Hazard Calculations for the WGCEP-2002 Earthquake Forecast Using OpenSHA and Distributed Object Technologies

Edward H. Field[1], Nitin Gupta[2], Vipin Gupta[2], Michael Blanpied[3], Philip Maechling[1], and Thomas H. Jordan[1]

INTRODUCTION

We present seismic-hazard calculations for what is arguably the most sophisticated earthquake forecast ever developed—the model by the 2002 Working Group on California Earthquake Probabilities (2003), or WGCEP-2002 as referred to hereafter. These calculations have been made possible by developments in both OpenSHA (Field *et al.*, 2003) and the Information Technology Research (ITR) Collaboration of the Southern California Earthquake Center (SCEC) (Jordan *et al.*, 2003). In particular, we demonstrate the use of a freely available, platform-independent, and graphical-user-interface–based application for computing hazard curves. This application utilizes distributed-object technologies, meaning the code running on the client's computer communicates over the Internet with code hosted elsewhere on server computers. This interoperability, which is invisible to the user, makes the application fast and executable without heavy installation requirements. Instructions on how to reproduce the calculations presented here are available under "Publications" at the Web site http://www.OpenSHA.org/.

OpenSHA

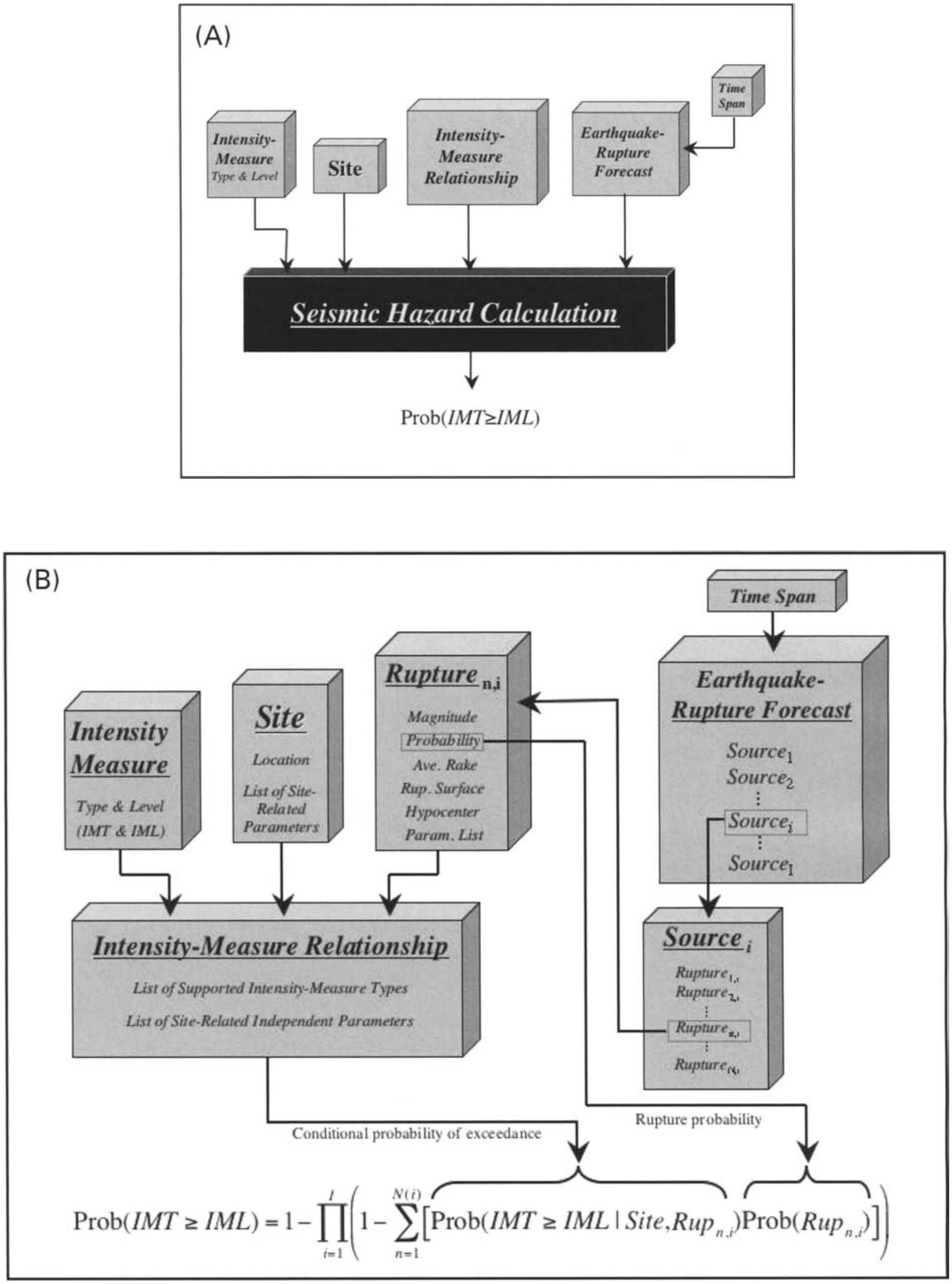
OpenSHA (Field *et al.*, 2003) is a collaborative effort between SCEC and the United States Geological Survey (USGS). The goal has been to develop a "community modeling environment" or "collaboratory" for seismic hazard analysis (SHA), where any arbitrarily complex model component can be plugged in for analysis (without having to change what is being plugged into). This is a departure from previous approaches in which PSHA code (typically FORTRAN) would be rewritten for each new application. Developing a more flexible "plug and play" environment is particularly important in that proper SHA requires that all viable models be considered in the analysis (SSHAC, 1997).

