An Empirical Model for Global Earthquake Fatality Estimation

Kishor Jaiswal,[a) M.EERI, and **David Wald,**[b) M.EERI

We analyzed mortality rates of earthquakes worldwide and developed a country/region-specific empirical model for earthquake fatality estimation within the U.S. Geological Survey's Prompt Assessment of Global Earthquakes for Response (PAGER) system. The earthquake fatality rate is defined as total killed divided by total population exposed at specific shaking intensity level. The total fatalities for a given earthquake are estimated by multiplying the number of people exposed at each shaking intensity level by the fatality rates for that level and then summing them at all relevant shaking intensities. The fatality rate is expressed in terms of a two-parameter lognormal cumulative distribution function of shaking intensity. The parameters are obtained for each country or a region by minimizing the residual error in hindcasting the total shaking-related deaths from earthquakes recorded between 1973 and 2007. A new global regionalization scheme is used to combine the fatality data across different countries with similar vulnerability traits. [DOI: 10.1193/1.3480331]

INTRODUCTION

Several loss estimation models and post-earthquake loss computation tools exist to estimate social (e.g., fatalities, injuries, and homelessness) and financial losses (both direct and indirect losses) resulting from strong earthquakes. These models typically use one of three general approaches to casualty estimation: empirical analytical, and hybrid (or semi-empirical). The regression-based empirical approach in general consists of performing statistical analysis on historical loss data using a chosen hazard-specific parameter (e.g., magnitude, intensity, population density) and deriving regression parameters that can be used for future loss estimation. The analytical approach involves a multi-step process consisting of seismic hazard analysis (estimating ground shaking in terms of peak ground motions or spectral response, and its uncertainty), exposure analysis (estimating human and economic exposure of the building stock), structural analysis (assessing structural response given the shaking hazard), damage analysis (estimating damage given the structural response), and loss analysis (estimating social and economic losses due to structural and nonstructural damage) (FEMA 2006). The hybrid (or semi-empirical) approach is generally a simplified analytical approach in which both structural response and damage analyses are combined by directly correlating structural damage or losses with macroseismic shaking intensity (Shiono et al. 1991, Murakami 1992, Shakhramanian et al. 2000, Jaiswal and Wald 2010).

[a) U.S. Geological Survey, Golden CO (contracted through Synergetics Incorporated), kjaiswal@usgs.gov
[b) U.S. Geological Survey, Golden CO

Earthquake Spectra, Volume 26, No. 4, pages 1017–1037, November 2010; © 2010, Earthquake Engineering Research Institute

Table 1. Fatality vs. exposure comparison for selected earthquakes worldwide

Country	Earthquake (Name, Date and Local Time)	Magnitude (M_w)	Population Exposure ($\geqslant$VII)	Fatalities
Iran	Bandar Abbas Earthquake Mar 21, 1977 (12.30 am)	6.7	2,700	167
	Bam Earthquake Dec 26, 2003 (5.26 am)	6.6	79,500	26,271
Indonesia	Kepulauan Alor Earthquake Nov 11, 2004 (5.26 am)	7.5	130,800	34
	Yogyakarta Earthquake 26 May 2006 (5.54 am)	6.4	4,434,500	5,749

Empirical modeling has been performed in a variety of different ways in the past depending upon the earthquake damage data available and assumptions about the correlative hazard. Some of the earliest casualty estimation studies were conducted by Japanese researchers such as Kawasumi (1951), who estimated a measure of earthquake danger and expectation of maximum intensity for large earthquakes in Japan. Similarly, early casualty estimation efforts in the United States were scenario-specific, based on estimation of the casualty rate per 100,000 people and use of engineering judgment (Algermissen et al. 1972). Ohta et al. (1983) developed an empirical relationship for estimating the number of casualties as a function of the number of completely destroyed houses and Oike (1991) proposed a relationship between earthquake magnitude and fatalities. A more recent attempt, based on an analysis of 450+ global earthquakes, obtains a log-linear relationship for fatalities as a function of magnitude and population density (Samardjieva and Badal 2002). Similarly, Nichols and Beavers (2003) studied a fatality catalog of the twentieth century and established a bounding function using the fatality count and the magnitude of an earthquake.

In general, the previously published casualty estimation models were derived using available data in hand for specific regions and by performing regression analysis using either single or multiple parameters. Most regression equations derived were dependent upon simplified hazard parameters, often just magnitude, so there was often a disconnect because the shaking and exposure characteristics were only implicit in the regression analyses. Also, most of the previously published casualty models were derived based on fatal earthquakes only and without validating them against the events that did not cause any fatalities.

Study of intensity-dependent population exposure suggests that earthquakes of similar size in the same country have caused uneven fatalities due to highly variable population exposure to shaking. Table 1 show that more people subjected to higher shaking intensities (e.g., intensity $\geqslant$ VII) in general causes increased number of casualties. Similarly, earthquake intensity dependent population exposure and the associated fatality rates drive the total number of deaths rather than earthquake's magnitude in itself. For example, the mortality rate of 6% and 33% in Iran and 0.026% and 0.13% in Indonesia, respectively, both cases indicating higher rates contrary to the magnitude of earthquake.

This observation is imperative in light of previously published empirical casualty estimation models wherein researchers (Samardjieva and Badal 2002, Nichols and Beavers 2003, Chen et al. 2005) have used earthquake magnitude as a parameter of regression analysis.

In this paper we present a new approach for calculating an earthquake fatality rate which is expressed as a function of macroseismic intensity. Macroseismic intensity is a spatially varying parameter and an indicator of direct impact of ground motion on the built environment. Magnitude-based empirical loss models are generally ineffective in capturing such spatial variability of exposure and associated losses unless they are derived for a specific region and for a unique seismogenic source.

In this retrospective study, we require (a) an account of total shaking-related fatalities for each earthquake and (b) an estimate of population exposed at each shaking intensity level during that earthquake.

Allen et al. (2009a) compiled a composite global catalog called PAGER-CAT that comprises over 140 fields including information pertaining to fatalities due to shaking and other secondary effects for earthquakes from the year 1900. By recreating ShakeMaps[3] for large earthquakes (magnitude 5.5 or greater in active tectonic regions and magnitude 4.5 in stable continental regions) that occurred since 1973 and overlaying them on global population maps developed by Oak Ridge National Laboratory (Landscan, Bhaduri et al. 2002), Allen et al. (2009b) developed an exposure catalog (referred as EXPO-CAT). This catalog contains estimated population exposure at discrete levels of instrumental shaking intensity at the time of earthquake obtained through correcting the present day population to the year of the earthquake by reversing country-specific population growth rates.

Both PAGER-CAT and EXPO-CAT catalogs provide essential data for the development of the PAGER empirical model. In the next section, we use those datasets to examine the variability of earthquake fatalities rates among the different countries. Our findings strongly suggest that there is a need to develop a country-specific empirical model.

