the Itanium II platform, the running speed is 20.3 times faster than the procedure without optimization, and it is 0.57% faster than the previous step of using the approximate method procedure. Because floating point operations are decreased similarly in this step, and decreased modestly, therefore, considering the floating point processing performance formidable Itanium II, the performance enhancement might be ignored.

3.4. Parallel Performance Testing

Performance testing of the parallel program is run on the DeepComp 6800 cluster. The running time on 1, 2, 4, 8, 16, 32 and 64 processors is tested. Four types of data distribution are used, including 1D block, 2D block, 1D cyclic and 2D cyclic. From Table 4, the performances using cyclic distributions are observably higher than those using block distribution.

Parallel speedup can be computed by the following formula:

Table 4

Running times of using 4 types of data distribution

Number of Processors	Running Time (sec)			
	1D block	2D block	1D cyclic	2D cyclic
1	66.960	66.960	66.960	66.960
2	38.764	38.764	34.612	34.612
4	22.119	20.504	18.368	18.438
8	12.665	12.264	10.373	10.286
16	7.804	7.743	6.266	6.258
32	5.143	5.410	4.400	4.436
64	3.885	4.501	3.834	3.637

Table 5

Speedups using 4 types of data distribution

Number of Processors	Speedup			
	1D block	2D block	1D cyclic	2D cyclic
1	1	1	1	1
2	1.727	1.727	1.935	1.935
4	3.027	3.266	3.645	3.632
8	5.287	5.460	6.455	6.510
16	8.580	8.648	10.686	10.700
32	13.020	12.377	15.218	15.095
64	17.236	14.877	17.465	18.411

$$\text{Speedup}(N) = \frac{T_1}{T_N}, \tag{6}$$

where T_1 is the running time on 1 processor, T_N is the running time on N processors. The speedups using 4 types of data distributions are listed in Table 5.

From this table, the speedups are much lower than the number of processors, especially when the number of processors is large. According to AMDAHL's law (1967), a program consists of 2 parts: the serial part and the parallel part. Assume in a program, the ratio of the serial part is α, the running time of the serial part is αT_1, the running time of the paralleled part is $(1-\alpha)\,\alpha T_1/N$, then the speedup formula can be expressed as:

$$Speedup(N) = \frac{T_1}{\alpha T_1 + \frac{(1-\alpha)T_1}{N}} = \frac{1}{\alpha + \frac{1-\alpha}{N}}. \tag{7}$$

From this formula we can see that the ratio of the serial part is larger, speedup is smaller, when the number of processors $N \to \infty$, the limitation of speedup is $1/\alpha$. In this program, the ratio of the serial part increases after optimization. According to the test on DeepComp 6800, the ratio of the measurable serial part is about 3%, thus the theoretical speedups are listed in Table 6.

The premise of these theoretical speedups is full load-balancing and omitting parallel overhead, such as synchronized and communication overhead. In Figure 4, speedup curves using 4 types of data distribution, ideal speedup and theoretical speedup are drawn.

From Figure 4, speedups using cyclic data distributions are close to the theoretical values. This shows that their load-balancing is better than block distributions; and the speedups using 1D cyclic distribution and 2D distributions have no significant difference,

Table 6

Theoretical speedups by Amdahl's law

Number of Processors	1	2	4	8	16	32	64
Theoretical speedup	1	1.942	3.670	6.612	11.034	16.580	22.145

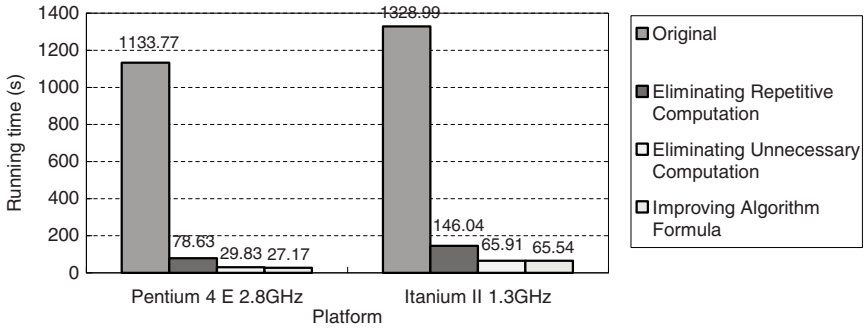


Figure 3
Running time of each step by the optimized procedure in two kinds of platforms.

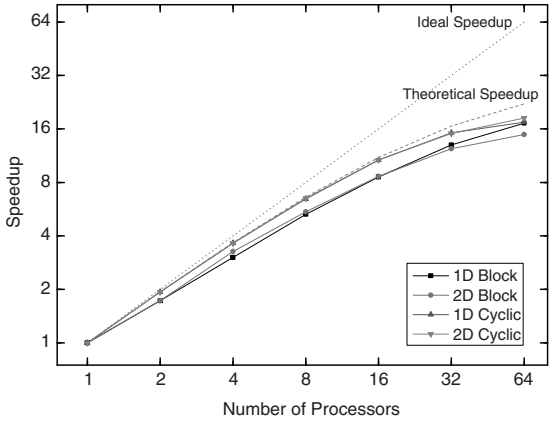


Figure 4
The testing result of parallel speedups.

when the number of processors is more than 32, 2D cyclic distribution is better. In 2 types of block distribution, when the number of processors is not more than 16, 2D block distribution's load-balancing is better than 1D block distribution's, but if the number of processors is more than 32, 1D block distribution's load-balancing is better.

4. Conclusion

After the parallel tempo-spatial scanning program is optimized, its running speed increases 20 to 40 times. With the optimization and parallel implementation of LURR, a wider range, longer time and higher accuracy earthquake prediction are possible. In

addition, this program requires less computation power, which makes it less dependent on high performance computers and decreases running cost.

Acknowledgements

This research was funded by the Ministry of Science and Technology of China (Grant No. 2005DKA64000-2), National Natural Science Foundation of China (Grant No. 60533020) and Research on high performance scientific computation (973 programs, Grant No. 2005CB321702)

References

Amdahl, G.M. Validity of the single-processor approach to achieving large scale computing capabilities. In *AFIPS Conf. Proc. 30* (Atlantic City, N.J., Apr. 18–20). (AFIPS Press, Reston, Va., 1967), pp. 483–485.

Mora, P., Place, D., Wang, Y., Yin, X., Peng, K., Weatherley, D., (2000), *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.

Yin, X.C., *A new approach to earthquake prediction*, Earth. Res. in China *3*, 1–7, (1987).

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., 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.

Zeyao Mo and Yuan Guoxing, *Parallel Programming Environment MPI (in Chinese)*. ISBN 7-03-009805-6 (Science Press 2001).

Zhang Yongxian, Yin, Xiangchu, and Peng Keyin, (2004), *Spatial and temporal variation of LURR and its implication for the tendency of earthquake occurrence in southern California*, *Pure Appl. Geophys 161*, 1–9.

Yin, Xiangchu *et al*. (1996) *The temporal variation of LURR in Kanto and other regions in Japan and its application to earthquake prediction*, *Earthq. Res. in China 10*(4), 381–385.

Zhuang, J.C. and Yin X.C. (1999), *Random distribution of the Load/Unload Response Ration(LURR) under Assumptions of Poisson Model*, Earthq. Res. in China *15*, 128–138. .

(Received March 21, 2007; revised March 21, 2007, accepted June 28, 2007)
Published Online First: April 2, 2008

Pure appl. geophys. 165 (2008) 761–775
0033–4553/08/030761–15
DOI 10.1007/s00024-008-0330-z

Pure and Applied Geophysics

Complexity in Sequences of Solar Flares and Earthquakes

VLADIMIR G. KOSSOBOKOV,[1] FABIO LEPRETI,[2,3] and VINCENZO CARBONE[2,3]

Abstract—In this paper the statistical properties of solar flares and earthquakes are compared by analyzing the energy distributions, the time series of energies and interevent times, and, above all, the distributions of interevent times per se. It is shown that the two phenomena have different statistics of scaling, and even the same phenomenon, when observed in different periods or at different locations, is characterized by different statistics that cannot be uniformly rescaled onto a single, universal curve. The results indicate apparent complexity of impulsive energy release processes, which neither follow a common behaviour nor could be attributed to a universal physical mechanism.

1. Introduction

Impulsive energy release events are observed in many natural systems. Solar flares (e.g., ASCHWANDEN, 2004 and refs. therein) and earthquakes (GUTENBERG and RICHTER, 1954) are certainly among the most remarkable examples of such processes. Solar flares are intense and sudden energy release events occurring in active regions of the solar atmosphere. The source of flares is the energy stored in highly stressed magnetic field configurations of the solar corona. During flares this energy is converted into accelerated particles, electromagnetic radiation emission, plasma motions, and heat. The amount of energy released in flares (reported to date) varies between roughly 10^{17} J and 10^{26} J. Contrary to solar flares, whose origin and behavior we can currently observe in detail, a coherent phenomenology on seismic events, which we evidence from their consequences, is lacking. Apparently, earthquakes occur through frictional sliding along the boundaries of highly stressed hierarchies of blocks of different sizes (from grains of rock about 10^{-3} m to tectonic plates up to 10^{7} m in linear dimension) that form the lithosphere of the Earth (KEILIS-BOROK, 1990). The amount of energy released in earthquakes varies from about 63 J in a detectable magnitude $M = -2$ event to more

[1] International Institute of Earthquake Prediction Theory and Mathematical Geophysics, Russian Academy of Sciences, 84/32 Profsoyuznaya, Moscow 117997, Russian Federation.

[2] Dipartimento di Fisica, Università della Calabria, Ponte P. Bucci 31/C, I-87036 Rende, CS, Italy.

[3] Consorzio Nazionale Interuniversitario per le Scienze Fisiche della Materia (CNISM), Unità di Cosenza, Ponte P. Bucci 31/C, I-87036 Rende, CS, Italy.

than 10^{18} J in a magnitude $M = 9$ mega-earthquake (e.g., the recent 26 December 2004 Sumatra-Andaman, $M_W = 9.3$ mega-thrust event is believed to release 4.3×10^{18} J).

Being recognized in the framework of plasma phenomena, the physics of solar flares is rather well established (e.g., ASCHWANDEN, 2004 and refs. therein), while understanding the physics of seismic events is more complicated (KEILIS-BOROK 1990; TURCOTTE, 1997; DAVIES, 1999; GABRIELOV *et al.*, 1999; KANAMORI and BRODSKY, 2001 and refs. therein) regardless of the longer history of instrumental observations. Some fundamental integral properties of seismic events are widely recognized in literature: (i) Following OMORI (1894) many seismologists claim, with slight modifications, that $n(t) \sim t^{-p}$, where t is the time span after the main shock, $n(t)$ is the rate of occurrence of aftershocks at time t, and p is a constant with a value of about 1; (ii) the number $N(M)$ of earthquakes with magnitude above a given threshold M in a given region behaves as $\log N(M) = a - bM$, where a and b are constants characterizing seismic activity and balance between earthquakes of different size in the region (GUTENBERG and RICHTER, 1954); (iii) earthquake epicenters follow fractal distribution in space (OKUBO and AKI, 1987; TURCOTTE, 1997). Although the Omori law remains a controversy (UTSU *et al.*, 1995), there is growing evidence that distributed seismic activity obeys the unified scaling law for earthquakes generalizing the Gutenberg-Richter relation as follows: $\log N(M,L) = A - B\,(M - 5) + C \log L$, where $N(M,L)$ is the expected annual number of earthquakes of magnitude M within a seismic locus of liner size L (KOSSOBOKOV and MAZHKENOV, 1988; KOSSOBOKOV and NEKRASOVA, 2005).

A statistical relation analogous to the Gutenberg-Richter law already has been found for solar flare energy flux (see e.g., LIN *et al.*, 1984; CROSBY *et al.*, 1993). Moreover, the interevent time Δt between solar flares has been shown to follow a power-law distribution $P(\Delta t) \sim \Delta t^{-\alpha}$ (BOFFETTA *et al.*, 1999) that suggests the Omori law analogue. Since the two statistical relations seem to be roughly shared by both processes, the existence of universality in the occurrence of these impulsive events was recently conjectured (DE ARCANGELIS et al., 2006).

In this paper, the statistical properties of solar flares and earthquakes are compared by analyzing the two sequences available from the Geostationary Operational Environmental Satellites (GOES) flare catalogue and the Southern California Seismic Network (SCSN) earthquake catalogue. The temporal evolution of the two processes and the probability distributions of event energies and interevent times are thoroughly investigated. Unlike many studies based on probability density functions or histograms (see e.g., BAK *et al.*, 2002; CORRAL, 2003; DE ARCANGELIS *et al.*, 2006), in order to reduce the ambiguities and misinterpretations related to the choice of the binning we use the cumulative probability distributions. This allows us to exploit the Kolmogorov-Smirnov nonparametric criterion, the K-S test, when comparing different distributions either of different phenomena, or in different periods of time, or at different locations. The paper is organized as follows: The considered catalogues of solar flares and earthquakes are briefly described in Section 2; the results of data analysis are presented in Section 3, and discussed in Section 4.