1. U.S. Geological Survey, Pasadena
2. SCEC
3. U.S. Geological Survey, Reston

As the name implies, OpenSHA is an open-source development, meaning anyone is free to examine the code. It is also object-oriented (enabling the plug-and-play modularity), platform-independent, accessible via a graphical user interface (GUI), and potentially network-distributed. These features have been achieved using the Java programming language. There is no requirement that all components be written in Java, however, as exemplified by the WGCEP-2002 model existing as FORTRAN code "wrapped" in Java (discussed more below).

The basic OpenSHA framework and computational sequence are illustrated in Figure 1. The two main model components are an Earthquake-Rupture Forecast (ERF), which gives an inventory of all possible earthquake-rupture events in the region, and their associated probabilities, over a specified time span and above some magnitude threshold; and an Intensity-Measure Relationship (IMR), which gives the probability that some Intensity Measure Type (such as peak ground acceleration) will exceed a specified value at a site given the occurrence of an earthquake-rupture event. Note that attenuation relationships are one particular type (subclass) of IMR, leaving the overall definition more general to accommodate models that use, for example, full waveform modeling. More details of the overall framework can be found in Field *et al.* (2003).

Using the basic design depicted in Figure 1, we have developed applications for computing hazard curves, hazard spectra, scenario ShakeMaps, and full hazard maps. For example, one can currently use our scenario-ShakeMap application to produce maps for any one of the hundreds to tens of thousands of ruptures defined in the WGCEP-2002 model (exact numbers depend on parameter settings), any one of the more than 1,000,000 ruptures defined for California in the 2002 National Seismic Hazard Map model (Frankel *et al.*, 2002), or any custom "hand built" rupture. One can choose an arbitrary IMR for the calculation, and site effects can be included if so desired. As exemplified in a recent study of the Puente Hills blind thrust fault in Los Angeles (Field *et al.*, 2004), the results can also be imported into HAZUS-MH (http://www.fema.gov/hazus) for loss estimation.



▲ **Figure 1.** (A) The fundamental elements needed to compute the probability that an intensity-measure type (IMT) will exceed a certain intensity-measure level (IML). (B) The calculation sequence inside the black box of (A). *Rupture* is short for *Probabilistic Earthquake Rupture*, and *Source* is shorthand for *Earthquake Source*. See Field *et al.* (2003) for more details.

Again, the main goal has been to enable any new ERF or IMR to plug into these applications without having to rewrite existing code. All the OpenSHA applications need to know is that the model component exists and where to find it. The IMR's that are currently available, all of which are attenuation relationships, include Abrahamson and Silva (1997), Boore *et al.* (1997), Campbell (1997), Sadigh *et al.* (1997), Spudich *et al.* (1999), Abrahamson (2000), Field (2000), Campbell and Bozorgnia (2003), and "ShakeMap-2003" (Wald, 2003). Available ERF's include the WGCEP (2002) model discussed here, the California component of both the 1996 and 2002 National Seismic Hazard Maps (Frankel *et al.*, 2000; Frankel *et al.*, 2002), a Short Term Earthquake Probability (STEP) model for California based on foreshock/aftershock statistics (Gerstenberger *et al.*, 2004), and a variety of generic models to implement relatively simple test cases. New IMR and ERF models are added on a regular basis.