EARTHQUAKE FATALITIES WORLDWIDE

Examination of global earthquake fatality data available through PAGER-CAT since 1900 shows that on a global scale, 76 percent of the total shaking-related deaths (not including deaths due to secondary effects such as fire, tsunami, liquefaction or landslide) were in China, Iran, Pakistan, and Turkey (Jaiswal et al. 2009). About 80 percent of the total shaking-related deaths since 1900 were due to only 25 earthquakes in 11 countries: China, Pakistan, Iran, Turkey, Italy, Chile, Armenia, Guatemala, India, Tajikistan, and Nepal. China experienced 122 fatal earthquakes since the year 1900 which claimed

---

[3] Refer to ShakeMap atlas development criteria at http://earthquake.usgs.gov/earthquakes/shakemap/atlas.php

604,000 lives. Seventy-five Iranian earthquakes caused a total of 161,000 fatalities, whereas Turkey experienced 64 fatal earthquakes that killed more than 85,000 people. In Indonesia, 62 fatal earthquakes have killed 11,000 people; more than 50 percent of these fatalities are attributed to the Yogyakarta earthquake of 26 May 2006, which caused 5,749 deaths. Pakistan has experienced the most devastating earthquake in recent times, the Kashmir earthquake of 2005, which killed more than 85,000 people. Countries such as Armenia, Nepal, Argentina, Romania, and Nicaragua have experienced very few deadly earthquakes, but the number of deaths in any single event is quite large compared to other countries. Although Japan and Taiwan have experienced 43 and 38 fatal earthquakes, respectively, the single deadliest earthquake in these countries claimed more than 80 and 40 percent of the total deaths since the beginning of the twentieth century, respectively. The United States has experienced 18 fatal earthquakes, but remarkably they caused only 270 deaths, averaging 15 deaths per event during the last 100 years. Large earthquakes in the twentieth century such as the 1906 San Francisco earthquake in California in the United States, the great Kanto earthquake of 1923 in Japan, and the 2004 Sumatra earthquake in Indonesia have claimed extremely high death tolls but mainly due to the non-shaking-related causes and hence are not included in the above analysis. The PAGER-CAT includes over 19,700 earthquakes for the period 1900–2007 out of which 882 were fatal (experienced one or more fatalities) and overall 78 countries worldwide have experienced these fatal earthquakes in that time period.

Earthquake fatality and exposure data of past earthquakes in general provide a useful basis for developing country-specific fatality rates that can be used for future earthquake fatality estimation. According to EXPO-CAT, many countries (roughly 30) have had a three or more deadly earthquakes recorded during 1973–2007. These earthquakes in addition to more frequent but nonfatal events can be used evaluate the fatality rates for future loss estimation. For the other countries, we provide a means of grouping regions of like-vulnerability in order to apply the empirical fatality estimation approach on a global basis.

OBJECTIVES

The objective of the PAGER empirical model is to estimate country-specific mean fatality rate as a function of macroseismic intensity (here, Modified Mercalli Intensity, MMI). Using this model, PAGER should be able to estimate total shaking-related fatalities in future earthquakes within an average of $\frac{1}{2}$ to 1 order of magnitude, with higher accuracy in highly fatal events. A procedure is necessary to quantify the total uncertainty associated with the fatality estimates and to portray both the expected value and uncertainty bounds effectively and efficiently for real-time applications.

Available PAGER calibration data include (a) population estimates at each macroseismic intensity level (as illustrated in Table 2) for historical earthquakes (5,600+ earthquakes both fatal and nonfatal that are listed in EXPO-CAT for the period 1973–2007), and (b) total shaking-related fatalities (also shown in Table 2) listed in PAGER-CAT catalog (which covers the period 1900–2007) for each earthquake that also appear in EXPO-CAT catalog.

Table 2. Snapshot of EXPO-CAT catalog showing population exposure estimates* at the time of earthquake at each MM intensity level and total recorded death for these selected Italian earthquakes taken from PAGER-CAT catalog

Country	Earthquake Date	MMI-V	MMI-VI	MMI-VII	MMI-VIII	MMI-IX+	Total Deaths
Italy	197605062000	17,460,864	1,246,533	228,060	79,406	41,275	965
Italy	197609150315	2,754,979	440,564	181,950	36,602	0	11
..	..	..	..	..	..	..	..
..	..	..	..	..	..	..	..
Italy	200411242259	1,313,135	161,735	51,217	0	0	0

*This information is publicly available through EXPO-CAT (version 2007-12) catalog which list the population exposure for the historical earthquakes occurred during the period 1973-2007.
Refer-http://earthquake.usgs.gov/research/data/pager/expocat.php

METHODOLOGY

In order to derive the fatality rate function one needs to compile the fatality rate statistics at each intensity level using the observations from past earthquakes. However, the fatality rates are rarely known or documented at specific shaking intensity levels in an earthquake and are difficult to compute directly given the limited understanding of seismic hazard, its spatial variation and population demographics. The indirect method of this paper consists of the development of fatality rate function in such a way that the total estimated deaths from all intensity levels matches the total recorded deaths for these earthquakes.

We need a functional form describing the fatality rate (i.e., population killed divided by total population exposed) which varies between 0 and 1.0, and is constantly increasing with increase of shaking intensity. Lognormal distributions have been widely used both in engineering applications for reliability analysis and more specifically for earthquake vulnerability/fragility analysis in the past (Singhal and Kiremidjian 1997, Yamaguchi and Yamazaki 2000, FEMA 2006). Yeo and Cornell (2003) parameterized the number of occupants killed at a given spectral acceleration (S_a) using a lognormal distribution. In the present investigation, the fatality rate (ν), which is solely a function of shaking intensity (S), is expressed in terms of a two-parameter lognormal cumulative distribution function as follows:

$$\nu(S_j) = \Phi\left[\frac{1}{\beta}\ln\left(\frac{S_j}{\theta}\right)\right] \tag{1}$$

where Φ is the standard normal cumulative distribution function, S_j is a set of discrete value of shaking intensity at level j (with cutoff at intensity V for $j=1$ and expressed in numeric values at 0.5 increments; for example, 5.0, 5.5,…9.0 intensity levels corresponds to the value of j as 1, 2, 3,…,9 respectively). The parameters θ and β corresponds to the mean and standard deviation of the natural logarithm of the intensity measure respectively (by definition, the logarithm of intensity measure is normally distributed). Let $P_i(S_j)$ denote an estimated population exposed to shaking intensity S at level j for an event i (also illustrated using Table 2). The expected number of fatalities E_i can be denoted as:

$$E_i = \sum_j \nu_i(S_j) P_i(S_j) \tag{2}$$

The fatality rate depends on the two free parameters, θ and β, of the cumulative distribution function of lognormal distribution. For each country or a geographic location k, if there are N historical fatal earthquakes, each event-specific fatality number could be used to determine the fatality rate by reconstructing the shaking intensity distribution for each earthquake and estimating population exposure within each intensity interval. If we suppose that O_i is the number of recorded deaths for an earthquake i, then we can determine the fatality rate in such a way that the residual error (that is, the error estimate between estimated and recorded deaths) is minimized. We can estimate the parameters of the distribution (Equation 1) by minimizing the $L2$ norm (square error) or G norm (log-residual error) as shown below:

$$\varepsilon_{1,k} = \sum_{i=1}^{N} (E_i - O_i)^2 \ \textit{or L2 norm} \tag{3a}$$

or

$$\varepsilon_{2,k} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} [\ln(E_i/O_i)]^2} \ \textit{or} \ \log - \textit{residual}(G)\textit{norm} \tag{3b}$$