2. Data

The catalogue of soft X-ray (SXR) solar flares used in this work was compiled from observations of the Geostationary Operational Environmental Satellites (GOES) in the wavelength band 1–8 Å. The catalogue consists of 62,212 events from September 1, 1975 to May 30, 2006 and covers the three most recent solar cycles. The catalogue and the details about the event selection criteria can be found at the ftp site ftp://ftp.ngdc.noaa.gov/. The GOES catalogue is the most complete data base on solar flare bursts currently available. In order to classify the intensity of flares detected by GOES, an index named *GOES X-ray Class* is used, which corresponds to the peak burst intensity I_p measured in the wavelength band 1–8 Å. The classes are defined in the following way: B class if $I_p < 10^{-3}$ erg s^{-1} cm^{-2}, C class if 10^{-3} erg s^{-1} $cm^{-2} < I_p < 10^{-2}$ erg s^{-1} cm^{-2}, M class if 10^{-2} erg s^{-1} $cm^{-2} < I_p < 10^{-1}$ erg s^{-1} cm^{-2}, and X class se $I_p > 10^{-1}$ erg s^{-1} cm^{-2}. The class thus constitutes one of the previous letters followed by a number which allows to identification of the value of I_p, that is, for example a C4.6 class means that $I_p = 4.6 \times 10^{-3}$ erg s^{-1} cm^{-2}. Beginning from January 1997, the integrated flux from event start to end is also provided in the catalogue. Figure 1 shows the integrated flux as a function of the peak burst intensity. A reasonably good proportionality between the two quantities is observed, even if deviations from proportionality can be noted, especially for low and intermediate energies. We can thus consider the X-ray class as a good proxy for energy flux.

In the main part of our analysis, we considered the 32,076 flares of class C2 or larger. This is because there is evidence for incompleteness of the catalogue for events of class smaller than C2, as is explained in the next Section. In order to perform a separation between the three cycles covered by the data set, we used the following simple criterion: the boundary between two successive cycles corresponds to the longest interevent times occurring between two C2+ flares. With this definition the three cycles cover the following time intervals respectively: The first one (solar cycle 21) from September 1, 1975 to September 8, 1986, the second one (solar cycle 22) from September 8, 1986 to August 22, 1996, and the third one (solar cycle 23) from August 22, 1996 until May 30, 2006.

The Southern California Earthquake Data Center (SCEDC) regularly compiles the Southern California Seismographic Network (SCSN) catalogue of hypocentral information. The catalogue started in 1932 and is being updated continuously. The data include more than 450,000 local and regional events with epicenters mostly (i.e., more than 99% of the total) between latitudes 32 and 37°N and longitudes between 114 and 122°W. For information on catalogue completeness and data sources, see http://www.data.scec.org/about/data_avail.html. In our study we consider catalogue completeness for the entire region of Southern California, i.e., so-called magnitude completeness of the region. This term should be distinguished from the so-called coverage completeness due to the specifics of spatial distribution of a seismographic network. The magnitude completeness is evidently different in relatively small areas at different locations, e.g., off-shore

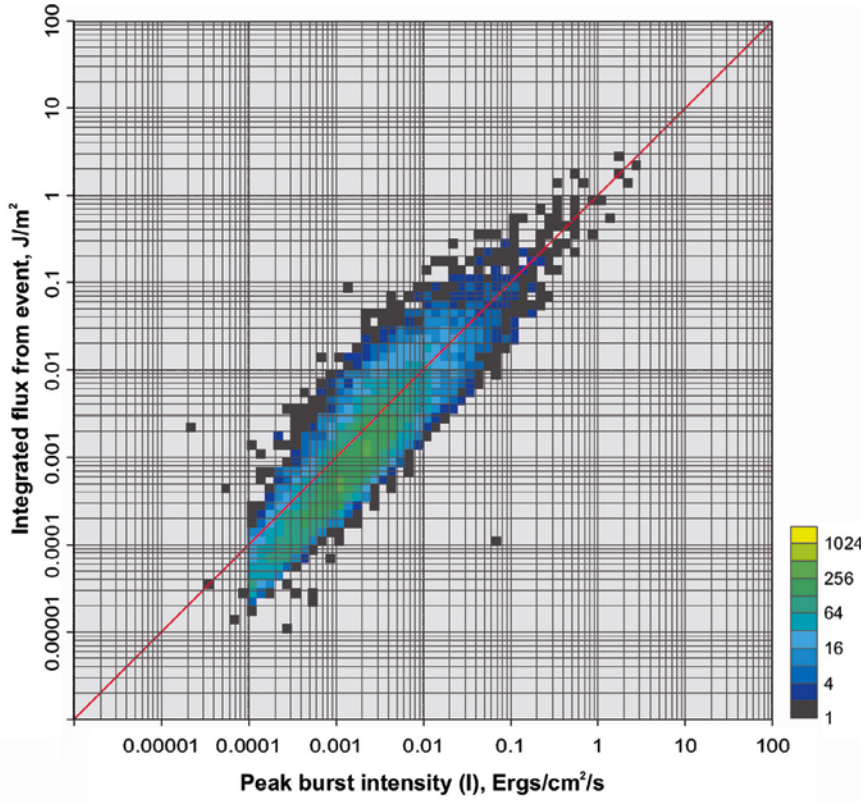


Figure 1
Integrated flux from event start to end for GOES flares (January 1997–May 2006) as a function of peak burst intensity.

Southern California or in the area of the Parkfield experiment. Moreover, in a small area the catalogue availability might be too short in time for a decent evaluation of the magnitude completeness. That is not the case of Southern California as a whole where the SCSN catalogue is believed to be reasonably complete for seismic events greater than or equal to M 3.0 since 1932, above M 1.8 since 1981, and for all magnitude events between the time periods of January 1984 through the present. In the present study we use the SCSN catalogue of earthquakes for the period 1986–2005, which contains 356,545 events. For the same reasons of completeness mentioned above for solar flares and illustrated in the next Section, we consider the 87,688 earthquakes with magnitude $M \geq 2$ in the main part of the analysis. We also consider the aftershock series of the six strong earthquakes in Southern California. The selection of the spatial and temporal span of the aftershock series has been made by inspection of the epicenter maps in dynamics of

seismic occurrences, so that the individually selected areas outline the evident clusters of epicenters while the individually selected durations include the times of evident activation of seismic activity next to the epicenter of the main shock. Such identification of the individual clusters of seismic events is similar to identification of the flare productive magnetic activity regions, which reference numbers reported in the GOES catalogue of solar flares link flares occurring in the same cluster of activity.

3. Data Analysis and Results

3.1. Gutenberg-Richter Distributions

As was mentioned in the Introduction, both earthquake magnitudes and solar flare energy fluxes are expected to be distributed in good agreement with the Gutenberg-Richter law. Figure 2 shows the cumulative histograms of the peak intensity of the GOES flares (upper plate) and of the magnitude of the SCSN earthquakes (lower plate). Along with the total histograms, the ones for the three solar cycles and for each of the twenty years in the SCSN catalogue are also given. (To facilitate inspection, straight lines with $b = 1$ are plotted on both plates.)

As mentioned in the Introduction, the low energy cut-offs where the histograms start deviating from the power-law Gutenberg-Richter behavior at low energies occur for flares below GOES C2 class and for seismic events below magnitude 2.0. These deviations indicate incompleteness of the catalogues in these regions and justify the thresholds chosen to select the events used in the rest of our analysis. Some small bumps are clearly visible in the flare distribution for the first cycle (for example for C2, M2 and X2 flares) due to the poor resolution of event size in the initial years, as will be clearer from the discussion of the next figure. It can also be seen that the earthquake distributions of different years show consistent differences between them (e.g., the 1992 distribution) due to the effect of the strongest events, as will be explained below.

3.2. Temporal Behavior

In order to have a first synoptic illustration of the two phenomena, we report in Figure 3 the time series of the interevent times between successive events for both catalogues, the time series of the peak GOES SXR intensities of solar flares and that of the SCSN seismic event magnitudes.

The most prominent feature observable for solar flares is the effect of the activity cycle resulting in an alternation of maximum periods, where the flaring activity is enhanced, and minimum periods, where activity assumes a much more sporadic character. It can also be seen that until roughly the end of 1980 the GOES classes (i.e., the peak intensities) were given with the only one significant digit. This lack of resolution is at the origin of the small bumps found in the cumulative peak SXR intensity histogram (previously shown) for the

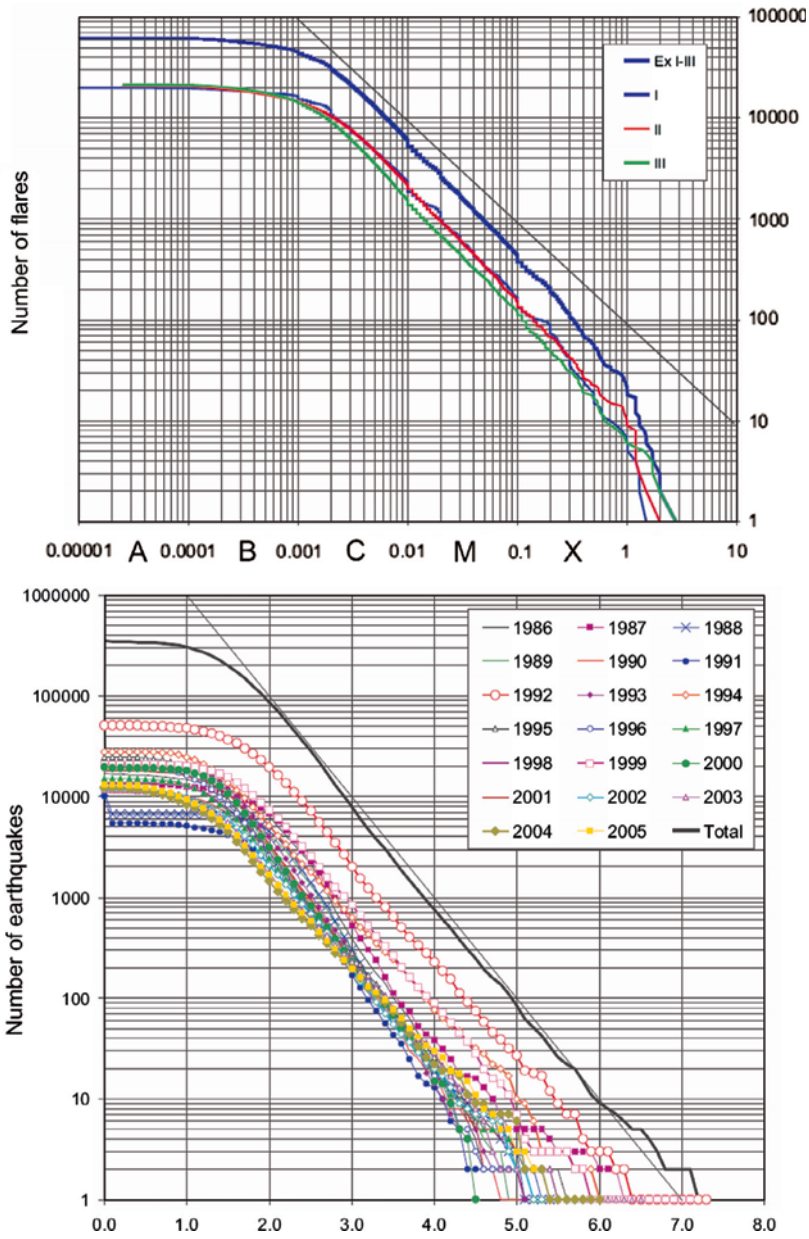


Figure 2
Cumulative histograms of the GOES peak intensity of flares and of the SCSN magnitude of seismic events.

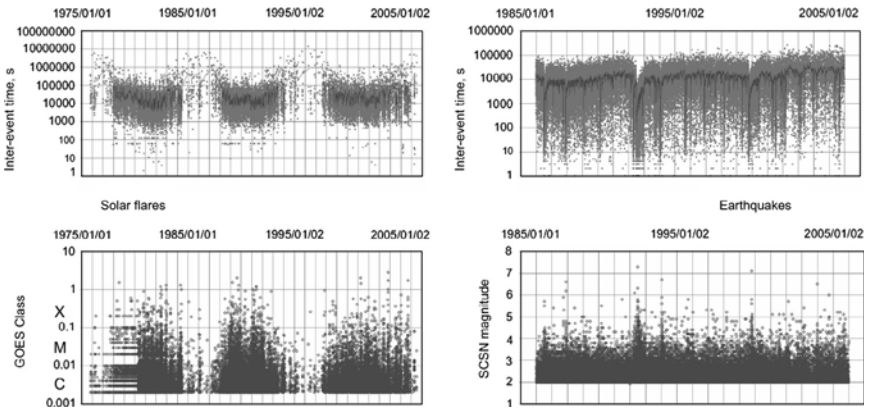


Figure 3
Time series of the interevent times between successive events for the GOES flares (upper-left) and for the SCSN seismic events (upper-right). The solid lines represent moving averages over 50 events. Time series of the peak GOES SXR intensity of flares (lower-left) and of the SCSN magnitude of seismic events (lower-right). Only flares with GOES class above C2 and earthquakes with magnitude $M \geq 2$ were considered.

solar cycle 21. For the seismic events, the time evolution is characterized by lengthy periods of stable, slowly varying activity interrupted by the occurrence of strong activity peaks corresponding to strong earthquakes with magnitude $M > 6$.