Code validation is an important part of any software development. To this end, OpenSHA has participated in the formal PSHA-validation exercises sponsored by the Pacific Earthquake Engineering Research (PEER) Center (http://www.pubs.asce.org/WWWdisplay.cgi?0409668). Furthermore, an advantage of the object-oriented (modular) framework is that each component can be validated independently, which is the approach we have taken. For example, each of the attenuation relationships listed above has been independently validated by Kenneth Campbell (written communication), and his results have been used to establish an automatic revalidation procedure that can be executed every time the code is modified (known as "JUnit" testing in Java). Details on how other OpenSHA components have been tested are provided in the documentation available at http://www.OpenSHA.org/.

THE WGCEP-2002 FORECAST

The ERF developed by the WGCEP-2002 is the latest in a series of time-dependent forecasts issued for the San Francisco Bay area (*e.g.*, WGCEP, 1988, 1990). This most recent model is novel in that it accounts for earthquake-induced stress changes, most notably the apparent stress shadow cast by the M 7.8 earthquake in 1906. Perhaps the most innovative aspect of the model, however, is the extensive treatment of epistemic uncertainties. This type of uncertainty is when only one possible value exists, it's just unknown at this time what that value is. This is in contrast to an aleatory uncertainty, where different values will occur with some relative frequency of occurrence.

The epistemic uncertainties that the WGCEP-2002 model accounts for include fault segment endpoints; seismogenic thicknesses; fault slip rates; the relative frequency of various multisegment ruptures; magnitude-area scaling relationships; the percentage of aseismic slip; magnitude frequency distributions; the amount of off-fault seismicity; and the rupture probability model applied to each fault. The latter uncertainty, the probability model used to predict the next rupture, was found by the working group to be the most influential in the forecast. The options include a time-independent Poisson model plus the following four time-dependent models: an "Empirical Poisson" model where long-term rates are scaled down to be more consistent with recent seismicity rates; the Brownian Passage Time (BPT) model, which depends on the date of the last event; a "BPT step" model, which accounts for earthquake-induced stress changes; and a "Time Predictable" model (available only on the San Andreas Fault).

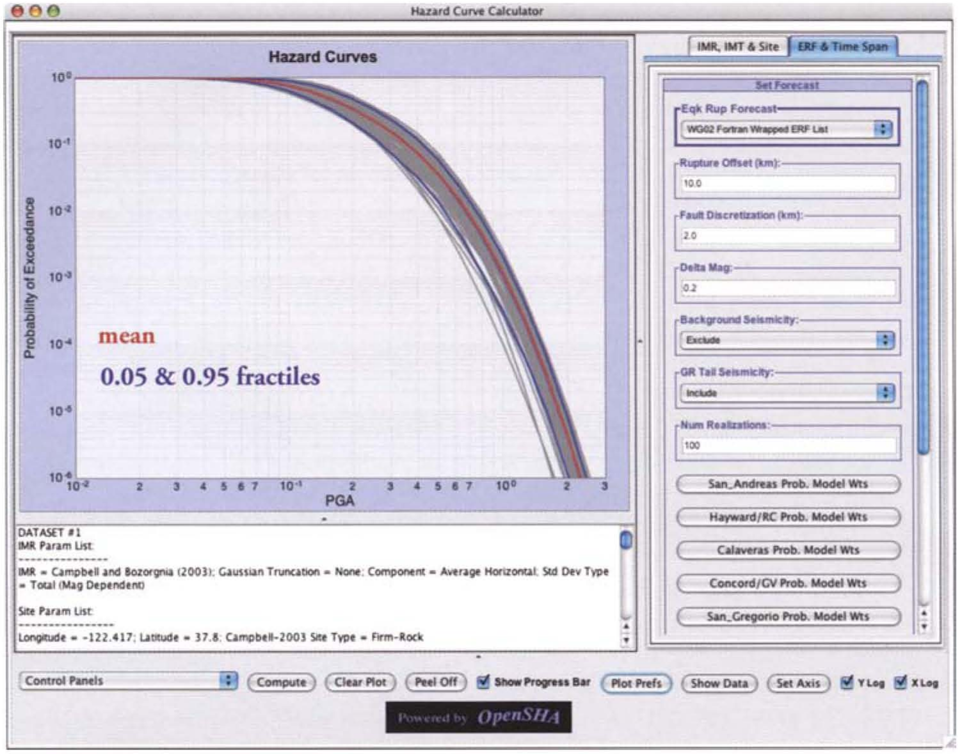
The WGCEP-2002 treatment of epistemic uncertainties is one of the reasons the model is arguably the most sophisticated ERF ever developed. The parameter values and associated probabilities, or "logic-tree branch weights" in the parlance of probabilistic SHA, are specified in a single input file to the FORTRAN code. Given the large number of possible combinations (outer branches of the logic tree), the code does not systematically evaluate each one. Rather, a Monte Carlo approach is taken where each realization of the model is based on randomly chosen parameter values (collectively defining a single, complete logic-tree branch). Each realization essentially provides the magnitude and probability of every possible earthquake rupture in the region, or, in other words, each realization provides a separate ERF. The desired number of realizations is specified in the input file, and one of the output files provides the information needed to construct that number of associated ERF's. Each ERF is assigned an equal probability of being the correct ERF.