The $L2$ norm provides a minimization such that in the case of high-fatality earthquakes, the squared differences tend to dominate the overall contribution of error. However, in the case of the G norm (Equation 3b), we take the squared natural logarithm of the ratio between the estimated and recorded deaths, which tends to reduce the contribution of high-fatality earthquakes in the total error term and generally better satisfies the fit to the more numerous low-fatality events. We need a norm that combines the advantages of both $L2$ and G norms and minimizes the misfit (through some trade-off) among low and high fatality earthquakes to estimate the parameters of the distribution function. For our study we use a combination of $L2$ and G norms as

$$\varepsilon_{3,k} = \ln\left(\sqrt{\frac{1}{N}\sum_{i=1}^{N} (E_i - O_i)^2}\right) + \sqrt{\frac{1}{N}\sum_{i=1}^{N} [\ln(E_i/O_i)]^2} \ \textit{or L2G norm} \tag{4}$$

We use a standard iterative search technique often called Nelder-Mead optimization available in MATLAB® (Ver., R2007a) for minimizing the objective function (Equation 4) with the two free parameters of the distribution function, θ and β. The combined-norm approach is simple and suitable for countries with at least three or more fatal earthquakes in the catalog, and thus it helps us obtain earthquake fatality rates for a large number of countries. Combining norms is commonplace in geophysical inverse problems, including jointly addressing error terms for both seismic amplitude and waveform

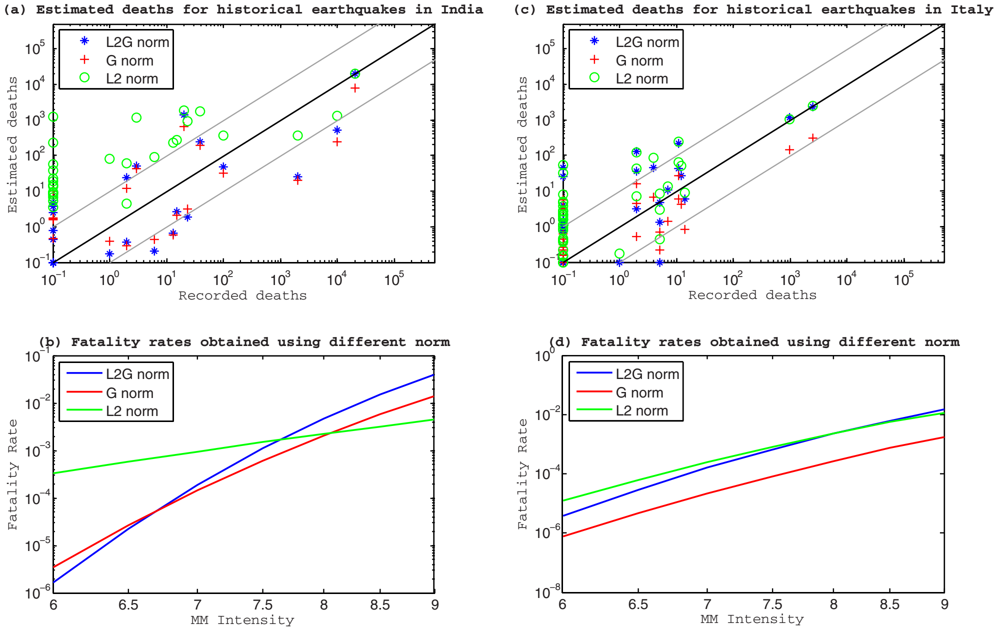

Figure 1. Empirical fatality model derived using historical (1973–2007) fatal earthquakes. Plots (a) and (c) show estimated and catalog recorded fatalities for India and Italy, respectively, obtained using the empirically derived fatalities rates shown in plot (b) and (d) derived using different norms. The dark lines shown in (a) and (c) are one-to-one estimates of fatalities, and light gray lines are one order of magnitude differences for estimated fatalities.

(e.g., Ji et al. 2002). Nevertheless the empirical approach presented herein is not limited to a specific norm or functional form assumed for fatality rate. The choice of the two-parameter lognormal cumulative distribution function and *L2G* minimizing norm is based on limited analysis as is demonstrated in the next section. Further, we measure the uncertainty associated in hindcasting the historical losses and use it in the forward sense for future loss prediction. In the following section, we discuss the global applicability of the proposed empirical model through a new regionalization scheme and then compare the fatality rates among the different countries or regions.

FATALITY MODEL ILLUSTRATION

We describe the development of empirical fatality rates using historical earthquakes in India and Italy as examples.

India: More than 150 large earthquakes have struck India since 1973, 28 of which were fatal and caused a total of 31,994 deaths. We show the development of an empirical model for India and a comparison of three norms namely *L2* norm (given by Equation 3a), *G* norm (given by Equation 3b), and a combination *L2G* norm (shown in Equation 4) in Figure 1a. As expected, the *L2* norm estimates the few deadliest earthquakes with higher accuracy than the *G* norm; the latter provides a better fit to the more numerous smaller events. Although the *L*2 norm estimates deadlier earthquakes better, it estimates on the order of 1,000 deaths for an earthquake that had no fatalities. The combination norm *L*2*G* provides a way to constrain both low- and high-fatality domains and suggests a model that can be used for future earthquake fatality estimates. The empirical fatality model with $\theta=11.53$ and $\beta=0.14$ for India indicates a rate of 1 death per 25 people exposed to shaking intensity IX and 1 death per 5,250 people exposed to intensity VII. The Bhuj earthquake of 2001 was one of the most devastating one in recent time in urban India and had caused widespread damage killing 20,023 people. It had an estimated population exposure of 212,000 at shaking intensity IX+, and about 982,600 at intensity VII which resulted in an estimated 20,337 deaths.

Italy: Forty-three earthquakes, of which 15 were fatal, were used to estimate the empirical model parameters for Italy. The largest earthquake that struck Italy since 1973 was the magnitude 6.9 Irpinia earthquake on Nov 23, 1980, which caused 2,483 deaths. The estimated population exposure was 37,200 people at shaking intensity IX and above and 250,180 at shaking intensity VIII. The fatality rate estimated using empirical model, with parameters $\theta=13.23$ and $\beta=0.18$, is shown in Figure 1b. The model corresponds to a fatality rate of one death per 68 people exposed to shaking intensity IX which reduced to one death per 6,310 people exposed at shaking intensity VII estimating 2,326 deaths for the Irpinia earthquake.

In general, the *L*2*G* norm works more like *G* norm for the case of India whereas it works more like *L*2 norm for the case of Italy. While generating the empirical fatality rate models for different countries, we found that the *L*2*G* norms fit the data well compared to other norms for respective country.

UNCERTAINTY ESTIMATION IN PREDICTED LOSSES

For a given country or region, we now know both the actual (i.e., recorded) deaths *O* for the historical earthquakes (from the PAGER-CAT earthquake catalog) as well as the empirical model-estimated deaths *E* using Equation 2 for these events (and shown in Figure 1). The objective here is to estimate the uncertainty associated with model's prediction for future earthquakes. We can study the overall dispersion associated with model's prediction for the past earthquakes in that country or region and then use such measure for determining the uncertainty associated with model's future estimates.