Another way to get an overall view of the two processes is to look at the frequencies of the GOES class and SCSN magnitude respectively as functions of time. This is done here by calculating the semi-annual number of flares in different GOES class intervals and the annual number of earthquakes in different magnitude intervals (Fig. 4).

A number of evident features can be recognized from the solar flare plot. In the central part of the cycles, around the solar maximum periods, intervals of roughly steady activity are observed. In these intervals flares below C1 class are mostly missing., due to the rise of the solar soft X-ray background associated with the maximum of magnetic activity. It can also be seen that the time evolution of flaring activity is quite different in the three cycles. For example, in the first cycle a quite steady and slow activity enhancement is followed by a much faster decay, while in the second cycle two successive enhancements and decreases can clearly be observed. The situation is even more complex for the third cycle, where three or four distinct peaks of activity seem to be present. Other remarkable features can be seen in the intervals between the cycles, which should be characterized by low magnetic activity. Between the first and the second cycle some clear activity bumps occur, while between the second and third cycle such enhancements are absent. The earthquake magnitude-frequency versus time shown in the lower panel of Figure 4 confirms once again that seismic events reported by SCSN display a near stationary background activity interrupted by the occurrences of strong earthquakes.

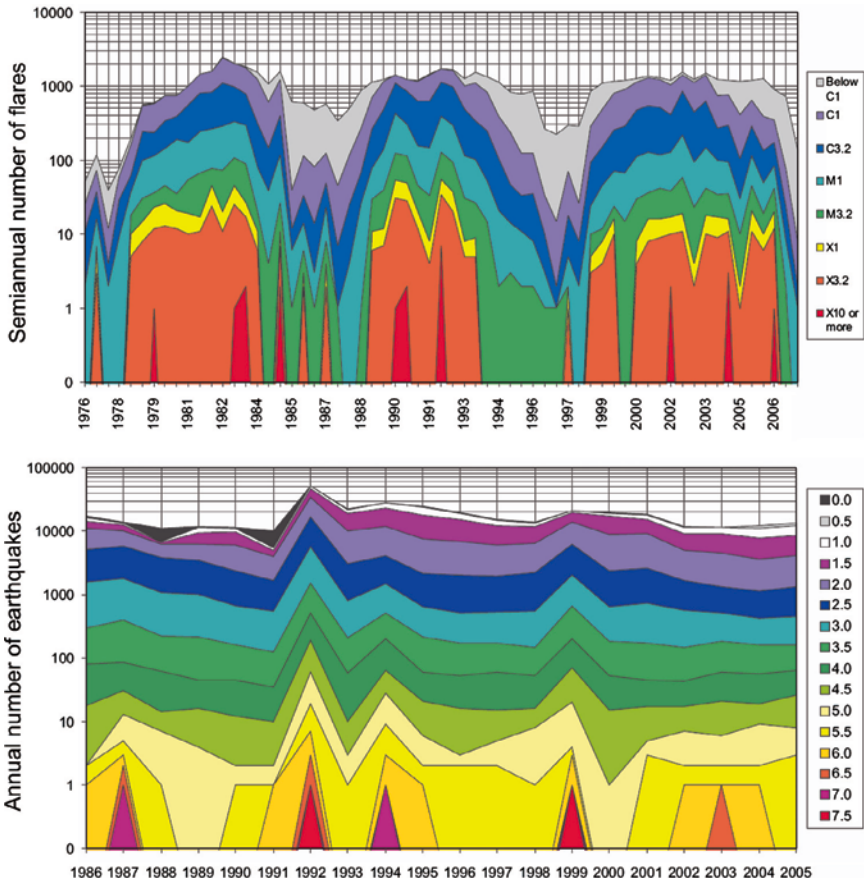


Figure 4

Semi-annual number of flares in different GOES class intervals as a function of time (top panel) and annual number of SCSN seismic events in different magnitude intervals as a function of time (bottom panel).

Figure 5 shows the accumulated number of events as a function of time both for flares and earthquakes. The accumulated peak X-ray intensity for solar flares and the so-called accumulated Benioff strain release for seismic events are also reported. The cumulative Benioff strain release is an integral measure of fracturing defined by the sum of the square root of the energy for consecutive events (see, e.g., VARNES, 1989; BUFE *et al.*, 1994).

The time series of the accumulated number of flares shows a rather smooth evolution, once again dominated by the solar cycle, with intervals of increase corresponding to the activity maxima and nearly flat intervals corresponding to the solar minima. A similar behavior is found in the accumulated peak X-ray intensity, although some differences in timing can be observed, for example in the last part of the third cycle where the

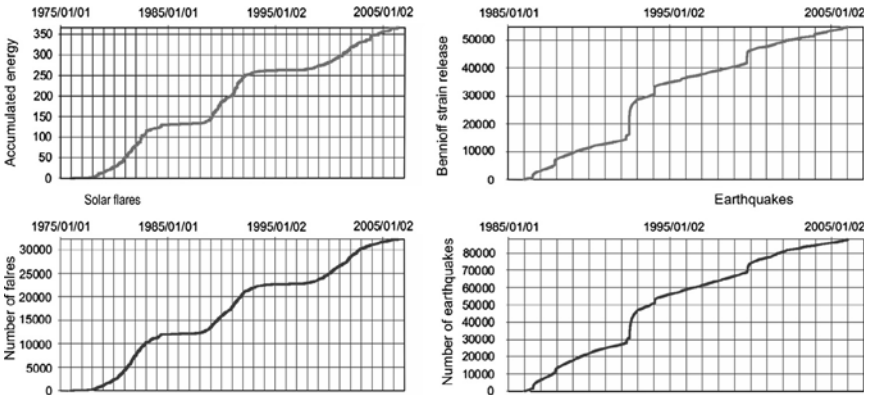


Figure 5
Accumulated number of events as a function of time for solar flares (lower-left) and earthquakes (lower-right). The accumulated peak X-ray intensity for solar flares (upper-left) and the accumulated Benioff strain release for seismic events (upper-right) are also shown.

accumulated number has become almost flat while the accumulated intensity is continuously increasing. On the other hand, the two plots for seismic events both display intervals of steady stationary increases interrupted by sudden jumps produced by strong earthquakes and their aftershocks.

3.3. Inter-Event Time Distributions

We consider now the statistics of interevent times, that is, the time intervals separating two consecutive events. In the upper panels of Figure 6 we report the cumulative distribution of interevent times for GOES flares, both in semi-logarithmic and bi-logarithmic scale. The cumulative distributions for the three different cycles are also shown. In the lower panels of Figure 6 the interevent cumulative distributions of SCSN seismic events are shown for the whole considered period and for each year separately. Although the total flare interevent distributions and those for the three cycles show quite similar power-law decay for times longer than 10^4 s, some differences can be seen even in the power-law domain.

The differences between the distributions for seismic events are significant beyond any doubt. It can be noted that the distributions of 1992 and 1999, which show the most evident departure from the other ones, are dominated by the effect of the 1992 Landers and 1999 Hector Mine earthquakes and their sequences of associated fore- and aftershocks.

In order to study the differences of flaring and seismic activity between different activity regions, we investigated the interevent time distributions of solar flares produced in separate flare productive magnetic activity regions and those of the aftershocks following the strongest earthquakes reported in the SCSN catalogue. The results are plotted in Figure 7. It can be seen that different activity spots produce quite different

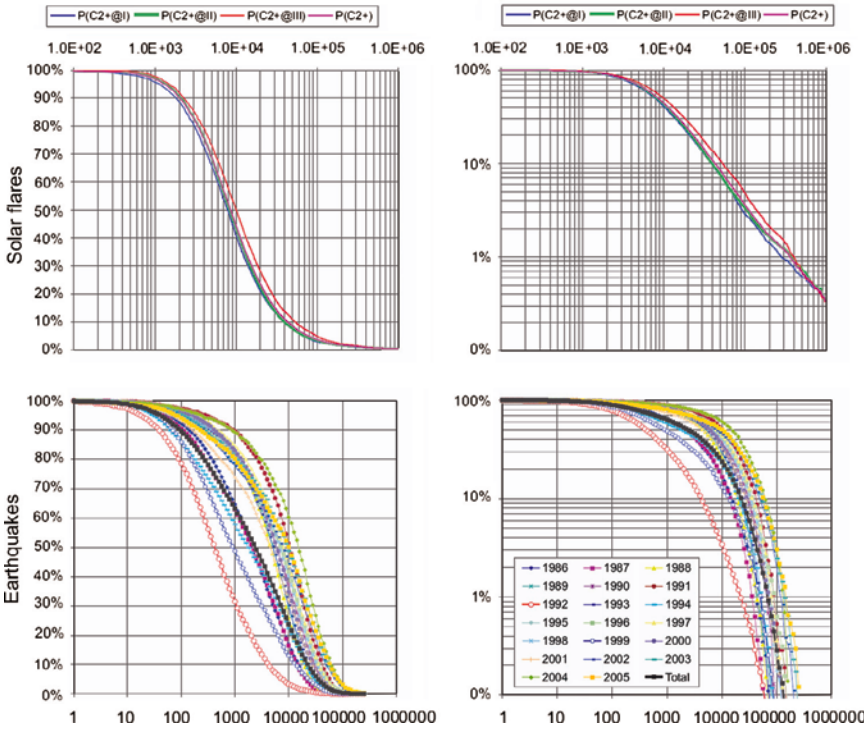


Figure 6

Cumulative distribution of interevent times for GOES flares (upper panels) and SCSN earthquakes (lower panels), both in semi-logarithmic (left panels) and bi-logarithmic scale (right panels). The four distributions shown for GOES flares refer to the whole period 1975–2006 (violet curve), to the first cycle 1975–1986 (blue curve), to the second cycle 1986–1996 (green curve), and to the third cycle 1996–2006 (red curve) respectively. The distributions of SCSN earthquakes are shown for the whole period 1986–2005 (black curve) and for each year separately.

interevent time distributions; the differences being stronger for seismic events than for solar flares. Note that the 1992 Joshua Tree and Landers earthquakes, which occurred at close locations and have shown up very similar distributions, are well separated in time.

To compare the statistical properties of interevent times of different impulsive energy release processes, we report (Fig. 8) the cumulative interevent time distribution of all GOES flares, GOES flares occurring in the same activity regions, all SCSN earthquakes, and aftershocks of the Landers earthquake along with the cumulative interevent time distribution of the unique long series of soft γ-ray flashes on the neutron star 1806–20 (the numbers are its celestial coordinates).

It is believed that flashes of energy radiated by a neutron star in the form of the soft γ-ray repeaters, SGRs, are probably generated by "starquakes" analogous to earthquakes

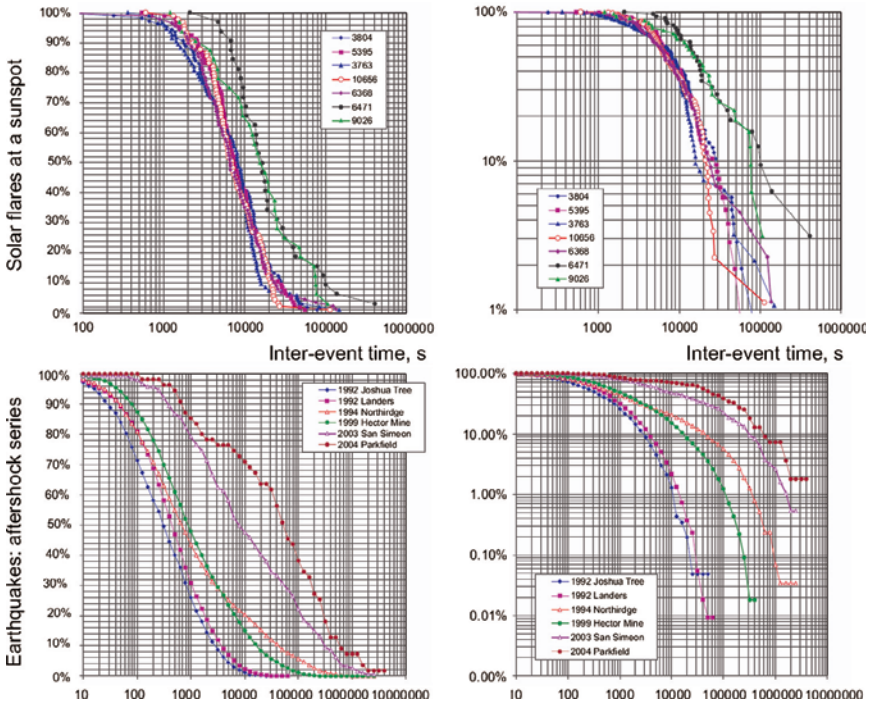


Figure 7
Cumulative distributions of interevent times for GOES flares occurring in some flare productive active regions (upper panels) and for aftershocks of some strong seismic events recorded in the SCSN catalogue (lower panels). The distributions are shown both in semi-logarithmic scale (left panels) and bi-logarithmic scale (right panels). The active zone numbers and main shock names are given in the legends.