The suite of ERF's generated by the code represents the overall uncertainty of the forecast. For example, based on 10,000 realizations, WGCEP-2002 reported that the total probability of a $M \geq 6.7$ occurring in the San Francisco Bay area between 2002 and 2032 is 0.62, with a 95% confidence interval of 0.37 to 0.87. WGCEP-2002 did not extend its analysis to full hazard curves (by pairing its ERF with an IMR), at least not for the time-dependent model. The opportunity to complete this analysis presented OpenSHA with the challenge of plugging this most sophisticated ERF into our framework without compromising our design goals.

MODEL ACCESS VIA DISTRIBUTED-COMPUTING TECHNOLOGIES

The primary difficulty in plugging the WGCEP-2002 model into OpenSHA was providing access to the FORTRAN code while maintaining the platform independence of our applications (the download and double-click convenience). The problem is that FORTRAN code must be recompiled for every platform, so bundling it in our applications was difficult.

The solution was to deploy the WGCEP-2002 model on a server that can be accessed over the Internet by any of the OpenSHA applications. Specifically, we deployed a network-accessible, Java-based ERF interface that configures and runs the FORTRAN code upon request (a "Java wrapped" version of the WGCEP-2002 model). The remote access was originally achieved using Java Servlet technologies (*e.g.*, http://novocode.com/doc/servlet-essentials/), but we subsequently



▲ **Figure 2.** Screenshot of our server-mode hazard curve calculator. Shown are 30-year PGA hazard curves for a site near downtown San Francisco (latitude 37.80°, longitude –122.42°) computed using the WGCEP-2002 forecast paired with the Campbell and Bozorgnia (2003) attenuation relationship (using their magnitude-dependent standard deviation and no Gaussian truncation). This site is classified as NEHRP B according to Wills *et al.* (2000), which translates to the Campbell and Bozorgnia (2003) "Firm Rock" category. One hundred iterations of the model were requested, giving the 100 individual curves plotted in gray (most of which overlap to form a solid gray region). The mean curve is plotted in red, and the 0.05 and 0.95 fractiles (or 90% confidence bounds) are plotted in blue. The following were the settings for various WGCEP-2002 adjustable parameters: Rupture Offset = 10 km; Fault Discretization = 2 km; Delta Mag = 0.2; Background Seismicity = "Exclude"; GR Tail Seismicity = "Include"; and the probability model weights for all faults were the preferred values defined by the working group (the defaults in the application). Readers can explore the influence of alternative parameter settings by running the application themselves (see instructions under "Publications" at http://www.OpenSHA.org).

adopted Java's Remote Method Invocation (RMI) capabilities (*e.g.*, http://java.sun.com/developer/technicalArticles/RMI/rmi/) as a more elegant solution to our particular problem. Recall that OpenSHA is an object-oriented, or modular, framework. RMI simply enables our Java modules, such as an ERF, to exist anywhere on the network. As discussed in the Electronic Seismologist column of this issue (Maechling *et al.*, 2005), we also explored the use of so-called Web Services, which have the advantage of being a more general, less-Java-specific solution. We found that Web Services are not yet up to the task of relaying the relatively complex information we are dealing with.