The logarithmic transformation of model-estimated and recorded losses for the regression helps to measure the linear scatter on a logarithmic scale. Let us transform the actual/recorded deaths for an earthquake *i* as $y_i=ln(O_i+0.5)$ and model-estimated deaths as $x_i=ln(E_i+0.5)$. A constant of 0.5 is added for all the observations to avoid negative infinity estimates when natural logarithm of zero is taken i.e., for instances of no fatalities. We can estimate the expected (or mean) value of actual loss $\mu_{\ln O|\ln E}$ by performing linear regression $\ln(O)=c+m\ln(E)$ on all earthquakes in that country or region. The regression parameters can be estimated using following.

$$c=\bar{y}-m\bar{x} \tag{5}$$

$$m = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \tag{6}$$

where the $\bar{x}$ & $\bar{y}$ are sample mean estimates obtained using both fatal and non-fatal earthquakes. Thus for any given earthquake with model-estimated deaths E, the regression now provides the mean or expected value of deaths.

Let ζ (called zeta) denote a variable (a measure of the variability or dispersion associated with actual deaths) representing normalized standard deviation of the logarithm of actual deaths given the logarithm of model-estimated deaths, which can vary by country. If this dispersion is constant between the areas of low and high losses in a scattergram for a given country, we can estimate this dispersion using following expression:

$$\zeta = \sqrt{\frac{1}{N-2}\sum_{i=1}^{N}\left[\ln(O_i) - \mu_{\ln O_i|\ln E_i}\right]^2} \tag{7}$$

If the conditional dispersion ζ varies between low and high losses, the regression would have to be obtained differently due to heteroscedasticity. A weighted or fuzzy regression or some standard method for reducing the heteroscedastic errors can be used.

We note that the error estimated in hindcasting the total shaking deaths using the empirical model already incorporates the variability that comes from (a) the uncertainty in estimated shaking hazard for each earthquake, (b) the uncertainty in the population exposure, and (c) possible (unnoticeable) errors in the number of recorded deaths in the catalog for these events. Variability in each of these inputs may have different effects depending upon the country under consideration (countries that experience frequent earthquakes vs. countries with relatively low seismic hazard, countries with high or low vulnerable building stock) or the nature of the constraints for shaking hazard estimates. Due to the lack of availability of fatality data by shaking intensity (most earthquake catalogs as shown in Table 2 contain only total reported deaths for any given earthquake), it is difficult to study the variability of estimated deaths at either each intensity level or within differing ranges of total fatalities (low fatal vs. high fatal earthquakes). For a limited set of earthquakes where such data could be gathered, the variability due to individual contributions (discussed above) can be studied, however this is beyond the scope of the present investigation.

REGIONALIZATION

In order to estimate the empirical fatality rate for countries with no or little fatality data, Jaiswal et al. (2009) proposed aggregation of fatal events at a regional level through a scheme that focuses on use of indicators that help associate countries with similar vulnerability. The regionalization scheme combines the information specific to geography, climatic similarities (based on the Köppen Climate scheme, discussed by Kottek et al. 2006), building inventory (Jaiswal and Wald 2008), and socioeconomic indicators defined using the Human Development Index (HDI) (UNDP 2008). The choices made in aggregating countries by these indicators, outlined below, are detailed and tabulated in Jaiswal et al. (2009).

HUMAN DEVELOPMENT INDEX

The Human Development Index (HDI) is an index measuring human development by combining indicators of life expectancy, educational attainment and income into a single statistic serving a frame of reference for both social and economic development for countries worldwide (http://hdr.undp.org/en/reports/global/hdr1990/). By setting a minimum and maximum for each dimensions called goalpost, the composite index provide a measure (a value between 0 and 1) of where each country stands in relation to these goalposts.

Socio-economic conditions affect the way people live and also tend to influence building construction and maintenance practices (refer Jaiswal et al, 2009 for further details). HDI provides a quick and combined measure of socio-economic conditions of human population around the globe. For example, the south central African countries such as Botsvana, Central African Republic, Chad, Congo, Ghana, Namibia, and others with HDI mostly between 0.3–0.6 were grouped with Yemen and Morocco (countries in the neighboring region) that have 0.56 and 0.65 respectively. We used this information qualitatively to group the countries together for development of regional fatality models. However, even with similar HDI indices, countries with dissimilar climates have different construction demands and require further consideration.

CLIMATE CLASSIFICATION

Climate is a considerable influence to the way people construct their homes. Since ancient times, building architecture characteristics (their configuration, sizes of opening) or their construction techniques (due to presence or absence of certain material) have been influenced by the local climate. In arid regions, due to diurnal swing in temperatures, buildings are generally constructed with flat roofs, small openings, and heavyweight materials without sufficient ductility due to lack of availability of timber. Similarly, a hot humid region often favors lightweight buildings which are breeze-penetrable but with pitched roof for wall shading. The size and position of openings in the walls also significantly affects the lateral load resistance capacity and are generally addressed in prevalent building construction standards and codes. The Köppen climate scheme (Kottek et al. 2006) is a recognized standard to demarcate the world according to their climate characteristics. We used the main climate characteristic as a reference to identify countries with similar climatic conditions when in agreement with other indicators. We identified and tabulated such commonalities (countries with similar climate, HDI and their most common building stock) and grouped the countries together accordingly. Jaiswal et al. (2009) provide a detailed discussion on HDI, climate classification, and building stock characteristics for each of the 20+ PAGER regions shown in Figure 2.

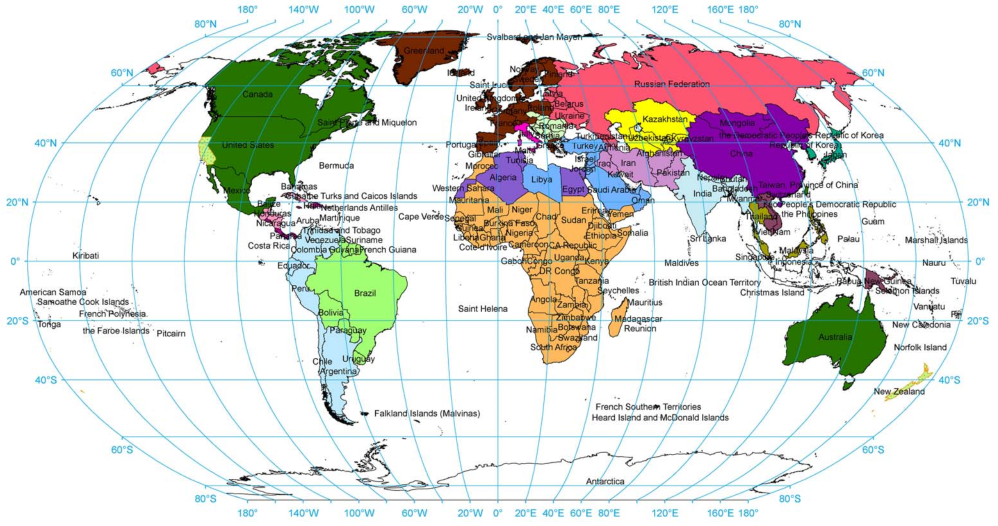

Figure 2. Proposed regionalization scheme for grouping the countries with like vulnerability traits.

EMPIRICAL MODEL PARAMETERS

Table 3 provides the empirical model parameters derived using Equation 1 and conditional standard deviation using Equation 5 for a selected country or a region using both fatal and non-fatal earthquakes. Interested readers are referred to Jaiswal et al. (2009) for detailed description of the estimated fatalities rates and their comparison for several other countries. This report also provides most recent version of empirical model parameters and its documentation as a part of Appendix II which can be downloaded at http://pubs.usgs.gov/of/2009/1136/.