(Norris *et al.*, 1991; Thompson and Duncan, 1995). The source of a starquake is a fracture in the neutron star crust, which may build up strain energy up to 10^{46} erg. The star 1806–20 is a dense ball about 20 km in diameter composed of neutrons. It weighs roughly as much as the Sun, has the period of rotation about 7.5 s and the magnetic field of 10^{15} Gauss. Its crust, made of a solid lattice of heavy nuclei with electrons flowing between, is 1 km thick. The crust of a neutron star is loaded by magnetic forces as the field drifts through. These forces cause fracturing of the crust and associated flashes of energy release. Starquakes are of special interest due to the extreme energies released in a single event. The 111 flashes on 1806–20 recorded during continuous observation from August 1978 to April 1985 follow power-law energy distribution and other earthquake-like statistics and behaviour (Kossobokov *et al.*, 2000).

The distributions presented in Figure 8 are very different. In order to check whether they can be rescaled onto a single curve, we used the Kolmogoroff-Smirnoff two-sample criterion. This test has the advantage of making no assumptions about the distribution of

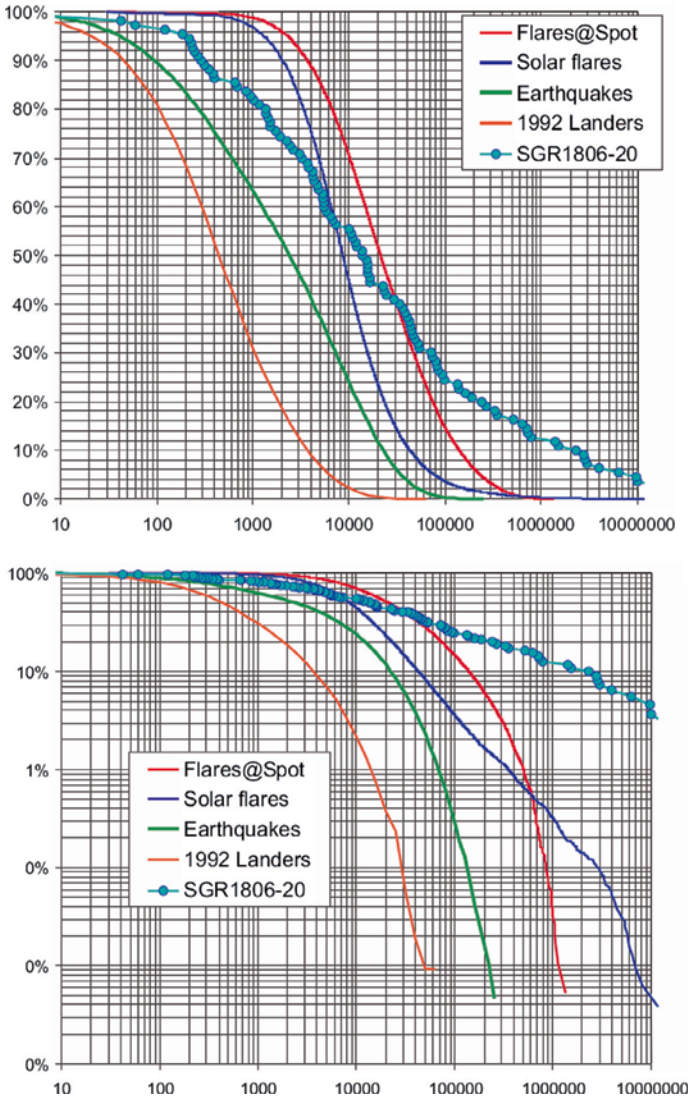


Figure 8

Cumulative interevent time distributions of all GOES flares (blue curve), GOES flares occurring in the same active regions (red curve), all SCSN earthquakes (green curve), and aftershocks of the Landers earthquake (orange curve). The distributions are shown both in semi-logarithmic scale (upper panel) and in bi-logarithmic scale (lower panel). The cumulative interevent time distribution of soft γ-ray flashes of the SGR 1806–20, attributed to starquakes produced by fractures in the neutron star crust, is also shown (light blue circles).

Table 1

Results of the K-S test on the interevent time distributions shown in Figure 8. The sample sizes are reported on the diagonal, the minimum values of $\lambda_{K\text{-}S}$ over the rescaling fits described in the text are given above the diagonal and the corresponding probabilities to reject the hypothesis of being drawn from the same distribution below the diagonal

	Flares	Flares at spot	SCSN	Landers	SGR1806–20
Flares	**32076**	3.435	8.648	2.071	0.636
Flares at spot	100.00%	**18878**	5.898	1.669	0.434
SCSN	100.00%	100.00%	**87688**	3.726	1.435
Landers	99.96%	99.26%	100.00%	**10706**	0.47
SGR1806–20	19.13%	0.92%	96.77%	2.24%	**110**

data. Moreover, it is widely accepted to be one of the most useful and general nonparametric methods for comparing two samples, as it is sensitive to differences in both location and shape of the empirical cumulative distribution functions of the two samples. The two sample Kolmogoroff-Smirnoff statistic $\lambda_{K\text{-}S}$ is defined as $\lambda_{K\text{-}S}(D,n,m) = [nm/(n + m)]^{1/2}D$, where $D = \max |P_{1,n}(\Delta t) - P_{2,m}(\Delta t)|$ is the maximum value of the absolute difference between the cumulative distributions $P_{1,n}(\Delta t)$ and $P_{2,m}(\Delta t)$ of the two samples, whose sizes are n and m, respectively. We calculated the minimum values of $\lambda_{K\text{-}S}$, for all the couples of distributions over all rescaling fits of the type $P'(\Delta t) = P(C\ \Delta t^{\alpha})$, where C and α are fitting constants. These minimum values are collected in Table 1 (above the diagonal) together with the corresponding probabilities to reject the hypothesis of being drawn from the same distribution (under the diagonal). The results indicate that only the associations of the "starquake" distribution, which is by far the sample of the smallest size, with the interevent time distribution either of all flares, or activity spot flares, or the 1992 Landers earthquake sequence cannot be rejected. For all the remaining cases, we can conclude with high confidence that the distributions cannot be rescaled onto the same curve. For example, the K-S test rejects a possibility that the interevent times between solar flares and those from a single active region are drawn from the same distribution at the confidence level starting with nine nines (i.e., 99.9999999%). In fact, the minimum of the maximal difference D between the cumulative distributions reaches 3.15%, when $C = 0.90$ and $\alpha = 1.112$, and due to the large sample sizes n and m implies $\lambda_{K\text{-}S}(D, n, m) = 3.435$ and the above-mentioned level of confidence.

4. Discussion

Our comparison of the statistical properties of solar flares and seismic events shows that, beside the expected global differences arising from the fact that solar flares are related to the periodic solar activity cycle, the two phenomena are characterized by different statistics of scaling. In fact, even the same phenomenon observed in different

periods or locations (i.e., different solar active regions or aftershocks of different earthquakes) can produce different probability distributions. Our analysis is mainly based on the cumulative distributions of interevent times. Analyzing separately the interevent time distributions of certain flare productive active regions, we have found that different active regions give rise to substantially different statistics of the interevent times. Remarkable differences are observed also in the interevent time statistics of aftershocks of some strong ($M > 6$) seismic events present in the SCSN catalogue.

After constructing a data set with the interevent times between consecutive flares occurring in the same active region, we have found that the interevent time distribution of these flares displays a very different shape with respect to the distribution obtained taking into account all solar flares. A clear power-law scaling ranging from 10^4 to 3×10^6 s is found in the distribution for all flares, while the distribution for flares in individually active regions has a more complex shape, which cannot be easily identified with a simple law. The physical origin of this difference is an interesting question to be addressed in the future. Similar complexity is observed in the interevent time statistics of the whole SCSN seismic event catalogue as well as of the aftershock sequences of individually strong earthquakes (e.g., the 1992 Landers event).

All these interevent time distributions have been compared to each other and also to the unique distribution of the interevent time between starquakes produced by fracturing of the neutron star crust. The main finding is that these distributions display a wide variety of shapes and cannot be uniformly rescaled onto a single, universal curve by natural two-parametric transforms of the interevent times. The small sample of starquakes, which are of special interest due to the extreme energies released in a single event and the uniqueness of the SGR1806–20 series, may appear similar to the others (the similarity appears to be rejected just in the only case of the SCSN sample, the largest among those considered) simply due to its size, which may not be enough for a confident claim of specificity. Evidently, our results do not support the presence of universality and the existence of a common physical mechanism at the origin of the energy release phenomena considered. The statistical analogies between these phenomena should be attributed rather to more general, integral properties of impulsive processes in critical, nonlinear systems.

References

Aschwanden, M., *Physics of the Solar Corona* (Springer-Verlag, Berlin 2004).

Bak, P., Christensen, K., Danon, L., and Scanlon, T. (2002), *Unified scaling law for earthquakes*, Phys. Rev. Lett. *88*, 178501.

Boffetta, G., Carbone, V., Giuliani, P., Veltri, P., and Vulpiani, A. (1999), *Power laws in solar flares: Self-organized criticality or turbulence?* Phys. Rev. Lett. *83*, 4662–4665.

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.

Corral, A. (2003), *Local distributions and rate fluctuations in a unified scaling law for earthquakes*, Phys. Rev. E *68*, 035102.

CROSBY, N.B., ASCHWANDEN, M.J., and DENNIS, B.R. (1993), *Frequency distributions and correlations of solar X-ray flare parameters*, Solar Phys. *143*, 275–299.

DAVIES, G.F., *Dynamic Earth: Plates, Plumes and Mantle Convection*, (Cambridge University Press, Cambridge 1999).

DE ARCANGELIS, L., GODANO, C., LIPPIELLO, E., and NICODEMI, M. (2006), *Universality in solar flare and earthquake occurrence*, Phys. Rev. Lett. *96*, 051102.

GABRIELOV, A., NEWMAN, W.I., and TURCOTTE, D.L. (1999), *An exactly soluble hierarchical clustering model: Inverse cascades, self-similarity, and scaling*, Phys. Rev. E *60*, 5293–5300.

GUTENBERG, B., and RICHTER, C.F., *Seismicity of the Earth*, 2nd ed. (Princeton University Press, Princeton, N.J. 1954).

KANAMORI, H. and BRODSKY, E.H. (2001), *The Physics of Earthquakes*, Physics Today, June 2001, 34–40.

KEILIS-BOROK, V.I. (1990), The lithosphere of the Earth as a nonlinear system with implications for earthquake prediction, Rev. Geophys. 28 (1), 19–34.

KOSSOBOKOV, V.G. and MAZHKENOV, S.A., *Spatial characteristics of similarity for earthquake sequences: Fractality of seismicity*, Lecture Notes of the Workshop on *Global Geophysical Informatics with Applications to Research in Earthquake Prediction and Reduction of Seismic Risk* (15 Nov.–16 Dec., 1988), 15 pp., (ICTP, Trieste 1988).

KOSSOBOKOV, V.G., KEILIS-BOROK, V.I., and CHENG, B. (2000), *Similarities of multiple fracturing on a neutron star and on the Earth*, Phys. Rev. E *61*, 3529–3533.

KOSSOBOKOV, V.G. and NEKRASOVA, A.K. (2005), *Temporal variations in the parameters of the Unified Scaling Law for Earthquakes in the eastern part of Honshu Island (Japan)*, Doklady Earth Sciences 405, 1352–1356.

LIN, R.P., SCHWARTZ, R.A., KANE, S.R., PELLING, R.M., and HURLEY, K.C. (1984), *Solar hard X-ray microflares*, Astrophys. J. *283*, 421–425.

NORRIS, J.P., HERZ, P., WOOD, K.S., and KOUVELIOTOU, C. (1991), *On the nature of soft gamma repeaters*, Astrophys. J. *366*, 240–252.

OKUBO, P.G., and AKI, K. (1987), *Fractal geometry in the San Andreas Fault system*, J. Geophys. Res. *92* (B1), 345–356.

OMORI, F. (1894), *On the after-shocks of earthquakes*, J. Coll. Sci. Imp. Univ. Tokyo *7*, 111–200.

THOMPSON, C., and DUNCAN, R. C. (1995), *The soft gamma repeaters as very strongly magnetized neutron stars – I. Radiative mechanism for outbursts*, Mon. Not. R. Astr. Soc. *275*, 255–300.

TURCOTTE, D.L., *Fractals and Chaos in Geology and Geophysics*, 2nd edition, (Cambridge University Press, Cambridge 1997).

UTSU, T., OGATA, Y., and MATSU'URA, R.S. (1995), *The centenary of the Omori formula for a decay law of aftershock activity*, J. Phys. Earth *43*, 1–33.

VARNES, D.J. (1989), *Predicting earthquakes by analyzing accelerating precursory seismic activity*, Pure Appl. Geophys. *130*, 661–686.