HAZARD CURVE CALCULATIONS

A screenshot of the OpenSHA hazard-curve calculator is shown in Figure 2. The first step in using the application is to choose an IMR the user would like to apply. The present options are any of the attenuation relationships listed above. The next step is to choose the type of intensity measure (the options depend on those supported by the chosen IMR). For example, one might choose peak ground acceleration (PGA), so the subsequent hazard curve will give the probability of exceeding various PGA levels. Next, the user sets the latitude and longitude for the site of interest, as well as any site-related parameters used by the selected IMR. If the site is located in California, one can utilize a control panel to set the site type automatically from the Wills *et al.* (2000) classification map. This panel gets the information from a Java Servlet, which currently resides on a computer at the University of Southern California (USC).

The next step is to choose an ERF for the calculation, in this case the model labeled "WGCEP-2002 Fortran Wrapped ERF List." Again, this is an "ERF List" because the model produces multiple ERF realizations (each of which is given an equal probability of being correct). The application then queries the chosen ERF (located on the server) for any user-adjustable parameters and lists these in the GUI for the user to set. The adjustable parameters provided by the present WGCEP-2002 model implementation include various discretization intervals; whether to include background and/or Gutenberg-Richter-tail seismicity (the latter being a distribution of smaller events added to the characteristic ruptures for each fault source); the number of ERF realizations desired; and the relative branch weights for the probability models used on each fault (the latter default to the values preferred by the working group). Because we are dealing with an ERF list, the application will compute a separate hazard curve for each ERF in the list. Therefore, we also have the option, specified via another control panel, to plot each curve separately, the mean, or any particular fractiles.

The final task before proceeding with the calculation is to set the time span for the forecast. Because the WGCEP-2002 model is time-dependent, the user must choose both a start time and duration. Presently, the user can select only 2002 as the start year, and the choice of duration is 1, 5, 10, 20, or 30 years (these limits are imposed by the WGCEP-2002 model itself, not by the OpenSHA implementation).

When one clicks "Compute" to generate the hazard curves, the application sends the WGCEP-2002 parameter settings chosen by the user to the server, which then configures the input file, runs the FORTRAN code, reads an output file, and builds the ERF list (all of this occurs in a matter of seconds). This ERF list is then combined, in a simplified manner of speaking, with the chosen IMR and intensity-measure type to compute the hazard curves. To maximize performance, these calculations are done on the server using a separate module. Each curve takes a few seconds to generate, the exact duration depending on the various parameter settings, so a progress bar is provided to keep the user abreast. The results are then passed back to the application and plotted.

Figure 2 shows 30-year PGA hazard curves for a site in downtown San Francisco from 100 realizations of the WGCEP-2002 model using the Campbell and Bozorgnia (2003) attenuation relationship. The mean probability of exceeding 0.5 g is ~0.03, with a 90% confidence bound of 0.01 to 0.05. Figure 3 shows mean hazard curves (from 100 iterations each) using the four attenuation relationships applied in the 2002 National Seismic Hazard Maps: Campbell and Bozorgnia (2003), Boore *et al.* (1997), Abrahamson and Silva (1997), and Sadigh *et al.* (1997). The range of mean exceedance probabilities at 0.5 g PGA obtained from these alternative models is 0.01 to 0.06. Thus, the epistemic uncertainties from the WGCEP-2002 model (Figure 2) are of the same order as those implied by the alternative attenuation relationships (Figure 3). The reason different attenuation relationships give different hazard estimates, in spite of being based on similar regression data, is that each extrapolates differently from existing observations to the relatively large-magnitude and short-distance events that dominate the hazard.

Figure 4 shows mean PGA hazard curves for the different durations supported by the ERF. Not surprisingly, the probability of exceedance increases with duration. The increase does not follow a simple Poissonian scaling, however, because the model includes conditional probabilities that depend on the dates of previous events (something worth keeping in mind in terms of performance-based design when the probability for different exposure times is considered). As discussed previously, WGCEP-2002 assigned relative weights for alternative probability models on each fault, and the results in Figures 2–4 reflect their preferences. Figure 5 compares mean curves obtained by giving the Empirical Poisson and BPT models exclusive weight (chosen because these represent the low and high regional-probability end members, respectively). The Empirical Poisson model gives a probability of ~0.023 at 0.5 g PGA, and the BPT model gives ~0.037. These relative values are consistent with the relative probability of occurrence values listed in Table 6.15 of WGCEP-2002. Note that the spread of values in Figure 5 is low compared to the individual curves in Figure 2, implying that the other epistemic uncertainties discussed above have a significant contribution (even though the probability model was found to be most influential).