Figure 3 demonstrates the fatality rate variation among the most and least vulnerable countries with a difference of several orders of magnitude between Iran and earthquake resistant regions of United States (California). As expected, the variation is larger at higher intensities between the most and least vulnerable countries and of constant order between countries that are expected to have comparable vulnerabilities.

Fatality rates below intensity VI are merely statistical extrapolation of rates obtained through functional form and are not used for fatality calculations. Similarly the fatality rates at intensity VI are subject to increased uncertainty given the uncertainties in estimating intensities (and their variabilities) during past and future events. At intensity IX, the model estimates one death for approximately 30,000 people exposed in California, and the rate increases to one fatality for every four person exposed at similar shaking intensity in Iran; thus, the total range of fatality rates globally approaches an astonishing four orders of magnitude. In Japan, the fatality rate corresponds to an estimated one death in every 330 people exposed at shaking intensity IX, it reduces to one death in every 20,100 at shaking intensity VIII and much lowest at lower intensities. Relatively

Table 3. Empirical model parameters derived using country-specific or using regionalization scheme by grouping earthquakes from several countries. The latest version of model parameters and documentation can be accessed at http://pubs.usgs.gov/of/2009/1136/.

Country	Total Events	θ	β	ζ	Region	Total Events	θ	β	ζ
Algeria	18	15.91	0.22	2.79	Brunei Darussalam, DPR. of Korea, R. Korea, Macao, Mongolia	119	10.40	0.10	2.03
Chile	26	40.93	0.44	1.90					
China	119	10.40	0.10	2.03					
Colombia	22	48.07	0.47	2.82					
El Salvador	7	26.62	0.32	2.17					
Georgia	7	26.49	0.33	0.99	Bahrain, Cyprus, Israel, Jordan, Kuwait, Lebanon, Libya, Oman, Palestine, Qatar, Saudi Arabia, UAE, Syria	87	11.05	0.10	1.99
Greece	30	21.48	0.28	1.92					
Guatemala	15	12.25	0.13	2.31					
India	28	11.53	0.14	2.28					
Indonesia	78	14.05	0.17	2.15					
Iran	93	9.58	0.10	2.60					
Italy	43	13.23	0.18	1.71	Bangladesh, Bhutan, Myanmar, Nepal, Sri Lanka	44	11.01	0.11	2.49
Japan	108	11.93	0.10	1.61					
Pakistan	23	9.71	0.10	2.62					
Peru	33	51.50	0.50	1.96	Hong Kong, Malaysia, Singapore, Thailand	32	16.04	0.18	1.85
Philippines	31	15.95	0.18	1.88					
Romania	6	17.50	0.24	2.16					
Taiwan	27	12.54	0.10	1.69	Armenia, Azerbaijan, Belarus, Estonia, Latvia, Lithuania, Russia, Ukraine	21	29.74	0.36	2.82
Turkey	81	10.97	0.10	1.95					
U.S.(California)	39	38.53	0.36	1.36					

high fatality rates at intensity IX in Japan are perhaps surprising but are understandable given the fact that 38% of the total dwelling stock is built before 1980s, this includes about 58% of traditional Japanese wood-frame buildings that suffered heavy losses during the 1995 Kobe earthquake (Jaiswal and Wald 2008). It is worth noting that due to the lack of sufficient large fatal earthquakes in eastern South America, former Soviet Union countries and Russian Federation, we tend to underestimate the overall seismic vulnerability of the region. The results obtained using the other two loss models implemented in the PAGER system may get preference for future earthquakes in this region.

The spatial variation of seismic hazard-independent mortality expressed as the number of fatalities expected per 1,000 people exposed at shaking intensity IX is illustrated in Figure 4. Clearly, future large earthquakes in populated areas of countries like Iran, Pakistan and other South Asian nations (shown as hot colors) will tend to produce the highest fatalities, whereas countries like the United States, Canada, and Australia (shown as cooler colors) remain less vulnerable, irrespective of their seismic hazard.

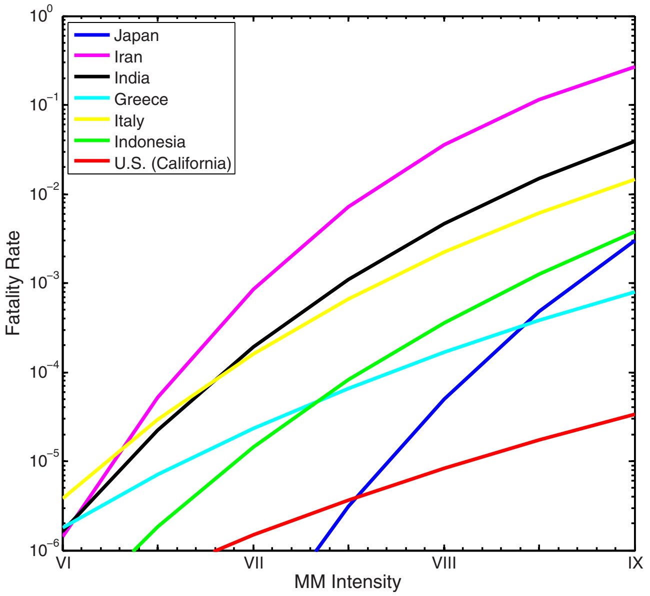

Figure 3. Fatality rates derived for selected countries using empirical model (v1.0 June, 2009).

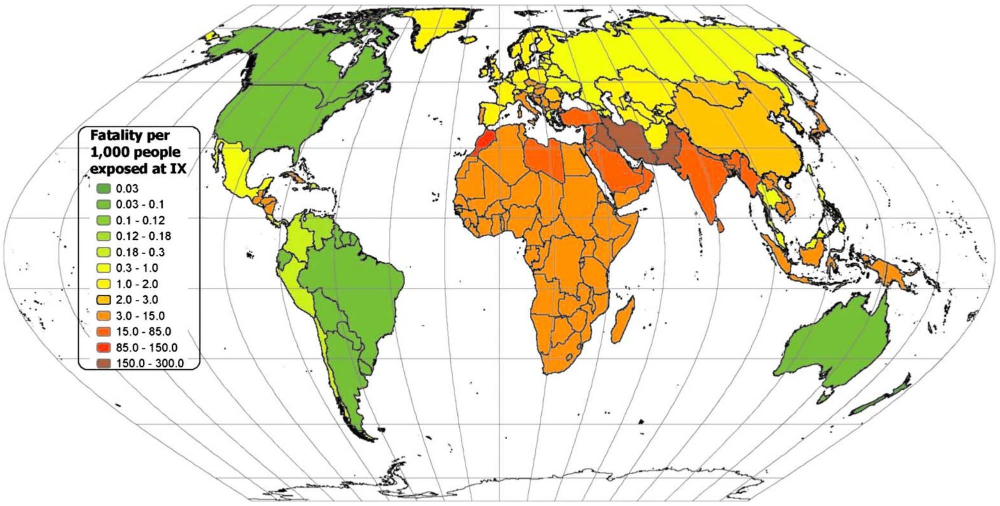

Figure 4. The PAGER empirical model showing earthquake fatalities estimated per 1,000 people exposed at MMI IX without any consideration of shaking hazard.

