Ground Motion-Based Testing of Seismic Hazard Models in the USA

Final Technical Report
USGS Award G11AP20024

Mark Stirling & Matt Gerstenberger
GNS Science
PO Box 30368
Lower Hutt
New Zealand
+64 4 5704794
m.stirling@gns.cri.nz, m.gerstenberger@gns.cri.nz

Term of Award: 01/01/2011 – 12/31/2011

Abstract

We compare the observed rates of exceedance of 0.1g (peak ground acceleration) at strong motion accelerograph sites contained in the Next Generation Attenuation (NGA) Flatfile (peer.berkeley.edu/nga/flatfile.html) to the predicted rates of exceedance at the sites from the national probabilistic seismic hazard (PSH) model for the USA. There is evidence of over-prediction and significant under-prediction at various sites in the model; however, because of data limitations and consideration of all results together there is no strong indication that the model is inconsistent with the data. The results are largely limited to California and for a mean observation and prediction length of 32 years. Additionally, the results are preliminary and future improvements to our analysis should include an augmentation of the strong-motion dataset with data not contained in the NGA flatfile. Dependency issues should also be addressed within the site observations (i.e. the same earthquakes recorded at multiple stations).

Introduction

The past decade has seen considerable effort directed towards developing methods to test probabilistic seismic hazard (PSH) models. Efforts have either focussed on testing individual components of a PSH model (e.g. predicted versus observed magnitudes and rates of earthquakes; e.g. Schorlemmer et al., 2007), or testing the hazard estimates of the model against observation (e.g. Stirling & Petersen, 2006; Stirling & Gerstenberger, 2010). The former approach has the advantage of being able to identify issues associated with individual components of the PSH model, whereas the cause of a discrepancy between predicted and observed hazard estimates has to be "unravelled" from the PSH model by way of sensitivity tests and the like. However, the latter approach has the conceptual simplicity of testing the "bottom-line" (PSH model output), and one we have pursued under this USGS Award G11AP20024.

In this final report we compare the instrumental record of strong-motion data recorded over the past few decades in the lower 48 states to the USA National PSH model (Petersen et al. 2008). Our approach is the same as our New Zealand-based study of the New Zealand National PSH model (Stirling & Gerstenberger, 2010). The New Zealand comparison comprised a statistical analysis of the differences between the observed number of exceedances at each strong-motion station to the predicted number of exceedances of the same ground motion levels at the station sites (Peak ground acceleration, PGA levels of 0.1 and 0.2g). The New Zealand study showed the PSH model predictions and the strong motion data to generally be consistent, except where aftershocks contribute to the observed exceedances. In this case the New Zealand PSH model significantly under-predicted the observed number of exceedances for both 0.1 and 0.2g.

Award G11AP20024 Activities

Our activities associated with USGS Award G11AP20024 comprise the following:

1. Data compilation and preliminary analysis while visiting USGS Golden for two weeks in September 2011.
2. Lunchtime seminar to USGS Golden scientists on our work, and leading a discussion on the recent SRL Opinion paper "Bad assumptions or bad luck: Why earthquake hazard maps need objective testing" by Seth Stein, Robert Geller, and Mian Liu (SRL 82-5, 623-626).
3. Publication of SRL Opinion paper "Earthquake hazard maps and objective testing: the hazard mapper's point of view", by Stirling (SRL 83-2, 231-232; see Appendix).
4. Undertook revised analysis in March 2012 after addressing errors encountered in (1).

Methodology and Data

Our methodology is the same as that used in our earlier New Zealand-based study (Stirling & Gerstenberger, 2010). Specifically, we utilize the instrumental record of PGA levels experienced at strong motion accelerograph stations (hereafter referred to as "sites") in the USA to calculate the annual rate of exceedance for those PGA levels, and then compare these historically-based hazard curves to the hazard curves calculated for the sites from the USA PSH model (Petersen et al. 2008). The sites are those available to us from the PEER Next Generation Attenuation (NGA) flatfile (peer.berkeley.edu/nga/flatfile.html). This is the definitive strong motion database used to develop the NGA ground motion prediction equations. A total of 193 stations, 66 unique earthquakes, and 205 station recordings are used in our analysis. The data are exclusively located in the western USA as a consequence of the high earthquake activity there. Statistical analysis of our comparison is done on the basis of individual site data, and also for all site data together.