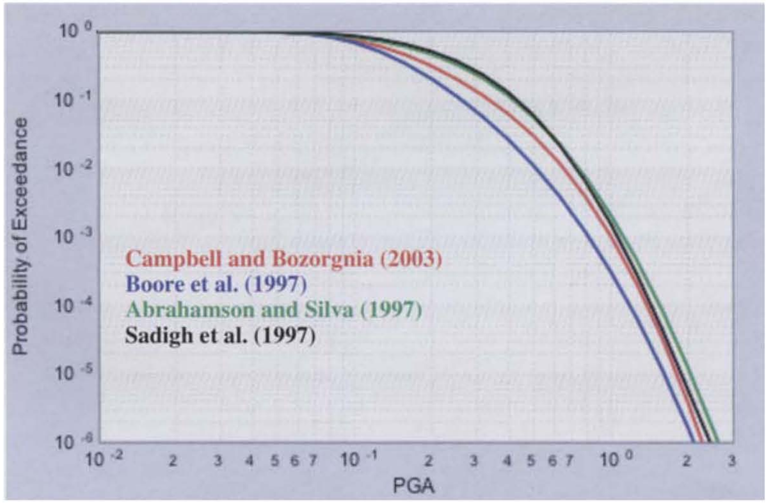
DISCUSSION

Our use of distributed-computing technologies to deploy the WGCEP-2002 model not only makes it accessible from multiple, platform-independent applications but also makes these applications relatively lightweight (because the WGCEP-2002 model never actually resides in the application, only its list of independent parameters). It also allows the WGCEP-2002 code to be run on a server that is generally much faster than the average personal computer.

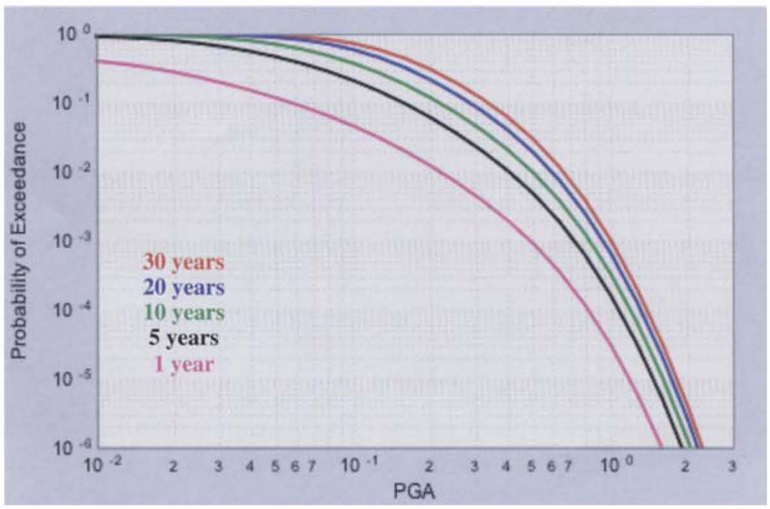
The examples shown here represent a small fraction of what can be done with the hazard-curve application. We have considered only one location, one site condition, one intensity measure type (PGA), and one ERF (WGCEP-2002), and we excluded the background seismicity in that ERF. Readers are invited to explore the many alternatives themselves. The important point is that we now have the tools needed to explore these options and, consequently, the ability to quantify SHA more fully. In particular, we can evaluate the influence of epistemic uncertainties as represented by different viable models, which in turn will enable us to focus on research topics needed to reduce those uncertainties and thereby improve our hazard estimates.

This and other OpenSHA applications are available at our Web site (http://www.OpenSHA.org/). Most of these are still in beta mode, meaning they should be reliable but may crash under certain circumstances. The plug-and-play flexibility allows for many possible model combinations. For example, in generating hazard curves like those above, the user has more than 20 different parameters that can be adjusted (from Gaussian truncation of the attenuation relationship, to setting a maximum source-site distance cut-off in the hazard curve calculation). Not all of the many possible parameter combinations can be validated separately. The most practical approach is to test a subset of combinations that maximizes some level of confidence. This is an ongoing process, the duration of which will depend on available resources. We believe our level of testing compares favorably to that of other available SHA codes, and our open-source approach ensures full disclosure and faster identification of software problems.