(Received December 22, 2006, revised March 30, 2007, accepted March 30, 2007)
Published Online First: April 2, 2008

Pure appl. geophys. 165 (2008) 777–795
0033–4553/08/030777–19
DOI 10.1007/s00024-008-0331-y

Pure and Applied Geophysics

Earthquakes: Recurrence and Interoccurrence Times

S.G. Abaimov,[1] D.L. Turcotte,[1] R. Shcherbakov,[5] J.B. Rundle,[2] G. Yakovlev,[2] C. Goltz,[1,3] and W.I. Newman[4]

Abstract—The purpose of this paper is to discuss the statistical distributions of recurrence times of earthquakes. Recurrence times are the time intervals between successive earthquakes at a specified location on a specified fault. Although a number of statistical distributions have been proposed for recurrence times, we argue in favor of the Weibull distribution. The Weibull distribution is the only distribution that has a scale-invariant hazard function. We consider three sets of characteristic earthquakes on the San Andreas fault: (1) The Parkfield earthquakes, (2) the sequence of earthquakes identified by paleoseismic studies at the Wrightwood site, and (3) an example of a sequence of micro-repeating earthquakes at a site near San Juan Bautista. In each case we make a comparison with the applicable Weibull distribution. The number of earthquakes in each of these sequences is too small to make definitive conclusions. To overcome this difficulty we consider a sequence of earthquakes obtained from a one million year "Virtual California" simulation of San Andreas earthquakes. Very good agreement with a Weibull distribution is found. We also obtain recurrence statistics for two other model studies. The first is a modified forest-fire model and the second is a slider-block model. In both cases good agreements with Weibull distributions are obtained. Our conclusion is that the Weibull distribution is the preferred distribution for estimating the risk of future earthquakes on the San Andreas fault and elsewhere.

Key words: Earthquakes, interoccurrence and recurrence statistics, Weibull distribution.

1. Introduction

The statistical distribution of interval times between earthquakes on a fault plays an important role in probabilistic hazard assessments. There are clearly two limiting cases that can be considered. The first is that earthquakes occur periodically. In this case the interval times between great earthquakes on segments of the San Andreas fault would

[1] Department of Geology, University of California, Davis, CA 95616, U.S.A.
E-mail: abaimov@geology.ucdavis.edu; turcotte@geology.ucdavis.edu; cgoltz@ucdavis.edu
[2] Center for Computational Science and Engineering, University of California, Davis, CA 95616, U.S.A.
Email: jbrundle@ucdavis.edu; gleb@cse.ucdavis.edu
[3] Institute for Geosciences, Department of Geophysics, Kiel University, 24118 Kiel, Germany.
E-mail: goltz@geophysik.uni-kiel.de
[4] Departments of Earth and Space Sciences, Physics and Astronomy, and Mathematics, University of California, Los Angeles, CA 90095, U.S.A. E-mail: win@ucla.edu
[5] Department of Earth Sciences, University of Western Ontario, London, Ontario, Canada, N6A 5B7.
E-mail: rshcherb@uwo.ca

be constant. This is clearly not the case. The second limiting case is that earthquakes occur randomly. In this case the distribution of interval times is exponential (Poissonian). This may be a reasonable approximation for the global occurrence of great earthquakes, but the exponential distribution of interval times is not generally valid. Aftershocks are an extreme example of such a deviation.

Before any further discussion, it is necessary to make a clear distinction between interoccurrence times and recurrence times. Interoccurrence times are the time intervals between earthquakes on all faults in a region with magnitudes greater than a specified value. Recurrence times are the time intervals between successive earthquakes on a single fault or fault segment. If these earthquakes occur on a major (plate boundary) fault they are generally referred to as characteristic earthquakes.

We first discuss interoccurrence times. All earthquakes in a specified region and specified time window with magnitudes greater than a specified magnitude are considered to be point events. Based on studies of all earthquakes in Southern California during a prescribed time interval, BAK *et al*. (2002) obtained a universal scaling for the statistical distribution of interoccurrence times. Subsequently, other studies of this type have been carried out (CARBONE *et al*., 2005; CORRAL 2003, 2004a,b, 2005a,b 2007; DAVIDSEN and GOLTZ 2004; LINDMAN *et al*., 2005; LIVINA *et al*., 2005a,b). SHCHERBAKOV *et al*. (2005) showed that this observed behavior for aftershocks can be explained by a nonhomogeneous Poisson process. The earthquakes occur randomly but the rate of occurrence is determined by Omori's law for the temporal decay of aftershock activity. Regional compilations of interoccurrence time statistics such as those given by BAK *et al*. (2002) are clearly dominated by the intervals between aftershocks.

If a point on a fault experiences a sequence of earthquakes, we refer to the intervals between these earthquakes as recurrence times. Of particular interest are the recurrence times on major faults. These faults experience the quasi-periodic occurrence of large earthquakes. They are usually referred to as characteristic earthquakes. Available evidence is that there is considerable variability in both the recurrence times and in the magnitudes of characteristic earthquakes. This variability can be attributed to the interactions between faults and fault segments. The statistical distribution of recurrence times is an important input into probabilistic seismic hazard assessments.

Various distributions have been used to fit observed earthquake recurrence time statistics. OGATA (1999) and UTSU (1984) have considered in some detail the Weibull, lognormal, gamma, exponential, and doubly exponential distributions. Many authors have applied the Weibull distribution to distributions of recurrence times between earthquakes. One of the earliest was HAGIWARA (1974). RIKITAKE (1976, 1982) applied the Weibull distribution to the observed recurrence times of great earthquakes at six subduction zones. RIKITAKE (1991) fit the Weibull distribution to great earthquakes in the Tokyo area. The lognormal distribution has also been used to correlate distributions of earthquake recurrence times. Arguments in its favor were presented by NISHENKO and BULAND (1987). It was also used in three formal assessments of future earthquake probabilities in California (JACKSON *et al*., 1995; WORKING GROUP ON CALIFORNIA

EARTHQUAKE PROBABILITIES, 1988, 1990). Another statistical distribution that has been used to correlate distributions of recurrence times is the Brownian passage-time distribution (MATTHEWS *et al*., 2002). This distribution was used in the most recent assessment of the earthquake hazard in the San Francisco area (WORKING GROUP ON CALIFORNIA EARTHQUAKE PROBABILITIES, 2003).

DAVIS *et al*. (1989), SORNETTE and KNOPOFF (1997), and CORAL (2005b) have also considered the paradox that for many statistical distributions that have been proposed for recurrence times, the mean waiting time for the next earthquake increases with an increase in the time since the last earthquake. They point out that this increase is associated with a thick tail of the distribution. The transition is the exponential distribution associated with a Poissonian distribution of time intervals between earthquakes on a fault. The Poissonian distribution has no memory and the mean waiting time to the next event does not depend on the time interval since the last event. Distributions that have "fatter" tails than the exponential distribution have an increase in mean waiting times with an increase in the time since the last event; an example is the lognormal distribution. Distributions that have "thinner" tails than the exponential have a decrease in mean waiting times with an increase in the time since the last event; an example is the Weibull distribution with an exponent greater than 1.

In statistics it is standard practice to test a variety of distributions and select the distribution that best fits the data using one of several error analyses. In statistical physics several distributions have been shown to have universal applicability. One example is the power-law (fractal) distribution (MANDELBROT, 1982; TURCOTTE, 1997). A second example is the Weibull distribution (ALTMANN *et al*., 2004; ALTMANN and KANTZ, 2005; BUNDE *et al*., 2003, 2004); in the physics literature it is known as the stretched exponential distribution if $\beta < 1$.

The purpose of this paper is to show that the Weibull distribution is the appropriate distribution for the recurrence times of characteristic earthquakes. We first introduce the Weibull distribution and a basis for its universality. We then show correlations of the distribution with several examples of characteristic earthquakes on the San Andreas fault. We also compare the distribution with the distribution of recurrence times for simulated characteristic earthquakes on the northern San Andreas fault. Finally we show the applicability of the distribution to several models that can be associated with characteristic earthquakes.

2. *Weibull Distribution*

The Weibull distribution is often used for the distribution of life times in engineering applications (MEEKER and ESCOBAR, 1991; WEIBULL, 1951). Examples include composite materials and fiber bundles. The probability distribution function (pdf) for a Weibull distribution is given by (PATEL *et al*., 1976)

$$p(t) = \frac{\beta}{\tau}\left(\frac{t}{\tau}\right)^{\beta-1} \exp\left[-\left(\frac{t}{\tau}\right)^{\beta}\right], \tag{1}$$

where β and τ are fitting parameters. The mean μ and the coefficient of variation C_V of the Weibull distribution are given by

$$\mu = \tau\, \Gamma\left(1 + \frac{1}{\beta}\right), \tag{2}$$

$$C_V = \left\{ \frac{\Gamma\left(1 + \frac{2}{\beta}\right)}{\left[\Gamma\left(1 + \frac{1}{\beta}\right)\right]^2} - 1 \right\}^{\frac{1}{2}}, \tag{3}$$

where $\Gamma(x)$ is the gamma function of x. The coefficient of variation C_V is the ratio of the standard deviation σ to the mean μ. The cumulative distribution function (cdf) for the Weibull distribution is given by

$$P(t) = 1 - \exp\left[-\left(\frac{t}{\tau}\right)^{\beta}\right], \tag{4}$$

where $P(t)$ is the fraction of the recurrence times that are shorter than t.

If $\beta = 1$ the Weibull distribution becomes the exponential (Poisson) distribution with $\sigma = \mu$ and $C_V = 1$. In the limit $\beta \to +\infty$ the Weibull distribution becomes a δ-function with $\sigma = C_V = 0$. In the range $0 < \beta < 1$ the Weibull distribution is often referred to as the stretched exponential distribution (Laherrere and Sornette, 1998).

An important property of the Weibull distribution is the power-law behavior of the hazard function

$$h(t_0) = \frac{\text{pdf}}{1 - \text{cdf}} = \frac{\beta}{\tau}\left(\frac{t_0}{\tau}\right)^{\beta-1}. \tag{5}$$

The hazard function $h(t_0)$ is the pdf that an event will occur at a time t_0 after the occurrence of the last event. For the Poisson distribution, $\beta = 1$, the hazard function is constant $h(t_0) = \tau^{-1}$ and there is no memory of the last event. For $\beta > 1$ the hazard rate increases as a power of the time t_0 since the last event. The hazard rate is expected to increase monotonically for a driven system.

Major plate boundary faults can be considered to be driven systems since the motions of the tectonic plates systematically increase the stress on these faults. Thus for such characteristic earthquakes it is expected that the hazard function must increase as the time since the last characteristic earthquake t_0 increases (Davis *et al.*, 1989; Sornette and Knopoff, 1997). For this to be the case, the tail of the distribution must be thinner than the exponential distribution. For the lognormal distribution the tail is thicker than the exponential and the hazard function decreases with time t_0. For the Brownian passage-time distribution the tail is exponential for large times and the hazard function becomes

constant. The Weibull distribution with $\beta > 1$ is the only distribution that has been applied to characteristic earthquakes with an increasing hazard function with increasing t_0.

We attribute the universality of the Weibull distribution to the scale invariant (power-law) dependence of the hazard rate. The Weibull distribution is the only distribution which does not have a characteristic value of the time t_0. For example, there is a characteristic value of the time t_0 in the hazard function for the lognormal distribution. It is the value of t_0 where $dh/dt_0 = 0$. We next consider several distributions of characteristic earthquakes on the San Andreas fault.

3. *Parkfield Earthquakes*

Probably the best studied sequence of characteristic earthquakes is the sequence that occurred on the Parkfield, California section of the San Andreas fault between 1857 and 2004 (BAKUN *et al*., 2005). This is because the slip rate is relatively high ($\approx$30 mm/year) and the earthquakes are relatively small ($m \approx 6.0$), thus the recurrence times are relatively short ($\approx$25 years). Also, this fault is subject to a near constant tectonic drive due to the relative motion between the Pacific and North American plates. Slip on the Parkfield section of the San Andreas fault occurred during $m \approx 6$ earthquakes that occurred in 1857, 1881, 1901, 1922, 1934, 1966, and 2004. The mean and coefficient of variation of these recurrence times are $\mu = 24.5$ years, and $C_V = 0.378$, respectively. The cumulative distribution of recurrence times $P(t)$ is given as a function of the recurrence time t in Figure 1. Also included in Figure 1 is the best log likelihood (-21.22) fit of the Weibull distribution with $\tau = 27.4 \pm 3.7$ years and $\beta = 3.21 \pm 1.03$

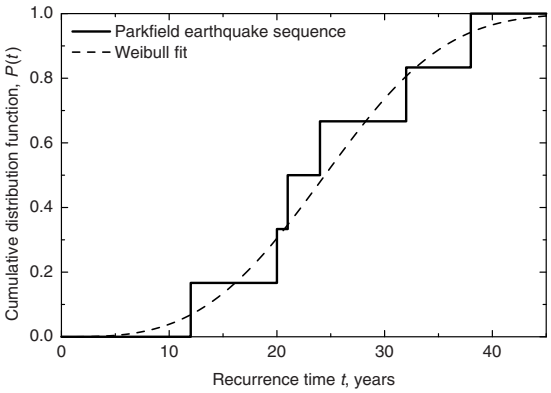


Figure 1
Cumulative distribution of recurrence times of Parkfield earthquakes. The continuous line is the distribution of six actual recurrence times. The discontinuous line is the best-fit Weibull distribution ($\mu = 24.5$ years, $\sigma = 9.25$ years, $C_V = 0.378$, $\tau = 27.4$ years, and $\beta = 3.21$).

(the error bars are 95% confidence limits). GONZALEZ *et al*. (2006) have also correlated the recurrence times of the Parkfield earthquakes.

4. Southern San Andreas

Sets of characteristic earthquakes on the southern section of the San Andreas fault have been obtained from paleoseismic studies. SIEH *et al*. (1989) obtained seven intervals between great earthquakes at the Pallett Creek site. The mean and the coefficient of variation of recurrence times are $\mu = 155$ years and $C_V = 0.645$. These authors fit a Weibull distribution to these data and found that $\tau = 166.1 \pm 44.5$ years and $\beta = 1.50 \pm 0.80$.