We obtain PGA data for the sites of 0.1g or greater from the NGA flatfile. PGA of 0.1g is our basis for selection for practical reasons as it represents the approximate lower limit of PGA considered important for engineering design (e.g. Reiter 1991). The choice of a higher level of PGA would also yield considerably fewer data for our analysis. The accelerographs have been recording for between three and 74 years (average of 32 years). A histogram of site recording times is shown in Figure 1. The hazard estimates used in our comparisons (predicted exceedances for the recording time of the site) are calculated according to the relevant site class for the site.

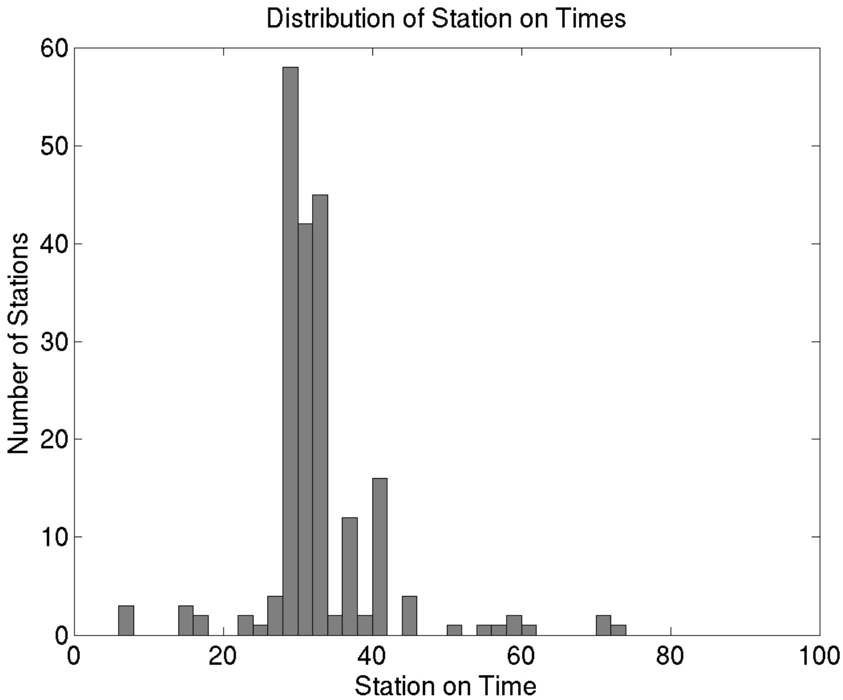

Figure 1. Histogram showing the time periods (in years) that the accelerograph stations (sites) have been operating.

Analysis and Results

The comparisons of observed to predicted number of exceedances are shown by the histograms in Figures 2 and 3. The histogram in Figure 2 shows the difference between the observed number of events in the time period of recording at each site and the number predicted by the PSH model for those sites and time periods. The comparison is repeated in terms of observed and predicted annual rates in Figure 3. The histograms show the observed and predicted numbers or rates to be similar in general, with the mode of the histogram slightly negative, i.e. observed slightly less than predicted. Also noticeable is the scatter of sites to the left of the histogram on Figure 2 and right on Figure 3, revealing the extent of departure of some observed exceedances from the predicted exceedances.