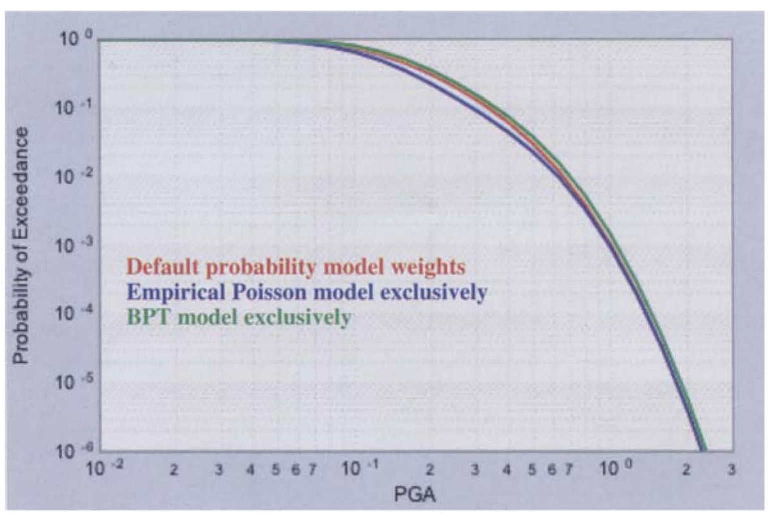
Finally, we wish to emphasize that all of the features discussed here, from utilizing an ERF list to deploying the



▲ **Figure 3.** Mean hazard curves as computed in Figure 1, but using the different attenuation relationships as labeled.



▲ **Figure 4.** Mean hazard curves as computed in Figure 1, but for the different forecast durations as labeled.



▲ **Figure 5.** Mean hazard curve as computed in Figure 1 (red), along with mean curves where the Empirical Poisson model and BPT models (blue and green, respectively) are given exclusive weight on all faults. Curves for the BPT-step and Poisson models being given exclusive weight are very close to that of the BPT model (not shown for clarity).

WGCEP-2002 model using distributed-computing technologies, are generic capabilities that can be used with other models. Furthermore, these distributed model components can exist anywhere on the Internet. We believe the computational capabilities presented in this paper are evidence of an emerging community-modeling environment, or collaboratory, for seismic hazard analysis.

ACKNOWLEDGMENTS

We would like to acknowledge very helpful reviews of this manuscript by Paul Reasenberg and Matt Gerstenberger. We are also grateful to Norm Abrahamson, Bruce Julian, and Paul Reasenberg for developing, debugging, and documenting the WGCEP-2002 FORTRAN code. This research was supported in part by the Southern California Earthquake Center (SCEC). SCEC is funded by NSF Cooperative Agreement EAR-0106924 and USGS Cooperative Agreement 02HQAG0008. The SCEC contribution number for this paper is 819.

REFERENCES

Abrahamson, N.A. (2000). Effects of rupture directivity on probabilistic seismic hazard analysis, *Proceedings of 6th International Conference on Seismic Zonation*, Palm Springs, Earthquake Engineering Research Institute.

Abrahamson, N. A. and W. Silva (1997). Empirical response spectral attenuation relations for shallow crustal earthquakes, *Seismological Research Letters* **68**, 94–127.

Boore, D. M., W. B. Joyner, and T. E. Fumal (1997). Equations for estimating horizontal response spectra and peak acceleration from western North American earthquakes: A summary of recent work, *Seismological Research Letters* **68**, 128–153.

Borcherdt, R. D. (1994). Estimates of site-dependent response spectra for design (methodology and justification), *Earthquake Spectra* **10**, 617–653.

Campbell, K. W. (1997). Empirical near-source attenuation relationships for horizontal and vertical components of peak ground acceleration, peak ground velocity, and pseudo-absolute acceleration response spectra, *Seismological Research Letters* **68**, 154–179.

Campbell, K. W. and Y. Bozorgnia (2003). Updated near-source ground motion (attenuation) relations for the horizontal and vertical components of peak ground acceleration and acceleration response spectra, *Bulletin of the Seismological Society of America* **93**, 314–331.

Field, E. H. (2000). A modified ground-motion attenuation relationship for southern California that accounts for detailed site classification and a basin-depth effect, *Bulletin of the Seismological Society of America* **90**, S209–S221.