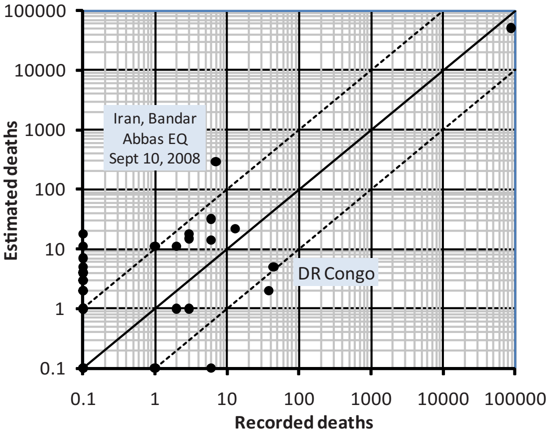

Figure 5. Fatality prediction using the empirical model for global earthquakes recorded in 2008. See text for details.

FATALITY ESTIMATION FOR RECENT EARTHQUAKES

The PAGER empirical model (v1.0) has been implemented within the lossPAGER (beta version) system since the beginning of 2008. Out of a total 139 earthquakes for which the automated PAGER system estimated losses, 100 were estimated to be nonfatal. Incidentally, there were no reported fatalities after these earthquakes (all are in the lower left corner of the log-log plot in Figure 5 as 0.1 deaths). Among the outliers, the model predicted fewer fatalities (~5) for an earthquake (M 5.4, 14 February 2008) in Democratic Republic of Congo (DRC) that killed 44 people (both in DRC and Rwanda) and overestimated fatalities (~295) for the Bandar Abbas earthquake of 10 September 2008 in southern Iran that killed 7 people. Nearly 80% of the total 139 earthquakes occurred in the countries for which country-specific models were available. Among the remaining events of which most occurred in Oceania, i.e., Pacific island countries, the group models were utilized to generate losses and the preliminary estimates were within an order of magnitude of the recorded deaths. The validation of other group models in general was limited as most of these countries were either non-seismic or they did not experience a sizable earthquake during the observation period.

For the most fatal earthquakes in the sample, the estimated deaths are within one order of magnitude of those observed. While in a strict predictive sense such a large ratio of predicted to observed deaths seems unimpressive, in reality the alerting level associated with such a prediction would be highly useful and on target for determining the appropriate level of response activation.

The empirical approach to fatality estimation is robust when used globally in comparison with other current loss models (the semi-empirical and analytical approaches), since it avoids the arduous requirements of detailing building inventory, vulnerability, indoor population exposure, and casualty rates based on specific damage states. However, the data and experience gained in developing the more physics-based models—particularly the building inventory and engineering-based vulnerability functions—help inform the empirical model development particularly as it pertains to constraining the empirical model in countries where loss data are lacking. The earthquake fatality rate functions consisting of two lognormal model parameters, an error term, the number of events, and its status (country or group) produced here for all the countries are available as an Excel spreadsheet (PAGER Implementation of Empirical Model.xls) as a part of appendix II of the USGS Open File Report (http://pubs.usgs.gov/of/2009/1136/).

FATALITY BASED ALERT

From a response perspective, the PAGER system proposes to produce fatality-based alerts with associated color-coded, logarithm-based estimated fatality domains: Green (no deaths); Yellow (1–100); Orange (100-1,000); Red (1,000–10,000+ fatalities). These alerts are meant to correspond to actionable responses, i.e., little to no response, local/regional, national, and international response mobilization, respectively. Wald et al. (2009, 2010) provides further details on specific choice of alert threshold governing into an actionable response, their colors or symbols while discussing an earthquake impact scale.

For an earthquake with expected (or mean) value of logarithm of fatality/loss estimate μ, the probability P that the actual death/loss d may be between predefined thresholds a and b, i.e., (a,b) is given as:

$$P(a < d \leqslant b) = \Phi\left[\frac{\ln(b) - \mu}{\zeta}\right] - \Phi\left[\frac{\ln(a) - \mu}{\zeta}\right] \tag{8}$$

The conditional dispersion ζ is obtained using Equation 7 which is unique for each country or region under consideration. The hypothesis that the uncertainty associated with expected value of logarithm of loss can be modeled using a normal distribution (with standard deviation ζ) is generally found to satisfy the Lilliefors test (Lilliefors 1967) at 5 percent significance level. Alternatively, the quantile estimates of deaths, i.e., the deaths D associated with different probability ranges $p \sim 10$, 50, and 90 percentile etc., can be represented as follows:

$$D = \exp[\zeta\Phi^{-1}(p) + \mu] \tag{9}$$

Following an earthquake, using model-estimated fatalities and expected value of loss, one can determine the relative likelihood of different PAGER alert thresholds using Equation 8. As discussed by Wald et al. (2010), the probabilities of estimated fatalities being in specified ranges remove undue attention on the median or mean value (which is not provided). In addition, the probability that the estimated deaths are within a particular alert threshold or in neighboring alert threshold can be provided to account for the uncertainty associated with PAGER alert.

Figure 6 provides the PAGER empirical loss model results obtained using v1.0 for the 2 September 2009 M7.0 Java earthquake in Indonesia consisting of a) a map showing population exposure overlain on shaking intensity contours, b) probabilistic assessment of fatality ranges, c) comparison of recorded vs. model-estimated fatalities for his-