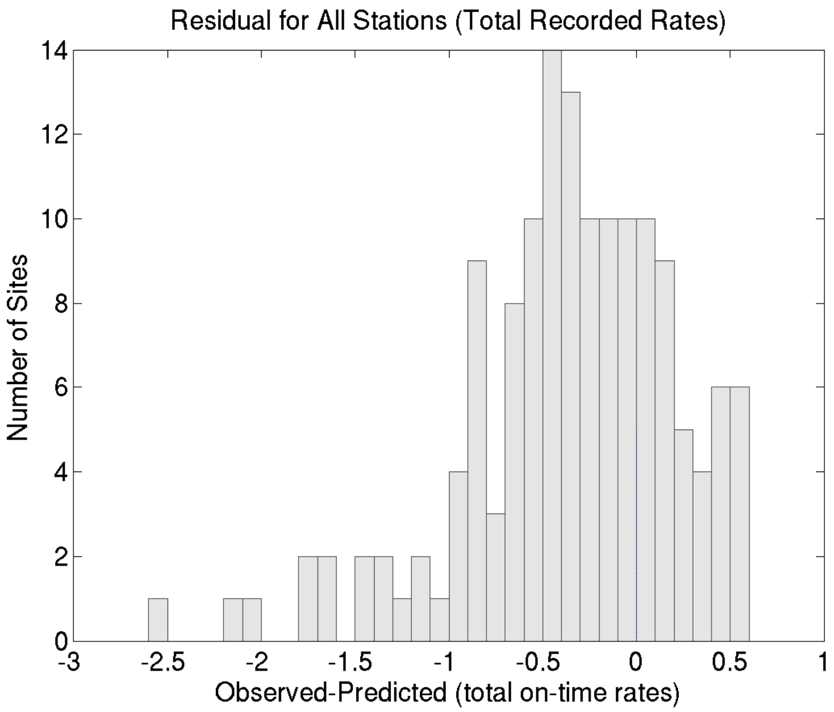

Figure 2. Histogram showing the difference between the observed and predicted number of exceedances of 0.1g at the sites for the time periods of recording at each site.

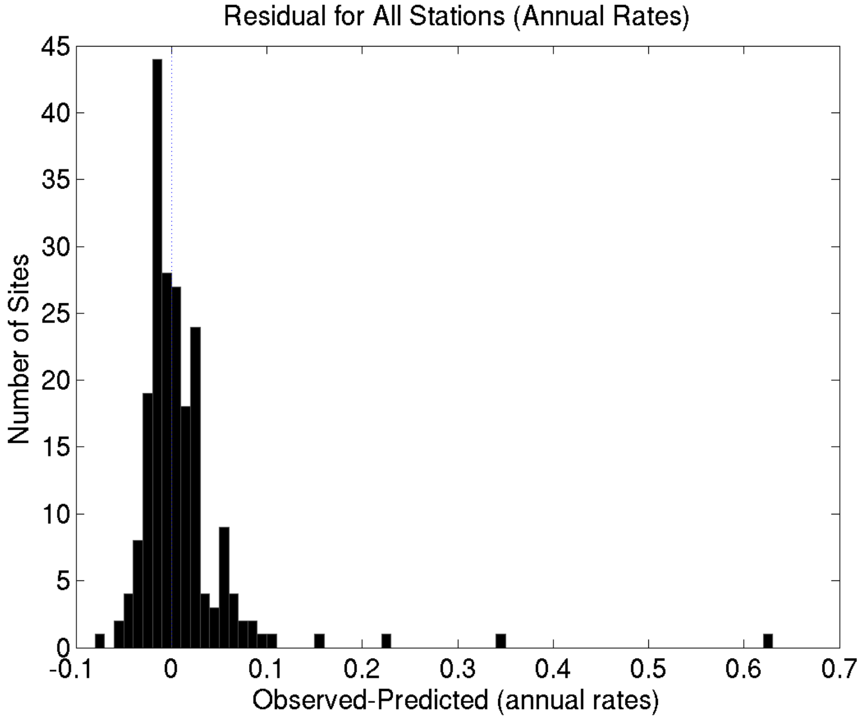

Figure 3. Histogram showing the difference between the observed and predicted rates of exceedance of 0.1g at the sites.