BIASI *et al*. (2002) have quantified the paleoseismic recurrence times at the Wrightwood site as well as the Pallett site on the San Andreas fault. At the Wrightwood site interval times are 110, 65, 31, 60, 70, 165, 101, 148, 224, 49, 150, 130, and 44 years. The mean and coefficient of variation of these recurrence times are $\mu = 103.6$ years and $C_V = 0.553$. The cumulative distribution of recurrence times $P(t)$ is given in Figure 2. Also included in Figure 2 is the best log likelihood (-69.55) fit of the Weibull distribution with $\tau = 117.5 \pm 17.1$ years and $\beta = 2.01 \pm 0.44$ (the error bars are 95% confidence limits).

5. Micro-Repeater Earthquake Sequence

Sequences of small earthquakes have been observed to occur on the San Andreas fault in central California (NADEAU *et al*., 1995; NADEAU and MCEVILLY, 2004). These small

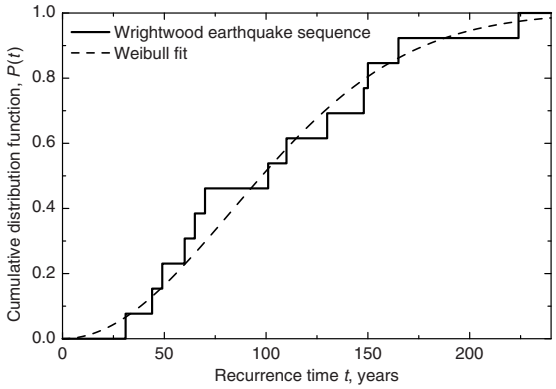


Figure 2
Cumulative distribution of recurrence times of great earthquakes at the Wrightwood site on the southern San Andreas fault. The continuous line is the distribution of 13 actual recurrence times. The discontinuous line is the best-fit Weibull distribution ($\mu = 103.6$ years, $C_V = 0.553$, $\tau = 117.5$ years, and $\beta = 2.01$).

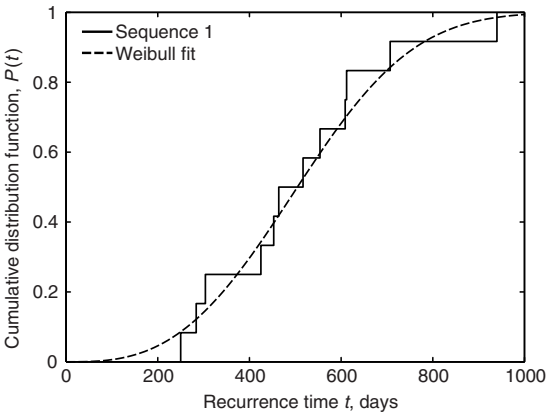


Figure 3
Cumulative distribution of recurrence times of 13 micro-repeating earthquakes on the San Andreas fault near San Juan Batista (mean magnitude $m = 1.36$). The continuous line is the distribution of 12 actual recurrence times. The discontinuous line is the best-fit Weibull distribution ($\mu = 510$ days, $C_V = 0.38$, $\tau = 572$ days, and $\beta = 2.91$).

earthquakes occur repetitively at a number of sites on the fault. Some 23 sequences have been identified. We will consider a sequence of 13 earthquakes with a mean magnitude $\bar{m} = 1.36$ that occurred at a site near San Juan Batista. The sequence of recurrence times for these earthquakes was 707, 250, 284, 517, 304, 554, 609, 453, 425, 940, 612, and 464 days. The mean and coefficient of variation of these recurrence times are $\mu = 510$ days and $C_V = 0.38$. The cumulative distribution of recurrence times $P(t)$ is given in Figure 3. Also included in Figure 3 is the best log likelihood (-79.58) fit of the Weibull distribution with $\tau = 572 \pm 60$ days and $\beta = 2.91 \pm 0.63$ (the error bars are 95% confidence limits).

6. *Virtual California Simulations*

Ideally the recurrence statistics of characteristic earthquakes would be used to determine the applicable distribution. Although the fits of the Weibull distribution to the three sets of recurrence times given in the previous three sections were quite good, the numbers of events in each set were not sufficient to establish definitively the validity of the Weibull distribution (SAVAGE, 1994).

In order to overcome this lack of data, several numerical simulations of earthquake statistics have been carried out. We give results using the Virtual California simulation (RUNDLE *et al*., 2004, 2005; YAKOVLEV *et al*., 2006). This model is a geometrically realistic numerical simulation of earthquakes occurring on the San Andreas fault system. It includes the major strike-slip faults in California and is composed of 650 fault

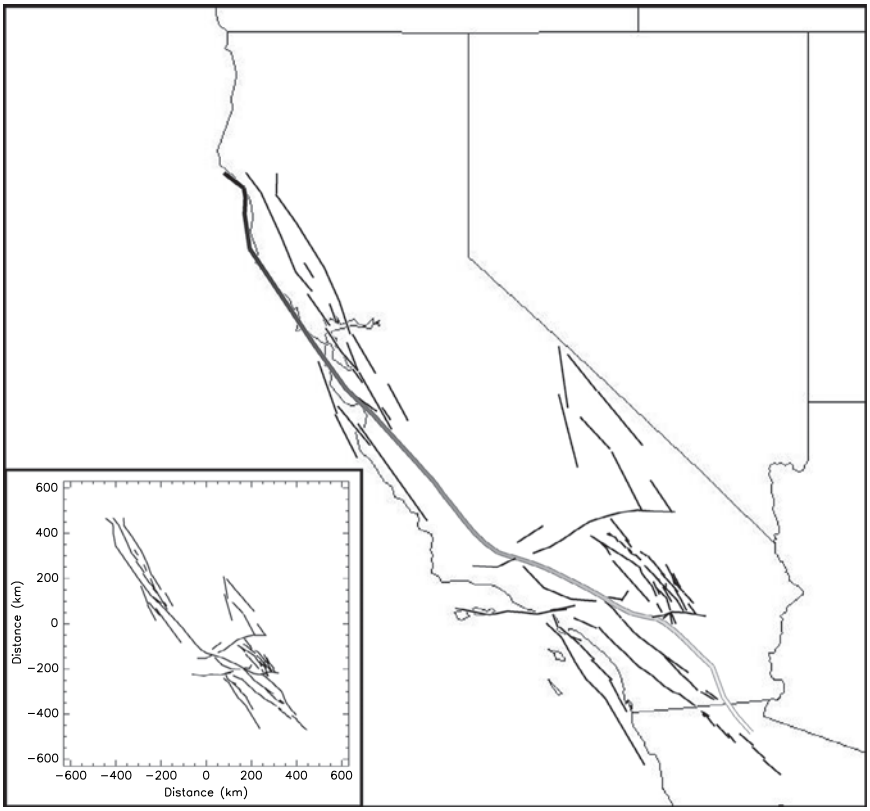


Figure 4
Fault segments making up Virtual California. The model has 650 fault segments, each approximately 10 km in length along-strike and 15 km in depth.

segments; each with a length along-strike of 10 km and a depth of 15 km, as shown in Figure 4. The fault segments are treated as dislocations and interact with each other elastically, utilizing dislocation theory. Virtual California is a backslip model, the accumulation of a slip deficit on each segment is prescribed using available data. The mean recurrence times of earthquakes on each segment are also prescribed using available data to give friction law parameters. As a specific example we consider $M > 7.5$ simulated earthquakes on the northern San Andreas fault (from San Jose to Cape Mendocino, the site of the 1906 San Francisco earthquake). From a 1,000,000 year simulation the mean recurrence time for 4606 overlapping simulated earthquakes on this section is $\mu = 217$ years and the coefficient of variation is $C_V = 0.528$. The cumulative distribution of recurrence times $P(t)$ is given in Figure 5. Also included in Figure 5 is the

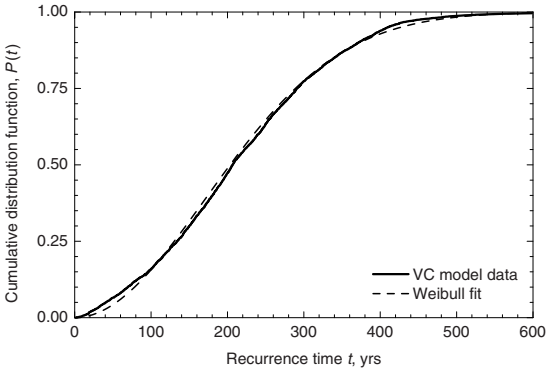


Figure 5
Cumulative distribution of recurrence times for 4606 simulated earthquakes with $M > 7.5$ on the northern San Andreas fault. The continuous line is the distribution of simulated recurrence times. The discontinuous line is the best-fit Weibull distribution ($\mu = 217$ years, $\sigma = 115$ years, $C_V = 0.528$, $\tau = 245$ years, and $\beta = 1.97$).

best fit of the Weibull distribution with $\tau = 245$ years and $\beta = 1.97$. Details of the quality of fit have been given by YAKOVLEV *et al*. (2006).

A similar simulation model "Standard Physical Earth Model" (SPEM) was developed by WARD (1992) and applied to characteristic earthquakes associated with subduction at the Middle American trench. This model was further developed and applied to the entire San Andreas fault system by GOES and WARD (1994), to the San Andreas system in Southern California by WARD (1996), and to the San Andreas system in Northern California by WARD (2000). These authors also correlated their interval statistics with the Weibull distribution, and for their California data found parameter values similar to those obtained from the Virtual California simulations. The two simulation models have many differences, yet the statistical distributions of recurrence times are quite similar. This would indicate that the recurrence time scaling may be a universal feature of seismicity in the sense that Gutenberg-Richter frequency-magnitude scaling is universal.

7. *Forest-Fire Model*

We next turn to two models that have been used to simulate the behavior of characteristic earthquakes, NEWMAN and TURCOTTE (2002) proposed a modification of the forest-fire model to simulate characteristic earthquakes. This model combines aspects of the forest-fire and site-percolation models. The forest-fire model (DROSSEL and SCHWABL, 1992) consists of a square grid of sites. At each time step, either an attempt is made to plant a model tree on a randomly selected site or a model match is dropped on the site. A tree is planted if a site is not occupied by a tree. If a match is dropped on a site occupied

by a tree, that tree and all adjacent trees are destroyed in a model fire. The sparking frequency is the inverse of the number of attempted tree drops before a model match is dropped. It is found that the frequency-area distribution of fires over a range of sizes is power-law with a slope near unity.

The site-percolation model can be regarded as the forest-fire model without fires. This is a transient problem as the number of trees on the grid increases monotonically until every site is occupied by a tree. However, there is a well-defined critical point during this evolution. This critical point occurs when a cluster of trees spans the grid. In the hybrid model, no model matches are dropped however a model fire occurs when a cluster spans the grid from one side to the other (either horizontally or vertically). When a tree cluster spans the grid, it ignites and all the trees in that cluster are removed from the grid. This is a quasi steady-state model like the forest-fire model, but it has only large (characteristic) fires. The variations in the number of residual trees after a fire result in a memory that influences the occurrence of future fires. There is considerable variability in fire sizes and the intervals between fires. The fires in this model are somewhat analogous to characteristic earthquakes on a major fault or fault segment. The planting of trees corresponds to the tectonic increase in stress. The variations in cluster structure correspond to stress variations in the crust. The variability of the sizes is analogous to the variability in characteristic earthquakes and the associated variability in the recurrence times between characteristic earthquakes. In this model, there is a memory of the structure of previous fires for any given fire, each characteristic earthquake is influenced by the structure of previous characteristic earthquakes.

NEWMAN *et al*. (2007) carried out a series of simulations on a 128×128 grid. The spanning fires were found to have a considerable range of sizes from about 2000 to 8000 burned trees. The time-average cumulative distribution of cluster sizes was found to be power-law with a slope near two, the same as for the forest-fire model. The recurrence times of the spanning fires were found to have a mean value $\mu = 7385$ time steps and the coefficient of variation $C_V = 0.394$. The cumulative distribution of recurrence times $P(t)$ is given in Figure 6. Also included in Figure 6 is the best fit of the Weibull distribution with $\tau = 8300$ and $\beta = 2.74$. Details of the quality of fit have been given by NEWMAN *et al*. (2007).

8. Slider-Block Model

BURRIDGE and KNOPOFF (1967) proposed the multiple slider-block model as a simple model for earthquake occurrence. CARLSON and LANGER (1989) showed that the multiple slider-block model exhibits self-organized criticality. The frequency-size distribution of slip events is fractal as is the case for earthquakes (TURCOTTE, 1997). We will now consider the interval statistics of slip events. We will obtain both the Poisson behavior of regional seismicity and the Weibull behavior of characteristic earthquakes. We utilize a variation of the linear slider-block model which CARLSON and LANGER (1989) used.