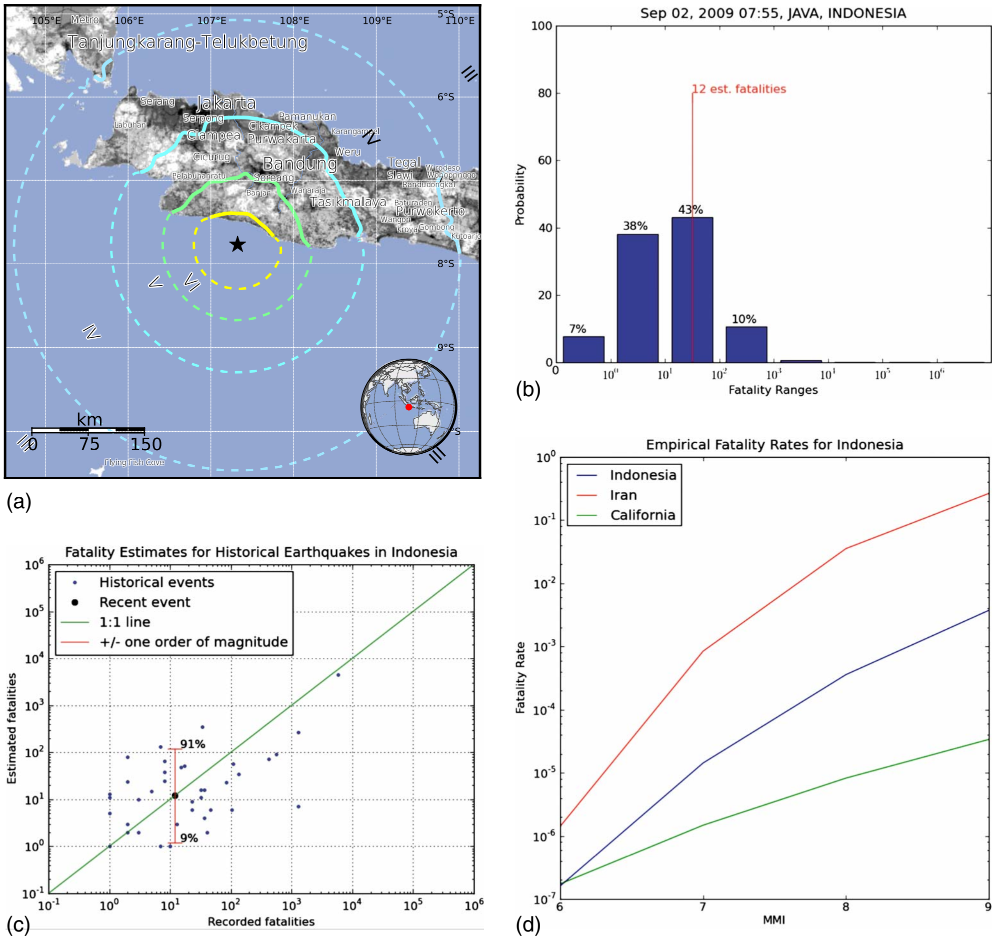

Figure 6. The automated empirical model-based fatality estimates for M7.0 Java Indonesia earthquake (7.77° S, 107.32° E) on 2 September 2009 at 07:55:01 UTC, created 50 minutes, 30 seconds after the earthquake (a) population exposure map, (b) median fatality estimate and the probabilistic ranges of fatality likelihood using v1.0 of empirical model (θ=14.05 β=0.17 and ζ =1.74), (c) recorded vs. estimated fatalities for past earthquakes, and (d) comparison of the country-specific fatality rate model with the least and most vulnerable countries worldwide.

torical earthquakes, and d) a plot showing comparison of the Indonesian fatality rates to those from the lowest and highest fatality rates determined globally in our model (California and Iran). In this, the empirical model-estimated fatalities are assumed to be median estimate (with 50% occurrence probability) and the uncertainty parameter ζ was estimated by measuring the scatter along the one-to-one line shown in Figure 6c. According to the empirical model results generated within minutes following this earthquake, there was a 43% probability that the earthquake caused between 10–99 fatalities (indicating yellow alert), and there was total 91% probability that it would be less than or equal to 100. There were 79 fatalities reported during the weeks following this earthquake and at least 37 of them were reported to be due to landslide (http://en.wikipedia.org/wiki/2009_West_Java_earthquake).

Quick assessments from the PAGER empirical model along with estimations from different models are vital for response agencies in the early hours after an earthquake, answering critical questions such as whether a particular earthquake requires the response and if so, at what level (local, regional, national, or international). With the addition of the empirical model to the PAGER system, we have also proposed a fatality-based alert scale which will provide an estimate of the likelihood of a range of fatalities caused by an earthquake in an effort to facilitate rapid response activities (Wald et al. 2010).

DISCUSSION

The PAGER system produces ShakeMaps shortly after all global earthquakes ($M \geqslant 5.5$), and estimates the population exposed to various levels of shaking intensities (Wald et al., 2008). In order to estimate the total number of fatalities from any given earthquake, we needed an estimate of the country or region-specific fatality rate as a function of shaking intensity. The model development consisted of estimating an empirical fatality rate (defined by two-parameter lognormal cumulative distribution function) for each country that minimizes the difference between the estimated deaths and the total recorded shaking-related deaths for each earthquake in the catalog using a numerical optimization technique. The estimated deaths are obtained using the assumed fatality rate model (bounded between MMI V–IX), earthquake-specific ShakeMap intensities, and the population exposure at different shaking intensities.

The main purpose of the regionalization scheme was to develop a fatality vulnerability model considering fatality data at a regional scale rather than at the country level. Countries that have sufficient fatal earthquakes maintain their own country-specific fatality model, and their historical earthquakes are utilized in developing a regional model that can be used for countries with no or few fatal earthquakes from 1973–2007. Further investigations are necessary in order to validate the applicability of regional empirical models for countries where there are no or few fatal earthquakes, particular where vulnerability assessment is difficult. Improvements are possible to the regionalization scheme proposed here, for example, separating South Africa from the African country group, Haiti from the group of Caribbean countries, Russia from the group of former Soviet countries, and differentiating between northern and southern China or northern versus central and southern India. However, such refinements must be informed by data other than earthquake fatalities alone, since these data are limited as it is and further subgrouping further reduces the data set. Efforts made to derive country-specific building inventories and vulnerability functions as part of the semi-empirical and analytical PAGER loss models (See Jaiswal and Wald 2008; Jaiswal and Wald 2010, Porter 2009 for details) have also helped inform grouping countries of like-vulnerability to fill gaps in the empirically-based fatality model for loss estimates worldwide, and these efforts will continue.

The present empirical approach does have certain deficiencies, specifically (1) it is country or region-specific, a resolution too coarse for some countries, (2) the approach does not quantify the errors associated with reported fatalities available through PAGER-CAT, errors associated with LandScan population database, and (3) uncertainties in estimating the population exposed per intensity level. Similarly the fatality rate is poorly constrained due to paucity of data in the countries with low-seismic hazard or countries that have only a few fatal earthquakes despite higher hazard. In order to account for this, the PAGER system is also informed by other loss models for casualty/loss estimation. Some of these uncertainties are not quantifiable without taking a more physical approach (which in turn entails additional uncertainties and limitation as discussed prior). For example, the evolution of building code and its adaptation or dramatic changes in building construction practices can cause considerable change in the earthquake fatality rates.

We envision that the addition of more recent earthquake observations (both nonfatal and fatal) and the refinements to hazard-specific constraints (for example, in terms of macroseismic intensities, selection of appropriate ground motion prediction equations, or new ground motion to intensity conversion relations) will help further refine the empirical fatality rates globally. This new information may require re-creation of ShakeMaps of past fatal earthquakes and recalibration of the proposed empirical-model parameters. In order to include such changes, we plan to regularly update the empirical model parameters and also make it available at the USGS website (http://pubs.usgs.gov/of/2009/1136/), as new data become available.

CONCLUSIONS

Based on our studies of global earthquake fatality data (1973–2007), we propose a new approach for estimating earthquake fatalities worldwide. A two-parameter lognormal distribution is used to express country-specific mean fatality rates as a function of MMI, without reference to other earthquake parameters (i.e., magnitude, location, or time of day). We also presented a comparison of fatality estimations based on the empirical model with the actual recorded fatalities for recent (2008) earthquakes and found a reasonable agreement for most of the events. Further investigations are necessary to estimate the uncertainty associated with fatality rate model (i.e., model parameters θ and β), to constrain the fatality rate given new observations through Bayesian updating, and to evaluate heteroschedastic errors, if possible, on loss estimates from the model.

Rapid fatality estimation using an empirical model provides an important opportunity to quickly and approximately assess the disaster potential of any earthquake worldwide. Empirical model estimates are crucial for most countries since it is very difficult to compile the building inventory, vulnerability, and time-dependent population exposures within each building class to the degree of accuracy needed for physics-based structural vulnerability models. Along with the other two candidate models (semi-empirical and analytical), the empirical model can be used not only as a post-earthquake rapid loss estimation tool within PAGER system but also as a model for computing losses from scenario earthquakes, thereby facilitating pre-disaster planning for major potential earthquakes anywhere in the world.