A quantitative comparison of the observed data to the PSH model is summarised in Figures 4 and 5. As in the study of Stirling & Gerstenberger (2010), we assume the predicted number of exceedances for each site is the mean of a Poisson distribution. From this we evaluate whether the PSH model forecast is consistent with the observed number of exceedances. This is a two sided test, and if the observation falls on either tail of the expected Poisson distribution then the observation may be considered unlikely given the prediction. A right-continuous Poisson distribution is used with a rejection level of 5%. Specifically, a comparison showing that observing at most the true observed number of exceedances has a probability of 0.025 or less indicates a significant over-prediction; a comparison showing that observing more than the true observed number of events (Observed – 1 for a right-continuous Poisson distribution, see Zechar et al. 2010 for details) has a probability of 0.975 or greater indicates a significant under-prediction. In both cases such an observation would result in a rejection of the site, or the total PSH model, with at least 95% confidence. Conversely, an observation within these bounds is not significantly inconsistent with the PSH model and the model will not be rejected. These probabilities are shown for all sites in Figures 4 and 5.

We see from Figures 4 and 5 that few site predictions are inconsistent with the observed data. In Figure 4 we see that in all cases the probabilities (P_{Lower}) are greater than 0.025, meaning that none of the observed numbers of exceedances are significantly less than the predicted number. In Figure 5, we see that that almost all of the probabilities (P_{Upper}) are less than 0.975. However, in this case, for five sites, the predicted numbers of exceedances are significantly less than the observations. Looking in detail at these sites we see that the observations are either from: 1) a swarm in the Mammoth region, or; 2) the Coalinga earthquake and its aftershock sequence. For both sequences the data contain recordings from multiple stations for the same earthquake, which may bias the result. Additionally, both the Mammoth sequence and the Coalinga aftershock sequence represent non-Poissonian behaviour, which the PSH model does not attempt to predict. For this reason it can be argued that these observations should be removed from the testing data set and the model will no longer significantly under-predict the observations at these sites. However, by leaving the data in the testing data set, we can gain a better understanding of where the deficiencies in the model are and where future improvements can be made.

An additional test was carried out on the total number of observed and predicted exceedances for all stations. Because of the variability in station recording times (Figure 1) we performed this test in two different ways. In the first test we compared the predicted rate to the observed rate for the station recording time. Calculating the test in this way biases it to the longer recording times. However it is the simplest test and most transparent way to test the model given the dataset. This test shows the total model to be consistent with the data (P_{Lower} = 0.64 and P_{Upper} = 0.74). In the alternate test we normalised the observations and predictions to the mean recording time (32 years). In this test the total model significantly under-predicts the data with a P_{Lower} of 0.99. The interpretation of this result is not so clear, but gives an indication that there is significant under-prediction in the model. However, this can be contrasted with the trend seen in both Figures 2 and 3 for the bulk of the sites to be over-predicted. Upon closer inspection of the results, it can be seen that only a few sites contribute to the significant under-prediction while many have a small (insignificant) over-prediction.

In summary, while there is some indication of over-prediction and significant under-prediction at various sites in the model, when all results are considered together there is no strong indication that the model is inconsistent with the data. Limitations of our analysis arise from the shortness of the recording time, with a mean of 32 years and the use of a PGA of 0.1g as a criterion for comparison. In other words, issues of non-stationarity of earthquake rates in the region (clustering, non-poissonian behaviour) will have influenced our analysis. It is worth noting that the observations at the majority of sites are insignificantly overpredicted (Fig. 3). While 32 years is a short time period, a longer period on the order of 50 years would not likely yield a significantly different result, given the large number of overpredicted sites. The 0.1g criterion also has an unclear relationship to the commonly-used hazard criteria (e.g. 10% in 50 years). Our observations may also be influenced by the dependencies of site observations (i.e. the same earthquakes recorded at multiple stations). While all of the above are likely to have contributed to the differences between the observed and predicted number of exceedances, we can still gain an understanding of the deficiencies in performance of the PSH model from our data set. Most of these explanations were invoked to explain discrepancies observed in the earlier New Zealand analysis (Stirling & Gerstenberger, 2010).