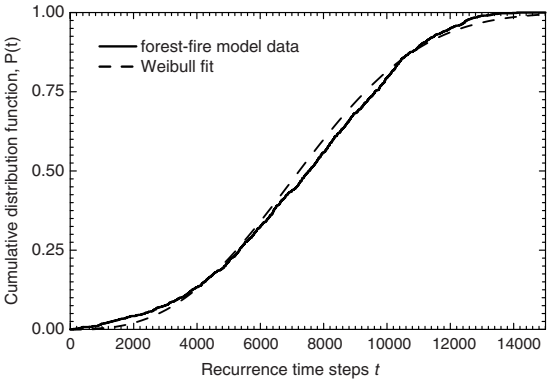


Figure 6
Cumulative distribution of recurrence intervals (in time steps) between model fires. The continuous line is the distribution of simulated recurrence times. The discontinuous line is the best-fit Weibull distribution ($\mu = 7385$, $\sigma = 2908$, $C_V = 0.394$, $\tau = 8300$, and $\beta = 2.74$).

We consider a linear chain of 100 slider blocks of mass m pulled over a surface at a constant velocity V_L by a loader plate as illustrated in Figure 7. Each block is connected to the loader plate by a spring with spring constant k_L. Adjacent blocks are connected to each other by springs with spring constant k_C.

The blocks interact with the surface through a static-dynamic friction law. The static stability of each slider-block is given by

$$k_L y_i + k_C(2y_i - y_{i-1} - y_{i+1}) < F_{Si}, \tag{6}$$

where F_{Si} is the maximum static friction force on block i, and y_i is the position of block i relative to the loader plate.

When the cumulative force from the springs connected to block i exceeds the maximum static friction F_{Si}, the block begins to slide. The dynamic slip of block i is controlled by the equation

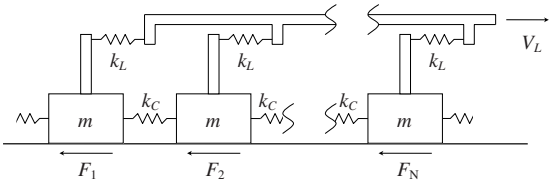


Figure 7
Illustration of the one-dimensional slider-block model. A linear array of N blocks of mass m are pulled along a surface by a constant velocity V_L loader plate. The loader plate is connected to each block with a loader spring with spring constant k_L and adjacent blocks are connected by springs with spring constant k_C. The frictional resisting forces are F_1, F_2, ... , F_N .

$$m\frac{d^2y_i}{dt^2} + k_L y_i + k_C(2y_i - y_{i-1} - y_{i+1}) = F_{Di} \tag{7}$$

where F_{Di} is the dynamic (sliding) frictional force on block i. The loader plate velocity is assumed to be much smaller than the slip velocity so that the movement of the loader plate is neglected during a slip event. The sliding of one block can trigger the instability of the other blocks forming a multiple block event.

It is convenient to introduce the nondimensional variables and parameters

$$\tau_f = t\sqrt{\frac{k_L}{m}},\ \tau_s = \frac{tk_L V_L}{F_S^{\text{ref}}},\ Y_i = \frac{k_L y_i}{F_S^{\text{ref}}},\ \phi = \frac{F_{Si}}{F_{Di}},\ \alpha = \frac{k_C}{k_L},\ \beta_i = \frac{F_{Si}}{F_S^{\text{ref}}}. \tag{8}$$

The ratio of static to dynamic friction ϕ is assumed to be the same for all blocks although the values themselves β_i vary from block to block with F_S^{ref} as a reference value of the static frictional force.

In terms of these nondimensional variables the static stability condition (6) becomes

$$Y_i + \alpha(2Y_i - Y_{i-1} - Y_{i+1}) < \beta_i \tag{9}$$

and the dynamic slip Eq. (7) becomes

$$\frac{d^2Y_i}{d\tau_f^2} + Y_i + \alpha(2Y_i - Y_{i-1} - Y_{i+1}) = \frac{\beta_i}{\phi}. \tag{10}$$

Before obtaining solutions, it is necessary to prescribe the parameters ϕ, α, and β_i. The parameter α is the stiffness of the system. We will give results for $\alpha = 3$ and $\alpha = 1000$. For $\alpha = 3$ the system is soft and there are no system-wide (100 block) events. For $\alpha = 1000$ the system is stiff and system-wide (100 block) events dominate. The ratio ϕ of static friction to dynamic friction is taken to be the same for all blocks $\phi = 1.5$, while the values of frictional parameters β_i are assigned to blocks with a uniform random distribution in the range $1 < \beta_i < 3.5$. This random variability in the system is a "noise" required to generate event variability in stiff systems.

We first give results for a soft system with $\alpha = 3$. The linear event size L is the number of blocks that participate in the event. Frequency-size statistics for 10,000 events are given in Figure 8. The smaller events are well approximated by a power-law relation with exponent -1.29. For larger event sizes there is a rollover and there are no system-wide (100 block) events. We next consider the interoccurrence time statistics for events. In order to do this we specify a threshold magnitude and consider events only larger than this threshold. The cumulative distribution $P(t)$ of interoccurrence times for events in the size range from $L = 10$ to 100 is given in Figure 9. Also included in this figure is the best log likelihood ($-1549.$) fit of the exponential (Poisson) distribution to these data, this is Eq. (4) with $\beta = 1$. To a good approximation the events occur randomly and are not correlated.

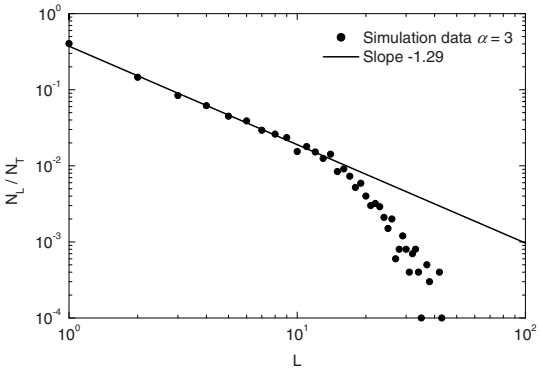


Figure 8
Frequency-size distribution of 10,000 slip events for a "soft" system with $\alpha = 3$. The ratio of the number of events N_L of event size L to the total number of events N_T is given as a function of L. The solid line is a power-law dependence with exponent -1.29.

WEATHERLEY and ABE (2004) have also studied the interoccurrence time statistics for a similar slider-block model. They correlate their results for a soft system with the generalized gamma distribution. The difference between their study and our study is that they considered all events, not just the larger events as we have done. WEATHERLEY and ABE (2004) also correlated the interoccurrence time statistics of a fully-dynamic fault model with the gamma distribution. WEATHERLEY (2006) made a similar correlation for

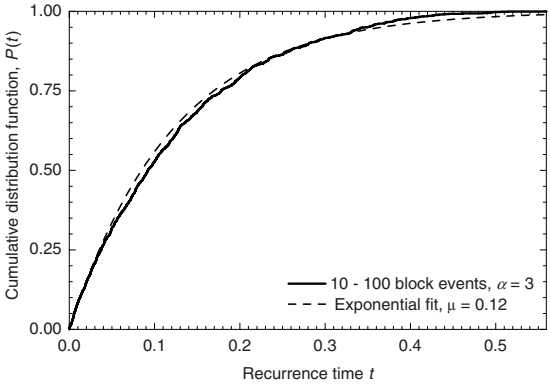


Figure 9
Cumulative distribution function $P(t)$ of interoccurrence times t for the events given in Figure 7 in the size range $L = 10$ to 100. The continuous line is the distribution of observed recurrence times. The dashed line is the best-fit exponential (Poisson) distribution.

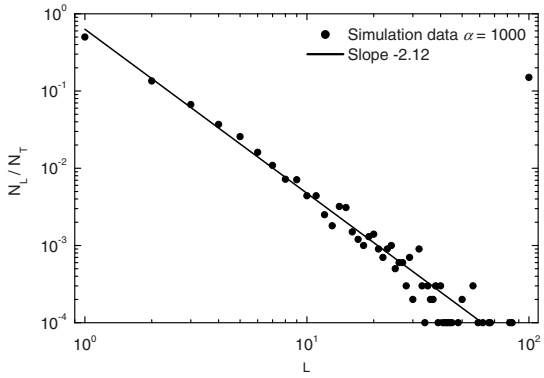


Figure 10
Frequency-size distribution of 10,000 slip events for a "stiff" system with α=1000. The ratio of the number of events N_L of event size L to the total number of events N_T is given as a function of L. The solid line is a power-law dependence with exponent -2.12.

the Olami-Feder-Christensen cellular automaton model. SANCHEZ *et al*. (2002) found that the interoccurrence time statistics for the sandpile model satisfy Poissonian statistics.

We next give results for a stiff system with $\alpha = 1000$. The motion organizes itself into the recurrence of system-wide (100 block) events separated by sets of small size events. Frequency-size statistics for 10,000 events are given in Figure 10. Again the smaller events are well approximated by a power-law relation with exponent -2.12. In this case there are about 1500 system-wide (100 block) events. We consider that these are equivalent to characteristic earthquakes. We next consider the recurrence time statistics for these events. The cumulative distribution of these recurrence times is given in Figure 11. Also included in this figure is the best log likelihood (-1779.) fit of the Weibull distribution (4) to these data obtained by taking $\tau = 0.206 \pm 0.002$ and $\beta = 2.60 \pm 0.05$. There is excellent agreement between the data and the Weibull distribution.

9. Discussion

A variety of self-organizing complex systems have been shown to exhibit power-law frequency-magnitude scaling of "avalanche" sizes (TURCOTTE, 1999a,b). Examples include "sandpile" models, forest-fire models, and slider-block models as well as natural phenomena such as landslides, wild fires, and earthquakes. This behavior appears to be "universal" under a wide range of conditions.

The question addressed in this paper is whether the Weibull recurrence time scaling also has universality, particularly with regard to earthquakes. The Weibull distribution is unique in that it has a scale invariant hazard function. The hazard function $h(t_0)$ is the pdf

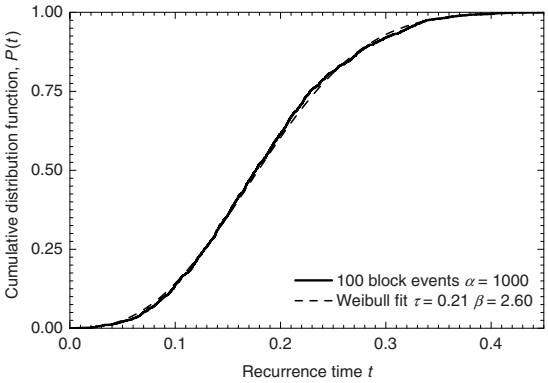


Figure 11
Cumulative distribution function $P(t)$ of recurrence times t for the 1500 system-wide (100 block) events with $\alpha = 1000$. The continuous line is the distribution of observed recurrence times. The dashed line is the best-fit Weibull distribution with $\tau = 0.21$ and $\beta = 2.60$.

that an event in a series of events at a specified location will occur at a time t_0 after the previous event. The hazard function for the Weibull distribution has a power-law dependence on t_0 as shown in Eq. (5). All other distributions have a characteristic time in their hazard function. For example, for the lognormal distribution, the characteristic time is the time when $dh/dt_0 = 0$, the hazard function initially increases with increasing t_0 and then decreases for larger values (SORNETTE and KNOPOFF, 1997).

In this paper we have considered three sets of characteristic earthquakes on the San Andreas fault: (1) the Parkfield earthquakes, (2) the sequence of 14 earthquakes identified by paleoseismic studies at the Wrightwood site, and (3) an example of a sequence of micro-repeating earthquakes at a site near San Juan Batista. In each case the distribution of recurrence times correlates well with the Weibull distribution. We also find that the distribution of recurrence times for simulated earthquakes on the northern San Andreas fault correlates well with the Weibull distribution.

In order to demonstrate the applicability of the Weibull distribution we consider the behavior of a slider-block model. We carry out a full dynamic simulation using a linear array of 100 slider blocks. The behavior of the model is governed by its "stiffness" α, the ratio of the connector spring constant to the loader spring constant as given in Eq. (8). For low values of α, soft systems, there are no system-wide events. For high values of α, stiff systems, large number of system-wide events occur in which all blocks slip simultaneously.

We find that the statistical distribution of recurrence times between system-wide events with $\alpha = 1000$ satisfies a Weibull distribution to a very good approximation. We have also considered the distribution of interoccurrence times for large events for a soft system with $\alpha = 3$. In this case the distribution is well approximated by an exponential

(Poisson) distribution. An event does not have a significant "memory" of prior events. This is consistent with prior studies of SOC models such as the sandpile model (SANCHEZ *et al*., 2002) in which exponential distributions were found.

The statistical distribution of recurrence times of characteristic earthquakes plays an important role in probabilistic hazard assessment. Specific examples are the series of assessments of California earthquake probabilities (JACKSON *et al*., 1995; WORKING GROUP ON CALIFORNIA EARTHQUAKE PROBABILITIES, 1988, 1990, 2003). In the first three of these studies the lognormal distribution was used. The use of this distribution was strongly criticized by a number of authors (DAVIS *et al*., 1989; SORNETTE and KNOPOFF, 1997). The hazard function for the lognormal distribution increases at first but then decreases. The longer it has been since the last earthquake, the longer the expected time to the next earthquake. There is a characteristic time when the time derivative of the hazard function is zero. The decrease in the hazard function does not appear to be acceptable. Because of this problem, the Brownian passage-time distribution (MATTHEWS *et al*., 2002) was used in the 2003 study. At large times the hazard rate becomes constant. The expected time for the next earthquake does not depend upon when the last earthquake accured. There is a characteristic time when the hazard function no longer increases. Again this behavior does not appear to be acceptable.