Field, E. H., T. H. Jordan, and C. A. Cornell (2003). OpenSHA: A Developing community-modeling environment for seismic hazard analysis, *Seismological Research Letters* **74**, 406–419.

Field, E. H., H. A. Seligson, N. Gupta, V. Gupta, T. H. Jordan, and K. Campbell (2004). Probabilistic loss estimates for a Puente Hills blind-thrust earthquake in Los Angeles, California, *Earthquake Spectra* (submitted).

Frankel, A. D., C. S. Mueller, T. P. Barnhard, E. V. Leyendecker, R. L. Wesson, S. C. Harmsen, F. W. Klein, D. M. Perkins, N. C. Dickman, S. L. Hanson, and M. Hopper (2000). USGS National Seismic Hazard Maps, *Earthquake Spectra* **16**, 1–20.

Frankel, A. D., M. D. Petersen, C. S. Mueller, K. M. Haller, R. L. Wheeler, E. V. Leyendecker, R. L. Wesson, S. C. Harmsen, C. H. Cramer, D. M. Perkins, and K. S. Rukstales (2002). *Documentation for the 2002 Update of the National Seismic Hazard Maps*, U.S. Geological Survey Open-File Report 02-420.

Gerstenberger, M., S. Wiemer, and L. Jones (2004). *Real-time Forecasts of Tomorrow's Earthquake in California: A New Mapping Tool*, U.S. Geological Survey Open-File Report (in review).

Jordan, T. H., P. J Maechling, and the SCEC/CME Collaboration (2003). The SCEC community modeling environment: An information infrastructure for system-level science, *Seismological Research Letters* **74**, 324–328.

Maechling, P. J., V. Gupta, N. Gupta, E. H. Field, D. Okaya, and T. H. Jordan (2005). Distributed seismic hazard analysis research using distributed computing in the SCEC Community Modeling Environment, *Seismological Research Letters* **76**, 177–181.

SSHAC (Senior Seismic Hazard Analysis Committee) (1997). *Recommendations for Probabilistic Seismic Hazard Analysis: Guidance on Uncertainty and Use of Experts*, U.S. Nuclear Regulatory Commission, U.S. Dept. of Energy, Electric Power Research Institute; NUREG/CR-6372, UCRL-ID-122160, Vol. 1–2. Also a review of the document by National Academy Press, Washington, DC, 73 pp.

Sadigh, K., C. Y. Chang, J. A. Egan, F. Makdisi, and R. R. Youngs (1997). Attenuation relationships for shallow crustal earthquakes based on California strong motion data, *Seismological Research Letters* **68**, 180–189.

Spudich, P., W. B. Joyner, A. G. Lindh, D. M. Boore, B. M. Margaris, and J. B. Fletcher (1999). SEA99: A revised ground motion prediction relation for use in extensional tectonic regimes, *Bulletin of the Seismological Society of America* **89**, 1,156–1,170.

WGCEP (Working Group on California Earthquake Probabilities) (1988). *Probabilities of Large Earthquakes Occurring in California on the San Andreas Fault*, USGS Open-File Report 88-393.

WGCEP (Working Group on California Earthquake Probabilities) (1990). *Probabilities of Large Earthquakes in the San Francisco Bay Region, California*, USGS Circular 1053.

WGCEP (Working Group on California Earthquake Probabilities) (2003). *Earthquake Probabilities in the San Francisco Bay Region: 2002–2031*, USGS Open-File Report 03-214.

Wald, D. J. (2003). Written communication on the attenuation relationship used for the USGS scenario ShakeMaps archived at http://quake.wr.usgs.gov/research/strongmotion/effects/shake/archive/scenario.html.

Wills, C. J., M. Petersen, W. A. Bryant, M. Reichle, G. J. Saucedo, S. Tan, G. Taylor, and J. Treiman (2000). A site conditions map for California based on geology and shear wave velocity, *Bulletin of the Seismological Society of America* **90**, S187–S208.

U.S. Geological Survey
525 S. Wilson Avenue
Pasadena, CA 91106
Telephone: +1-626-583-7814
E-mail: **field@usgs.gov**
(E.H.F.)

Southern California Earthquake Center
University of Southern California
Los Angeles, CA 90089-0742
(N.G., V.G., P.M., T.H.J.)

U.S. Geological Survey
12201 Sunrise Valley Drive
Reston, VA 20192
(M.B.)