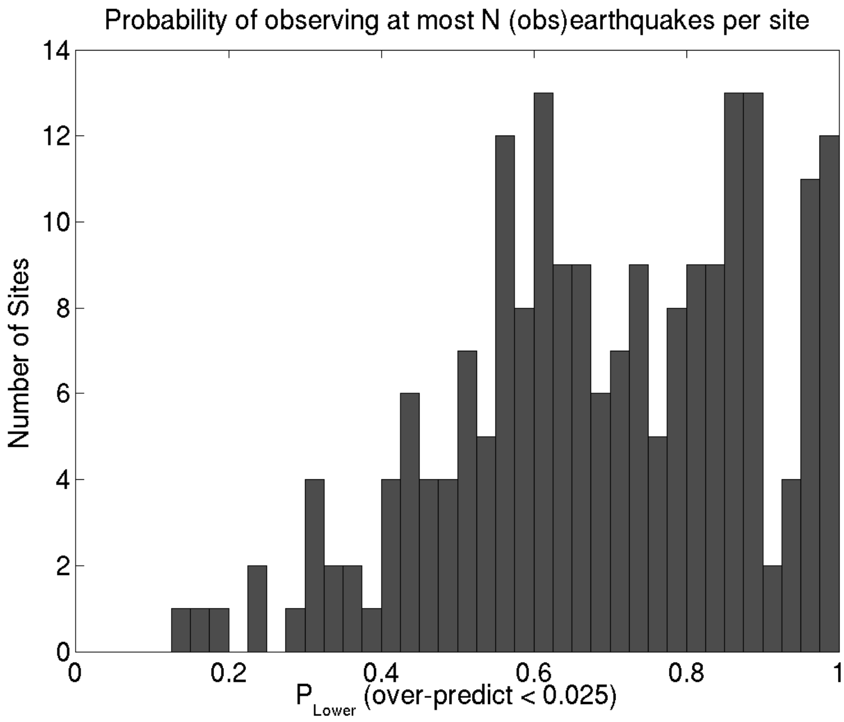

Figure 4. Histogram showing the distribution of P_{Lower}. See the text for explanation.