Many of the early studies of the recurrence time statistics of characteristic earthquakes utilized the Weibull distribution (HAGIWARA, 1974; RIKITAKE, 1976, 1982). In general the number of data points in any sequence of recurrence times is too small to differentiate between alternative distributions (SAVAGE, 1994; SORNETTE and KNOPOFF, 1997). In order to overcome this difficulty YAKOVLEV *et al*. (2006) utilized data from a one million year simulation of earthquakes on the northern San Andreas fault. This study showed conclusively that the Weibull distribution gives a better fit to the data than the lognormal and Brownian passage-time distributions. Their fit of the Weibull distribution to the simulated data is given in Figure 5.

It is important to make a clear distinction between the pdf of the recurrence times and the hazard function. There is clearly a mean return time for characteristic earthquakes, and this is incorporated in proposed distributions. The mean recurrence time of the Weibull distribution has been given in Eq. (2). As defined in Eq. (5) the hazard function can be scale-invariant (and is for the Weibull distribution) even though the distribution of recurrence times is thin tailed. We argue that this scale invariance of the hazard function leads to the universal applicability of the Weibull distribution to many self-organizing complex systems as shown in this paper.

Acknowledgements

This research has been supported by NSF Grants. ATM 0327799 (SGA, DLT, RS, WIN) and U.S. Department of Energy Grant DE-FG02-04ER15568 (JBR). We would like to thank Alvaro Corral and Steffen Abe for very helpful reviews.

REFERENCES

ALTMANN, E.G., DA SILVA, E.C., and CALDAS, I.L. (2004), *Recurrence time statistics for finite size intervals*, Chaos *14*(4), 975–981.

ALTMANN, E.G. and KANTZ, H. (2005), *Recurrence time analysis, long-term correlations, and extreme events*, Phys. Rev. E *71*(5), 056106.

BAK, P., CHRISTENSEN, K., DANON, L., and SCANLON, T. (2002), *Unified scaling law for earthquakes*, Phys. Rev. Lett. *88*(17), 178501.

BAKUN, W.H., AAGAARD, B., DOST, B., ELLSWORTH, W.L., HARDEBECK, J.L., HARRIS, R.A., *et al*. (2005), *Implications for prediction and hazard assessment from the 2004 Parkfield earthquake*, Nature *437*(7061), 969–974.

BIASI, G.P., WELDON, R.J., FUMAL, T.E., and SEITZ, G.G. (2002), *Paleoseismic event dating and the conditional probability of large earthquakes on the southern San Andreas fault, California*, Bull. Seismol. Soc. Am. *92*(7), 2761–2781.

BUNDE, A., EICHNER, J.F., HAVLIN, S., and KANTELHARDT, J.W. (2003), *The effect of long-term correlations on the return periods of rare events*, Physica A *330*(1–2), 1–7.

BUNDE, A., EICHNER, J.F., HAVLIN, S., and KANTELHARDT, J.W. (2004), *Return intervals of rare events in records with long-term persistence*, Physica A *342*(1–2), 308–314.

BURRIDGE, R., and KNOPOFF, L. (1967), *Model and Theoretical Seismicity*, Bull. Seismol. Soc. Am. *57*, 341–371.

CARBONE, V., SORRISO-VALVO, L., HARABAGLIA, P., and GUERRA, I. (2005), *Unified scaling law for waiting times between seismic events*, Europhys. Lett. *71*(6), 1036–1042.

CARLSON, J.M., and LANGER, J.S. (1989), *Mechanical model of an earthquake fault*, Phys. Rev. A *40*(11), 6470–6484.

CORRAL, A. (2003), *Local distributions and rate fluctuations in a unified scaling law for earthquakes*, Phys. Rev. E *68*(3), 035102(R).

CORRAL, A. (2004a), *Long-term clustering, scaling, and universality in the temporal occurrence of earthquakes*, Phys. Rev. Lett. *92*(10), 108501.

CORRAL, A. (2004b), *Universal local versus unified global scaling laws in the statistics of seismicity*, Physica A *340*(4), 590–597.

CORRAL, A. (2005a), *Mixing of rescaled data and Bayesian inference for earthquake recurrence times*, Nonlinear Proc. Geophys. *12*(1), 89–100.

CORRAL, A. (2005b), *Time-decreasing hazard and increasing time until the next earthquake*, Phys. Rev. E *71*(1), 017101.

CORRAL, A. (2007), *Statistical features of earthquake temporal occurrence*, Lect. Notes Phys. *705*, 191–221.

DAVIDSEN, J., and GOLTZ, C. (2004), *Are seismic waiting time distributions universal*?, Geophys. Res. Lett. *31*(21), L21612.

DAVIS, P.M., JACKSON, D.D., and KAGAN, Y.Y. (1989), The *longer it has been since the last earthquake, the longer the expected time till the next*, Bull. Seismol. Soc. Am. *79*(5), 1439–1456.

DROSSEL, B. and SCHWABL, F. (1992), *Self-organized critical forest-fire model*, Phys. Rev. Lett. *69*(11), 1629–1632.

GOES, S.D.B., and WARD, S.N. (1994), *Synthetic seismicity for the San Andreas fault*, Ann. Geofisica *37*, 1495–1513.

GONZALEZ, A., GOMEZ, J.B., and PACHECO, A.F. (2006), *Updating Seismic Hazard at Parkfield*, J. Seismol. *10*(2), 131–135.

HAGIWARA, Y. (1974), *Probability of earthquake occurrence as obtained from a Weibull distribution analysis of crustal strain*, Tectonophys. *23*(3), 313–318.

JACKSON, D.D., AKI, K., CORNELL, C.A., DIETERICH, J.H., HENYEY, T.L., MAHDYIAR, M. *et al*. (1995), *Seismic hazards in Southern California - Probable earthquakes, 1994 to 2024*, Bull. Seismol. Soc. Am. *85*(2), 379–439.

LAHERRERE, J. and SORNETTE, D. (1998), *Stretched exponential distributions in nature and economy: "Fat Tails" with characteristic scales*, Eur. Phys. J. B *2*(4), 525–539.

LINDMAN, M., JONSDOTTIR, K., ROBERTS, R., LUND, B., and BODVARSSON, R. (2005), *Earthquakes descaled: On waiting time distributions and scaling laws*, Phys. Rev. Lett. *94*(10), 108501.

LIVINA, V., HAVLIN, S., and BUNDE, A. (2005a), *Memory in the occurrence of earthquakes*, Phys. Rev. Lett. *95*(20), 208501.

LIVINA, V., TUZOV, S., HAVLIN, S., and BUNDE, A. (2005b), *Recurrence intervals between earthquakes strongly depend on history*, Physica A *348*, 591–595.

MANDELBROT, B., *The Fractal Geometry of Nature* (Freeman, San Francisco, 1982).

MATTHEWS, M.V., ELLSWORTH, W.L., and REASENBERG, P.A. (2002), *A Brownian model for recurrent earthquakes*, Bull. Seismol. Soc. Am. *92*(6), 2233–2250.

MEEKER, W.Q., and ESCOBAR, L.A., *Statistical Methods for Reliability Data* (John Wiley, New York, 1991).

NADEAU, R.M., FOXALL, W., and MCEVILLY, T.V. (1995), *Clustering and periodic recurrence of microearthquakes on the San-Andreas fault at Parkfield, California*, Science *267*(5197), 503–507.

NADEAU, R.M., and MCEVILLY, T.V. (2004), *Periodic pulsing of characteristic microearthquakes on the San Andreas fault*, Science *303*(5655), 220–222.

NEWMAN, W.I. and TURCOTTE, D.L. (2002), *A simple model for the earthquake cycle combining self-organized complexity with critical point behavior*, Nonlinear Proc. Geophys. *9*(5–6), 453–461.

NEWMAN, W.I., TURCOTTE, D.L., SHCHERBAKOV, R., RUNDLE, J.B., RUNDLE, P.B., and YAKOVLEV, G. (2007), *Recurrence time statistics and the Weibull distribution: earthquakes and models*, Phys. Rev. E, submitted.

NISHENKO, S.P. and BULAND, R. (1987), *A generic recurrence interval distribution for earthquake forecasting*, Bull. Seismol. Soc. Am. *77*(4), 1382–1399.

OGATA, Y. (1999), *Estimating the hazard of rupture using uncertain occurrence times of paleoearthquakes*, J. Geophys. Res. *104*(B8), 17995–18014.

PATEL, J.K., KAPADIA, C.H., and OWEN, D.B., *Handbook of Statistical Distributions* (Marcel Dekker, New York, 1976).

RIKITAKE, T. (1976), *Recurrence of great earthquakes at subduction zones*, Tectonophys., *35*(4), 335–362.

RIKITAKE, T., *Earthquake Forecasting and Warning* (D. Reidel Publishing Co., Dordrecht, 1982).

RIKITAKE, T. (1991), *Assessment of earthquake hazard in the Tokyo area, Japan*, Tectonophys. *199*(1), 121–131.

RUNDLE, J.B., RUNDLE, P.B., DONNELLAN, A., and FOX, G. (2004), *Gutenberg-Richter statistics in topologically realistic system-level earthquake stress-evolution simulations*, Earth Planets Space *56*(8), 761–771.

RUNDLE, J.B., RUNDLE, P.B., DONNELLAN, A., TURCOTTE, D.L., SHCHERBAKOV, R., LI, P., *et al.* (2005), *A simulation-based approach to forecasting the next great San Francisco earthquake*, Proc. Natl. Acad. Sci. U.S.A. *102*(43), 15363–15367.

SANCHEZ, R., NEWMAN, D.E., and CARRERAS, B.A. (2002), *Waiting-time statistics of self-organized-criticality systems*, Phys. Rev. Lett., *88*(6), 068302.

SAVAGE, J.C. (1994), *Empirical earthquake probabilities from observed recurrence intervals*, Bull. Seismol. Soc. Am. *84*(1), 219–221.

SHCHERBAKOV, R., YAKOVLEV, G., TURCOTTE, D.L., and RUNDLE, J.B. (2005), *Model for the distribution of aftershock interoccurrence times*, Phys. Rev. Lett. *95*, 218501.

SIEH, K., STUIVER, M., and BRILLINGER, D. (1989), *A more precise chronology of earthquakes produced by the San-Andreas fault in Southern California*, J. Geophys. Res. *94*(B1), 603–623.

SORNETTE, D. and KNOPOFF, L. (1997), *The paradox of the expected time until the next earthquake*, Bull. Seismol. Soc. Am. *87*(4), 789–798.

TURCOTTE, D.L., *Fractals and Chaos in Geology and Geophysics* (2nd ed.) (Cambridge Univ. Press, Cambridge, 1997).

TURCOTTE, D.L. (1999a), *Seismicity and self-organized criticality*, Phys. Earth Planet. Inter. *111*(3–4), 275–293.

TURCOTTE, D.L. (1999b), *Self-organized criticality*, Rep. Prog. Phys. *62*(10), 1377–1429.

UTSU, T. (1984), *Estimation of parameters for recurrence models of earthquakes*, Bull. Earthquake Res. Insti.Univ. Tokyo *59*, 53–66.

WARD, S.N. (1992), *An application of synthetic seismicity in earthquake statistics: The middle America trench*, J. Geophys. Res. *97*(B5), 6675–6682.

WARD, S.N. (1996), *A synthetic seismicity model for Southern California: Cycles, probabilities, and hazard*, J. Geophys. Res. *101*(B10), 22393–22418.

WARD, S.N. (2000), *San Francisco Bay area earthquake simulations: A step toward a standard physical earthquake model*, Bull. Seismol. Soc. Am. *90*(2), 370–386.

WEATHERLEY, D. (2006), *Recurrence interval statistics of cellular automaton seismicity models*, Pure Appl. Geophys. *163*(9), 1933–1947.

WEATHERLEY, D., and ABE, S. (2004), *Earthquake statistics in a block slider model and a fully dynamic fault model*, Nonlinear Proc. Geophys. *11*(5–6), 553–560.

WEIBULL, W. (1951), *A statistical distribution function of wide applicability*, J. Appl. Mech. *18*(3), 293–297.

WORKING GROUP ON CALIFORNIA EARTHQUAKE PROBABILITIES (1988), *Probabilities of Large Earthquakes Occurring in California on the San Andreas Fault*, U.S. Geological Survey (Open-File Report No. 88–398).

WORKING GROUP ON CALIFORNIA EARTHQUAKE PROBABILITIES (1990), *Probabilities of Large Earthquakes in the San Francisco Bay Region, California*, U.S. Geological Survey (Circular No. 1053).

WORKING GROUP ON CALIFORNIA EARTHQUAKE PROBABILITIES (2003), *Earthquake Probabilities in the San Francisco Bay Region, 2002–2031*, U.S. Geological Survey (Open-File Report No. 2003–214).

YAKOVLEV, G., TURCOTTE, D.L., RUNDLE, J.B., and RUNDLE, P.B. (2006), *Simulation based distributions of earthquake recurrence times on the San Andreas fault system*, Bull. Seismol. Soc. Am. *96*, 1995–2007.

(Received August 2006, Revised March 9, 2007, accepted March 21, 2007)
Published Online First: April 2, 2008