ACKNOWLEDGMENTS

The development of EXPO-CAT and PAGER-CAT database was particularly motivated and primarily relevant to this investigation. The investigation has greatly benefited from stimulating discussions and comments from Paul Earle and Keith Porter. We appreciate the constructive comments and suggestions made by Jim Dewey and David Perkins that helped improve the manuscript. Thanks to Hyeuk Ryu, Kristin Marano and Margaret Hopper for their grammatical review and to M. Hearne who performed the software implementation of the empirical model in the PAGER system for real-time loss computation. The constructive comments from four anonymous reviewers have also greatly benefited this manuscript.

REFERENCES

Algermissen, S. T., Rinehart, W. A., Dewey, J. W., Steinbrugge, K. V., Lagorio, H. J., Degenkolb, H. J., Cluff, L. S., McClure, F. E., Scott, S. N., and Gordon, R. F., 1972. *A Study of Earthquake Losses in the San Fransisco Bay Area: Data and Analysis*, prepared for the Federal Disaster Assistance Administration, Department of Housing and Urban Development, 220 pp.

Allen, T. I., Marano, K., Earle, P. S., and Wald, D. J., 2009a. PAGER-CAT: A composite earthquake catalog for calibrating global fatality models, *Seismol. Res. Lett.* **80**, 57–62.

Allen, T. I., Wald, D. J., Earle, P. S., Marano, K. D., Hotovec, A. J., Lin, K., and Hearne, M., 2009b. An atlas of ShakeMaps and population exposure catalog for earthquake loss modeling, *Bull. Earthquake Eng.* **7**, 701–718.

Bhaduri, B., Bright, E., Coleman, P., and Dobson, J., 2002. LandScan—Locating people is what matters, *Geoinformatics* **5**, 34–37.

Chen, Q-F., Mi, H., and Huang, J., 2005. A simplified approach to earthquake risk in mainland China, *Pure Appl. Geophys.* **162**, 1255–1269.

Federal Emergency Management Agency (FEMA), 2006. *HAZUS-MH MR2 Technical Manual*. Washington, D. C., http://www.fema.gov/plan/prevent/hazus/hz_manuals.shtm (22 Aug 2008).

Jaiswal, K., and Wald, D. J., 2008. *Creating a Global Building Inventory for Earthquake Loss Assessment and Risk Management*, USGS Open File Report, OF 2008-1160, 103 pp., (http://pubs.usgs.gov/of/2008/1160/).

Jaiswal, K., Wald, D. J., and Hearne, M., 2009. *Estimating Casualties for Large Earthquakes Worldwide Using an Empirical Approach*, USGS Open File Report, OF 2009-1136, 83 pp., (http://pubs.usgs.gov/of/2009/1136/).

Jaiswal, K., and Wald, D. J., 2010. Development of a semi-empirical loss model within the USGS Prompt Assessment of Global Earthquakes for Response (PAGER) System, in *Proc. 9th US National and 10th Canadian Conference on Earthquake Engineering*, Paper No. 1095, July 25–29, Toronto, Canada.

Ji, C., Wald, D. J., and Helmberger, D. V., 2002. Source description of the 1999 Hector Mine, California, earthquake; Part I: Wavelet domain inversion theory and resolution analysis, *Bull. Seismol. Soc. Am.* **92**, 1192–1207.

Kawasumi, H., 1951. Measures of earthquake danger and the expectancy of maximum intensity throughout Japan, as inferred from the seismic activity in historical times, *Bull. Earthquake*

Res. Inst. **29**, 469–482.

Kottek, M., Grieser, J., Beck, C., Rudolf, B., and Rubel, F., 2006. World map of the Koppen-Geiger climate classification updated, *Meteorol. Z.* **15**, 259–263.

Lilliefors, H., 1967. On the Kolmogorov-Smirnov test for normality with mean and variance unknown, *J. Am. Stat. Assoc.* **62**, 399–402.

Murakami, H. O., 1992. A simulation model to estimate human loss for occupants of collapsed buildings in an earthquake, in *Proc. Tenth World Conference on Earthquake Engineering*, Madrid, Spain, 5969–5974.

Nichols, J. M., and Beavers, J. E., 2003. Development and calibration of an earthquake fatality function, *Earthquake Spectra* **19**, 605–633.

Ohta, Y., Goto, N., and Ohashi, H., 1983. An empirical construction of equations for estimating number of victims by earthquakes, *Zisin II* **36**, 463–466.

Oike, K., 1991. A discussion on the relation between magnitude and number of deaths by earthquakes, in *Proc. of the Int. Seminar on Earthquake and Hazard Mitigation Technology*, Tsukuba, Japan, 333–341.

Porter, K., 2009. Cracking an open safe: HAZUS vulnerability functions in terms of structure-independent spectral acceleration, *Earthquake Spectra* **25**, 361–378.

Samardjieva, E., and Badal, J., 2002. Estimation of the expected number of casualties caused by strong earthquakes, *Bull. Seismol. Soc. Am.* **92**, 2310–2322.

Shakhramanian, M. A., Larionov, V. I., Nigmetov, G. M., and Sutschev, S. P., 2000. *Assessment of the Seismic Risk and Forecasting Consequences of Earthquakes While Solving Problems on Population Rescue (Theory and Practice)*, Russian Civil Defense and Disaster Management Research Institute, Moscow, 180 pp.

Shiono, K., Krimgold, F., and Ohta, Y., 1991. A method for the estimation of earthquake fatalities and its applicability to the global macro-zonation of human casualty risk, in *Proc. Fourth International Conference on Seismic Zonation*, Stanford, CA, III, 277–284.

Singhal, A., and Kiremidjian, A. S., 1997. *A Method for Earthquake Motion-Damage Relationships with Application to Reinforced Concrete Frames*, Technical Report *NCEER 97-0008*, Buffalo, NY: National Center for Earthquake Engineering Research, 218 pp.

United Nations Development Program (UNDP), 2008. *Human Development Indices: A Statistical Update 2008* at http://hdr.undp.org/en/mediacentre/news/title,15493,en.html (last visited Aug 26, 2009).

Wald, D. J., Earle, P. S., Allen, T. I., Jaiswal, K. S., Porter, K. A., and Hearne, M., 2008. Development of the U. S. Geological Survey's PAGER system (Prompt Assessment of Global Earthquakes for Response), in *Proc. 14th World Conference on Earthquake Engineering*, October 2008, Beijing, China, Paper No. 10-0008.

Wald, D. J., Jaiswal, K. S., Marano, K., and Bausch, D., 2009. Developing casualty and impact alert protocols based on the USGS Prompt Assessment of Global Earthquakes for Response (PAGER) system. *Second International Workshop on Disaster*, 15–16 June 2009, Cambridge, UK.

Wald, D. J., Jaiswal, K. S., Marano, K., and Bausch, D., 2010. An Earthquake Impact Scale, Natural Hazards Review, in review.

Yamaguchi, N., and Yamazaki, F., 2000. Fragility curves for buildings in Japan based on damage surveys after the 1995 Kobe earthquake. *Proc. 12th World Conference on Earthquake Engineering*, New Zealand, January, 2000, Paper No. 2451.

Yeo, G. L., and Cornell, C. A., 2003. Building-specific seismic fatality estimation methodology, in *Proc. 9th International Conference on Applications of Statistics and Probability in Civil Engineering* (ICASP9), San Francisco, CA, July 2003.

(Received 27 August 2009; accepted 2 April 2010)