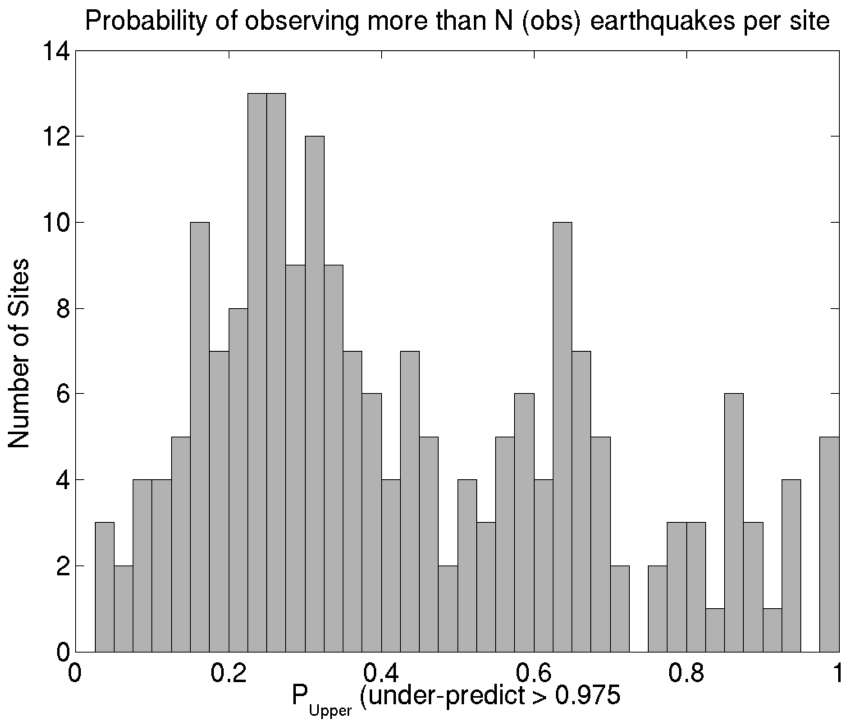

Figure 5. Histogram showing the distribution of P_{Upper}. See the text for explanation.

Conclusions

We have compared the observed rates of exceedance of 0.1g at strong motion accelerograph sites contained in the Next Generation Attenuation (NGA) Flatfile to the predicted rates of exceedance at the sites from the national PSH model for the USA. There is evidence of over-prediction and significant under-prediction at various sites in the model; however, because of data limitations and consideration of all results together there is no strong indication that the model is inconsistent with the data. These results are preliminary, in that future improvements to our analysis should include an augmentation of the strong-motion dataset with data not contained in the NGA flatfile, and dependency and clustering issues within the site observations should also be addressed. Finally, it should be kept in mind that the aim of such tests is to learn about deficiencies in the PSH model that can be improved.

References

Petersen, M.D., A.D. Frankel, S.C. Harmsen, C.S. Mueller, K.M. Haller, R.L. Wheeler, R.L. Wesson, Y. Zeng, O.S. Boyd, D.M. Perkins, N. Luco, E.H. Field, C.J. Wills, and K.S. Rukstales (2008). Documentation for the 2008 Update of the United States National Seismic Hazard maps, *U.S. Geol. Survey OFR 2008-1128*, pp. 60 and 11 appendices.

Reiter, L., (1991). Earthquake hazard analysis: Issues and Insights. 254 pp.

Schorlemmer, D., Gerstenberger,M., Wiemer, S., Jackson, D., and Rhoades. D. (2007). Earthquake Likelihood Model Testing. Seismological Research Letters, Vol. 78, No. 1, 17-29.

Stirling, M.W., Petersen, M.D. (2006). Comparison of the historical record of earthquake hazard to seismic hazard models for New Zealand and the continental United States. *Bulletin of the Seismological Society of America* 96, 1978-1994

Stirling, M.W. (2012). Earthquake Hazard Maps and Objective Testing: The Hazard Mapper's Point of View. Opinion. *Seismological Research Letters*, 83(2). 231-232.

Stirling, M.W., and Gerstenberger M. (2010). Ground motion-based testing of seismic hazard models in New Zealand. *Bulletin of the Seismological Society of America* 100, 1407-1414.

Stein, S., Geller, R., Liu, M. (2011). Bad assumptions or bad luck: Why earthquake hazard maps need objective testing. Opinion. *Seismological Research Letters* 82(5), 623-626.

Zecher, J., Gerstenberger, M., and Rhoades, D. (2010). Likelihood-Based Tests for Evaluating Space–Rate–Magnitude Earthquake Forecasts. *Bulletin of the Seismological Society of America* 100, 1184-1195.

OPINION

Earthquake Hazard Maps and Objective Testing: The Hazard Mapper's Point of View

The recent *SRL* Opinion article titled "Bad Assumptions or Bad Luck: Why Earthquake Hazard Maps Need Objective Testing," by Seth Stein, Robert Geller, and Mian Liu (*SRL* 82(5), 623–626) suggests that probabilistic seismic hazard (PSH) models have been inadequate in terms of forecasting recent devastating earthquakes, and expresses a need for objective testing of these models. This response comes from the perspective of a hazard mapper.

PSH MODELS: INTENDED USAGE

First, I'd like to clarify that the PSH models referred to by Stein and his colleagues are not forecasting tools. They are tools developed to provide estimates of hazard for long return periods (*e.g.*, hundreds to thousands of years) for engineering design and planning purposes. They are not developed to provide short-term (*e.g.*, months to years) probabilities for impending earthquakes. Examples of appropriate PSH model applications are in building construction (typically 500 to 2,500 year return periods) and in nuclear facility and hydro-dam developments (typically ≥10,000 year return periods). Hazard maps for these return periods typically show large differences in hazard across regions like the western United States and New Zealand, reflecting differences in the expected future activity of earthquake sources across the regions. The differences seem logical, given that one would expect sites close to major plate boundary faults to experience more earthquakes in the long term than sites further away. This is useful information for engineering and planning, including the development of loadings standards like the International Building Code. The PSH-derived hazard estimates can also be disaggregated to identify the most likely (or most unlikely) earthquake scenarios for the site or region in question, and these scenarios are often used by regional authorities and others to plan for future earthquake hazards. However, to give these PSH models the ability to provide actual short-term earthquake forecasts would require the integration of relevant forecasting models into the PSH model framework. Promising relevant efforts have been happening in California, New Zealand, and elsewhere, but it is still early days in terms of a substantial update to standard PSH methodology.

To give these PSH models the ability to provide actual short-term earthquake forecasts would require the integration of relevant forecasting models into the PSH model framework. Promising relevant efforts have been happening in California, New Zealand, and elsewhere, but it is still early days in terms of a substantial update to standard PSH methodology.

Stein *et al.* do raise some perfectly valid issues with regard to the performance of the relevant PSH models. While I have said that the models are not intended to be used as forecasting tools, it is true that model parameters like maximum magnitude and expected ground motions should adequately encompass any event observed in the particular region. In this respect it is true that Japanese PSH models underestimated the magnitude of the M_w 9, 11 March 2011 Tohoku earthquake. In New Zealand, the M_w 7.1, 4 September 2010 Darfield, Canterbury earthquake occurred on a previously unknown fault, reflecting a partial lack of knowledge about that part of New Zealand. However, the earthquake was to an extent accounted for in the distributed or background seismicity model, which has a maximum magnitude set at M_w 7.2 in the area of the earthquake. The main purpose of a distributed seismicity model is to allow for the occurrence of earthquakes on unknown sources, which is exactly what happened in New Zealand. Some modern PSH models have gone the extra step of incorporating comprehensive epistemic uncertainties into every component of the model to account for all possible surprise events. The Californian UCERF3 model, for instance, allows virtually every possible combination of rupture geometry on the fault sources and uses seismological and geodetic data to define a range of distributed seismicity rates.

PSH MODEL TESTING

Finally, it is helpful to report that research focused on the objective testing of PSH models has been progressing for some years. The Collaboratory for the Study of Earthquake Predictability (CSEP) has been developing testing strategies

doi: 10.1785/gssrl.83.2.231

and methods for a wide variety of applications, and collaborative work has also focused on developing ground motion–based tests of the New Zealand and U.S. national seismic hazard models. The Global Earthquake Model (GEM), a worldwide seismic hazard and loss modeling initiative, is including testing and evaluation as an integral part of the overall model development. The Yucca Mountain seismic hazard modeling efforts have developed an innovative "points in hazardspace" approach to considering all viable constraints on ground motions for long return periods. GEM and Yucca Mountain stand as examples of a more holistic approach to PSH modeling and are therefore examples of what needs to happen more widely in the future. It will, after all, be the holistic, versatile, and tested models that best stand the test of time.

GEM and Yucca Mountain stand as examples of a more holistic approach to PSH modeling and are therefore examples of what needs to happen more widely in the future.

Discussions with Pilar Villamor, Ned Field, Matt Gerstenberger, and Nicolas Pondard on this article were very helpful.

Mark W. Stirling
GNS Science
P.O. Box 30368
Lower Hutt, New Zealand
m.stirling@gns.cri.nz